MODERN ANTENNA HANDBOOK

Edited by

Constantine A. Balanis

WILEY

A JOHN WILEY & SONS, INC., PUBLICATION

MODERN ANTENNA HANDBOOK

MODERN ANTENNA HANDBOOK

Edited by

Constantine A. Balanis

WILEY

A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc.
Published simultaneously in Canada

Library of Congress Cataloging-in-Publication Data:

Balanis, Constantine A., 1938–
Modern antenna handbook / Constantine A. Balanis.
p. cm.
Includes index.
ISBN 978-0-470-03634-1 (cloth)
1. Antennas (Electronics) I. Title.
TK7871.6.B354 2008
621.382′4—dc22
2007050162

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

To my teachers, students, colleagues, relatives and friends

CONTENTS

PREFACE

Antennas and antenna systems are the eyes and ears of wireless communication systems, which have experienced an unprecedented rapid expansion. Part of this advancement has been attributed to contributions of antenna technology. These wireless systems, no matter how simple or complex, cannot operate efficiently unless they utilize transmitting and receiving elements/antennas to efficiently radiate and receive the waves that carry the information. This is analogous to humans whose daily schedule and contributions are largely controlled by the efficiency of their eyes and ears.

The future of the communication systems is even more challenging, and their efficacy will depend on what we, as antenna engineers and scientists, can invent and contribute. In fact, some of the future services and performances of wireless communication may be dependant on and limited by antenna designs which will require our imagination and vision to push the outer limits of the laws of physics. For example, handheld mobile units, which in 2008 numbered nearly 1 billion subscribers, are ubiquitous from the smallest rural village to the largest urban city and provide numerous services, such as voice, video, email, news, weather, stock quotes, GPS, TV, satellite, wireless LAN, Bluetooth, WiFi, WiMax, Radio Frequency ID (RFID), and so on. In the very near future, the number of services will expand and the number of units will increase at an almost exponential rate as everyone, from the youngest to the oldest around the world, will possess one. To integrate all these services into a single unit, and to provide them reliably and efficiently, we must develop broadband antenna designs to accommodate these demands, while at the same time provide esthetic visualization of the devices. Using single antenna elements for each of these services will not "cut the mustard." Therefore we are faced with unique challenges, and we must respond.

To meet the demands of this rapidly evolving technology, the *Modern Antenna Handbook* provides leadership towards these challenges, as seen by some of our visionary contributors. The handbook covers a wide range of topics, from the fundamental parameters of antennas to antennas for mobile wireless communications and medical applications. The information is practical in nature and is intended to be useful to practicing engineers and scientists, researchers, educators, consultants and students. There are a plethora of antenna elements and all could not be covered with some depth in a book with space constraints. Also, many antenna elements and designs are seen to be static and have been covered in other publications. We opted to include some of the classic antenna topics, to make the handbook self contained, and to focus more on contemporary topics and modern technologies, including wireless communication, materials and structures, techniques and computational methods relating to antennas. The 33 included chapters are all new, contemporary, provide breadth and depth, and are detailed. The chapters have been authored by 68 leading international experts who have contributed to the unprecedented advancement of antenna technology, and include coverage of current and future antenna-related vision.

Computational methods, which have been instrumental in the rapid evolutionary advancement of the modeling, simulation, animation, and visualization of antenna design, are an integral part of this handbook. In the past few years, Computational Electromagnetic Methods (CEM) have basically served as the working "slide rule" of the 1950s and 1960s. There is hardly any modern and efficient antenna design whose operation has not been influenced, modeled, simulated, and optimized using one or more computational methods. We expect that CEM will play an even bigger role in the future.

In addition to antenna topics, the handbook also covers modern and contemporary technologies such as metamaterials, micro-electromechanical systems (MEMS), frequency selective surfaces (FSS), radar cross section (RCS) and their applications to antennas.

The handbook is basically subdivided into six main themes.

PART I : FUNDAMENTAL PARAMETERS AND DEFINITIONS OF ANTENNAS

This part introduces the fundamental parameters and definitions of antennas. It is probably one of the most important chapters because it prepares the reader for the understanding of the various topics and parameters that are used throughout the handbook.

PART II : ANTENNA ELEMENTS

Some of the most basic and classic antenna configurations, from the wire-type to low-profile and apertures, such as monopoles, dipoles, loops, microstrips, apertures, reflectors, frequency independent, leaky-wave, reconfigurable, wideband, traveling, small, and fractal elements are included in this second part.

PART III: ARRAYS AND SYNTHESIS METHODS

Arrays are probably the most versatile antenna designs with a wide range of applications from ground-based to space-borne. Three chapters have been devoted to this theme, including smart antennas and adaptive arrays which we expect to play a key role in the advancement of the wireless communication technology. Some unique and intriguing designs of arrays for wideband operation are part of this section. Synthesis methods, although classic, will always be part of any antenna technology advancement.

PART IV: STRUCTURES AND TECHNIQUES RELATED TO ANTENNAS

Advanced materials and structures, such as metamaterials and artificial magnetic conductors, are modern technologies which already have been contributing factors to the evolution of antennas. In addition, MEMS, FSS, and RCS are contemporary devices and designs which are now integral parts of antenna technology. Measurement techniques, both conventional and near-to-far field, are also included.

PART V: ANTENNA APPLICATIONS

The applications emphasized in this theme are more toward wireless communications, including mobile units, base stations, multiple-input multiple-output (MIMO) techniques, and medical and biomedical applications. These will play a pivotal role in the advancement of wireless communication system efficiency, in terms of increasing capacity and range.

PART VI: METHODS OF ANALYSIS, MODELING AND SIMULATION

Five chapters are devoted to advanced numerical/computational methods targeted primarily for the analysis, modeling, simulation, design, animation and visualization of antennas. These methods have been an integral part of modern antenna technology, and include Integral Equations/Method of Moments, Finite-Difference Time-Domain (FDTD), Finite Element Method (FEM), Genetic Algorithms (GA) and Neural Networks.

There are many people that have contributed to the completion of this publication, and I would like to extend my sincere appreciation for their interest, contributions, cooperation, and support. First of all, I would like to thank the 68 authors, all of national and international acclaim, who volunteered to contribute to this handbook. The contents reflect their present and future vision of antenna technology. The interest and support from Wiley-Interscience was very critical from the genesis to the final production of this handbook. I would like to acknowledge the vision, dedication for excellence, and support provided within Wiley-Interscience by George Telecki (Editor) and the assistance from Rachel Witmer, Melissa Valentine, and the expert production led by Lisa Morano Van Horn.

Last, but not list, I would like to thank my family, Helen, Renie and Stephanie, for their patience, support, and encouragement over the years for all my challenging undertakings, including this one.

CONSTANTINE A. BALANIS

Tempe, Arizona,
May 2008

CONTRIBUTORS

Dr. Marco A. Antoniades, The Edwards S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ontario, Canada

Alejandro Aragón-Zavala, Department of Electronic Engineering, Technologic of Monterey, Cangus Querétano, Mexico

Professor Constantine A. Balanis, Department of Electrical Engineering, Arizona State University, Tempe, Arizona

Jodie Bell, Hawaii Center for Advanced Communications, University of Hawaii at Manoa, Honolulu, Hawaii

Professor Paolo Bernardi, Department of Electronic Engineering, University of Rome "La Sapienza," Rome, Italy

Professor Jennifer T. Bernhard, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois

Dr. Steven R. Best, MITRE, Bedford, Massachusetts

Dr. Trevor S. Bird, CSIRO ICT Centre, Sydney, New South Wales, Australia

Craig R. Birtcher, Department of Electrical Engineering, Arizona State University, Tempe, Arizona

Professor Marta Cavagnaro, Department of Electronic Engineering, University of Rome "La Sapienza," Rome, Italy

Nuri Celik, Hawaii Center for Advanced Communications, University of Hawaii at Manoa, Honolulu, Hawaii

Professor Christos G. Christodoulou, Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, New Mexico

William F. Croswell, formerly of Harris Corporation, Melbourne, Florida

Dr. Thomas Cwik, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California

Tim Durham, Harris Corporation, Melbourne, Florida

Professor George V. Eleftheriades, The Edwards S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ontario, Canada

A. Ross Forsyth, CSIRO ICT Centre, Sydney, New South Wales, Australia; now with Australia Telescope National Facility, Sydney, New South Wales, Australia

Michael H. Francis, National Institute of Standards and Technology, Boulder, Colorado

Paul Friederich, School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, Georgia

Professor Kyohei Fujimoto, University of Tsukuba, Tsukuba, Japan

Professor Cynthia M. Furse, Department of Electrical and Computer Engineering, University of Utah, Salt Lake City, Utah

Professor Om P. Ghandi, Department of Electrical and Computer Engineering, University of Utah, Salt Lake City, Utah

Dr. Christophe Granet, CSIRO ICT Centre, Sydney, New South Wales, Australia; now with BAE Systems Australia Ltd., Sydney, New South Wales, Australia

Jerzy Guterman, Institute of Telecommunications, Lisbon, Portugal

Dr. Randy L. Haupt, Applied Research Laboratory, The Pennsylvania State University, State College, Pennsylvania

Dr. John Huang, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California

Professor Gregory H. Huff, Department of Electrical and Computer Engineering, Texas A&M University, College Station, Texas

Dr. William A. Imbriale, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California

Professor Magdy F. Iskander, Hawaii Center for Advanced Communications, University of Hawaii at Manoa, Honolulu, Hawaii

Dr. Graeme L. James, CSIRO ICT Centre, Sydney New South Wales, Australia

Professor David R. Jackson, Department of Electrical and Computer Engineering, University of Houston, Houston, Texas

Professor Michael A. Jensen, Department of Electrical and Computer Engineering, Brigham Young University, Provo, Utah

Professor Jian-Ming Jin, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois

Mark Jones, Harris Corporation, Melbourne, Florida; now with Ansoft, Pittsburgh, Pennsylvania

Dr. Oren B. Kesler, Denmar Technical Services, Willis, Texas

Wayne Kim, Hawaii Center for Advanced Communications, University of Hawaii at Manoa, Honolulu, Hawaii

Professor Gianluca Lazzi, Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, North Carolina

Stanislav Licul, Maxtena, Inc., Blacksburg, Virginia

Professor James C. Lin, Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, Illinois

Dr. Zheng Lou, Department of Electrical and Computer Engineering, University of Illinois of Urbana-Champaign, Urbana, Illinois

Dr. James G. Maloney, Georgia Tech Research Institute, Georgia Institute of Technology, Atlanta, Georgia

Professor A. A. Moreira, Institute of Telecommunications, Lisbon, Portugal

Professor Hisamatsu Nakano, Department of Electronic Informatics, Faculty of Engineering, Hosei University, Koganei, Tokyo

Professor Sima Noghanian, Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, Manitoba, Canada

Dr. Arthur A. Oliner, University Professor Emeritus, Department of Electrical Engineering, Polytechnic University, Brooklyn, New York

Dr. Bo Pan, Wionics Research, Real Tek Semiconductor Group, Irvine, California

Professor John Papapolymerou, School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, Georgia

Douglas Pasquan, Lockheed Martin Corporation, Bethesda, Maryland

Professor Amalendu Patnaik, Department of Electronics and Computer Engineering, Indian Institute of Technology-Roorkee, Roorkee, India

Professor C. Peixeiro, Institute of Telecommunications, Lisbon, Portugal

Larry Pellett, Lockheed Martin Aeronautics, Palmdale, California

Professor Andrew F. Peterson, School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, Georgia

Professor Stefano Pisa, Department of Electronic Engineering, University of Rome "La Sapienza," Rome, Italy

Dr. Emanuele Piuzzi, Department of Electronic Engineering, University of Rome "La Sapienza," Rome, Italy

Professor Yahya Rahmat-Samii, Department of Electrical Engineering, University of California at Los Angeles, Los Angeles, California

Dr. Douglas Riley, Northrup Grumman, Space Technology, Redondo Beach, California

Dr. Norma Riley, Northrup Grumman Space Technology, Redondo Beach, California

Dr. Simon R. Saunders, Real Wireless Limited, West Sussex, United Kingdom

Professor Daniel Schaubert, Antennas and Propagation Laboratory, Electrical and Computer Engineering, University of Massachusetts, Amherst, Massachusetts

Professor Lotfollah Shafai, Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, Manitoba, Canada

Dr. Daniel F. Sievenpiper, HRL Laboratories, LLC, Malibu, California

Professor Glenn S. Smith, School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, Georgia

Professor Warren Stutzman, Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, Virginia

Professor Manos M. Tentzeris, School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, Georgia

Professor George V. Tsoulos, Department of Telecommunications Science and Technology, University of Peloponnese, Tripoli, Greece

Jon W. Wallace, School of Engineering and Science, Jacobs University, Bremen, Germany

Ronald C. Wittmann, National Institute for Standards and Technology, Boulder, Colorado

Professor Zhengqing Yun, Hawaii Center for Advanced Communications, University of Hawaii at Manoa, Honolulu, Hawaii

PART I

INTRODUCTION

CHAPTER 1

Fundamental Parameters and Definitions for Antennas

CONSTANTINE A. BALANIS

1.1 INTRODUCTION

To describe the performance of an antenna, definitions of various parameters are necessary. Some of the parameters are interrelated and not all of them need be specified for complete description of the antenna performance. Parameter definitions are given in this chapter. Many of those in quotation marks are from the *IEEE Standard Definitions of Terms for Antennas* (IEEE Std 145-1983).[†] This is a revision of the IEEE Std 145-1973. A more detailed discussion can be found in Ref. 1.

1.2 RADIATION PATTERN

An antenna *radiation pattern* or *antenna pattern* is defined as "a mathematical function or a graphical representation of the radiation properties of the antenna as a function of space coordinates. In most cases, the radiation pattern is determined in the far-field region and is represented as a function of the directional coordinates. Radiation properties include power flux density, radiation intensity, field strength, directivity, phase, or polarization." The radiation property of most concern is the two- or three-dimensional spatial distribution of radiated energy as a function of the observer's position along a path or surface of constant radius. A convenient set of coordinates is shown in Figure 1.1. A trace of the received electric (magnetic) field at a constant radius is called the amplitude *field pattern*. On the other hand, a graph of the spatial variation of the power density along a constant radius is called an amplitude *power pattern*.

Often the *field* and *power* patterns are normalized with respect to their maximum value, yielding *normalized field* and *power patterns*. Also, the power pattern is usually plotted on a logarithmic scale or more commonly in decibels (dB). This scale is usually desirable because a logarithmic scale can accentuate in more detail those parts of the

[†]*IEEE Transactions on Antennas and Propagation*, Vol. AP-17, No. 3, May 1969; Vol. AP-22, No. 1, January 1974; and Vol. AP-31, No. 6, Part II, November 1983.

Modern Antenna Handbook. Edited by Constantine A. Balanis

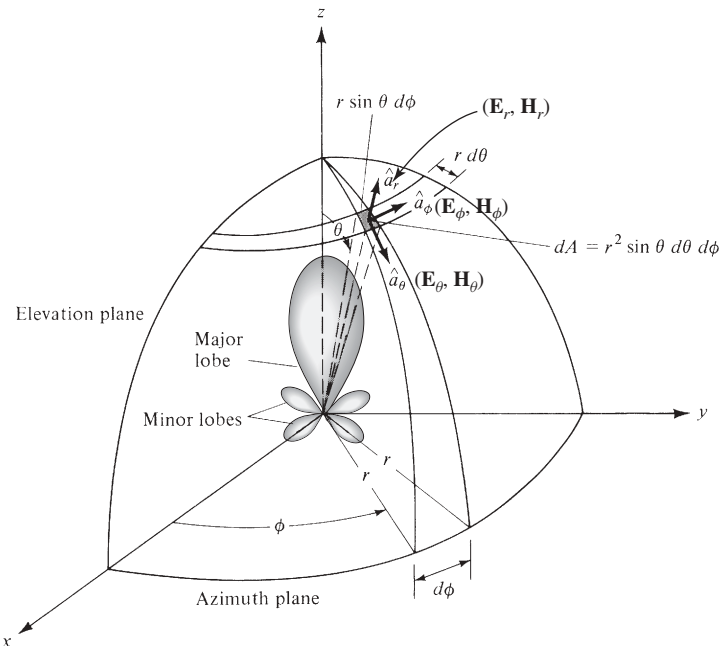
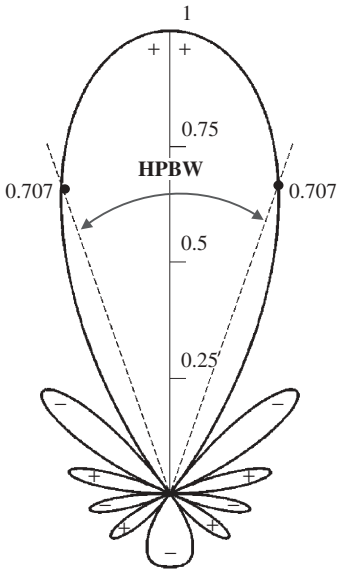


Figure 1.1 Coordinate system for antenna analysis.

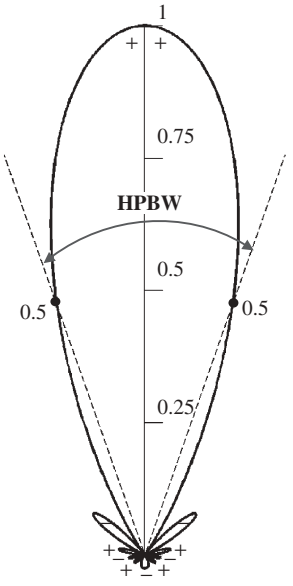
pattern that have very low values, which later we refer to as minor lobes. For an antenna, (1) the *field* pattern in (*linear scale*) typically represents a plot of the magnitude of the electric or magnetic field as a function of the angular space; (2) the *power* pattern in (*linear scale*) typically represents a plot of the square of the magnitude of the electric or magnetic field as a function of the angular space; and (3) the *power* pattern in (*dB*) represents the magnitude of the electric or magnetic field, in decibels, as a function of the angular space.

To demonstrate this, the two-dimensional normalized field pattern (*plotted in linear scale*), power pattern (*plotted in linear scale*), and power pattern (*plotted on a logarithmic dB scale*) of a 10-element linear antenna array of isotropic sources, with a spacing of $d = 0.25\lambda$ between the elements, are shown in Figure 1.2. *In this and subsequent patterns, the plus* (+) *and minus* (−) *signs in the lobes indicate the relative polarization of the amplitude between the various lobes, which changes (alternates) as the nulls are crossed*. To find the points where the pattern achieves its half-power (−3 dB points), relative to the maximum value of the pattern, you set the value of (1) *the field pattern at 0.707 value of its maximum, as shown in Figure 1.2a;* (2) *the power pattern (in a linear scale) at its 0.5 value of its maximum, as shown in Figure 1.2b; and* (3) *the power pattern (in dB) at −3 dB value of its maximum, as shown in Figure 1.2c*. All three patterns yield the same angular separation between the two half-power points, 38.64°, on their respective patterns, *referred to as HPBW* and illustrated in Figure 1.2. This is discussed in detail in Section 1.5.

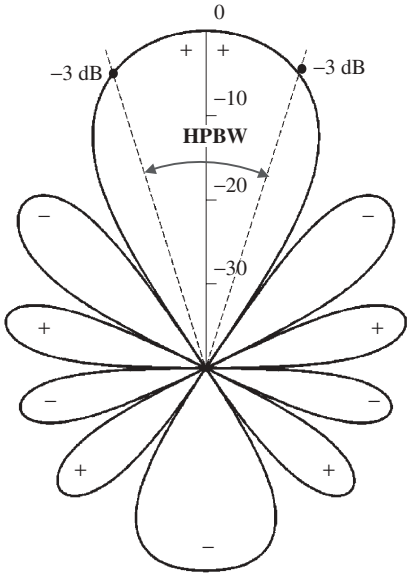
In practice, the three-dimensional pattern is measured and recorded in a series of two-dimensional patterns. However, for most practical applications, a few plots of the pattern as a function of θ for some particular values of ϕ, plus a few plots as a function of ϕ for some particular values of θ, give most of the useful and needed information.



(a) Field pattern (*in linear scale*)



(b) Power pattern (*in linear scale*)



(c) Power pattern (*in dB*)

Figure 1.2 Two-dimensional normalized *field* pattern (*linear scale*), *power* pattern (*linear scale*), and *power* pattern (in *dB*) of a 10-element linear array with a spacing of $d = 0.25\lambda$.

1.2.1 Radiation Pattern Lobes

Various parts of a radiation pattern are referred to as *lobes*, which may be subclassified into *major* or *main, minor, side*, and *back* lobes.

A *radiation lobe* is a "portion of the radiation pattern bounded by regions of relatively weak radiation intensity." Figure 1.3a demonstrates a symmetrical three-dimensional polar pattern with a number of radiation lobes. Some are of greater radiation intensity than others, but all are classified as lobes. Figure 1.3b illustrates a linear two-dimensional pattern (one plane of Figure 1.3a) where the same pattern characteristics are indicated.

MATLAB-based computer programs, designated as *polar* and *spherical*, have been developed and are included in the CD of [1]. These programs can be used to plot the two-dimensional patterns, both polar and semipolar (in *linear* and *dB scales*), in polar form and spherical three-dimensional patterns (in *linear* and *dB scales*). A description

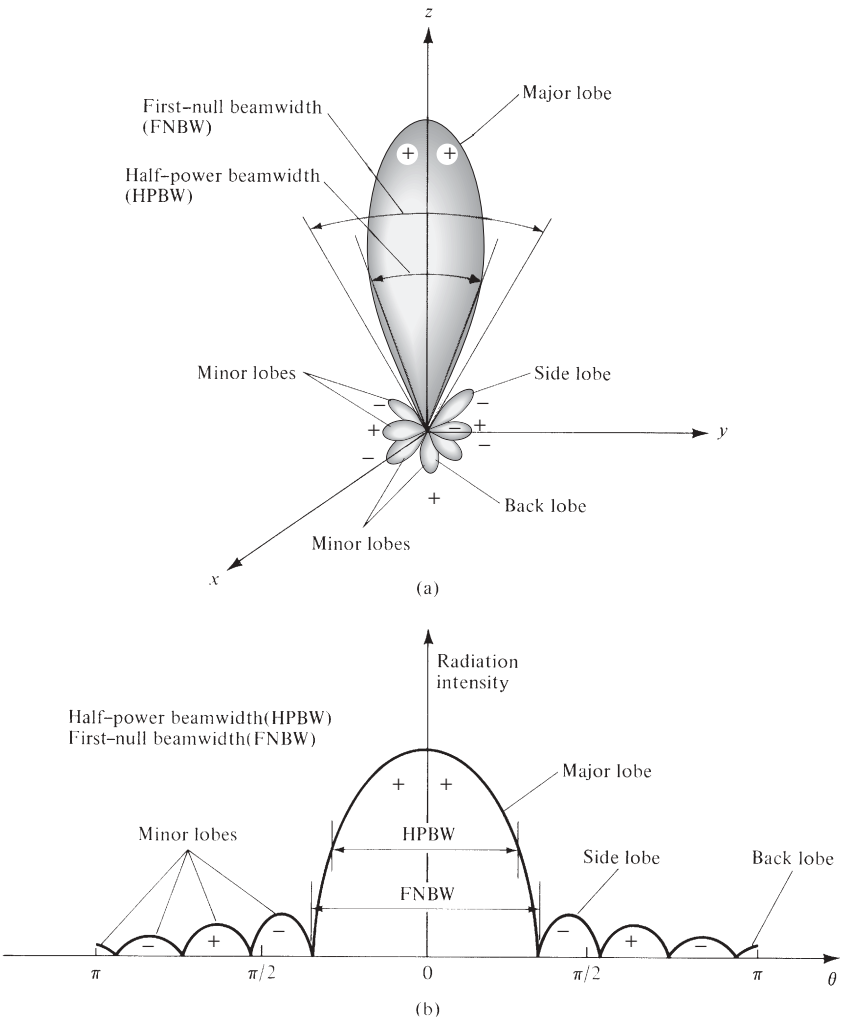


Figure 1.3 (a) Radiation lobes and beamwidths of an antenna pattern. (b) Linear plot of power pattern and its associated lobes and beamwidths.

of these programs is found in the CD attached to Ref. 1. Other programs that have been developed for plotting rectangular and polar plots are those of Refs. 1–5.

A *major lobe* (also called main beam) is defined as "the radiation lobe containing the direction of maximum radiation." In Figure 1.3 the major lobe is pointing in the $\theta = 0$ direction. In some antennas, such as split-beam antennas, there may exist more than one major lobe. A *minor lobe* is any lobe except a major lobe. In Figures 1.3a and 1.3b all the lobes with the exception of the major can be classified as minor lobes. A *side lobe* is "a radiation lobe in any direction other than the intended lobe." (Usually a side lobe is adjacent to the main lobe and occupies the hemisphere in the direction of the main beam.) A *back lobe* is "a radiation lobe whose axis makes an angle of approximately 180° with respect to the beam of an antenna." Usually it refers to a minor lobe that occupies the hemisphere in a direction opposite to that of the major (main) lobe.

Minor lobes usually represent radiation in undesired directions, and they should be minimized. Side lobes are normally the largest of the minor lobes. The level of minor lobes is usually expressed as a ratio of the power density in the lobe in question to that of the major lobe. This ratio is often termed the side lobe ratio or side lobe level. Side lobe levels of -20 dB or smaller are usually not desirable in many applications. Attainment of a side lobe level smaller than -30 dB usually requires very careful design and construction. In most radar systems, low side lobe ratios are very important to minimize false target indications through the side lobes.

A normalized three-dimensional far-field amplitude pattern, plotted on a linear scale, of a 10-element linear antenna array of isotropic sources with a spacing of $d = 0.25\lambda$ and progressive phase shift $\beta = -0.6\pi$ between the elements is shown in Figure 1.4. It

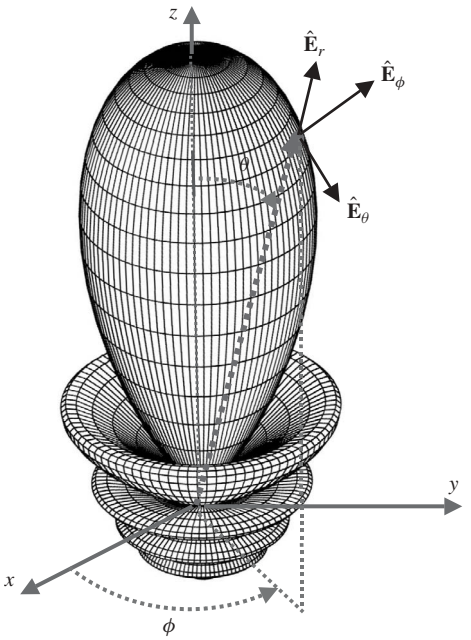


Figure 1.4 Normalized three-dimensional amplitude *field* pattern (in *linear scale*) of a 10-element linear array antenna with a uniform spacing of $d = 0.25\lambda$ and progressive phase shift $\beta = -0.6\pi$ between the elements.

is evident that this pattern has one major lobe, five minor lobes, and one back lobe. The level of the side lobe is about -9 dB relative to the maximum. A detailed presentation of arrays is found in Chapter 6 of Ref. 1. For an amplitude pattern of an antenna, there would be, in general, three electric-field components (E_r, E_θ, E_ϕ) at each observation point on the surface of a sphere of constant radius $r = r_c$, as shown in Figure 1.1. In the far field, the radial E_r component for all antennas is zero or, vanishingly small compared to either one, or both, of the other two components (see Section 3.6 of Chapter 3 of Ref. 1). Some antennas, depending on their geometry and also observation distance, may have only one, two, or all three components. In general, the magnitude of the total electric field would be $|\mathbf{E}| = \sqrt{|E_r|^2 + |E_\theta|^2 + |E_\phi|^2}$. The radial distance in Figure 1.4, and similar ones, represents the magnitude of $|\mathbf{E}|$.

1.2.2 Isotropic, Directional, and Omnidirectional Patterns

An *isotropic* radiator is defined as "a hypothetical lossless antenna having equal radiation in all directions." Although it is ideal and not physically realizable, it is often taken as a reference for expressing the directive properties of actual antennas. A *directional* antenna is one "having the property of radiating or receiving electromagnetic waves more effectively in some directions than in others. This term is usually applied to an antenna whose maximum directivity is significantly greater than that of a half-wave dipole." Examples of antennas with directional radiation patterns are shown in Figures 1.5 and 1.6. It is seen that the pattern in Figure 1.6 is nondirectional in the azimuth plane ($f(\phi)$, $\theta = \pi/2$) and directional in the elevation plane ($g(\theta)$, $\phi =$ constant). This type of a pattern is designated as *omnidirectional*, and it is defined as one "having an essentially nondirectional pattern in a given plane (in this case in azimuth) and a directional pattern in any orthogonal plane (in this case in elevation)." An *omnidirectional* pattern is then a special type of a *directional* pattern.

1.2.3 Principal Patterns

For a linearly polarized antenna, performance is often described in terms of its principal E- and H-plane patterns. The *E-plane* is defined as "the plane containing the electric-field vector and the direction of maximum radiation," and the *H-plane* as "the plane containing the magnetic-field vector and the direction of maximum radiation." Although it is very difficult to illustrate the principal patterns without considering a specific example, it is the usual practice to orient most antennas so that at least one of the principal plane patterns coincides with one of the geometrical principal planes. An illustration is shown in Figure 1.5. For this example, the x-z plane (elevation plane; $\phi = 0$) is the principal E-plane and the x-y plane (azimuthal plane; $\theta = \pi/2$) is the principal H-plane. Other coordinate orientations can be selected.

The omnidirectional pattern of Figure 1.6 has an infinite number of principal E-planes (elevation planes; $\phi = \phi_c$) and one principal H-plane (azimuthal plane; $\theta = 90^\circ$).

1.2.4 Field Regions

The space surrounding an antenna is usually subdivided into three regions: (1) *reactive near-field*, (2) *radiating near-field* (*Fresnel*), and (3) *far-field* (*Fraunhofer*) regions as shown in Figure 1.7. These regions are so designated to identify the field structure in

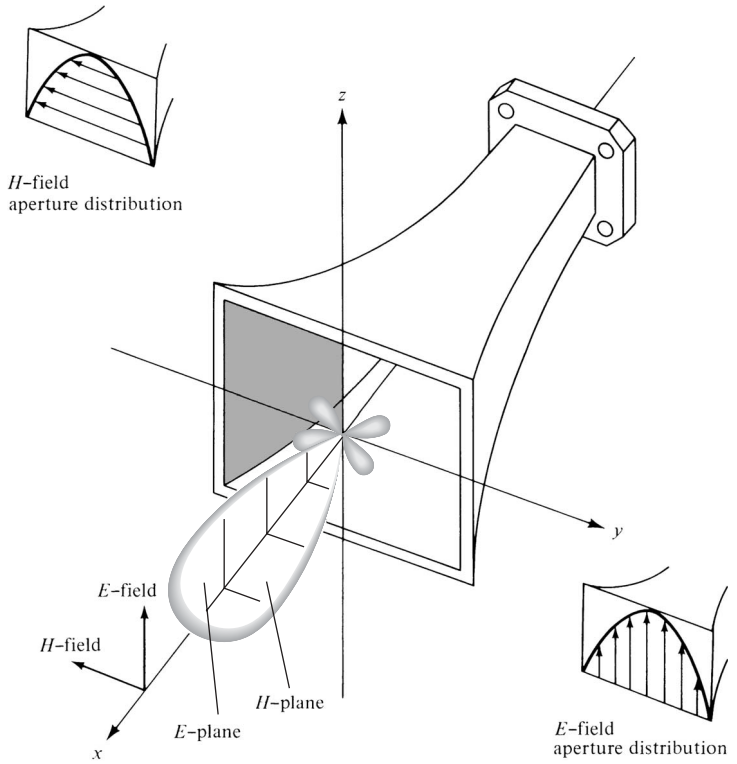


Figure 1.5 Principal E- and H-plane patterns for a pyramidal horn antenna.

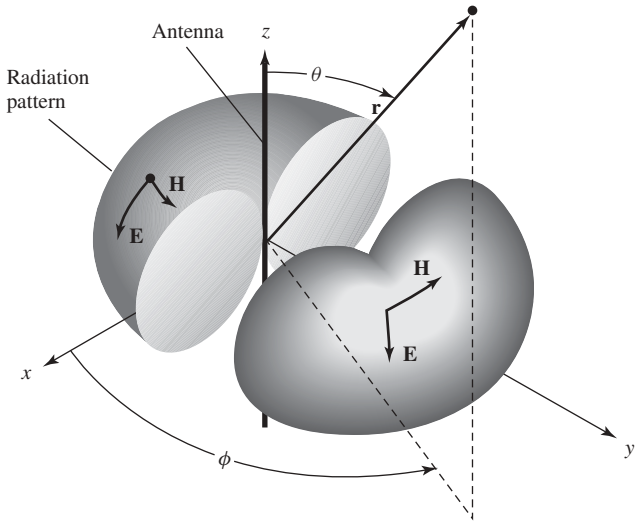


Figure 1.6 Omnidirectional antenna pattern.

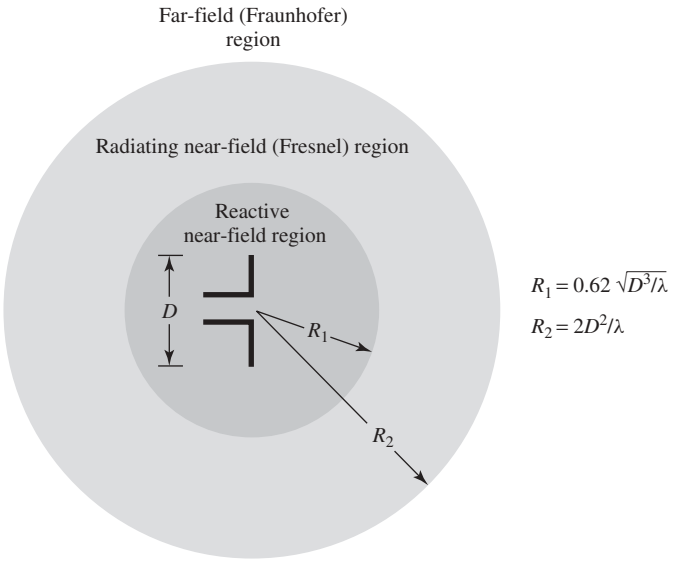


Figure 1.7 Field regions of an antenna.

each. Although no abrupt changes in the field configurations are noted as the boundaries are crossed, there are distinct differences among them. The boundaries separating these regions are not unique, although various criteria have been established and are commonly used to identify the regions.

Reactive near-field region is defined as "that portion of the near-field region immediately surrounding the antenna wherein the reactive field predominates." For most antennas, the outer boundary of this region is commonly taken to exist at a distance $R < 0.62\sqrt{D^3/\lambda}$ from the antenna surface, where λ is the wavelength and D is the largest dimension of the antenna. "For a very short dipole, or equivalent radiator, the outer boundary is commonly taken to exist at a distance $\lambda/2\pi$ from the antenna surface."

Radiating near-field (Fresnel) region is defined as "that region of the field of an antenna between the reactive near-field region and the far-field region wherein radiation fields predominate and wherein the angular field distribution is dependent upon the distance from the antenna. If the antenna has a maximum dimension that is not large compared to the wavelength, this region may not exist. For an antenna focused at infinity, the radiating near-field region is sometimes referred to as the Fresnel region on the basis of analogy to optical terminology. If the antenna has a maximum overall dimension which is very small compared to the wavelength, this field region may not exist." The inner boundary is taken to be the distance $R \geq 0.62\sqrt{D^3/\lambda}$ and the outer boundary the distance $R < 2D^2/\lambda$, where D is the largest[†] dimension of the antenna. This criterion is based on a maximum phase error of $\pi/8$. In this region the field pattern is, in general, a function of the radial distance and the radial field component may be appreciable.

Far-field (Fraunhofer) region is defined as "that region of the field of an antenna where the angular field distribution is essentially independent of the distance from the

[†]To be valid, D must also be large compared to the wavelength ($D > \lambda$).

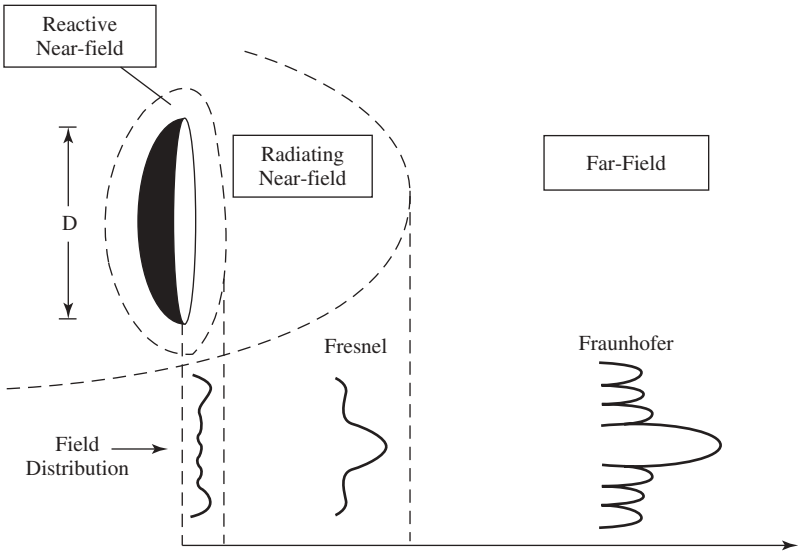


Figure 1.8 Typical changes of antenna amplitude pattern shape from reactive near field toward the far field. (From: Y. Rahmat-Samii, L. I. Williams, and R. G. Yoccarino, The UCLA bi-polar planar-near-field antenna measurement and diagnostics range, *IEEE Antennas Propag. Mag*., Vol. 37, No. 6, December 1995. Copyright © 1995 IEEE.)

antenna. If the antenna has a maximum[†] overall dimension D, the far-field region is commonly taken to exist at distances greater than $2D^2/\lambda$ from the antenna, λ being the wavelength. The far-field patterns of certain antennas, such as multibeam reflector antennas, are sensitive to variations in phase over their apertures. For these antennas $2D^2/\lambda$ may be inadequate. In physical media, if the antenna has a maximum overall dimension, D, which is large compared to $\pi/|\gamma|$, the far-field region can be taken to begin approximately at a distance equal to $|\gamma|D^2/\pi$ from the antenna, γ being the propagation constant in the medium. For an antenna focused at infinity, the far-field region is sometimes referred to as the Fraunhofer region on the basis of analogy to optical terminology." In this region, the field components are essentially transverse and the angular distribution is independent of the radial distance where the measurements are made. The inner boundary is taken to be the radial distance $R = 2D^2/\lambda$ and the outer one at infinity.

The amplitude pattern of an antenna, as the observation distance is varied from the reactive near field to the far field, changes in shape because of variations of the fields, both magnitude and phase. A typical progression of the shape of an antenna, with the largest dimension D, is shown in Figure 1.8. It is apparent that in the reactive near-field region the pattern is more spread out and nearly uniform, with slight variations. As the observation is moved to the radiating near-field region (Fresnel), the pattern begins to smooth and form lobes. In the far-field region (Fraunhofer), the pattern is well formed, usually consisting of few minor lobes and one, or more, major lobes.

[†]To be valid, D must also be large compared to the wavelength ($D > \lambda$).

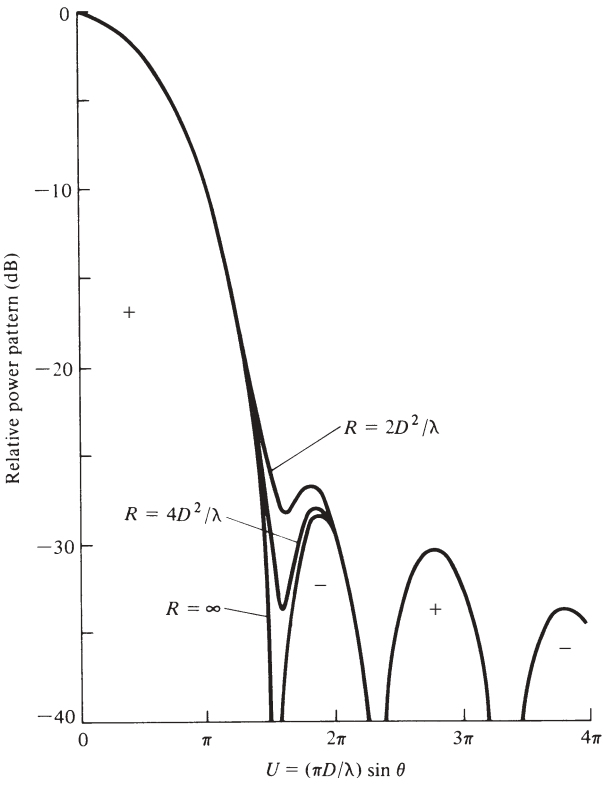


Figure 1.9 Calculated radiation patterns of a paraboloid antenna for different distances from the antenna. (From Ref. 6.)

To illustrate the pattern variation as a function of radial distance beyond the minimum $2D^2/\lambda$ far-field distance, in Figure 1.9 we have included three patterns of a parabolic reflector calculated at distances of $R = 2D^2/\lambda$, $4D^2/\lambda$, and infinity [6]. It is observed that the patterns are almost identical, except for some differences in the pattern structure around the first null and at a level below 25 dB. Because infinite distances are not realizable in practice, the most commonly used criterion for minimum distance of far-field observations is $2D^2/\lambda$.

1.2.5 Radian and Steradian

The measure of a plane angle is a radian. One *radian* is defined as the plane angle with its vertex at the center of a circle of radius r that is subtended by an arc whose length is the radius r. A graphical illustration is shown in Figure 1.10a. Since the circumference of a circle of radius r is $C = 2\pi r$, there are 2π rads $(2\pi r/r)$ in a full circle.

The measure of a solid angle is a steradian. One *steradian* is defined as the solid angle with its vertex at the center of a sphere of radius r that is subtended by a spherical surface area equal to that of a square with each side of length r. A graphical illustration is shown in Figure 1.10b. Since the area of a sphere of radius r is $A = 4\pi r^2$, there are 4π sr $(4\pi r^2/r^2)$ in a closed sphere.

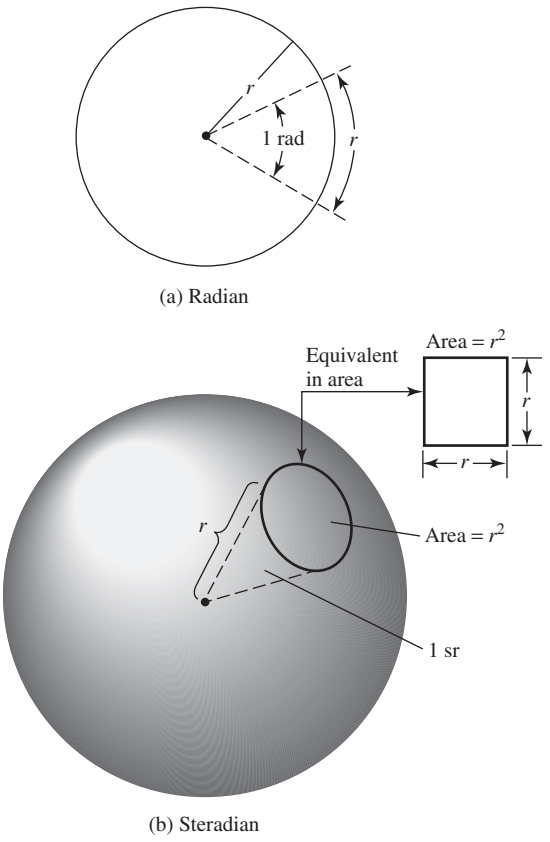


Figure 1.10 Geometrical arrangements for defining a radian and a steradian.

The infinitesimal area dA on the surface of a sphere of radius r, shown in Figure 1.1, is given by

$$dA = r^2 \sin\theta \, d\theta \, d\phi \quad (\mathrm{m}^2) \tag{1.1}$$

Therefore the element of solid angle $d\Omega$ of a sphere can be written

$$d\Omega = \frac{dA}{r^2} = \sin\theta \, d\theta \, d\phi \quad (\mathrm{sr}) \tag{1.2}$$

1.3 RADIATION POWER DENSITY

Electromagnetic waves are used to transport information through a wireless medium or a guiding structure, from one point to the other. It is then natural to assume that power and energy are associated with electromagnetic fields. The quantity used to describe the power associated with an electromagnetic wave is the instantaneous Poynting vector defined as

$$\mathscr{W} = \mathscr{E} \times \mathscr{H} \tag{1.3}$$

where

$\mathscr{W}$ = instantaneous Poynting vector (W/m^2)
$\mathscr{E}$ = instantaneous electric-field intensity (V/m)
$\mathscr{H}$ = instantaneous magnetic-field intensity (A/m)

Note that script letters are used to denote instantaneous fields and quantities, while roman letters are used to represent their complex counterparts.

Since the Poynting vector is a power density, the total power crossing a closed surface can be obtained by integrating the normal component of the Poynting vector over the entire surface. In equation form

$$\mathscr{P} = \oiint_s \mathscr{W} \cdot d\mathbf{s} = \oiint_s \mathscr{W} \cdot \hat{\mathbf{n}}\, da \tag{1.4}$$

where

$\mathscr{P}$ = instantaneous total power (W)
$\hat{\mathbf{n}}$ = unit vector normal to the surface
da = infinitesimal area of the closed surface (m^2)

The time-average Poynting vector (average power density) can be written

$$\boxed{\mathbf{W}_{\text{av}}(x, y, z) = [\mathscr{W}(x, y, z; t)]_{\text{av}} = \frac{1}{2}\text{Re}(\mathbf{E} \times \mathbf{H}^*)} \quad (\text{W/m}^2) \tag{1.5}$$

The $\frac{1}{2}$ factor appears in Eq. (1.5) because the **E** and **H** fields represent peak values, and it should be omitted for RMS values. Based on the definition of Eq. (1.5), the average power radiated by an antenna (radiated power) can be written

$$\boxed{\begin{aligned} P_{\text{rad}} = P_{\text{av}} &= \oiint_s \mathbf{W}_{\text{rad}} \cdot d\mathbf{s} = \oiint_s \mathbf{W}_{\text{av}} \cdot \hat{\mathbf{n}}\, da \\ &= \frac{1}{2} \oiint_s \text{Re}(\mathbf{E} \times \mathbf{H}^*) \cdot d\mathbf{s} \end{aligned}} \tag{1.6}$$

1.4 RADIATION INTENSITY

Radiation intensity in a given direction is defined as "the power radiated from an antenna per unit solid angle." The radiation intensity is a far-field parameter, and it can be obtained by simply multiplying the radiation density by the square of the distance. In mathematical form it is expressed as

$$\boxed{U = r^2 W_{\text{rad}}} \tag{1.7}$$

where

U = radiation intensity (W/unit solid angle)
W_{rad} = radiation density (W/m^2)

The radiation intensity is also related to the far-zone electric field of an antenna, referring to Figure 1.4, by

$$U(\theta,\phi) = \frac{r^2}{2\eta}|\mathbf{E}(r,\theta,\phi)|^2 \simeq \frac{r^2}{2\eta}[|E_\theta(r,\theta,\phi)|^2 + |E_\phi(r,\theta,\phi)|^2] \tag{1.7a}$$

$$\simeq \frac{1}{2\eta}[|E_\theta^0(\theta,\phi)|^2 + |E_\phi^0(\theta,\phi)|^2]$$

where

$\mathbf{E}(r,\theta,\phi)$ = far-zone electric-field intensity of the antenna $= \mathbf{E}^0(\theta,\phi)\frac{e^{-jkr}}{r}$
E_θ, E_ϕ = far-zone electric-field components of the antenna
η = intrinsic impedance of the medium

The radical electric-field component (E_r) is assumed, if present, to be small in the far zone. Thus the power pattern is also a measure of the radiation intensity.

The total power is obtained by integrating the radiation intensity, as given by Eq. (1.7), over the entire solid angle of 4π. Thus

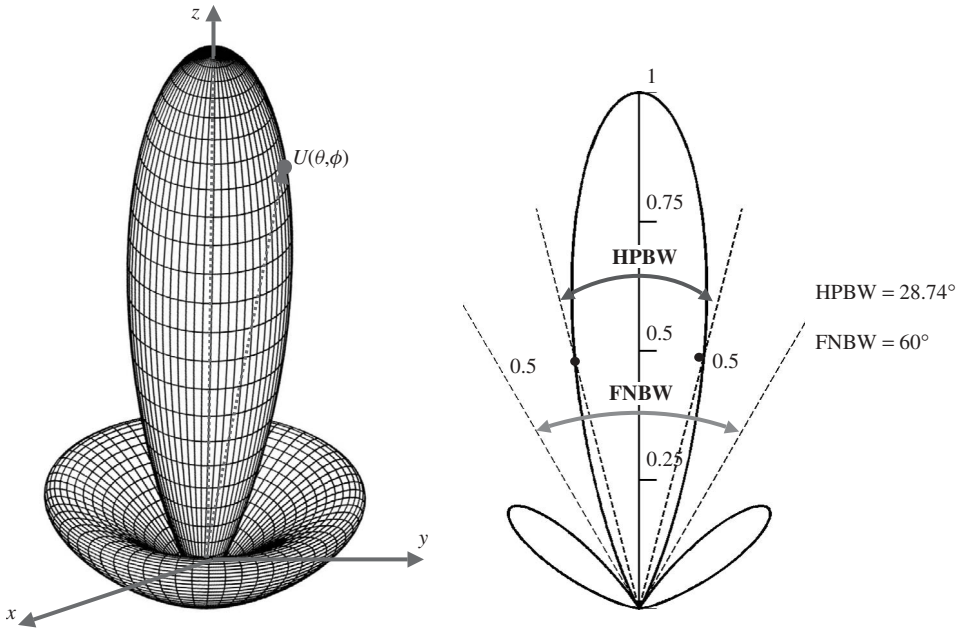
$$P_{\text{rad}} = \oiint_{\Omega} U\, d\Omega = \int_0^{2\pi}\int_0^{\pi} U \sin\theta\, d\theta\, d\phi \tag{1.8}$$

where $d\Omega$ = element of solid angle $= \sin\theta\, d\theta\, d\phi$.

1.5 BEAMWIDTH

Associated with the pattern of an antenna is a parameter designated as *beamwidth*. The *beamwidth* of a pattern is defined as the angular separation between two identical points on opposite sides of the pattern maximum. In an antenna pattern, there are a number of beamwidths. One of the most widely used beamwidths is the *half-power beamwidth* (*HPBW*), which is defined by IEEE as: "In a plane containing the direction of the maximum of a beam, the angle between the two directions in which the radiation intensity is one-half value of the beam." This is demonstrated in Figure 1.2. Another important beamwidth is the angular separation between the first nulls of the pattern, and it is referred to as the *first-null beamwidth* (*FNBW*). Both the *HPBW* and *FNBW* are demonstrated for the pattern in Figure 1.11. Other beamwidths are those where the pattern is -10 dB from the maximum, or any other value. However, in practice, the term *beamwidth*, with no other identification, usually refers to the *HPBW*.

The beamwidth of an antenna is a very important figure-of-merit and often is used as a trade-off between it and the side lobe level; that is, as the beamwidth decreases, the side lobe increases and vice versa. In addition, the beamwidth of the antenna is also used



(a) Three dimensional (b) Two dimensional

Figure 1.11 Three- and two-dimensional power patterns (in linear scale) of $U(\theta) = \cos^2(\theta)\cos^2(3\theta)$.

to describe the resolution capabilities of the antenna to distinguish between two adjacent radiating sources or radar targets. The most common resolution criterion states that *the resolution capability of an antenna to distinguish between two sources is equal to half the first-null beamwidth (FNBW/2), which is usually used to approximate the half-power beamwidth (HPBW)* [7, 8]. That is, two sources separated by angular distances equal to or greater than FNBW/2 $\approx$ HPBW of an antenna with a uniform distribution can be resolved. If the separation is smaller, then the antenna will tend to smooth the angular separation distance.

1.6 DIRECTIVITY

In the 1983 version of the *IEEE Standard Definitions of Terms for Antennas*, there was a substantive change in the definition of *directivity*, compared to the definition of the 1973 version. Basically the term *directivity* in the 1983 version has been used to replace the term *directive gain* of the 1973 version. In the 1983 version the term *directive gain* has been deprecated. According to the authors of the 1983 standards, "this change brings this standard in line with common usage among antenna engineers and with other international standards, notably those of the International Electrotechnical Commission (IEC)." Therefore *directivity of an antenna* is defined as "the ratio of the radiation intensity in a given direction from the antenna to the radiation intensity averaged over all directions. The average radiation intensity is equal to the total power radiated by the antenna divided

by 4π. If the direction is not specified, the direction of maximum radiation intensity is implied." Stated more simply, the directivity of a nonisotropic source is equal to the ratio of its radiation intensity in a given direction over that of an isotropic source. In mathematical form, it can be written

$$\boxed{D = \frac{U}{U_0} = \frac{4\pi U}{P_{\text{rad}}}} \tag{1.9}$$

If the direction is not specified, it implies the direction of maximum radiation intensity (maximum directivity) expressed as

$$\boxed{D_{\max} = D_0 = \frac{U|_{\max}}{U_0} = \frac{U_{\max}}{U_0} = \frac{4\pi U_{\max}}{P_{\text{rad}}}} \tag{1.9a}$$

where

D = directivity (dimensionless)
D_0 = maximum directivity (dimensionless)
U = radiation intensity (W/unit solid angle)
$U_{\max}$ = maximum radiation intensity (W/unit solid angle)
U_0 = radiation intensity of isotropic source (W/unit solid angle)
P_{rad} = total radiated power (W)

For an isotropic source, it is very obvious from Eq. (1.9) or (1.9a) that the directivity is unity since U, $U_{\max}$, and U_0 are all equal to each other.

For antennas with orthogonal polarization components, we define the *partial directivity of an antenna for a given polarization in a given direction* as "that part of the radiation intensity corresponding to a given polarization divided by the total radiation intensity averaged over all directions." With this definition for the partial directivity, then in a given direction "the total directivity is the sum of the partial directivities for any two orthogonal polarizations." For a spherical coordinate system, the total maximum directivity D_0 for the orthogonal θ and ϕ components of an antenna can be written

$$D_0 = D_\theta + D_\phi \tag{1.10}$$

while the partial directivities D_θ and D_ϕ are expressed as

$$D_\theta = \frac{4\pi U_\theta}{(P_{\text{rad}})_\theta + (P_{\text{rad}})_\phi} \tag{1.10a}$$

$$D_\phi = \frac{4\pi U_\phi}{(P_{\text{rad}})_\theta + (P_{\text{rad}})_\phi} \tag{1.10b}$$

where

U_θ = radiation intensity in a given direction contained in θ field component
U_ϕ = radiation intensity in a given direction contained in ϕ field component
$(P_{\text{rad}})_\theta$ = radiated power in all directions contained in θ field component
$(P_{\text{rad}})_\phi$ = radiated power in all directions contained in ϕ field component

The directivity of an isotropic source is unity since its power is radiated equally well in all directions. *For all other sources, the maximum directivity will always be greater than unity, and it is a relative "figure-of-merit" that gives an indication of the directional properties of the antenna as compared with those of an isotropic source*. In equation form, this is indicated in Eq. (1.9a). The directivity can be smaller than unity; in fact it can be equal to zero. *The values of directivity will be equal to or greater than zero and equal to or less than the maximum directivity* $(0 \le D \le D_0)$.

A more general expression for the directivity can be developed to include sources with radiation patterns that may be functions of both spherical coordinate angles θ and ϕ. The radiation intensity of an antenna can be written

$$U = B_0 F(\theta, \phi) \simeq \frac{1}{2\eta}[|E_\theta^0(\theta, \phi)|^2 + |E_\phi^0(\theta, \phi)|^2] \tag{1.11}$$

where B_0 is a constant, and E_θ^0 and E_ϕ^0 are the antenna's far-zone electric-field components. The maximum value of Eq. (1.11) is given by

$$U_{\max} = B_0 F(\theta, \phi)|_{\max} = B_0 F_{\max}(\theta, \phi) \tag{1.11a}$$

The maximum directivity can be written

$$\boxed{D_0 = \frac{4\pi}{\left(\int_0^{2\pi} \int_0^{\pi} F(\theta, \phi) \sin\theta \, d\theta \, d\phi\right) \Big/ F(\theta, \phi)|_{\max}} = \frac{4\pi}{\Omega_A}} \tag{1.12}$$

where Ω_A is the *beam solid angle*, and it is given by

$$\Omega_A = \frac{1}{F(\theta, \phi)|_{\max}} \int_0^{2\pi} \int_0^{\pi} F(\theta, \phi) \sin\theta \, d\theta \, d\phi = \int_0^{2\pi} \int_0^{\pi} F_n(\theta, \phi) \sin\theta \, d\theta \, d\phi \tag{1.12a}$$

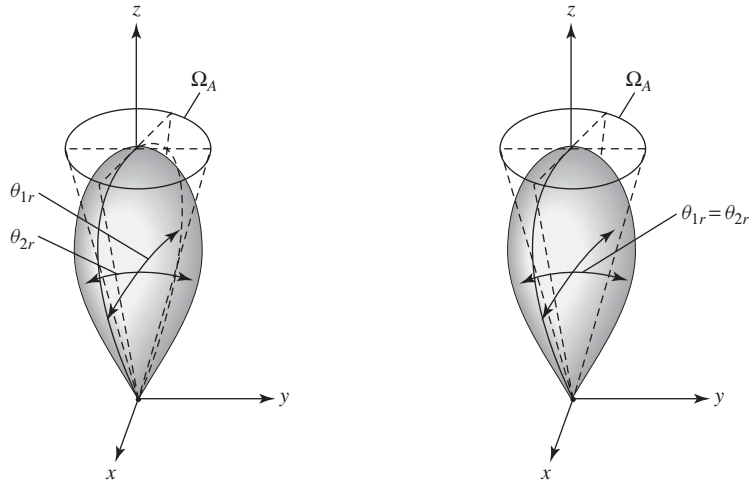
$$F_n(\theta, \phi) = \frac{F(\theta, \phi)}{F(\theta, \phi)|_{\max}} \tag{1.12b}$$

Dividing by $F(\theta, \phi)|_{\max}$ merely normalizes the radiation intensity $F(\theta, \phi)$, and it makes its maximum value unity.

The beam solid angle Ω_A is defined as the solid angle through which all the power of the antenna would flow if its radiation intensity is constant (and equal to the maximum value of U) for all angles within Ω_A.

1.6.1 Directional Patterns

Instead of using the exact expression of Eq. (1.12) to compute the directivity, it is often convenient to derive simpler expressions, even if they are approximate, to compute the directivity. These can also be used for design purposes. For antennas with one narrow major lobe and very negligible minor lobes, the beam solid angle is approximately equal to the product of the half-power beamwidths in two perpendicular planes [7] shown in Figure 1.12(a). For a rotationally symmetric pattern, the half-power beamwidths in any two perpendicular planes are the same, as illustrated in Figure 1.12(b).



(a) Nonsymmetrical pattern (b) Symmetrical pattern

Figure 1.12 Beam solid angles for nonsymmetrical and symmetrical radiation patterns.

With this approximation, Eq. (1.12) can be approximated by

$$\boxed{D_0 = \frac{4\pi}{\Omega_A} \simeq \frac{4\pi}{\Theta_{1r}\Theta_{2r}}} \tag{1.13}$$

The beam solid angle Ω_A has been approximated by

$$\boxed{\Omega_A \simeq \Theta_{1r}\Theta_{2r}} \tag{1.13a}$$

where

Θ_{1r} = half-power beamwidth in one plane (rad)
Θ_{2r} = half-power beamwidth in a plane at a right angle to the other (rad)

If the beamwidths are known in degrees, Eq. (1.13) can be written

$$D_0 \simeq \frac{4\pi(180/\pi)^2}{\Theta_{1d}\Theta_{2d}} = \frac{41,253}{\Theta_{1d}\Theta_{2d}} \tag{1.14}$$

where

Θ_{1d} = half-power beamwidth in one plane (degrees)
Θ_{2d} = half-power beamwidth in a plane at a right angle to the other (degrees)

For planar arrays, a better approximation to Eq. (1.14) is [9]

$$\boxed{D_0 \simeq \frac{32,400}{\Omega_A(\text{degrees})^2} = \frac{32,400}{\Theta_{1d}\Theta_{2d}}} \tag{1.14a}$$

The validity of Eqs. (1.13) and (1.14) is based on a pattern that has only one major lobe and any minor lobes, if present, should be of very low intensity. For a pattern with two identical major lobes, the value of the maximum directivity using Eq. (1.13) or (1.14) will be twice its actual value. For patterns with significant minor lobes, the values of maximum directivity obtained using Eq. (1.13) or (1.14), which neglect any minor lobes, will usually be too high.

Many times it is desirable to express the directivity in decibels (dB) instead of dimensionless quantities. The expressions for converting the dimensionless quantities of directivity and maximum directivity to decibels (dB) are

$$D(\text{dB}) = 10\log_{10}[D(\text{dimensionless})] \tag{1.15a}$$

$$D_0(\text{dB}) = 10\log_{10}[D_0(\text{dimensionless})] \tag{1.15b}$$

It has also been proposed [10] that the maximum directivity of an antenna can also be obtained approximately by using the formula

$$\frac{1}{D_0} = \frac{1}{2}\left(\frac{1}{D_1} + \frac{1}{D_2}\right) \tag{1.16}$$

where

$$D_1 \simeq \frac{1}{\left(\frac{1}{2\ln 2}\int_0^{\Theta_{1r}/2}\sin\theta d\theta\right)} \simeq \frac{16\ln 2}{\Theta_{1r}^2} \tag{1.16a}$$

$$D_2 \simeq \frac{1}{\left(\frac{1}{2\ln 2}\int_0^{\Theta_{2r}/2}\sin\theta d\theta\right)} \simeq \frac{16\ln 2}{\Theta_{2r}^2} \tag{1.16b}$$

Θ_{1r} and Θ_{2r} are the half-power beamwidths (in radians) of the E and H planes, respectively. Formula (1.16) will be referred to as the arithmetic mean of the maximum directivity. Using Eqs. (1.16a) and (1.16b) we can write Eq. (1.16) as

$$\frac{1}{D_0} \simeq \frac{1}{2\ln 2}\left(\frac{\Theta_{1r}^2}{16} + \frac{\Theta_{2r}^2}{16}\right) = \frac{\Theta_{1r}^2 + \Theta_{2r}^2}{32\ln 2} \tag{1.17}$$

or

$$\boxed{D_0 \simeq \frac{32\ln 2}{\Theta_{1r}^2 + \Theta_{2r}^2} = \frac{22.181}{\Theta_{1r}^2 + \Theta_{2r}^2}} \tag{1.17a}$$

$$D_0 \simeq \frac{22.181(180/\pi)^2}{\Theta_{1d}^2 + \Theta_{2d}^2} = \frac{72{,}815}{\Theta_{1d}^2 + \Theta_{2d}^2} \tag{1.17b}$$

where Θ_{1d} and Θ_{2d} are the half-power beamwidths in degrees.

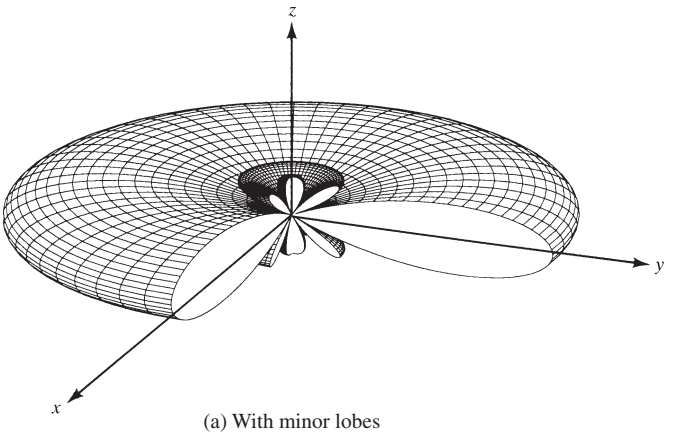
1.6.2 Omnidirectional Patterns

Some antennas (such as dipoles, loops, and broadside arrays) exhibit omnidirectional patterns, as illustrated by the three-dimensional patterns in Figure 1.13. Approximate directivity formulas have been derived [11, 12] for antennas with omnidirectional patterns similar to the ones shown in Figure 1.13. The approximate directivity formula for an omnidirectional pattern as a function of the pattern half-power beamwidth (in degrees), which is reported by McDonald in [11], was derived based on the array factor of a broadside collinear array, and it is given by

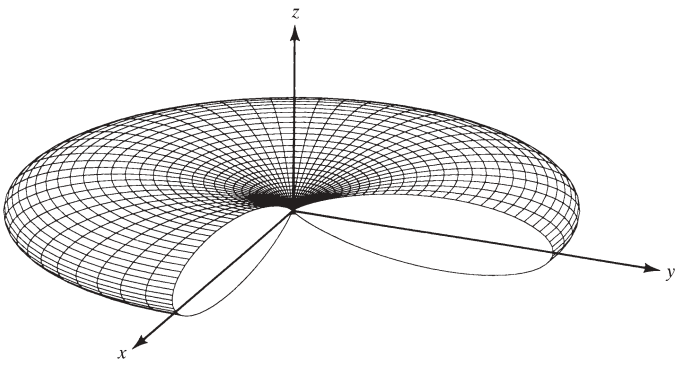
$$D_0 \simeq \frac{101}{\text{HPBW (degrees)} - 0.0027[\text{HPBW (degrees)}]^2} \tag{1.18a}$$

However, that reported by Pozar [12] is given by

$$D_0 \simeq -172.4 + 191\sqrt{0.818 + 1/\text{HPBW (degrees)}} \tag{1.18b}$$



(a) With minor lobes



(b) Without minor lobes

Figure 1.13 Omnidirectional patterns with and without minor lobes.

1.7 NUMERICAL TECHNIQUES

For most practical antennas, their radiation patterns are so complex that closed-form mathematical expressions are not available. Even in those cases where expressions are available, their form is so complex that integration to find the radiated power, required to compute the maximum directivity, cannot be performed. Instead of using the approximate expressions of Kraus, Tai and Pereira, McDonald, or Pozar, alternate and more accurate techniques may be desirable. With the high speed computer systems now available, the answer may be to apply numerical methods.

Let us assume that the radiation intensity of a given antenna is given by

$$U = B_0 \, \mathrm{F}(\theta, \phi) \tag{1.19}$$

where B_0 is a constant. The directivity for such a system is given, in general, by

$$D_0 = \frac{4\pi U_{\max}}{P_{\text{rad}}} \tag{1.20}$$

where

$$P_{\text{rad}} = B_0 \int_0^{2\pi} \left(\int_0^{\pi} \mathrm{F}(\theta, \phi) \sin\theta \, d\theta \right) d\phi \tag{1.20a}$$

For N uniform divisions over the π interval of θ and M uniform divisions over the 2π interval of ϕ, the digital form of the radiated power (Eq. (1.20a)) can be written as

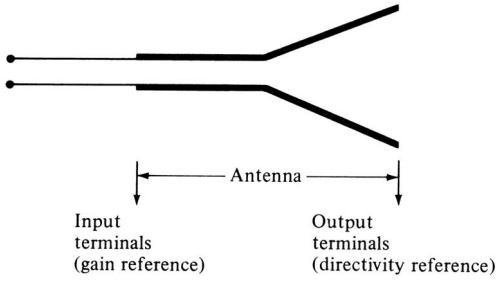
$$\boxed{P_{\text{rad}} = B_0 \left(\frac{\pi}{N}\right) \left(\frac{2\pi}{M}\right) \sum_{j=1}^{M} \left(\sum_{i=1}^{N} F(\theta_i, \phi_j) \sin\theta_i \right)} \tag{1.21}$$

where θ_i and ϕ_j represent the discrete values of θ and ϕ.

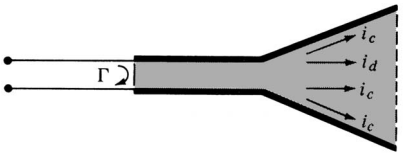
A MATLAB and FORTRAN computer program called *Directivity* has been developed to compute the maximum directivity of any antenna whose radiation intensity is $U = F(\theta, \phi)$ based on the formulation of Eq. (1.21). The intensity function F does not have to be a function of both θ and ϕ. The program is included in the CD attached to [1]. It contains a *subroutine* for which the intensity factor $U = F(\theta, \phi)$ for the required application must be specified by the user. As an illustration, the antenna intensity $U = \sin\theta \sin^2\phi$ has been inserted in the subroutine. In addition, the upper and lower limits of θ and ϕ must be specified for each application of the same pattern.

1.8 ANTENNA EFFICIENCY

Associated with an antenna are a number of efficiencies that can be defined using Figure 1.14. The total antenna efficiency e_0 is used to take into account losses at the input terminals and within the structure of the antenna. Such losses may be due, referring to Figure 1.14(b), to (1) reflections because of the mismatch between the transmission line and the antenna and (2) I^2R losses (conduction and dielectric).



(a) Antenna reference terminals



(b) Reflection, conduction, and dielectric losses

Figure 1.14 Reference terminals and losses of an antenna.

In general, the overall efficiency can be written

$$\boxed{e_0 = e_r e_c e_d} \tag{1.22}$$

where

e_0 = total efficiency (dimensionless)
e_r = reflection (mismatch) efficiency $= (1 - |\Gamma|^2)$ (dimensionless)
e_c = conduction efficiency (dimensionless)
e_d = dielectric efficiency (dimensionless)
Γ = voltage reflection coefficient at the input terminals of the antenna [$\Gamma = (Z_{\text{in}} - Z_0)/(Z_{\text{in}} + Z_0)$ where Z_{in} = antenna input impedance and Z_0 = characteristic impedance of the transmission line]
VSWR = voltage standing wave ratio $= (1 + |\Gamma|)/(1 - |\Gamma|)$

Usually e_c and e_d are very difficult to compute, but they can be determined experimentally. Even by measurements they cannot be separated, and it is usually more convenient to write Eq. (1.22) as

$$e_0 = e_r e_{cd} = e_{cd}(1 - |\Gamma|^2) \tag{1.23}$$

where $e_{cd} = e_c e_d$ = antenna radiation efficiency, which is used to relate the gain and directivity.

1.9 GAIN

Another useful measure describing the performance of an antenna is the *gain*. Although the gain of the antenna is closely related to the directivity, it is a measure that takes into account the efficiency of the antenna as well as its directional capabilities.

Gain of an antenna (in a given direction) is defined as "the ratio of the intensity, in a given direction, to the radiation intensity that would be obtained if the power accepted by the antenna were radiated isotropically. The radiation intensity corresponding to the isotropically radiated power is equal to the power accepted (input) by the antenna divided by 4π."

In most cases we deal with *relative gain*, which is defined as "the ratio of the power gain in a given direction to the power gain of a reference antenna in its referenced direction." The power input must be the same for both antennas. The reference antenna is usually a dipole, horn, or any other antenna whose gain can be calculated or it is known. In most cases, however, the reference antenna is a *lossless isotropic source*. Thus

$$G = \frac{4\pi U(\theta, \phi)}{P_{\text{in}}(\text{lossless isotropic source})} \quad \text{(dimensionless)} \tag{1.24}$$

When the direction is not stated, the power gain is usually taken in the direction of maximum radiation.

Referring to Figure 1.14(a), we can write that the total radiated power (P_{rad}) is related to the total input power (P_{in}) by

$$P_{\text{rad}} = e_{cd} P_{\text{in}} \tag{1.25}$$

where e_{cd} is the antenna radiation efficiency (dimensionless), which is defined in Eqs. (1.22) and (1.23). According to the IEEE Standards, "gain does not include losses arising from impedance mismatches (reflection losses) and polarization mismatches (losses)."

Here we define two gains: one, referred to as *gain* (G), and the other, referred to as *absolute gain* (G_{abs}), that also takes into account the reflection/mismatch losses represented in both Eqs. (1.22) and (1.23).

Using Eq. (1.25) reduces Eq. (1.24) to

$$G(\theta, \phi) = e_{cd} \left(4\pi \frac{U(\theta, \phi)}{P_{\text{rad}}} \right) \tag{1.26}$$

which is related to the directivity of Eq. (1.9) by

$$\boxed{G(\theta, \phi) = e_{cd} D(\theta, \phi)} \tag{1.27}$$

In a similar manner, the maximum value of the gain is related to the maximum directivity of Eq. (1.9a) and (1.12) by

$$\boxed{G_0 = G(\theta, \phi)|_{\max} = e_{cd} D(\theta, \phi)|_{\max} = e_{cd} D_0} \tag{1.27a}$$

While Eq. (1.25) does take into account the losses of the antenna element itself, *it does not take into account the losses when the antenna element is connected to a transmission line*, as shown in Figure 1.14. These connection losses are usually referred to as *reflections (mismatch) losses*, and they are taken into account by introducing a reflection (mismatch) efficiency e_r, which is related to the reflection coefficient as shown

in Eq. (1.23) or $e_r = (1 - |\Gamma|^2)$. Thus we can introduce an *absolute gain* G_{abs} that takes into account the reflection/mismatch losses (due to the connection of the antenna element to the transmission line), and it can be written

$$\boxed{\begin{aligned} G_{\text{abs}}(\theta, \phi) &= e_r G(\theta, \phi) = (1 - |\Gamma|^2) G(\theta, \phi) \\ &= e_r e_{cd} D(\theta, \phi) = e_o D(\theta, \phi) \end{aligned}} \tag{1.28}$$

where e_o is the overall efficiency as defined in Eqs. (1.22) and (1.23). Similarly, the *maximum absolute* gain $G_{0\text{abs}}$ of Eq. (1.28) is related to the maximum directivity D_0 by

$$\boxed{\begin{aligned} G_{0\text{abs}} &= G_{\text{abs}}(\theta, \phi)|_{\max} = e_r G(\theta, \phi)|_{\max} = (1 - |\Gamma|^2) G(\theta, \phi)|_{\max} \\ &= e_r e_{cd} D(\theta, \phi)|_{\max} = e_o D(\theta, \phi)|_{\max} = e_o D_0 \end{aligned}} \tag{1.28a}$$

If the antenna is matched to the transmission line, that is, the antenna input impedance Z_{in} *is equal to the characteristic impedance* Z_0 *of the line* ($|\Gamma| = 0$), *then the two gains are equal* ($G_{\text{abs}} = G$).

As was done with the directivity, we can define the *partial gain of an antenna for a given polarization in a given direction* as "that part of the radiation intensity corresponding to a given polarization divided by the total radiation intensity that would be obtained if the power accepted by the antenna were radiated isotropically." With this definition for the partial gain, then, in a given direction, "the total gain is the sum of the partial gains for any two orthogonal polarizations." For a spherical coordinate system, the total maximum gain G_0 for the orthogonal θ and ϕ components of an antenna can be written, in a similar form as was the maximum directivity in Eqs. (1.10), (1.10a) and (1.10b), as

$$G_0 = G_\theta + G_\phi \tag{1.29}$$

while the partial gains G_θ and G_ϕ are expressed as

$$G_\theta = \frac{4\pi U_\theta}{P_{\text{in}}} \tag{1.29a}$$

$$G_\phi = \frac{4\pi U_\phi}{P_{\text{in}}} \tag{1.29b}$$

where

U_θ = radiation intensity in a given direction contained in E_θ field component
U_ϕ = radiation intensity in a given direction contained in E_ϕ field component
P_{in} = total input (accepted) power

For many practical antennas an approximate formula for the gain, corresponding to Eq. (1.14) or (1.14a) for the directivity, is

$$\boxed{G_0 \simeq \frac{30{,}000}{\Theta_{1d}\Theta_{2d}}} \tag{1.30}$$

In practice, whenever the term "gain" is used, it usually refers to the *maximum gain* as defined by Eq. (1.27a) or (1.28a).

Usually the gain is given in terms of decibels instead of the dimensionless quantity of Eq. (1.27a). The conversion formula is given by

$$G_0(\text{dB}) = 10\log_{10}[e_{cd}\,D_0\,(\text{dimensionless})] \tag{1.31}$$

1.10 BEAM EFFICIENCY

Another parameter that is frequently used to judge the quality of transmitting and receiving antennas is the *beam efficiency*. For an antenna with its major lobe directed along the z-axis ($\theta = 0$), as shown in Figure 1.1, the beam efficiency (BE) is defined by

$$\text{BE} = \frac{\text{power transmitted (received) within cone angle } \theta_1}{\text{power transmitted (received) by the antenna}} \quad \text{(dimensionless)} \tag{1.32}$$

where θ_1 is the half-angle of the cone within which the percentage of the total power is to be found. Equation (1.32) can be written

$$\text{BE} = \frac{\displaystyle\int_0^{2\pi}\int_0^{\theta_1} U(\theta,\phi)\sin\theta\,d\theta\,d\phi}{\displaystyle\int_0^{2\pi}\int_0^{\pi} U(\theta,\phi)\sin\theta\,d\theta\,d\phi} \tag{1.33}$$

If θ_1 is chosen as the angle where the first null or minimum occurs (see Figure 1.1), then the beam efficiency will indicate the amount of power in the major lobe compared to the total power. A very high beam efficiency (between the nulls or minimums), usually in the high 90s, is necessary for antennas used in radiometry, astronomy, radar, and other applications where received signals through the minor lobes must be minimized.

1.11 BANDWIDTH

The *bandwidth* of an antenna is defined as "the range of frequencies within which the performance of the antenna, with respect to some characteristic, conforms to a specified standard." The bandwidth can be considered to be the range of frequencies, on either side of a center frequency (usually the resonance frequency for a dipole), where the antenna characteristics (such as input impedance, pattern, beamwidth, polarization, side lobe level, gain, beam direction, radiation efficiency) are within an acceptable value of those at the center frequency. For broadband antennas, the bandwidth is usually expressed as the ratio of the upper-to-lower frequencies of acceptable operation. For example, a 10:1 bandwidth indicates that the upper frequency is 10 times greater than the lower. For narrowband antennas, the bandwidth is expressed as a percentage of the frequency difference (upper minus lower) over the center frequency of the bandwidth. For example, a 5% bandwidth indicates that the frequency difference of acceptable operation is 5% of the center frequency of the bandwidth.

Because the characteristics (input impedance, pattern, gain, polarization, etc.) of an antenna do not necessarily vary in the same manner or are not even critically affected by the frequency, there is no unique characterization of the bandwidth. The specifications are set in each case to meet the needs of the particular application. Usually there is a distinction made between pattern and input impedance variations. Accordingly *pattern bandwidth* and *impedance bandwidth* are used to emphasize this distinction. Associated with pattern bandwidth are gain, side lobe level, beamwidth, polarization, and beam direction while input impedance and radiation efficiency are related to impedance bandwidth. For example, the pattern of a linear dipole with overall length less than a half-wavelength ($l < \lambda/2$) is insensitive to frequency. The limiting factor for this antenna is its impedance, and its bandwidth can be formulated in terms of the Q. The Q of antennas or arrays with dimensions large compared to the wavelength, excluding superdirective designs, is near unity. Therefore the bandwidth is usually formulated in terms of beamwidth, side lobe level, and pattern characteristics. For intermediate length antennas, the bandwidth may be limited by either pattern or impedance variations, depending on the particular application. For these antennas, a 2:1 bandwidth indicates a good design. For others, large bandwidths are needed. Antennas with very large bandwidths (like 40:1 or greater) have been designed in recent years. These are known as *frequency-independent* antennas.

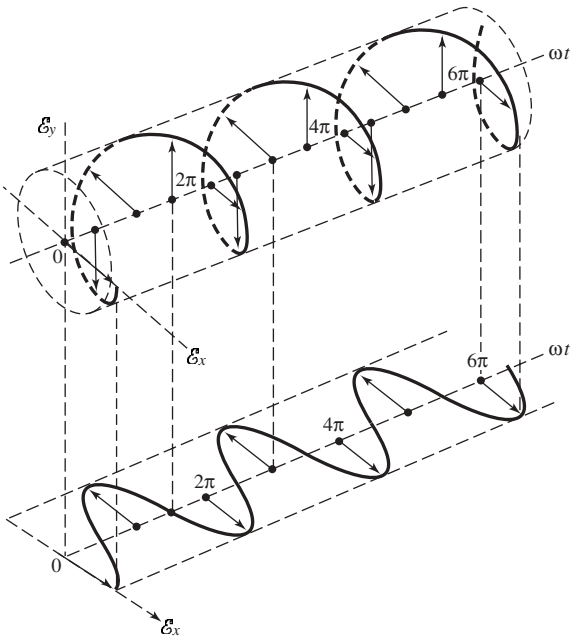
The above discussion presumes that the coupling networks (transformers, baluns, etc.) and/or the dimensions of the antenna are not altered in any manner as the frequency is changed. It is possible to increase the acceptable frequency range of a narrowband antenna if proper adjustments can be made on the critical dimensions of the antenna and/or on the coupling networks as the frequency is changed. Although not an easy or possible task in general, there are applications where this can be accomplished. The most common examples are the antenna of a car radio and the "rabbit ears" of a television. Both usually have adjustable lengths that can be used to tune the antenna for better reception.

1.12 POLARIZATION

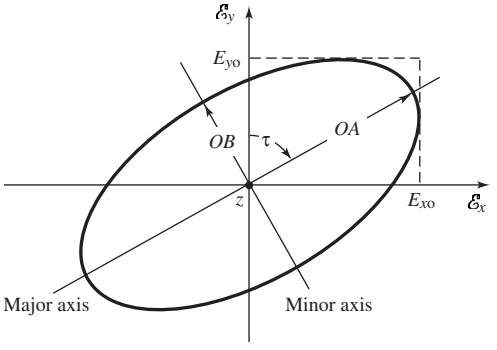
Polarization of an antenna in a given direction is defined as "the polarization of the wave transmitted (radiated) by the antenna. *Note:* When the direction is not stated, the polarization is taken to be the polarization in the direction of maximum gain." In practice, polarization of the radiated energy varies with the direction from the center of the antenna, so that different parts of the pattern may have different polarizations.

Polarization of a radiated wave is defined as "that property of an electromagnetic wave describing the time-varying direction and relative magnitude of the electric-field vector; specifically, the figure traced as a function of time by the extremity of the vector at a fixed location in space, and the sense in which it is traced, *as observed along the direction of propagation*." Polarization then is the curve traced by the end point of the arrow (vector) representing the instantaneous electric field. The field must be observed along the direction of propagation. A typical trace as a function of time is shown in Figure 1.15.

The polarization of a wave can be defined in terms of a wave *radiated (transmitted)* or *received* by an antenna in a given direction. The polarization of a wave *radiated* by an antenna in a specified direction at a point in the far field is defined as "the polarization of the (locally) plane wave which is used to represent the radiated wave at that point. At



(a) Rotation of wave



(b) Polarization ellipse

Figure 1.15 Rotation of a plane electromagnetic wave and its polarization ellipse at $z = 0$ as a function of time.

any point in the far field of an antenna the radiated wave can be represented by a plane wave whose electric-field strength is the same as that of the wave and whose direction of propagation is in the radial direction from the antenna. As the radial distance approaches infinity, the radius of curvature of the radiated wave's phase front also approaches infinity and thus in any specified direction the wave appears locally as a plane wave." This is a far-field characteristic of waves radiated by all practical antennas. The polarization of a wave *received* by an antenna is defined as the "polarization of a plane wave, incident from a given direction and having a given power flux density, which results in maximum available power at the antenna terminals."

Polarization may be classified as *linear*, *circular*, or *elliptical*. If the vector that describes the electric field at a point in space as a function of time is always directed along a line, the field is said to be *linearly* polarized. In general, however, the figure that the electric field traces is an ellipse, and the field is said to be elliptically polarized. Linear and circular polarizations are special cases of elliptical, and they can be obtained when the ellipse becomes a straight line or a circle, respectively. The figure of the electric field is traced in a *clockwise* (CW) or *counterclockwise* (CCW) sense. *Clockwise* rotation of the electric-field vector is also designated as *right-hand polarization* and *counterclockwise* as *left-hand polarization*.

In general, the polarization characteristics of an antenna can be represented by its *polarization pattern* whose one definition is "the spatial distribution of the polarizations of a field vector excited (radiated) by an antenna taken over its radiation sphere. When describing the polarizations over the radiation sphere, or portion of it, reference lines shall be specified over the sphere, in order to measure the tilt angles (see tilt angle) of the polarization ellipses and the direction of polarization for linear polarizations. An obvious choice, though by no means the only one, is a family of lines tangent at each point on the sphere to either the θ or ϕ coordinate line associated with a spherical coordinate system of the radiation sphere. At each point on the radiation sphere the polarization is usually resolved into a pair of orthogonal polarizations, the *co-polarization* and *cross polarization*. To accomplish this, the co-polarization must be specified at each point on the radiation sphere. . . . *Co-polarization* represents the polarization the antenna is intended to radiate (receive) while *Cross polarization* represents the polarization orthogonal to a specified polarization which is usually the co-polarization.

"For certain linearly polarized antennas, it is common practice to define the co-polarization in the following manner: First specify the orientation of the co-polar electric-field vector at a pole of the radiation sphere. Then, for all other directions of interest (points on the radiation sphere), require that the angle that the co-polar electric-field vector makes with each great circle line through the pole remain constant over that circle, the angle being that at the pole.

"In practice, the axis of the antenna's main beam should be directed along the polar axis of the radiation sphere. The antenna is then appropriately oriented about this axis to align the direction of its polarization with that of the defined co-polarization at the pole. . . .This manner of defining co-polarization can be extended to the case of elliptical polarization by defining the constant angles using the major axes of the polarization ellipses rather than the co-polar electric-field vector. The sense of polarization (rotation) must also be specified."

The polarization of the wave radiated by the antenna can also be represented on the Poincaré sphere [13–16]. Each point on the Poincaré sphere represents a unique polarization. The north pole represents left circular polarization, the south pole represents right circular, and points along the equator represent linear polarization of different tilt angles. All other points on the Poincaré sphere represent elliptical polarization. For details, see Figure 17.24 of Chapter 17 [1].

1.12.1 Linear, Circular, and Elliptical Polarizations

We summarize the discussion on polarization by stating the general characteristics and the *necessary and sufficient* conditions that the wave must have in order to possess *linear, circular*, or *elliptical* polarization.

Linear Polarization A *time-harmonic wave is linearly polarized at a given point in space if the electric-field (or magnetic-field) vector at that point is always oriented along the same straight line at every instant of time*. This is accomplished if the field vector (electric or magnetic) possesses the following:

1. Only one component, or
2. Two orthogonal linear components that are in time phase or 180° (or multiples of 180°) out-of-phase.

Circular Polarization *A time-harmonic wave is circularly polarized at a given point in space if the electric (or magnetic) field vector at that point traces a circle as a function of time*.

The *necessary and sufficient* conditions to accomplish this are if the field vector (electric or magnetic) possesses *all* of the following:

1. The field must have two orthogonal linear components, and
2. The two components must have the same magnitude, and
3. The two components must have a time-phase difference of odd multiples of 90°.

The sense of rotation is always determined by rotating the phase-leading component toward the phase-lagging component and observing the field rotation as the wave is viewed as it travels away from the observer. If the rotation is clockwise, the wave is right-hand (or clockwise) circularly polarized; if the rotation is counterclockwise, the wave is left-hand (or counterclockwise) circularly polarized. The rotation of the phase-leading component toward the phase-lagging component should be done along the angular separation between the two components that is less than 180°. *Phases equal to or greater than* 0° *and less than* 180° *should be considered leading whereas those equal to or greater than* 180° *and less than* 360° *should be considered lagging*.

Elliptical Polarization *A time-harmonic wave is elliptically polarized if the tip of the field vector (electric or magnetic) traces an elliptical locus in space. At various instants of time the field vector changes continuously with time in such a manner as to describe an elliptical locus. It is right-hand (clockwise) elliptically polarized if the field vector rotates clockwise, and it is left-hand (counterclockwise) elliptically polarized if the field vector of the ellipse rotates counterclockwise* [13]. The sense of rotation is determined using the same rules as for the circular polarization. In addition to the sense of rotation, elliptically polarized waves are also specified by their axial ratio whose magnitude is the ratio of the major to the minor axis.

A wave is elliptically polarized if it is not linearly or circularly polarized. Although linear and circular polarizations are special cases of elliptical, usually in practice elliptical polarization refers to other than linear or circular. The *necessary and sufficient* conditions to accomplish this are if the field vector (electric or magnetic) possesses *all* of the following:

1. The field must have two orthogonal linear components, and
2. The two components can be of the same or different magnitude.

3. (a) If the two components are not of the same magnitude, the time-phase difference between the two components must not be 0° or multiples of 180° (because it will then be linear). (b) If the two components are of the same magnitude, the time-phase difference between the two components must not be odd multiples of 90° (because it will then be circular).

If the wave is elliptically polarized with two components not of the same magnitude but with odd multiples of 90° time-phase difference, the polarization ellipse will not be tilted but it will be aligned with the principal axes of the field components. The major axis of the ellipse will align with the axis of the field component that is the larger of the two, while the minor axis of the ellipse will align with the axis of the field component that is the smaller of the two.

For elliptical polarization of a wave traveling along the negative z axis, the curve traced at a given z position as a function of time is, in general, a tilted ellipse, as shown in Figure 1.15(b). The ratio of the major axis to the minor axis is referred to as the *axial ratio* (AR), and it is equal to

$$\text{AR} = \frac{\text{major axis}}{\text{minor axis}} = \frac{OA}{OB}, \quad 1 \le \text{AR} \le \infty \tag{1.34}$$

where

$$OA = \left[\tfrac{1}{2}\{E_{xo}^2 + E_{yo}^2 + [E_{xo}^4 + E_{yo}^4 + 2E_{xo}^2 E_{yo}^2 \cos(2\Delta\phi)]^{1/2}\}\right]^{1/2} \tag{1.34a}$$

$$OB = \left[\tfrac{1}{2}\{E_{xo}^2 + E_{yo}^2 - [E_{xo}^4 + E_{yo}^4 + 2E_{xo}^2 E_{yo}^2 \cos(2\Delta\phi)]^{1/2}\}\right]^{1/2} \tag{1.34b}$$

where E_{xo} and E_{yo} represent, respectively, the maximum magnitudes of the two electric field components while $\Delta\phi$ is the time-phase difference between them. The tilt of the ellipse, *relative to the y axis*, is represented by the angle τ given by

$$\tau = \frac{\pi}{2} - \frac{1}{2}\tan^{-1}\left[\frac{2E_{xo}E_{yo}}{E_{xo}^2 - E_{yo}^2}\cos(\Delta\phi)\right] \tag{1.35}$$

When the ellipse is aligned with the principal axes [$\tau = n\pi/2$, $n = 0, 1, 2, \ldots$], the major (minor) axis is equal to $E_{xo}(E_{yo})$ or $E_{yo}(E_{xo})$ and the axial ratio is equal to E_{xo}/E_{yo} or E_{yo}/E_{xo}.

1.12.2 Polarization Loss Factor and Efficiency

In general, the polarization of the receiving antenna will not be the same as the polarization of the incoming (incident) wave. This is commonly stated as "polarization mismatch." The amount of power extracted by the antenna from the incoming signal will not be maximum because of the polarization loss. Assuming that the electric field of the incoming wave can be written

$$\mathbf{E}_i = \hat{\boldsymbol{\rho}}_w E_i \tag{1.36}$$

where $\hat{\boldsymbol{\rho}}_w$ is the unit vector of the wave, and the polarization of the electric field of the receiving antenna can be expressed as

$$\mathbf{E}_a = \hat{\boldsymbol{\rho}}_a E_a \tag{1.37}$$

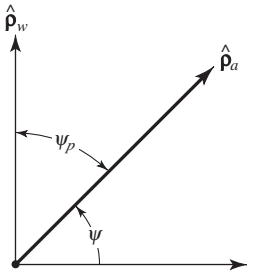


Figure 1.16 Polarization unit vectors of incident wave ($\hat{\boldsymbol{\rho}}_w$) and antenna ($\hat{\boldsymbol{\rho}}_a$), and polarization loss factor (PLF).

where $\hat{\boldsymbol{\rho}}_a$ is its unit vector (polarization vector), the polarization loss can be taken into account by introducing a *polarization loss factor* (PLF). It is defined, based on the polarization of the antenna in its transmitting mode, as

$$\text{PLF} = |\hat{\boldsymbol{\rho}}_w \cdot \hat{\boldsymbol{\rho}}_a|^2 = |\cos \psi_p|^2 \quad \text{(dimensionless)} \tag{1.38}$$

where ψ_p is the angle between the two unit vectors. The relative alignment of the polarization of the incoming wave and of the antenna is shown in Figure 1.16. If the antenna is polarization matched, its PLF will be unity and the antenna will extract maximum power from the incoming wave.

Another figure of merit that is used to describe the polarization characteristics of a wave and that of an antenna is the *polarization efficiency* (*polarization mismatch* or *loss factor*), which is defined as "the ratio of the power received by an antenna from a given plane wave of arbitrary polarization to the power that would be received by the same antenna from a plane wave of the same power flux density and direction of propagation, whose state of polarization has been adjusted for a maximum received power." This is similar to the PLF and it is expressed as

$$p_e = \frac{|\boldsymbol{\ell}_e \cdot \mathbf{E}^{\text{inc}}|^2}{|\boldsymbol{\ell}_e|^2 |\mathbf{E}^{\text{inc}}|^2} \tag{1.39}$$

where

$\boldsymbol{\ell}_e$ = vector effective length of the antenna
$\mathbf{E}^{\text{inc}}$ = incident electric field

The vector effective length $\boldsymbol{\ell}_e$ of the antenna has not yet been defined, and it is introduced in Section 1.15. It is a vector that describes the polarization characteristics of the antenna. Both the PLF and p_e lead to the same answers.

The conjugate (*) is not used in Eq. (1.38) or (1.39) so that a right-hand circularly polarized incident wave (when viewed in its direction of propagation) is matched to a

right-hand circularly polarized receiving antenna (when its polarization is determined in the transmitting mode). Similarly, a left-hand circularly polarized wave will be matched to a left-hand circularly polarized antenna.

Based on the definitions of the wave transmitted and received by an antenna, the polarization of an antenna in the *receiving* mode is related to that in the *transmitting* mode as follows:

1. "In the same plane of polarization, the polarization ellipses have the same axial ratio, the same sense of polarization (rotation) and the same spatial orientation.
2. "Since their senses of polarization and spatial orientation are specified by viewing their polarization ellipses in the respective directions in which they are propagating, one should note that:
   (a) Although their senses of polarization are the same, they would appear to be opposite if both waves were viewed in the same direction.
   (b) Their tilt angles are such that they are the negative of one another with respect to a common reference."

Since the polarization of an antenna will almost always be defined in its transmitting mode, according to the IEEE Std 145-1983, "the receiving polarization may be used to specify the polarization characteristic of a nonreciprocal antenna which may transmit and receive arbitrarily different polarizations."

The polarization loss must always be taken into account in the link calculations design of a communication system because in some cases it may be a very critical factor. Link calculations of communication systems for outer space explorations are very stringent because of limitations in spacecraft weight. In such cases, power is a limiting consideration. The design must properly take into account all loss factors to ensure a successful operation of the system.

An antenna that is elliptically polarized is that composed of two crossed dipoles, as shown in Figure 1.17. The two crossed dipoles provide the two orthogonal field components that are not necessarily of the same field intensity toward all observation angles. If the two dipoles are identical, the field intensity of each along zenith (perpendicular to the plane of the two dipoles) would be of the same intensity. Also, if the two dipoles were fed with a 90° time-phase difference (phase quadrature), the polarization along zenith would be circular and elliptical toward other directions. One way to obtain the 90° time-phase difference $\Delta\phi$ between the two orthogonal field components, radiated respectively by the two dipoles, is by feeding one of the two dipoles with a transmission line that is $\lambda/4$ longer or shorter than that of the other ($\Delta\phi = k\,\Delta\ell = (2\pi/\lambda)(\lambda/4) = \pi/2$). One of the lengths (longer or shorter) will provide right-hand (CW) rotation while the other will provide left-hand (CCW) rotation.

1.13 INPUT IMPEDANCE

Input impedance is defined as "the impedance presented by an antenna at its terminals or the ratio of the voltage to current at a pair of terminals or the ratio of the appropriate components of the electric to magnetic fields at a point." In this section we are primarily interested in the input impedance at a pair of terminals that are the input terminals of the antenna. In Figure 1.18a these terminals are designated as $a-b$. The ratio of the

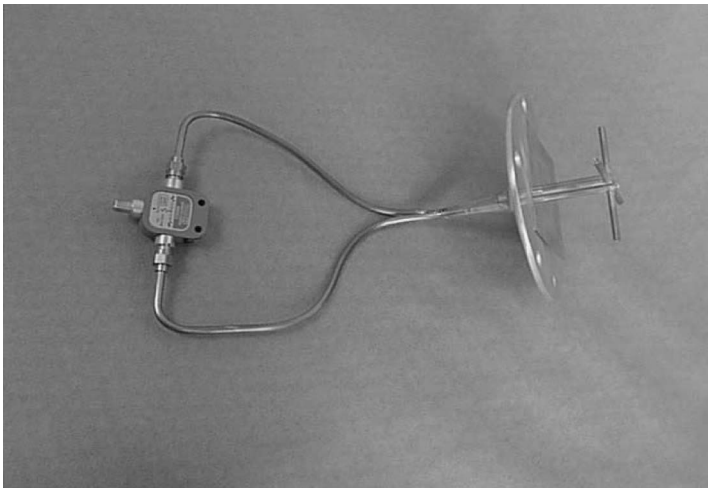


Figure 1.17 Geometry of elliptically polarized cross-dipole antenna.

voltage to current at these terminals, with no load attached, defines the impedance of the antenna as

$$\boxed{Z_A = R_A + jX_A} \tag{1.40}$$

where

Z_A = antenna impedance at terminals $a-b$ (ohms)
R_A = antenna resistance at terminals $a-b$ (ohms)
X_A = antenna reactance at terminals $a-b$ (ohms)

In general, the resistive part of Eq. (1.40) consists of two components; that is,

$$\boxed{R_A = R_r + R_L} \tag{1.41}$$

where

R_r = radiation resistance of the antenna
R_L = loss resistance of the antenna

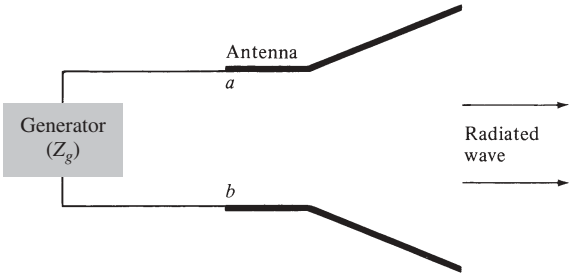
The radiation resistance is used to represent the power delivered to the antenna for radiation.

If we assume that the antenna is attached to a generator with internal impedance

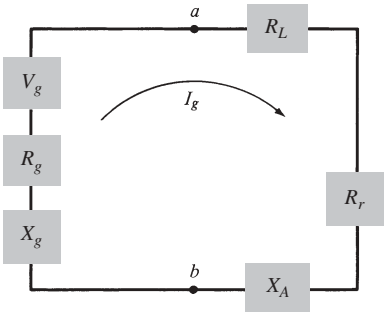
$$Z_g = R_g + jX_g \tag{1.42}$$

where

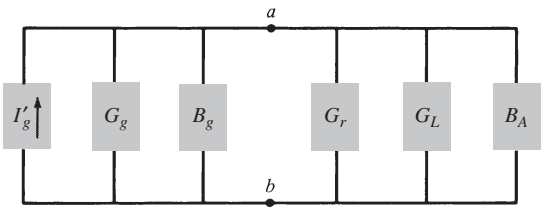
R_g = resistance of generator impedance (ohms)
X_g = reactance of generator impedance (ohms)



(a) Antenna in transmitting mode



(b) Thevenin equivalent



(c) Norton equivalent

Figure 1.18 Transmitting antenna and its equivalent circuits.

and the antenna is used in the transmitting mode, we can represent the antenna and generator by an equivalent circuit† shown in Figure 1.18b.

The maximum power delivered to the antenna occurs when we have conjugate matching; that is, when

$$R_r + R_L = R_g \tag{1.43a}$$

$$X_A = -X_g \tag{1.43b}$$

Under conjugate matching, of the power that is provided by the generator, half is dissipated as heat in the internal resistance (R_g) of the generator and the other half is delivered

†This circuit can be used to represent small and simple antennas. It cannot be used for antennas with lossy dielectric or antennas over lossy ground because their loss resistance cannot be represented in series with the radiation resistance.

to the antenna. This only happens when we have *conjugate matching*. Of the power that is delivered to the antenna, part is radiated through the mechanism provided by the radiation resistance and the other is dissipated as heat, which influences part of the overall efficiency of the antenna. If the antenna is lossless and matched to the transmission line ($e_o = 1$), then half of the total power supplied by the generator is radiated by the antenna during conjugate matching, and the other half is dissipated as heat in the generator. Thus to radiate half of the available power through R_r you must dissipate the other half as heat in the generator through R_g. These two powers are, respectively, analogous to the power transferred to the load and the power scattered by the antenna in the receiving mode. In Figure 1.18 it is assumed that the generator is directly connected to the antenna. If there is a transmission line between the two, which is usually the case, then Z_g represents the equivalent impedance of the generator transferred to the input terminals of the antenna using the impedance transfer equation. If, in addition, the transmission line is lossy, then the available power to be radiated by the antenna will be reduced by the losses of the transmission line. Figure 1.18c illustrates the Norton equivalent of the antenna and its source in the transmitting mode.

The input impedance of an antenna is generally a function of frequency. Thus the antenna will be matched to the interconnecting transmission line and other associated equipment only within a bandwidth. In addition, the input impedance of the antenna depends on many factors including its geometry, its method of excitation, and its proximity to surrounding objects. Because of their complex geometries, only a limited number of practical antennas have been investigated analytically. For many others, the input impedance has been determined experimentally.

1.14 ANTENNA RADIATION EFFICIENCY

The antenna efficiency that takes into account the reflection, conduction, and dielectric losses was discussed in Section 1.8. The conduction and dielectric losses of an antenna are very difficult to compute and in most cases they are measured. Even with measurements, they are difficult to separate and they are usually lumped together to form the e_{cd} efficiency. The resistance R_L is used to represent the conduction–dielectric losses.

The *conduction–dielectric efficiency* e_{cd} is defined as *the ratio of the power delivered to the radiation resistance* R_r *to the power delivered to* R_r *and* R_L.

The radiation efficiency can be written

$$\boxed{e_{cd} = \frac{R_r}{R_L + R_r} \quad \text{(dimensionless)}} \tag{1.44}$$

For a metal rod of length l and uniform cross-sectional area A, the dc resistance is given by

$$R_{\text{dc}} = \frac{1}{\sigma}\frac{l}{A} \quad \text{(ohms)} \tag{1.45a}$$

If the skin depth δ ($\delta = \sqrt{2/\omega\mu_0\sigma}$) of the metal is very small compared to the smallest diagonal of the cross section of the rod, the current is confined to a thin layer near the

conductor surface. Therefore the high frequency resistance, based on a *uniform current distribution*, can be written

$$R_{\text{hf}} = \frac{l}{P} R_s = \frac{l}{P}\sqrt{\frac{\omega\mu_0}{2\sigma}} \quad \text{(ohms)} \tag{1.45b}$$

where P is the perimeter of the cross section of the rod ($P = C = 2\pi b$ for a circular wire of radius b), R_s is the conductor surface resistance, ω is the angular frequency, μ_0 is the permeability of free space, and σ is the conductivity of the metal. For a $\lambda/2$ dipole with a sinusoidal current distribution $R_L = \frac{1}{2}R_{\text{hf}}$, where R_{hf} is given by Eq. (1.45b).

1.15 ANTENNA VECTOR EFFECTIVE LENGTH AND EQUIVALENT AREAS

An antenna in the receiving mode, whether it is in the form of a wire, horn, aperture, array, or dielectric rod, is used to capture (collect) electromagnetic waves and to extract power from them, as shown in Figures 1.19. For each antenna, an equivalent length and a number of equivalent areas can then be defined.

These equivalent quantities are used to describe the receiving characteristics of an antenna, whether it be a linear or an aperture type, when a wave is incident on the antenna.

1.15.1 Vector Effective Length

The effective length of an antenna, whether it be a linear or an aperture antenna, is a quantity that is used to determine the voltage induced on the open-circuit terminals of the antenna when a wave impinges on it. The vector effective length $\boldsymbol{\ell}_e$ for an antenna is usually a complex vector quantity represented by

$$\boldsymbol{\ell}_e(\theta, \phi) = \hat{\mathbf{a}}_\theta l_\theta(\theta, \phi) + \hat{\mathbf{a}}_\phi l_\phi(\theta, \phi) \tag{1.46}$$

It should be noted that it is also referred to as the *effective height*. It is a far-field quantity and it is related to the *far-zone* field $\mathbf{E}_a$ radiated by the antenna, with current I_{in} in its terminals, by [13–18]

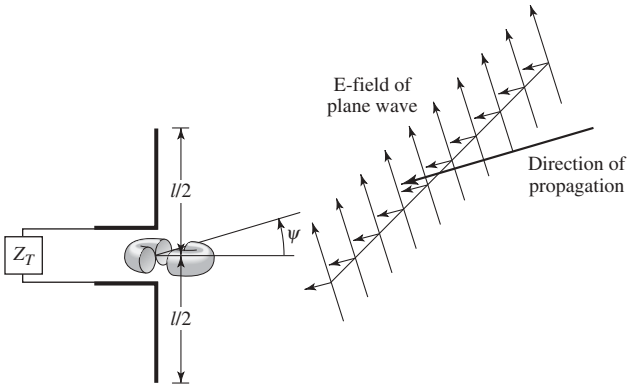
$$\mathbf{E}_a = \hat{\mathbf{a}}_\theta E_\theta + \hat{\mathbf{a}}_\phi E_\phi = -j\eta\frac{kI_{\text{in}}}{4\pi r}\boldsymbol{\ell}_e e^{-jkr} \tag{1.47}$$

The effective length represents the antenna in its transmitting and receiving modes, and it is particularly useful in relating the open-circuit voltage V_{oc} of receiving antennas. This relation can be expressed as

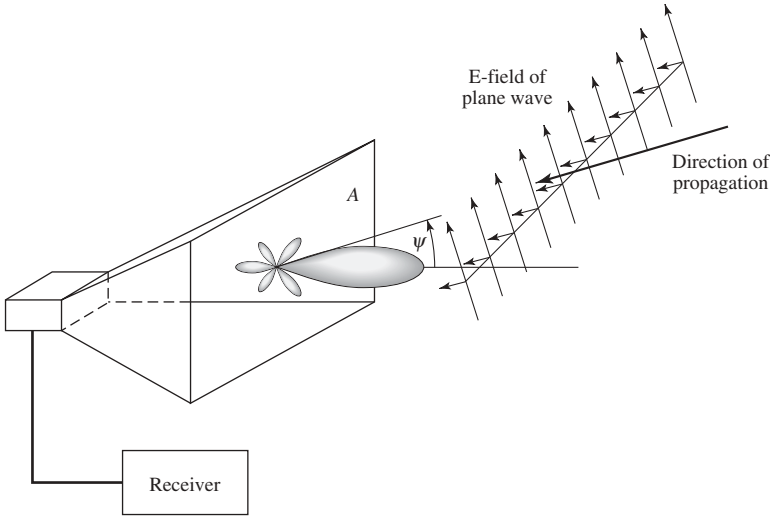
$$V_{\text{oc}} = \mathbf{E}^i \cdot \boldsymbol{\ell}_e \tag{1.48}$$

where

V_{oc} = open-circuit voltage at antenna terminals
$\mathbf{E}^i$ = incident electric field
$\boldsymbol{\ell}_e$ = vector effective length



(a) Dipole antenna in receiving mode



(b) Aperture antenna in receiving mode

Figure 1.19 Uniform plane wave incident on dipole and aperture antennas.

In Eq. (1.48) V_{oc} can be thought of as the voltage induced in a linear antenna of length $\boldsymbol{\ell}_e$ when $\boldsymbol{\ell}_e$ and $\mathbf{E}^i$ are linearly polarized [19, 20]. From relation Eq. (1.48) the *effective length of a linearly polarized antenna receiving a plane wave in a given direction* is defined as "the ratio of the magnitude of the open-circuit voltage developed at the terminals of the antenna to the magnitude of the electric-field strength in the direction of the antenna polarization. Alternatively, the effective length is the length of a thin straight conductor oriented perpendicular to the given direction and parallel to the antenna polarization, having a uniform current equal to that at the antenna terminals and producing the same far-field strength as the antenna in that direction."

In addition, as shown in Section 1.12.2, the antenna vector effective length is used to determine the polarization efficiency of the antenna.

1.15.2 Antenna Equivalent Areas

With each antenna, we can associate a number of equivalent areas. These are used to describe the power capturing characteristics of the antenna when a wave impinges on it. One of these equivalent areas is the *effective area (aperture)*, which in a given direction is defined as "the ratio of the available power at the terminals of a receiving antenna to the power flux density of a plane wave incident on the antenna from that direction, the wave being polarization-matched to the antenna. If the direction is not specified, the direction of maximum radiation intensity is implied." In equation form it is written as

$$A_e = \frac{P_T}{W_i} = \frac{|I_T|^2 R_T/2}{W_i} \tag{1.49}$$

where

A_e = effective area (effective aperture) (m^2)
P_T = power delivered to the load (W)
W_i = power density of incident wave (W/m^2)

The effective aperture is the area that when multiplied by the incident power density gives the power delivered to the load. We can write Eq. (1.49) as

$$A_e = \frac{|V_T|^2}{2W_i}\left(\frac{R_T}{(R_r + R_L + R_T)^2 + (X_A + X_T)^2}\right) \tag{1.50}$$

Under conditions of maximum power transfer (*conjugate matching*), $R_r + R_L = R_T$ and $X_A = -X_T$, the effective area of Eq. (1.50) reduces to the maximum effective aperture given by

$$A_{em} = \frac{|V_T|^2}{8W_i}\left(\frac{R_T}{(R_L + R_r)^2}\right) = \frac{|V_T|^2}{8W_i}\left(\frac{1}{R_r + R_L}\right) \tag{1.51}$$

1.16 MAXIMUM DIRECTIVITY AND MAXIMUM EFFECTIVE AREA

In general then, the *maximum effective area* (A_{em}) *of any antenna is related to its maximum directivity* (D_0) *by* [1]

$$\boxed{A_{em} = \frac{\lambda^2}{4\pi} D_0} \tag{1.52}$$

Thus when Eq. (1.52) is multiplied by the power density of the incident wave it leads to the maximum power that can be delivered to the load. This assumes that there are no conduction-dielectirc losses (radiation efficiency e_{cd} is unity), the antenna is matched to the load (reflection efficiency e_r is unity), and the polarization of the impinging wave matches that of the antenna (polarization loss factor PLF and polarization efficiency p_e are unity). If there are losses associated with an antenna, its maximum effective aperture of Eq. (1.52) must be modified to account for conduction-dielectric losses (radiation efficiency). Thus

$$A_{em} = e_{cd}\left(\frac{\lambda^2}{4\pi}\right) D_0 \tag{1.53}$$

The maximum value of Eq. (1.53) assumes that the antenna is matched to the load and the incoming wave is polarization-matched to the antenna. If reflection and polarization losses are also included, then the maximum effective area of Eq. (1.53) is represented by

$$\begin{aligned} A_{em} &= e_0 \left(\frac{\lambda^2}{4\pi}\right) D_0 |\hat{\boldsymbol{\rho}}_w \cdot \hat{\boldsymbol{\rho}}_a|^2 \\ &= e_{cd}(1 - |\Gamma|^2) \left(\frac{\lambda^2}{4\pi}\right) D_0 |\hat{\boldsymbol{\rho}}_w \cdot \hat{\boldsymbol{\rho}}_a|^2 \end{aligned} \tag{1.54}$$

1.17 FRIIS TRANSMISSION EQUATION AND RADAR RANGE EQUATION

The analysis and design of radar and communications systems often require the use of the *Friis transmission equation* and the *radar range equation*. Because of the importance [21] of the two equations, a few pages will be devoted to them.

1.17.1 Friis Transmission Equation

The Friis transmission equation relates the power received P_r to the power transmitted P_t between two antennas separated by a distance $R > 2D^2/\lambda$, where D is the largest dimension of either antenna. Referring to Figure 1.20, we can write the ratio of received power P_r to transmitted power P_t as

$$\frac{P_r}{P_t} = e_t e_r \frac{\lambda^2 D_t(\theta_t, \phi_t) D_r(\theta_r, \phi_r)}{(4\pi R)^2} \tag{1.55}$$

where

e_t = radiation efficiency of transmitting antenna
e_r = radiation efficiency of receiving antenna
D_t = directivity of transmitting antenna
D_r = directivity of receiving antenna

The power received based on Eq. (1.55) assumes that the transmitting and receiving antennas are matched to their respective lines or loads (reflection efficiencies are unity)

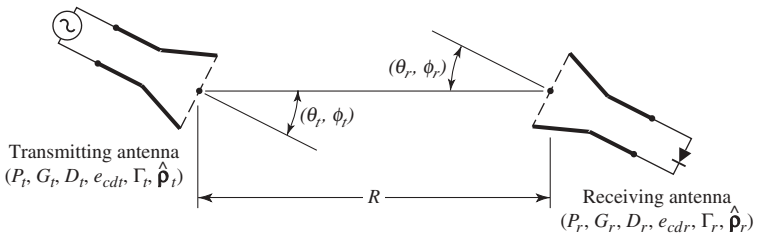


Figure 1.20 Geometrical orientation of transmitting and receiving antennas for Friis transmission equation.

and the polarization of the receiving antenna is polarization-matched to the impinging wave (polarization loss factor and polarization efficiency are unity). If these two factors are also included, then the ratio of the received to the input power of Eq. (1.55) is represented by

$$\boxed{\frac{P_r}{P_t} = e_{cdt}e_{cdr}(1-|\Gamma_t|^2)(1-|\Gamma_r|^2)\left(\frac{\lambda}{4\pi R}\right)^2 D_t(\theta_t,\phi_t)D_r(\theta_r,\phi_r)|\hat{\boldsymbol{\rho}}_t\cdot\hat{\boldsymbol{\rho}}_r|^2} \tag{1.56}$$

For reflection and polarization-matched antennas aligned for maximum directional radiation and reception, Eq. (1.56) reduces to

$$\boxed{\frac{P_r}{P_t} = \left(\frac{\lambda}{4\pi R}\right)^2 G_{0t}G_{0r}} \tag{1.57}$$

Equation Eq. (1.55) or (1.56), or (1.57) is known as the *Friis transmission equation*, and it relates the power P_r (delivered to the receiver load) to the input power of the transmitting antenna P_t. The term $(\lambda/4\pi R)^2$ is called the *free-space loss factor*, and it takes into account the losses due to the spherical spreading of the energy by the antenna.

1.17.2 Radar Range Equation

Now let us assume that the transmitted power is incident on a target, as shown in Figure 1.21. We now introduce a quantity known as the *radar cross section* or *echo area* (σ) of a target, which is defined as *the area intercepting that amount of power which, when scattered isotropically, produces at the receiver a density that is equal to that scattered by the actual target* [1]. In equation form,

$$\lim_{R\to\infty}\left(\frac{\sigma W_i}{4\pi R^2}\right) = W_s \tag{1.58}$$

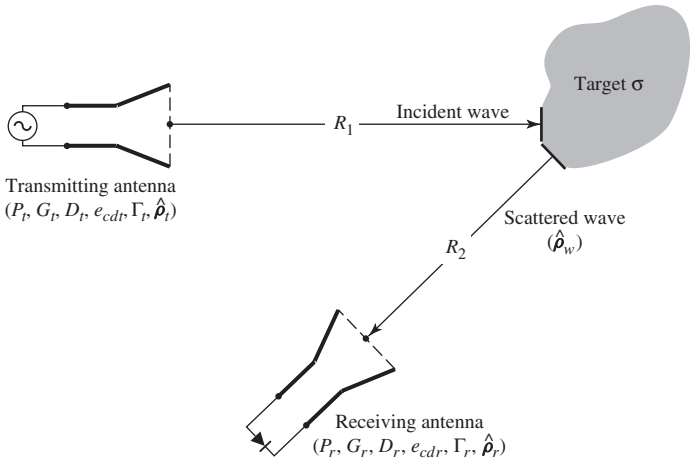


Figure 1.21 Geometrical arrangement of transmitter, target, and receiver for radar range equation.

or

$$\boxed{\begin{aligned}\sigma &= \lim_{R\to\infty}\left(4\pi R^2\frac{W_s}{W_i}\right) = \lim_{R\to\infty}\left(4\pi R^2\frac{|\mathbf{E}^s|^2}{|\mathbf{E}^i|^2}\right)\\ &= \lim_{R\to\infty}\left(4\pi R^2\frac{|\mathbf{H}^s|^2}{|\mathbf{H}^i|^2}\right)\end{aligned}} \tag{1.59}$$

where

σ = radar cross section or echo area (m^2)
R = observation distance from target (m)
W_i = incident power density (W/m^2)
W_s = scattered power density (W/m^2)
$\mathbf{E}^i(\mathbf{E}^s)$ = incident (scattered) electric field (V/m)
$\mathbf{H}^i(\mathbf{H}^s)$ = incident (scattered) magnetic field (A/m)

Any of the definitions in Eq. (1.59) can be used to derive the radar cross section of any antenna or target. For some polarization one of the definitions based either on the power density, electric field, or magnetic field may simplify the derivation, although all should give the same answers [13].

The ratio of received power P_r to transmitted power P_t, that has been scattered by the target with a radar cross section of σ, can be written

$$\boxed{\frac{P_r}{P_t} = e_{cdt}e_{cdr}\sigma\frac{D_t(\theta_t,\phi_t)D_r(\theta_r,\phi_r)}{4\pi}\left(\frac{\lambda}{4\pi R_1R_2}\right)^2} \tag{1.60}$$

Expression (1.60) is used to relate the received power to the input power, and it takes into account only conduction–dielectric losses (radiation efficiency) of the transmitting and receiving antennas. It does not include reflection losses (reflection efficiency) and polarization losses (polarization loss factor or polarization efficiency). If these two losses are also included, then Eq. (1.60) must be expressed as

$$\boxed{\begin{aligned}\frac{P_r}{P_t} &= e_{cdt}e_{cdr}(1-|\Gamma_t|^2)(1-|\Gamma_r|^2)\sigma\frac{D_t(\theta_t,\phi_t)D_r(\theta_r,\phi_r)}{4\pi}\\ &\quad\times\left(\frac{\lambda}{4\pi R_1R_2}\right)^2|\hat{\boldsymbol{\rho}}_w\cdot\hat{\boldsymbol{\rho}}_r|^2\end{aligned}} \tag{1.61}$$

where

$\hat{\boldsymbol{\rho}}_w$ = polarization unit vector of the scattered waves
$\hat{\boldsymbol{\rho}}_r$ = polarization unit vector of the receiving antenna

For polarization-matched antennas aligned for maximum directional radiation and reception, Eq. (1.61) reduces to

$$\boxed{\frac{P_r}{P_t} = \sigma\frac{G_{0t}G_{0r}}{4\pi}\left(\frac{\lambda}{4\pi R_1R_2}\right)^2} \tag{1.62}$$

Equation (1.60), or (1.61), or (1.62) is known as the *radar range equation*. It relates the power P_r (delivered to the receiver load) to the input power P_t transmitted by an antenna, after it has been scattered by a target with a radar cross section (echo area) of σ.

1.17.3 Antenna Radar Cross Section

The radar cross section, usually referred to as RCS, is a far-field parameter, which is used to characterize the scattering properties of a radar target. For a target, there is *monostatic* or *backscattering RCS* when the transmitter and receiver of Figure 1.21 are at the same location, and a *bistatic RCS* when the transmitter and receiver are not at the same location. In designing low-observable or low-profile (stealth) targets, it is the parameter that you attempt to minimize. For complex targets (such as aircraft, spacecraft, missiles, ships, tanks, or automobiles) it is a complex parameter to derive. In general, the RCS of a target is a function of the polarization of the incident wave, the angle of incidence, the angle of observation, the geometry of the target, the electrical properties of the target, and the frequency of operation. The units of RCS of three-dimensional targets are meters squared (m^2) or for normalized values decibels per squared meter (dBsm) or RCS per squared wavelength in decibels (RCS/λ^2 in dB). Representative values of some typical targets are shown in Table 1.1 [22]. Although the frequency was not stated [22], these numbers could be representative at X-band.

The RCS of a target can be controlled using primarily two basic methods: *shaping* and the use of *materials*. Shaping is used to attempt to direct the scattered energy toward directions other than the desired. However, for many targets shaping has to be compromised in order to meet other requirements, such as aerodynamic specifications for flying targets. Materials are used to trap the incident energy within the target and to dissipate part of the energy as heat or to direct it toward directions other than the desired. Usually both methods, shaping and materials, are used together in order to optimize the performance of a radar target. One of the "golden rules" to observe in order to achieve low RCS is to "*round corners, avoid flat and concave surfaces, and use material treatment in flare spots*."

TABLE 1.1 RCS of Some Typical Targets

	Typical RCSs [22]	
Object	RCS (m^2)	RCS (dBsm)
Pickup truck	200	23
Automobile	100	20
Jumbo jet airliner	100	20
Large bomber or commercial jet	40	16
Cabin cruiser boat	10	10
Large fighter aircraft	6	7.78
Small fighter aircraft or four-passenger jet	2	3
Adult male	1	0
Conventional winged missile	0.5	−3
Bird	0.01	−20
Insect	0.00001	−50
Advanced tactical fighter	0.000001	−60

There are many methods of analysis to predict the RCS of a target [13, 22–33]. Some of them are full-wave methods, others are designated as asymptotic methods, either low frequency or high frequency, and some are considered as numerical methods. The methods of analysis are often contingent on the shape, size, and material composition of the target. Some targets, because of their geometrical complexity, are often simplified and are decomposed into a number of basic shapes (such as strips, plates, cylinders, cones, wedges), which when put together represent a very good replica of the actual target. This has been used extensively and proved to be a very good approach. The topic is very extensive to be treated here in any detail, and the reader is referred to the literature [13, 22–33]. There are a plethora of references but because of space limitations, only a limited number are included here to get the reader started on the subject. Chapter 21 in this and book is devoted to the antenna scattering and design considerations.

Antennas individually are radar targets that many exhibit large radar cross section. In many applications, antennas are mounted on the surface of other complex targets (such as aircraft, spacecraft, satellites, missiles, or automobiles) and become part of the overall radar target. In such configurations, many antennas, especially aperture types (such as waveguides and horns) become large contributors to the total RCS, monostatic or bistatic, of the target. Therefore in designing low-observable targets, the antenna type, location, and contributions become important considerations of the overall design.

The scattering and transmitting (radiation) characteristics of an antenna are related [34–36]. There are various methods that can be used to analyze the fields scattered by an antenna. The presentation here parallels that in Refs. 23 and 37–40. In general, the electric field scattered by an antenna with a load impedance Z_L can be expressed by

$$\mathbf{E}^s(Z_L) = \mathbf{E}^s(0) - \frac{I_s}{I_t}\frac{Z_L}{Z_L + Z_A}\mathbf{E}^t \tag{1.63}$$

where

$\mathbf{E}^s(Z_L)$ = electric field scattered by antenna with a load Z_L
$\mathbf{E}^s(0)$ = electric field scattered by short-circuited antenna ($Z_L = 0$)
I_s = short-circuited current induced by the incident field on the antenna with $Z_L = 0$
I_t = antenna current in transmitting mode
$Z_A = R_A + jX_A$ = antenna input impedance
$\mathbf{E}^t$ = electric field radiated by the antenna in transmitting mode

By defining an antenna reflection coefficient of

$$\Gamma_A = \frac{Z_L - Z_A}{Z_L + Z_A} \tag{1.64}$$

the scattered field of Eq. (1.63) can be written

$$\mathbf{E}^s(Z_L) = \mathbf{E}^s(0) - \frac{I_s}{I_t}\frac{1}{2}(1 + \mathbf{\Gamma}_A)\mathbf{E}^t \tag{1.65}$$

Therefore according to Eq. (1.65) the scattered field by an antenna with a load Z_L is equal to the scattered field when the antenna is short-circuited ($Z_L = 0$) minus a term related to the reflection coefficient and the field transmitted by the antenna.

Green has expressed the field scattered by an antenna terminated with a load Z_L in a more convenient form that allows it to be separated into the *structural* and *antenna mode* scattering terms [23, 37–40]. This is accomplished by assuming that the antenna is loaded with a conjugate-matched impedance ($Z_L = Z_A^*$). Doing this and using Eq. (1.63) generates another equation for the field scattered by the antenna with a load $Z_L = Z_A^*$. When this new equation is subtracted from Eq. (1.63) it eliminates the short-circuited scattered field, and we can write that the field scattered by the antenna with a load Z_L is

$$\mathbf{E}^s(Z_L) = \mathbf{E}^s(Z_A^*) - \frac{I_s}{I_t}\frac{\boldsymbol{\Gamma}^* Z_A}{2R_A}\mathbf{E}^t \tag{1.66}$$

$$\boldsymbol{\Gamma}^* = \frac{Z_L - Z_A^*}{Z_L + Z_A^*} \tag{1.66a}$$

where

$\mathbf{E}^s(Z_L)$ = electric field scattered by the antenna with load Z_L
$\mathbf{E}^s(Z_A^*)$ = electric field scattered by the antenna with a conjugate-matched load
$I(Z_A^*)$ = current induced by the incident wave at the antenna terminals matched with a conjugate impedance load
Γ^* = conjugate-matched reflection coefficient
Z_L = load impedance attached to antenna terminals

For the short-circuited case and the conjugate-matched transmitting (radiating) case, the product of their currents and antenna impedance are related by [34]

$$I_s Z_A = I_m^*(Z_A + Z_A^*) = 2R_A I_m^* \tag{1.67}$$

where I_m^* is the scattering current when the antenna is conjugate-matched ($Z_L = Z_A^*$). Substituting Eq. (1.67) into (1.66) for I_s reduces Eq. (1.66) to

$$\mathbf{E}^s(Z_L) = \mathbf{E}^s(Z_A^*) - \frac{I_m^*}{I_t}\boldsymbol{\Gamma}^*\mathbf{E}^t \tag{1.68}$$

It can also be shown that if the antenna is matched with a load Z_A (instead of Z_A^*), then Eq. (1.68) can be written

$$\mathbf{E}^s(Z_L) = \mathbf{E}^s(Z_A) - \frac{I_m}{I_t}\Gamma_A\mathbf{E}^t \tag{1.69}$$

Therefore the field scattered by an antenna loaded with an impedance Z_L is related to the field radiated by the antenna in the transmitting mode in three different ways, as shown by Eqs. (1.65), (1.68), and (1.69). According to Eq. (1.65) the field scattered by an antenna when it is loaded with an impedance Z_L is equal to the field scattered by the antenna when it is short-circuited ($Z_L = 0$) minus a term related to the antenna reflection coefficient and the field transmitted by the antenna. In addition, according to Eq. (1.68), the field scattered by an antenna when it is terminated with an impedance Z_L is equal

to the field scattered by the antenna when it is conjugate-matched with an impedance Z_A^* minus the field transmitted (radiated) times the conjugate reflection coefficient. The second term is weighted by the two currents. Alternatively, according to Eq. (1.69), the field scattered by the antenna when it is terminated with an impedance Z_L is equal to the field scattered by the antenna when it is matched with an impedance Z_A minus the field transmitted (radiated) times the reflection coefficient weighted by the two currents.

In Eq. (1.68) the first term consists of the *structural* scattering term and the second of the *antenna mode* scattering term. The *structural* scattering term is introduced by the currents induced on the surface of the antenna by the incident field when the antenna is conjugate-matched, and it is independent of the load impedance. The *antenna mode* scattering term is only a function of the radiation characteristics of the antenna, and its scattering pattern is the square of the antenna radiation pattern. The antenna mode depends on the power absorbed in the load of a lossless antenna and the power that is radiated by the antenna due to a load mismatch. This term vanishes when the antenna is conjugate-matched.

From the scattered field expression of Eq. (1.65), it can be shown that the total radar cross section of the antenna terminated with a load Z_L can be written as [40]

$$\sigma = |\sqrt{\sigma^s} - (1 + \Gamma_A)\sqrt{\sigma^a} e^{j\phi_r}|^2 \tag{1.70}$$

where

σ = total RCS with antenna terminated with Z_L
σ^s = RCS due to structural term
σ^a = RCS due to antenna mode term
ϕ_r = relative phase between the structural and antenna mode terms

If the antenna is short-circuited ($\Gamma_A = -1$), then according to Eq. (1.70)

$$\sigma_{\text{short}} = \sigma^s \tag{1.71}$$

If the antenna is open-circuited ($\Gamma_A = +1$), then according to Eq. (1.70)

$$\sigma_{\text{open}} = |\sqrt{\sigma^s} - 2\sqrt{\sigma^a} e^{j\phi_r}|^2 = \sigma_{\text{residual}} \tag{1.72}$$

Lastly, if the antenna is matched $Z_L = Z_A (\Gamma_A = 0)$, then according to (1.70)

$$\sigma_{\text{match}} = |\sqrt{\sigma^s} - \sqrt{\sigma^a} e^{j\phi_r}|^2 \tag{1.73}$$

Therefore under matched conditions, according to Eq. (1.73), the range of values (minimum to maximum) of the radar cross section is

$$|\sigma^s - \sigma^a| \le \sigma \le |\sigma^s + \sigma^a| \tag{1.74}$$

The minimum value occurs when the two RCSs are in phase while the maximum occurs when they are out of phase.

To produce a zero RCS, Eq. (1.70) must vanish. This is accomplished if

$$\mathrm{Re}(\Gamma_A) = -1 + \cos\phi_r \sqrt{\sigma^s/\sigma^a} \tag{1.75a}$$

$$\mathrm{Im}(\Gamma_A) = -\sin\phi_r \sqrt{\sigma^s/\sigma^a} \tag{1.75b}$$

Assuming positive values of resistances, the real value of Γ_A cannot be greater than unity. Thus there are some cases where the RCS cannot be reduced to zero by choosing Z_L. Because Z_A can be complex, there is no limit on the imaginary part of Γ_A.

In general, the structural and antenna mode scattering terms are very difficult to predict and usually require that the antenna is solved as a boundary-value problem. However, these two terms have been obtained experimentally utilizing the Smith chart [37–39].

For a monostatic system the receiving and transmitting antennas are collocated. In addition, if the antennas are identical ($G_{0r} = G_{0t} = G_0$) and are polarization-matched ($P_r = P_t = 1$), the total radar cross section of the antenna for backscattering can be written as

$$\sigma = \frac{\lambda_0^2}{4\pi} G_0^2 |A - \Gamma^*|^2 \tag{1.76}$$

where A is a complex parameter independent of the load.

If the antenna is a thin dipole, then $A \simeq 1$ and Eq. (1.76) reduces to

$$\begin{aligned} \sigma &\simeq \frac{\lambda_0^2}{4\pi} G_0^2 |1 - \Gamma^*|^2 = \frac{\lambda_0^2}{4\pi} G_0^2 \left| 1 - \frac{Z_L - Z_A^*}{Z_L + Z_A} \right| \\ &= \frac{\lambda_0^2}{4\pi} G_0^2 \left| \frac{2R_A}{Z_L + Z_A} \right|^2 \end{aligned} \tag{1.77}$$

If in addition we assume that the dipole length is $l = \lambda_0/2$ and is short-circuited ($Z_L = 0$), then the normalized radar cross section of Eq. (1.77) is equal to

$$\frac{\sigma}{\lambda_0^2} \simeq \frac{G_0^2}{\pi} = \frac{(1.643)^2}{\pi} = 0.8593 \simeq 0.86 \tag{1.78}$$

which agrees with the experimental corresponding maximum monostatic value of Figure 1.22 and those reported in the literature [41, 42].

Shown in Figure 1.22 is the measured E-plane monostatic RCS of a half-wavelength dipole when it is matched to a load, short-circuited (straight wire), and open-circuited (gap at the feed). The aspect angle is measured from the normal to the wire. As expected, the RCS is a function of the observation (aspect) angle. Also it is apparent that there are appreciable differences between the three responses. For the short-circuited case, the maximum value is approximately -24 dBsm, which closely agrees with the computed value of -22.5 dBsm using Eq. (1.78). Similar responses for the monostatic RCS of a pyramidal horn are shown in Figure 1.23(a) for the E-plane and in Figure 1.23(b) for the H-plane. The antenna is a commercial X-band (8.2–12.4 GHz) 20-dB standard gain horn with aperture dimension of 9.2 cm by 12.4 cm. The length of the horn is 25.6 cm. As for the dipole, there are differences between the three responses for each plane. It is seen that the short-circuited response exhibits the largest return.

Antenna RCS from model measurements [43] and microstrip patches [44, 45] have been reported.

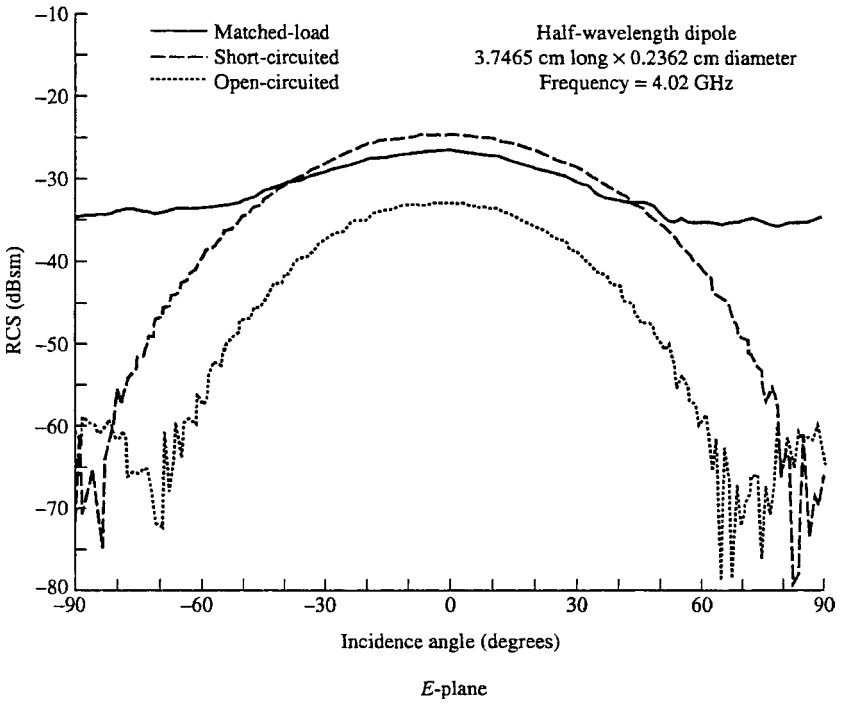


Figure 1.22 E-plane monostatic RCS ($\sigma_{\theta\theta}$) versus incidence angle for a half-wavelength dipole.

1.18 ANTENNA TEMPERATURE

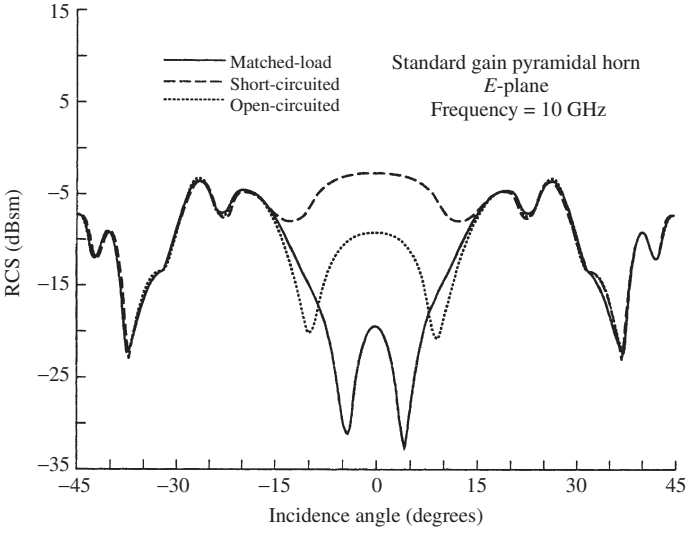
Every object with a physical temperature above absolute zero (0 K $= -273^{\circ}$C) radiates energy [8]. The amount of energy radiated is usually represented by an equivalent temperature T_B, better known as brightness temperature, and it is defined as

$$T_B(\theta, \phi) = \epsilon(\theta, \phi)T_m = (1 - |\Gamma|^2)T_m \tag{1.79}$$

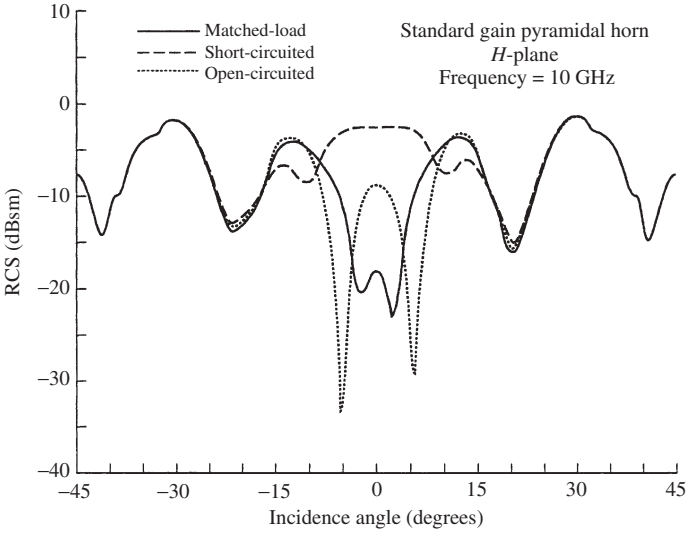
where

T_B = brightness temperature (equivalent temperature; K)
ϵ = emissivity (dimensionless)
T_m = molecular (physical) temperature (K)
$\Gamma(\theta, \phi)$ = reflection coefficient of the surface for the polarization of the wave

Since the values of emissivity are $0 \leq \epsilon \leq 1$, the maximum value the brightness temperature can achieve is equal to the molecular temperature. Usually the emissivity is a function of the frequency of operation, polarization of the emitted energy, and molecular structure of the object. Some of the better natural emitters of energy at microwave frequencies are (1) the ground with equivalent temperature of about 300 K and (2) the



(a) E-plane ($\sigma_{\theta\theta}$)



(b) H-plane ($\sigma_{\phi\phi}$)

Figure 1.23 E- and H-plane monostatic RCS versus incidence angle for a pyramidal horn.

sky with equivalent temperature of about 5 K when looking toward zenith and about 100–150 K toward the horizon.

The brightness temperature emitted by the different sources is intercepted by antennas, and it appears at their terminals as an antenna temperature. The temperature appearing at the terminals of an antenna is that given by Eq. (1.79), after it is weighted by the gain pattern of the antenna. In equation form, this can be written

$$T_A = \frac{\int_0^{2\pi}\int_0^{\pi} T_B(\theta,\phi)G(\theta,\phi)\sin\theta\, d\theta\, d\phi}{\int_0^{2\pi}\int_0^{\pi} G(\theta,\phi)\sin\theta\, d\theta\, d\phi} \tag{1.80}$$

where

T_A = antenna temperature (effective noise temperature of the antenna radiation resistance; K)
$G(\theta,\phi)$ = gain (power) pattern of the antenna

Assuming no losses or other contributions between the antenna and the receiver, the noise power transferred to the receiver is given by

$$P_r = kT_A\,\Delta f \tag{1.81}$$

where

P_r = antenna noise power (W)
k = Boltzmann's constant (1.38×10^{-23} J/K)
T_A = antenna temperature (K)
Δf = bandwidth (Hz)

If the antenna and transmission line are maintained at certain physical temperatures, and the transmission line between the antenna and receiver is lossy, the antenna temperature T_A as seen by the receiver through Eq. (1.81) must be modified to include the other contributions and the line losses. If the antenna itself is maintained at a certain physical temperature T_p and a transmission line of length l, constant physical temperature T_0 throughout its length, and uniform attenuation of α (Np/unit length) is used to connect an antenna to a receiver, as shown in Figure 1.24, the effective antenna temperature at the receiver terminals is given by

$$T_a = T_A e^{-2\alpha l} + T_{AP}e^{-2\alpha l} + T_0(1 - e^{-2\alpha l}) \tag{1.82}$$

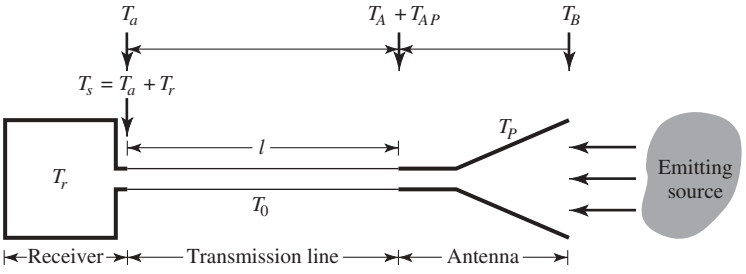


Figure 1.24 Antenna, transmission line, and receiver arrangement for system noise power calculation.

where

$$T_{AP} = \left(\frac{1}{e_A} - 1\right) T_p \tag{1.82a}$$

T_a = antenna temperature at the receiver terminals (K)
T_A = antenna noise temperature at the antenna terminals (Eq. (1.80)) (K)
T_{AP} = antenna temperature at the antenna terminals due to physical temperature (Eq. (1.82a)) (K)
T_p = antenna physical temperature (K)
α = attenuation coefficient of transmission line (Np/m)
e_A = thermal efficiency of antenna (dimensionless)
l = length of transmission line (m)
T_0 = physical temperature of the transmission line (K)

The antenna noise power of Eq. (1.81) must also be modified and written as

$$P_r = kT_a \, \Delta f \tag{1.83}$$

where T_a is the antenna temperature at the receiver input as given by Eq. (1.82).

If the receiver itself has a certain noise temperature T_r (due to thermal noise in the receiver components), the *system noise power at the receiver terminals* is given by

$$P_s = k(T_a + T_r) \, \Delta f = kT_s \, \Delta f \tag{1.84}$$

where

P_s = system noise power (at receiver terminals)
T_a = antenna noise temperature (at receiver terminals)
T_r = receiver noise temperature (at receiver terminals)
$T_s = T_a + T_r$ = effective system noise temperature (at receiver terminals)

A graphical relation of all the parameters is shown in Figure 1.24. The effective system noise temperature T_s of radio astronomy antennas and receivers varies from very few degrees (typically $\simeq 10$ K) to thousands of Kelvins depending on the type of antenna, receiver, and frequency of operation. Antenna temperature changes at the antenna terminals, due to variations in the target emissions, may be as small as a fraction of one degree. To detect such changes, the receiver must be very sensitive and be able to differentiate changes of a fraction of a degree.

A summary of the pertinent parameters and associated formulas and equation numbers for this chapter are listed in Table 1.2.

TABLE 1.2 Summary of Important Parameters and Associated Formulas and Equation Numbers

Parameter	Formula	Equation Number
Infinitesimal area of sphere	$dA = r^2 \sin\theta \, d\theta \, d\phi$	(1.1)
Elemental solid angle of sphere	$d\Omega = \sin\theta \, d\theta \, d\phi$	(1.2)
Average power density	$\mathbf{W}_{\text{av}} = \frac{1}{2}\text{Re}(\mathbf{E} \times \mathbf{H}^*)$	(1.5)
Radiated power/average radiated power	$P_{\text{rad}} = P_{\text{av}} = \oiint_S \mathbf{W}_{\text{av}} \cdot d\mathbf{s} = \frac{1}{2} \oiint_S \text{Re}(\mathbf{E} \times \mathbf{H}^*) \cdot d\mathbf{s}$	(1.6)
Radiation density of isotropic radiator	$W_0 = \dfrac{P_{\text{rad}}}{4\pi r^2}$	
Radiation intensity (far field)	$U = r^2 W_{\text{rad}} = B_0 F(\theta, \phi)$ $\simeq \dfrac{r^2}{2\eta}[\lvert E_\theta(r, \theta, \phi)\rvert^2 + \lvert E_\phi(r, \theta, \phi)\rvert^2]$	(1.7), (1.7a)
Directivity $D(\theta, \phi)$	$D = \dfrac{U}{U_0} = \dfrac{4\pi U}{P_{\text{rad}}} = \dfrac{4\pi}{\Omega_A}$	(1.9), (1.12)
Beam solid angle Ω_A	$\Omega_A = \displaystyle\int_0^{2\pi} \int_0^{\pi} F_n(\theta, \phi) \sin\theta d\theta d\phi$ $F_n(\theta, \phi) = \frac{F(\theta, \phi)}{\lvert F(\theta, \phi)\rvert_{\max}}$	(1.12a) (1.12b)
Maximum directivity D_0	$D_{\max} = D_0 = \dfrac{U_{\max}}{U_0} = \dfrac{4\pi U_{\max}}{P_{\text{rad}}}$	(1.9a)
Partial directivities D_θ,D_ϕ	$D_0 = D_\theta + D_\phi$	(1.10)
	$D_\theta = \dfrac{4\pi U_\theta}{P_{\text{rad}}} = \dfrac{4\pi U_\theta}{(P_{\text{rad}})_\theta + (P_{\text{rad}})_\phi}$	(1.10a)
	$D_\phi = \dfrac{4\pi U_\phi}{P_{\text{rad}}} = \dfrac{4\pi U_\phi}{(P_{\text{rad}})_\theta + (P_{\text{rad}})_\phi}$	(1.10b)
Approximate maximum directivity (*one main lobe pattern*)	$D_0 \simeq \dfrac{4\pi}{\Theta_{1r}\Theta_{2r}} = \dfrac{41,253}{\Theta_{1d}\Theta_{2d}}$ (Kraus)	(1.13), (1.14)
	$D_0 \simeq \dfrac{32 \ln 2}{\Theta_{1r}^2 + \Theta_{2r}^2} = \dfrac{22.181}{\Theta_{1r}^2 + \Theta_{2r}^2} = \dfrac{72,815}{\Theta_{1d}^2 + \Theta_{2d}^2}$ (Tai–Pereira)	(1.17a), (1.17b)
Approximate maximum directivity (*omnidirectional pattern*)	$D_0 \simeq \dfrac{101}{\text{HPBW (degrees)} - 0.0027[\text{HPBW (degrees)}]^2}$ (McDonald)	(1.18a)
	$D_0 \simeq -172.4 + 191\sqrt{0.818 + \dfrac{1}{\text{HPBW (degrees)}}}$ (Pozar)	(1.18b)

TABLE 1.2 (*Continued*)

Parameter	Formula	Equation Number
Gain $G(\theta,\phi)$	$G = \dfrac{4\pi U(\theta,\phi)}{P_{\text{in}}} = e_{cd}\left(\dfrac{4\pi U(\theta,\phi)}{P_{\text{rad}}}\right) = e_{cd}D(\theta,\phi)$ $P_{\text{rad}} = e_{cd}P_{\text{in}}$	(1.24), (1.27), (1.25)
Antenna radiation efficiency e_{cd}	$e_{cd} = \dfrac{R_r}{R_r + R_L}$	(1.44)
Loss resistance R_L (*straight wire/uniform current*)	$R_L = R_{\text{hf}} = \dfrac{l}{P}\sqrt{\dfrac{\omega\mu_0}{2\sigma}}$	(1.45b)
Loss resistance R_L (*straight wire/λ/2 dipole*)	$R_L = \dfrac{l}{2P}\sqrt{\dfrac{\omega\mu_0}{2\sigma}}$	
Maximum gain G_0	$G_0 = e_{cd}D_{\max} = e_{cd}D_0$	(1.27a)
Partial gains G_θ, G_ϕ	$G_0 = G_\theta + G_\phi$ $G_\theta = \dfrac{4\pi U_\theta}{P_{\text{in}}}, \quad G_\phi = \dfrac{4\pi U_\phi}{P_{\text{in}}}$	(1.29) (1.29a) (1.29b)
Absolute gain G_{abs}	$G_{\text{abs}} = e_r G(\theta,\phi) = e_r e_{cd} D(\theta,\phi)$ $= (1-\lvert\Gamma\rvert^2)e_{cd}D(\theta,\phi) = e_0 D(\theta,\phi)$	(1.28) (1.28b)
Total antenna efficiency e_0	$e_0 = e_r e_c e_d = e_r e_{cd} = (1-\lvert\Gamma\rvert^2)e_{cd}$	(1.22), (1.23)
Reflection efficiency e_r	$e_r = (1-\lvert\Gamma\rvert^2)$	(1.23)
Beam efficiency BE	$\text{BE} = \dfrac{\int_0^{2\pi}\int_0^{\theta_1} U(\theta,\phi)\sin\,\theta\;d\theta\;d\phi}{\int_0^{2\pi}\int_0^{\pi} U(\theta,\phi)\sin\,\theta\;d\theta\;d\phi}$	(1.33)
Polarization loss factor (PLF)	$\text{PLF} = \lvert\hat{\boldsymbol{\rho}}_w \cdot \hat{\boldsymbol{\rho}}_a\rvert^2$	(1.38)
Vector effective length $\boldsymbol{\ell}_e(\theta,\phi)$	$\boldsymbol{\ell}_e(\theta,\phi) = \hat{\mathbf{a}}_\theta l_\theta(\theta,\phi) + \hat{\mathbf{a}}_\phi l_\phi(\theta,\phi)$	(1.46)
Polarization efficiency p_e	$p_e = \dfrac{\lvert\boldsymbol{\ell}_e \cdot \mathbf{E}^{\text{inc}}\rvert^2}{\lvert\boldsymbol{\ell}_e\rvert^2\lvert\mathbf{E}^{\text{inc}}\rvert^2}$	(1.39)
Antenna impedance Z_A	$Z_A = R_A + jX_A = (R_r + R_L) + jX_A$	(1.40), (1.41)
Maximum effective area A_{em}	$A_{em} = \dfrac{\lvert V_T\rvert^2}{8W_i}\left(\dfrac{1}{R_r+R_L}\right) = e_{cd}\left(\dfrac{\lambda^2}{4\pi}\right)D_0\lvert\hat{\boldsymbol{\rho}}_w \cdot \hat{\boldsymbol{\rho}}_a\rvert^2$ $= \left(\dfrac{\lambda^2}{4\pi}\right)G_0\lvert\hat{\boldsymbol{\rho}}_w \cdot \hat{\boldsymbol{\rho}}_a\rvert^2$	(1.51), (1.53), (1.54)

(*continued*)

TABLE 1.2 (*Continued*)

Parameter	Formula	Equation Number
Aperture efficiency ε_{ap}	$\varepsilon_{ap} = \dfrac{A_{em}}{A_p} = \dfrac{\text{maximum effective area}}{\text{physical area}}$	
Friis transmission equation	$\dfrac{P_r}{P_t} = \left(\dfrac{\lambda}{4\pi R}\right)^2 G_{0t} G_{0r} \lvert\hat{\boldsymbol{\rho}}_t \cdot \hat{\boldsymbol{\rho}}_r\rvert^2$	(1.56), (1.57)
Radar range equation	$\dfrac{P_r}{P_t} = \sigma \dfrac{G_{0t} G_{0r}}{4\pi} \left(\dfrac{\lambda}{4\pi R_1 R_2}\right)^2 \lvert\hat{\boldsymbol{\rho}}_w \cdot \hat{\boldsymbol{\rho}}_r\rvert^2$	(1.61), (1.62)
Radar cross section (RCS)	$\sigma = \lim\limits_{R\to\infty}\left(4\pi R^2 \dfrac{W_s}{W_i}\right) = \lim\limits_{R\to\infty}\left(4\pi R^2 \dfrac{\lvert\mathbf{E}^s\rvert^2}{\lvert\mathbf{E}^i\rvert^2}\right) = \lim\limits_{R\to\infty}\left(4\pi R^2 \dfrac{\lvert\mathbf{H}^s\rvert^2}{\lvert\mathbf{H}^i\rvert^2}\right)$	(1.59)
Brightness temperature $T_B(\theta,\phi)$	$T_B(\theta,\phi) = \varepsilon(\theta,\phi) T_m = (1 - \lvert\Gamma\rvert^2) T_m$	(1.79)
Antenna temperature T_A	$T_A = \dfrac{\int_0^{2\pi}\int_0^{\pi} T_B(\theta,\phi) G(\theta,\phi) \sin\theta \, d\theta \, d\phi}{\int_0^{2\pi}\int_0^{\pi} G(\theta,\phi) \sin\theta \, d\theta \, d\phi}$	(1.80)

REFERENCES

1. C. A. Balanis, *Antenna Theory: Analysis and Design* (3rd eds.), John Wiley & Sons, Hoboken, NJ, 2005.
2. W. R. Scott Jr., A general program for plotting three-dimensional antenna patterns, *IEEE Antennas Propag. Soc. Newslett.*, pp. 6–11, December 1989.
3. A. Z. Elsherbeni and C. D. Taylor, Jr., Interactive antenna pattern visualization, *in Software Book in Electromagnetics*, Vol. II, CAEME Center for Multimedia Education, University of Utah, 1995, Chap. 8, pp. 367–410.
4. A. Z. Elsherbeni and P. H. Ginn, Interactive analysis of antenna arrays, *in Software Book in Electromagnetics*, Vol. II, CAEME Center for Multimedia Education, University of Utah, 1995, Chap. 6, pp. 337–366.
5. J. Romeu and R. Pujol, Array, *in Software Book in Electromagnetics*, Vol. II, CAEME Center for Multimedia Education, University of Utah, 1995, Chap. 12, pp. 467–481.
6. J. S. Hollis, T. J. Lyon, and L. Clayton, Jr. (Eds.), *Microwave Antenna Measurements*, Scientific-Atlanta, Inc., July 1970.
7. J. D. Kraus and R. J. Marhefka, *Antennas*, McGraw-Hill, New York, 2002.
8. J. D. Kraus, *Radio Astronomy*, McGraw-Hill, New york 1966.
9. R. S. Elliott, Beamwidth and directivity of large scanning arrays, *Microwave J.*, pp. 74–82, January 1964.
10. C.-T. Tai and C. S. Pereira, An approximate formula for calculating the directivity of an antenna, *IEEE Trans. Antennas Propag.*, Vol. 24, No. 2, pp. 235–236, March 1976.

11. N. A. McDonald, Approximate relationship between directivity and beamwidth for broadside collinear arrays, *IEEE Trans. Antennas Propag.*, Vol. 26, No. 2, pp. 340–341, March 1978.
12. D. M. Pozar, Directivity of omnidirectional antennas, *IEEE Antennas Propag. Mag.*, Vol. 35, No. 5, pp. 50–51, October 1993.
13. C. A. Balanis, *Advanced Engineering Electromagnetics*, John Wiley & Sons, Hoboken, NJ, 1989.
14. H. Poincaré, *Theorie Mathematique de la Limiere*, Georges Carre, Paris, 1892.
15. G. A. Deschamps, Part II—Geometrical representation of the polarization of a plane electromagnetic wave, *Proc. IRE*, Vol. 39, pp. 540–544, May 1951.
16. E. F. Bolinder, Geometrical analysis of partially polarized electromagnetic waves, *IEEE Trans. Antennas Propag.*, Vol. 15, No. 1, pp. 37–40, January 1967.
17. G. A. Deschamps and P. E. Mast, Poincaré sphere representation of partially polarized fields, *IEEE Trans. Antennas Propag.*, Vol. 21, No. 4, pp. 474–478, July 1973.
18. G. Sinclair, The transmission and reflection of elliptically polarized waves, *Proc. IRE*, Vol. 38, pp. 148–151, February 1950.
19. C. A. Balanis, Antenna theory: a review, *Proc. IEEE*, Vol. 80, No. 1, pp. 7–23, January 1992.
20. R. E. Collin, *Antennas and Radiowave Propagation*, McGraw-Hill, New York, 1985.
21. M. I. Skolnik, *Radar Systems*, Chapter 2, McGraw-Hill, New York, 1962.
22. J. A. Adam, How to design an invisible' aircraft, *IEEE Spectrum*, pp. 26–31, April 1988.
23. G. T. Ruck, D. E. Barrick, W. D. Stuart, and C. K. Krichbaum, *Radar Cross Section Handbook*, Vols. 1 and 2, Plenum Press, New York, 1970.
24. M. I. Skolnik (Ed.), *Radar Handbook*, McGraw-Hill, New York, 1970, Chap. 27, Sec. 6, pp. 27-19–27-40.
25. J. W. Crispin, Jr. and K. M. Siegel, *Methods of Radar Cross Section Analysis*, Academic Press, New York, 1968.
26. J. J. Bowman, T. B. A. Senior, and P. L. Uslenghi (Eds.), *Electromagnetic and Acoustic Scattering by Simple Shapes*, North-Holland., Amsterdam, The Netherland: 1969.
27. E. F. Knott, M. T. Turley, and J. F. Shaeffer, *Radar Cross Section*, Artech House, Norwood, MA, 1985.
28. A. K. Bhattacharya and D. L. Sengupta, *Radar Cross Section Analysis and Control*, Artech House, Norwood, MA, 1991.
29. A. F. Maffett, *Topics for a Statistical Description of Radar Cross Section*, John Wiley & Sons, Hoboken, NJ, 1989.
30. Special issue, *Proc. IEEE*, Vol. 53, No. 8, August 1965.
31. Special issue, *Proc. IEEE*, Vol. 77, No. 5, May 1989.
32. Special issue, *IEEE Trans. Antennas Propag.*, Vol. 37, No. 5, May 1989.
33. W. R. Stone (Ed.), *Radar Cross Sections of Complex Objects*, IEEE Press, Piscataway, NJ, 1989.
34. A. F. Stevenson, Relations between the transmitting and receiving properties of antennas, *Q. Appl. Math.*, pp. 369–384, January 1948.
35. R. F. Harrington, Theory of loaded scatterers, *Proc. IEE* (British), Vol. 111, pp. 617–623, April 1964.
36. R. E. Collin, The receiving antenna, in *Antenna Theory, Part I*, R. E. Collin and F. J. Zucker (Eds.), McGraw-Hill, New York, 1969.
37. R. B. Green, The Effect of Antenna Installations on the Echo Area of an Object, Report No. 1109-3, ElectroScience Laboratory, Ohio State University, Columbus, OH, September 1961.
38. R. B. Green, Scattering from conjugate-matched antennas, *IEEE Trans. Antennas Propagat.*, Vol. 14, No. 1, pp. 17–21, January 1996.

39. R. J. Garbacz, the Determination of Antenna Parameters by Scattering Cross-Section Measurements, III. Antenna Scattering Cross Section, Report No. 1223-10, Antenna Laboratory, Ohio State University, Columbus, November 1962.

40. R. C. Hansen, Relationships between antennas as scatterers and as radiators, *Proc. IEEE*, Vol. 77, no. 5, pp. 659–662, May 1989.

41. S. H. Dike and D. D. King, Absorption gain and backscattering cross section of the cylindrical antenna, *Proc. IRE*, 40, 1952.

42. J. Sevick, Experimental and Theoretical Results on the Backscattering Cross Section of Coupled Antennas, Tech. Report No. 150, Cruft Laboratory, Harvard University, May 1952.

43. D. L. Moffatt, Determination of Antenna Scattering Properties From Model Measurements, Report No. 1223-12, Antenna Laboratory, Ohio State University, January 1964.

44. J. T. Aberle, Analysis of Probe-Fed Circular Microstrip Antennas, Ph.D. Dissertation, University of Massachusetts, Amherst, 1989.

45. J. T. Aberle, D. M. Pozar, and C. R. Birtcher, Evalution of input impedance and radar cross section of probe-fed microstrip patch elements using an accurate feed model, *IEEE Trans. Antennas Propagat*., 39, no. 12, pp. 1691–1696, December 1991.

PART II

ANTENNA ELEMENTS

CHAPTER 2

Wire Elements: Dipoles, Monopoles, and Loops

CYNTHIA M. FURSE, OM P. GANDHI and GIANLUCA LAZZI

2.1 INTRODUCTION

Dipole, monopole, and loop antennas and their associated arrays are the most common antennas used for communication systems, broadcasting, and measurement of electric and magnetic fields. This chapter describes the basic nature of these antennas and some of their applications. Variations such as biconical and bow-tie antennas, slot dipoles, folded dipoles, sleeve dipoles, and shunt-fed dipoles are also described.

2.2 DIPOLE AND MONOPOLE ANTENNAS

The simplest type of wire antennas are the dipole and monopole antennas. A dipole antenna is most commonly a linear metallic wire or rod with a feed point at the center as shown in Figure 2.1a. Most often, a dipole antenna has two symmetrical radiating arms. A monopole antenna is a single radiating arm as shown in Figure 2.1b; however, monopoles are most often used above a full or partial ground plane as shown in Figure 2.1c. The reflections from the ground plane produce a "virtual monopole" below the ground, as shown by the dotted line, so a monopole antenna above a perfect ground plane can be evaluated in much the same way as a dipole antenna. Thus the basics of dipole and monopole antennas are discussed side by side in this chapter.

Because of the symmetry of dipoles and monopoles relative to the x-y plane containing the feed point, the resultant radiation is independent of ϕ (rotationally symmetric about the z-axis) and is shown in Figure 2.2. The shape of the lobe(s) depends on the length of the antenna, as discussed in the next section. In general, however, dipole and monopole antennas are used for applications where the radiation is desired in the x-y plane of the antenna.

2.2.1 Infinitesimal Dipole (Hertzian Dipole)

An infinitesimal dipole (length $L<<\lambda$) is a linear element that is assumed to be short enough that the current (I) is constant along its length. This is also called a Hertzian

Modern Antenna Handbook. Edited by Constantine A. Balanis

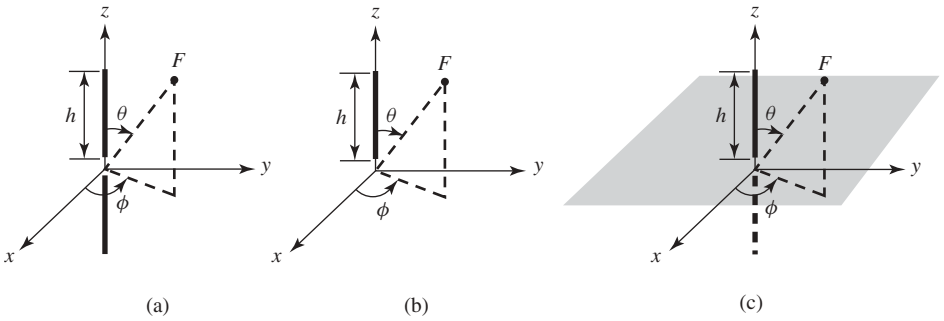


Figure 2.1 (a) Dipole antenna, (b) monopole antenna, and (c) monopole above a ground plane showing the virtual monopole below.

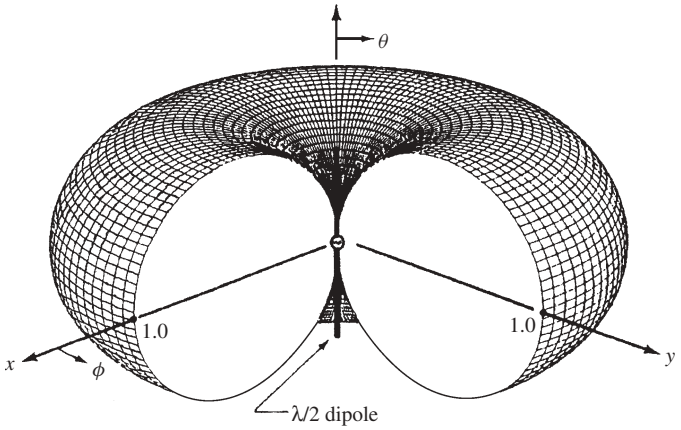


Figure 2.2 Radiation pattern for an infinitesimal (or Hertzian) dipole. (From Ref. 1.)

dipole. The electric- and magnetic-field components of the Hertzian dipole are [2]

$$\mathbf{H} = \frac{1}{4\pi} IL \sin\theta \left(\frac{jk_0}{r} + \frac{1}{r^2} \right) e^{-jk_0 r} \hat{\boldsymbol{\phi}} \tag{2.1}$$

$$\mathbf{E} = \frac{j\eta_0 IL}{2\pi k_0} \cos\theta \left(\frac{jk_0}{r^2} + \frac{1}{r^3} \right) e^{-jk_0 r} \hat{\mathbf{r}} - \frac{j\eta_0 IL}{4\pi k_0} \sin\theta \left(-\frac{k_0^2}{r} + \frac{jk_0}{r^2} + \frac{1}{r^3} \right) e^{-jk_0 r} \hat{\boldsymbol{\theta}} \tag{2.2}$$

where $\eta_0 = (\mu_0/\varepsilon_0)^{1/2}$ is the intrinsic impedance ($= 377\ \Omega$) for free space and $k_0 = \omega (\mu_0 \varepsilon_0)^{1/2}$ is the propagation constant ($= \omega/c$ where c is the velocity of light). The fields are seen to decay rapidly ($1/r^3$ variation, where r is the distance from the antenna) very near the antenna, and less rapidly ($1/r$ variation) farther away. The fields with terms $1/r^2$ and $1/r^3$ (the *induction terms*) provide energy that is stored near the antenna. This can be

used for applications where heat is desired near the antenna such as for cardiac ablation or hyperthermia, applications that are discussed in Section 2.2.7.5. The fields with $1/r$ variation (the *radiation terms*) provide energy propagation away from the antenna. These are the fields used for communication applications. The distance away from the antenna where the induction and radiation terms are equal is $r = \lambda/2\pi$. When $r < \lambda/2\pi$ this is the *near field* of the antenna, and the induction terms dominate. When $r > \lambda/2\pi$ this is the *far field*, and the radiation terms dominate. In the far field, the wave propagation is in the transverse electromagnetic (TEM) mode, which is characteristic of far-field radiation from finite structures.

The far-zone radiated fields of the Hertzian dipole follow from Eqs. (2.1) and (2.2) by retaining the $1/r$ varying terms:

$$\mathbf{H} = \frac{j}{4\pi r} I L \sin\theta e^{-jk_0 r} \hat{\boldsymbol{\phi}} \tag{2.3}$$

$$\mathbf{E} = \frac{j\eta_0}{4\pi r} I L \sin\theta e^{-jk_0 r} \hat{\boldsymbol{\theta}} \tag{2.4}$$

As expected for TEM wave propagation, the **E** and **H** fields are perpendicular to each other and to the outward propagation in the $\hat{\mathbf{r}}$ direction. Also the ratio of $E/H = \eta_0 = (\mu_0/\varepsilon_0)^{1/2}$, which is the intrinsic impedance of free space.

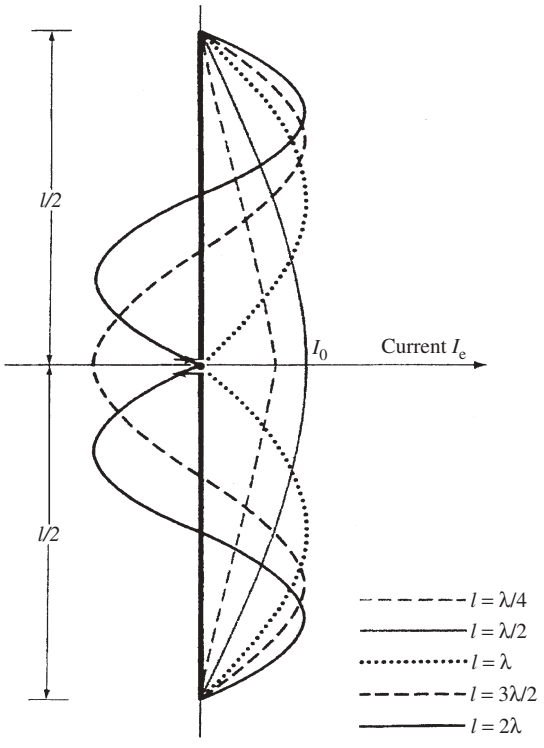
The radiation pattern of the Hertzian dipole is shown in Figure 2.2 and exhibits the classical symmetry expected of dipole antennas, being both independent of ϕ and symmetric about the x-y plane through the center (feed point) of the dipole. The magnitude of the total radiated power is $P_{\text{rad}} = 40\pi^2 I_0^2 (L/\lambda)^2$, where I_0 is the current on the dipole. From Eqs. (2.3) and (2.4) it is interesting to note that even for this constant-current infinitesimal dipole, the radiated power density is proportional to $\sin^2\theta$. Hence it is maximum for $\theta = 90^\circ$ (i.e., in the x-y plane normal to the orientation of the dipole) and zero for the directions along the length of the dipole ($\theta = 0^\circ$ and 180°). The latter property for zero radiation along the length of the dipole is seen for all linear antennas regardless of length. This follows from the fact that a linear antenna of finite length may be considered to be composed of a set of infinitesimal dipoles. Since the infinitesimal dipoles do not create **E** and **H** fields or radiated power density for the $\theta = 0^\circ$ and 180° directions, the sum of many of them also does not have power in these directions.

2.2.2 Linear Dipole Antennas and Monopoles Above a Perfect Ground Plane

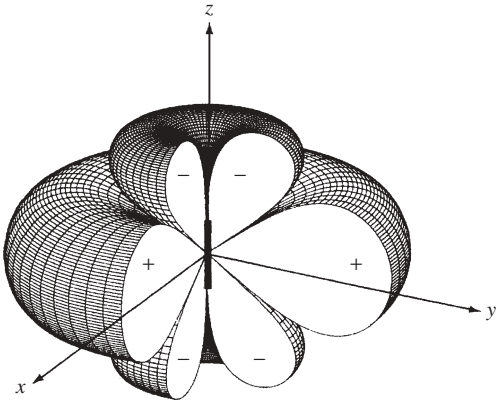
The geometry of a linear dipole antenna of length $2h$ (or a monopole antenna of length h sitting on a perfect ground plane) is shown in Figure 2.1. The current distribution is sinusoidal and is given by

$$I(z) = \frac{I(0)}{\sin kh} \sin k(h - |z|) \quad \text{for} - h < z < h \tag{2.5}$$

where $I(0)$ is the current at the feed point of the antenna, h is the half-length of the antenna, and $k = \omega(\mu\varepsilon)^{1/2}$ is the propagation constant in the material surrounding the dipole. The current distributions for several lengths of dipole antennas are shown in Figure 2.3. These current distributions are the same for the monopole and its virtual reflection below the ground plane.



(a)



Three dimensional

(b)

Figure 2.3 (a) Current distributions and (b) radiation patterns for a 1.25λ dipole antenna. Monopole antennas above a ground have currents and radiation patterns of the same form as dipole antennas, but the currents and fields below the ground are "virtual" and therefore do not actually exist. (From Ref. 3.)

The electric and magnetic fields around the dipole are calculated by modeling the antenna as a series of Hertzian dipoles having different current strengths shown by the patterns in Figure 2.1 and integrating the fields from each of these elements. The resultant fields far from the antenna at a distance r_0 are

$$\mathbf{E} = \frac{j\eta I(0)}{2\pi r_0 \sin(kh)} F(\theta) e^{j(\omega t - kr_0)} \hat{\boldsymbol{\theta}} \tag{2.6}$$

and

$$\mathbf{H} = \frac{jI(0)}{2\pi r_0 \sin(kh)} F(\theta) e^{j(\omega t - kr_0)} \hat{\boldsymbol{\phi}} \tag{2.7}$$

where $\eta = (\mu/\varepsilon)^{1/2}$ and where the θ dependence of the radiated fields $F(\theta)$ is called the pattern factor and is given by the following:

$$F(\theta) = \frac{\cos(kh\cos(\theta)) - \cos kh}{\sin\theta} \tag{2.8}$$

The radiated power density (*radiation pattern*) is given by

$$P(\theta) = \frac{\mathbf{E}\cdot\mathbf{E}^*}{2\eta} = \frac{\eta I^2(0)}{8\pi^2 r^2 \sin^2(kh)} F^2(\theta) \tag{2.9}$$

Using $\eta = \eta_0 = 120\pi$, this can also be expressed in terms of the total radiated power $W(= I^2(0)R_a/2)$ and the feed-point resistance R_a as follows:

$$P(\theta) = \frac{30}{\pi r_0^2} \frac{W}{R_a} \frac{F^2(\theta)}{\sin^2 kh} \tag{2.10}$$

The normalized radiation patterns are shown in Figure 2.3b for several different lengths of dipoles and monopoles above ground. The radiation patterns for the dipole and monopole are the same, except that the monopole does not radiate in the lower half-plane and therefore radiates twice the power in the upper half-plane. Mathematically this is accounted for, because the feed point resistance R_a of the monopole is half that of the dipole, as will be seen later in this section.

The directivity of a dipole antenna is the power density in the direction of maximum radiation (at a specified distance r_0) normalized by the power density of an isotropic radiator. Directivity is given by

$$D = \frac{P_{\max}}{P_0} = \frac{F^2(\theta)_{\max}}{\frac{1}{2}\int_0^\pi F^2(\theta)\sin\theta\, d\theta} = \frac{120}{R_a} \frac{F^2(\theta)|_{\max}}{\sin^2 kh} \tag{2.11}$$

where $P_0 = W/4\pi r^2$ is the power density of an isotropic radiator.

Because monopole antennas help reduce the length of the necessary dipole by a factor of 2 and result in directivities that are twice as large, vertical monopole antennas above ground are extensively used for amplitude modulated (AM) broadcasting in the frequency range 535–1605 kHz. For this application, the wavelengths are long, on the

order of 200–600 m, and the monopole antennas are immensely helpful in reducing the required height. It is necessary, of course, to create a good conductivity ground, which for dry or rocky soil conditions is often obtained by burying a conducting screen made of radially spread metal wires with angular separations of 2–3° that extend to a radius at least equal to the height of the antenna, but preferably 20–50% larger than this minimal requirement. This screen, called a counterpoise, is often buried a few inches below the surface of the natural ground but may also be left slightly above ground for rocky or otherwise difficult terrain.

A graph of the variation of the feed-point resistance and reactance of a monopole antenna above ground is given in Figure 2.4 as a function of length h/λ, where λ is the

TABLE 2.1 Relationships Between Monopole and Dipole Antennas

	Monopole Above Ground Length $= h$	Corresponding Dipole of Twice Length $L = 2h$
Radiation pattern	Same as that for the dipole but only for angle $0 \le \theta \le 90^\circ$	
Feed-point resistance R_a	R_a—$_{\text{monopole}} = \frac{1}{2}R_{a,d}(2h)$	$R_{a,d}$: function of length
	X_a—$_{\text{monopole}} = \frac{1}{2}X_{a,d}(2h)$	$L = 2h$ (see Figure 2.2)
Feed-point reactance X_a	D_a—$_{\text{monopole}} = \frac{1}{2}[2D_{a,d}(2h)]$	$X_{a,d}$: function of length $L = 2h$ (see Figure 2.2)
Directivity D_a		$D_{a,d}$: function of length $L = 2h$

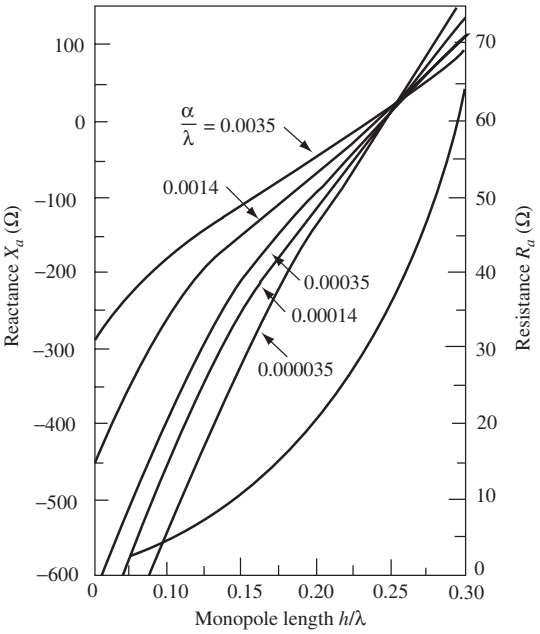


Figure 2.4 Variation of feed-point R_a and reactance X_a for an end-fed monopole above ground as a function of height h/λ (From Ref. 4, with permission of Pearson Education, Inc., Upper Saddle River, NJ.)

free-space wavelength at the radiation frequency [4]. Note that the reactance X_a depends on the conductor radius a, whereas the feed-point resistance R_a is relatively independent of conductor radius a for thin antennas ($a/\lambda \leq 1$). The input resistance R_a of a center-fed dipole antenna of length $2h$ is twice that of an end-fed monopole of length h.

The ohmic losses of a dipole antenna (given by $I^2(0)R_{\text{ohmic}}/2$) are quite small, particularly for $h/\lambda > 0.1$. Ohmic losses for a monopole above ground are half those of a dipole antenna. The resultant antenna radiation efficiencies (given by $R_a/(R_a + R_{\text{ohmic}})$) are on the order of 90–99%.

Monopole antennas are also the antennas of choice for very low frequency (VLF) (3–30 kHz) and low frequency (LF) (30–300 kHz) communication systems. For these applications, the height h of the antenna is generally a small fraction of the wavelength. From Figure 2.4, for small values of h/λ, the feed-point resistance R_a is very small, on the order of a few ohms, and the ohmic losses resulting from the surface resistance of the antenna can be significant by comparison. This results in reduced antenna radiation efficiency.

Top loading of the monopole antenna illustrated in Figure 2.5 is often used to improve the feed-point or antenna-equivalent resistance R_a. Because the top set of wires acts as a capacitance to ground rather than the open end of the antenna, the current at the top of the antenna ($z = h$) no longer needs to go to zero as it does for an open-ended monopole (from Eq. (2.5), $I = 0$ at $z = h$). The upper region of the antenna can therefore support a substantial radio frequency (RF) current, resulting in improved radiation efficiency. The antenna-equivalent resistance $R_a = 395\ h^2/\lambda^2$ for open-ended short monopoles may be improved by a factor of 2–3 by use of top loading, resulting in higher radiation efficiency.

It should be mentioned that the various concepts discussed here are also usable at higher frequencies. Open-ended monopole antennas are, in fact, used up to ultrahigh frequency (UHF) and microwave frequencies, where metallic conducting sheets, either solid or in the form of a screen, may be used in lieu of the ground to create the image antenna. A variety of top-loading "hats" illustrated in Figure 2.6 may also be used at these higher frequencies.

Physical dipoles and monopoles act slightly different than predicted, because they have some finite thickness, and the ends of the wire capacitively couple to air. This effectively makes the antenna electrically longer than its physical length by 2%–9%.

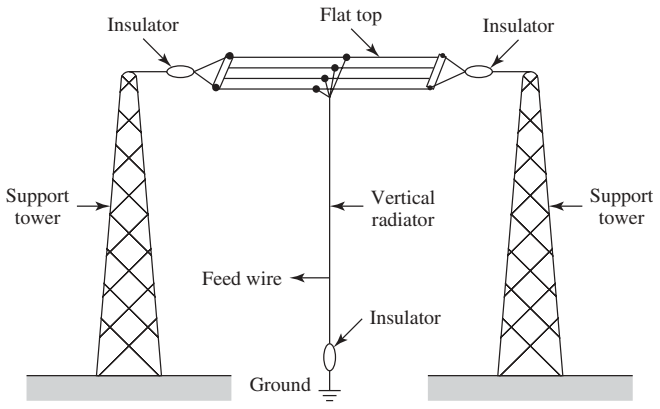


Figure 2.5 A top-loaded short-monopole antenna typical of VLF/LF communication systems.

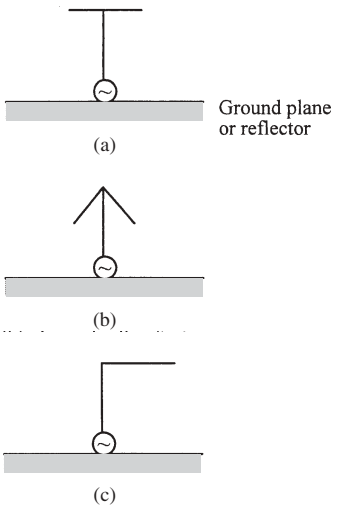


Figure 2.6 Some typical geometries used for top-loaded monopole antennas: (a) plate, (b) umbrella of wires, and (c) inverted L monopole.

For a half-wave dipole (length $= 2\ h = \lambda/2$), for instance, the physical length must be slightly shortened in order to create a resonant length antenna ($X_a = 0$). Table 2.2 shows the wire lengths required to produce a resonant half-wave dipole. This shortening factor varies from 2% to 9%, depending on the thickness of the dipole.

Since a dipole or monopole antenna is a physically resonant structure, its feed-point impedance (particularly the reactance X_a) varies greatly with frequency. The input impedance of a dipole antenna is shown in Figure 2.7. Thus these antennas have a fairly narrow bandwidth. The VSWR of a dipole antenna as a function of frequency and wire thickness is shown in Figure 2.8 for an antenna that would be half-wave resonant at 300 MHz. Using a measure of "usable bandwidth" that the measured VSWR should be less than 2:1, this antenna has bandwidths of $310 - 262 = 48$ MHz for the thicker wire and $304 - 280 = 24$ MHz for the thinner wire. As fractions of the design frequency (300 MHz), the bandwidths are 16% and 8%, respectively.

2.2.3 Effect of Imperfect Ground Plane on Monopole Antennas

Section 2.2.2. describes the performance of a monopole antenna above a perfect ground; however, its performance is degraded if the ground plane is resistive, curved, or finite

TABLE 2.2 Wire Lengths Required to Produce a Resonant Half-Wave Dipole for a Wire Diameter of 2*a* and a Length *l*

Length to Diameter Ratio, $L/2a$	Percent Shortening Required	Resonant Length	Dipole Thickness Class
5000	2	0.49λ	Very thin
50	5	0.475λ	Thin
10	9	0.455λ	Thick

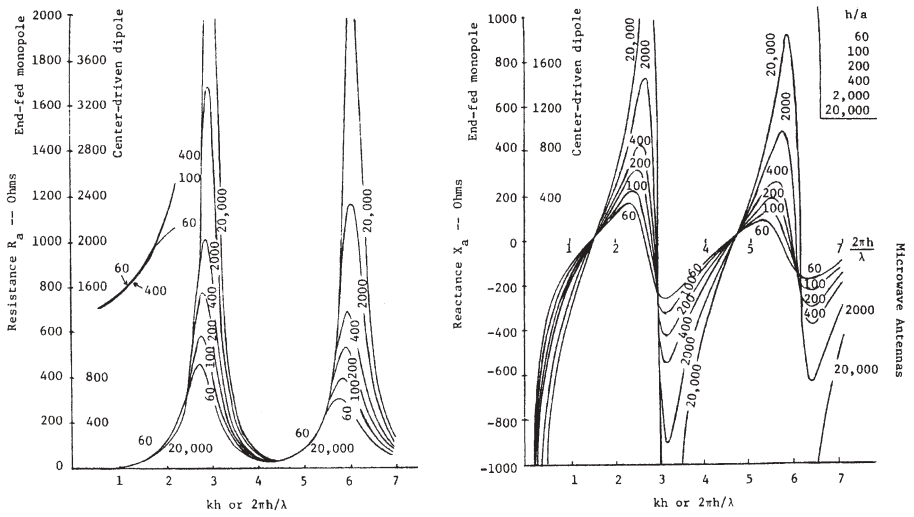


Figure 2.7 Input impedance of dipole. (From Ref. 4, with permission of Pearson Education, Inc., Upper Saddle River, NJ.)

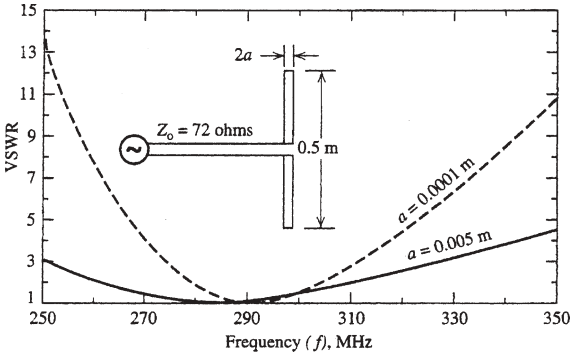


Figure 2.8 VSWR of a dipole antenna as a function of frequency and wire thickness. (From Ref. 5.)

in size. The effect of finite conductivity of the ground plane is important in large scale applications such as AM broadcasting, where the ground plane is created by burying individual wires in the ground in a radial (wheel spokes) pattern around the antenna. In this case the power density at the surface of the earth diminishes more rapidly than the square of the distance r from the antenna, effectively partially absorbing rather than ideally reflecting the radiated power [6]. Higher frequencies and regions with lower conductivity soil such as sandy or rocky lands and urban areas experience greater power reduction. When the antennas are used for transmission in excess of about 1000 miles, the curvature of the surface of the earth also reduces the radiated power [6].

A ground plane of diameter 10 wavelengths or larger has a fairly small effect on the feed-point impedance of a monopole antenna as compared to the values given in Figure 2.4, but a finite-size ground plane has a fairly significant effect on both the

feed-point impedance and radiation pattern [7]. An example of the effect on the radiation pattern is illustrated in Figure 2.24a for the radiation pattern of a monopole antenna of a cellular telephone in free space. Since this monopole antenna is mounted on a plastic-covered metal box of finite dimensions, only a finite-size ground plane of box shape is used for this radiator. One effect is to alter the radiation pattern (see Figure 2.24a) from a perfect circle in the azimuthal plane to one that is crooked with a lower gain in the sector where the extent of the ground plane is minimal, that is, on the side where the antenna shown by a circle in the insert of Figure 2.24a is mounted. Also the gain for this quarter-wave monopole antenna of about 1 dBi is considerably lower than 5.16 dBi (a factor of 3.28) for the same antenna mounted on a flat, infinite ground plane. The actual reason for this considerably reduced gain is that part of the energy is also radiated in the lower half-space for this vertical monopole antenna above the box while none would be radiated for this antenna mounted above an infinite ground plane.

2.2.4 Specialized Dipole Antennas

2.2.4.1 Slot Dipole A slot dipole antenna is a dual to the linear dipole antenna. A slot antenna is produced by creating a thin slot of length h in a conducting metal sheet and feeding it in the center of the slot. The radiation pattern of a slot antenna is identical to that of the linear dipole of the same length (see Figure 2.3) except that orientations of the **E** and **H** fields are interchanged. This means that the power pattern is the same, even though the orientation of the fields is different. Also the feed-point impedance Z_s of a slot antenna is related to that of the dual linear antenna by the following equation:

$$Z_s = \frac{\eta^2}{4Z_a} \tag{2.12}$$

where Z_s is the impedance of the slot and Z_a is the impedance of the dual linear antenna. Cavity-backed slot antennas with integrated matching networks are among the smallest (physical size/electrical size) antennas available [8].

2.2.4.2 Biconical Dipoles A biconical dipole, such as shown in Figure 2.9a, is commonly used for broadband applications. Typical flare angles θ are between 30° and 60°. The exact flare angle is not critical, so it is generally chosen so that the impedance of the dipole nearly matches the impedance of the feed line to which it is connected. The impedance of the biconical dipole varies as a function of wavelength and flare angle, with a relatively flat impedance response for wide flare angles. Hence this antenna is broader band than a simple linear dipole. Some variations of this method of using flaring to increase bandwidth are the flat bow-tie antenna (Figure 2.9b) (which may be built on a printed circuit board) and the wire version of the biconical antenna (Figure 2.9c).

2.2.4.3 Folded Dipole Antennas A folded dipole antenna is shown in Figure 2.10. The dipole is created by joining two cylindrical dipoles at the ends and driving the entire structure by a transmission line (often a two-wire transmission line) at the center of one arm as shown. The feed-point impedance of a folded dipole of two identical-diameter arms is four times as large as for an unfolded dipole of the same length. This can actually be advantageous, since the feed-point resistance may now be comparable to the characteristic impedance, Z_0, of the transmission or feed line. The reactance of the antenna may easily be compensated by using a lumped element with a reactance that is

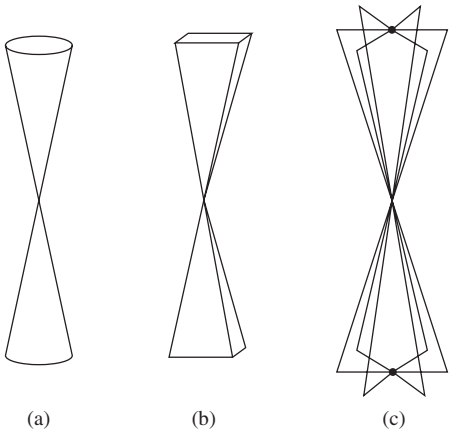


Figure 2.9 Biconical dipole antenna and variations: (a) biconical dipole antenna, (b) flat bow-tie antenna, and (c) wire version of biconical dipole antenna.

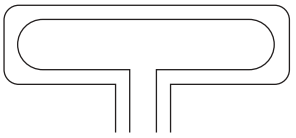


Figure 2.10 Folded dipole antenna.

the negative of the reactance at the terminals of the folded dipole antenna or else by using a foreshortened antenna length to resonant length arms so that $X_a = 0$ (see Table 2.1).

2.2.4.4 Shunt-fed Dipoles Matching networks of reactive elements are generally required to match the feed-point impedance ($R_a + jX_a$) of center-fed dipoles to transmission lines. Typically these lines have characteristic impedance on the order of 300–600 Ω, and a thin half-wave dipole has impedance $Z = 73 + j42\ \Omega$. To alleviate the need for matching networks, the dipoles are at times shunt-fed at symmetric locations off the center point as shown in Figure 2.11. This procedure using either the delta match (Figure 2.11a) or the T-match (Figure 2.11b) is often used for half-wave dipoles ($2h = \lambda/2$) with A and B dimensions that are typically on the order of 0.10λ to 0.15λ.

2.2.4.5 Sleeve Dipole The sleeve dipole antenna and its equivalent electrical model are shown in Figure 2.12. In practice, this antenna is built from a coaxial line with the outside conductor and insulation stripped away from the center conductor, which is left protruding. The outer conductor is connected to the ground plane, and the image produced by the ground plane creates an equivalent sleeve dipole antenna. These dipoles are useful, because they have a broadband VSWR over nearly an octave of bandwidth.

2.2.5 Dipole Antenna Arrays

Dipole antennas and arrays of dipole antennas are used for short-wave (3–30 MHz) and for VHF and UHF (30–900 MHz) radio and TV broadcasting. If directional communication

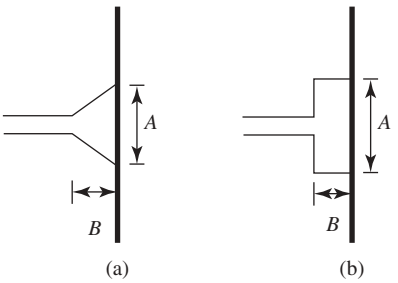


Figure 2.11 Shunt-fed dipoles: (a) delta match and (b) T-match.

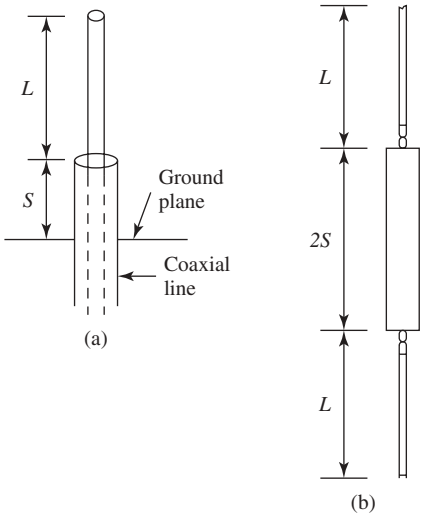


Figure 2.12 Sleeve dipole antenna: (a) physical model and (b) equivalent electrical model.

is desired such as for short-wave radio transmission via the ionosphere, a phased array of horizontal dipoles may be used mounted above a ground plane. The spacing is chosen to send the major lobe of radiation toward the sky at a suitable angle to reflect off the ionosphere and provide broadcast coverage over the desired service area.

For VHF and UHF radio and TV broadcasting over a 360° azimuthal angle, colinearly mounted vertical dipoles that are excited in-phase with each other are often used. Two examples of this are shown in Figure 2.13. An example variation of this is a three- to eight-bay turnstile antenna used for TV broadcasting, shown in Figure 2.14a. Each of the turnstiles is made of two perpendicular slot antennas as shown in Figure 2.14b.

2.2.5.1 Log-Periodic Antennas For broadband applications log-periodic antennas are commonly used as both transmitting and receiving antennas. The bandwidth is easily controlled by adjusting the relative lengths of the longest and shortest elements in the array. The geometry of a log-periodic array is shown in Figure 2.15a, which shows how the "phase-reversal" feed system for this antenna is constructed. The equivalent antenna model of this array is shown in Figure 2.15b. The elements of the array are dipole

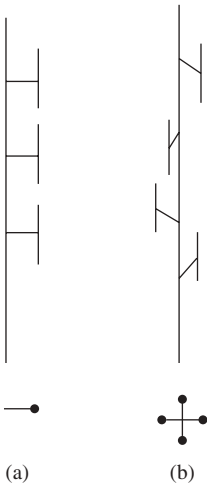


Figure 2.13 Colinearly mounted vertical dipoles for VHF and UHF radio and TV broadcasting: (a) pole-mounted array of colinear dipoles and (b) vertical dipoles spaced around a pole.

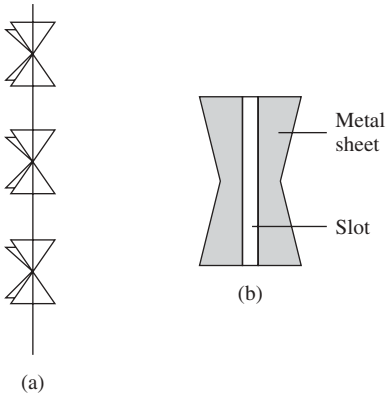


Figure 2.14 Variation on colinearly mounted vertical dipoles: (a) turnstile antenna used for TV broadcasting and (b) one of two perpendicular slot antennas that comprise each turnstile.

antennas that increase in both length and spacing according to the formula

$$\tau = \frac{R_{n+1}}{R_n} = \frac{d_{n+1}}{d_n} \tag{2.13}$$

where $\tau = \mathrm{f}_n / f_{n+1}$ is the ratio of the resonant frequencies f_n and f_{n+1} of the adjacent dipole elements. Since lengths and spacings are interrelated, the choice of one initial value controls the design of the remaining elements. The spacing between one half-wave dipole and its adjacent shorter neighbor is given by

$$\sigma = \frac{d_n}{2L_n} = \frac{(1-\tau)}{4} \cot \alpha \tag{2.14}$$

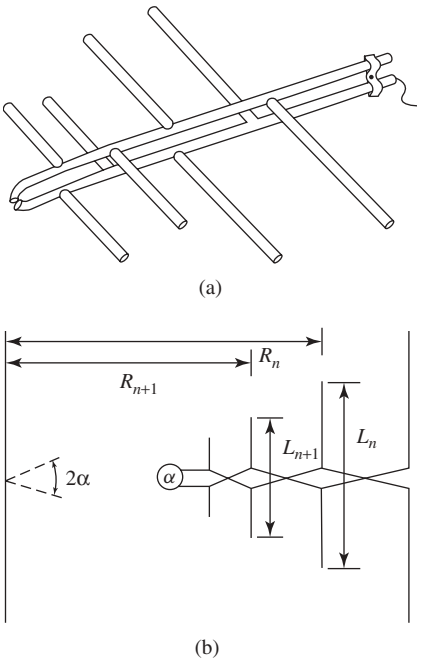


Figure 2.15 Log-periodic dipole array. (a) Geometry of a log-periodic array showing how the "phase-reversal" feed system for this antenna is constructed. (From Ref. 2.) (b) Equivalent antenna model of the log-periodic array.

Log-periodic arrays are generally constructed with small values of α ($10^\circ \leq \alpha \leq 45^\circ$ [3]) and large values of τ ($0.95 \leq \tau \leq 0.7$ [3]) that essentially gives a traveling wave propagating to the left in the backfire direction, away from the antenna array. The nature of this array is that only the elements that are approximately half-wavelength long radiate, and since they are radiating to the left, the smaller elements do not interfere with them. This is accomplished by the phase reversal of the feeds. An array that is built without the phase reversal radiates in the endfire direction. The interference of the longer elements to the right of radiating elements results in spurious reflections and erratic impedance behavior, known as "end effect."

An effective way to further increase the bandwidth of a log-periodic antenna is to change from dipole elements to elements with individual broader bandwidths, similar to changing from a dipole antenna to a biconical antenna. This is accomplished for log-periodic arrays (LPAs) by using a configuration of wires such as shown in Figure 2.16, where each element is a sawtooth element and therefore has broader bandwidth than the individual dipole elements.

2.2.5.2 Broadband Dipole Curtain Arrays A broadband dipole curtain such as shown in Figure 2.17 is commonly used for high power (100–500-kW) HF ionospheric broadcasting and short-wave broadcasting stations. The curtain is composed of several dipoles, usually half-wavelength long, mounted horizontally or vertically in a rectangular or square array, often backed by a reflecting plane or wire mesh. This array has several desirable features including high gain, broad bandwidth, independent control of horizontal

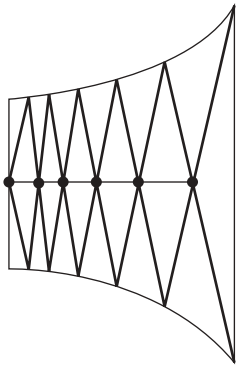


Figure 2.16 LPA with sawtooth wire elements for increased bandwidth.

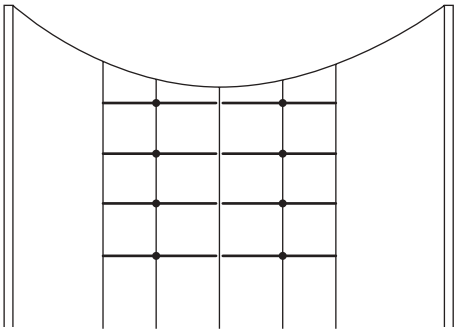


Figure 2.17 A broadband dipole curtain.

and vertical radiation patterns, ease of matching (low VSWR), and the ability to broadcast efficiently. Using a phased-feeds system, this array allows beam steering of the radiation pattern in both the azimuthal and elevation planes, providing a very high degree of flexibility.

2.2.5.3 Yagi–Uda Dipole Array Yagi–Uda arrays are commonly used as general purpose antennas from 3–3000 MHz, in particular, as home TV antennas. They are inexpensive, have reasonable bandwidth, and have gains up to 17 dBi or more if multiple arrays are used. The have unidirectional beams with moderate side lobes [2].

A typical Yagi–Uda array is shown in Figure 2.18. This array is a simple endfire array of dipole antennas, where only one of the elements is driven and the rest are parasitic. The parasitic elements operate as either reflectors or directors. In general [2], the longest antenna, which is about $\lambda/2$ in length, is the main reflector and is generally spaced $\lambda/4$ in back of the driven dipole. The feed element is commonly a folded dipole antenna 0.45λ to 0.49λ long. Adding directors, which are generally 0.4λ to 0.45λ long, to the front of the driven element increases the gain of the array. The directors are not always of the same length, diameter, or spacing. Common arrays have 6–12 directors and at most two reflectors. Additional improvements in gain by adding more elements are limited; however, arrays have been designed with 30–40 elements [3]. A gain (relative

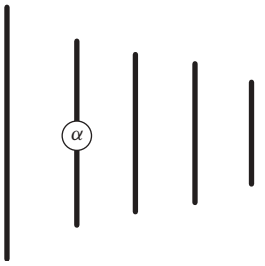
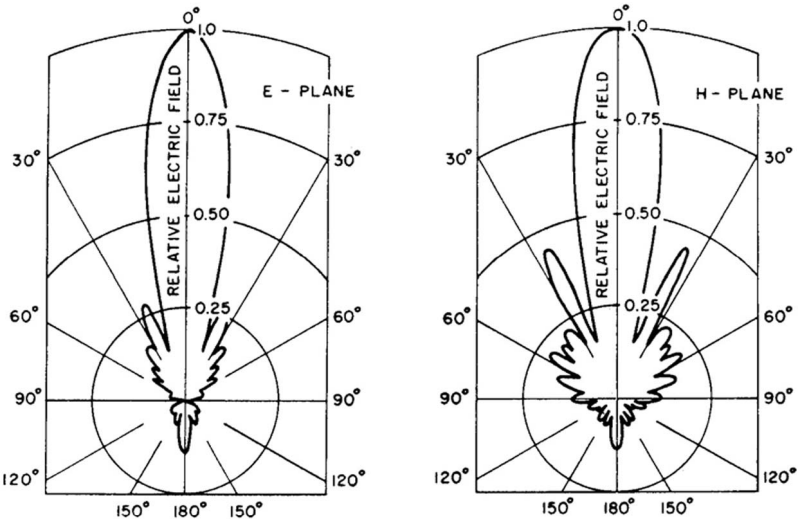


Figure 2.18 Yagi–Uda array.

to isotropic) of 5–9 per wavelength of array length is typical for Yagi–Uda arrays, for an overall gain of 50–54 (14.8–17.3 dB).

The Yagi–Uda array is characterized by a main lobe of radiation in the direction of the director elements and small side lobes. The beamwidth is small, generally 30–60° [3]. Typical E- and H-plane patterns of a Yagi–Uda array are shown in Figure 2.19. Typically, the performance of a Yagi-Uda array is computed using numerical techniques [10]. For the simple case where all of the elements are approximately the same size, the electric-field pattern can be computed from the array factors of the various elements.

The input impedance of a Yagi–Uda array is often small. For example, for a 15-element array with reflector length = 0.5λ, director spacing = 0.34λ, and director length = 0.406λ, the input impedance is 12, 22, 32, 50, or 62 Ω for reflector spacings of 0.10λ, 0.13λ, 0.15λ, 0.18λ, and 0.25λ, respectively. This can make matching to



E-plane *H*-plane

Figure 2.19 Typical E- and H-plane patterns of a Yagi–Uda array. Total number of elements = 27, number of directors = 25, number of reflectors = 1, number of driven elements = 1, total length of reflector = 0.5λ, total length of feeder = 0.47λ, total length of each director = 0.406λ, spacing between reflector and feeder = 0.125λ, spacing between adjacent directors = 0.34λ, radius of wires = 0.003λ. (From Ref. 9.)

typical transmission lines (50, 75, or 300 Ω) difficult. Folded dipoles used for the driven element are therefore used to boost the input impedance by a factor of 4 or more.

Extensive studies of the design of Yagi–Uda arrays have been made [11, 12], and tables are provided to optimize the Yagi–Uda array for a desired gain.

2.2.5.4 Crossed Dipoles for Circular Polarization For applications that require a circularly polarized antenna such as TV and FM broadcasts and space communications, at least two dipoles, each of which has a linear polarization, must be combined in an array, often referred to as *crossed dipoles*. In a crossed dipole configuration, dipoles are mounted perpendicular to each other for circular polarization or at other angles for elliptical polarization. Currents are fed 90° out of phase between the two dipoles. These can also be used as probes for sensing vector fields to isolate individual components of the electric field. Adaptations of the crossed dipole are shown in Figure 2.20a, b. Dipole arrays such as the Yagi–Uda can also be combined to provide circular polarization, as shown in Figure 2.20.

2.2.6 Specialized Monopole Antennas

Because a monopole antenna above ground or similar reflector acts as a dipole antenna, several of the concepts that are useful for dipole antennas are also used for the monopole antenna. Some of these concepts that are often used in practice are discussed next.

2.2.6.1 Folded Monopoles A folded monopole antenna is illustrated in Figure 2.21a, where the feed point is connected to arm 1, and arm 2 is grounded at the bottom end. For dipoles, this provides a method of increasing the feed-point impedance given in the following relationship [5]:

$$Z_i = (1 + c)^2 (R_a + jX_a) \tag{2.15}$$

where R_a and X_a plotted in Figure 2.4 are the resistance and reactance of the single arm (not folded) monopole antenna, and the factor c is given approximately by

$$c \approx \frac{\ln(d/a_1)}{\ln(d/a_2)} \tag{2.16}$$

In Eq. (2.16), a_1 and a_2 are radii of wires or rods used for arms 1 and 2, respectively, and d is the center-to-center spacing between the two arms of the antenna, generally much smaller than the height of each of the arms. For a folded monopole antenna where equal radii arms are used, the feed-point impedance is therefore four times higher than that for a monopole antenna that is not folded. This can be truly advantageous, because the feed-point resistance may now be comparable to the characteristic impedance Z_0 of the transmission or feeder line, and the reactance of the antenna may easily be compensated by using a lumped element with a reactance that is the negative of the reactance $(1 + c)^2 X_a$ at the terminals of the folded monopole antenna.

2.2.6.2 Shunt-Fed Monopoles A typical arrangement of a shunt-fed grounded monopole is shown in Figure 2.21b. This corresponds to one-half of a delta match that is often used for dipoles. Because the base of the monopole is grounded for this arrangement, the antenna is fed typically 15–20% farther up on the antenna; the exact

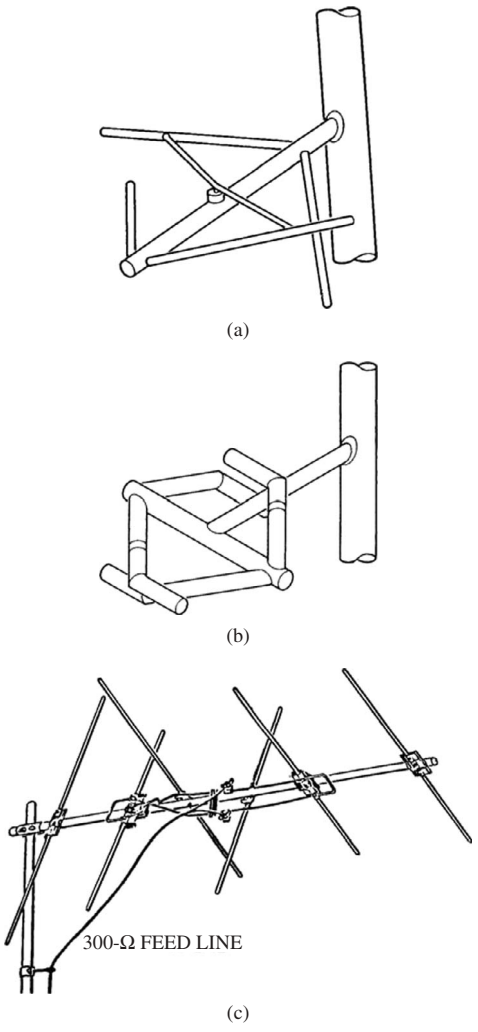


Figure 2.20 Cross-dipole applications for circular or elliptical polarization: (a) two shunt-fed slanted V-dipoles, (b) series-fed slanted dipoles, and (c) circularly polarized Yagi–Uda array. (From Ref. 2.)

location is determined by matching the feeder line of characteristic impedance Z_0, which is typically 200–600 Ω. Shunt-fed monopole antennas were often used in the early days of AM broadcasting when high quality insulators were not generally available and it was convenient to ground the tall steel/aluminum masts being used for monopole antennas. With the advent of high compression strength insulators, the masts are now more typically mounted on these insulators, and the monopole towers are fed at the base of these antennas.

2.2.6.3 Parasitic Monopoles Much like the Yagi–Uda arrays of dipoles [5], parasitic monopoles of grounded elements are used to help direct the beams away from the reflector monopoles in the direction of director monopoles. An illustration of a

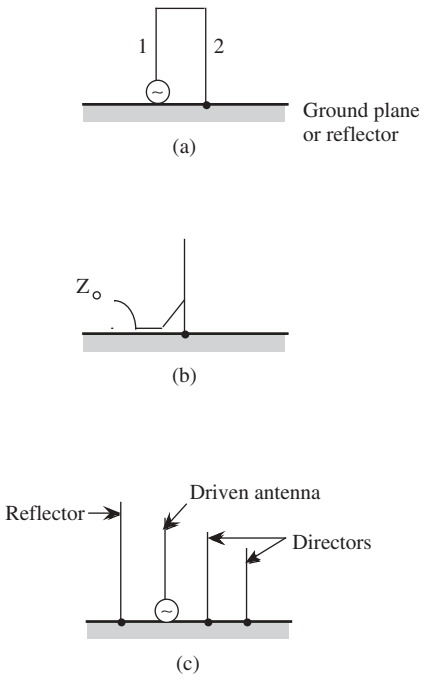


Figure 2.21 Some special adaptations of monopole antennas: (a) folded monopole antenna, (b) shunt-fed grounded monopole, and (c) four-element Yagi–Uda array of monopoles.

four-element Yagi–Uda array of monopoles is shown in Figure 2.21c. Forward gains on the order of 10–12 dB with front-to-back ratios of 5–8 dB may be obtained using such arrangements. One or more of the parasitic monopoles have also been occasionally used for beam shaping by AM broadcast stations. Two-element antenna arrays consisting of a parasitic reflector monopole and a driven monopole antenna have also been proposed for use in hand-held wireless telephones (e.g., see Ref. 13). If they are carefully designed, such antennas may allow electromagnetic fields to be partially directed away from the human head.

2.2.6.4 Arrays of Monopole Antennas Like arrays of dipole elements, monopole arrays of equal or arbitrarily spaced and phased elements may be used for directing principal lobe(s) of radiation in desired directions. Arrays of monopole antennas are used for VLF/LF communications, for AM broadcasting, and, in general, for all frequencies up to microwave frequencies. AM broadcasters for nonclear-channel stations have, for example, used this technique to alter the day and night radiation patterns by varying the phases and the magnitudes of excitations in the various elements.

2.2.6.5 Conical Monopole A conical monopole, realized as a cone above a perfect ground plane, is commonly used because of its broadband properties. The cone may be either solid or built from a series of wires above a ground plane. Similar to conventional monopoles, the ground may also be built from a series of radial wires as shown in Figure 2.22.

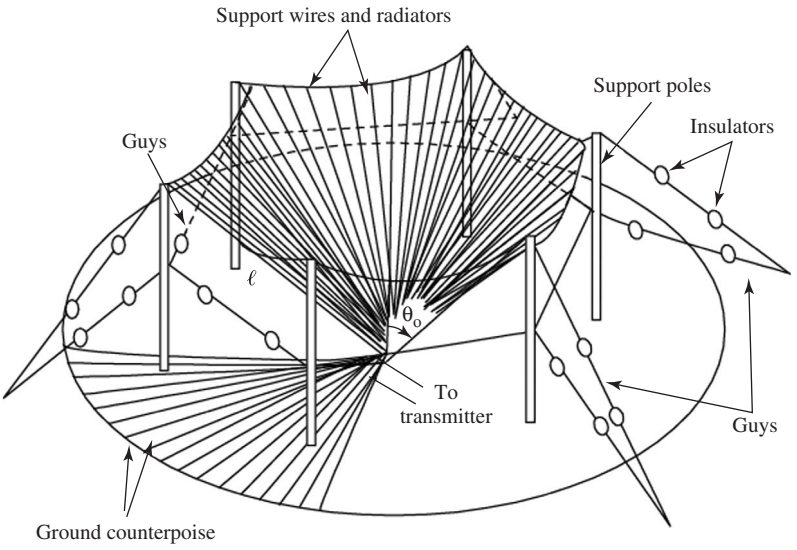


Figure 2.22 A conical monopole.

The analysis of this antenna may be done using either rigorous analytical solutions [14, 15] or an approximate method [15]. The rigorous approach is based on a series expansion of transverse magnetic (TM) modes and is computationally complex. Numerical methods, based on the assumption that the antenna is built from a series of wires, have commonly been used for specific monopole arrangements.

An example of a conical monopole constructed of 60 evenly spaced wires above a ground plane is an antenna that has been in use by the U.S. Navy [16]. For use in the HF band of 2–14 MHz, the dimensions of the monopole are $\ell = 30.6$ m, $a = 0.001$ m, $\theta_0 = 45^\circ$ (see Figure 2.22). The radius of the ground is approximately 37.5 m, constructed of 120 radial wires over average soil ($\varepsilon_r = 15$ and $\sigma = 0.012$ S/m), and the radius of the screen wire is 0.002 m. This antenna performs over a 7:1 frequency band (2–14 MHz) with an input resistance varying from 44 to 64 Ω [15]. Because of the fairly narrow range of variation of input resistance, matching of this antenna to a 50-Ω line is possible over at least a 7:1 frequency range.

2.2.7 Modern Applications of Monopole and Dipole Antennas

2.2.7.1 Monopole Antennas for Car Radio and Mobile Communications

Monopole antennas used for car radio and mobile communications commonly rely on the body of the vehicle to provide the ground plane. Variations in the shape of the vehicle and placement of the antenna affect the radiation pattern and performance of these antennas [17]. For an antenna mounted on the automobile, it would no longer radiate isotropically in the horizontal plane but may radiate with a slightly higher gain in directions where the automobile body extends farther, thus simulating a larger width ground plane in such directions.

For communication antennas in general, it is often desirable to have the shortest possible antenna. Although a quarter-wave monopole is easier to match, a shorter antenna

has an associated capacitive reactance. This can, however, be canceled out by adding a loading coil (inductance) to the base of the antenna.

Another arrangement often used is to place an inductive coil at the center of the antenna [18]. Because of the altered current distribution along the length of the antenna (similar to a top-loaded monopole), this increases the feed-point impedance of the end-fed antenna to roughly twice the value, making it easier to match to such an antenna. The value of the inductance needed at the center is approximately double that which would be needed at the ends. This consequently increases the coil resistance. In order to reduce the inductance of the loading coil required to match the antenna at either the center or top, a capacitive load (a small metal ball) may be added to the tip of the antenna.

2.2.7.2 Monopole Antennas for Personal Wireless Devices A monopole antenna with or without a helix has commonly been used for personal wireless devices, including cellular telephones [19]. Examples of antennas commonly used for personal communication system (PCS) devices are shown in Figure 2.23. Antennas of different lengths, typically from a quarter-wavelength to a half-wavelength at the irradiation frequency, have been used for hand-held devices that operate at 800–900 MHz for cellular telephones and 1800–1900 MHz for PCS devices. Often, these monopole antennas are in the form of a whip antenna that can be pulled out to its full length during a telephone conversation. At times, a small length helical antenna is permanently mounted on the top of the handset through which the monopole antenna can be retracted and to which this

Figure 2.23 Examples of stationary and retractable/telescopic monopole antennas used in commercial cellular telephones.

monopole connects when it is completely pulled out, giving a monopole loaded with a helix at the bottom as the radiating antenna during telephone usage. Of necessity, these antennas use a finite-size reflecting plane, which is generally the metallic shielding box used either for the top RF section, or at times for the entire handset. Several authors have calculated and/or measured the radiation patterns of such monopole antennas mounted on a handset held in air or against the human head [20–22].

Given in Figure 2.24 are the computed and measured radiation patterns in the azimuthal plane for a typical commercial telephone at the center-band frequency of 835 MHz in

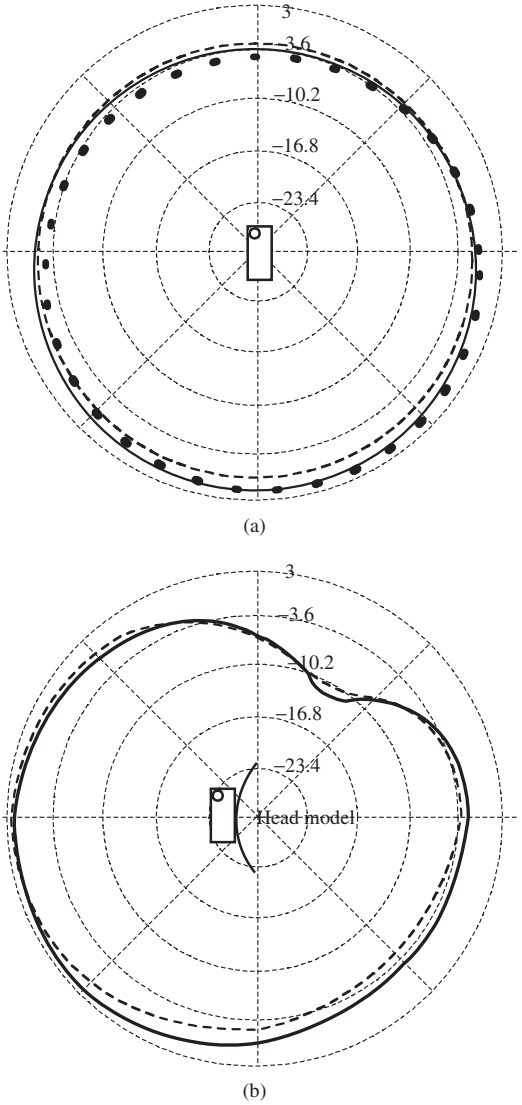


Figure 2.24 The computed and measured radiation patterns (in dBi) in the azimuthal plane for a commercial cellular telephone at 835 MHz (a) in free space and (b) held against the human head. (From Ref. 22.)

free space and in the presence of the human head, respectively [22]. This telephone uses a short helical antenna of diameter 4.2 mm, pitch 1.4 mm, and total length 16 mm. Dimensions of the handset are similar to those for many telephones and are approximately 2.4 cm × 5.2 cm × 14.7 cm along the three axes, respectively. Fairly similar radiation patterns are also obtained for antennas using either monopoles or monopole–helix combinations. In the presence of the human head, gains on the order of 1.0–3.0 dBi are obtained away from the human head, with gains in directions through the head that are substantially lower due to absorption in the tissues of the head. Radiation efficiencies for monopole antennas used in personal wireless devices are typically on the order of 30–50% with 30–50% of the power being absorbed in the head and another 5–15% of the power being absorbed in the hand [20, 23].

2.2.7.3 *Printed Monopole and Dipole Antennas*

Monopoles and dipoles are not limited to wire devices. Printed monopoles [24] and dipoles [25] have also been designed and are in many cases easier to fabricate than a wire-type device. Figure 2.25 shows one such printed monopole used for dual-band wireless local area network coverage in the Industrial, Scientific, and Medical (ISM) bands of 2.4–2.4835 GHz and 5.15–5.825 GHz, and Figure 2.26 shows a printed dipole used for measurement of electric fields for evaluation of cell phone compliance with RF exposure guidelines.

2.2.7.4 *Monopole Antennas for Medical Imaging*

Confocal imaging for breast cancer detection is an exciting application of antenna arrays in medical imaging. This method typically uses a single antenna scanned in a flat array pattern above the breast or a cylindrical array of very small broadband antennas [26]. For planar imaging, the patient lies face up, and the antenna is physically scanned in a plane above the breast [27–29]. For cylindrical imaging, the patient lies face down, with the breast extending into the cylindrical array through a hole in the table [30, 31]. Matching fluid surrounding

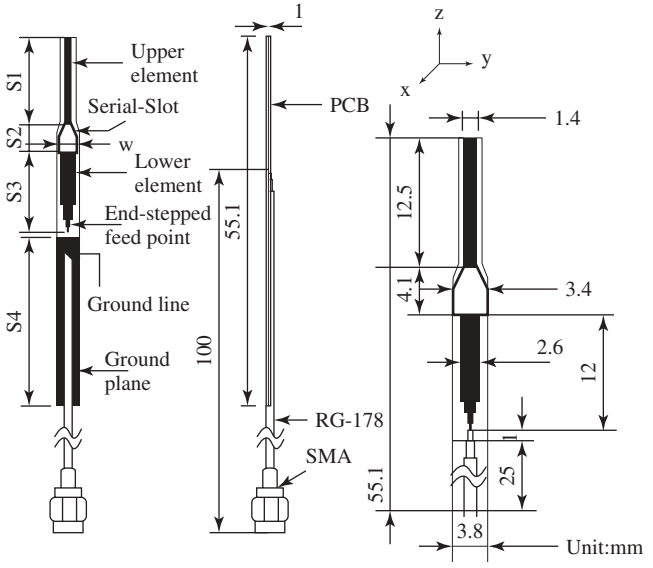


Figure 2.25 Geometry and dimensions of the proposed dual ISM-band antenna. (From Ref. 24.)

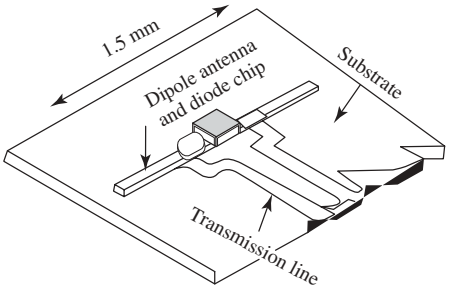


Figure 2.26 Miniature printed dipole antenna. (From Ref. 25.)

the breast tissue, similar to that used for microwave tomography, is suggested in this case. Both methods provide similar results [31]. One antenna in the array transmits an ultrawideband (UWB) pulse, which propagates into the breast tissue, where it is reflected off significant electrical discontinuities, and is received in parallel by the other antennas in the array. Knowing the physical spacing between the array elements, the different delays between the transmit antenna, scattering point, and receiving antenna can be calculated geometrically. The received pulses representing a specific point in space can then be time delayed appropriately for each antenna, added up, and integrated to indicate the magnitude of the scattered energy from that point in space. This is effectively correlating the signals received from that point at all antennas.

The antennas used for confocal imaging must be ultrawideband and small enough to fit within the relatively small array area. Resolution of less than 1 cm requires a bandwidth of at least 5 GHz. The lossy nature of tissue attenuates high frequency signals, limiting the upper frequency to about 10 GHz. Initially, resistively loaded bow tie antennas were suggested for the planar configuration [27–29, 32, 33], while dipole antennas were suggested for the cylindrical system [30, 31]. Resistively loaded V dipoles have also been proposed [34]. In the cylindrical configuration, multiple antennas are present in the array, although they are not simultaneously active. In the planar system, a single antenna is scanned over the surface, creating a synthetic antenna aperture. In order to overcome the inherent inefficiency of resistively loaded antennas, a modified ridged horn antenna operating from 1 to 11 GHz has been introduced [35]. Most of the antennas are designed to observe copolarized reflections from the breast tissue; however, using two resistively loaded bow-tie antennas in the shape of a Maltese cross (Figure 2.27) has also been proposed to pick up the cross-polarized reflections [29]. Cross-polarized reflections from simple tumor models were also examined in Refs. 35 and 36.

The antenna shown in Figure 2.27 [35] consists of two cross-polarized bow-tie antenna elements, an octagonal cavity behind the bow-tie elements, and a metal flange attached to the cavity. The broadband bow ties have flare angles of 45°. They are 1.67 cm long, which is a half-wavelength at 3 GHz in fat (similar to breast tissue). The octagonal cavity blocks waves radiated away from the breast tissue. The cavity is approximated as a circular waveguide filled with fat material for matching and size reduction. The first cutoff frequency is set to be 2 GHz for 2–4-GHz operation. The cavity length is a quarter-wavelength, which is 11 mm at 3 GHz. The flange consists of an inner and an outer component and is designed to block unwanted waves such as surface waves. The antenna performance does not change significantly when the flange size is varied

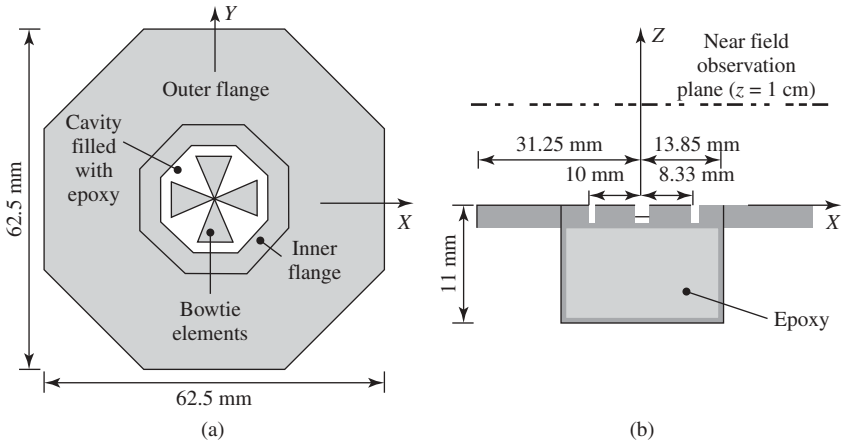


Figure 2.27 Cross-polarized antenna for confocal imaging. The properties of the substance inside the cavity and the medium outside the antenna are similar to fat ($\varepsilon_r = 9$ $\sigma = 0{:}2$ S/m). (From Ref. 35.)

between 10 and 6.25 cm; therefore the width of the outer flange is set to be 6.25 cm. The inner flange is designed to prevent possible electric field overshoot at the inner corners of the opening of the octagonal cavity or at the ends of the bow-tie elements. A slotline bow-tie antenna has also been proposed [37].

A resistively loaded monopole antenna shown in Figure 2.28 and suitable for use in a cylindrical array was proposed in Refs. [38–40]. Based on the Wu–King design [41, 42], this antenna was designed to be usable from 1 to 10 GHz immersed in canola oil ($\varepsilon_r = 3.0$) for matching to breast tissue. The antenna is fabricated using high frequency chip resistors (Vishay 0603HF) (Malvern, PA) soldered to a high frequency substrate (Rogers RO3203 series) (Rogers Corporation, Chandler, AZ). The substrate ($\varepsilon_r = 3.02$ and $\sigma = 0.001$ S/m) has electrical properties similar to those of canola oil. The antenna

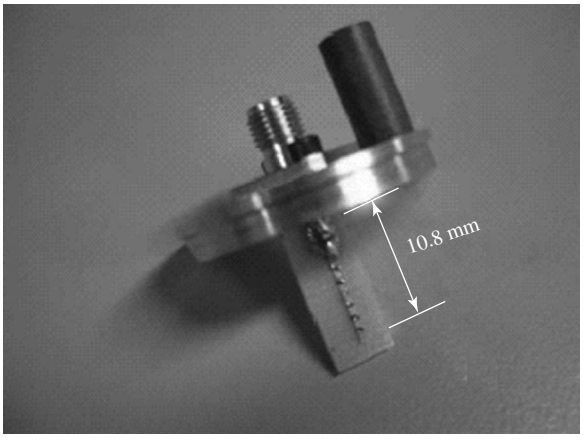


Figure 2.28 Fabricated resistively loaded monopole antenna soldered to an SMA connector and attached to a metal plug. (From Ref. 40.)

is soldered to a subminiature A (SMA) connector and attached to a metal plug for connection into the oil-filled test canister.

2.2.7.5 *Monopole Antennas for Hyperthermia Treatment of Cancer*

End-fed monopole antennas excited by a coaxial line have been used for a number of medical applications such as hyperthermia for cancer therapy [26] and the cardiac ablation catheter [43]. For these applications, both straight wire and small-diameter helix monopoles have been used.

Hyperthermia (HT) [44, 45] is a method of treating cancer by heating the body. The tissue is typically heated to 41–45°C for 30–60 min. Often this involves focusing the energy on the tumor region, relying on the tumor to be more sensitive to heat than the surrounding healthy tissue. HT has also been shown to increase the effectiveness of radiation and chemotherapy [46, 47]. The most commonly used frequencies for hyperthermia are 433, 915, and 2450 MHz. The type of antenna or antenna array used for HT depends on whether it is to be administered superficially, interstitially, or deep body.

Interstitial applicators for HT are typically monopole antennas made from coaxial cables with the center conductor extending beyond the outer ground shield of the cable [48]. These antennas have a tear-drop-shaped radiation pattern, so the majority of the heating is near the feed point of the antenna (where the ground shield stops), leaving the tip of the antenna extended beyond the usable heating range. The heating distribution can be made more uniform by varying the width of the conductor [49] or adding a choke to the antenna [50]. The heating pattern can be adjusted within the array by phasing the antenna elements [50] or by using nonuniform insulation [51].

2.3 LOOP ANTENNAS

2.3.1 Small Loop Antennas

A small loop antenna can be analyzed in much the same way as the infinitesimal dipole antenna described in Section 2.2.1. In fact, they are "dual" antennas, where the electric and magnetic fields are reversed as shown in Figure 2.29. A dual current relationship between the dipole and loop antennas is also used [5]. This relationship is

$$I^e L = I^m L = j\omega\mu I S$$

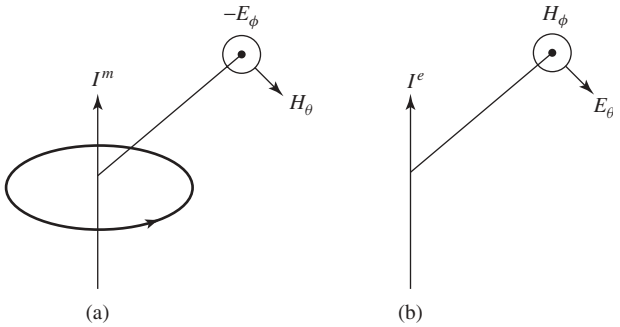


Figure 2.29 Dual radiation field components of (a) small loops and (b) infinitesimal dipole antennas.

where I^e is the electric current on the infinitesimal dipole of length L, I^m is the virtual magnetic current on the loop antenna, and I is the actual current on the loop of cross-sectional area S. The actual shape (circular, square, etc.) of a small loop does not matter, since the current I is assumed to be constant on this small loop. For a small circular loop $S=\pi a^2$. The electric and magnetic fields for a small loop antenna in the x-y plane are given by

$$\mathbf{E} = -\frac{1}{4\pi}(j\omega\mu IS)\sin\theta e^{-j\beta_0 r}\left(\frac{jk_0}{r}+\frac{1}{r^2}\right)\hat{\boldsymbol{\phi}} \tag{2.17}$$

$$\mathbf{H} = \frac{j\eta_0(j\omega\mu IS)}{2\pi k_0}\cos\theta\left(\frac{j\beta_0}{r^2}+\frac{1}{r^3}\right)e^{-jk_0 r}\hat{\mathbf{r}} - \frac{j\eta_0(j\omega\mu IS)}{4\pi k_0}\sin\theta\left(-\frac{k_0^2}{r}+\frac{jk_0}{r^2}+\frac{1}{r^3}\right)e^{-jk_0 r}\hat{\boldsymbol{\theta}} \tag{2.18}$$

Comparing these to Eqs. (2.1) and (2.2) for the infinitesimal dipole, the duality of these two antennas is clear. Similarly (comparing to Eqs (2.3) and (2.4)), their radiated fields (the $1/r$ terms) are

$$\mathbf{E} = \frac{\eta_0 k^2 IS}{4\pi r}\sin\theta e^{-jk_0 r}\hat{\boldsymbol{\phi}} \tag{2.19}$$

$$\mathbf{H} = -\frac{k^2 IS}{4\pi r}\sin\theta e^{-jk_0 r}\hat{\boldsymbol{\theta}} \tag{2.20}$$

Note that the radiation for a small loop is in the same direction as for the infinitesimal dipole. Although perhaps not initially intuitive, the small loop radiates no power directly above or below the loop. The maximum power is radiated directly out the sides of the loop, symmetric in the x-y plane. The directivity of the small loop is 3/2 [3].

The radiation resistance for a small loop antenna is

$$R_r = 20(k^2S)^2 \approx 31,200\left(\frac{S}{\lambda^2}\right)^2 \ \Omega \tag{2.21}$$

The total impedance of the small loop is a combination of its radiation resistance R_r, internal resistance R_i, internal inductance L_i, and external inductance L_e. Then $\mathrm{Z}=R_r+R_i+j\omega(L_i+L_e)$. This resistance is so small that the efficiency of small loop antennas is very low. They are therefore poor radiators and normally are used only in the receiving mode, such as for pagers and radios, where the signal-to-noise ratio is more important than efficiency. One method of increasing the radiation resistance is to use multiple coexistent loops made from several turns of wire wound into N multiple loops. The radiation resistance is then multiplied by N^2. Typically, a capacitor is placed in parallel with the loop to help tune out its inductance.

An approximate equation for the ohmic losses of a multiple-turn loop antenna is [3]

$$R_{\text{ohmic}} = \frac{Na}{b}R_s\left(\frac{R_p}{R_o}+1\right) \tag{2.22}$$

where a is the loop radius, b is the wire radius, N is the number of turns, R_p is the ohmic resistance per unit length due to the proximity of the other loops, R_o is the ohmic skin effect resistance per unit length ($= NR_s/2\pi b$), and R_s is the surface impedance of the conductor ($= (0.5\ \omega\mu_0/\sigma)^{1/2}$).

2.3.2 Ferrite Loaded Loops

In an effort to increase the radiation resistance of a loop antenna, the loop is wound around a ferrite core. The radiation resistance of this *ferrite loop* is then

$$R_f = 20\pi^2 \left(\frac{C}{\lambda}\right)^4 \mu_{cr}^2$$

where C is the circumference of the loop, and μ_{cr} is the effective relative magnetic permeability of the ferrite core. This depends on the core material and its length and diameter [52]. Because of their small size, ferrite loaded loops are often used in pocket transistor radios. In addition to providing a receiving antenna, the inductance it provides can be used in parallel with the RF amplifier tuning capacitor to create the tuned receiver circuit.

2.3.3 Larger Diameter Loops

Increasing the diameter of the loop antenna but keeping the current constant changes the radiation pattern, similar to expanding the length of a dipole antenna with sinusoidal current distribution. The radiated (far-zone) electric and magnetic fields for a circular loop of diameter a carrying constant current I are

$$\mathbf{E} = \frac{\eta_0 akI}{2r} J_1(ka\sin\theta)\hat{\boldsymbol{\phi}}$$

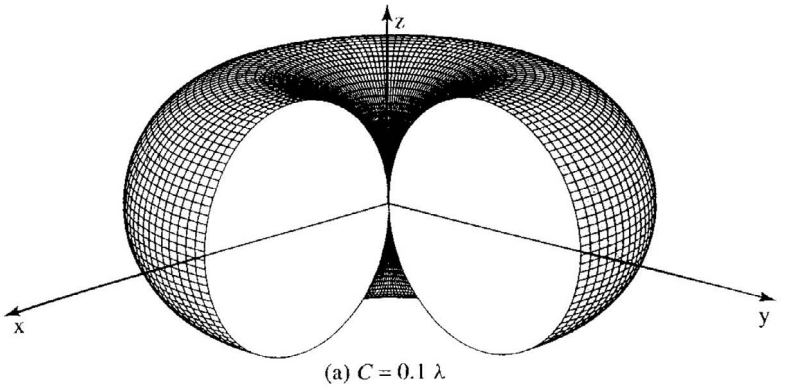
$$\mathbf{H} = -\frac{akI}{2r} J_1(ka\sin\theta)\hat{\boldsymbol{\theta}}$$

For small diameter loops, these expressions reduce to those shown in Eqs. (2.19) and (2.20). The radiation patterns for small and large diameter loops are shown in Figure 2.30.

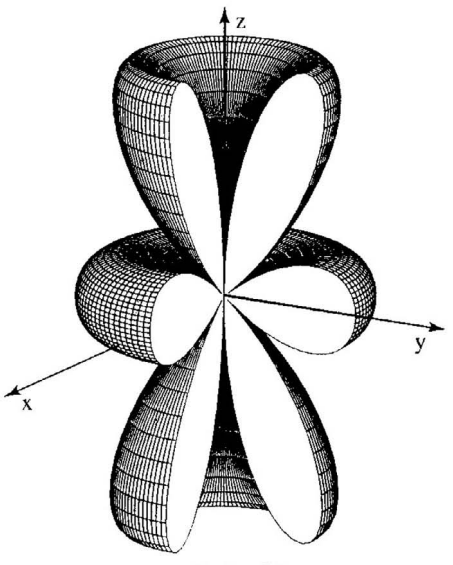
For larger loops, the current distribution can no longer be considered constant. Instead, it can be considered to be represented by the Fourier series

$$I(\phi) = I_0 + 2\sum_{n=1}^{M} I_n \cos(n\phi)$$

The field distributions are very complex for these large loops and can be found in Ref. 53. The direction of maximum radiation changes from its plane ($\theta = 90^\circ$) when the loop is small to its axis ($\theta = 0^\circ$ and 180°) when the circumference of the loop is about one wavelength. At one wavelength circumference, the directivity is about 3.4, as long as the wire is thin compared to the circumference of the loop [54]. The one-wavelength loop is commonly used in the design of helix antennas, which can be modeled as an array of loops. The impedance of circular loop antennas is shown in Figure 2.31.



(a) $C = 0.1\ \lambda$

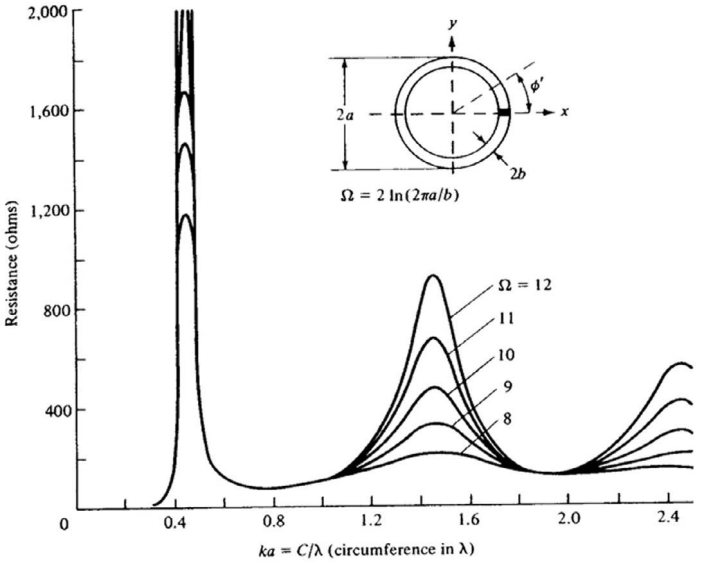


(b) $C = 5\ \lambda$

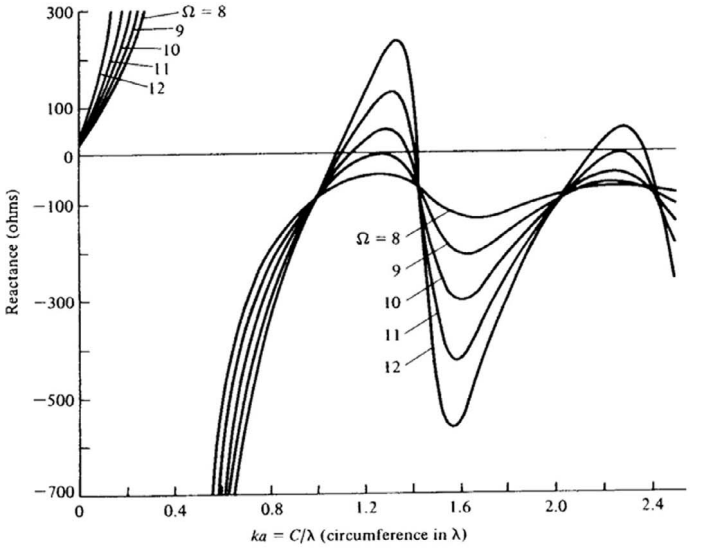
Figure 2.30 Three-Dimensional radiation patterns for a circular loop with constant current distribution: (a) circumference = 0.1 wavelength and (b) circumference = 5 wavelengths. (From Ref. 3.)

2.3.4 Applications of Loop Antennas

2.3.4.1 Wireless Telemetry for Implantable Devices An interesting application of loops is the wireless telemetry for medical devices. Since the invention of the first pacemaker almost fifty years ago [56], wireless implantable devices to treat a wide range of diseases and disabilities have gone through remarkable improvements. Currently, devices to treat or alleviate braycardia, chronic back pain, epilepsy, deafness, paralysis, and blindness, to name a few, are the subject of research in the biomedical field [57–67]. Common to these implanted devices is the need to receive data and/or power from a device external to the human body and often to transmit data to the external device regarding important information on the state of the patient and the implanted device. Neural prosthetic devices, in particular, present particular challenges since they may



(a) Resistance



(b) Reactance

Figure 2.31 Input impedance of circular loop antennas with Fourier current distributions. (From Ref. 55.)

need to stimulate continuously tens or hundreds of sites and thus possibly require more power and bandwidth than other implantable devices.

In the majority of wireless telemetry cases for implantable devices, inductively coupled coils are used [68]. In these, the time-varying magnetic field generated by the primary coil concatenates with the secondary coil, which results in an induced current in the implanted

coil. Frequencies are often lower than 50 MHz to ensure that the presence of the human body does not obstruct significantly the coupling between the coils. In this case, the most important parameters for the design of the telemetry system are the self- and mutual inductances of the coils. Several methods can be used to determine these parameters depending on the frequency of interest and the geometrical shape of the coils. For the simplest traditional coil geometries (circle, square), analytical approximations of self- and mutual inductances are often used, whereas more sophisticated methods (such as the partial inductance method and similar methods) can be used in the case of geometrically complex coils. A measure of the quality of coupling between two coils is given by the coupling coefficient K between two coils ($0 \leq K \leq 1$), defined by

$$K = \frac{M_{12}}{\sqrt{L_1 L_2}} \tag{2.23}$$

with M_{12} the mutual coupling between coils 1 and 2, and L_i the self-inductance of coil i. To maximize the power delivered to the load in these applications, usually a capacitor is inserted in parallel to the inductance of the coil and the resistance of the load to form a parallel resonant LC circuit. Many other parameters may affect the design of inductively coupled coils for biomedical telemetry systems, such as implant size, maximum power, temperature increase in the implanted device, and specific absorption rate of power (SAR given in W/kg) induced in the tissue. It is beyond the scope of this chapter to explore the details of the design of such systems, and the reader is referred to specific publications on the topic for additional insights into wireless telemetry systems (e.g., see Ref. 68 and references therein).

2.3.4.2 Vector Antennas Loop antennas have received considerable attention lately in the development of compact wireless communication systems when integrated with other antennas such as dipoles. This use of loop antennas leads to the possible development of compact multiple-input multiple-output (MIMO) systems which can find applications not only in communication systems, but also in direction of arrival estimation, sensor networks, and imaging. While, traditionally, the increase in channel capacity in a communication system has been achieved through spatial arrays of antennas, similar increases in channel capacity have been obtained through the use of "vector antennas" consisting of collocated loops and dipoles that can respond to more than one component of the electromagnetic field. One example of such antennas is given in Ref. 69, where three- and four-element vector antennas, consisting of one loop and two or three dipoles, respectively, are employed in a MIMO system. A number of research efforts that address the characterization of systems of collocated loops and dipoles can be found in the literature (e.g., Refs. [69–74]). In Ref. 69, for example, it is shown experimentally that in a rich multipath scattering environment systems with three- and four-element vector antennas at both the transmitter and receiver support three and four times more information, respectively, as compared to conventional systems consisting of sensors with single antennas. Figure 2.32 shows an example of a three-element vector antenna. In this particular system, since the loop must have a constant current distribution to retain the radiation characteristics of a magnetic dipole, the loop was realized by means of four pie-shaped sectors fed in phase at their corners [75]. This arrangement ensured that the currents were directed in opposite directions along adjacent feed lines, thus effectively nullifying any spurious radiation. Ultrawideband versions of such systems have also been developed for imaging purposes.

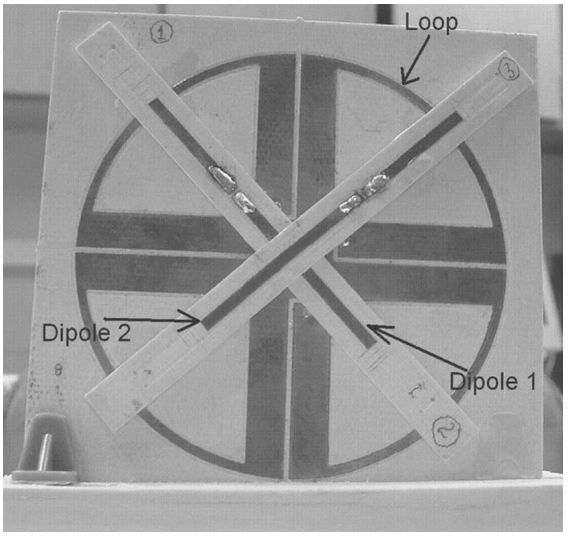


Figure 2.32 Vector antenna.

2.3.4.3 Magnetic Field Probes Loop antennas are commonly used as pick-up probes for magnetic fields. The shielded loop antennas [76] shown in Figures 2.33 and 2.34 are typical probe antenna designs. Figure 2.33 shows an unbalanced design, and Figure 2.34 shows a balanced design. The effective terminals of this type of antenna are at the gap AB. The inner conductor and the shield form a coaxial transmission line of length $h + \pi b$ connecting the gap at AB to the load impedance Z_s. Thus the effective input impedance seen at the terminal gap AB is the load impedance Z_s transformed by the transmission line of $h + \pi b$. The receiving loop antenna is analyzed by considering

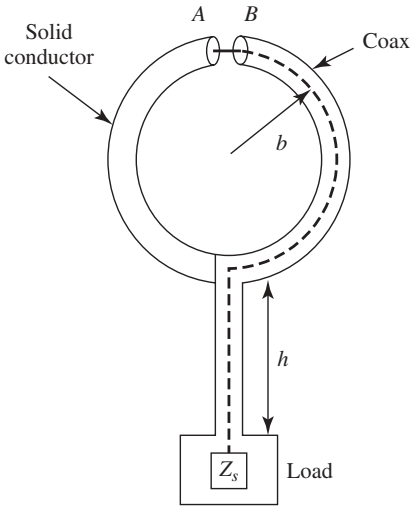


Figure 2.33 Shielded loop antenna made with solid conductor.

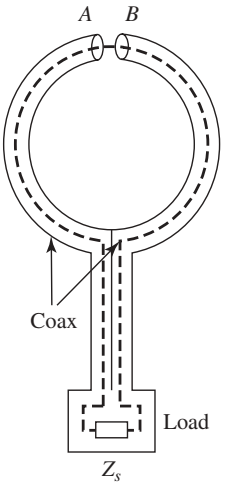


Figure 2.34 Shielded loop antenna made with coax.

the loop and transmission line separately. The incident external field produces a current on the external surface of the shield. The current passes through the effective input impedance Z_L (at terminals AB), producing a voltage across it, V_L. This voltage is then transformed to V_s (across load impedance Z_s) through the transmission line.

REFERENCES

1. C. A. Balanis, Antenna theory: a review, *Proc. IEEE*, Vol. 80, No. 1, pp. 7–23, January 1992.
2. R.C. Johnson (Ed.), *Antenna Engineering Handbook*, 3rd ed., McGraw-Hill, New York, 1993.
3. C. Balanis, *Antenna Theory, Analysis and Design*, 3rd ed., John Wiley & Sons, Hoboken, NJ, 2005.
4. E. C. Jordan and K. G. Balmain, *Electromagnetic Waves and Radiating Systems*, 2nd ed., Prentice Hall, Englewood Cliffs, NJ, 1968.
5. W. L. Stutzman and G. A. Thiele, *Antenna Theory and Design*, John Wiley & Sons, Hoboken, NJ, 1991.
6. ITT Handbook, *Reference Data for Radio Engineers*, 5th ed., Howard W. Sams & Co., Indianapolis, IN, 1973, Chap. 26.
7. R. C. Johnson and H. Jasik, *Antenna Engineering Handbook*, 2nd ed., McGraw-Hill, New York, 1984, Chap. 4.
8. W. Hong, N. Behdad, and K. Sarabandi, Size reduction of cavity-backed slot antennas, *IEEE Trans. Antennas Propag.*, Vol. 54, No. 5, pp. 1461–1466, May 1996.
9. G. A. Thiele, Analysis of Yagi–Uda antennas, *IEEE Trans. Antennas Propag.*, Vol. 17, pp. 24–31, 1969.
10. G. A. Thiele, Analysis of Yagi–Uda type antennas, *IEEE Trans. Antennas Propag.*, Vol. 14, No. 5, pp. 648–649, September 1966.
11. D. K. Cheng and C. A. Chen, Optimum spacings for Yagi–Uda arrays, *IEEE Trans. Antennas Propag.*, Vol. 21, No. 5, pp. 615–623, September 1973.
12. C. A. Chen and D. K. Cheng, Optimum element lengths for Yagi–Uda arrays, *IEEE Trans. Antennas Propag.*, Vol. 23, No. 1, pp. 8–15, January 1975.

13. O. P. Gandhi and J. Y. Chen, Electromagnetic absorption in human head from experimental 6-GHz hand-held transceivers, *IEEE Trans. Electromagn. Compat.*, Vol. 37, pp. 547–558, 1995.
14. C. T. Tai, On the theory of biconical antennas, *J. Appl. Phys.*, Vol. 10, pp. 1155–1160, 1948.
15. R. W. P. King, *Theory of Linear Antennas*, Harvard University Press, Cambridge, MA, 1956.
16. M. T. Ma and K. P. Spies, *A Simplified Formulation Computation for Conical Monopole Antennas*, U.S. Department of Commerce Report OT 74–53, 1974.
17. D. E. Hicks, *CB Radio Antennas*, Howard W. Sams & Co., Indianapolis, IN, 1967.
18. *The A.R.R.L. Antenna Book*, The American Radio Relay League, West Hartford, CT, 1956.
19. R. Luebbers et al., FDTD calculation of radiation patterns, impedance and gain for a monopole antenna on a conducting box, *IEEE Trans. Antennas Propag.*, Vol. 40, pp. 1577–1583, 1992.
20. M. A. Jensen and Y. Rahmat-Samii, EM interaction of handset antennas and a human in personal communications, *Proc. IEEE*, Vol. 83, pp. 7–17, 1995.
21. M. Okoniewski and M. A. Stuchly, A study of the handset antenna and human body interaction, *IEEE Trans. Microwave Theory Tech.*, Vol. 44, pp. 1855–1864, 1996.
22. G. Lazzi et al., Comparison of FDTD-computed and radiation patterns of commercial mobile telephones in the human head, *IEEE Trans. Antennas Propag.*, Vol. 46, pp. 943–944, 1998.
23. O. P. Gandhi, G. Lazzi, and C. M. Furse, Electromagnetic absorption in the human head and neck for mobile telephones at 835 and 1900 MHz, *IEEE Trans. Microwave Theory Tech.*, Vol. 44, pp. 1884–1897, 1996.
24. I.-Fong Chen, Chia-Mei Peng, and Sheng-Chieh Liang, Single layer printed monopole antenna for dual ISM-band operation, *IEEE Trans. Antennas Propag.*, Vol. 53, No. 4, pp. 1270–1273, April 2005.
25. H. I. Bassen, Electric field probes for cellular phone dosimetry, in Proceedings of the 19th International Congress IEEE/EMBS Society, October 30–November 2, 1997, Chicago, IL.
26. J.-C. Camart et al., Coaxial antenna array for 915 MHz interstitial hyperthermia: design and modelization—power deposition and heating pattern—phased array, *IEEE Trans. Microwave Theory Tech.*, Vol. 40, pp. 2243–2250, 1992.
27. E. C. Fear, S. C Hagness, P. M. Meaney, M. Okieniewski, and M. Stuchly, Enhancing breast cancer detection using near field imaging, *IEEE Microwave Mag.*, pp. 48–56, March 2002.
28. S. C. Hagness, A. Taflove, and J. E. Bridges, Two-dimensional FDTD analysis of a pulsed microwave confocal system for breast cancer detection: fixed-focus and antenna-array sensors, *IEEE Trans. Biomed. Eng.*, Vol. 45, pp. 1470–1479, December 1998.
29. S. C. Hagness, A. Taflove, and J. E. Bridges, Three-dimensional FDTD analysis of a pulsed microwave confocal system for breast cancer detection: design of an antenna-array element, *IEEE Trans. Antennas Propag.*, Vol. 47, pp. 783–791, May 1999.
30. X. Li and S. C. Hagness, A confocal microwave imaging algorithm for breast cancer detection, *IEEE MicrowaveWireless Components. Lett.*, Vol. 11, pp. 130–132, March 2001.
31. E. Fear and M. Stuchly, Microwave system for breast tumor detection, *IEEE Microwave Guided Wave Lett.*, Vol. 9, pp. 470–472, November 1999.
32. Xing Yun, Elise C. Fear, and Ronald H. Johnston, Compact antenna for radar-based breast cancer detection, *IEEE Trans. Antennas Propag.*, Vol. 53, No. 8, pp. 2374–2380, Aug. 2005.
33. M. A. Hernandez-Lopez, M. Pantoja, M. Fernandez, S. Garcia, A. Bretones, R. Martin, and R. Gomez, Design of an ultra-broadband V antenna for microwave detection of breast tumors, *Microwave Optical Technol. Lett.*, Vol. 34, No. 3, pp. 164–166, August. 2002.
34. S. C. Hagness, A. Taflove, and J. E. Bridges, Wideband ultralow reverberation antenna for biological sensing, *Electron. Lett.*, Vol. 33, No. 19, pp. 1594–1595, September 1997.

35. E. C. Fear and M. A. Stuchly, Microwave detection of breast cancer, *IEEE Trans. Microwave Theory Tech.*, Vol. 48, pp. 1854–1863, November 2000.
36. X. Li, S. C. Hagness, M. K. Choi, and D.W.W. Choi, Numerical and experimental investigation of an ultrawideband ridged pyramidal horn antenna with curved launching plane for pulse radiation, *IEEE Antennas Wireless Propag. Lett.*, Vol. 2, pp. 259–262, 2003.
37. X. Yun, E. C. Fear, and R. H. Johnston, Radar-based microwave imaging for breast cancer detection: tumor sensing with cross-polarized reflections, *IEEE Antennas Propag. Soc. Symp. Dig.*, Vol. 3, pp. 2432–2435, 2004.
38. C. J. Shannon, E. C. Fear, and M. Okoniewski, Dielectric-filled slotline bowtie antenna for breast cancer detection, *Electron. Lett.*, Vol. 41, No. 7, March 2001.
39. J. M. Sill and E. C. Fear, Tissue sensing adaptive radar for breast cancer detection: a study of immersion liquid, *Electron. Lett.*, Vol. 41, No. 3, pp. 113–115, February 2005.
40. J. M. Sill and E. C. Fear, Tissue sensing adaptive radar for breast cancer detection: preliminary experimental results, in *IEEE MTT-S International Microwave Symposium Digest*, Long Beach, CA, June 2005.
41. T. Wu and R. King, The cylindrical antenna with nonreflecting resistive loading, *IEEE Trans. Antennas Propag.*, Vol. 13, No. 3, pp. 369–373, May 1965.
42. T. Wu and R. King, Corrections to "The cylindrical antenna with nonreflecting resistive loading", *IEEE Trans. Antennas Propag.*, Vol. 13, No. 11, p. 998, November 1965.
43. S. Labonte et al., Monopole antennas for microwave catheter absorption, *IEEE Trans. Microwave Theory Tech.*, Vol. 44, pp. 1832–1840, 1996.
44. Special Issue of *IEEE Trans. Microwave Theory Tech.*, Vol. 34, 1986.
45. C. H. Durney and M. F. Iskander, in *Antenna Handbook*, Y. T. Lo and S. W. Lee (Eds.), Van Nostrand, New York, 1993.
46. P. K. Sneed and T. L. Phillips, Combining hyperthermia and radiation: How beneficial?, *Oncology*, Vol. 5, pp. 99–108, 1991.
47. C. C. Vernon, J. W. Hand, S. B. Field, et al., Radiotherapy with or without hyperthermia in the treatment of superficial localized breast cancer: results from five randomized controlled trials, *Int. J. Radiat. Oncol. Biol. Phys.*, Vol. 35, pp. 731–44, 1996.
48. P. F. Turner, Interstitial equal-phased arrays for EM hyperthermia, *IEEE Trans. Microwave Theory Tech.*, Vol. 34, No. 5, pp. 572–578, May 1986.
49. C. M. Furse and M. F. Iskander, Three-dimensional electromagnetic power deposition in tumors using interstitial antenna arrays, *IEEE Trans. Biomed. Eng.*, Vol. 36, pp. 977–986, October 1989.
50. R. D. Nevels, G. D. Arndt, G. W. Raffoul, J. R. Carl, and A. Pacifico, Microwave catheter design, *IEEE Trans. Biomed. Eng.*, Vol. 45, pp. 885–890, July 1998.
51. C. Manry, S. L. Broschat, C.-K. Chou, and J. A. McDougall, An eccentrically coated asymmetric antenna applicator for intracavity hyperthermia treatment of cancer, *IEEE Trans. Biomed. Eng*., Vol. 39, No. 9, pp. 935–942, September 1992.
52. E. A. Wolff, *Antenna Analysis*, John Wiley & Sons, Hoboken, NJ, 1966.
53. D. H. Werner, An exact integration procedure for vector potentials of thin circular loop antennas, *IEEE Trans. Antennas Propag.*, Vol. 44, No. 2, pp. 157–165, February 1996.
54. G. S. Smith, Loop antennas, in *Antenna Engineering Handbook*, 3rd ed., McGraw-Hill, New York, 1993, Chap. 5.
55. J. E. Storer, Impedance of thin-wire loop antennas, *AIEE Trans.*, Vol. 75, November 1956.
56. W. Greatbatch and C. F. Holmes, History of implantable devices, *IEEE Eng. Med. Biol.*, pp. 38–42, September 1991.

57. D. J. Woolons, To beat or not to beat: the history and development of heart pacemakers, *IEE J. Eng. Sci. Education*, Vol. 4, No. 6, pp. 259–268, December 1995.

58. R. Allan, Medtronic sets the pace with implantable electronics, *Electron. Design*, Vol. 51, No. 24, pp. 52–56, October 2003.

59. F. A. Spelman, The past, present, and future of cochlear prostheses, *IEEE Eng. Med. Biol.*, pp. 27–33, May 1999.

60. J. P. Rauschecker and R. V. Shannon, Sending sound to the brain, *Science*, Vol. 295, pp. 1025–1029, February 2002.

61. J. D. Weiland and M. S. Humayun, A biomimetic retinal stimulating array, *IEEE Eng. Med. Biol. Mag.*, Vol. 24, pp. 14–21, September 2005.

62. P. Walter, Z. F. Kisvarday, M. Gortz, N. Alteheld, G. Rossler, T. Stieglitz, and U. T. Eysel, Cortical activation via an implanted wireless retinal prosthesis, *Invest. Ophthalmol. Visual Sci.*, Vol. 46, pp. 1780–1785, May 2005.

63. E. Margalit, M. Maia, J. D. Weiland, R. J. Greenberg, G. Y. Fujii, G. Torres, D. V. Piyathaisere, T. M. O'Hearn, W. Liu, G. Lazzi, G. Dagnelie, D. A. Scribner, E. de Juan, and M. S. Humayun, Retinal prosthesis for the blind, *Survey Ophthalmol.*, Vol. 47, pp. 335–356, July 2002.

64. R. A. Normann, E. M. Maynard, K. S. Guilloty, and D. J. Warren, Cortical implants for the blind, *IEEE Spectrum*, pp. 54–59, May 1996.

65. E. Zrenner, Will retinal implants restore vision?, *Science*, Vol. 295, pp. 1022–1025, February 2002.

66. M. A. L. Nicolelis, Brain–machine interfaces to restore function and probe neural circuits, *Nat. Rev. Neurosci.*, Vol. 4, pp. 417–422, May 2003.

67. J. K. Chapin and K. A. Moxon (Eds.), *Neural Prostheses for Restoration of Sensory and Motor Function*, CRC Press, Boca Raton, FL, 2000.

68. M. Ghovanloo and G. Lazzi, Transcutaneous magnetic coupling of power and data, in *Wiley Encyclopedia of Biomedical Engineering*, M. Akay (Ed.), John Wiley & Sons, Hoboken, NJ, 2006.

69. A. Konanur, K. Gosalia, S. Krishnamurthy, B. Hughes, and G. Lazzi, Compact MIMO systems employing vector antennas for increased wireless channel capacity, *IEEE Trans. Microwave Theory Tech.*, Vol. 53, pp. 1837–1844, June 2005.

70. M. R. Andrews, P. P. Mitra, and R. de Carvalho, Tripling the capacity of wireless communications using electromagnetic polarization, *Nature*, Vol. 409, pp. 316–318, January 2001.

71. A. S. Y. Poon, R. W. Brodersen, and D. N. C. Tse, Degrees of freedom in multiple antenna channels: a signal space approach, *IEEE Trans. Inf. Theory*, Vol. 51, No. 2, pp. 523–536, February 2005.

72. T. Svantesson, M. A. Jensen, and J. W. Wallace, Analysis of electromagnetic field polarizations in multiantenna systems, *IEEE Trans. Wireless Commun.*, Vol. 3, No. 2, pp. 641–646, March 2004.

73. J. Andersen and B. Getu, The MIMO cube—a compact MIMO antenna, in *5th Wireless Personal Multimedia Communications International Symposium*, Vol. 1, pp. 112–114, October 2002.

74. D. D. Stancil, A. Berson, J. P. V. Hof, R. Negi, S. Sheth, and P. Patel, Doubling wireless capacity using co-polarized, colocated electric and magnetic dipoles, *Electron. Lett.*, Vol. 38, No. 14, pp. 746–747, July 2002.

75. A. G. Kandonian, Three new antenna types and their applications, *Waves Electrons*, February 1946.

76. G. S. Smith, Loop antennas, in *Antenna Engineering Handbook*, R.C. Johnson (Ed.), McGraw-Hill, New York, 1993, Chap. 5.

77. E. C. Fear and M. A. Stuchly, Microwave breast tumor detection: antenna design and characterization, *IEEE Antennas Propag. Soc. Symp. Dig.*, Vol. 2, pp. 1076–1079, 2000.

78. J. M. Sill and E. C. Fear, Tissue sensing adaptive radar for breast cancer detection—experimental investigation of simple tumor models, *IEEE Trans. Microwave Theory Tech.*, Vol. 53, No. 11, pp. 3312–3319, November 2005.

CHAPTER 3

Aperture Antennas: Waveguides and Horns

CHRISTOPHE GRANET, GRAEME L. JAMES, and A. ROSS FORSYTH

3.1 INTRODUCTION

3.1.1 Aperture Antennas

The most common type of aperture antenna is that of open-ended radiating waveguides including horn-type antennas; the latter being essentially flared-out waveguides. An associated aperture antenna is the radiating slot cut into the wall of a closed waveguide structure. Aperture antennas can be used directly in applications such as a source antenna for radiation pattern measurements or as an accurate gain standard. They also find use, either alone or in an array, in point-to-point radio communication links and are widely used as feeds in reflector antenna systems.

Whatever the application in mind, there are a number of general properties to be considered when designing aperture antennas. For example, common to all applications is the required radiation pattern characteristics including beamwidth and the degree of sidelobe level and front-to-back ratio that can be tolerated. Another feature that is now often required is dual-polarization operation and this places demands on the cross-polarization radiation performance of the antenna. Other important attributes of the radiation pattern include gain, the phase center and the main beam efficiency. Gain expresses the increase in power radiated in a given direction by the antenna to the average power radiated by an isotropic radiator emitting the same total power. For some horn-type antennas gain can be calculated accurately and these find use as gain standards. The phase center of an antenna is defined as the on-axis position that most closely approximates the center of curvature of the radiating wavefront. It is not a unique position and depends on a number of factors including the pattern beamwidth and the frequency of operation. Knowing the phase center of an aperture antenna is important if it is to be used as a feed on a reflector antenna or as an element in an array. With regard to the main beam efficiency, it is defined as the ratio of the power radiated within a specified cone angle θ from boresight to the total power radiated by the antenna. It is an important parameter in designing feeds and has direct implications to the commonly used figure-of-merit parameter, G/T, where G represents the gain of the antenna and T the antenna temperature.

Modern Antenna Handbook. Edited by Constantine A. Balanis

Aside from the radiation pattern characteristics, the other major electromagnetic design consideration is the return loss, or mismatch, as seen from the exciting waveguide. For horn-type aperture antennas, mismatch created by their flared-out waveguide geometry is always a matter of concern but not the mismatch to free space as this is usually negligible given that most horn apertures are somewhat in excess of a wavelength in dimension. For radiating waveguides and slots the reverse is the case and, given their generally smaller aperture dimension, the mismatch to free space can be of concern.

Whatever the requirements are for a given application, the bandwidth over which an aperture antenna is expected to operate has been increasing over the years and this can pose quite a challenge to the antenna designer. To assist in this design process, we present here a collection of basic useful design data collected over many years of research both from our own experience and those of others. Except for a brief outline of the classical approach to analyzing radiation from an aperture, we have purposefully left out theoretical analysis and complex electromagnetic theories as these topics are already covered in numerous textbooks (e.g., see Refs. 1–4). Instead, we have concentrated on gathering useful practical information to assist in the initial design of an aperture antenna for a specific purpose. Although we have tried to keep it simple, the notions of radiation pattern, gain, return loss, cross-polarization, and other basic properties of antennas are assumed to be understood. This chapter is therefore structured as a starting point where the reader, already knowledgeable in the field of antenna design (but not necessarily aperture antennas), can find simple equations and some basic rules from past experience which will enable a first prototype to be designed for a specific application. It will often be desirable, and perhaps necessary, to fine-tune this design further. This may be achieved by referring to more specific detailed design data available in the literature (such as the references already mentioned) or, more directly, by making use of one of the many standard numerical analysis software packages commercially available. While these can be slow in application, their use is appropriate here since the purpose is to refine a usable starting-point design deduced from the basic data given in this chapter.

We have also given consideration to manufacturing and overall mechanical aspects necessary to realize the antennas we discuss. These are topics that are rarely approached in the literature but in reality are crucial to the success or failure of a project. It is pointless to design an excellently performing antenna system from the electromagnetic point-of-view if it is not physically possible to manufacture it (at least at a reasonable cost) and to test it. This chapter therefore provides some basic fabrication information, including what material to use and how to approach the manufacturing issue. It reinforces the need for antenna designers to have a basic understanding of mechanical issues and have a good relationship with their mechanical engineers and machinists if projects are truly to succeed.

Before proceeding with applications, design data, and fabrication issues, we need to identify some key specifications that will govern the choice of an appropriate antenna.

3.1.2 Key Specifications

Each application will come with a number of specifications, and while overall specifications are often given for a complete system, it is necessary for the antenna designer

to extract those key specifications that will dictate or point toward the choice of a suitable antenna. Important general specifications for the antenna design include the following.

3.1.2.1 Bandwidth The frequency band over which the system is to operate is usually specified as a range (for single-band applications) or ranges (for multiple-band applications) of frequencies where the antenna has to satisfy a required return loss (or VSWR) and provide a radiation pattern adequate for the application. The latter will often include a specification on the cross-polarization isolation, especially where dual polarization operation is mandated. From this information, one can ascertain if the application is narrowband or wideband. The notion of "narrowband" or "wideband" is a somewhat ill-defined concept and the literature abounds in claims of "wideband" for what other researchers would call narrowband. To be specific, consider an antenna required to operate from $f_{\min}$ to $f_{\max}$, the bandwidth, in general, is calculated as

$$\%Bandwidth = 200\frac{(f_{\max} - f_{\min})}{(f_{\max} + f_{\min})} \tag{3.1}$$

Using this definition, a bandwidth below 40% is generally considered in this chapter to be narrowband whereas a bandwidth over 40% is considered to be wideband. For multiband applications, it is usual to consider the individual bands in isolation and apply Eq. (3.1) to each band independently. Thus in some multiband applications, one band may be classed as narrowband while another may be classed as wideband.

Often, the specifications will provide a "design frequency," that is, a frequency for which the design is to be optimum. If the application is required to operate in only one frequency band, our practice is to design the antenna for a design frequency of $f_c = \sqrt{f_{\min} f_{\max}}$ for narrowband operations and $f_c = 1.2 f_{\min}$ for wideband operations. For multiband systems a design frequency should be specified for each band and, as before, treated independently.

3.1.2.2 Gain/Antenna Efficiency It is common to specify the gain or antenna efficiency performance requirement over the required band(s). Of importance here is the gain of the whole antenna system, not just the gain of the aperture antenna in isolation. It is necessary therefore to take into account the mismatch and ohmic losses of the different components making up the entire antenna system. A well designed aperture antenna fed by a poorly designed feed system will remain a poorly performing system and vice versa.

3.1.2.3 The Working Environment An indication of the working environment in which the antenna is likely to operate is important. This will dictate the choice of materials and also the mechanical aspects of the system. One consideration is the proximity of potential interference to the system from neighboring antennas and other structures. An antenna working perfectly well in isolation may give completely unsatisfactory performance when placed in an environment with other nearby scatterers. This can include blockage from structure surrounding the antenna or a poor choice of radome that is fitted afterwards. Another consideration is any special needs for the antenna such as

pressurization, high-power handling, low passive intermodulation levels, robustness, and low-loss requirements.

3.1.2.4 Cost An indication of the cost of the antenna is often overlooked but there are often compromises that both the client and the antenna designer have to agree upon for an acceptable price to be reached. Cost sensitive issues can include choice of materials, finish, and the overall performance design level acceptable within a given budget.

3.2 COMMON APPLICATIONS OF APERTURE ANTENNAS

In this section, we discuss briefly some common applications of aperture antennas. This is not an exhaustive list of all possible applications as this would be too big an undertaking, but it is a selection of those applications most frequently faced by antenna engineers.

3.2.1 Standard-Gain Horns

Accurate antenna gain measurements are important in a wide range of applications such as satellite communications, remote sensing, electromagnetic compatibility measurements, and radar. To evaluate the gain of an antenna, it is customary to compare its performance against a well calibrated "standard." So-called standard-gain horns are used for this purpose.

The gain of a standard-gain horn is ideally measured inside a temperature-controlled anechoic chamber using the three-antenna extrapolation technique [5]. This involves measuring the power transmitted between the three pairs of antennas as a function of distance and then extrapolating these results to obtain the true far-field values for each antenna. The complex reflection coefficients of the antennas are also measured to correct for mismatches. An uncertainty of ± 0.05 dB is typical for most standard-gain horn antennas measured under these conditions. Several companies provide horns of this type but since this process is expensive, they are usually measured accurately at a few specially chosen frequencies only with the calculated gain being provided over the remaining useful bandwidth of the horn.

The most common type of horn used as a standard-gain horn is the pyramidal horn shown in Figure 3.1 (see also Section 3.4.2). The rectangular geometry of this horn enables easy manufacture and results in a low-cost antenna. Conical horns (Figure 3.2; see also Section 3.4.3) can also be used as standard-gain horns where the antenna gain, $G_{\mathrm{dBi}} > 15$ dBi. Often overlooked is the conical corrugated horn (Figure 3.3, see also Section 3.4.5) as a standard-gain horn when $G_{\mathrm{dBi}} > 12$ dBi. In calculating the gain of all of these horns it is assumed that the aperture terminates in an infinite flange. In practice, of course, horns have a finite flange, often just the thickness of the metal wall, and this leads to some inaccuracy in the calculated gain. The calculated results for the corrugated horn will, in principle, be more accurate since flange effects are far less severe for this type of horn (at least for large aperture horns considered here when $G_{\mathrm{dBi}} > 12$ dBi) than for the other horn types. This is a consequence of the electromagnetic field in the corrugated horn becoming zero, or close to zero, as the aperture edge is approached, which in turn will minimize any flange effects. This aspect is also largely the case with the dual-mode horn (Figure 3.4, see also Section 3.4.4) but only over a much reduced bandwidth around the center design frequency.

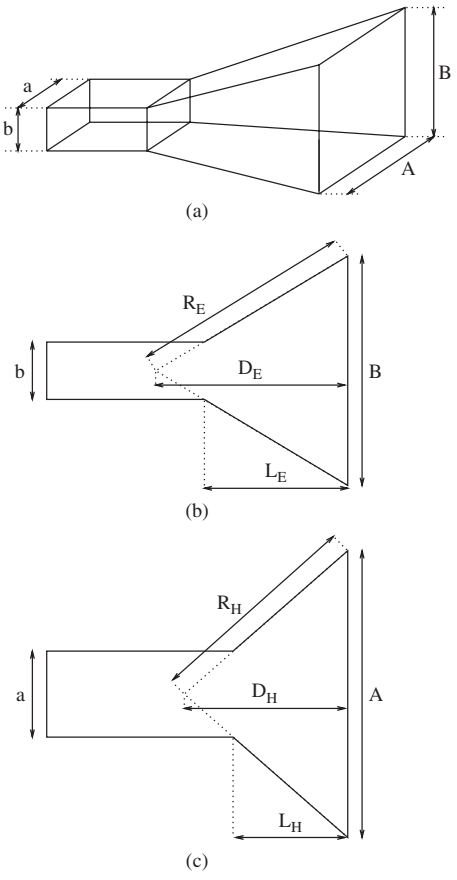


Figure 3.1 (a) Pyramidal horn, (b) *E*-plane view, and (c) *H*-plane view.

3.2.2 Prime-Focus Feeds

Prime-focus feeds encompass the types of feeds used mainly for circular-symmetrical or offset reflectors (see Chapter 8 of Ref. 3). The most common type of reflector used in the industry is the circular-symmetrical parabolic reflector defined typically by its *f/D* ratio, where f is the focal length of the parabola and D is the aperture diameter (see Figure 3.5). The *f/D* ratio provides the antenna designer with an idea of the subtended angle of the antenna and also a rough description of the dish: *f/D* values of less than 0.35 tend to be labeled as "deep reflectors." Note that often an offset-parabolic configuration, as in Figure 3.6 is specified [6, 7] [see also Chapter 8 of Ref. 3] and the offset parameter influences the pointing of the feed. It is interesting to point out that most commercial companies have a range of reflectors varying in diameter but with the same *f/D* ratio. It enables them to use the same feeds on all these antennas irrespective of the diameter size.

In the case of a circular-symmetrical parabolic reflector, for a given *f/D* ratio, the half-subtended angle at the feed (θ_e) can be calculated using the following formula:

$$\theta_e = \arctan\left(\frac{8f/D}{16(f/D)^2 - 1}\right) \tag{3.2}$$

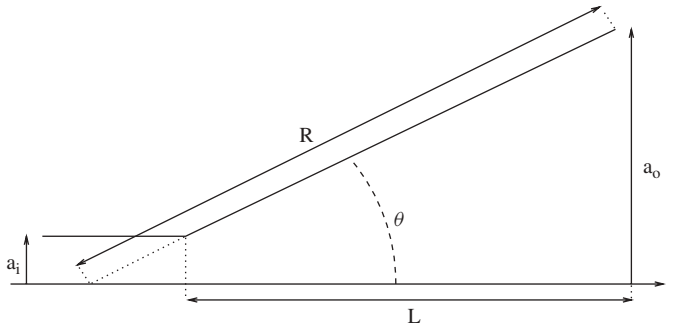


Figure 3.2 Conical horn geometry.

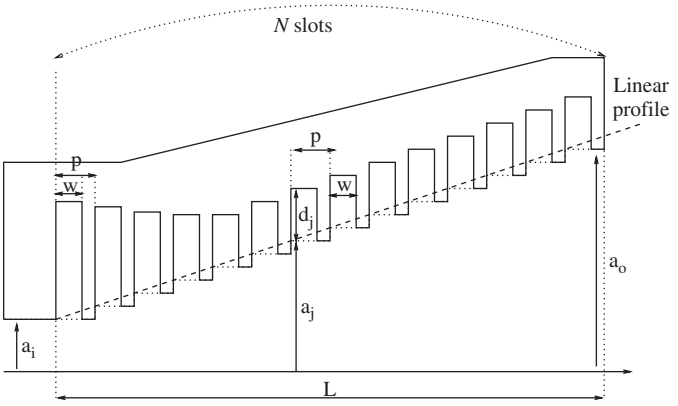


Figure 3.3 Conical corrugated horn.

In the case of an offset parabolic reflector, the offset geometry makes the definition of a single "half-subtended" angle (θ_e) more complex, but an approximate value for the angle can be defined by the average of the three following angles (see Figure 3.6):

$$\theta_L = 2\arctan\left(\frac{D+2h}{4f}\right) - \arctan\left(\frac{2f(D+2h)}{4f^2 - h(D+h)}\right) + \arctan\left(\frac{2fD}{4f^2 + h(D+h)}\right)$$

$$\theta_U = \arctan\left(\frac{2f(D+2h)}{4f^2 - h(D+h)}\right) + \arctan\left(\frac{2fD}{4f^2 + h(D+h)}\right) - 2\arctan\left(\frac{D+2h}{4f}\right)$$

$$\theta_Q = \arcsin\left[\frac{\frac{D}{2}}{\sqrt{\left(h+\frac{D}{2}\right)^2 + \left(\frac{D}{2}\right)^2 + \left(\frac{(2h+D)^2 + D^2}{16f} - f\right)^2}}\right]$$

$$\theta_e \approx \frac{|\theta_L| + |\theta_U| + 2|\theta_Q|}{4} \tag{3.3}$$

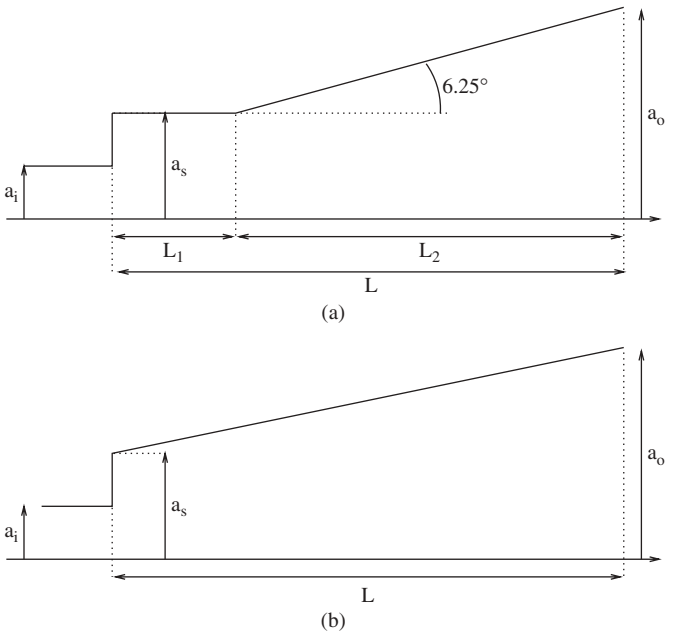


Figure 3.4 Dual-mode horn geometries: (a) classical "Potter" horn and (b) general case.

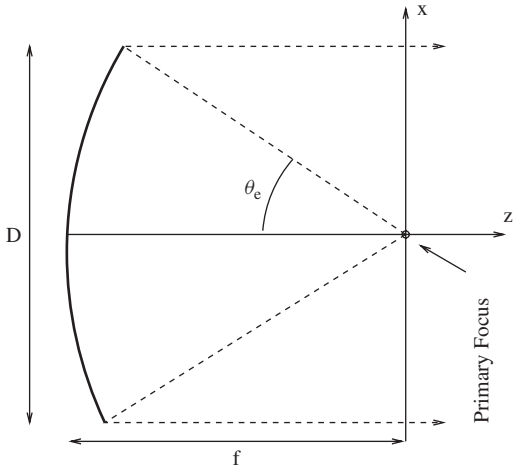


Figure 3.5 Geometry of a circular-symmetrical parabolic reflector.

This approximate value of θ_e is then used to work out the basic geometry of the feed by choosing the appropriate feed pattern taper at this angle. Note that antenna companies generally design their reflectors with values of *f/D* between 0.35 and 0.65 (i.e., $42^\circ \leq \theta_e \leq 71^\circ$), therefore most prime-focus feeds have a broad radiation pattern.

Among the most common types of prime-focus feeds used in the industry are the axially corrugated horns and the choke-ring waveguide (see Figure 3.7 and Sections 3.4.6

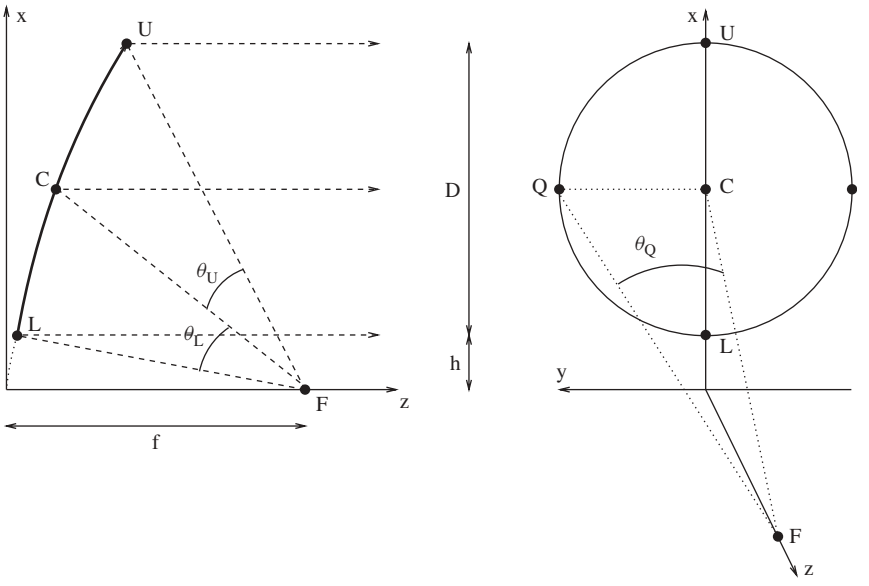


Figure 3.6 Offset parabolic reflector.

and 3.4.7). As an initial guide [8], design the horn to have a −12-dB taper at θ_e. In most cases, the reflector will be in the farfield of the feed and using the far-field patterns of the horn for design purposes is usually valid.

3.2.3 Secondary-Focus Feeds

The term "secondary focus" refers to feeds used to illuminate the subreflector of a dual-reflector antenna. There are numerous combinations of dual-reflector antennas but by far the most common is the Cassegrain antenna (see Chapter 8 of Ref. 3), (Figure 3.8), made of a parabolic main reflector and a hyperbolic subreflector. The main reflectors and subreflectors are either "classical," that is, exact parabolas and hyperbolas, or "shaped" by means of numerical techniques to enhance or tailor the radiation performance of the antenna to a specified performance. From a feed design point of view, the important angle to consider is the half-subtended angle θ_e from the axis of the antenna to the edge of the subreflector. In practice, this angle is usually much smaller than that for the prime-focus case and is typically in the range $10^\circ - 30^\circ$. Therefore the radiation pattern of the feed needs to be appropriately directive with the most common types of secondary feeds being the conical horn (see Figure 3.2 and Section 3.4.3), the dual-mode horn (see Figure 3.4 and Section 3.4.4) and the conical corrugated horn (see Figure 3.3 and Section 3.4.5).

If the application does not require high cross-polarization isolation, then a simple conical horn might be sufficient. If the radiation pattern of the feed is required to have Gaussian-type radiation properties with low cross polarization, then it will be necessary to consider a dual-mode horn for narrowband applications or a conical corrugated horn for wideband operations. For low sidelobes and optimal *G/T* the conical corrugated horn is often the better choice. As an initial guide [8], design the horn to have a −12-dB taper at θ_e. Secondary-focus feeds being mostly directive (i.e., with a large aperture size in

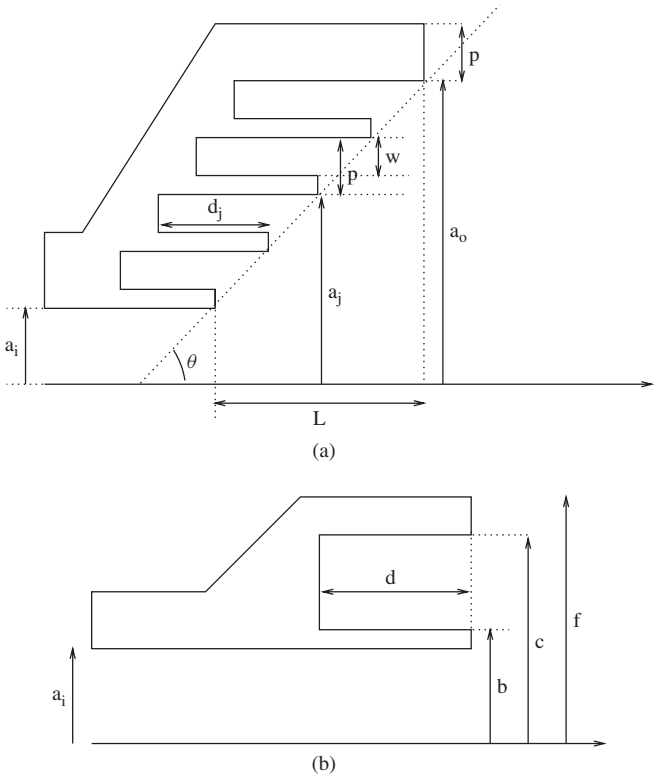


Figure 3.7 Geometry of (a) an axially corrugated conical horn and (b) a choke-ring waveguide.

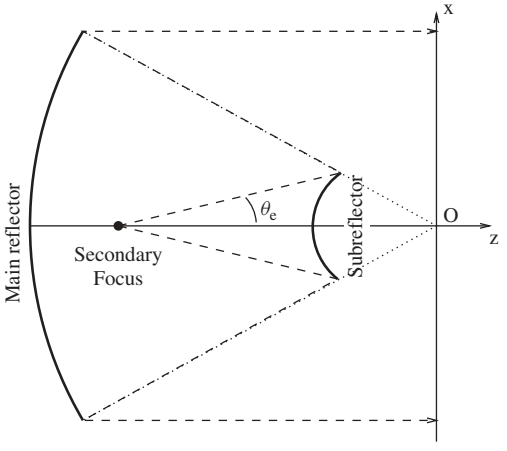


Figure 3.8 Geometry of a Cassegrain antenna.

terms of wavelength), it is necessary to determine if the subreflector lies in the far field or the near field of the feed and use the appropriate Fresnel-zone correction to calculate the antenna radiation pattern [9].

3.2.4 Direct Radiators

Direct radiators are horns or arrays that are used to radiate directly toward either a target or another antenna. There are many different applications for direct radiators and the main ones are:

- *Range Illuminators.* To measure antennas, you need to receive (transmit) from (to) the antenna under test at the required frequency. Mostly these tests are undertaken in an antenna test range, either enclosed as in an anechoic chamber or outside in dedicated outdoor facilities. For measurements on an outdoor test range, reflections from the ground and surrounding building or structures often demand as directive a range-illuminator antenna as possible and in this situation reflector-type antennas are more suitable than a horn radiator. However, for an indoor range, depending on the accuracy required from the measurements, a simple feed, like the pyramidal horn, or a more complex feed, like the corrugated horn, is appropriate. It is important to realize that the cross-polarization performance of the range illuminator will limit the accuracy at which you can measure the cross-polarization performance of the antenna under test. Also, the radiation pattern of the range illuminator will react with the geometry of the anechoic chamber and provide reflections that, even though small, can increase the errors of the measurement process [10].
- *Point-to-Point Communication.* This is where highly directive horns or reflector antennas are pointed at each other to operate as a receive/transmit network for data transfer. As is obvious for this application, a good line of sight is essential.
- *Earth Illumination from Satellites.* In this application, either a compact, directive horn with low sidelobes is used to illuminate the full earth efficiently as seen from a satellite [11] or a reflector antenna system is used to illuminate a "footprint" on the earth; that is, a specific region is illuminated with an appropriate power distribution.

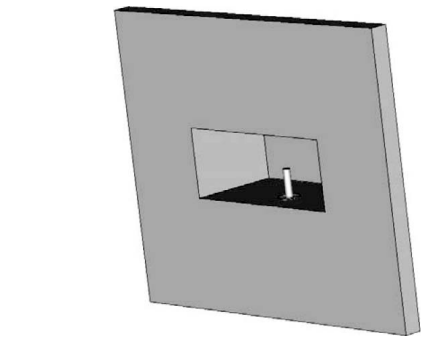
3.2.5 Array Elements

While any of the aperture antennas discussed here can be used as an array element, there are restrictions in practice. One of these is the physical size of the aperture (and indeed the length of the antenna). To avoid grating lobes one must have adjacent elements separated by not much more than half-a-wavelength, which is simply not physically possible with many of the antennas described here. There are applications, however, where the main interest is radiation near boresite (e.g., see Ref. 12) and grating lobes may not be of major concern. Nonetheless, in all cases issues such as the array factor, mutual coupling between neighboring elements, and other aspects of the array environment need to be addressed. To this end, the reader is referred to Chapter 12 of this book for further details on array design. There is, however, one type of array of particular interest to aperture antennas: the slotted waveguide array.

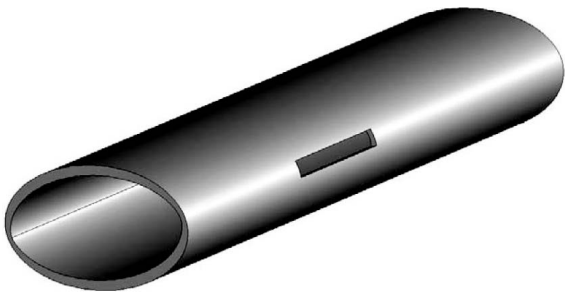
3.2.5.1 Slot Antennas: General Comments A slot antenna can be defined as any narrow aperture in a metal surface that is excited to radiate. While in general the aperture has no restriction in geometry—and in the past a number of different shaped slots have been used for specific purposes—the most common type of slot used is the simple narrow slot, where the slot width (the narrow dimension) is small compared to the wavelength. If this slot is assumed to lie in an infinite ground plane then it is complementary to a flat wire antenna having the same dimensions as the slot. Thus if the slot is excited at the same point as the wire antenna, the field radiated by the slot is obtained from the solution for the flat wire antenna simply by interchanging the electric and magnetic fields.

A slot in an infinite ground plane radiates on both sides of the plane. In many applications radiation is required on one side only and one way to achieve this is to back the slot by a cavity, as shown in Figure 3.9a. Of course, in practice, the ground plane will be finite in size and the radiation pattern will be affected to a degree dependent on the location of the slot to the edge of the plane. In other cases, such as a slot antenna on the body of an aircraft or missile, the slot will reside on a closed structure approximated by a cone or cylinder, as in Figure 3.9b, and radiation from the slot will be modified by the curvature of the surrounding structure. A widespread application of a radiating slot is where it is cut into the wall of a waveguide. While this can be any type of waveguide, the most common use is with a rectangular waveguide, where the three possible slot configurations are shown in Figure 3.10.

Unlike the other aperture antenna types discussed in this chapter, slot antennas do not easily lend themselves to meaningful simple design formula. A slot in an infinite



(a)



(b)

Figure 3.9 (a) cavity-backed slot in a finite ground plane and (b) slot cut in a cylindrical structure.

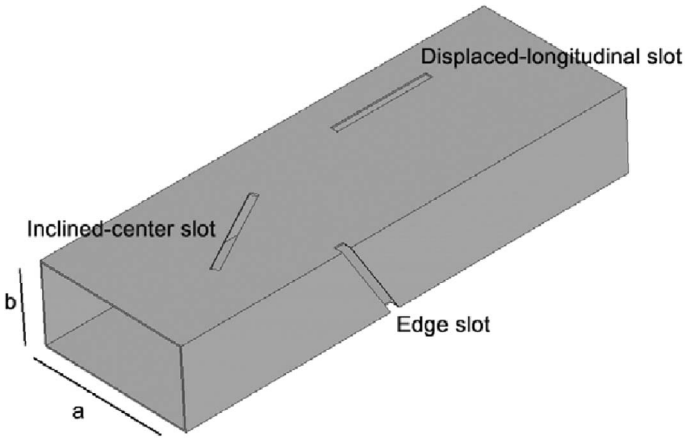


Figure 3.10 Various types of waveguide slots.

ground plane is readily characterized by the equivalent flat wire antenna, but when it is cavity backed in a practical application the solution is no longer a simple one [13–15]. Furthermore, when this antenna is embedded in a finite structure, edge and curvature effects need to be addressed. While these and other applications of slot antennas have been developed extensively over many years, deducing simple design formulas from this body of work has remained elusive.

3.2.5.2 Slotted Waveguide Arrays The slot array versions of the slot configurations shown in Figure 3.10 are given in Figure 3.11. In all cases the slots are cut to resonate at the desired (center) frequency and alternate either in inclination (in the case of edge slots or inclined slots in the broad wall) or in location (either side of the center line in the case of the longitudinal slots) to counteract the 180° phase shift of the waveguide mode at each (waveguide) half-wavelength. The array can operate as either a resonant array or a traveling-wave array. In the former case the array terminates in a short circuit and radiates an "averaged" broadside pattern, whereas a traveling-wave array terminates in a matched load. In this case the main beam radiates at a frequency-dependent angle, θ, given by

$$\theta = \sin^{-1}\left(\frac{\lambda}{\lambda_g} - \frac{\lambda}{2\delta_y}\right) \tag{3.4}$$

where λ is the free-space wavelength and λ_g is the waveguide mode wavelength.

While in principle broadside radiation is possible at a given frequency where $\delta_y = \lambda_g/2$, at this spacing, coherency of slot reflections can result in severe mismatch at the input port. Therefore a slot spacing closer to $\lambda/2$ is chosen, giving a beam angle close to broadside. Note that the traveling-wave array should be well matched otherwise the reflected power will radiate an unwanted sidelobe at an angle of $-\theta$.

Waveguide slot arrays are favored for their relatively simple construction and the "in-built" phasing of the elements by the waveguide mode. Disadvantages include their inherent narrow bandwidth, especially for the resonant array, and the difficulty of providing for dual polarization, although a number of designs have been attempted for the

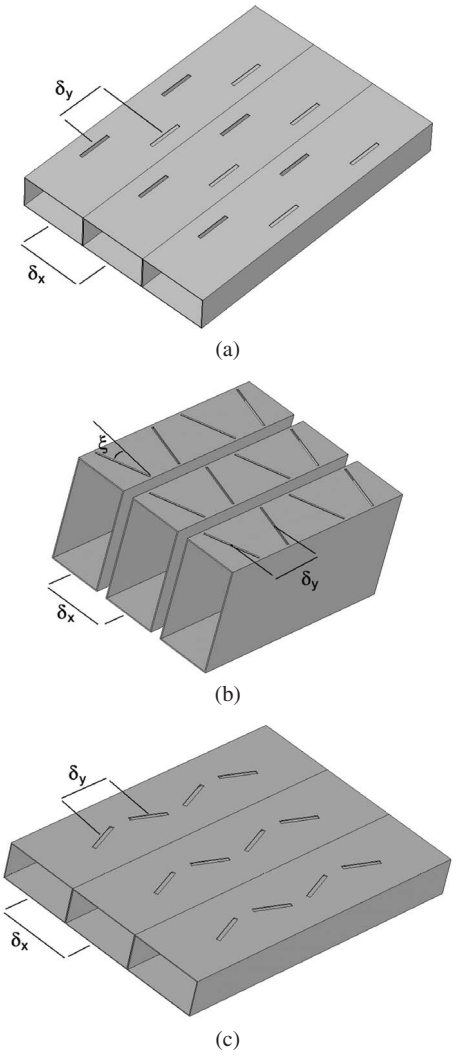


Figure 3.11 (a) Displaced-longitudinal slot array, (b) edge-slot array, and (c) inclined-center slot array.

latter (e.g., see Refs. 16–18) with varying degrees of success. With regard to the slot types, edge-slot arrays are commonly used given that the element spacing, δ_x, can be kept within a half-wavelength to avoid grating lobes in wide-angle scanning applications. This is not possible with the two broadside wall arrays. The inclined slot and edge-slot array produce an unwanted cross-polar component, which could be a serious drawback in some applications. Of the three slot types, the broad-wall inclined slot has the most disadvantages and is usually avoided in practice.

To shape the radiated power distribution along the line of elements (in effect, the aperture distribution along the array), one can vary the position of the slots in order to change the value of the radiation conductance in line with the required power distribution.

In the case of the longitudinal broad-wall slots, the radiation conductance, G_{rad}, is given to a good approximation by [19]

$$G_{\text{rad}} = \frac{2.1a\lambda_g}{b\lambda}\cos^2\left(\frac{\pi\lambda}{2\lambda_g}\right)\sin^2\left(\frac{\delta\pi}{a}\right) \tag{3.5}$$

where δ is the slot displacement from the guide center. The equivalent expression for edge slots is given by

$$G_{\text{rad}} = \frac{30\lambda^3\lambda_g}{73\pi a^3 b}\left[\frac{\sin(\xi)\cos\left(\frac{\pi\lambda\sin(\xi)}{2\lambda_g}\right)}{1-\left(\frac{\lambda\sin(\xi)}{\lambda_g}\right)^2}\right]^2 \tag{3.6}$$

Where ξ is the inclination angle of the slot. The case of the broad-wall inclined slot is of little interest here.

Aside from rectangular waveguides, other waveguide types used in slot arrays include the ridge waveguide [20, 21], where the aim is to reduce the broad-wall dimension sufficiently to minimize grating lobes in a scanning array of longitudinal slots. TEM-mode coaxial waveguides and circular waveguides (supporting either the TE_{11} or TM_{01} mode depending on the required radiation pattern) have also been used as the basis for slotted waveguide arrays [22]. Another application of slot arrays has been the work of Ando and colleagues (e.g., see Ref. 23) in designing radial line slot antennas for both circular and linear polarization.

3.2.6 Less Common Applications of Aperture Antennas

We mention briefly here other applications where aperture antennas, especially horns, are required or where exceptions to the rules as presented earlier apply.

- *Tracking Feeds.* These feeds are designed to enable a target to be tracked. An example is an Earth-station antenna that is required to maintain a link to a given satellite. The dominant mode is used to receive the signal information while the higher order modes are used to provide the tracking information. The higher order modes perform the tracking function by using radiation patterns that have a null on axis. Tracking feeds need to be designed to propagate all the modes from the point at which the tracking and signal information is split-up and abstracted [2]. Some other examples of tracking feeds are presented in Refs. 24 and 25.
- *Probes.* These are simple circular or rectangular waveguide working in the dominant mode used mainly for antenna measurements in near-field scanners. A number of commercial companies supply circular- or rectangular-waveguide probes and the easiest solution is to buy directly from a supplier. However, should the need arise to design one, be aware that the wall thickness at the aperture of the probe has a significant effect on the radiation pattern [26].
- *Monopulse.* Monopulse feed systems are often used in radar applications. Monopulse is the term used to describe a means whereby radar detection and tracking information are derived from the same radar return, that is, simultaneous lobe comparison. In a radar system, when a reflector is used, monopulse can be

achieved with the use of a four-horn feed [27] or a multimode horn feed with four feeding waveguides [28, 29].

- *Prime-Focus Feeds for Large f/D Paraboloid Reflectors.* For these reflectors, the angle θ_e can become small and a highly directive "secondary-focus" feed is required. These reflectors are sometimes used for very small aperture terminal (VSAT) antennas or for applications where very low antenna sidelobes are required. In the latter case, the horn is designed to under illuminate the reflector antenna with a taper of -20 to -25 dB at θ_e instead of the more conventional -12 dB.
- *Compact Dual-Reflector Systems.* Here the feed can be located close to the subreflector as in a compact Cassegrain system or in a splash-plate reflector design. In these cases, the half-subtended angle to the edge of the subreflector is large and a standard primary-focus feed is required. Care needs to be taken to understand the effect of the close proximity of the subreflector to the performance of the feed, especially the match. It might be necessary, for example, to include a matching "tip" on the subreflector [30, 31].

3.3 SPECIFIC ENVIRONMENTS

3.3.1 Space Applications

An important aspect to consider for space applications is the premium on space on a spacecraft. The cost of launching a satellite is strongly related to the size and the weight of the spacecraft and every effort in reducing both these quantities is crucial. However, there is always a trade-off between compactness and performance as it is not always possible to design a compact antenna with the required performance. Note that "compact" in this instance relates to the physical size of the antenna, not the size in wavelengths. Much effort is needed to reduce the size and weight of, say, an L-band (1-GHz) or an S-band (2-GHz) antenna because of its sheer physical size, but not nearly so much effort would be required for an equivalent Ka-band (20- to 30-GHz) antenna given that the amount of design effort would probably, in practice, only provide a gain of a few tens of grams and would therefore not be cost effective.

Another important factor for on-board satellite applications is that satellites tend to be small and require a number of antennas to perform their missions. There are thus risks of interference and obstruction from surrounding structures, including antennas. Therefore satellite antennas tend to have low sidelobes to limit these interactions. A global approach to the design of the antenna platform is necessary to make sure the specifications are met.

Another point to keep in mind is the choice of materials. This will be covered in Section 3.5 from the fabrication and mechanical point of view, but the choice of materials, especially dielectric materials, is important from an electromagnetics point of view as well. A number of dielectric materials are "space qualified" in that they have been designed and approved for space applications. For example, their "degassing" rate is small to minimize their loss of properties over time in space. They also need to withstand the harsh changes in temperature from extremely hot when the satellite is in full sun to extremely cold when the satellite is shadowed by the Earth. Furthermore, the antenna and its components have to be sturdy enough to withstand the vibrations and extreme forces exerted during the launch. For further reading on satellite antennas see Ref. 32.

3.3.2 Harsh Environments

What defines a harsh environment is basically conditions that could be harmful to the antenna itself to the extent that it adversely affects its performance. Some of the electromagnetics issues are highlighted briefly here while fabrication and mechanical design aspects are considered in Section 3.5.

Some adverse conditions an antenna may have to deal with include the following.

- *Water.* It is important to keep water out of the system as it will degrade the performance of the antenna and can damage components. Also, for regions where heavy rainfalls are the norm, high frequency operations (e.g., Ka-band) can suffer and become almost completely ineffective during heavy rainfall.
- *Ice and Snow.* The formation of ice and snow on antennas will severely degrade the performance and, depending on where the antenna is to be used, some kind of anti-icing system is required.
- *Maritime Applications.* As for satellite applications, the available space on the deck of a ship is limited and there are numerous possible scatterers near the antennas. For these applications, the same global approach as for satellite antennas is necessary. If access to the satellite is necessary from a ship, a complex fast system of pointing and tracking is usually necessary and this has repercussions on the mechanical aspect of the antenna.
- *Wind.* If the antenna is to operate in a windy environment, the mechanical design has to be such that the antenna is given a "wind-factor" rating. This is related to the survivability of the antenna itself but also to the performance of the antenna. For example, for Earth stations required to maintain contact with a satellite, there may be a necessity to keep the antenna steady under heavy wind loading and to prevent damage to the antenna.
- *Vibrations and High Shocks.* This is especially important for flyaway-type antenna systems and military applications. From an antenna designer point of view, it means the need to avoid designing components with small pieces that could break easily and to design the antenna to have a performance tolerance with respect to small discrepancies in shape. Parameter studies are therefore necessary to see if certain parts of the design have a significant influence on the performance of the system and convey that information to the mechanical designer.

3.4 DESIGN DATA

In this section, we present some simple procedures to provide an initial design for a range of standard waveguide and horn radiators that can be tailored further to meet specific needs. A number of examples are given with accurately computed radiation patterns using up to three different methods.

- *Method 1.* The mode-matching method [33, 34] is used to calculate the propagation of the waveguide modes through the antenna. To calculate the radiation pattern we use the classical approach outlined in the next section, where the aperture is assumed to be located in an infinite ground plane.

- *Method 2.* This is the same as Method 1 but assumes the radiating aperture has a 2-mm thick flange to take into account the effect of a finite flange.
- *Method 3.* The results are calculated using *CST MicroWave Studio* to provide an example of a commercially available electromagnetics analysis software package that can simulate accurately the radiation pattern of a horn antenna, including the effects of the horn body. We acknowledge, of course, that *CST MicroWave Studio* is not the only commercial, or necessarily the best, package available to do this. For example, Ansoft Corporation's *HFSS* is another commonly used commercial software package.

The performance of some of the examples in terms of gain, return loss, and cross-polarization isolation calculated by Method 2 is also shown. These results help define what type of bandwidth is covered by each type of horn and indicate which one would fulfil the requirements of a given application.

The horns presented here are the main examples of a large number of horn types available. For further reading on horn types, refer to Refs. 2 and 3, or for a more general view on antenna feeds in general, see, for example, Refs. 35 and 36.

Again, we would like to emphasize that the following design procedures are only meant to provide you with a "starting point" that can be optimized to meet your specific requirements. There are numerous ways to optimize a horn antenna and a literature search will provide you with a number of techniques to optimize each of the basic horn geometries showcased here.

3.4.1 Basic Radiation Characteristics

The classical approach in calculating the field radiated by an aperture is by means of an equivalent source distribution in the plane of the aperture. To place the underlying theoretical analysis in perspective, consider first the general case of a source distribution of magnetic currents $\mathbf{M}(\mathbf{r}')$ and electric currents $\mathbf{J}(\mathbf{r}')$ within a volume V' bounded by the surface S' as illustrated in Figure. 3.12a. The solution for the electromagnetic field, $\mathbf{E}_1$, $\mathbf{H}_1$, in the source-free region external to the volume V' is given by [37, p. 10]

$$\begin{aligned}\mathbf{E}_1(\mathbf{r}) &= -jk\int_{V'}\left(\mathbf{M}(\mathbf{r}')\times\hat{\mathbf{R}}+\sqrt{\frac{\mu}{\varepsilon}}[\mathbf{J}(\mathbf{r}')-\{\mathbf{J}(\mathbf{r}')\bullet\hat{\mathbf{R}}\}\hat{\mathbf{R}}]\right)G(\mathbf{r},\mathbf{r}')\,dV'+0\left(\frac{1}{R^2}\right)\\ \mathbf{H}_1(\mathbf{r}) &= jk\int_{V'}\left(\mathbf{J}(\mathbf{r}')\times\hat{\mathbf{R}}-\sqrt{\frac{\varepsilon}{\mu}}[\mathbf{M}(\mathbf{r}')-\{\mathbf{M}(\mathbf{r}')\bullet\hat{\mathbf{R}}\}\hat{\mathbf{R}}]\right)G(\mathbf{r},\mathbf{r}')\,dV'+0\left(\frac{1}{R^2}\right)\end{aligned} \tag{3.7}$$

where μ and ε are the permeability and permittivity of free space, $R\hat{\mathbf{R}}=\mathbf{r}-\mathbf{r}'$ and $G(\mathbf{r},\mathbf{r}')$ is the free-space Green's function

$$G(\mathbf{r},\mathbf{r}')=\frac{\exp(-jk|\mathbf{r}-\mathbf{r}'|)}{4\pi|\mathbf{r}-\mathbf{r}'|}$$

In formulating radiation problems it is often convenient to replace the actual sources of the field with an equivalent source distribution. For the original problem shown in Figure 3.12a, if we now specify a source-free field $\mathbf{E}_2$, $\mathbf{H}_2$ internal to the volume V' as

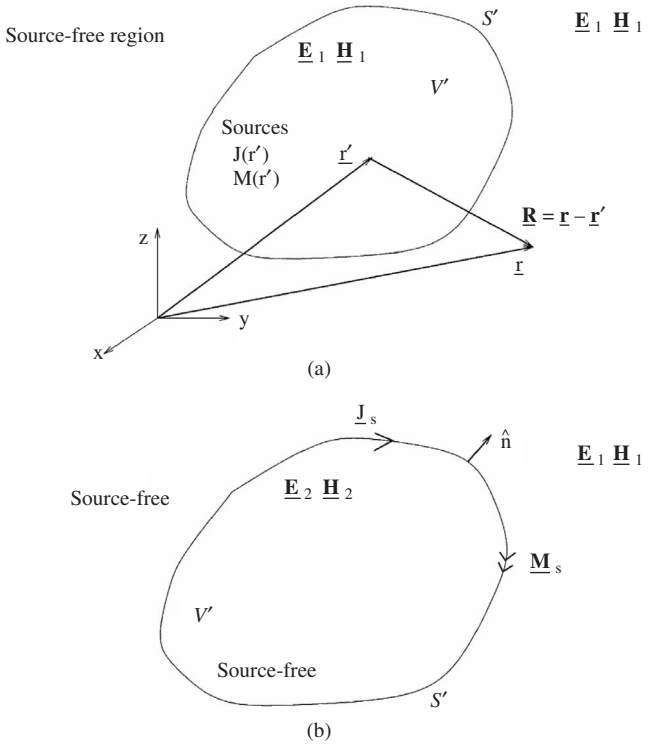


Figure 3.12 (a) Source distribution contained within a volume V' (b) equivalent source distribution.

shown in Figure 3.12b while maintaining the original field external to V', then on the bounding surface S' one finds there must exist surface currents

$$\mathbf{J}_s = \hat{\mathbf{n}} \times (\mathbf{H}_1 - \mathbf{H}_2) \quad \mathbf{M}_s = (\mathbf{E}_1 - \mathbf{E}_2) \times \hat{\mathbf{n}} \tag{3.8}$$

to account for the discontinuity in the tangential components of the field. We can now substitute these sources into eq. (3.7) and perform the integration over S' to obtain the field $\mathbf{E}_2$, $\mathbf{H}_2$ internal to V' and $\mathbf{E}_1$, $\mathbf{H}_1$ external to V'. Thus the equivalent source distribution of Eq. (3.8) has produced the same field external to the volume as the original sources contained within it. Since the specification of the field $\mathbf{E}_2$, $\mathbf{H}_2$ is arbitrary, it is often advantageous to choose a null field for the region containing the actual sources. The surface currents now become

$$\mathbf{J}_s = \hat{\mathbf{n}} \times \mathbf{H}_1 \qquad \mathbf{M}_s = \mathbf{E}_1 \times \hat{\mathbf{n}}$$

With a null field chosen for $\mathbf{E}_2$, $\mathbf{H}_2$ we are at liberty to surround S' with a perfect conductor (electric or magnetic) to remove one of the equivalent source distributions. In such cases the remaining currents radiate in the presence of a perfectly conducting obstacle and the electromagnetic field cannot be determined from Eq. (3.7) as the medium is no longer homogeneous. However, if we choose that the surface S' divides all of

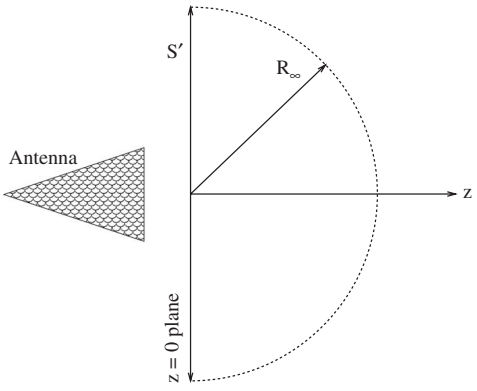


Figure 3.13 Infinite plane for aperture field method.

space into two regions by coinciding with the $z = 0$ plane with the sources contained in the half-space $z < 0$, then the perfectly conducting infinite plane at $z = 0$ can be replaced by the image of the surface currents. This gives a current distribution of either $\mathbf{J}_s = 2\hat{\mathbf{n}} \times \mathbf{H}_1$ or $\mathbf{M}_s = 2\mathbf{E}_1 \times \hat{\mathbf{n}}$ (depending on the conductor used) radiating into an unbounded homogeneous medium. The field for $z > 0$ can then be calculated using eq. (3.7). This is the basis of the *aperture field method* used extensively in the past to evaluate the radiated field from an aperture antenna. The $z = 0$ plane lies either in the aperture plane of the antenna or at some position in front of it, as shown in Figure 3.13. We then have three possible solutions for the field in the $z > 0$ space derived from Eq. (3.7) for $kR \gg 1$.

Solution I: $\mathbf{J}_s = 2\hat{\mathbf{n}} \times \mathbf{H}_1 \qquad \mathbf{M}_s = 0$

$$\begin{aligned} \mathbf{E}_{\mathrm{I}}(\mathbf{r}) &= -jk \int_{S'} \left(\sqrt{\frac{\mu}{\varepsilon}} \left[\mathbf{J}_s(\mathbf{r}') - \left\{ \mathbf{J}_s(\mathbf{r}') \bullet \hat{\mathbf{R}} \right\} \hat{\mathbf{R}} \right] \right) G(\mathbf{r}, \mathbf{r}')\, dS' \\ \mathbf{H}_{\mathrm{I}}(\mathbf{r}) &= jk \int_{S'} \left(\mathbf{J}_s(\mathbf{r}') \times \hat{\mathbf{R}} \right) G(\mathbf{r}, \mathbf{r}')\, dS' \end{aligned} \tag{3.9}$$

Solution II: $\mathbf{J}_s = 0 \quad \mathbf{M}_s = 2\mathbf{E}_1 \times \hat{\mathbf{n}}$

$$\begin{aligned} \mathbf{E}_{\mathrm{II}}(\mathbf{r}) &= -jk \int_{S'} \left(\mathbf{M}_s(\mathbf{r}') \times \hat{\mathbf{R}} \right) G(\mathbf{r}, \mathbf{r}')\, dS' \\ \mathbf{H}_{\mathrm{II}}(\mathbf{r}) &= -jk \int_{S'} \left(\sqrt{\frac{\varepsilon}{\mu}} \left[\mathbf{M}_s(\mathbf{r}') - \left\{ \mathbf{M}_s(\mathbf{r}') \bullet \hat{\mathbf{R}} \right\} \hat{\mathbf{R}} \right] \right) G(\mathbf{r}, \mathbf{r}')\, dS' \end{aligned} \tag{3.10}$$

Solution III: $\mathbf{J}_s = \hat{\mathbf{n}} \times \mathbf{H}_1 \qquad \mathbf{M}_s = \mathbf{E}_1 \times \hat{\mathbf{n}}$

$$\begin{aligned} \mathbf{E}_{\mathrm{III}} &= \tfrac{1}{2}(\mathbf{E}_{\mathrm{I}} + \mathbf{E}_{\mathrm{II}}) \\ \mathbf{H}_{\mathrm{III}} &= \tfrac{1}{2}(\mathbf{H}_{\mathrm{I}} + \mathbf{H}_{\mathrm{II}}) \end{aligned} \tag{3.11}$$

Provided the fields are known exactly along the $z = 0$ plane, then identical results will be yielded by these three solutions. In this situation it is preferable to use either Eq. (3.9) or (3.10) since Eq. (3.11) requires both the electric and magnetic fields tangential to the chosen plane. In practice, however, it is necessary to approximate these fields so the three formulations will, in general, yield differing results. For a radiating aperture, Solution I often gives the most accurate results since this solution assumes the aperture is surrounded by a perfectly electrically conducting infinite flange. In reality, of course, the metal flange surrounding the aperture will be finite in size, but if this flange extends about half-a-wavelength or greater beyond the aperture, then surprising accurate corrections can be made to Solution I by using the geometrical theory of diffraction (GTD) in conjunction with the aperture field method. (An example can be found in Section 7.5 of Ref. 37.) When the surrounding flange is somewhat smaller than half-a-wavelength in extent, approximate high frequency techniques such as the GTD cannot be relied on and, in this situation, Solution III is often the better approximation to use. The remaining option, Solution II, where a perfect magnetic conductor is assumed to surround the aperture, is rarely, if ever, used in predicting the radiated field from an aperture antenna.

There are basically two ways to specify the radiation characteristics of a horn antenna; one is based on the required gain from the given horn while the other is based on the required shape of the main beam of the radiation pattern. While some applications call for specific shaping of the main beam as a function of rotational (azimuthal) direction, such as a fan-shaped beam, the large majority of applications call for a simple "cigar-shaped" rotationally invariant (symmetrical) main beam. In this case the radiation pattern is determined by the beamwidth (or half-beamwidth) for a given illumination level (or taper) below the maximum value. There are of course other characteristics, such as sidelobe levels and cross-polarization performance, which are of interest, but these are determined largely by the choice of horn type.

In the case of most high performance horns, the radiation pattern over a given frequency range (defined later for different horn types) is essentially symmetrical and Gaussian shaped in nature, thereby enabling us to estimate to a good approximation the half-beamwidth of the radiation pattern for a given taper. Specifying the taper in decibels (dB) at a given angle (θ_o) as shown in Figure 3.14, the relative (Gaussian) radiation pattern can be expressed as

$$Relative\ power_{\mathrm{dB}}(\theta) = 20\log\{\exp[-\alpha(\theta/\theta_o)^2]\} \tag{3.12}$$

where

$$\alpha = \frac{-Taper\ \ln(10)}{20} \qquad \text{(note that } Taper \text{ is a negative value, i.e., } -12 \text{ dB)}$$

(*Note*: As a -12-dB taper is often used as the feed pattern illumination toward the edge of a reflector antenna at an angle θ_e (in degrees.) [8], the required gain for the feed is approximated by

$$Gain_{-12\ \mathrm{dB\ at}\ \theta_e} \approx 9.7219\log_{10}\left(\frac{0.6\ln(10)}{\theta_e}\right) + 43.539\ \mathrm{dBi} \quad \text{for } 10^\circ \le \theta_e \le 80^\circ$$

Aside from the Gaussian-type radiation pattern, the so-called cosine-to-the-power-q type relative radiation pattern is sometimes specified for the feed [38]. In this case, we use

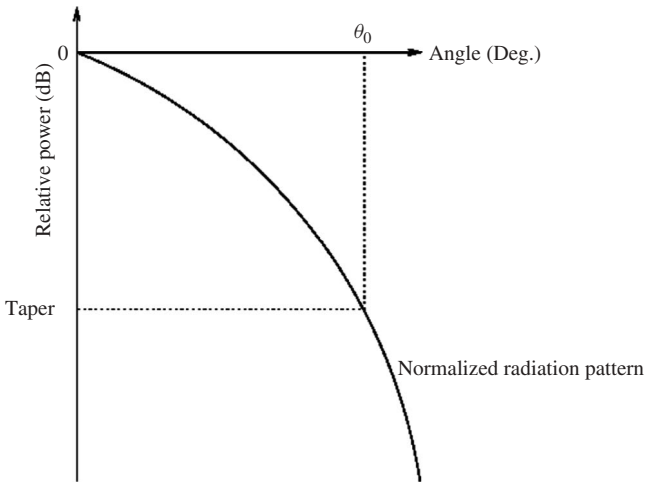


Figure 3.14 Relative radiation pattern specification.

the formula

$$Relative\ \ power_{\mathrm{dB}}(\theta) = 20\log\lfloor(\cos\theta)^q\rfloor \tag{3.13}$$

where

$$q = \frac{Taper\ \ \ln(10)}{20\ln(\cos\theta_o)} \qquad \text{(as before, } Taper \text{ is a negative value)}$$

(Note that the directivity of a cosine-to-the-power-q type radiation pattern can be derived in close-form as $D_{\mathrm{dB}} = 10\ \log_{10}(2(2q+1))$ [Ref. 39].)

Basically, Gaussian and "cosine-to-the-power-q patterns are very similar and specifying the radiation pattern using either of these functions leads, essentially, to equivalent solutions. Our own preference is to specify Gaussian radiation patterns and this definition will be used for the most part within this text.

In practice, tapers of -10, -12, -15, and -18 dB are commonly required and in Figure 3.15 we plot for these tapers the resultant half-beamwidth and gain values for a Gaussian radiation pattern. We also plot the -3-dB taper used in many textbooks. This, however, has very little practical application and is only shown here for completeness. Using Figure 3.15, we can approximate the half-beamwidth for a chosen taper of a horn that has a given gain or, inversely, estimate the gain of a horn from the required half-beamwidth at the chosen taper.

The radiation pattern of a horn with a circular symmetry has its maximum of cross-polarization level in the 45°-plane and, in the ideal case, no cross polarization in the E- and H-planes. It is therefore customary to represent the radiation pattern of a horn by plotting the copolar E-, H-, and 45°-planes and only the cross polarization in the 45°-plane. Note that for the examples of the pyramidal horns shown next, the cross-polarization maximum is not necessarily present in the 45°-plane, but the level of cross polarization shown in that plane gives a good indication of the overall cross-polarization isolation.

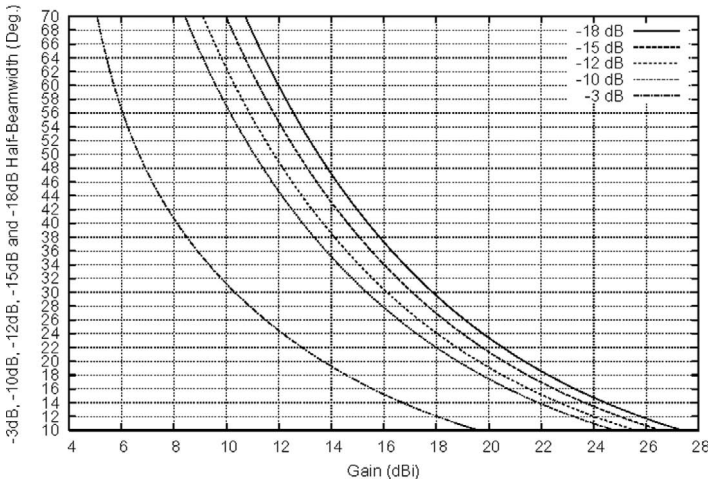


Figure 3.15 Approximate half-beamwidth as a function of gain for a nominally Gaussian radiation pattern.

3.4.2 Pyramidal Horns

The basic geometry of a pyramidal horn is shown in Figure 3.1. To design a pyramidal horn for a given design frequency and a given gain, it is necessary to choose the appropriate input waveguide dimensions for that frequency band. Most antenna designers choose from an extensive list of standard waveguide band sizes provided by waveguide manufacturers. For a given frequency (i.e., a given wavelength λ), a given gain expressed in dBi, G_{dBi}, choose the standard rectangular waveguide (that corresponds to the frequency band you require) with dimensions a and b (see Figure 3.1) and then follow the procedure below [40] to generate the dimensions of a horn that will closely match your requirements. Note that the quantities a, b, and λ are expressed in meters.

$$G_l = 10^{G_{\text{dBi}}/10} \tag{3.14}$$

$$A = 0.096aG_l^{0.232} + 0.42\lambda G_l^{0.503} - 0.193b \tag{3.15}$$

$$R_H = A\sqrt{\frac{1}{4} + \left(\frac{A}{3\lambda}\right)^2} \tag{3.16}$$

$$L_H = (A-a)\sqrt{\left(\frac{R_H}{A}\right)^2 - \frac{1}{4}} \tag{3.17}$$

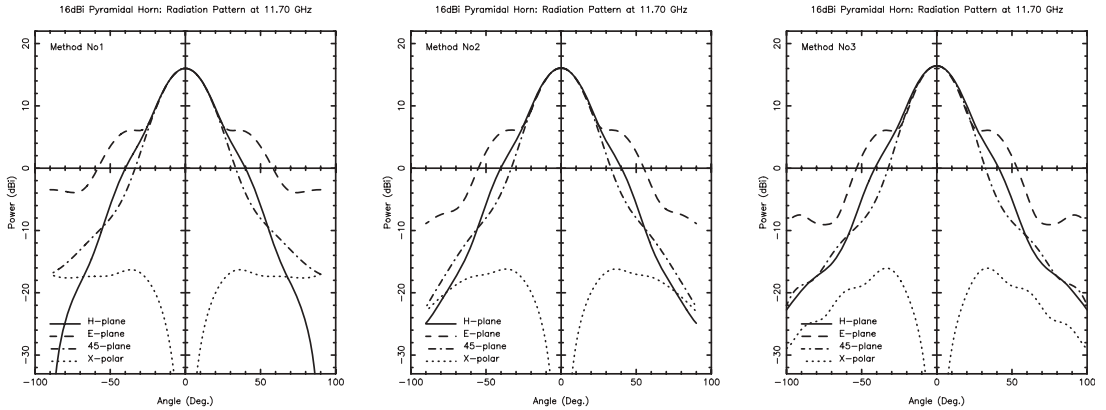
$$D_H = \sqrt{R_H^2 - \left(\frac{A}{2}\right)^2} \tag{3.18}$$

$$B = \tfrac{1}{2}\left[b + \sqrt{b^2 + 8L_H\lambda}\right] \tag{3.19}$$

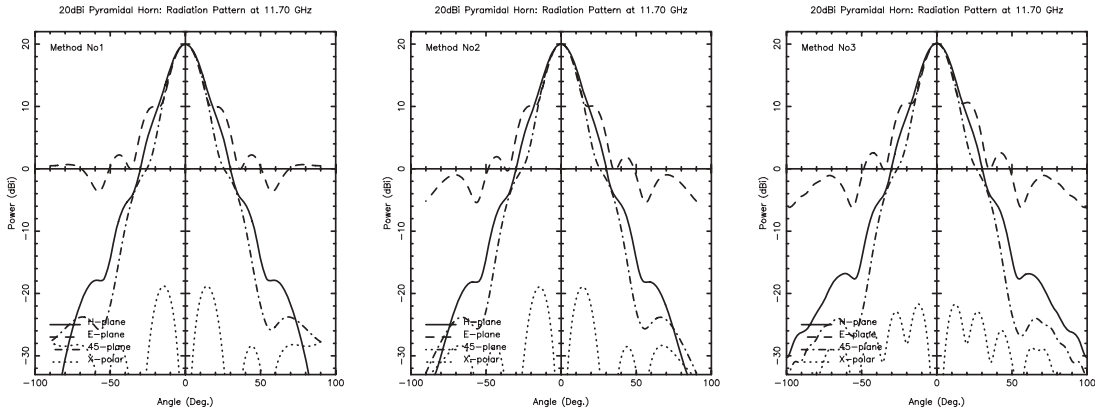
$$R_E = \frac{B}{2}\sqrt{1 + \left(\frac{B}{\lambda}\right)^2} \tag{3.20}$$

$$L_E = (B - b)\sqrt{\left(\frac{R_E}{B}\right)^2 - \frac{1}{4}} \tag{3.21}$$

$$D_E = \sqrt{R_E^2 - \left(\frac{B}{2}\right)^2} \tag{3.22}$$



(a)



(b)

Input Waveguide: WR75: a = 19.05 mm and b = 9.525 mm

Specification	A (mm)	B (mm)	R_E (mm) D_E (mm)	R_H (mm) D_H (mm)	$L_E = L_H$ (mm)	G_{dBi} (calculated)
16 dBi gain at 11.70 GHz	71.12	54.68	64.42 58.33	74.79 65.80	48.17	Method 1:16.01 Method 2:16.08 Method 3:16.38
20 dBi gain at 11.70 GHz	112.60	88.70	159.80 153.52	174.28 164.94	137.03	Method 1:19.97 Method 2:20.01 Method 3:20.14

(c)

Figure 3.16 Pyramidal horn examples.

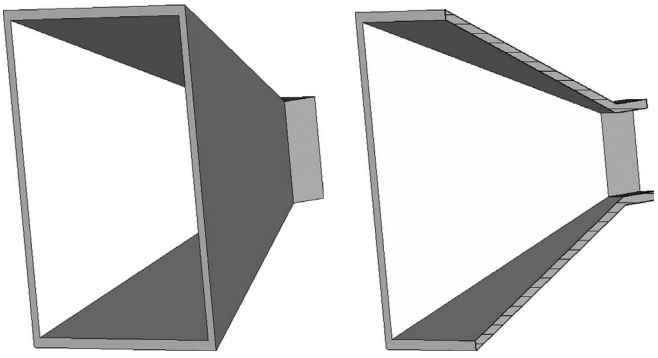


Figure 3.17 Geometry of the nominal 16-dBi pyramidal horn: (a) full view and (b) longitudinal section.

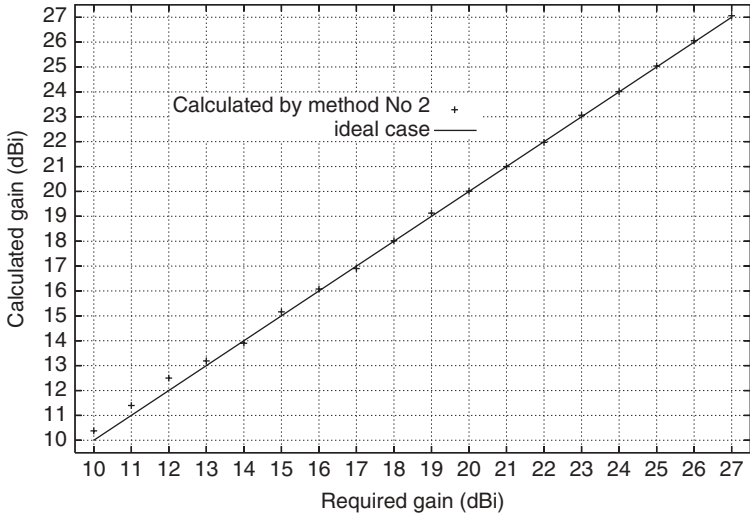


Figure 3.18 Accuracy of the simple design procedure for pyramidal horns.

The above method is generally accurate and consequently valuable in giving a useful design starting point. One of its limitations, however, is that it assumes that the aperture of the horn lies in an infinite ground plane. Ideally, it would be desirable to include the edge diffraction effects at the aperture in the calculations when finalizing the design or, alternatively, have the horn fully tested to determine the gain accurately.

Two examples of pyramidal horns designed using the above procedure are given in Figures 3.16 and 3.17. A plot of the accuracy of this simple design procedure compared to Method 2 in calculating the gain (at 11.70 GHz with a WR75 input waveguide) over the range of 10–27 dBi range is shown in Figure 3.18. Note that due to their nature, pyramidal horns do not have a well defined circular-symmetrical pattern and have significantly different E- and H-plane beamwidths. This affects the stability of the

phase center of the horn and its position at each frequency is a compromise between the phase fronts seen in the E- and H-planes. For that reason, pyramidal horns are rarely used as feeds for high performance reflector antennas.

3.4.3 Conical Horns

The geometry of a conical horn antenna is shown in Figure 3.2. To design a conical horn for a given gain (G_{dBi}) at a given design frequency [given wavelength λ (m)], use the following procedure [1]:

$$a_i = \frac{3\lambda}{2\pi} \tag{3.23}$$

$$a_o = \frac{\lambda}{2\pi}\sqrt{10^{(G_{\text{dBi}}+2.91)/10}} \tag{3.24}$$

$$R = \frac{4a_o^2}{3\lambda} \tag{3.25}$$

$$\theta = \arcsin\left(\frac{a_o}{R}\right) \tag{3.26}$$

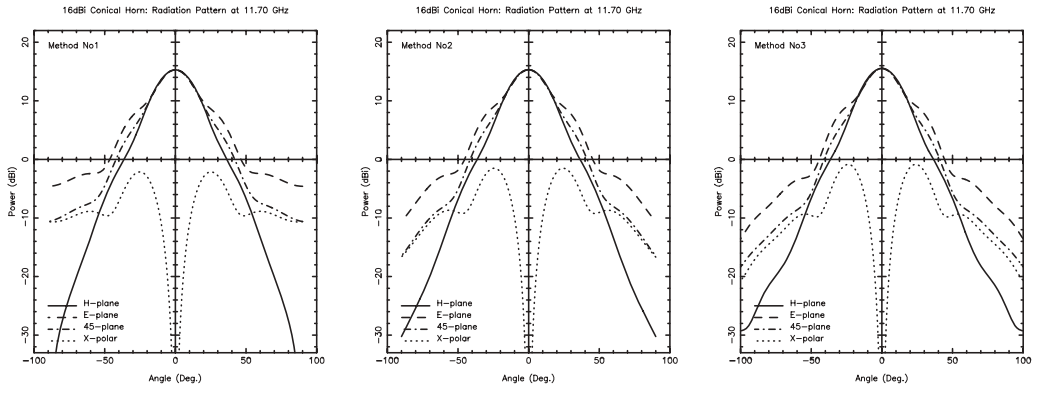
$$L = \frac{a_o - a_i}{\tan\theta} \tag{3.27}$$

The above method is reasonably accurate for horns with gains above 15 dBi and therefore valuable in giving a useful design starting point. As for the pyramidal horn, one of its limitations is in assuming that the aperture of the horn lies in an infinite ground plane. Again, it would be desirable to include the edge diffraction effects at the aperture in the calculations when finalizing the design or, alternatively, have the horn fully tested to determine the gain accurately.

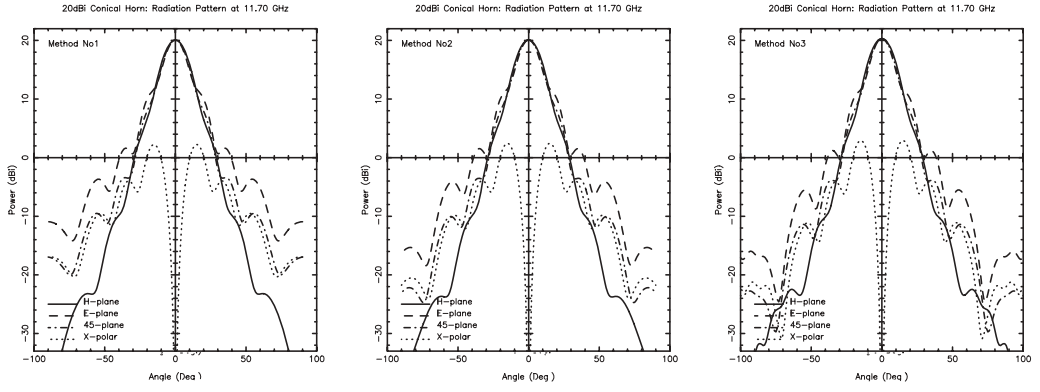
Two examples of horns designed using the above procedure are given in Figures 3.19 and 3.20. A plot of the accuracy of this simple design procedure compared to Method 2 in calculating gain over the range of 15–25 dBi is shown in Figure 3.21. One thing to keep in mind when deciding on using a conical horn is that there is a significant difference in the beamwidths of the principal planes, resulting in somewhat poor cross-polarization isolation. To demonstrate these properties, we have analyzed the nominal 16-dBi and 20-dBi horns over the 9.5–15-GHz frequency band (i.e., a relative frequency band of 0.812–1.282 if 11.70 GHz is taken as the design frequency) and summarized the results in terms of return loss, cross-polarization isolation, and gain in Figure 3.22 while the performance in terms of beamwidths is shown in Figure 3.23. This will have a direct negative repercussion on the stability of the phase center of the horn. So, if you require your horn to maintain a good pattern symmetry over your frequency band of interest, a conical horn might not be adequate.

3.4.4 Dual-Mode Horns

The classical dual-mode horn is commonly known as a "Potter horn," named after P. D. Potter, who introduced these horns as early as 1963 [41]. The simplified geometry of a classical Potter horn is shown in Figure 3.4a and the geometries and radiation patterns of the three examples presented by Potter in the original paper [41] are shown in Figure 3.24.



(a)



(b)

Specification	a_i (mm)	a_o (mm)	R (mm)	L (mm)	θ (deg.)	G_{dBi} (calculated)
16 dBi gain at 11.70 GHz	12.23	35.97	67.33	37.56	32.29	Method 1:15.28 Method 2:15.31 Method 3:15.53
20 dBi gain at 11.70 GHz	12.23	57.01	169.13	125.06	19.70	Method 1:20.05 Method 2:20.08 Method 3:20.31

(c)

Figure 3.19 Conical horn examples.

Over the years, improvements have been made on the design of dual-mode horns and one especially has proved to be very effective [42]. The geometry of the dual-mode horn antenna we consider here is shown in Figure 3.4b and is based on Ref. 42. To design a dual-mode horn for a given gain (15 dBi $\leq G_{\text{dBi}} \leq$ 23 dBi) at a given frequency [given wavelength λ (m)], use the following procedure that we have adapted from Ref. 42:

$$a_i = 0.51\lambda \tag{3.28}$$

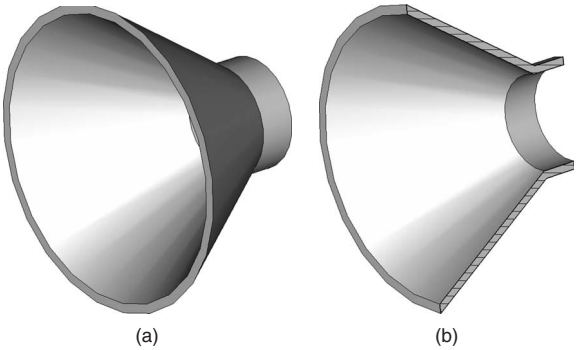


Figure 3.20 Geometry of the nominal 16-dBi conical horn: (a) Full view and (b) longitudinal section.

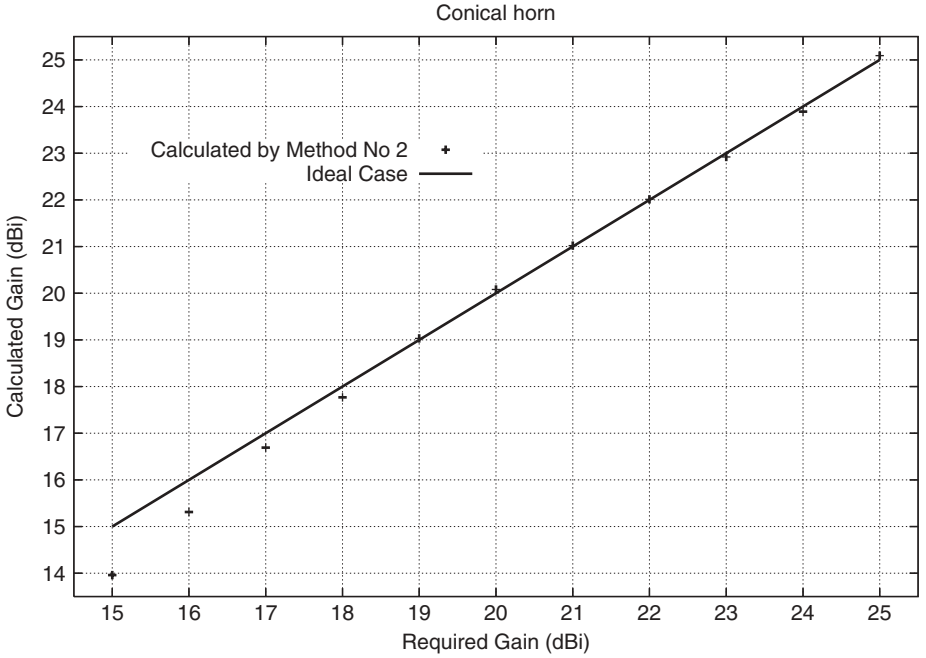


Figure 3.21 Accuracy of the simple design procedure for conical horns.

$$a_s = 0.65\lambda \tag{3.29}$$

$$a_o = (-6.074 + 1.174\ G_{\text{dBi}} - 0.07079\ G_{\text{dBi}}^2 + 0.001604\ G_{\text{dBi}}^3)\lambda \tag{3.30}$$

$$L = (3.5a_o/\lambda - 0.5)\lambda \tag{3.31}$$

The above procedure is accurate for horns with gains in the range of 15–23 dBi, but again, it assumes that the aperture of the horn lies in an infinite ground plane, so as before, it would be desirable to include the edge diffraction effects at the aperture in

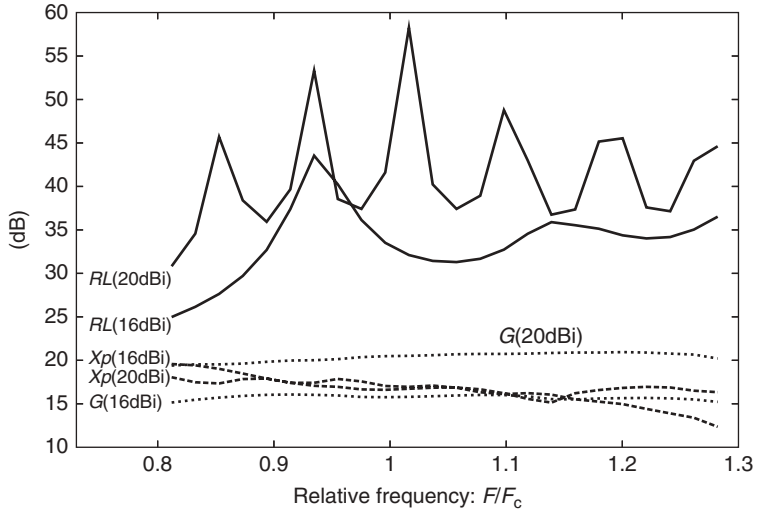


Figure 3.22 Performance of the nominal 16-dBi and 20-dBi conical horns with frequency (R_L, return loss; X_p, cross-polarization isolation; G, gain in dBi).

the calculations when finalizing the design or, alternatively, have the horn fully tested to determine the gain accurately.

Two examples of dual-mode horns designed using the above procedure are given in Figures 3.25 and 3.26. The performance with frequency of the nominal 20-dBi horn is shown in Figure 3.27 in terms of return loss, gain, and cross-polarization isolation, while the effects on the beamwidth are shown in Figure 3.28. A plot of the accuracy of this simple design procedure compared to Method 2 in calculating the gain over the range of 15–23 dBi is shown in Figure 3.29.

As a further validation of the above procedure, we consider the classical Potter horn No. 3, with a gain of 21.1 dBi. This horn has an overall length of 12.32 λ while a dual-mode horn (Figure 3.4b) designed using the above procedure has an overall length of 7.37 λ, being a reduction in length of about 40%. A comparison of both radiation patterns can be found in Figure 3.30 where it can be seen that both horns have good radiation pattern characteristics.

3.4.5 Conical Corrugated Horns

The design options for a corrugated horn are considerably more complicated than for its smooth-walled counterpart. For example, the match from a smooth-walled exciting waveguide to the corrugated surface is not a trivial matter and a number of options are available depending on the application. Space does not allow us here to go into sufficient detailed design to cover adequately all the aspects of corrugated horn design. However, one can find in Ref. 43 a companion paper to this chapter where basic data to design a variety of corrugated horns are presented.

The geometry of a basic conical corrugated horn antenna is shown in Figure 3.3. To design a conical corrugated horn for a given gain (13 dBi $\leq G_{\text{dBi}} \leq$ 22 dBi) at a given

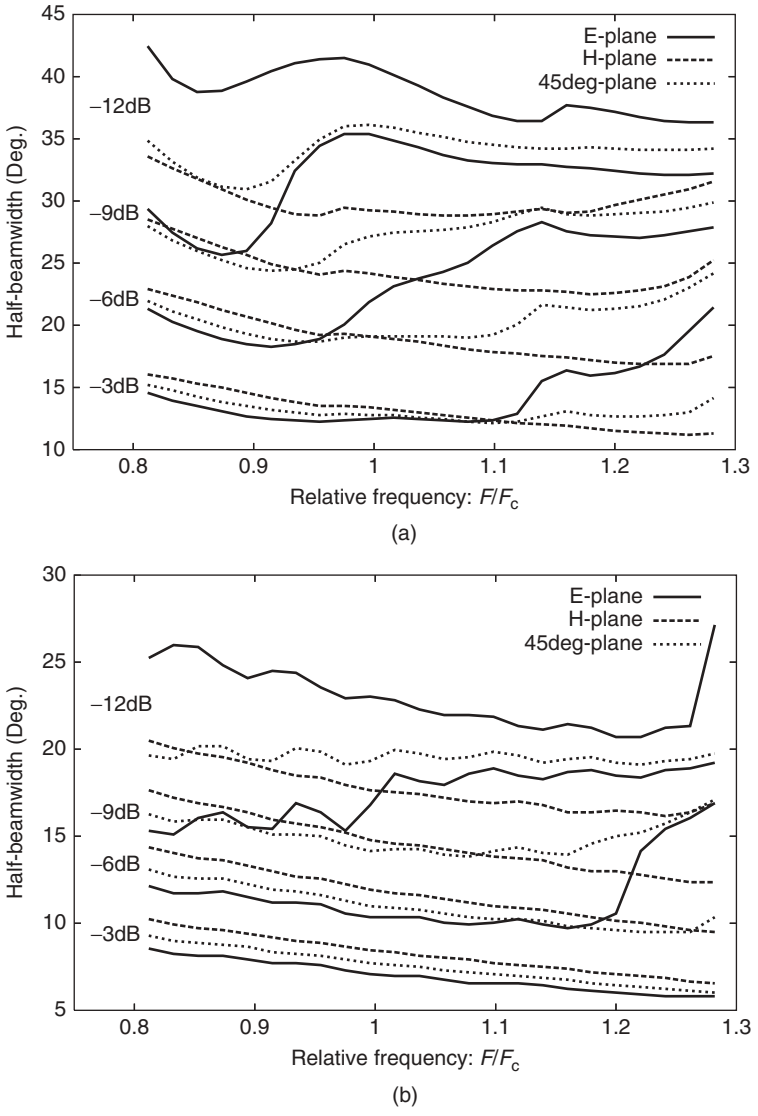


Figure 3.23 Half-beamwidth performance of the nominal 16-dBi and 20-dBi conical horns with frequency: (a) 16-dBi horn and (b) 20-dBi horn.

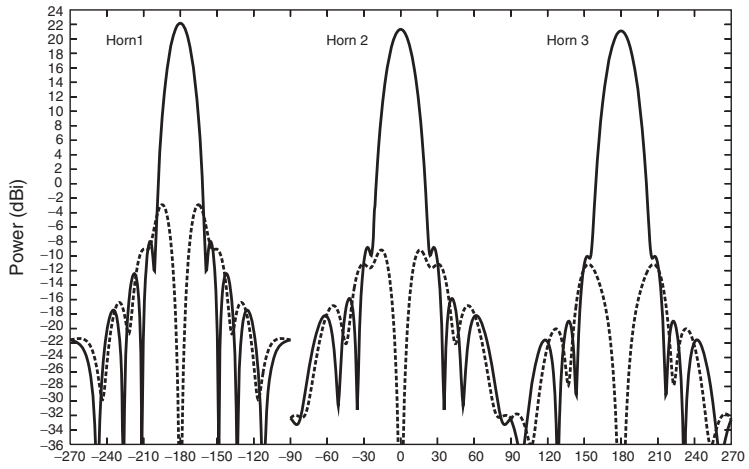
frequency (given wavelength λ (m)), use the following procedure:

$$a_i = 3\lambda/2\pi \tag{3.32}$$

$$\alpha = 192.351 - 17.7364G_{\text{dBi}} + 0.61362G_{\text{dBi}}^2 - 0.007712G_{\text{dBi}}^3 \tag{3.33}$$

$$a_o = (8.72704 - 0.740515\alpha + 0.0295435\alpha^2 - 0.00055165\alpha^3 + 0.00000387765\alpha^4)\lambda \tag{3.34}$$

$$p = \lambda/8 \tag{3.35}$$



(a)

Parameters	Horn 1	Horn 2	Horn 3
a_i (λ)	0.508	0.508	0.508
a_s (λ)	0.651	0.651	0.651
L_1 (λ)	0.203	0.325	0.439
L_2 (λ)	15.367	13.623	11.879
a_o (λ)	2.334	2.143	1.952
Gain (dBi)	22.2	21.3	21.1

(b)

Figure 3.24 (a) Radiation patterns in the 45°-plane (offseted by −180° and 180° for horns 1, 2, and 3) and (b) simplified geometries of the three original Potter horns 41.

$$w = 0.8p \tag{3.36}$$

$$N_{\text{Slots}} = \text{Nearest integer of } (4a_o/p) \tag{3.37}$$

$$L = N_{\text{Slots}}\, p \tag{3.38}$$

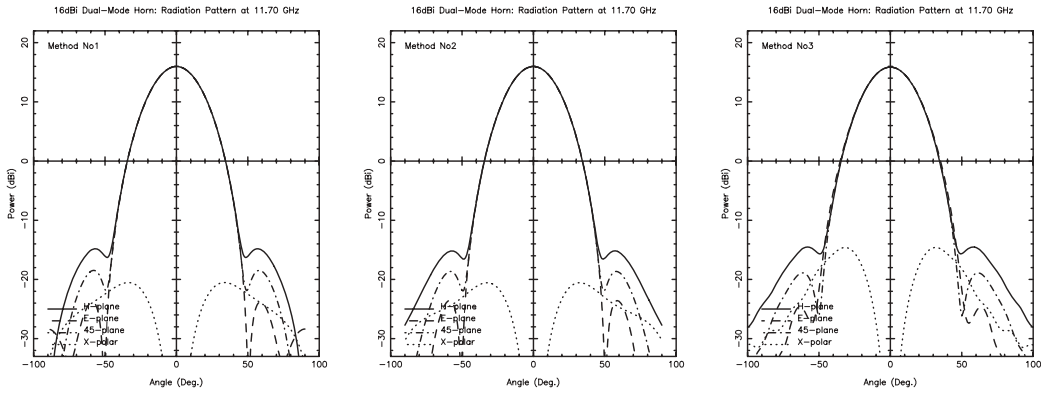
$$N_{MC} = 5 \tag{3.39}$$

$$a_j = a_i + (j-1)\frac{(a_o - a_i)}{(N_{\text{slots}} - 1)} \quad \text{for } 1 \le j \le N_{\text{Slots}} \tag{3.40}$$

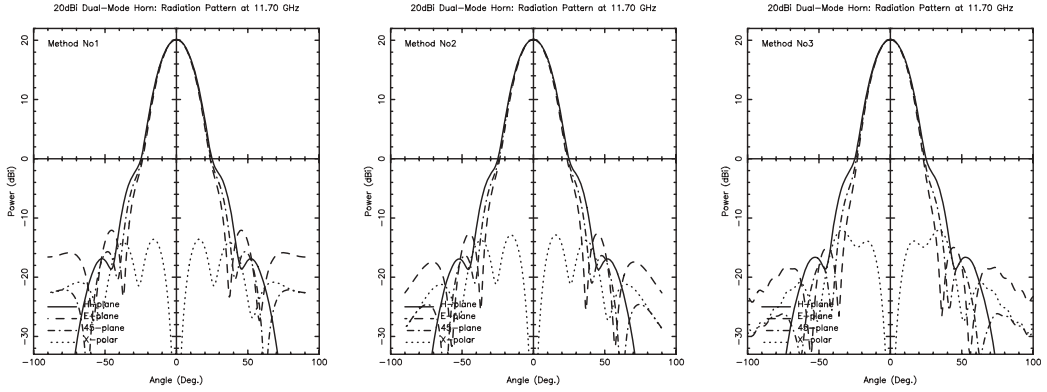
$$d_j = \left\{0.42 - \frac{j-1}{N_{MC}}\left(0.42 - \frac{1}{4}\exp\left[\frac{1}{2.114(2\pi a_j/\lambda)^{1.134}}\right]\right)\right\}\lambda \quad \text{for } 1 \le j \le N_{MC} \tag{3.41}$$

$$d_j = \frac{\lambda}{4}\exp\left[\frac{1}{2.114(2\pi a_j/\lambda)^{1.134}}\right] \quad \text{for } N_{MC}+1 \le j \le N_{\text{slots}} \tag{3.42}$$

Two examples of horns designed using the above procedure are given in Figures 3.31 and 3.32. The performance with frequency of the nominal 20-dBi horn is shown in Figure 3.33 in terms of return loss, gain, and cross-polarization isolation, while the



(a)



(b)

Specification	a_i (mm)	a_s (mm)	a_o (mm)	L (mm)	G_{dBi} (calculated)
16 dBi gain at 11.70 GHz	13.07	16.66	29.67	91.02	Method 1:15.98 Method 2:15.98 Method 3:15.87
20 dBi gain at 11.70 GHz	13.07	16.66	49.25	159.56	Method 1:20.13 Method 2:20.15 Method 3:20.14

(c)

Figure 3.25 Dual-mode horn examples.

effects on the beamwidth are shown in Figure 3.34. A plot of the accuracy of this simple design procedure compared to Method 2 in calculating gain over the range of 13–22 dBi is shown in Figure 3.35.

3.4.6 Low-Gain Axially Corrugated Horns

The geometry of an axially corrugated conical horn antenna is shown in Figure 3.7. To design this horn type for a given gain (10.5 dBi $\leq G_{\text{dBi}} \leq$ 14.5 dBi) at a given frequency

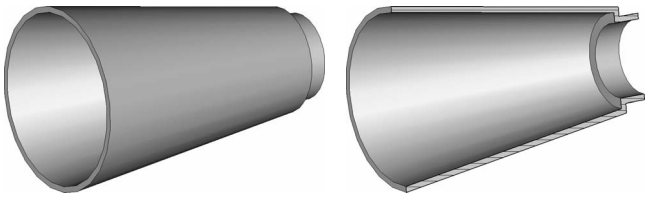


Figure 3.26 Geometry of the nominal 16-dBi dual-mode horn: (a) fullview and (b) longitudinal section.

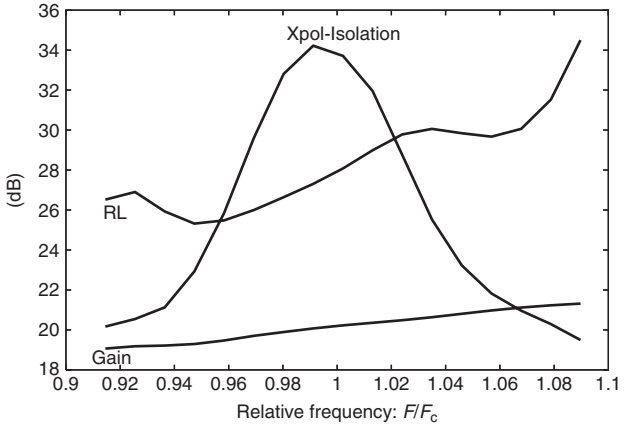


Figure 3.27 Performance with frequency of the nominal 20-dBi dual-mode horn.

(given wavelength λ (m)), use, as a starting point, the following simple procedure:

$$a_i = 3\lambda/2\pi \tag{3.43}$$

$$N_{\text{slots}} = \text{Nearest integer of } (-343.325 + 84.7229\ G_{\text{dBi}} - 6.99153\ G_{\text{dBi}}^2 + 0.194452\ G_{\text{dBi}}^3) \tag{3.44}$$

$$\theta = 45^\circ \tag{3.45}$$

$$p = \lambda/8 \tag{3.46}$$

$$w = 0.8p \tag{3.47}$$

$$L = N_{\text{slots}}\, p \tag{3.48}$$

$$a_j = a_i + jp \quad \text{for } 1 \le j \le N_{\text{slots}} \tag{3.49}$$

$$d_j = \frac{\lambda}{4} \exp\left[\frac{1}{2.114(2\pi a_j/\lambda)^{1.134}}\right] \quad \text{for } 1 \le j \le N_{\text{Slots}} \tag{3.50}$$

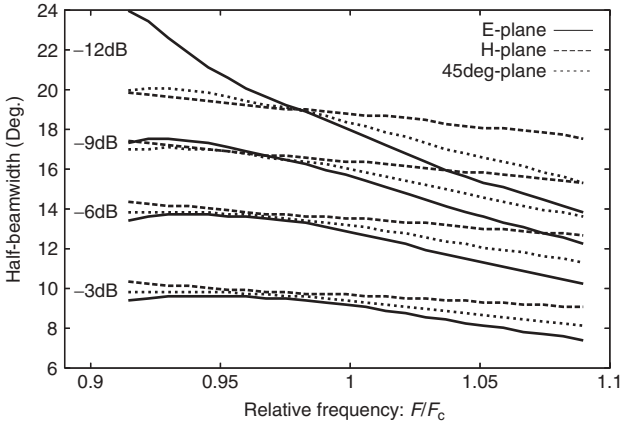


Figure 3.28 Beamwidth performance with frequency of the nominal 20-dBi dual-mode horn.

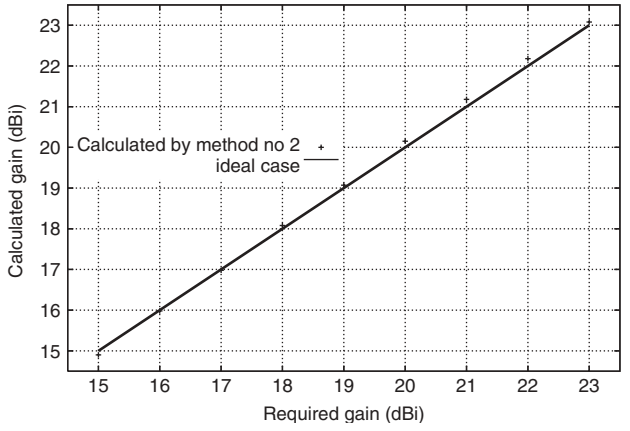


Figure 3.29 Accuracy of the simple design procedure for dual-mode horns.

Two examples of horns designed using the above procedure are given in Figure 3.36 and Figure 3.37. The performance with frequency of the nominal 14-dBi horn is shown in Figure 3.38 in terms of return loss, gain, and cross-polarization isolation. A plot of the accuracy of this simple design procedure compared to Method 2 in calculating the gain over the gain range of 10.5–14.5 dBi is shown in Figure 3.39. (Note that our procedure gives a "staircase-approximation" of the desired gain, with each step corresponding to the addition of another corrugation.)

3.4.7 Choke-Ring Waveguides

A choke-ring waveguide or choke horn is basically a circular waveguide with a choke at its aperture. Most horns of this type are used as feeds for prime-focus paraboloid reflectors with low f/D ratios (<0.5). The choke is designed to improve the radiation pattern of the waveguide. In many applications, only one choke is necessary, but

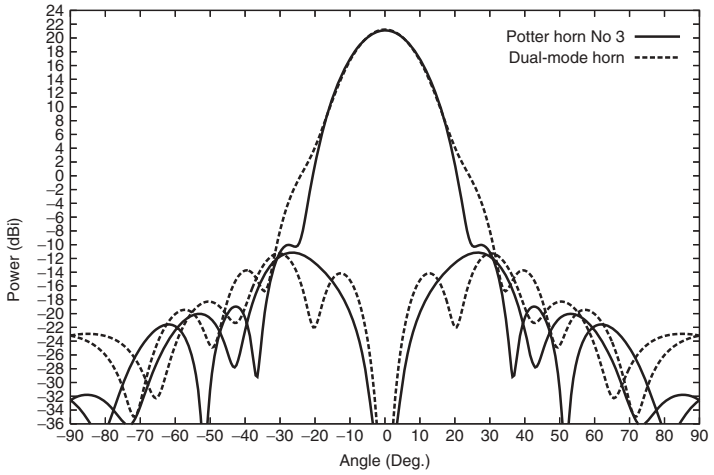


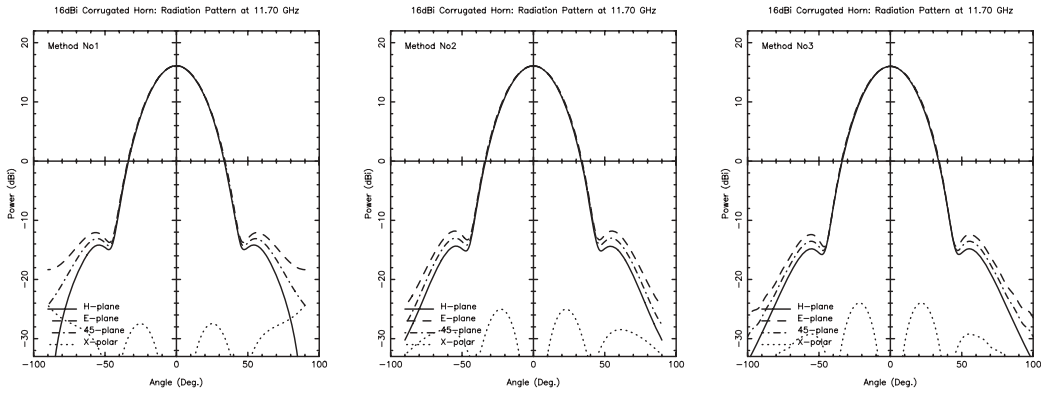
Figure 3.30 Radiation pattern comparison in the 45°-plane between the classical Potter horn No. 3 and a dual-band horn designed for a nominal 21.1-dBi gain.

depending on the application, it might be necessary to have several chokes. We will only consider here the simple example of a circular waveguide with a single choke (Figure 3.7b) and give two examples. The standard design for a choke is to make the depth of the choke a quarter of the free-space wavelength at the design frequency (see Choke waveguide No. 1, Figure 3.40a). This type of horn tends to give acceptable radiation performance over a 20–25% bandwidth. The second example (see Choke waveguide No. 2, Figure 3.40a) is a horn that we have optimized to provide effective illumination of a paraboloid reflector with an f/D of 0.345 over a 50% bandwidth (the standard bandwidth of commercial circular-to-rectangular waveguide transitions). The radiation patterns at the design frequency and the performance with frequency in terms of gain, return loss, and cross-polarization isolation for both horns are shown in Figures 3.40b and 3.41 while the half-beamwidths at −3, −6, −9, −12, and −15 dB of both horns are shown in Figure 3.42.

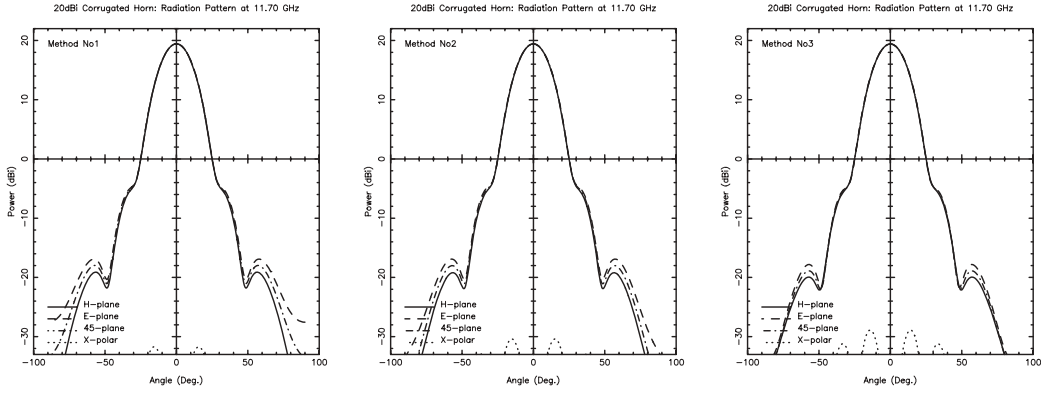
3.5 FABRICATION

Many texts on antennas and RF engineering quite rightly mention the issues of mechanical design and fabrication of RF equipment. However, many of these treatments are intended to broadly outline the principles of mechanical design and fabrication for RF designers, with little emphasis on providing guidance for fabrication shop staff in the specific techniques used in RF hardware fabrication. Not all organizations involved in the design of RF equipment have in-house fabrication facilities available, and most resort to out-sourcing manufacture to general fabrication shops, where the finer points of RF hardware manufacture may not be familiar.

In this chapter we have directed much of the discussion of fabrication toward practical issues of use to the fabricators, in the hope that, working together with RF design staff, the fabrication shop may be able to make worthwhile contributions to the overall process of creating effective RF equipment.



(a)



(b)

Common parameters: $a_i = 23.12$ mm, $p = 2.3$ mm

Specification	a_o (mm)	N_{Slots}	L (mm)	G_{dBi} (calculated)
16 dBi gain at 11.70 GHz	30.73	38	121.71	Method 1:16.11 Method 2:16.09 Method 3:16.01
20 dBi gain at 11.70 GHz	46.60	58	185.77	Method 1:19.43 Method 2:19.43 Method 3:19.41

(c)

Figure 3.31 Conical corrugated horn examples.

3.5.1 Materials

Materials for fabrication of aperture antennas are typically metals, with occasional use of dielectrics. Metals used are typically alloys of aluminum or copper, while dielectrics may range from rigid lightweight foams and plastics to hard ceramics. Fabrication data for specialized materials is best sought from the manufacturer, but some general properties of common materials that find regular use in aperture antennas are given next.

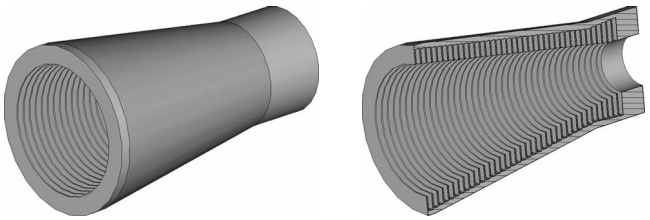


Figure 3.32 Geometry of the nominal 16-dBi corrugated horn: (a) full view and (b) longitudinal section.

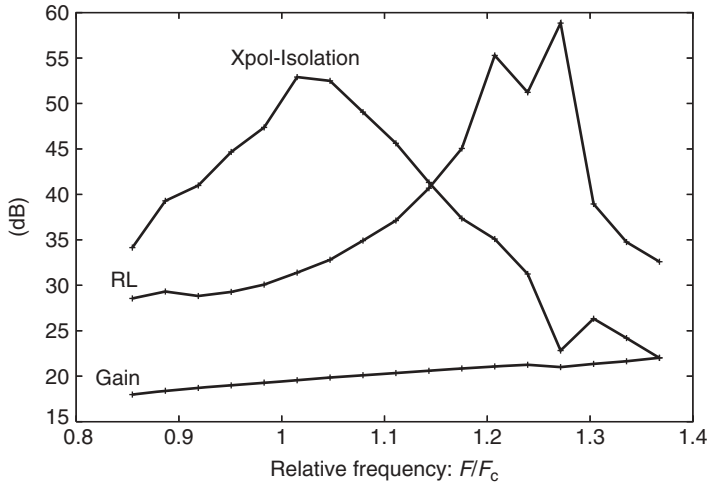


Figure 3.33 Performance with frequency of the nominal 20-dBi conical corrugated horn.

3.5.1.1 Aluminum Alloys Aluminum alloys can broadly be divided into two categories: wrought alloys supplied in the form of sheets, plates, bars, and extruded shapes, and casting alloys, intended for melting and pouring into molds to produce castings. Some alloys can be strengthened and hardened by heat-treatment processes, while others can have their physical properties controlled by mechanical work-hardening processes such as rolling or forging. Heat-treatable wrought alloys are typically supplied and used in the heat-treated condition, while heat-treatable casting alloys must be heat-treated after casting.

More detailed information may be found in Refs. 44 and 45.

Wrought Alloys

- *5005-H34.* A low strength nonheat-treatable aluminum–magnesium alloy with excellent corrosion resistance, supplied in the form of sheets and thin plates. Used

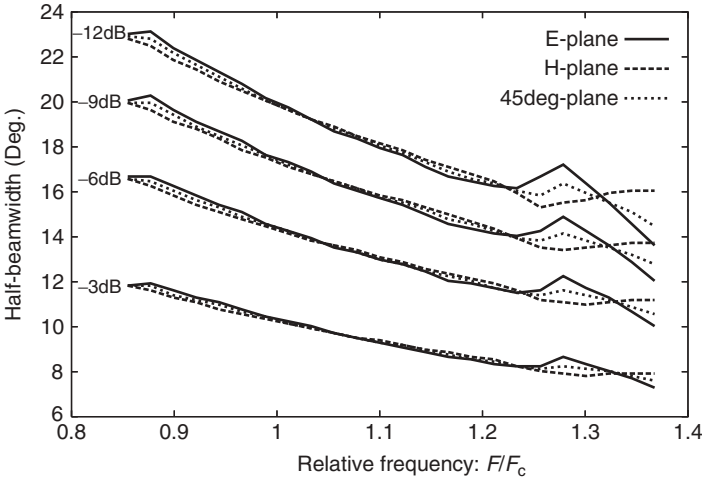


Figure 3.34 Beamwidth performance with frequency of the nominal 20-dBi conical corrugated horn.

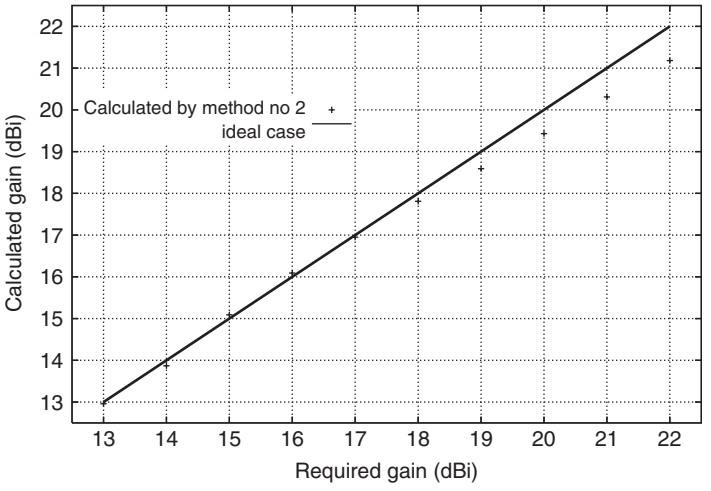
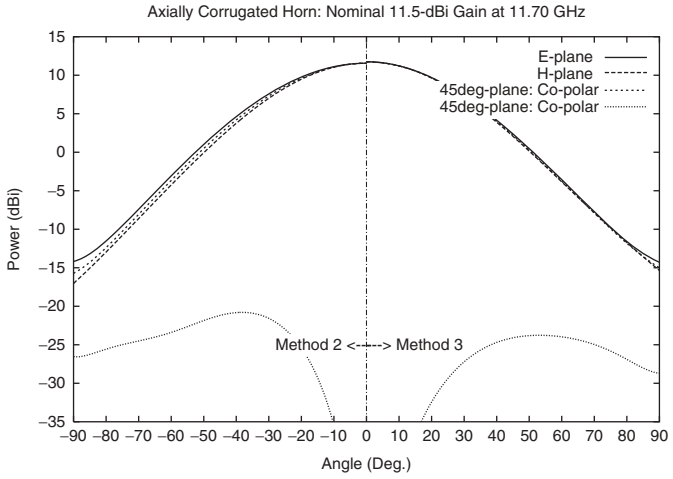


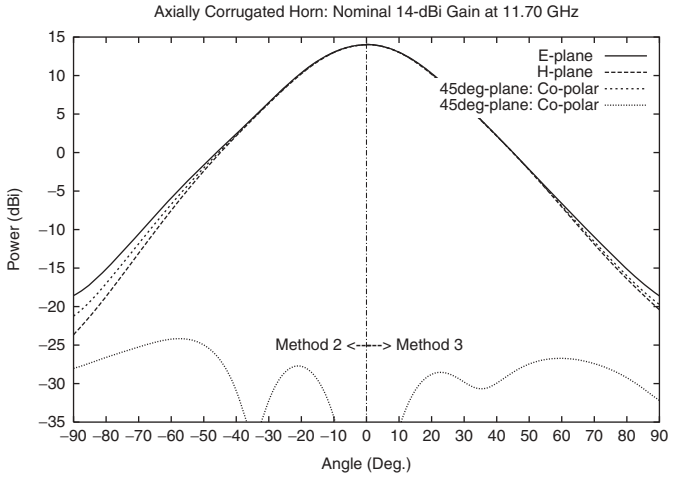
Figure 3.35 Accuracy of the simple design procedure for conical corrugated horns.

for fabrication of sheet metal or light plate structures. Readily formed, welded, and brazed.

- *5083-H112, 5083-H116, 5083-H32, 5083-H34.* A moderate strength non-heat-treatable aluminum-magnesium alloy, supplied in heavy sheets, plates, and bars, for machined components. Readily machined and welded. Not as corrosion resistant as alloy 6061.
- *6061-T6, 6061-T651.* A moderate strength heat-treatable aluminum–magnesium–silicon alloy supplied in the form of plates and bars for machined parts. Very readily machined, welded, and brazed, with very good corrosion resistance. Welding and



(a)



(b)

Common parameters: $a_i = 12.23$ mm, $p = 3.2$ mm

Specification	a_o (mm)	N_{Slots}	L (mm)	G_{dBi} (calculated)
11.5 dBi gain at 11.70 GHz	18.63	2	18.63	Method 2:11.56 Method 3:11.74
14 dBi gain at 11.70 GHz	31.43	6	31.43	Method 2:13.99 Method 3:14.03

(c)

Figure 3.36 Axially corrugated conical horn examples.

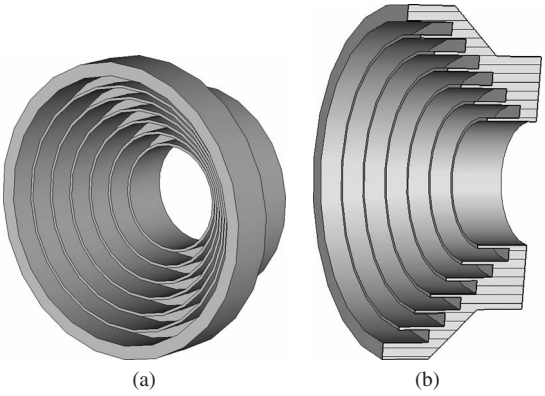


Figure 3.37 Geometry of the nominal 14-dBi axially corrugated horn: (a) full view and (b) longitudinal section.

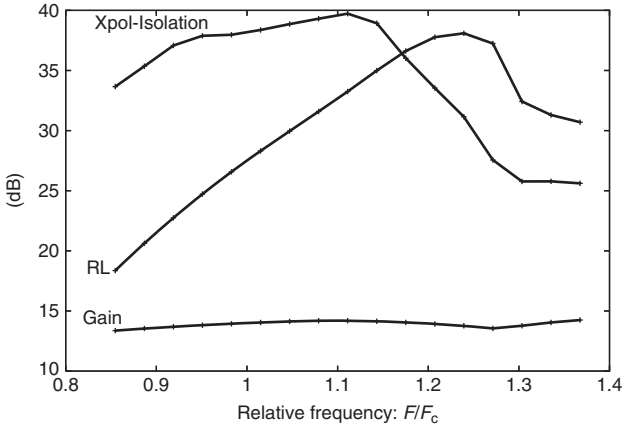


Figure 3.38 Performance with frequency of the axially corrugated 14-dBi horn.

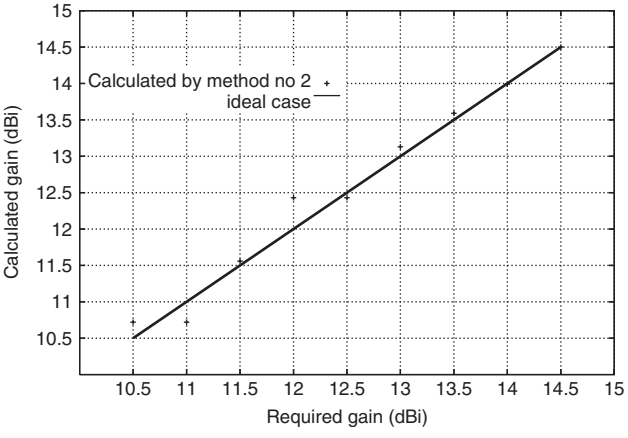
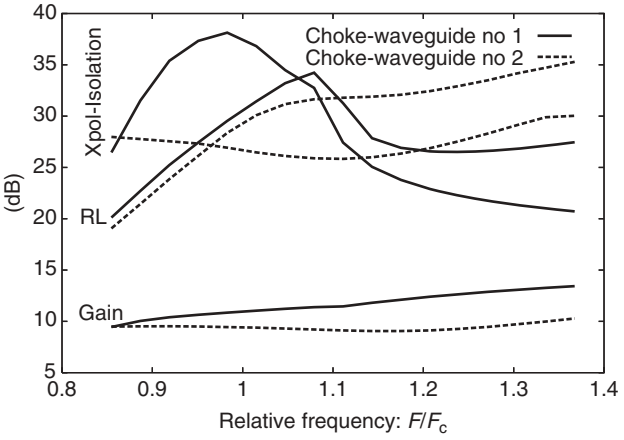


Figure 3.39 Accuracy of the simple design procedure for axially corrugated horns.

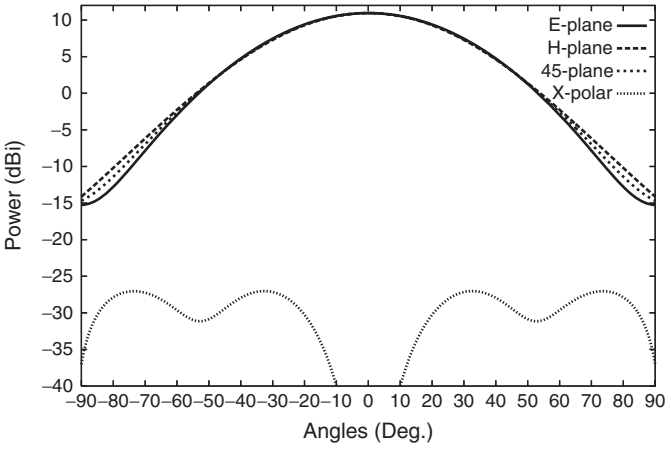
Parameters	Choke waveguide No. 1	Choke waveguide No. 2
$a_i(\lambda)$	0.5500	0.4293
$d(\lambda)$	0.2500	0.4488
$b(\lambda)$	0.5750	0.4722
$c(\lambda)$	0.6750	0.8079
$f(\lambda)$	0.7375	0.8859

(a)



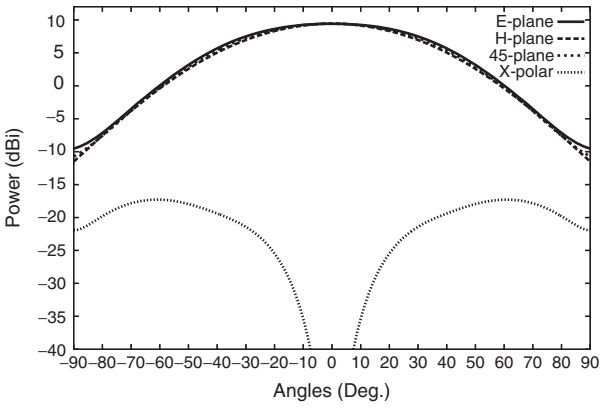
(b)

Figure 3.40 Simple choke waveguide with a single choke: (a) two simple examples and (b) performance of both horns with frequency.



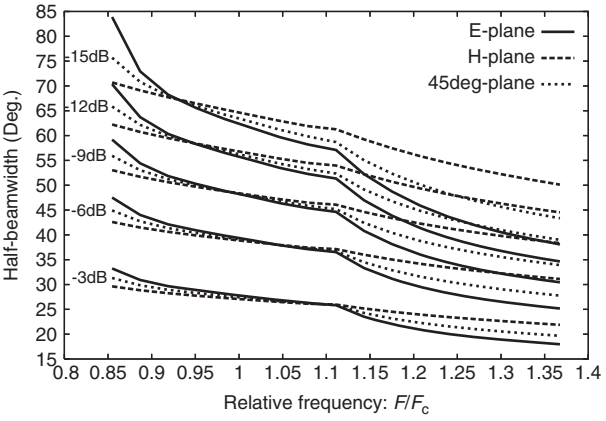
(a)

Figure 3.41 Radiation pattern at the design frequency: (a) Choke waveguide No. 1 and (b) Choke waveguide No. 2.

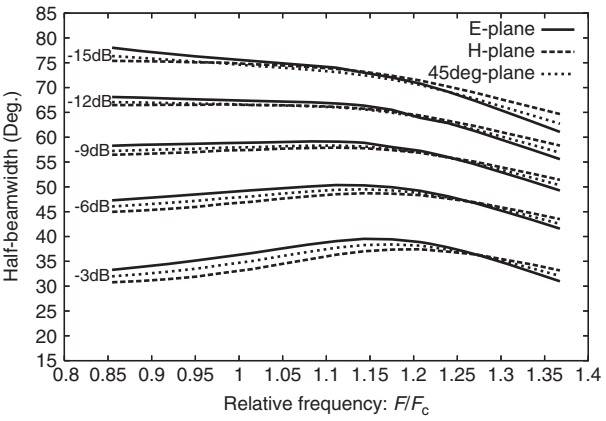


(b)

Figure 3.41 (*Continued*).



(a)



(b)

Figure 3.42 Half-beamwidth performance with frequency: (a) Choke waveguide No. 1 and (b) Choke waveguide No. 2.

brazing affect the heat-treatment state of the material, after which the machinability deteriorates. Heavily welded or brazed fabrications should be re-heat-treated back to the T6 temper before final machining.

- *7075-T6, 7075-T651.* A very high strength heat-treatable aluminum–magnesium–zinc alloy supplied in plates and bars. Machinable but not readily weldable. Used for parts carrying heavy structural loads.

Casting alloys

- *Alloys 356 and A356.* An excellent general-purpose heat-treatable aluminum–silicon alloy for sand castings. Castings are readily weldable and have excellent corrosion resistance and good machinability, though like all high silicon–aluminum casting alloys, it can be abrasive to cutting tools. Carbide tools are necessary to prevent excessive tool wear.
- *Alloys 360 and A360.* An aluminum–silicon die casting alloy for intricate parts that can be released easily from the casting dies. Good corrosion resistance and reasonably machinable with carbide tools.
- *Alloys 413 and A413.* An aluminum–silicon die casting alloy for large intricate parts. Very good corrosion resistance and reasonably machinable with carbide tools.

3.5.1.2 Copper Alloys Copper alloys for use in aperture antennas are usually wrought materials in the form of sheets, plates, and bars. Cast copper alloys are sometimes seen elsewhere in antenna systems—in waveguide components, for example. Some commonly used copper alloys are listed below.

- *UNS C11000 Electrolytic Tough Pitch Copper.* General-purpose engineering copper, containing a small amount of oxygen as an impurity. Available in sheets, plates, and bars. Very readily formed, soldered, and brazed. Welding can cause porosity. Can be machined, though problems with finish and chip control may be encountered. Special cutting lubricants suitable for copper may be required.
- *UNS C10100 Oxygen-Free Electronic Copper, UNS C10200 Oxygen-Free High Conductivity Copper.* These materials are essentially pure copper, with improved electrical conductivity compared to UNS C11000. Very readily formed, soldered, brazed and welded. Machining as per UNS C11000.
- *UNS C14500 Tellurium Bearing Copper, UNS C14700 Sulfur Bearing Copper.* These copper alloys have additions to improve machinability, though at some expense to electrical properties. Much more readily machined than UNS C11000 and UNS C10100/10200.

Other dilute copper alloys containing beryllium are commonly used for some microwave components such as waveguide castings, for which they are well suited. Their use in low volume fabrication should be approached with caution, however, due to the adverse health risks associated with beryllium.

3.5.1.3 Brass Alloys

- *UNS C36000 Free Cutting Brass, UNS C38500 Architectural Bronze.* Highly machinable brass alloys available in sheets, bars, and extrusions. Readily formed and soldered. Good corrosion resistance.

3.5.2 Coatings and Surface Treatments

While aperture antenna components can be used with the materials left in their as-manufactured state, often some form of coating or surface treatment can be used to enhance some property or aspect of performance. There are a few different types of surface treatments that are applicable to aperture antenna components.

- *Plating.* A coating of one or more layers of materials of different composition from the parent metal are deposited on the surface of the part by electrolytic or chemical means. Common applications are the coatings of precious metals such as gold and silver to improve surface conductivity and/or minimize the potential for corrosion. Plating processes vary widely for different parent materials, so advice should be sought from the plating service provider before deciding on an application.
- *Chemical Conversion Coatings.* A surface is converted by chemical means to a compound with different characteristics. Chromate conversion coatings [46] are highly corrosion resistant, electrically conductive, and of pleasing appearance. They are widely used for aluminum and zinc alloy components in antenna engineering. Chemical cleaning processes for copper and copper alloy parts often involve chromate stages to enhance corrosion protection and maintain a bright, clean appearance.
- *Passivation.* An existing surface or surface layer is chemically cleaned or decontaminated to make it more inert. The process is commonly applied to stainless steel components such as fasteners, to prevent rust-staining that can occur from contamination of the stainless surface by steel from the production machinery [47].

Paint is often applied as an external finish to enhance the appearance and improve the corrosion resistance of finished parts and assemblies, but not in electrically active areas. The types of paint available are innumerable, but one group that finds regular use in antenna fabrication are the infrared-reflective coatings. These paints strongly reflect infrared radiation and present a matt appearance to minimize solar heating and specular reflection of solar radiation, particularly on reflector antennas and their feeds.

3.5.3 Corrosion

Most metals are subject to corrosion in certain circumstances and this should be considered in the design and fabrication of an aperture antenna that is to be used for any length of time or in an environment exposed to weather or other adverse conditions.

While materials such as copper and aluminum alloys are quite corrosion resistant by themselves, and may be made even more so by the surface treatment processes mentioned in Section 3.5.2, certain combinations of dissimilar materials can cause serious corrosion problems by galvanic action.

When designing assemblies using combinations of dissimilar materials, reference should be made to an electropotential chart for the materials and environment concerned [48]. Arrangements using materials widely separated on an electropotential chart should be avoided, although where that is not practical, such as the use of stainless steel fasteners to secure assemblies made from aluminum alloys, the surface area of the more cathodic material in the assembly, in this example the stainless steel fasteners, should be kept to a minimum. Postassembly treatment such as paint may be applied to help protect the assembly from environmental conditions that may promote corrosion.

It is sometimes possible to plate a highly anodic material with a more cathodic one to reduce the potential for corrosion. Plating aluminum parts with silver, for example, makes the assembly fully compatible with stainless steel fasteners in all but the most extreme environments.

3.5.4 Machining Operations

Much has been written describing machining processes for the benefit of microwave engineers, but there is not much information about the important mechanical aspects of microwave structures available to a machinist not experienced in that field. A number of relatively simple but important aspects of microwave component fabrication that are sometimes overlooked in interactions between antenna and RF designers and fabrication shops are worth discussing.

3.5.4.1 Special Tools While many parts for aperture antennas can be machined using commercial turning and milling tools with carbide inserts, there are some instances where a specially made tool offers significant advantages. The machining of corrugated horns is an example. Commercial carbide grooving inserts work extremely well but often do not have square ends or sharp corners, and the clamping arrangements for these inserts do not always allow the tools to fit into confined areas at the input end of a horn.

In these cases, special purpose tools can be made from steel bar and brazed-on carbide inserts, with the cutting edges and angles ground after brazing on a tool-and-cutter grinder. These tools can be designed to ensure the largest possible boring bar can be fitted into the smallest possible space without fouling, which is of great assistance in achieving high metal removal rates with freedom from chatter. Examples of specially made boring bars and milling cutters are discussed in Section 3.5.8.

3.5.4.2 Filleted Internal Corners The machining of a pocket or cavity on a milling machine naturally produces a cavity with rounded corners with a radius equal to that of the milling cutter.

Such cavities should be designed in consultation with the machinist who will be manufacturing the parts to determine realistically achievable dimensions for such corner radii. Deep cavities will take longer to machine if small radii are specified, and long milling cutters are prone to chatter when they enter filleted corners.

CNC machines can reduce the tendency to chatter in internal corners by machining a radius larger than that of the cutter, which reduces the cutter's contact area while machining the fillet.

3.5.4.3 Swarf Control and Removal The chips generated in machining operations (swarf) can be a significant cause of problems in the machining of antenna components. Many wrought aluminum and copper alloys are known as "long chip" materials from their tendency to produce tenacious ribbons of swarf. These can become wrapped around cutting tools and around the job, damaging previously machined surfaces, interfering with coolant access, and occasionally causing tool breakage and scrapping of the part.

Many modern carbide tooling inserts are provided with "chip breaker" geometries to break long ribbons of swarf into short chips, and these should be used where suitable. Chip breaker inserts, however, are usually designed to function correctly at the high speeds and high feedrates used in CNC machine tools. At the lower speeds and moderate feedrates

necessary to machine intricate items, especially on manually operated machines, the chip breaking function of commercial inserts can be ineffective. In these cases special purpose cutting tools can be ground with more aggressive chip breakers, or with some trial and error can be ground to eject continuous ribbons of swarf safely from the machining area.

Turning of the internal features of horn antennas is an operation where particular attention must be paid to swarf control. Techniques such as the use of extreme flood coolant, interrupting the feed to break swarf, and regular withdrawals of the tool to clear the working area may be necessary, particularly on CNC machines where the operator has no "feel" or sight of the process.

Corrugated horns are especially adept at trapping swarf in the corrugations. Much care should be taken in the finishing and examination of corrugated horns to detect and remove trapped swarf.

3.5.4.4 Chamfering and Deburring It is a common practice in many engineering machine shops to lightly round or chamfer all sharp edges to minimize handling injuries. This can be detrimental to the performance of some microwave devices, and it must be made clear to all concerned in manufacture that edges not specifically noted to be chamfered must be deburred but otherwise left sharp. Manufacturing drawings should carry a clear general notation to this effect.

With this in mind, machining operations can sometimes be tailored to minimize generation of burrs. Methods such as climb-milling so the cutter teeth enter toward component edges rather than outward from edges, and taking finishing cuts twice, especially on CNC machines, are useful in reducing the need for manual removal of burrs.

Where manual deburring is unavoidable, it should be done with a soft tool that can break away burrs without removing parent material.

3.5.4.5 Surface Finish The finish on machined surfaces can be important to the functioning of an antenna component, especially at higher frequencies. A surface finish adequate to the task should be specified before manufacture, and this requirement should be agreed upon with the manufacturer. This does not mean, however, specifying the best possible finish on the part of the designer, or attempting to achieve the best possible finish on the part of the manufacturer. Where coarse finishes are adequate for the task, they should be used to simplify fabrication and reduce manufacturing costs.

3.5.4.6 Component Handling Due to their relative softness, components made from aluminum and copper alloys require a level of careful handling not always found in general engineering shops, to avoid damage by scratching and denting.

Clean areas should be set aside for storage of components during manufacture, and care should be taken with part handling both in and out of machine tools. Even such seemingly harmless practices as sliding a component across a bench top or accidentally knocking it against a chuck jaw during chucking can irreparably scratch or dent a waveguide contact face.

These precautions should be discussed with the manufacturing staff and any external contractors, such as electroplaters, before commencing work.

3.5.5 Welding, Brazing, and Soldering

Some aperture antenna components may require fabrication by welding, brazing, or soldering, either from flat or preformed sheet metal parts, from machined parts, or

combinations of both. A typical example might be a pyramidal horn made from flat sheets, joined along the edges and fitted with a waveguide flange at the input. In choosing and applying a joining process, a number of issues should be considered [49].

First, consider the suitability of process for the materials to be joined. Many aluminum alloys are readily welded or brazed with the proper processes, fluxes, and filler materials, while soldering is difficult. Copper alloys are very readily brazed and soldered, while welding can cause porosity problems.

Second, consider the effect of the process on the parts being joined. Welding requires intense localized heating, which can lead to distortion of the parts being joined through thermal expansion stresses. Brazing can be performed with either localized or generalized heating. Generalized heating can minimize thermal stresses during brazing, but can be difficult when parts or assemblies are very large. Any welding or brazing process will cause a change in the physical properties of a heat-treated or work-hardened material.

Third, consider the strength required in the assembly. Welding and brazing processes usually result in joints approaching the strength of the parent metal. Soldered joints, being carried out with low melting point filler metals, are typically much weaker than the parent metal.

3.5.6 Bolted Joints

Nearly all aperture antennas will have at least one bolted joint, even if it is only the attachment of the transmitting or receiving waveguide. Wherever possible, existing recognized standards for bolted joint configurations should be used. Waveguide flanges are a good example. If nonstandard joint configurations are necessary, such as between sections of a horn, a number of practices should be observed.

Any bolted connection in a microwave device must ensure adequate electrical contact at the internal edges to present an uninterrupted current path for the signal. While in many cases this can be achieved by simply ensuring the contact surfaces of both parts are flat and free from scratches and burrs, in more critical applications the methods shown in Figures 3.43 and 3.44 may be applied.

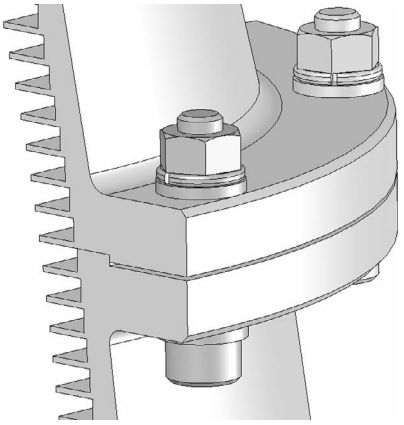


Figure 3.43 Flange joint between sections of a corrugated horn, showing gap at periphery to ensure contact at bore.

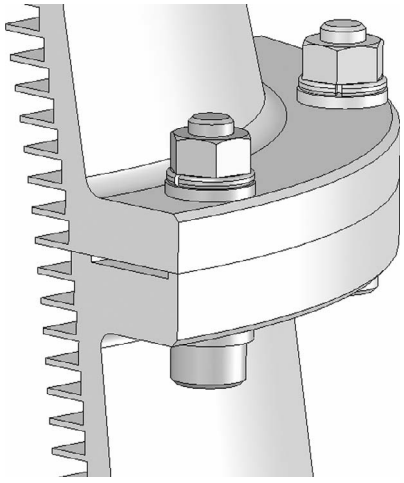


Figure 3.44 Contact flange joint between sections of a corrugated horn, showing relieved area in the center, and contact at both the periphery and the bore.

Figure 3.43 shows a "crush" flange joint for use between two sections of a horn. The contacting faces are at the inside wall, while the remainder of the joint has a small but definite gap. The joint is dimensioned so that the gap will always be present regardless of the variations in sizes permitted by manufacturing tolerances. A secondary feature of this arrangement is the spigot and socket fitting of the two parts, which ensures concentric alignment without the need for dowels.

A drawback of this design is the open edge into which water can penetrate, possibly leading to corrosion if the assembly is left unprotected. The gap is shown oversize for clarity in Figure 3.43. In reality, this gap should be approximately 0.1–0.2 mm.

Figure 3.44 shows a "contact" or "relieved" flange. This is a variation of a truly flat flange with part of the joint area removed, leaving two contact faces at the inner and outer edges with the bolts in between. In this arrangement the bolting load is shared by both faces so the contact pressure at the inner wall is reduced, even more so if the bolts are closer to the outer edge, and provision must be made for dowels to ensure alignment. Unlike the crush flange, both surfaces of a contact flange may be lapped against a flat reference for enhanced contact to minimize passive intermodulation (see Section 3.5.7), and the outer contact provides improved protection from corrosion.

The overall size and configuration of a bolted joint should be designed so that off-the-shelf parts and conventional assembly tools can be used and to provide a neat appearance. In Figures 3.43 and 3.44, the flange thicknesses have been chosen to ensure that when a standard length cap screw is fitted with the appropriate nuts and washers, approximately two threads of the screw protrude from the nut, and there is sufficient room around the nut to fit a standard spanner without fouling adjacent surfaces.

It will be noted that in these figures the assemblies have been designed with cap screws and nuts, rather than cap screws and tapped holes. This is a choice that may depend on the ultimate use of the product and the quantity to be made. Use of screws and nuts increases the component count in the assembly but it eliminates a tapping operation in one of the parts. Further more, in the event the assembly is used in a

corrosive environment, disassembly by cutting or breaking off corroded nuts and bolts will be found far easier than removing broken bolts from tapped holes.

If the assembly is to be a commercial product, it is worth applying aspects of one of the many standards available for mechanical assemblies. It may, for example, enhance a product's reliability if requirements for mechanical assemblies made from soft materials such as copper and aluminum alloys are applied. Among other things these define the correct use of flat washers, lock washers and nuts and, where parts require disassembly for maintenance, the use of thread reinforcements such as helical inserts in tapped holes.

Lastly, it is a good practice to keep a library of joint designs and configurations that have proved effective so they can be reapplied to new designs without the need for additional engineering tasks.

3.5.7 Special Conditions

3.5.7.1 Pressurization Antenna systems designed for use in outdoor environments, particularly transmit systems handling significant power, are often lightly pressurized with dehumidified air, dry nitrogen, or other gas to ensure water cannot gain entry. Sealing such systems to hold internal pressure requires an impermeable radome covering the aperture and seals at each joint to prevent leakage and excessive air or gas consumption.

Most standard waveguide flanges are available in sealed configurations using either standard industrial O-rings or specially molded silicone rubber seals, and these should be used on pressurized systems wherever waveguide attachments are made.

Other joints within a system will have to be designed to incorporate O-rings. In an environment where the seal is to protect the system against weather and to retain air or nitrogen, O-rings of silicone or nitrile rubber compounds may be used. These are inexpensive and readily available. O-ring suppliers can provide data for the proportions of the seal grooves to be used for low pressure static seals.

An arrangement for a joint design often used by the authors is shown in Figure 3.45. It is an intermediate joint in a machined horn, of the type shown earlier in Figure 3.43. In this instance a groove for an O-ring is machined in one of the parts, on the same diameter as the spigot locating the two sections.

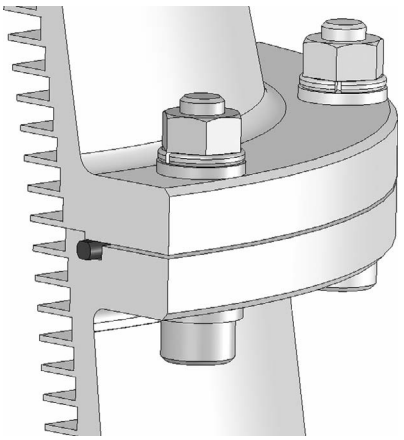


Figure 3.45 Flange joint between sections of a corrugated horn, showing O-ring for sealing of pressurization.

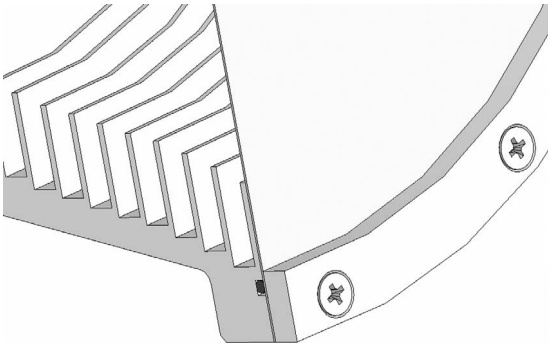


Figure 3.46 Radome secured by flat-faced ring and sealed by O-ring.

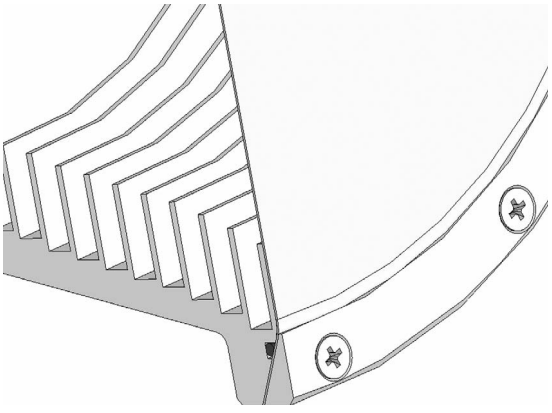


Figure 3.47 Radome secured by taper-faced ring and sealed by O-ring.

Where an antenna aperture must be sealed by a radome, the arrangement shown in Figure 3.46 is often used. This design is quite simple to apply, but in some instances the step formed by the radome attachment ring and the outer surface of the radome can collect a small pool of rainwater, dirt, or snow, which can affect antenna performance. The design shown in Figure 3.47 was developed by the authors to avoid this problem. While this design is slightly more difficult to produce, the step has been largely eliminated and the act of tightening the radome attachment screws draws the edges of the radome down the tapered face of the horn aperture and stretches it tightly. Note that although a corrugated horn is shown in Figures 3.46 and 3.47, these techniques can be applied to any aperture.

In designing the proportions of O-ring grooves for sealing, particularly on large diameter joints, an attempt should be made to use grooves of slightly larger diameter than those recommended for industrial use of O-rings, so that the O-rings have to be stretched a small amount to fit the groove during assembly. This will retain the O-ring in the groove during assembly, even in inverted positions, which can greatly simplify an often awkward assembly task, especially when performed at the vertex of a large antenna.

In most applications of pressurized antenna systems, a vent must be provided to allow for purging of atmospheric air after assembly. This may be in the form of a small industrial bleed or drain valve and should be placed in a convenient location.

The pressure applied to sealed and pressurized systems is not high, usually not more than 200–300 mm of water, but control of leakage is important if gas is not to be wasted or excessive power consumed by compressors and dehumidifiers. An antenna application may specify a maximum allowable leakage rate based on these factors. The authors have measured the leakage rate of pressurized systems by measuring system pressure of a sealed system with a water manometer, and timing the fall of pressure from 50 mm H_2O above the nominal system pressure to 50 mm H_2O below. A simple calculation based on the static volume of the pressurized system and the mean pressure during test gives the leakage rate.

3.5.7.2 Passive Intermodulation Achieving low incidence of passive intermodulation (PIM) in an antenna system is as much a result of good hardware design and fabrication as it is good microwave design. While this is a complex subject, some general principles can be applied during hardware design and fabrication.

The number of parts and therefore joints in an assembly must be kept to an absolute minimum, and the same material should be used throughout the assembly. Each joint is a potential source of PIM, as are contacts between dissimilar materials.

Where possible, complex devices should be manufactured in one piece, but where this is not feasible joints using brazing processes with filler metals of similar composition to the parent parts may be satisfactory. Such joining processes must be capable of producing smooth uninterrupted surfaces on the inside of the assembly.

Bolted joints require the utmost care in manufacture and assembly to ensure continuous and full contact between parts. Even precision machined surfaces will not necessarily meet this requirement so faces for bolted joints are usually lapped before assembly, first to a flat reference lapping plate and then to each other, and again before reassembly after being dismantled. Lapping should be carried out with a very fine nonembedding lapping compound and should proceed until full face contact is evident on both surfaces of a joint. Of course, all traces of lapping compound or residues must be removed from the parts before assembly.

Assembly steps, bolt tightening sequences, and torques may be critical in achieving a successful low PIM assembly and should be documented and rigidly applied once a joint design has been successfully verified.

3.5.7.3 Spaceflight Hardware The extreme environment of space imposes some stringent requirements on hardware design and fabrication. Materials used in flight hardware are subject to large temperature variations with rapid cycling, solar radiation, extremely high vacuum, and severe shock and vibration during launch.

Extremes of temperature can lead to problems when designing assemblies made from differing materials. The effects of differential expansion must be assessed to ensure that assemblies and fasteners do not become either loose or overloaded at the limits of temperature that will be encountered.

The performance of chosen materials must be assessed for the effects of radiation and high vacuum. Many materials degrade, lose mass, or outgas condensable vapors under these conditions. Space flight authorities such as NASA maintain databases of materials qualified for space flight applications and their performance under these conditions.

3.5.8 Examples

3.5.8.1 Machined Stepped or Corrugated Horns Corrugated horns are among the more difficult aperture antenna components to manufacture. Figure 3.48 shows a cross-sectional view of a profiled Ku-band corrugated horn, designed and made by the authors, which incorporates many of the physical features discussed previously, such as a sealed crush-type intermediate joint and a taper-flanged radome ring. The use of two main sections for this design allows an integral mounting flange to be incorporated at a convenient location and significantly reduces the depth to which boring bars or other tools must reach to machine the deep grooves at the small diameter input.

The authors have used two approaches to the manufacture of devices such as this: turning and milling. In both cases it is very worthwhile to make special tools for the job. Boring bars, in particular need to be as short and stiff as possible to avoid excessive chatter and deflection. Figure 3.49 shows how a CAD package can be used to design the shank of a special boring bar for the maximum possible size while still clearing both the front and rear walls during turning.

Figure 3.50 shows two such special boring bars for machining corrugations, made from steel bar stock with brazed-on carbide blanks. The carbide blanks were ground after brazing using a tool-and-cutter grinder. Note that the boring bar on the left has a

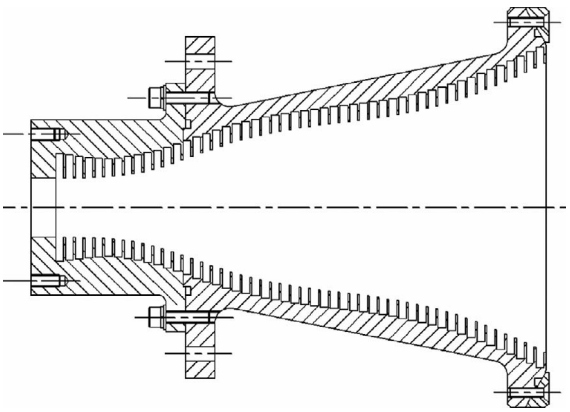


Figure 3.48 Profiled corrugated horn example.

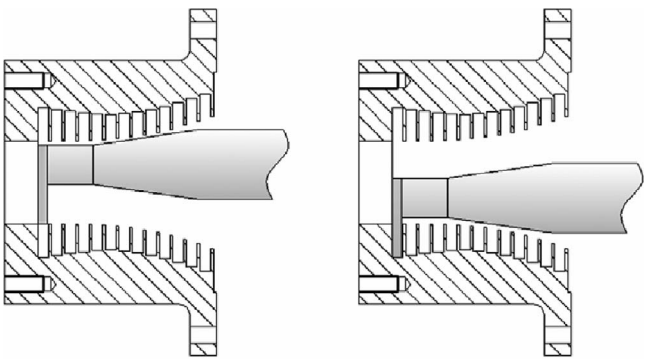


Figure 3.49 Designing a grooving boring bar shank for clearance and maximum stiffness.

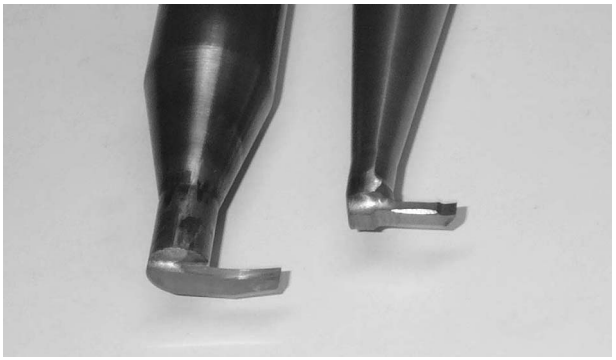


Figure 3.50 Grooving boring bars for turning corrugations in corrugated horns.

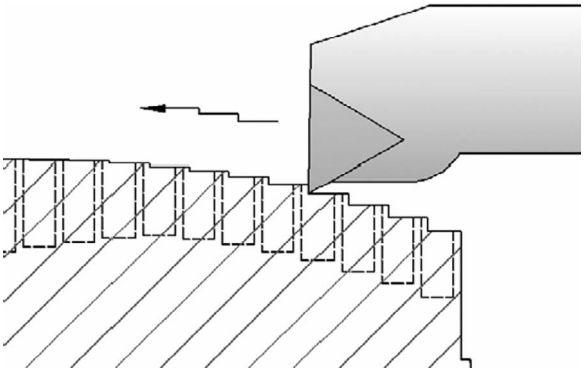


Figure 3.51 Boring steps to minor diameter of fins.

smoothly curved top face on the carbide to promote chip flow out of the corrugation, while the other has a chip breaker geometry.

For precision applications, boring bars such as these should be ground slightly narrower than the groove width to leave an allowance for a light finish cut on each side. Figure 3.51 shows the turning of a series of steps to the finished minor diameter of each fin using a commercial boring bar with a throwaway carbide insert, and Figures 3.52 and 3.53 show the roughing and finish cuts for the grooves with a specially made boring bar.

It will probably be found that any grooving boring bar will tend to chatter at the bottom of the corrugation when the infeed is stopped, leaving a poor finish. This can be prevented by reducing the lathe spindle speed at the end of each cut. On manually operated lathes, the spindle can be stopped and rotated slowly by hand for a few revolutions, and on CNC machines the spindle can be programmed to a low speed, $\sim$10 rpm or so, followed by a dwell for a few revolutions to wash out any deflection of the boring bar and avoid chatter.

The milling method is often used by the authors for machining larger sections of horns on a CNC machining center, where the workpiece can be very rigidly clamped on support blocks to the machine's worktable. The rotating tool obviously occupies more

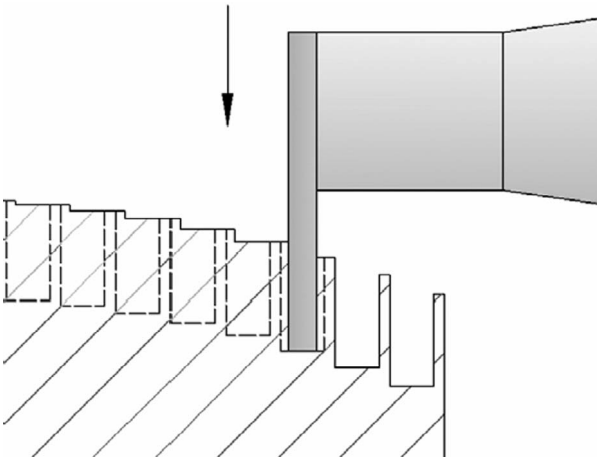


Figure 3.52 Roughing plunge cut for corrugation.

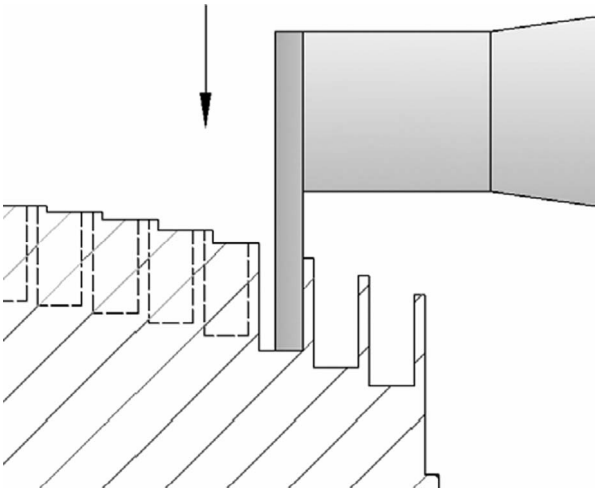


Figure 3.53 Finishing plunge cut for one side of corrugation.

space than the single-sided tool used for turning and boring, so this method is not always capable of machining features near the input of a horn antenna.

A distinct advantage of CNC milling over turning is swarf control. Milling swarf inherently consists of short chips that are easily removed by the copious coolant supply available on CNC machines, which flows out from under the horn between the support blocks. Unlike turning, the stationary workpiece ensures swarf is not driven into the corrugations by centrifugal forces.

A specially made grooving cutter such as the one shown in Figure 3.54 is used to mill corrugations. Like the grooving boring bars described earlier, this cutter is made from steel with carbide inserts brazed in place, and ground after brazing. The cutter is drilled for coolant from the machine's high pressure through-spindle coolant supply to

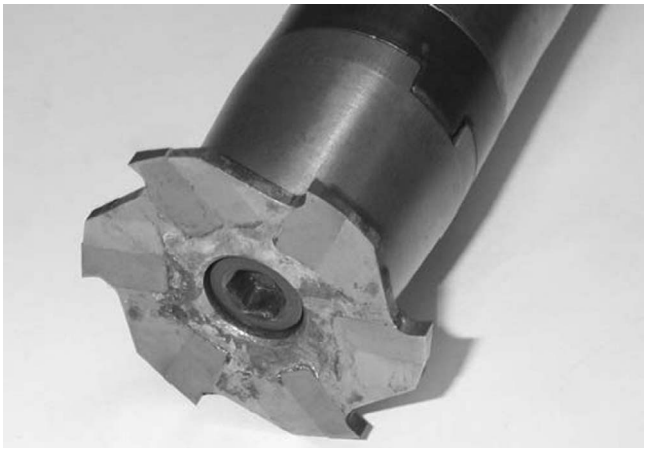


Figure 3.54 Grooving cutter for milling corrugated horns.

Figure 3.55 Grooving cutter and extended machine arbor.

spray into each tooth gullet to clear swarf from the working area. The cutter is mounted for use on a long machine arbor, as shown in Figure 3.55.

In practice, very high machining speeds can be achieved with good finish and complete freedom from chatter. Figure 3.56 shows a number of ~260-mm diameter Ku-band corrugated horn sections being milled at a spindle speed of 10,000 rpm.

The CNC program used for this work was written by the authors using the macro-language capabilities of the EIA/ISO CNC machine tool language. A subprogram controlling one roughing and two finishing cuts for each corrugation is run from a main program containing a list of sizes for all the corrugations. Each set of sizes is read sequentially to the subprogram as local variables before execution.

Figure 3.57 illustrates the order of operations for each of the three cuts made by the subroutine. A curved in-feed move at low feedrate to the full depth is followed by a fast circular path cut at maximum feedrate. A curved exit move returns the tool to the starting point.

Figure 3.56 CNC milling parts for corrugated feed horns at 10,000 rpm.

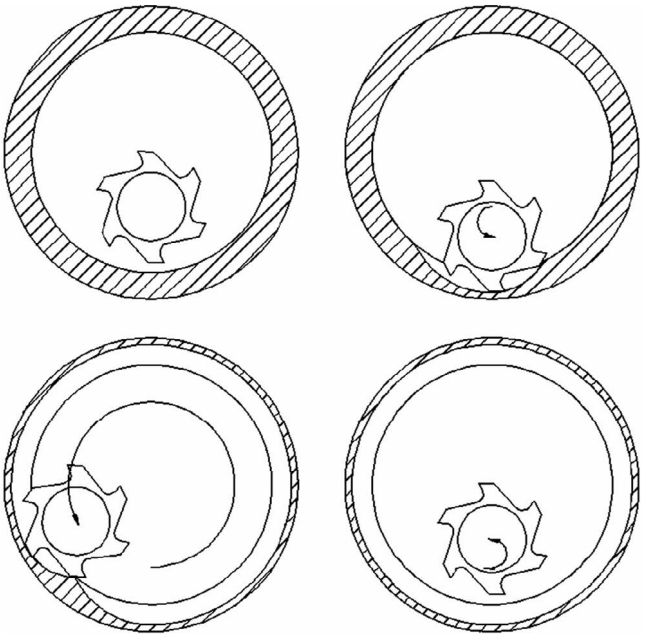


Figure 3.57 Movements to mill a circular corrugation.

3.5.8.2 Fabricated Horns Rectangular horns, especially for large sizes, are often fabricated from flat sheet sections, joined by welding, brazing, or soldering, depending on the strength and accuracy required. A typical rectangular horn shape is illustrated in Figure 3.17. Both welding and torch brazing processes can induce distortion in flat parts due to the high temperatures involved. Where complete freedom from distortion is necessary, brazing in a furnace can be effective, or a low temperature process such as soldering can be employed.

The typical joint configurations for these alternate joining methods are quite different. Figure 3.58 shows a section through a corner joint of a welded fabrication. Note that

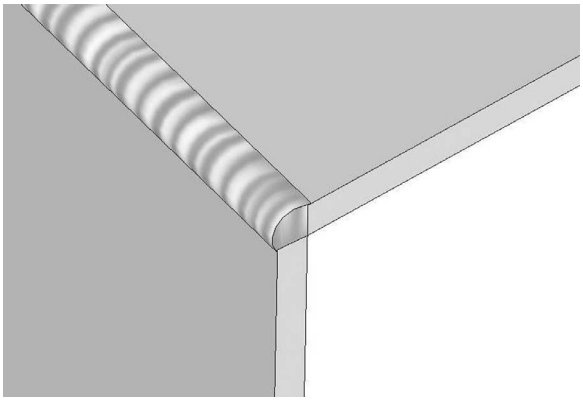


Figure 3.58 A welded joint in a fabricated rectangular horn.

the two members do not overlap significantly. This helps to ensure complete penetration of the weld pool toward the inner surfaces during welding to minimize gaps or crevices along the inside of the joint.

Backup bars, sometimes known as chills, may be placed on the inside of the joint to prevent burn-through of the weld pool and can be used to support the parts in correct alignment during welding and help minimize distortion. Backup bars, however, increase the welding current required as the additional mass of the bar acts as a significant heatsink.

It should be noted that welded joints of this type cannot always be performed in aluminum by simply fusing the parent parts together without the use of additional filler metal. Some aluminum alloys, notably 6061, will invariably crack if this is attempted without use of a compatible filler [49].

Brazed joints require one of the sections to overlap the other, as shown in Figure 3.59. In this case, the overlapping section is extended past the outside surface of the joint to create a lip to which filler metal can be applied by hand from the outside during torch brazing. The lip can be used as a location for preplaced filler wire during furnace brazing.

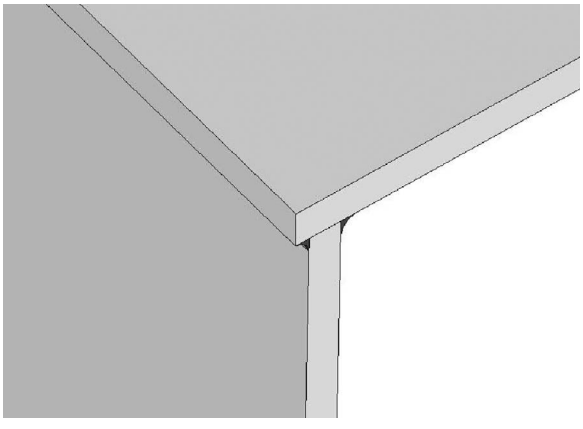


Figure 3.59 Brazed joint in a fabricated rectangular horn, showing overlap of parts.

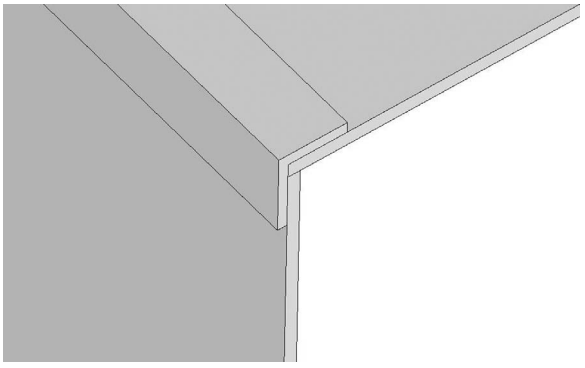


Figure 3.60 Soldered joint in a fabricated rectangular horn, showing use of reinforcement.

The filler metal flows through the joint by capillary action to form a small fillet on both sides.

Soldered construction also requires overlap of parts, but in this case the low strength of the filler metal usually requires additional reinforcement of the joint in the form of cover strips. Figure 3.60 shows a soldered joint and illustrates the use of a reinforcement of formed angle placed along the outside of the joint to increase the surface area of the solder joint.

3.5.8.3 Transitions Waveguide transitions are often needed to connect waveguides to the inputs of aperture antennas. Where the transition provides a change in diameter of a circular waveguide, a turned component can be used. Occasionally, however, a transition must also provide a change in waveguide shape. A method that is often used to make such devices is wire electric discharge machining, or wire EDM.

In this process an electrical discharge between a tightly stretched wire and the workpiece is used to burn a slot in the workpiece, and so to saw out a hole or shape. The wire

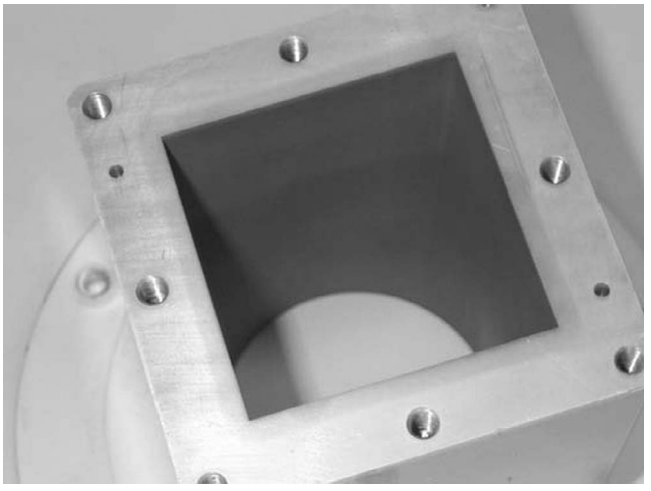


Figure 3.61 A circular-to-square transition cut by wire EDM.

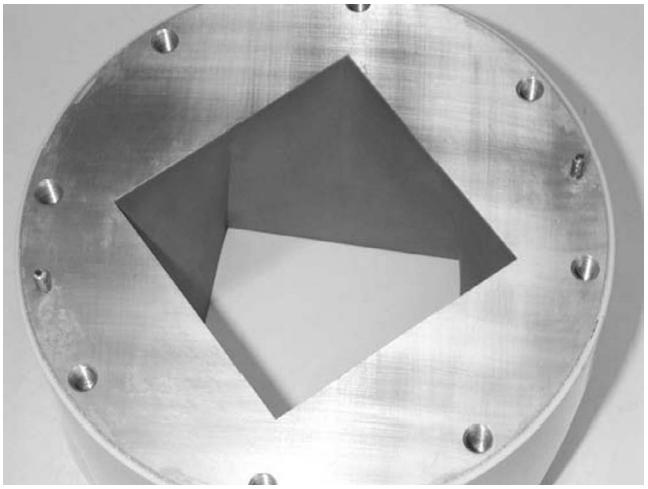


Figure 3.62 A square-to-square 45° 3-dB splitter cut by wire EDM.

is also partly consumed in the process and is continually replaced from a reel of new wire. By simultaneous manipulation of wire guides at each end, various shapes can be economically produced. Figures 3.61 and 3.62 illustrate two variations, a circular-to-square transition and a square-to-square 45° 3-dB splitter.

REFERENCES

1. C. A. Balanis, Horn antennas, in *Antenna Handbook*, Y. T. Lo and S.W. Lee (Eds.), Van Nostrand, New York 1988; Chap. 8.
2. A. D. Olver, P. J. B. Clarricoats, A. A. Kishk and L. Shafai. *Microwave Horns and Feeds*, IEE Electromagnetic Wave Series 39, Peter Peregrinus Ltd., London, 1994.
3. T. A. Milligan, *Modern Antenna Design*, 2nd ed., John Wiley & Sons, Hoboken, NJ, 2005.
4. A. W. Rudge, K. Milne, A. D. Olver and P. Knight, (Eds.), *The Handbook of Antenna Design*, IEE Electromagnetic Wave Series 16, Peter Peregrinus Ltd., London, 1982.
5. K. MacReynolds, and M. H. Francis, Antenna gain measurements: the three-antenna extrapolation method, *AMTA Conference,* pp. 370–375, 1999.
6. M. Uhm, A. Shishlov, and K. Park, Offset-paraboloid geometry: relations for practical use. *IEEE Antennas Propag. Mag.* Vol. 38, No. 3, pp. 77–79, 1996.
7. W. Stutzman and M. Terada, Design of offset-parabolic-reflector antennas for low cross-polar and low sidelobes, *IEEE Antennas Propag. Mag.*, Vol. 35, No.6, pp. 46–49, 1993.
8. G. L. James and B. MacA. Thomas, Comparison of *G/T* between dual-reflector and primary-focus antennas, *Electron. Lett.*, Vol. 16, No.8, pp. 286–288, 1980.
9. V. Hombach and E. Kühn, Complete dual-offset reflector antenna analysis including near-field, paint-layer and CFRP-structure effects, *IEEE Trans. Antennas Propag.*, Vol. 37, No. 9, pp. 1093–1101, 1989.
10. R. P. Flam. Antennas for optimum illumination of anechoic chambers, in *International Conference on Antennas and Propagation, ICAP'87*, IEE Publications, Newyork, 1987, pp. 351–354.
11. C. Granet, T. S. Bird and G. L. James, Compact multimode horn with low sidelobes for global earth coverage, *IEEE Trans. Antennas Propag.*, Vol. 48, No. 7, pp. 1125–1133, 2000.

12. L. Staveley-Smith, W. E. Wilson, T. S. Bird, M. J. Disney, R. D. Ekers, K. C. Freeman, R. F. Haynes, M. W. Sinclair, R. A. Vaile, R. L. Webster, A. E. Wright, The Parkes 21cm multibeam receiver, *Publ. Astron. Soc. Austral.* Vol. 14, No. 3, pp. 243–248, 1996.
13. J. Galejs, Admittance of a rectangular slot which is backed by a rectangular cavity, *IEEE Trans. Antennas Propag.*, Vol. 11, No. 2, pp. 119–126, 1963.
14. C. R. Cockrell, The input admittance of the rectangular cavity-backed slot antenna, *IEEE Trans. Antennas Propag.*, Vol. 24, No. 3, pp. 288–294, 1976.
15. S. A. Long, A mathematical model for the impedance of the cavity-backed slot antenna, *IEEE Trans. Antennas Propag.*, Vol. 25, No. 6, pp. 829–833, 1977.
16. A. J. Sangster, Cirularly polarised linear waveguide array, *IEEE Trans. Antennas Propag.*, Vol. 21, No. 5, pp. 704–705, 1973.
17. M. E. Armstrong and N. G. Alexopoulos, On the design of a circularly polarised waveguide narrow wall linear array, *IEEE Trans. Antennas Propag.*, Vol. 23, No. 2, pp. 244–250, 1975.
18. J. S. Ajiola, D. M. Joe and J. L. McFarland, Slot radiators in septated waveguide, *IEEE Trans. Antennas Propag.*, Vol. 32, No. 3, pp. 247–251, 1984.
19. R. C. Johnson (Ed.), *Antenna Engineering Handbook*, 3rd ed, McGraw-Hill, New York, 1993. Chap. 9.
20. D. Y. Kim and R. S. Elliott, A design procedure for slot arrays fed by single-ridge waveguide, *IEEE Trans. Antennas Propag.*, Vol. 36, No. 11, pp. 1531–1536, 1988.
21. J. Green, H. Shnitkin and P. J. Bertalan, Asymmetric ridge waveguide radiating element for a scanned planar array, *IEEE Trans. Antennas Propag.*, Vol. 38, No. 8, pp. 1161–1165, 1990.
22. R. C. Johnson, (Ed.), *Antenna Engineering Handbook*, 3rd ed., McGraw-Hill, New York, 1993, Chap 28, pp. 28-29– 28-33.
23. M. Takahashi, M. Ando, N. Goto, Y. Numano, M. Suzuki, Y. Okazaki, and T. Yoshimoto, Dual circularly polarised radial line slot antennas, *IEEE Trans. Antennas Propag.*, Vol. 43, No. 8, pp. 874–876, 1995.
24. P. D. Patel, Inexpensive multi-mode satellite tracking feed antenna, *IEE Proc.*, Vol. 135, Pt. H, No. 6, pp. 381–386, 1988.
25. S. Drabovitch, A new tracking modes coupler using a corrugated feed for satellite communication earth-station antenna, *European Microwave Conference*, pp. 165–168, 1976.
26. G. L. James and K. J. Greene, Effect of wall thickness on radiation from circular waveguides, *Electron. Lett.*, Vol. 14, No 4, pp. 90–91, 1978.
27. K. M. Lee and R. S. Chu, Design and analysis of a multimode feed horn for a monopulse feed, *IEEE Trans. Antennas Propag.*, Vol. 36, pp. 171–181, 1988.
28. R. D. Elliot and P. J. B. Clarricoats, Corrugated waveguide monopulse feed, *Electron. Lett.*, Vol. 16, No. 9, pp. 324–325, 1980.
29. H. Z. Zhang, C. Granet and M. A. Sprey, A compact Ku-band monopulse horn, *Microwave Optical Technol. Lett.*, Vol. 34, pp. 9–13, 2002.
30. R. Corkish, The use of conical tips to improve the impedance matching of Cassegrain subreflectors, *Microwave Optical Technol. Lett.*, Vol. 3, No. 9, pp. 310–313, 1990.
31. M. A. Sprey and T. S. Bird, Minimise pattern degradation when subreflector vertex shaping is used to improve feed match, *Electron. Lett.*, Vol. 32 No. 9, pp. 781–782, 1996.
32. C. C. Han and Y. Hwang, Satellite antennas, in *Antenna Handbook*, Y. T. Lo and S.W. Lee (Eds.), Van Nostrand Reinhold, New York, 1988, Chap. 21.
33. G. L. James, Mode-matching techniques in analysing waveguide structures, *Trends Microwave Theory Tech.*, Vol. 1, pp. 103–121, 1990.
34. G. L. James, Analysis and design of TE11 to HE11 corrugated cylindrical waveguide mode converters, *IEEE Trans. Microwave Theory Tech.*, Vol. MTT-29, pp. 1059–1066, 1981.

35. T. S. Bird, Antenna feeds, in *Encyclopedia of RF and Microwave Engineering*, Vol. 1, John Wiley & Sons, Hoboken, NJ, pp. 2005, 185–217.

36. T. S. Bird and G. L. James, Design and practice of reflector and feed systems in the 1990s, in *Review of Radio Science 1996–1999*, W. R. Stone, (Ed.), URSI, pp. 81–117, 1999.

37. G. L. James, *Geometrical Theory of Diffraction for Electromagnetic Waves*, 3rd ed. rev., IEE Electromagnetic Waves Series 1, Peter Peregrinus Ltd., London, 1986.

38. S. J. Blank, A note on reflector-feed element realizability and optimality, *IEEE Antennas Propag. Mag.*, Vol. 39, No. 2, pp. 68–71, April 1997.

39. Y. Rahmat-Samii, P. Cramer, K. Woo and S. W. Lee, Realizable feed-element patterns for multibeam reflector antenna analysis, *IEEE Trans. Antennas Propag.*, Vol. 29, No 6, pp. 961–963, 1981.

40. K. Guney and H. Hancer, Improved formulas for narrow and wide aperture dimensions of optimum gain pyramidal horn, *Int. J. RF Microwave Computer-aided Eng.*, Vol. 13, No. 3, pp. 239–245, 2003.

41. P. D. Potter, A new horn antenna with suppressed sidelobes and equal beamwidth, *Microwave J.*, Vol. 6, pp. 71–78, June 1963.

42. S. P. Skobolev, B.-J. Ku, A.V. Shishlov and D.-S. Ahn, Optimum geometry and performance of a dual-mode horn modification, *IEEE Antennas Propag. Mag.*, Vol. 43, No. 1, pp. 90–93, February 2001.

43. C. Granet and G. L. James, Design of corrugated horns: a primer, *IEEE Antennas Propag. Mag.*, Vol. 47, No. 2, pp. 76–84, April 2005.

44. *ASM Handbook, Volume 2, Properties and Selection: Nonferrous Alloys and Special-Purpose Materials*, ASM International, 1990.

45. *Aluminum Casting Technology*, 2nd ed., American Foundrymen's Society, 1993.

46. *MIL-STD 5541, Chemical Conversion Coatings on Aluminum and Aluminum Alloys.*

47. *ASTM A967-01e1 Standard Specification for Chemical Passivation Treatments for Stainless Parts.*

48. *MIL-STB 889B Dissimilar Metals.*

49. *ASM Handbook, Volume 6, Welding, Brazing and Soldering*, ASM International, 1993.

CHAPTER 4

Microstrip Antennas: Analysis, Design, and Application

JOHN HUANG

4.1 INTRODUCTION

Since the invention of the microstrip antenna a half-century ago [1, 2], the demand for its application [3–9] has been increasing rapidly, especially within the past two decades. Because of the microstrip antenna's many unique and attractive properties, there seems to be little doubt that it will continue to find many applications in the future. These properties include low profile, light weight, compact and conformable to mounting structure, easy fabrication and integratable with solid-state devices. The results of these properties contributed to the success of microstrip antennas not only in military applications such as aircraft, missiles, rockets, and spacecraft but also in commercial areas such as mobile satellite communications, terrestrial cellular communications, direct broadcast satellite (DBS) system, global positioning system (GPS), remote sensing, and hyperthermia. Although the microstrip antenna is generally known for its shortcoming of narrow bandwidth, recent technology advances have improved its bandwidth from a few percent to tens of percent. To understand the microstrip antenna's performance and to simplify its design process, several numerical analysis techniques have been developed and converted to computer-aided design (CAD) tools. Some of these analysis techniques also allow the designer to gain physical insight into the antenna's electrical operating mechanism. It is the purpose of this chapter to discuss some of the microstrip antenna's technical features, its advantages and disadvantages, substrate material considerations (in particular, for space application), excitation techniques, polarization behaviors, bandwidth characteristics, and miniaturization techniques. Discussion of the physical mechanisms of the various microstrip antennas is emphasized throughout the chapter. Analysis techniques, design processes and CAD tools are briefly presented. Several recent interesting applications of the microstrip antenna are also highlighted.

Modern Antenna Handbook. Edited by Constantine A. Balanis

4.2 TECHNICAL BACKGROUND

This section presents the technical background of the microstrip antenna, which is separated into three areas: features of the microstrip antenna, advantage and disadvantage trade-offs, and material considerations.

4.2.1 Features of the Microstrip Antenna

At the early stage of its development, the microstrip antenna [10, 11], as shown in Figure 4.1, is generally a single-layer design and consists of a radiating metallic patch or an array of patches situated on one side of a thin, nonconducting, substrate panel with a metallic ground plane situated on the other side of the panel. The metallic patch is normally made of thin copper foil or is copper-foil plated with a corrosion resistive metal, such as gold, tin, or nickel. Each patch can be designed with a variety of shapes, with the most popular shapes being rectangular or circular. The substrate panel generally has a thickness in the range of 0.01–0.05 free-space wavelength (λ_0). It is used primarily to provide proper spacing and mechanical support between the patch and its ground plane. It is also often used with high dielectric-constant material to load the patch and reduce its size. The substrate material should be low in insertion loss with a loss tangent of less than 0.005, in particular, for large array application. Generally, substrate materials [11] can be separated into three categories in accordance with their dielectric constant:

1. Having a relative dielectric constant (ε_r) in the range of 1.0–2.0. This type of material can be air, polystyrene foam, or dielectric honeycomb.
2. Having ε_r in the range of 2.0–4.0 with material consisting mostly of fiberglass reinforced Teflon.

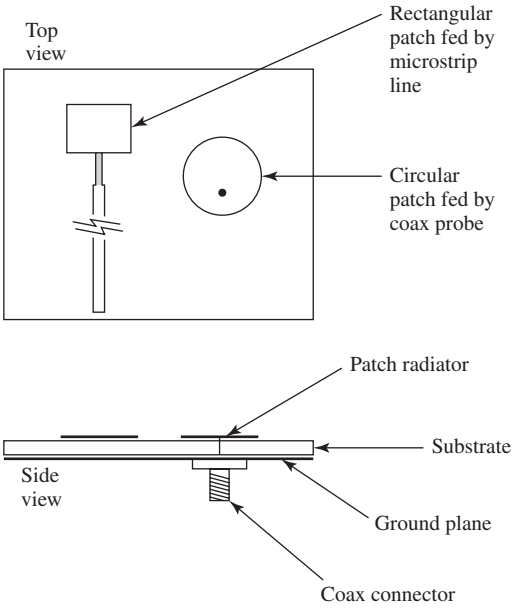


Figure 4.1 Configuration of microstrip patch elements.

3. With an ε_r between 4 and 10. The material can consist of ceramic, quartz, or alumina.

Although there are materials with ε_r much higher than 10, one should be careful in using these materials. As to be discussed later, they can significantly reduce the antenna's radiation efficiency.

Although a microstrip antenna can be excited by various methods (described in Section 4.4), a single microstrip patch can be simply excited either by a coaxial probe or by a microstrip transmission line as shown in Figure 4.1. For an array of microstrip patches, the patches can be combined either with microstrip lines located on the same side of the patches or with microstrip lines/striplines designed on separate layers placed behind the ground plane. For the separate-layer configuration, each patch and its feed line are electrically connected either by a small-diameter metal post or by an aperture coupling slot [12]. Regardless of the different layer configurations, tens or hundreds of patch elements in an array can be fabricated by a single, low cost chemical etching process, and each single-patch element does not need to be fabricated individually (as many other types of radiating elements do), which will lead to an overall lower antenna manufacturing cost.

4.2.2 Advantage and Disadvantage Trade-offs

There are advantages as well as disadvantages associated with the microstrip antenna. By understanding them well, one can readily design a microstrip antenna with optimum efficiency, minimum risk, and lower cost for a particular application.

The advantages of microstrip antennas when compared to conventional antennas (helix, horn, reflector, etc.) are the following.

1. The extremely low profile of the microstrip antenna makes it lightweight and it occupies very little volume of the structure or vehicle on which it is mounted. It can be conformally mounted onto a curved surface so it is aesthetically appealing and aerodynamically sound. Large aperture microstrip arrays on flat panels can be made mechanically foldable for space application [13, 14].
2. The patch element or an array of patch elements, when produced in large quantities, can be fabricated with a simple etching process, which can lead to greatly reduced fabrication cost. The patch element can also be integrated or made monolithic with other microwave active/passive components.
3. Multiple-frequency operation is possible by using either stacked patches [15] or a patch with loaded pin [16] or a stub [17].
4. There are other miscellaneous advantages, such as the low antenna radar cross section (RCS) when conformally mounted on aircraft or missiles, and the microstrip antenna technology can be combined with the reflectarray technology [18] to achieve very large aperture without any complex and RF lossy beamformer.

The disadvantages of the microstrip antennas are the following

1. A single-patch microstrip antenna with a thin substrate (thickness less than 0.02 free-space wavelength) generally has a narrow bandwidth of less than 5%. However, with technology advancement, up to 50% bandwidths have been achieved.

The bandwidth-widening techniques include multiple stacked patches, thicker substrate with aperture slot coupling [19, 20], external matching circuits [21], a sequential rotation element arrangement [22, 23], parasitic coupling [24], U-slot feed [25, 26], and L-shaped probe feed [27, 28]. It is generally true that wider bandwidth is achieved with the sacrifice of increased antenna physical volume.

2. The microstrip antenna can handle relatively lower RF power due to the small separation between the radiating patch and its ground plane (equivalent to small separation between two electrodes). Generally, a few tens of watts of average power or less is considered safe. However, depending on the substrate thickness, metal edge sharpness, and the frequency of operation, a few kilowatts of peak power for microstrip lines at X-band have been reported [29]. It should be noted that, for space application, the power-handling capability is generally less than that for ground application due to a mechanism called multipacting breakdown [30].
3. The microstrip array generally has a larger ohmic insertion loss than other types of antennas of equivalent aperture size. This ohmic loss mostly occurs in the dielectric substrate and the metal conductor of the microstrip line power dividing circuit. It should be noted that a single-patch element generally incurs very little loss because it is only one-half wavelength long. The loss in the power dividing circuit of a microstrip array can be minimized by using several approaches, such as the series feed power divider lines [11, 31], waveguide and microstrip combined power dividers, and honeycomb or foam low loss substrates. For very large arrays, transmit/receive (T/R) amplifier modules can be used on elements or subarrays to mitigate the effect of large insertion loss.

4.2.3 Material Considerations

The purpose of the substrate material of a microstrip antenna is primarily to provide mechanical support for the radiating patch elements and to maintain the required precision spacing between the patch and its ground plane. With higher dielectric constant of the substrate material, the patch size can also be reduced due to a loading effect (to be discussed later). Certainly, with reduced antenna volume, higher dielectric constant also reduces bandwidth. There are a variety of substrate materials. As discussed in Section 4.2.1, the relative dielectric constant of these materials can be anywhere from 1 to 10. Materials with dielectric constant higher than 10 should be used with care. They can significantly reduce the radiation efficiency by having overly small antenna volumes. The most popular type of material is Teflon based with a relative dielectric constant between 2 and 3. This Teflon-based material, also named PTFE (polytetrafluoroethylene), has a structure form very similar to the fiberglass material used for digital circuit boards but has a much lower loss tangent or insertion loss. The selection of the appropriate material for a microstrip antenna should be based on the desired patch size, bandwidth, insertion loss, thermal stability, cost, and so on. For commercial application, cost is one of the most important criteria in determining the substrate type. For example, a single patch or an array of a few elements may be fabricated on a low cost fiberglass material at the L-band frequency, while a 20-element array at 30 GHz may have to use higher cost, but lower loss, Teflon-based material. For a large number of array elements at lower microwave frequencies (below 20 GHz), a dielectric honeycomb or foam panel may be used as substrate to minimize insertion loss, antenna mass, and material cost

with increased bandwidth performance. A detailed discussion of substrate material can be found in Ref. 11.

When the microstrip antenna is used for space application, its substrate material must survive three major effects related to space environment: radiation exposure, material outgassing, and temperature change. Exposure to cosmic high energy radiation is an important factor in space applications. Cosmic radiations, such as beta, gamma, and X-ray, are similar to nuclear radiation in many respects. They can damage materials after the prolonged exposure typical of a long space mission. Outgassing is another phenomenon of concern for material in space. Outgassing will cause a material to lose its mass in the form of gases or volatile condensable matter when subject to a vacuum, especially when it is heated as the antenna is exposed to sunlight in space. Losing mass will certainly affect the material's mechanical and electrical properties. The effect of temperature in space on electrical and physical properties of the substrate material must be taken into consideration when designing a microstrip antenna. Since space is a vacuum without conduction medium, the temperature of an object could be extremely cold (e.g., -100°C) when it is not exposed to sunlight or it could become very hot, (e.g., $+100^{\circ}$C.) when it is directly illuminated by the sun over a period of time. The effects of these extreme temperatures could cause changes in the microstrip substrate material, including the dielectric constant (ε) and substrate thickness, which together could cause an impedance change of the microstrip patch or transmission line.

4.3 ANALYSIS AND DESIGN

4.3.1 Analysis Techniques

The main reason for developing an analytic model for the microstrip antenna is to provide a means of designing the antenna without costly and tedious experimental iteration. Also, it may allow the designer to discover the physical mechanisms of how the microstrip antenna operates. With an analysis technique, the engineer should be able to predict the antenna performance qualities, such as the input impedance, resonant frequency, bandwidth, radiation patterns, and efficiency. There are many different analysis techniques that have been developed for analyzing the microstrip antennas. However, the most popular ones can be separated into five groups: transmission-line circuit model, multimode cavity model, moment method, finite-difference time-domain (FDTD) method, and finite-element method. They are briefly discussed below:

4.3.1.1 Transmission-Line Circuit Model A microstrip patch, operating at its fundamental mode, is essentially a $\frac{1}{2}\lambda$-long microstrip transmission line and can be represented by an equivalent circuit network [32, 33]. For a rectangular or square patch, its radiation is basically generated from its two edges with two equivalent slots along the resonating dimension, as shown in Figure 4.2. Thus the microstrip radiator can be characterized by two slots separated by a transmission line, where each slot is represented by a parallel circuit of conductance *(G)* and susceptance *(B)*. The complete patch antenna can be represented by the equivalent network shown in Figure 4.3 [32]. This transmission-line model is simple, intuitively appealing, and computationally fast, but it suffers from limited accuracy. For example, this model lacks the radiation from the nonradiating edges of the patch, and it has no mutual coupling between the two radiating

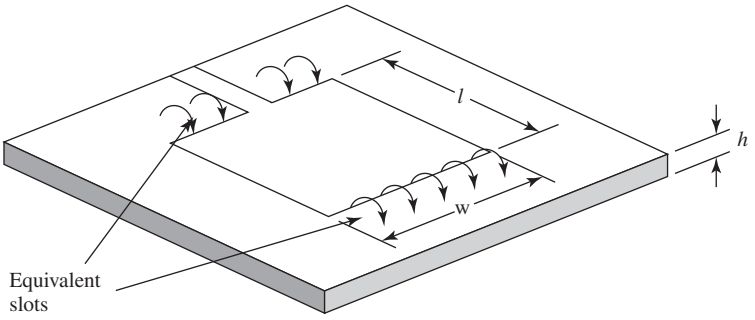


Figure 4.2 Microstrip patch radiation source represented by two equivalent slots.

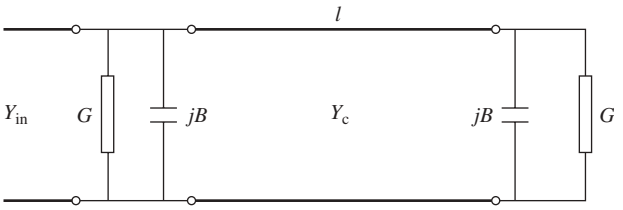


Figure 4.3 Equivalent circuit of a microstrip patch element.

slots. Although this model has led to a much improved version [33], it lacks the flexibility and generalization of analyzing other patch shapes.

4.3.1.2 Multimode Cavity Model Any microstrip radiator can be thought of as an open cavity bounded by the patch and its ground plane. The open edges can also be represented by radiating magnetic walls. Such a cavity will support multiple discrete modes similar to that of a completely enclosed metallic cavity. As an example, for a rectangular or square patch with relative dielectric constant ε_r, substrate thickness h, and patch dimensions $L \times W$ (see Figure 4.2), the total electric field in the cavity can be expressed as the sum of the fields associated with each sinusoidal mode [34]:

$$E_z(x, y) = \sum_m \sum_n C_{mn} \cdot \cos\left(\frac{m\pi}{L}\right) x \cdot \cos\left(\frac{n\pi}{W}\right) y \tag{4.1}$$

where C_{mn} is a constant that depends on the feed location, L and W dimensions, and dielectric constant. Due to the very thin substrate, the fields are assumed to be z-directed only, with no variation in the z-direction. The most interesting dominant mode is the TM_{10} mode, which could be obtained if the dimension L is approximately $\lambda_g/2$ (λ_g is the effective wavelength in the dielectric). The field variation underneath the patch for this fundamental mode is illustrated in Figure 4.4, and the radiating fringing fields are shown in Figure 4.5a. These figures indicate that, along the central line orthogonal to the resonant direction (x-direction), it is a null field region underneath the patch. This is why one is able to place shorting pins or additional feed probes along this central line without disturbing the performance of the patch of the original feed. This is also why two orthogonally placed feed probes can achieve dual-linear polarization without much

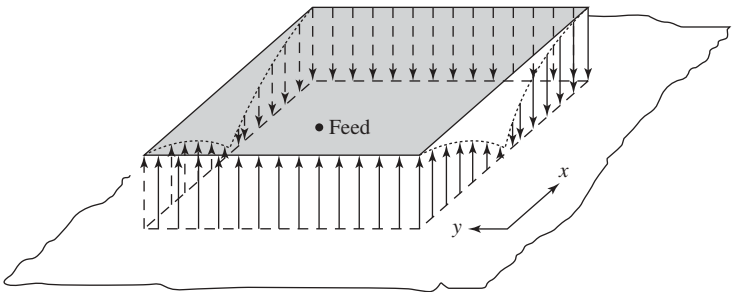


Figure 4.4 Fundamental-mode electric-field configuration underneath a rectangular patch.

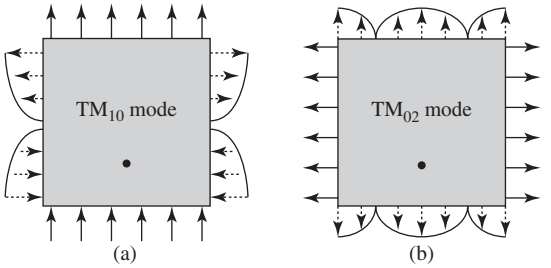


Figure 4.5 Fringing fields for fundamental-mode TM_{10} and higher order mode TM_{02}.

cross-talk between the two probes. In Figure 4.5a, the left- and right-edge fringing fields do not contribute much to far fields due to their oscillatory behavior and, hence, cancel each other in the far field. The top- and bottom-edge fringing fields are the primary contributors to the far-field radiation of a patch. This is further illustrated in Figure 4.2 with two equivalent slots. Thus the basic radiating mechanism of a patch (rectangular or circular) consists of two radiating slots spaced about $1/2\lambda_g$ apart. A secondary higher order mode that does contribute significantly to the cross-polarization radiation is the TM_{02} mode. This mode, shown in Figure 4.5b, has the left and right edges contributing to the far-field radiation but with lower magnitude than the TM_{10} mode. One should know that the top- and bottom-edge fringing fields contribute to the co polarization radiation in both the E and H-planes, while the left- and right-edge fringing fields yield the cross polarization radiation only in the H-plane pattern as illustrated in Figure 4.6. In the E-plane, the left- and right-edge fringing fields always cancel each other (note the arrow directions of the fringing fields). For a circular patch [35], although radial modes and angular modes are involved, the radiation mechanism is very similar to that of a rectangular patch.

By knowing the total fields at the edges of the patch from all modes, the equivalent edge magnetic currents can be determined and integrated to find the total far-field radiation patterns. By knowing the total radiated power and the input power, one can also determine the input impedance. The cavity model technique allows one to determine the mode structure underneath the patch, and therefore its physical mechanisms are more easily understood, such as its resonating and cross-polarization behaviors. However, because it assumes the field has no z-variation, its solution is not very accurate, especially when

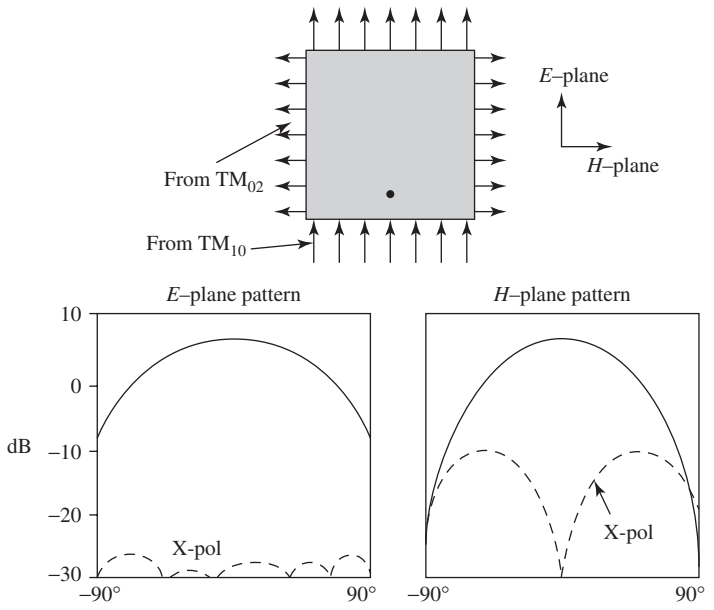


Figure 4.6 Basic E- and H-plane pattern shapes from a rectangular patch.

the substrate becomes thick (for wider bandwidth consideration). Also, the calculation of mutual coupling between patches in an array environment is very tedious and inaccurate.

4.3.1.3 Moment Method The radiated fields of a microstrip antenna can be determined by integrating all the electrical currents on its metallic surfaces via the integral equation approach whose solution is obtained by the so-called moment method. This integral equation approach [36–39] is analyzed by first solving the vector potential $\mathbf{A}(x,y,z)$, which satisfies the wave equation with J_s being the patch surface current:

$$\nabla^2\mathbf{A}^{\mathrm{I}} + k^2\mathbf{A}^{\mathrm{I}} = -juJ_S(x, y) \quad \text{in the dielectric (region I)} \tag{4.2}$$

and

$$\nabla^2\mathbf{A}^{\mathrm{II}} + k_0^2\mathbf{A}^{\mathrm{II}} = 0 \quad \text{infree space (region II)} \tag{4.3}$$

then the vector potential may be given as

$$\mathbf{A}^{\mathrm{I,II}}(x, y, z) = \iint_{\text{patch}} \mathbf{J}_S(x', y') \bullet \overline{\overline{G}}^{\mathrm{I,II}}(x, y, z/x', y', z')\, dx'\, dy' \tag{4.4}$$

where $\overline{\overline{G}}^{\mathrm{I,II}}$ is the dyadic Green's function for regions I and II. Region I contains the substrate, while region II is the free-space area above the substrate. The electric field $\mathbf{E}$ everywhere is given by

$$\mathbf{E}(x, y, z) = -j\omega\mathbf{A} + \frac{j\omega}{k^2}\nabla(\nabla \bullet \mathbf{A}) \tag{4.5}$$

By weighting the Green's function of Eq. (4.4) with the unknown electrical current density and integrating over the patch, the radiated electric or magnetic field can be calculated anywhere outside the dielectric. An integral equation for the unknown current is obtained by forcing the total tangential electric field on the patch surface to zero. Using the proper basis and testing functions for the unknown current, the integral equation is then discretized and reduced to a matrix equation:

$$[E] = [Z_{mn}][J] \tag{4.6}$$

where the impedance matrix element has the form

$$Z_{mn} = \int_x \int_y \int_{x'} \int_{y'} \int_{k_x} \int_{k_y} J^m(x, y) \cdot G(k_x, k_y) \cdot J^n(x', y') \cdot e^{-jk_x(x-x')} \cdot e^{-jk_y(y-y')} dk_y \, dk_x \, dy' \, dx' \, dy \, dx \tag{4.7}$$

where $G(k_x,k_y)$ is the Fourier transform of the Green's function given in Eq. (4.4), J^m is the mth expansion mode, and J^n is the nth weighting or testing mode. Equation (4.7) has been solved by two different approaches. One uses the space-domain approach [38, 39], where the spectral variables k_x and k_y are transformed to spatial polar coordinates α and β. The other approach uses the spectral-domain approach [36, 37], where the spectral integrations in Eq. (4.7) are done in closed form and result in an integral in the spectral domain only. Nevertheless, both approaches are derived to solve, via the method of moment and matrix inversion, for the patch surface current, which is then used to determine the properties of the microstrip antenna, such as the input impedance and radiation patterns. The moment method, a two-dimensional (2D) integration technique, is considered very accurate and includes the effects of mutual coupling between two surface current elements as well as the surface wave effect in the dielectric. It is computationally more time consuming than the transmission-line model and the cavity model. However, it is more computationally efficient than the three-dimensional (3D) technique to be discussed next.

4.3.1.4 Finite-Difference Time-Domain (FDTD) Method

The previous moment method is basically a two-dimensional solver. It solves for the 2D surface current on the microstrip patch. The FDTD method, on the other hand, is a three-dimensional solver. It solves for the electromagnetic fields in a 3D volumetric space. Thus it can solve more complex problems with 3D interfaces and connections, such as the multilayer microstrip antenna with complicated multilayer connections. However, it suffers from laborious computation time and is not suitable (with current computer capability) for solving large microstrip array problems. The FDTD method [40–42] uses Yee's algorithm [43] to discretize Maxwell's equation in 3D space and in time. The volume space of interest is discretized into many cubic cells and the E- and H-fields are then solved through Maxwell equations with given boundary conditions from cell to adjacent cells. This is illustrated briefly in Maxwell's curl equations:

$$\mu \cdot \frac{\partial \mathbf{H}}{\partial t} = -\nabla \times \mathbf{E} \tag{4.8}$$

$$\varepsilon \cdot \frac{\partial \mathbf{E}}{\partial t} = \nabla \times \mathbf{H} \tag{4.9}$$

With time and space discretized, the E- and H-fields are interlaced within the spatial 3D grid. For example, Eq. (4.9) can be discretized for the x-directed E-field:

$$E_x^{n+1}(i, j, k) = E_x^n(i, j, k) + \frac{\Delta t}{\varepsilon}\left(\frac{H_z^{n+1/2}(i, j+1, k) - H_z^{n-1/2}(i, j, k)}{\Delta y}\right) - \frac{\Delta t}{\varepsilon}\left(\frac{H_y^{n+1/2}(i, j, k+1) - H_y^{n-1/2}(i, j, k)}{\Delta z}\right) \tag{4.10}$$

where Δx, Δy, and Δz are the space steps in the x-, y-, and z-directions, and Δt is the time step. The same discretization can be carried out for Eq. (4.8).

Now Maxwell's equations have been replaced by a set of computer recognizable finite-difference equations, which can be solved sequentially from cube to cube once the known boundary conditions are applied. Certainly, this cube-to-cube solver cannot continue indefinitely outside the volume of interest and must be terminated. However, the fields will bounce back from any terminating boundary (which does not happen in reality) and disturb the correct solution. The solution is to use the electromagnetic absorbing boundaries to be set up outside the areas of interest and to absorb all outgoing fields. One significant advantage of the FDTD method is that, by discretizing time, one is able to see on a computer screen how the field is actually traveling and radiating in time sequence in a complicated antenna/circuit configuration.

4.3.1.5 Finite-Element Method (FEM) This method is also a three-dimensional solver that can best be described by a set of implementation steps [44]. First, one should define the electromagnetic boundary-value problem by an appropriate partial differential equation (PDE). Second, one obtains a variational formulation [45] for the PDE in terms of an energy-related functional or weighted residual expressions [46]. Third, one subdivides the field regions into discrete subregions (finite elements), such as triangles and quadrilaterals. Fourth, one chooses a trial or approximate solution (polynomial) defined in terms of nodal values (boundary points between elements) of the solution yet to be determined for each element. Fifth, one minimizes the functional (set function derivative to zero) with respect to the nodal value potentials. Finally, the resulting set of algebraic equations is solved and the required field problem solution is obtained. The primary difference between the FEM and the FDTD method is that the FDTD solves the problem from cell to cell with cell size serving as the approximating potential, while the FEM uses an approximate solution for each entire element. Thus, to achieve accuracy, the FDTD method cell size must be small and generally uniform, while the FEM element size can be large or small depending on the geometry or variation of the field. Both methods are computationally time consuming for electrically large structures. The FDTD method spends less computing time on each cell but with more cells, while the FEM spends more time on each element but with fewer elements. The implementation of the FEM is more complicated when compared to the FDTD method. The FDTD method is more straight forward, while the FEM requires a finer analytical development of the formulation before implementation, a deeper knowledge of linear algebra methods, and a more involved preprocessing procedure. Although the FEM is more complex, it is more versatile and flexible in modeling complex geometries. It yields more stable and accurate solutions and can handle nonhomogeneous materials. A very popular commercial

software, High-Frequency System Simulator (HFSS), uses the FEM to solve many electrically small but complex structures. Microstrip antennas or arrays with a small number of elements can certainly be solved by the FEM or FDTD method, but often the 2D moment method solution will suffice with faster computing speed.

4.3.2 Design Methodology

The previous section presented different techniques to analyze the microstrip antenna. To ease the design process, these different analysis techniques have been developed into several user-friendly computer-aided design (CAD) tools by several institutions. However, an analysis technique or a CAD tool, by itself, cannot generate an antenna design. It can only analyze a design and provide calculated performance results for a design. The basic and initial antenna design has to originate from human experience, knowledge, and innovation, even though an optimum and accurate design often cannot be achieved without an analysis tool. Figure 4.7 depicts a typical microstrip antenna development process. The block labeled "Computer Analysis Software" represents the central processing unit into which a human must enter the proper design data to initiate the design process. The block labeled "Antenna Design Techniques" represents the knowledge for generating a set of preliminary input design data, which is the main subject of this section. It includes techniques to design patch elements, array configurations, and power division transmission lines, which are separately discussed next. The first step in designing a microstrip array should be the element design.

4.3.2.1 Patch Element Design Patch elements come in various shapes, such as rectangular, square, circular, annular ring, triangular, pentagonal, and square or circular with perturbed truncations. These different shapes can often be used to meet various challenging requirements. For example, the rectangular patch, used for linearly polarized applications, can achieve slightly wider bandwidth than the square or circular patch. However, the square or circular patch, unlike the rectangular patch, can be excited orthogonally by two feeds to achieve circular polarization. In addition, the circular patch can be designed to excite higher order modes for generating different-shaped patterns

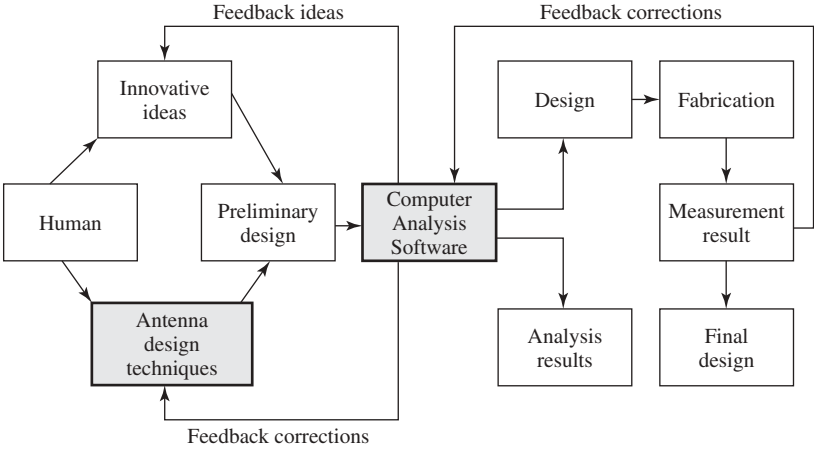


Figure 4.7 Microstrip antenna development procedures.

[47, 48]. One disadvantage of the square and circular patches when compared to the rectangular patch is that they are more susceptible to cross-polarization excitation when used as linearly polarized elements. This is because, due to their identical two orthogonal dimensions, it is easier to excite orthogonal resonance (cross-polarization) from spurious coupling or feed asymmetry. The pentagonal patch, as well as the square or circular patch with a small perturbation, can be used to generate circular polarization with only a single feed [11], which is often a desirable feature when simplicity and low insertion loss are required.

It should be noted that all these patch shapes can be accurately analyzed and designed by the full-wave moment method discussed in Section 4.3.1.3. However, designing a patch using the moment method or any other rigorous technique requires a priori knowledge of the approximate size of the patch so that appropriate dimensions, rather than random numbers, can be input to the analysis computer code. With a few iterations of the computer code, the designer should be able to determine the precise dimensions of the microstrip antenna. Once the dimensions are known, other parameters (e.g., input impedance, bandwidth, radiation patterns) can be accurately computed by the full-wave moment method. The above-mentioned a priori knowledge of the approximate patch size can be acquired through experience or derived by simple closed-form equations if available. Fortunately, the two most popular and often used patch shapes, rectangular (or square) and circular, do have simple closed-form equations available. These equations, in predicting the resonant frequency, substrate thickness, and dielectric constant, can generally achieve an accuracy of within 2%. For the fundamental-mode rectangular patch, the simple equation [10, 30] is given by

$$f = \frac{c}{2(L+h)\sqrt{\varepsilon_e}} \tag{4.11}$$

where

$$\varepsilon_e = \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r - 1}{2}\left(1 + \frac{12h}{w}\right)^{-1/2} \tag{4.12}$$

f is the resonant frequency, c is the speed of light, L is the patch resonant length, h is the substrate height, ε_r is the relative dielectric constant of the substrate, and w is the patch nonresonant width.

For the circular patch with TM_{mn} mode, the simple design equation is given by [10, 49]

$$f = \frac{\chi_{mn} c}{2\pi a_e \sqrt{\varepsilon_e}} \tag{4.13}$$

where

$$a_e = a\left\{1 + \frac{2h}{\pi a \varepsilon_r}\left[\ln\left(\frac{\pi a}{2h}\right) + 1.7726\right]\right\}^{1/2} \tag{4.14}$$

f, c, h, and ε_r are as defined for the rectangular patch design equation, a is the patch's physical radius, X_{mn} is the mth zero of the derivative of Bessel's function of order n, n represents the angular mode number, and m is the radial mode number.

There is no significant difference in performance between a fundamental-mode rectangular patch and a fundamental-mode circular patch. A circular patch does have the

advantage of offering higher order mode performance with different diameters and differently shaped radiation patterns [47, 48]. These patterns can be either linearly or circularly polarized, depending on the configuration of the feed excitations.

4.3.2.2 Array Configuration Design Before carrying out a detailed design, it is critically important to lay out the most suitable array configuration for a particular application. Array configuration variables include series feed or parallel feed, single layer versus multiple layers, substrate thickness, dielectric constant, array size, patch element shape and element spacing. The selection of the proper configuration depends on many factors, such as the required antenna gain, bandwidth, insertion loss, beam angle, grating/sidelobe level, polarization, and power-handling capability. Several important microstrip array configurations that often challenge the skills of antenna designers are presented next.

Series Feed In a series feed configuration [11, 50], multiple elements are arranged linearly and fed serially by a single transmission line. Multiples of these linear arrays can then be connected together serially or in parallel to form a two-dimensional planar array. Figure 4.8 illustrates two different configurations of the series feed method. The in-line feed [51] has the transmission line serially connected to two ports of each patch and is sometimes called the two-port series feed. The out-of-line feed [29] has the line connected to one port of each patch and is thus called a one-port series feed. The in-line feed array occupies the smallest real estate with the lowest insertion loss but generally has the least polarization control and the narrowest bandwidth. The in-line feed as shown in Figure 4.8 is generally more suitable for generating linear polarization than circular polarization. It has the narrowest bandwidth because the line goes through the patches, and thus the phase between adjacent elements is not only a function of line length but also of the patches' input impedances. Since the patches are amplitude weighted with different input impedances, the phases will be different for different elements and will change more drastically as frequency changes due to the narrowband characteristic of the patches.

The series feed can also be classified into two other configurations: resonant and traveling wave [11, 50]. In a resonant array, the impedances at the junctions of the transmission lines and patch elements are not matched. The elements are spaced multiple integrals of one wavelength apart so that the multiply bounced waves, caused by mismatches, will radiate into space in phase coherence in the broadside direction. Because

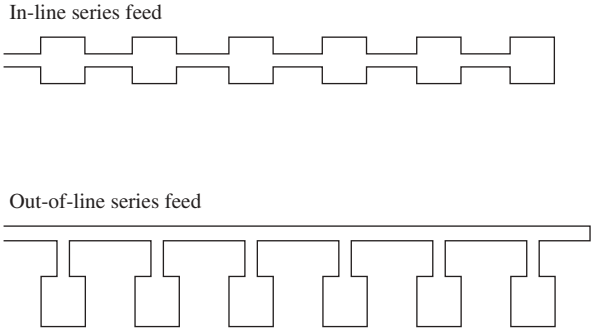


Figure 4.8 Series-fed microstrip arrays.

of this single- or multiple-wavelength element spacing, the beam of the resonant array is always pointed broadside. For the same reason, the bandwidth of a resonant array is very narrow, generally less than 1%. With a slight change in frequency, the one-wavelength spacing no longer exists, thereby causing the multiply bounced waves not to radiate coherently but, instead, to travel back to the input port as mismatched energy. Both the in-line and out-of-line feed arrays can be designed to be of the resonant type.

For the traveling-wave array type, the impedances of the transmission lines and the patches are generally all matched, and the element spacing can be one wavelength for broadside radiation, or less than one wavelength for off-broadside radiation. Because the energy travels toward the end of the array without multiple reflections, there is generally a small amount of energy remaining after the last element. This remaining energy can be either absorbed by a matched load or reflected back to be reradiated in phase for broadside radiation [31]. The array can also be designed such that the last element radiates all of the remaining energy [31]. The traveling-wave array has a wider impedance bandwidth, but its main beam will change in direction as frequency changes. A general rule of thumb for the frequency-scanned beam of a traveling-wave array is one degree of beam scan per 1% of frequency change. For an instantaneous wideband signal, such as a pulsed system, a beam broadening effect will occur. Both the in-line and out-of-line series-fed arrays of Figure 4.8 can be designed as the traveling-wave type. There are also other forms of series-fed microstrip arrays: chain, comb line, rampart line, Franklin, and coupled dipole [11, 50]. These arrays operate similarly to the arrays shown in Figure 4.8, except that they use microstrip radiators with different radiating mechanisms.

For a series-fed array, regardless of whether it is resonant type or traveling-wave type, it can be fed either from the end or from the center of the array. The end-fed array will encounter beam squint as frequency is changed and thus is sometimes utilized as a frequency-scanned array. For a center-fed array, the beam squint can be avoided as frequency deviates from its designed center frequency; however, it will suffer from a quick gain drop due to the beam split of the two half-beams.

Parallel Feed The parallel feed, also called the corporate feed [52], is illustrated in Figure 4.9 where the patch elements are fed in parallel by the power division transmission lines. The transmission line divides into two branches and each branch divides again until it reaches the patch elements. In a broadside-radiating array, all the parallel division lines have the same length. In a series-fed array, the insertion loss is generally less than that of a parallel-fed array because most of the insertion loss occurs in the transmission line at

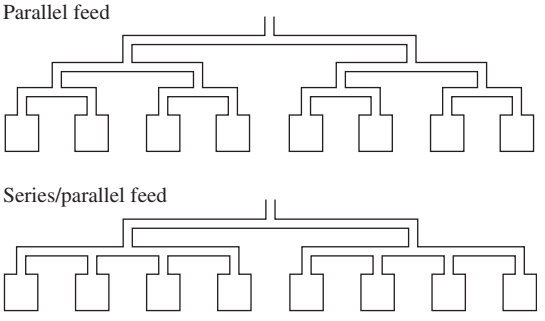


Figure 4.9 Configurations of parallel feed and hybrid parallel/series feed microstrip arrays.

the first few elements of the series-fed array and very little power remains toward the end of the array. Most of the power has already been radiated by the time the end elements are reached. Figures 4.10 and 4.11 are curves [53] for losses of parallel-fed arrays versus substrate height for dielectric constants of 1.06 and 2.32, respectively. A relatively small array with 16 elements and a large array with 256 elements are both given in these curves. It can be seen that the loss of a parallel-fed microstrip array can easily reach several

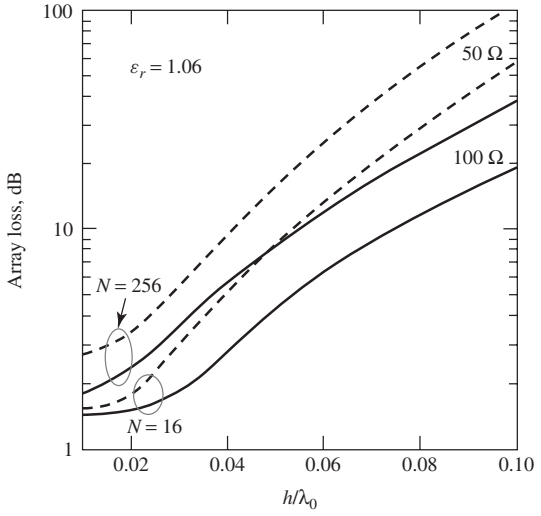


Figure 4.10 Array loss versus substrate height for parallel-fed array with substrate dielectric constant of 1.06. 16 and 256 elements as well as for line impedances of 50 ohms and 100 ohms. (From Ref. 53 with permission from IEE).

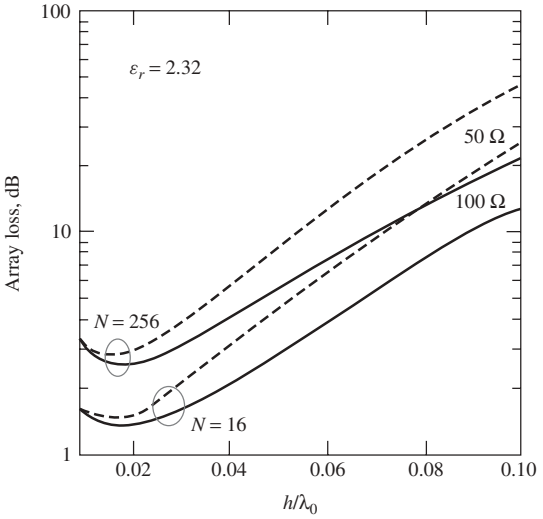


Figure 4.11 Array loss versus substrate height for parallel-fed array with substrate dielectric constant of 2.32. Plots are shown for 16 and 256 elements as well as for line impedances of 50 ohms and 100 ohms. (From Ref. 53, with permission from IEE).

decibels. Despite its relatively higher insertion loss, the parallel-fed array does have one significant advantage over the series-fed array, which is its wideband performance. Since all elements in a parallel-fed array could be fed by equal-length transmission lines, when the frequency changes, the relative phases between all elements will remain the same and thus no beam squint will occur. The bandwidth of a parallel-fed microstrip array is limited by two factors: the bandwidth of the patch element and the impedance matching circuit of the power dividing transmission lines, such as the quarter-wave transformer. A series-fed array can achieve a bandwidth on the order of 1% or less, while a parallel-fed array can achieve a bandwidth of 15% or more, depending on the design. Another advantage of the parallel-fed array is the relative ease with which both amplitude and phase for each element can be designed independently, while in a series-fed array one element's change will generally impact all other elements.

Hybrid Series/Parallel Feed An example of a hybrid series/parallel-fed array is depicted in Figure 4.9, where a combination of series and parallel feed lines is used. In a hybrid array [31], the smaller series-fed subarray has a broader beamwidth, which will suffer only a small gain degradation due to beam squint with frequency change. Hence a hybrid array will achieve a wider bandwidth than a purely series-fed array having the same aperture size. Of course, because of its partial parallel feed, the insertion loss of a hybrid array is higher than that of a purely series-fed array. This hybrid technique gives the designer an opportunity to make design trade-offs between bandwidth and insertion loss.

Regardless of whether the array is parallel or series fed, two recently developed arraying techniques can be employed to significantly improve the array's performance. The first is to reduce cross-polarization radiation in a planar array by oppositely exciting adjacent rows or columns of elements in phase and in orientation [31], as shown in Figure 4.12a. Another technique is shown in Figure 4.12b for a circularly polarized array, in which every adjacent four elements placed in a rectangular lattice can be sequentially arranged in both phase and orientation to achieve good circular polarization over a wide bandwidth [22, 23].

Single-Layer or Multilayer Design A microstrip array can be designed in either a single-layer or multilayer configuration. The factors that determine this choice are complexity and cost, sidelobe/cross-polarization level, number of discrete components,

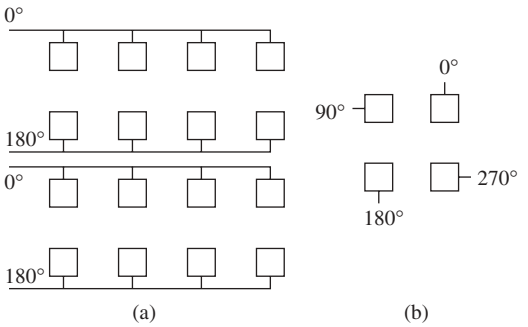


Figure 4.12 (a) Microstrip array with rows excited by opposite phases and orientations and (b) sequentially arranged four-element subarray for CP application.

polarization diversity, bandwidth, and so on. When the given electrical requirements are relaxed, a single-layer design will generally suffice. If all transmission lines and patch elements are etched on the same layer, it will have a low manufacturing cost. However, when extremely low sidelobe or cross-polarization radiation (e.g., less than -30 dB) is required, the double-layer design seems to be the better choice. With all transmission-lines etched on the second layer behind the radiating patch layer, the ground plane in the middle will shield most of the leakage radiation of the lines from the patch radiation. This leakage radiation becomes more pronounced when discrete components, such as MMIC T/R modules and phase shifters, are placed in the transmission line circuits. Thus it is more desirable to place all discrete components behind the radiating layer in a multilayer configuration. When dual-linear or dual-circular polarization is required with high polarization isolation, it is often more desirable to design the feed circuits of the two polarizations on two separate layers, as shown in Figure 4.13. When a radiating patch having a thick substrate is used to achieve a wider bandwidth, it is best to design the transmission lines on a separate layer because the lines may become too wide to be practical if designed on the same thick layer as the radiating patches. In other cases, when an extremely wide bandwidth requirement can be met only by using multiple stacked patches [19], the multilayer design becomes the obvious choice. With the advancement of the aperture-coupling technique that allows the transmission line to feed the patch, the multilayer design becomes much more feasible than those using many feed-through pins.

Other Array Configurations When designing a microstrip array, various antenna parameters, such as substrate thickness, dielectric constant, and element spacing, can all play important roles in determining an array's performance. Substrate thickness determines bandwidth, as well as the antenna's power-handling capability [29]. The thicker the substrate, the more power it can handle. For ground applications, a thicker microstrip antenna ($>0.05\lambda_0$ thick) can generally handle several hundred to a few thousand watts of peak power. For space applications, due to the effect of multipacting breakdown [30],

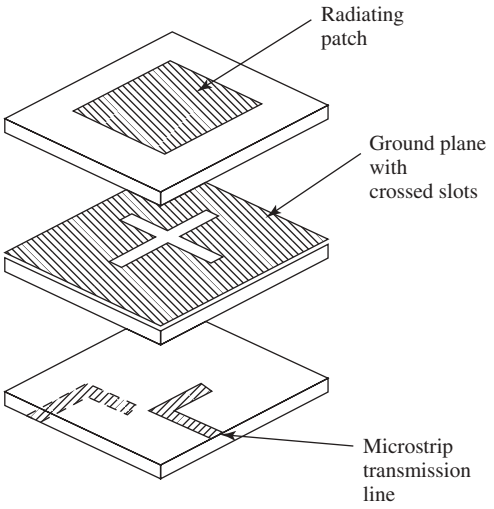


Figure 4.13 Multilayer dual-polarized microstrip patch element.

only tens of watts can be handled. The dielectric constant of the substrate material also affects the bandwidth: the higher the dielectric constant, the narrower the bandwidth. Because of the loading effect, a higher dielectric constant reduces the patch resonant size and hence increases the element beamwidth. A wider element beamwidth is desirable for a large-angle-scanning phased array. Another important array design parameter is element spacing. It is often desirable to design a microstrip array with larger element spacing so that more real estate can be made available for transmission lines and discrete components. However, to avoid the formation of high grating lobes, element spacing is limited to less than $1\lambda_0$ for broadside beam design and less than $0.6\lambda_0$ for a wide-angle scanned beam. In designing a wide-angle scanned microstrip phased array, substrate thickness, dielectric constant, and element spacing are all important parameters that need to be considered for reducing mutual coupling effects and avoiding scan blindness [54].

4.3.2.3 Power Division Transmission-Line Design

One of the principal shortcomings of a microstrip array with a coplanar feed network is its relatively large insertion loss, especially when the array is electrically large or when it is operating at a higher frequency. Most of the losses occur in the power division transmission line's dielectric substrate at microwave frequencies. At millimeter-wave frequencies, the loss in the copper lines becomes significant. It is thus crucially important to minimize insertion loss when designing the power division transmission lines. In order to minimize insertion loss, the following principles should be observed: the impedances of the power division lines should be matched throughout the circuit; low loss material should be used for the substrate; at higher frequencies, the roughness of the metal surfaces that face the substrate should be minimized; and the array configuration should be designed to minimize line length (as described in Section 4.3.2.1). A detailed discussion of most impedance-matching techniques for microstrip power division circuits is given in Ref. 55.

4.3.3 CAD Tools

In Section 4.3.2 on design methodology, a typical microstrip antenna development process is depicted in a block diagram (see Figure 4.7), where it indicates the need for computer-aided design (CAD) software. Although, as indicated, all CAD tools available today can only provide analysis and not a design, they do assist significantly in achieving the final design. For example, an engineer generates an initial design and then inputs the design dimensions and configuration into a CAD tool to calculate a set of performance results, such as input return loss, and radiation patterns. Generally, the initial results will not meet the given requirements, particularly, for a complicated design. The engineer, with his/her experience and knowledge, will perform corrections on the design and then input to the CAD tool again as indicated in Figure 4.7. This iterative process may take several times until satisfactory results are achieved. In the old days when CAD tools were not available, the engineer could only perform hardware verification of his/her design, which could take many iterations. This hardware verification step certainly took a significantly longer time, with higher cost than computer simulation. For a large array, the cost of iterative hardware verification soars with array size and complexity. Academic researchers have been prolific in generating analytic and numerical solutions for a wide variety of microstrip antennas and arrays, often with a high degree of accuracy and efficiency. But this area of work is generally performed primarily for graduate student theses or publications, and the software is seldom completely written, validated, or documented

for other users. Researchers in industry may be more pragmatic when developing comparable solutions for a specific antenna geometry, but such software is often considered proprietary.

It is clear that the demand/supply for CAD tools is unavoidable. The first commercial CAD tool for microstrip antennas became available about 15 years ago and, in the past decade, the number of commercial tools has mushroomed with more than ten available in the world. Table 4.1 lists some commercial software packages that can be used for microstrip antenna analysis and design.

Among these CAD tools, Ensemble and IE3D, which use the full-wave moment method, are the most popular ones. These two PC-based software packages were on the market much earlier than the other ones for microstrip antenna application. Through upgrades and modifications, they became more efficient, less prone to errors, and with more capabilities. Designs with multilayer, conductive via connections, finite ground plane and soon can all be accurately analyzed. With a 1-GB RAM capability, a current PC, by using either Ensemble or IE3D, can handle a microstrip array with approximately 30 elements and some microstrip power division lines. Some of the other software packages, which use finite difference time-domain (FDTD) or finite-element (FE) methods, take a three-dimensional approach by modeling the entire antenna space, including dielectric, metal components, and some surrounding volume. This approach allows a high degree of versatility for treating arbitrary geometries, including inhomogeneous dielectrics and irregularly shaped structures, but the price paid is computer time. With a current PC, only a few patch elements can be calculated. Regardless of the method used, future advancement in CAD tools is vested in two areas: PCs with high capacity and faster computation time and more efficient mathematical algorithms. With these advancements, large microstrip arrays can be more effectively analyzed and designed.

One important conclusion should be made here for all CAD users: although CAD software can be an invaluable analysis/design tool, it is not a substitute for design experience or a thorough understanding of the principles of operation of microstrip antennas and arrays. While microstrip antenna design is based on solid science, it also retains a strong component of understanding the physical mechanisms and a creative problem-solving approach that can only come from experience. It also can be concluded that, at least for the near future, CAD tools will continue to aid, rather than actually replace, the experienced designers.

TABLE 4.1 Some Commercially Available Microstrip Antenna CAD Tools

Software Name	Theoretical Model	Company
Ensemble (Designer)	Moment method	Ansoft
IE3D	Moment method	Zeland
Momentum	Moment method	HP
EM	Moment method	Sonnet
PiCasso	Moment method/genetic	EMAG
FEKO	Moment method	EMSS
PCAAD	Cavity model	Antenna Design Associates, Inc
Micropatch	Segmentation	Microstrip Designs, Inc.
Microwave Studio (MAFIA)	FDTD	CST
Fidelity	FDTD	Zeland
HFSS	Finite element	Ansoft

4.4 FEED/EXCITATION METHODS

A microstrip patch radiator can be fed or excited to radiate by many techniques; several common ones are listed and briefly discussed next.

4.4.1 Coax Probe Feed

A microstrip patch as shown in Figure 4.1 can be fed by a 50-ohm coax probe from behind the ground plane, where the flange of the coax probe (outer conductor) is soldered to the ground plane. The center conductor pin penetrates through the substrate and the patch and is then soldered to the top of the patch. The location of the probe should be at a 50-ohm point of the patch to achieve impedance matching. There are various types of coax probes for different frequency ranges. Type N, TNC, or BNC can be used for VHF, UHF, or low microwave frequencies. OSM or OSSM can be used throughout microwave frequencies. OSSM, OS-50, or K-connector should be used for the millimeter-wave frequency range.

4.4.2 Coax Probe with Capacitive Feed

For wider bandwidth (5–15%) applications, thicker substrate is generally used. If a regular coax probe were used, a larger inductance would be introduced, which results in impedance mismatch. In other words, the electrical field confined in the small cylindrical space of the coax cannot suddenly transition into the large spacing of the patch. To cancel the inductance occurring at the feed, capacitive reactance must be introduced. One method is to use a capacitive disk [56] as shown in Figure 4.14 where the patch is not physically connected to the probe. Another method is to use a "tear-drop" shaped [57] or a cylindrical shaped probe [58] as illustrated in Figure 4.15. With this method the probe is soldered to the patch, where mechanical rigidity may be offered for some applications.

4.4.3 Microstrip-Line Feed

As illustrated in Figure 4.1, a microstrip patch can be connected directly to a microstrip transmission line. At the edge of a patch, impedance is generally much higher than

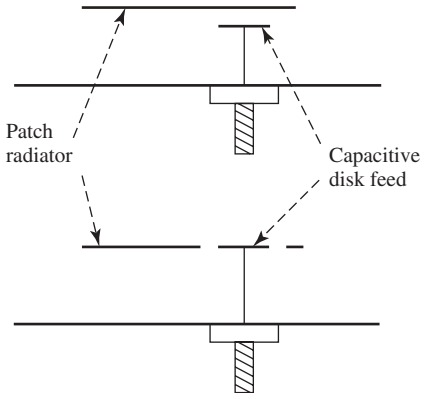


Figure 4.14 Two different capacitive feed methods for relatively thick substrates.

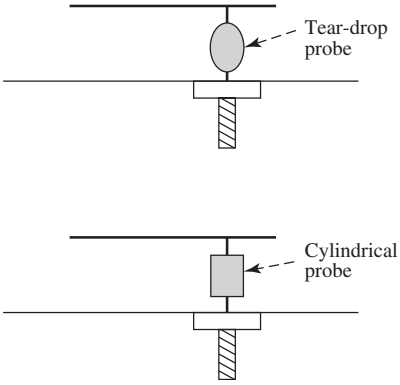


Figure 4.15 Tear-drop and cylindrical-shaped feed probes for relatively thick substrates.

50 ohms (e.g., 200 ohms). To avoid impedance mismatch, sections of quarter-wavelength-long impedance transformers [55] can be used to transform a large input impedance to a 50-ohm line. With this feed approach, an array of patch elements and their microstrip power division lines can all be designed and chemically etched on the same substrate with relatively lower fabrication cost per element. However, the leakage radiation of the transmission lines, in some cases, may be large enough to raise the sidelobe or cross-polarization levels of the array radiation.

4.4.4 Proximity-Coupled Microstrip Line Feed

An open-ended microstrip line can be used to feed a patch radiator through proximity coupling. For example, the open end of a 100-ohm line can be placed underneath the patch at its 100-ohm location as shown in Figure 4.16. The open-ended microstrip line can also be placed in parallel and very close to the edge of a patch, as shown in Figure 4.17, to achieve excitation through fringe-field coupling [59]. Both these methods will avoid any soldering connection, which in some cases could achieve better mechanical reliability.

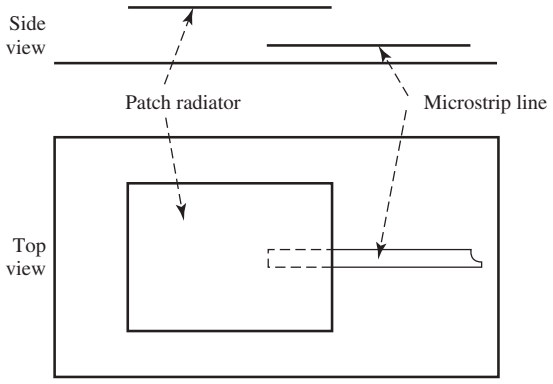


Figure 4.16 Non contact proximity feed from underneath the patch.

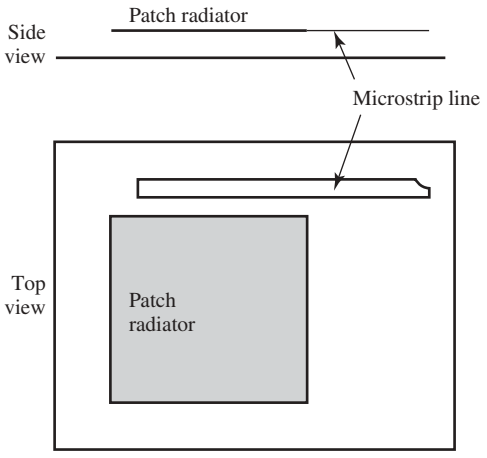


Figure 4.17 Non contact proximity feed from edge of patch.

4.4.5 Aperture-Coupled Feed

An open-ended microstrip line or stripline transmission line can be placed on one side of the ground plane to excite a patch radiator situated on the other side through an opening slot in the ground plane. This slot-coupling or aperture-coupling technique [12], as shown in Figure 4.18, can be used to avoid a soldering connection, as well as to avoid leakage radiation of the lines that interferes with the patch radiation. In addition, this feed method allows the patch to achieve wide bandwidth (>10%) with a thick substrate or extremely wide bandwidth (>30%) with stacked parasitic patches [60, 61]. The extra bandwidth achieved by this method when compared to the coax probe feed is generated by the coupling slot, which is also a resonator and a radiator. When two resonators (slot and patch) have slightly different sizes, a wider bandwidth is achieved. Another advantage of the noncontacting feeds (proximity-coupled and aperture-coupled) is the reduction

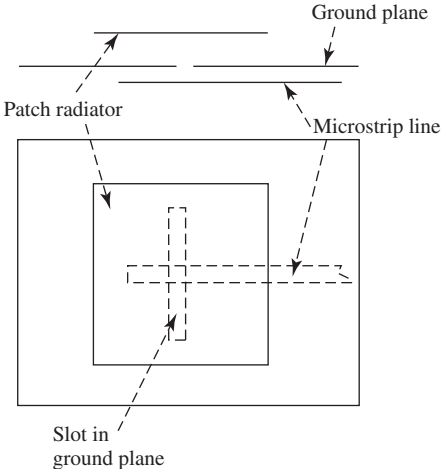


Figure 4.18 Patch fed by aperture-coupling slot.

of passive intermodulation distortion due to other harmonic frequencies created by nonlinear devices present in the circuit.

4.4.6 Coax Probe With a U-Slot Feed

A unique combination of a regular coax probe and a U-shaped slot on the patch has been developed to achieve very wide bandwidth (>30%). This U-slot feed [25, 26] as illustrated in Figure 4.19 is designed with a certain slot width to provide the needed capacitance for canceling the inductance introduced by the relatively thick substrate. The U-slot also provides two different resonant patch sizes to achieve the very wide bandwidth.

4.5 DUAL-POLARIZATION AND CIRCULAR-POLARIZATION TECHNIQUES

4.5.1 Dual Polarization

For either a square patch or a circular patch with conventional thickness and fundamental mode, the two orthogonal points on the patch (along the two cross center lines of a square patch) are generally isolated from each other. This can be explained by referring to Figure 4.4, where the field at the orthogonal region of the feed is always zero. Thus a second feed probe can be placed at the orthogonal region of the first probe without encountering significant field coupling. This, also explained in Section 4.3.1.2, is the basic reason why a single square or circular patch can be excited at its two orthogonal locations, as indicated in Figure 4.20, to achieve dual-linear polarization. Most of the excitation techniques presented in Section 4.4 with two orthogonal feeds can be used here to achieve dual polarization. The two orthogonal feeds do not need to be the same excitation technique. For example, one excitation could use a microstrip line feed, while the orthogonal excitation could use an aperture-coupling slot feed. One must be aware that, with a square or a circular patch, two different orthogonal feeds may cause the patch to resonate at two slightly different frequencies. In other words, one may have to use a slightly rectangular or elliptical patch to achieve the same resonant frequency for two different feed techniques.

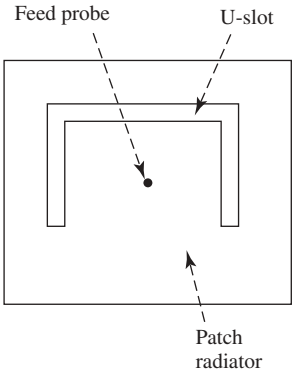


Figure 4.19 Patch fed by a U-slot for wide bandwidth.

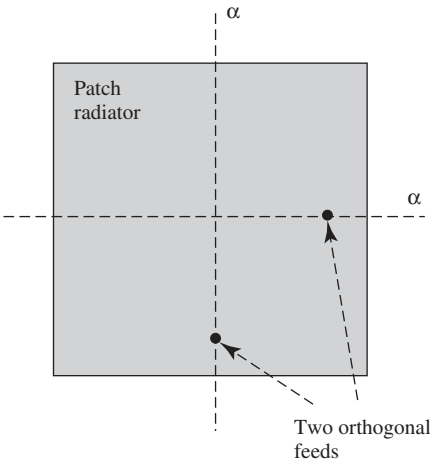


Figure 4.20 Patch with two orthogonal feeds for dual polarizations.

The above orthogonal excitation techniques for achieving dual polarization are suitable for a conventional patch with relatively small substrate thickness. For thick substrate, instead of two, four feed probes are needed with electrical phases arranged as 0°, 180°, 0°, 180°. Due to the presence of stronger higher order modes in a thick substrate, using only two orthogonal feeds will result in a large amount of mutual coupling. By using four feeds [62] as shown in Figure 4.21, one pair of oppositely located feeds with 0° and 180° phases will cause the higher order modes to cancel each other (see Figure 4.5b) while reinforcing the fundamental modes (see Figure 4.5a). In doing so, not only are the cross-polarization radiations canceled, but good isolation between the two pairs of feeds is also achieveds.

For a dual-polarization array application using a relatively thick substrate, higher order modes can be canceled by use a single feed per element. This can be achieved by having each pair of adjacent elements use oppositely located feeds with opposite phases as shown in Figure 4.22b (refer to Figures 4.5 and 4.6). It is apparent, as explained in the figure, that the cross-polarizations in the far-field distance will be canceled in both the

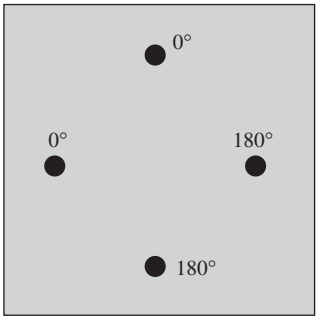


Figure 4.21 Four probes with 0°, 180°, 0°, 180° phase arrangement for dual polarization with relatively thick substrate.

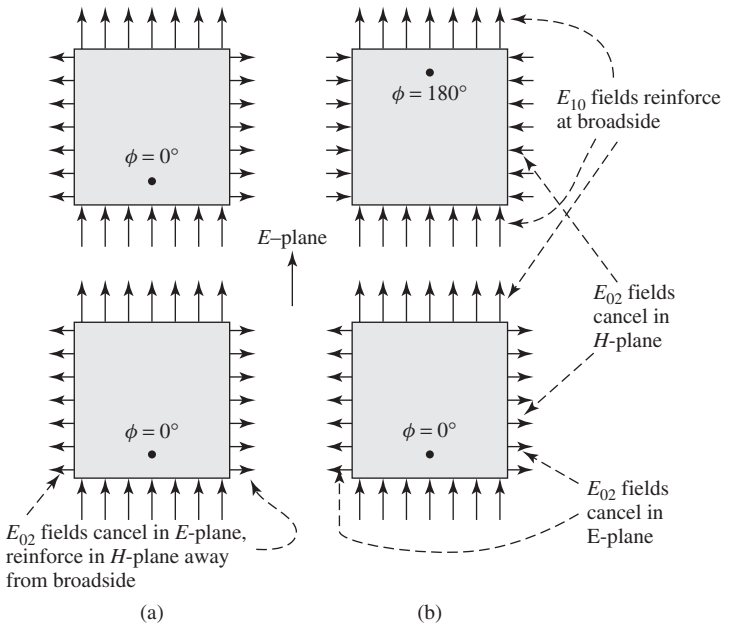


Figure 4.22 (a) Two-element subarray with conventional feed method and (b) two-element subarray feeds with 0° and 180° phases and opposite orientations to cancel cross-polarization radiation.

E- and H-planes, while the co polarization fields reinforce each other. To further illustrate the principle, a previously developed dual polarized 2×2 microstrip array [63] is presented in Figures 4.23 and 4.24. Both the V-port (vertical polarization) and the H-port (horizontal polarization) feed locations are offset from the centers of the feed transmission lines to achieve the required 180° phase differential. This 180° phase differential of each port is not only for the purpose of suppressing the cross-polarization as explained above; it is also for canceling the energy coupled in through the patches' cavities from

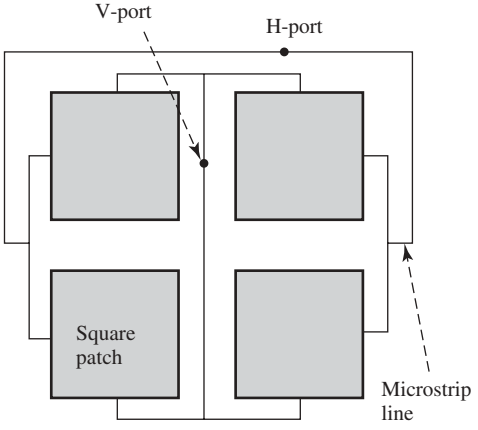


Figure 4.23 A dual-polarized 2×2 microstrip array.

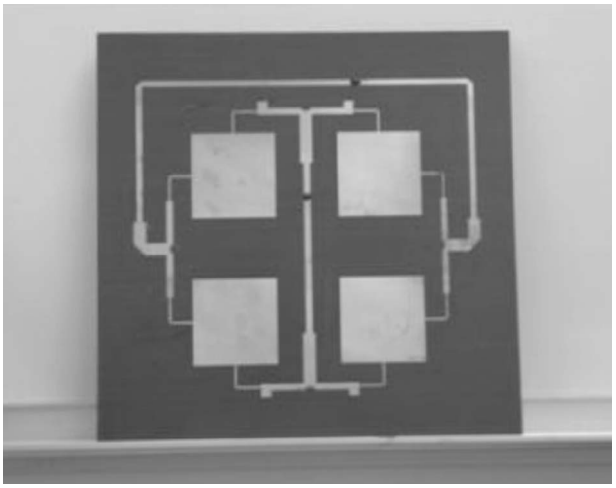


Figure 4.24 Photo of an L-band 2 × 2 dual-polarized microstrip array with low cross-polarization and high port isolation.

its orthogonal port. This cancellation does not happen in the patch or in space but occurs in the microstrip transmission line at the input port. It is this cancellation that provides the high isolation between the two input ports. For the antenna shown in Figure 4.24, the measured port isolation is below −40 dB across the bandwidth of the antenna. The worst cross-polarization level is −28 dB below the copolarization peak. For arrays larger than 2 × 2, the above principle has also been successfully applied by several researchers [31, 64].

4.5.2 Circular Polarization

Circular polarization (CP) from a microstrip antenna can generally be achieved by either a square or a circular patch with two orthogonal feeds having equal amplitudes and 90° phase differential. As illustrated in Figure 4.25, the square patch with its two orthogonal microstrip line feeds in this case will provide a right-hand circular polarization (RHCP). Similar to the dual-linear polarization case, a CP patch can be excited by various feed

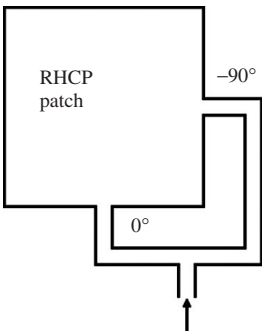


Figure 4.25 Circularly polarized patch with two orthogonal feeds.

techniques, as presented in Section 4.3, with two orthogonal excitations. A single CP patch with relatively thick substrate can also be excited by four feeds [65] with phases arranged as 0°, 90°, 180°, 270° to suppress cross-polarization level. A circular patch with diameter larger than a half-wavelength can be excited to radiate a higher order mode [47] with conical pattern shape (zero radiation in broadside direction). The higher the mode order, the larger the patch diameter required. In any case, four feeds arranged in special angular orientations and phases are needed to provide good CP radiation.

A CP patch can also be achieved by a single excitation with a small portion of the patch perturbed. For examples, CP can be generated by a square patch with its two diagonal corners truncated [8, 66] as shown in Figure 4.26, by a square or circular patch with a tilted center slot [8] as shown in Figure 4.27, or by a circular patch with two opposing edges indented with notches [67] as shown in Fig 4.28a or extended with pads [68] as shown in Fig 4.28b. A slight rectangular patch with a single feed located at its diagonal line [69], as shown in Figure 4.29, can also generate CP radiation. All these single-feed techniques are called the perturbation method. The small perturbation has to be just the right amount at the desired frequency to produce two orthogonal polarizations with the same amplitude but with a 90° phase differential. As a result, the CP bandwidths

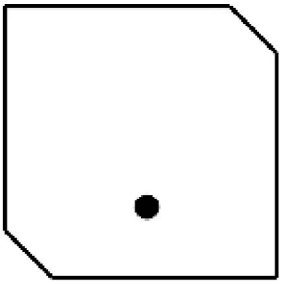


Figure 4.26 Circularly polarized patch with a single feed and truncated corners.

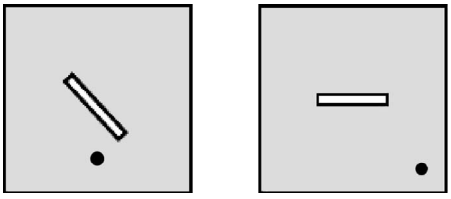


Figure 4.27 Circularly polarized patch with a single feed and a slot.

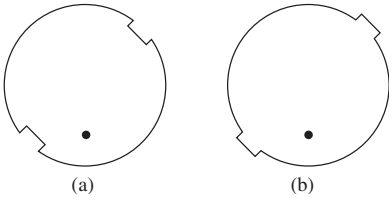


Figure 4.28 Circularly polarized circular patch with a single feed and (a) two indents and (b) two pads.

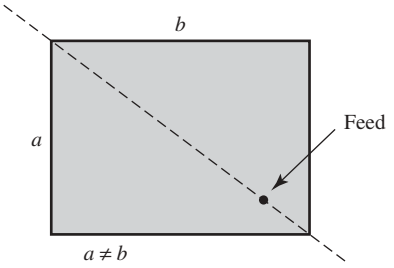


Figure 4.29 Circularly polarized patch (slightly rectangular) with a single feed.

of these single-feed methods, regardless of their impedance bandwidths, are extremely narrow (generally around 0.5%). However, these single-feed methods yield simplicity with reduced line loss.

In array application, CP performance, such as CP bandwidth and cross-polarization level, can be improved by sequentially arranging (orientations and phases) four neighboring CP elements as shown in Figure 4.30. Four single-fed CP elements can also be sequentially arranged [22] to improve the CP bandwidth as shown in Figure 4.31. This improvement of CP performance is based on the same reason as explained in Section 4.5.1 where the cross-polarization is canceled in a pair of oppositely located patches with opposite phases. In a large array, each of its four neighboring elements can even be linearly polarized but arranged sequentially [23], as shown in Figure 4.32, to provide CP radiation.

4.6 BROADBAND AND DUAL-BAND TECHNIQUES

4.6.1 Broadband Techniques

The microstrip antenna is basically a resonating cavity with open side walls. It is well known that a closed cavity with fixed dimensions indicates narrow bandwidth behavior. Thus the microstrip antenna also behaves as a narrow band device. Figure 4.33

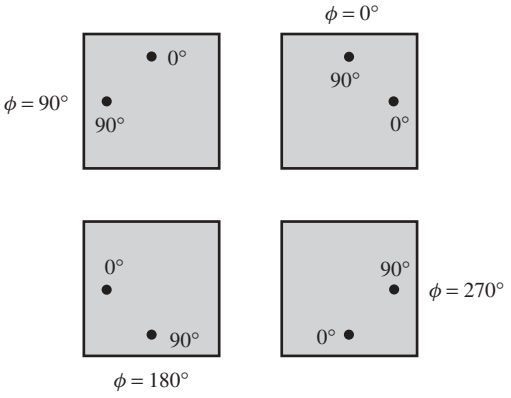


Figure 4.30 Circularly polarized 2 × 2 subarray with sequential rotation, Each element has two orthogonal feeds.

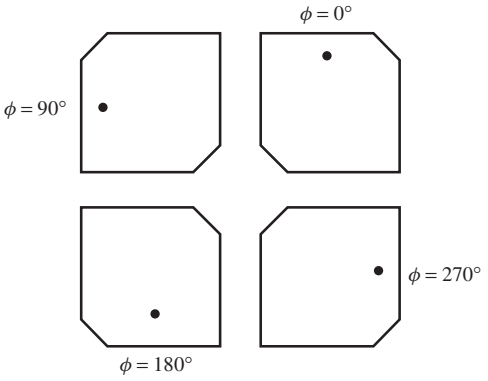


Figure 4.31 Circularly polarized 2×2 subarray with sequential rotation, Each element is a single-feed CP patch.

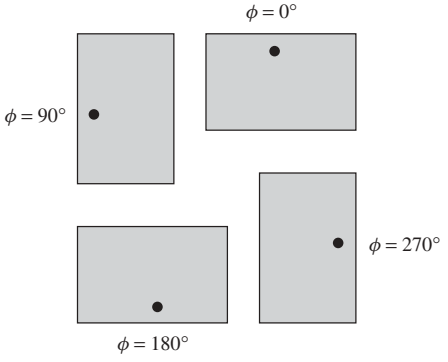


Figure 4.32 Circularly polarized 2×2 subarray with sequential rotation, Each element is a linearly polarized patch.

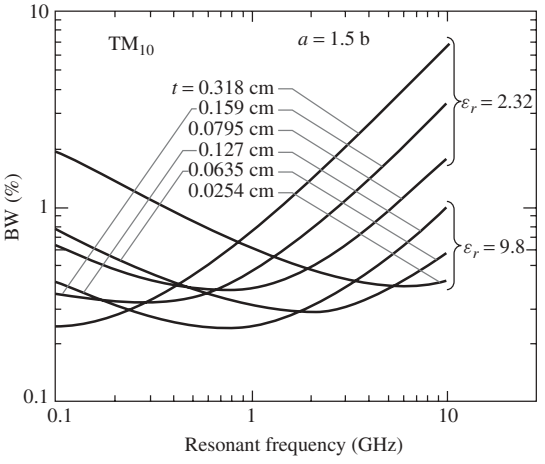


Figure 4.33 Bandwidths of rectangular microstrip patches with various substrate thicknesses and two dielectric constants. (From Ref. 11, with permission from Peter Peregrinus Ltd.)

is a plot of bandwidth versus operating frequency [11] for a typical rectangular patch with various thicknesses of the substrates and two different relative dielectric constants (2.32 and 9.8). This figure demonstrates that the bandwidth of a single patch with nominal substrate thickness has a bandwidth of less than 5%. However, due to the patch's open side walls, as the thickness of this cavity is increased, the bandwidth of this open cavity can be significantly increased. To illustrate this point, a rectangular patch is designed with x, y, z dimensions with x being the resonating dimension and z the cavity thickness. When the cavity thickness increases, as indicated in Figure 4.34 as well as in Figure 4.2 and 4.4, the height of the fringing field also increases. It is obvious from this figure that the degree of freedom for the resonant frequency to change is proportional to L_2—L_1, which is greater for the thicker substrate than for the shallow cavity. This bandwidth increase can also be explained from the antenna's quality factor *(Q)*. The Q of a rectangular patch [70] is inversely proportional to the cavity thickness *(h)*:

$$Q = \frac{c\sqrt{\varepsilon_e}}{4 f_r h} \tag{4.15}$$

and the bandwidth (BW) increases as Q reduces:

$$BW \approx \frac{f_r}{Q} \approx \frac{4 f_r^2 h}{c\sqrt{\varepsilon_e}} \tag{4.16}$$

From the above equation, one can conclude that a larger BW is possible for a microstrip antenna by choosing a thicker substrate and/or by lowering the value of the dielectric constant, ε_r. However, as the thickness is increased, the feed line or feed probe will encounter an impedance matching issue. Generally, a large reactance (inductance) is introduced by the feed. In the case of the microstrip line feed, an impedance matching circuit [21] can be used to balance out the large reactance. In the case of a feed probe, a capacitive feed can be used to cancel the excessive inductance as discussed in Section 4.4.2.

There is a limitation to how much the BW can be increased by increasing the substrate thickness. The maximum achievable BW is about 15% with the thickness equal to about

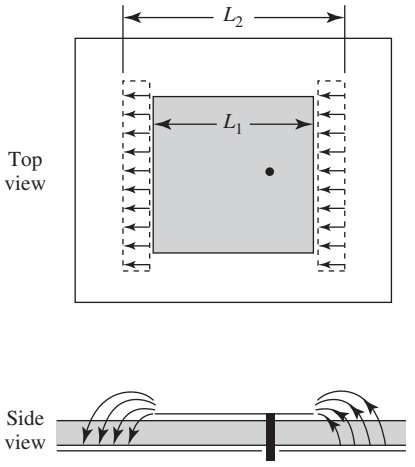


Figure 4.34 Primary fringing fields of a rectangular patch.

0.15 free-space wavelength. Upon further increase in the substrate thickness, the field will not cohere efficiently between the patch and the ground plane. One method to increase the BW beyond 15% is to introduce one or two more parasitic patches [71] on top of the patch being excited. A dual-stacked patch configuration is shown in Figure 4.35. Maximum BW of 25% has been achieved by this configuration. This configuration shows that the bottom patch is being driven by a coax probe; it is also possible to achieve similar performance by exciting the top patch by a coax probe that is sent through a clearance hole at the center of the bottom patch [72]. Another method to further increase the BW is to use a thick substrate with an aperture-coupling slot feed [12] as presented in Section 4.4.5. By combining the aperture-coupling slot method with a set of dual-stacked patches, as shown in Figure 4.36, a wide BW of 33% was reported [60, 61].

Two recent developments that achieved wide BW are worth mentioning here. One is to use a coax probe feed with a U-slot [25, 26] as shown in Figure 4.19. and discussed in Section 4.4.6. The other is to use a thick substrate augmented with an L-shaped probe [27, 28] as shown in Figure 4.37. Although the first method has difficulty in achieving dual or circular polarization, both methods achieved a BW of more than 30%.

To achieve wide BW by using thick substrate, one should be careful that, although the impedance bandwidth is widened, the radiation patterns may be distorted at the band edges. This is because the higher order modes may become noticeably stronger at the band edges. In this case for linear polarization, two feeds, instead of one, with

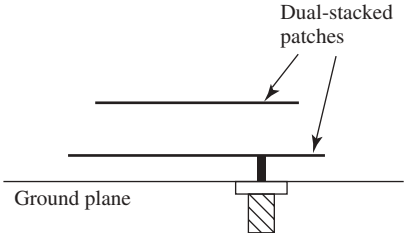


Figure 4.35 Probe-fed dual-stacked patches.

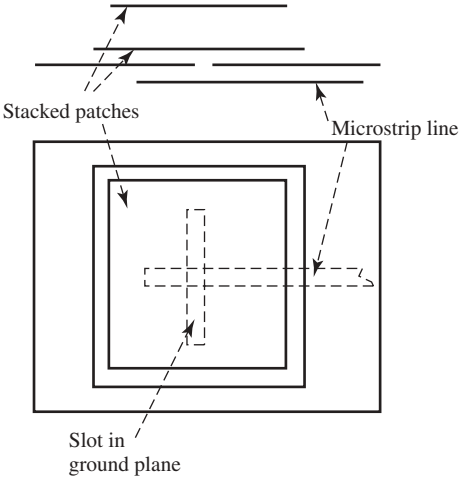


Figure 4.36 Aperture-coupled slot-fed dual-stacked patches.

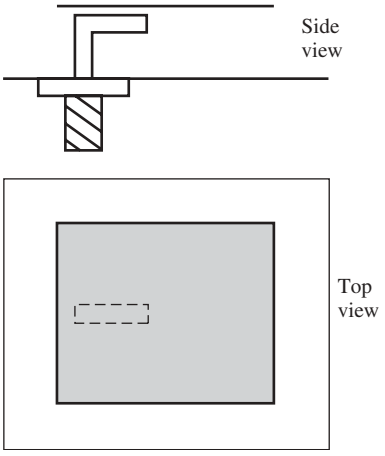


Figure 4.37 Thick patch fed by an L-shaped probe.

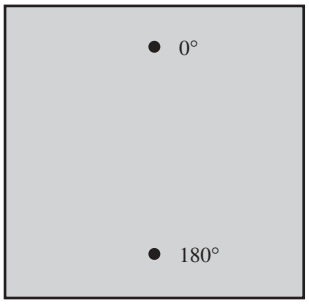


Figure 4.38 Two feeds with opposite locations and opposite phases to cancel higher order modes for wideband application.

opposite locations and opposite phases as shown in Figures 4.22 and 4.38 are needed to cancel some of the higher order modes [62]. For circular polarization, four feeds with 0°, 90°, 180°, 270° phases as shown in Figure 4.21 can be used to achieve nondistorted symmetrical patterns [52]. In one case [58], a single patch with thick air substrate and four capacitive feeds achieved good circular polarization and radiation patterns across a BW of 43%.

4.6.2 Dual-Band Techniques

In many applications, two or more small BWs are required over a large frequency range. A microstrip antenna can be specially designed to meet this requirement. Several approaches that can be used to achieve variously required dual bands are briefly discussed next.

4.6.2.1 Stacked Patches In Section 4.6.1, parasitically stacked dual patches are used to achieve broad BW. In those cases, broadbands, in essence, are achieved by having two resonant bands located close to each other. Here two parasitically coupled patches

are stacked together to achieve two separate, but relatively narrow, bands [15]. Because this stacked-patch approach relies completely on mutual coupling, there is a limit to the separation width of the dual bands. In other words, the sizes of the two patches cannot be grossly different and the separation distance between the two patches cannot be too large, otherwise mutual coupling will not be strong enough to achieve two resonances with equal magnitudes. The maximum band separation by this stacked-patch approach is about 20%.

4.6.2.2 Tunable Single Patch The previous dual-band technique requires two layers of patches, which in some array applications may be too costly to implement. Here a single-patch element can be specially tuned to have two separate narrowbands. There are several methods to achieve a dual-resonance single patch. One method is to add a quarter-wavelength-long stub [17] to the radiating edge of a patch, as shown in Figure 4.39. Due to the appropriate length of the stub, it introduces a second resonance to the patch. Another method is to add one or more shorting posts [16] to the patch, as shown in Figure 4.40. In doing so, an inductive reactance is loaded to the patch at the frequency away from the patch's resonance and causes the patch to cancel its capacitance and to form a second resonance. A voltage-controllable varactor diode can be inserted

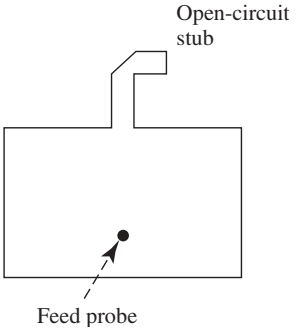


Figure 4.39 Open-circuit stub to achieve dual-band resonances.

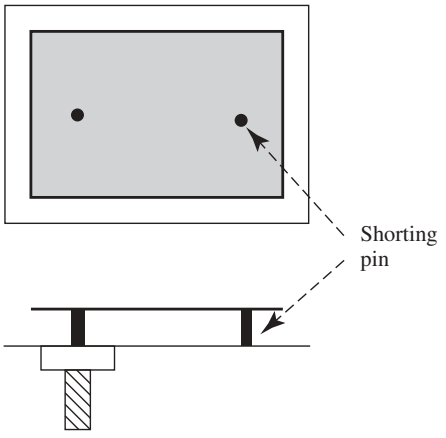


Figure 4.40 Shorting pin to achieve dual-band resonances.

into the shorting post to instantaneously change the two resonant frequencies [73]. In other words, by varying the varactor diode's voltage, the dual frequencies can be instantaneously varied over a small range.

4.6.2.3 Dual-Fed Single Patch For the above dual-resonant techniques, if one frequency is used for the transmit signal and the other is used for the receive signal, as in the case of a communication system, a diplexer or a circulator is generally needed to isolate the two signals. To eliminate the use of a diplexer or circulator, separate transmit and receive channels can be implemented with a single patch having two orthogonal feeds. As shown in Figure 4.41, a rectangular patch with two orthogonal feeds has its one dimension designed to resonate at the transmit frequency with one linear polarization, while the other dimension is designed for the receive frequency with an orthogonal linear polarization. An elliptical patch can also be used to achieve the same results. A shorting pin can be placed at the center of a patch to suppress some high order modes and thus achieve better isolation, particularly for the case of a relatively thick substrate. For a square or circular patch, dual resonance can also be achieved with two orthogonal feeds by introducing a slot at the center of the patch as shown in Figure 4.42. In the vertical direction of this figure, the electrical current on the patch has to travel a longer distance to reach the opposite edge than in the horizontal direction. This is the reason why this square patch can achieve two different resonant frequencies. All these dual-band techniques will yield two orthogonal polarized fields and thus have limited applications.

4.6.2.4 Widely Separated Dual Bands Often in a radar or communication system, two widely separated bands (e.g., L- and C-bands), are required to share the same aperture and thus to achieve savings in mass and real estate. There are several methods that

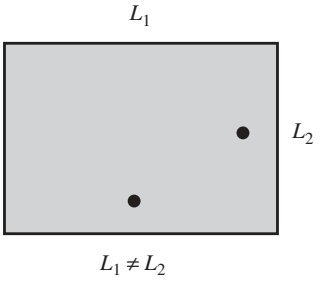


Figure 4.41 Rectangular patch having two orthogonal feeds with dual-frequency operation.

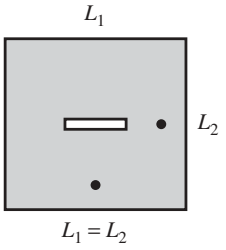


Figure 4.42 Square patch having two orthogonal feeds with dual-frequency operation.

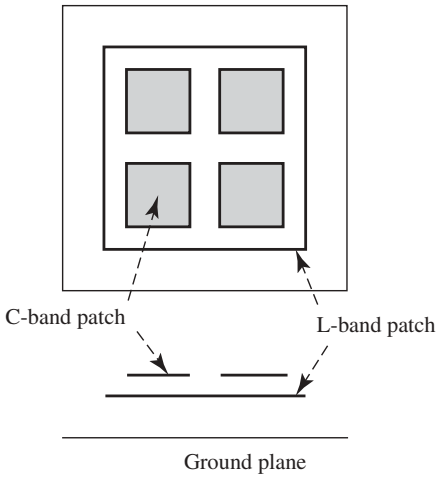


Figure 4.43 Dual-band with multilayer approach; higher frequency patches are situated on top of the lower frequency patch.

allow a microstrip antenna to achieve this requirement. One is to use a two-layer design as shown in Figure 4.43, where several C-band patches can be placed on top of a single L-band patch. Another method, shown in Figure 4.44, is to place a single L-band dichroic/meshed patch on top of several X-band patches [74, 75]. In this case, the top L-band layer acts as a frequency-selective surface to allow the X-band signal to transmit through, while the X-band patches act as part of the ground plane to the L-band patch. In a recent development, a microstrip reflectarray [76], having a layer of X-band annular ring elements placed on top of another layer of Ka-band annular rings, was tested with successful results.

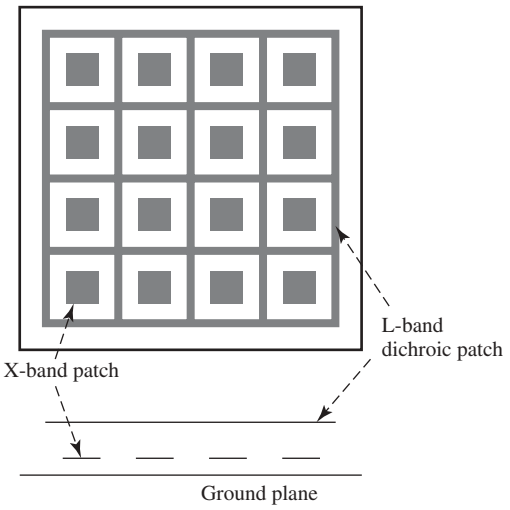


Figure 4.44 Dual-band multilayer approach; lower frequency dichroic patch is situated on top of the higher frequency patches.

4.7 ANTENNA MINIATURIZATION TECHNIQUES

Microstrip antennas are well known for their small size and light weight. However, in many applications, further reduction of the size and weight is desirable. For example, a L-band antenna at 1.5 GHz is required to fit into a small hand-held cellular phone and a UHF antenna at 400 MHz needs to be mounted onto a small size Mars rover vehicle. Both examples call for additional size reduction if a microstrip antenna is to be used. There are several basic miniaturization techniques that can further reduce the size of a microstrip antenna and they are briefly discussed next.

1. *Loading the Substrate with High-Dielectric-Constant Material.* The most popular technique in reducing the size of a microstrip antenna is to use a high-dielectric-constant (ε_r) material for its substrate. In doing so, the guided wavelength underneath the patch is reduced, and hence the resonating patch size is also reduced. The reduction ratio [10] is approximately related to the square root of ε_r. A conventional patch antenna, using an ε_r of 2 or 3, has a dimension of about $0.35\lambda_0$. With an ε_r of 10, the patch dimension reduces to about $0.2\lambda_0$. The antenna designer must use caution when using an ε_r much higher than 10. Generally, if the resonant dimension becomes smaller than $0.15\lambda_0$, the radiator starts to lose its radiation efficiency. Even though the antenna may achieve a very good impedance match with low input return loss, the input power may become stored reactive power instead of being radiated, resulting in low antenna gain. This size reduction limit is also true for all of the following techniques.

2. *Size Reduction by Using Slots.* To reduce the radiating patch size, slots of various configurations can be introduced either in the radiating patch or in the ground plane underneath the patch [77, 78]. In doing so, the current on the patch and the field underneath the patch will resonate from one edge of the patch and take the longer path around the slots to reach the opposite edge. This longer path, in essence, reduces the resonant frequency or the physical size of the antenna. Depending on the length of the slots, up to 20% size reduction can be achieved. As an example, a recently developed circular patch antenna [79] using an ε_r of 10 with four slots, shown in Figure 4.45, achieved a patch diameter of 11 cm ($0.15\lambda_0$) at the frequency of 400 MHz. Four slots, instead of two, are used to achieve either circular or dual polarization.

3. *Patch Folding Technique.* This technique is to fold [80] a single-layer patch antenna (including substrate and ground plane) to form a two-layer structure, thus significantly reducing the planar dimension. The configurations of folding a half-wave patch and a quarter-wave patch are shown in Figure 4.46. Although the patch is folded, the electrical field between the patch and the ground plane still travels the same resonant distance as in the unfolded patch. Thus both folded and unfolded patches will resonate at the same frequency. As an example, by folding a quarter-wave patch with a substrate ε_r of 10, an actual developed antenna achieved an overall size of 11.4 cm $\times$ 3.8 cm $\times$ 1.5 cm at the frequency of 400 MHz.

4. *Inverted-F and Shorting-Pin Methods.* The size reduction technique of using a planar inverted-F configuration [81] is shown in Figure 4.47. The key feature is to use a small shorting plate or pin as indicated in the figure. The width dimension (W) of the shorting plate is significantly smaller than L_1 ($\sim 0.2\ L_1$) and the dimensions of L_1 and L_2 are each on the order of $\frac{1}{8}\ \lambda_0$. Another technique, very similar to the inverted-F method, uses a circular patch with a shorting pin [82] as illustrated in Figure 4.48. An actual developed patch using this technique achieved a diameter of $0.1\lambda_0$ with an ε_r of

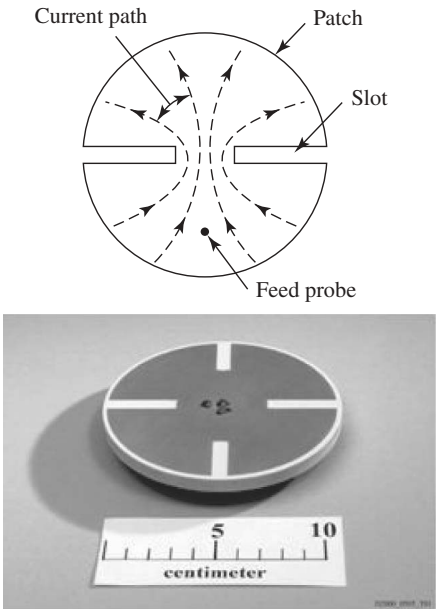


Figure 4.45 A miniaturized UHF circular patch with high-dielectric-constant substrate material and slots.

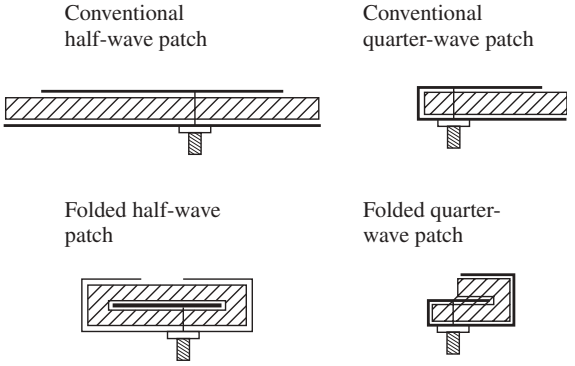


Figure 4.46 Size reduction by folding a half-wave or a quarter-wave patch.

4.81. Both the inverted-F and the shorting-pin methods have fields underneath the patch that bounce back-and-forth due to the presence of the shorting plate or pin. Once the bouncing distance reaches a half wave-length, the field starts to radiate. Due to these multiple bounces, the physical size of the patch is reduced. However, because of the non-unidirectional bounces and the fact that fields can radiate out from almost all edges of the patch, this type of antenna generally exhibits a high level of cross-polarization radiation. Of course, in certain applications, a high cross-polarization field is not a concern and actually could be a desirable feature, such as in the case of cellular phone communication in a multipath environment.

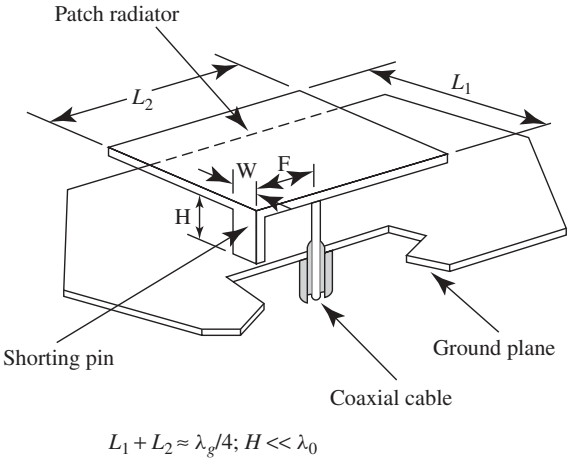


Figure 4.47 Size reduction by using an inverted-F patch.

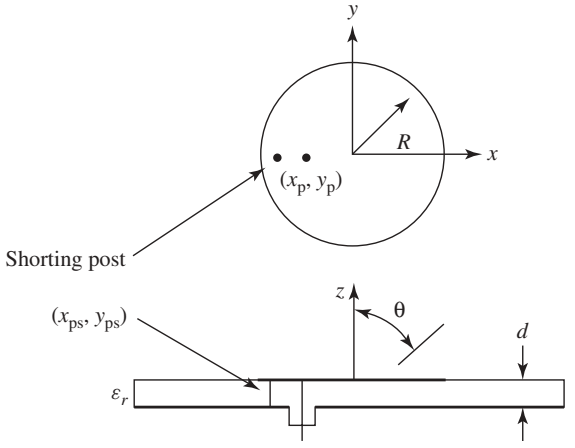


Figure 4.48 Size reduction of a circular patch by using a shorting pin.

5. *Back-to-Back Patch.* In most applications, the size of a patch antenna is generally limited by its large ground plane and the radiation is only on one side of the ground plane. However, if the ground plane is reduced to almost the same size as the patch by placing a second patch on the other side of the ground plane, an omni directional pattern can be formed with a relatively small and thin antenna. This is the so-called back-to-back antenna [83]. As shown in Figure 4.49, the antenna can be fed by a simple coplanar waveguide line in the ground plane. Another similar approach [84] is shown in Figure 4.50 as an array, where patch radiators and feed lines on one side of the ground plane are reversed in positions compared to the other side, and hence with radiators on both sides an azimuthal omnidirectional pattern is formed.

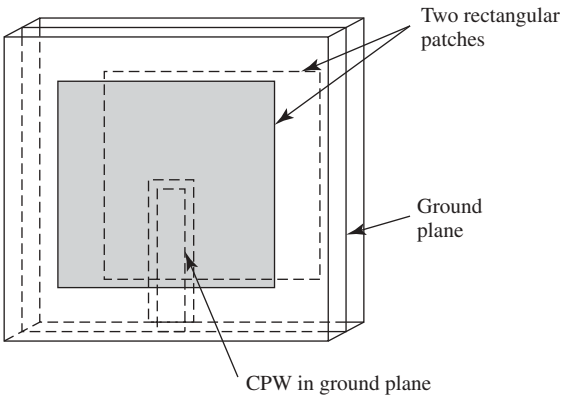


Figure 4.49 Back-to-back patch for omnidirectional pattern.

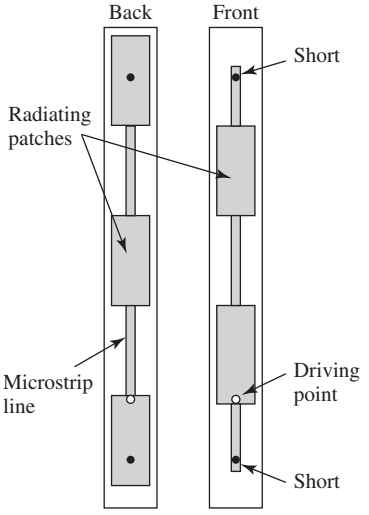


Figure 4.50 Back-to-back omnidirectional microstrip array.

4.8 SUMMARY

In this chapter, the key features of the microstrip antenna, along with its advantages and disadvantages, are discussed. Considerations in selecting a substrate material are also presented. The design methodologies and various analysis techniques for microstrip antennas have been briefly presented. In particular, the full-wave moment method is the most popular technique due to its computation accuracy and efficiency. Practical design techniques for the array configuration, radiating patch element, and power division lines are thoroughly discussed. Various patch excitation methods are also given. Important techniques for the microstrip antenna to generate dual-linear polarization and circular polarization, as well as to provide broadband and dual band, are presented.

It is expected that, because of their small size and low mass, the demand for microstrip antennas in future applications will continue to increase. On the other hand, there is also an unabated demand for improving the performance of microstrip antennas and arrays, such as widening of the bandwidth, reduction of the insertion loss, and improving the computation efficiency to handle large-size arrays. By utilizing the analysis techniques and design methods presented in this chapter, in conjunction with innovative ideas and knowing the physical mechanisms, the performance of microstrip antennas can be further enhanced to broaden their applications in the future.

REFERENCES

1. G. A. Deschamp, Microstrip microwave antennas, in *Proceedings of the Antenna Applications Symposium*, September 1953.
2. H. Gutton and G. Baissinot, Flat aerial for ultra high frequencies, French Patent No. 703113, 1955.
3. R. E. Munson, Conformal microstrip antennas and microstrip phased arrays, *IEEE Trans. Antennas Propag.*, Vol. 22, pp. 74–78, January 1974.
4. K. R. Carver and J. W. Mink, Microstrip antenna technology, *IEEE Trans. Antennas Propag.*, Vol. 29, pp. 2–24, January. 1981.
5. R. J. Mailloux, J. McIlvenna, and N. Kernweis, Microstrip array technology, *IEEE Trans. Antennas Propag.*, Vol. 29, pp. 25–38, January 1981.
6. J. Q. Howell, Microstrip antenna, *IEEE Trans. Antennas Propag.*, Vol. 23, pp. 90–93, January 1975.
7. A. G. Derneryd, Linear microstrip array antennas, Chalmer University of Technology, Goteborge, Sweden, Tech. Rep. 7505, October 1975.
8. J. L. Kerr, Microstrip polarization techniques, in *Proceedings of the Antenna Applications Symposium*, Allerton Park, IL, September. 1978.
9. J. Huang, Microstrip antennas for commercial applications, in *Microstrip Antennas*, D. M. Pozar and D. H. Schaubert, (Eds.), IEEE, Piscataway, NJ, 1995, pp. 371–379.
10. I. J. Bahl and P. Bhartia, *Microstrip Antennas*, Artech House, Norwood, MA, 1980.
11. J. R. James and P. S. Hall, *Handbook of Microstrip Antennas*, Peter Peregrinus Ltd., London, 1989.
12. D. M. Pozar, Microstrip antenna aperture-coupled to a microstripline, *Electron. Lett.*, Vol. 21, pp. 49–50, January 1985.
13. L. Murphy, SeaSAT and SIR-A microstrip antennas, in *Proceedings of the Workshop on Printed Circuit Antenna Technology*, paper 18, October 1979.
14. J. Huang, M. Lou, B. C. Lopez, and E. Gama, Foldable frame-supported thin-membrane array, in *Proceedings of the International Symposium on Antennas and Propagation (ISAP)*, Fukuoka, Japan, August 2000, pp. 213–216.
15. S. A. Long and M. D. Walton, A dual-frequency stacked circular disc antenna, in Digest of the International Symposium of the Antennas Propagation Society, College Park, MD, May 1978, pp. 260–263.
16. S. S. Zhong and Y. T. Lo, Single-element rectangular microstrip antenna for dual-frequency operation, *Electron. Lett.*, Vol. 19, pp. 298–300, 1983.
17. J. L. Kerr, Terminated microstrip antenna, in *Proceedings of the Antenna Applications Symposium*, Allerton Park, IL, September. 1978.

18. J. Huang, Microstrip reflectarray, in *IEEE AP-S/URSI Symposium*, London, Ontario, Canada, June 1991, pp. 612–615.
19. F. Crog and D. M. Pozar, Millimeter wave design of wideband aperture coupled stacked microstrip antennas, *IEEE Trans. Antennas Propag.*, Vol. 39, pp. 1770–1776, December 1991.
20. R. Q. Lee, K. F. Lee, and J. Bobinchak, Characteristics of a two-layer electromagnetically coupled rectangular patch antenna, *IEEE Trans. Antennas Propag.*, Vol. 38, pp. 1298–1302, August. 1990.
21. H. F. Pues and A. R. Van De Capelle, An impedance-matching technique for increasing the bandwidth of microstrip antennas, *IEEE Trans. Antennas Propag.*, Vol. 37, pp. 1345–1354, November 1989.
22. T. Teshirogi, M. Tanaka, and W. Chujo, Wideband circularly polarized array with sequential rotation, in *Proceedings of the International Symposium Antennas and Propagation (ISAP)*, Tokyo, Japan, August 1985, pp. 117–120
23. J. Huang, A technique for an array to generate circular polarization with linearly polarized elements, *IEEE Trans. Antennas Propag.*, Vol. 34, pp. 1113–1123, 1986.
24. D. H. Schaubert and F. G. Farrar, Some conformal printed circuit antenna designs, in *Proceedings of the Workshop 20 Printed Circuit Antenna Technology*, Las Cruses, New Mexico, October 1979.
25. T. Huynh and K. F. Lee, Single-layer single-patch wideband microstrip antenna, *Electron. Lett.* Vol. 31, No. 16, pp. 1310–1312, 1995.
26. K. F. Lee, K. M. Luk, K. F. Tong, and T. Huyuh, Experimental study of the rectangular patch with a U-shaped slot, in *IEEE AP-S/URSI Symposium*, 1996, Vol. 1, pp. 10–13,
27. K. M. Luk, C. L. Mak, Y. L. Chow, and K. F. Lee, Broadband microstrip patch antenna, *Electron. Lett.* Vol. 34, pp. 1442–1443, 1998.
28. C. L. Mak, K. M. Luk, K. F. Lee, and Y. L. Chow, Experimental study of a microstrip antenna with an L-shaped probe, *IEEE Trans. Antennas Propag.*, Vol. 48, pp. 777–783, May 2000.
29. H. Howe, Jr., Stripline is alive and well, *Microwave J.* Vol. 14, p. 25, July 1971.
30. R. Woo, Final report on RF voltage breakdown in coax transmission lines, *JPL Tech. Report 32-1500*, October 1970.
31. J. Huang, A parallel-series-fed microstrip array with high efficiency and low cross-polarization, *Microwave Optical Tech. Lett.*, Vol. 5, pp. 230–233, May 1992.
32. A. G. Derneryd, A theoretical investigation of the rectangular microstrip antenna, *IEEE Trans. Antennas Propag.*, Vol. 26, pp. 532–535, July 1978.
33. H. Pues and A. Van de Capelle, Accurate transmission-line model for the rectangular microstrip antenna, *Proc. IEE*, Vol. 131, pp. 334–340, December 1984.
34. W. F. Richards, Y. T. Lo, and D. Harrison, An improved theory for microstrip antennas and applications, *IEEE Trans. Antennas Propag.*, Vol. 29, pp. 38–46, January 1981.
35. Y. S. Wu and F. J. Rosenbaum, Mode chart for microstrip resonators, *IEEE Trans. Microwave TheoryTech.*, Vol. 21, pp. 487–489, 1973.
36. M. D. Deshpande and M. C. Bailey, Input impedance of microstrip antennas, *IEEE Trans. Antennas Propag.*, Vol. 31, pp. 740–747, September 1983.
37. D. M. Pozar, Input impedance and mutual coupling of rectangular microstrip antennas, *IEEE Trans. Antennas Propag.*, Vol. 30, pp. 1191–1196, November 1982.
38. I. E. Rana and N. G. Alexopoulos, Current distribution and input impedance of printed dipoles, *IEEE Trans. Antennas Propag.*, Vol. 29, pp. 99–105, January 1981.
39. J. R. Mosig and F. E. Gardiol, General integral equation formulation for microstrip antennas and scatterers, *Proc. IEE*, Vol. 132, P. H, pp. 424–432, 1985.

40. A. Reineix and B. Jecko, Analysis of microstrip patch antennas using finite difference time domain method, *IEEE Trans. Antennas Propag.*, Vol. 37, pp. 1361–1368, November. 1989.
41. X. Zhang and K. K. Mei, Time-domain finite-difference approach to the calculation of the frequency-dependent characteristics of microstrip discontinuities, *IEEE Trans. Microwave Theory Tech.*, Vol. 36, pp. 1775–1787, 1988.
42. C. Wu, K. L. Wu, Z. Q. Bi, and J. Litva, Accurate characterization of planar printed antennas using finite-difference time-domain method, *IEEE Trans. Antennas Propag.*, Vol. 40, pp. 526–534, May 1992.
43. K. S. Yee, Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media, *IEEE Trans. Antennas Propag.*, Vol. 14, pp. 302–307, 1966.
44. M. V. K. Chari and S. J. Salon, *Numerical Methods in Electromagnetism*, Academic Press, New York, 2000.
45. F. B. Hildebrand, *Advanced Calculus for Applications*, Prentice Hall, Englewood Cliffs, NJ, 1962.
46. H. C. Maric and G. Carey, *Introduction to Finite Element Analysis—Theory and Applications*, McGraw-Hill, New York, 1973.
47. J. Huang, Circularly polarized conical patterns from circular microstrip antennas, *IEEE Trans. Antennas Propag.*, Vol. 32, pp. 991–994, September 1984.
48. G. Kumar and L. Shafai, Generation of conical patterns from circular patch antennas and their performance, *Can. Elect. Eng. J.*, Vol. 10, pp. 108–112, 1985.
49. J. Watkins, Circular resonant structures in microstrip, *Electron. Lett.*, Vol. 5, No. 21, p. 524. October 1969.
50. J. R. James, P. S. Hall, and C. Wood, *Microstrip Antenna: Theory and Design*, Peter Peregrinus, Ltd., London, 1981.
51. T. Metzler, Microstrip series array, *IEEE Trans. Antennas Propag.*, Vol. 29, pp. 174–178, January 1981.
52. T. S. Horng and N. G. Alexopoulos, Corporate feed design for microstrip arrays, *IEEE Trans. Antennas Propag.*, Vol. 41, pp. 1615–1624, December 1993.
53. P. S. Hall and C. M. Hall, Coplanar corporate feed effects in microstrip patch array design, *Proc. IEE*, pp. 180–186, June 1988.
54. D. M. Pozar, Scanning characteristics of infinite arrays of printed antenna subarrays, *IEEE Trans. Antennas Propag.*, Vol. 40, pp. 666–674, June 1992.
55. J. Huang and D. M. Pozar, *Chapter 3 in Advances in Microstrip and Printed Antennas*, H. F. Lee and W. Chen, (Eds)., John Wiley & Sons, Hoboken, NJ 1997.
56. P. S. Hall, Probe compensation in thick microstrip patches, *Electron. Lett.*, Vol. 23, pp. 606–607, 1987.
57. J. Huang, Stripline feed for a microstrip array of patch elements with teardrop shaped probes, U.S. Patent 4,973,972, November 1990.
58. R. F. Thomas and J. Huang, Ultra-wideband UHF microstrip array for GeoSAR application, in *IEEE Antennas and Propagation Society International Symposium*, Atlanta, June 1998, pp. 2096–2099.
59. P. Bhartia, K. V. S. Rao, and R. S. Tomar, *Millimeter-Wave Microstrip and Printed Circuit Antennas*, Artech House, Boston, 1991, p.p 134 and 248.
60. R. Q. Lee, K. F. Lee, and J. Bobinchak, Characteristics of a two-layer electromagnetically coupled rectangular patch antenna, *Electron. Lett.*, Vol. 23, pp. 1070–1072, 1987.
61. J. F. Zurcher and F. E. Gardiol, *Broadband Patch Antenna*, Artech House, London, 1995, pp. 51–53.

62. T. Chiba, Y. Suzuki, and N. Miyano, Suppression of higher order modes and cross polarized component of microstrip antennas, *IEEE Antennas Propag. Symp. Dig*., pp 285–288, May 1982.

63. J. Huang, Dual polarized microstrip array with high isolation and low cross polarization, *Microwave Optical Technol. Lett.*, Vol. 41, pp 99–103, February 1991.

64. K. Woelders and J. Granholm, Cross-polarization and sidelobe suppression in dual linear polarization antenna arrays, *IEEE Trans. Antennas Propag.*, Vol. 45, pp. 1727–1740, December 1997.

65. T. Chiba, Y. Suzuki, N. Miyano, S. Muira, and S. Ohmori, A phased array antenna using microstrip patch antenna, in *12th European Microwave Conference*, Finland, September 1982, pp. 472–477.

66. M. Haneishi and S. Yoshida, A design method of circularly polarized rectangular microstrip antenna by one-point feed, *Electron. Commun. In Japan*, Vol. 54, pp. 46–54, 1981.

67. M. Haneishi, S. Yoshida, and N. Goto, Broadband microstrip array composed of single feed type circularly polarized microstrip elements, *IEEE Antennas Propag. Symp*., pp 160–163, May 1982.

68. L. T. Ostwald and C. W. Garvin, Microstrip command and telemetry antennas for communications technology satellite, in *IEE International Conference on Antennas for Aircraft and Spacecraft*, London, 1975, pp. 217–222.

69. Y. T. Lo and W. F. Richards, A perturbation approach to the design of circularly polarized microstrip antennas, in *IEEE Antennas and Propagation Society International Symposium*, Los Angeles, June 1981, pp. 339–342.

70. J. Vandensande, H. Pues, and A. Van de Capelle, Calculation of the bandwidth of microstrip resonator antennas, in *Proceedings of the 9th European Microwave Conference*, 1979, pp. 116–119.

71. A. Sabban, New broadband stacked two-layer microstrip antenna, *IEEE Antennas Propag. Symp*., pp. 63–66, 1983.

72. S. Sensiper, D. Williams, and J. P. McKone, An integrated global positioning satellite antenna low noise amplifier system, IEE ICAP conference publication 274, pp. 51–54, 1987.

73. P. Bhartia and I. J. Bahl, A frequency agile microstrip antenna, *IEEE Antennas Propag. Symp*., May pp. 304–307, 1982.

74. J. R. James and G. Andrasic, Dichroic dual band microstrip array, *Electron. Lett.*, Vol. 22, pp. 1040–1042, 1986.

75. D. M. Pozar and S. D. Targonski, A shared-aperture dual-band dual-polarized microstrip array, *IEEE Trans. Antennas Propag.*, Vol. 49, pp. 150–157, February 2001.

76. C. Han, J. Huang, and K. Chang, A high efficiency offset-fed X/Ka-dual-band reflectarray using thin membranes, *IEEE Trans. Antennas Propag.*, Vol. 53, pp. 2792–2798, September 2005.

77. S. A. Bokhari, J. F. Zuercher, J. R. Mosig, and F. E. Gardiol, A small microstrip patch antenna with a convenient tuning option, *IEEE Trans. Antennas Propag.*, Vol. 44, pp. 1521–1528, November 1996.

78. K. L. Wong, *Compact and Broadband Microstrip Antennas*, John Wiley & Sons, Hoboken, NJ, 2002.

79. J. Huang, Miniaturized UHF microstrip antenna for a Mars mission, in *IEEE Antennas and Propagation Society International Symposium*, Boston, July 2001, Vol. 4, pp. 486–489.

80. R. Chair, K. M. Luk, and K. F. Lee, Miniature multi-layer shorted patch antenna, *Electron. Lett.*, Vol. 36, pp. 3–4, January 2000.

81. H. Taga and K. Tsunekawa, Performance analysis of a build-in planar inverted-F antenna for 800 MHz band portable radio units, *IEEE Trans. Selected Areas Communi.* Vol. SAC-5, pp. 921–929, June 1987.

82. R. Waterhouse, Small microstrip patch antenna, *Electron. Lett.*, Vol. 31, pp. 604–605, April 1995.

83. H. Iwasaki, A microstrip array antenna with omni-directional pattern fed by a CPW, in *IEEE Antennas and Propagation Society International Symposium*, Baltimore, July 1996, pp. 1912–1915.

84. R. Bancroft and B. Bateman, An omni-directional planar microstrip antenna, *IEEE Trans. Antennas Propag.*, Vol. 52, pp. 3151–3153, November 2004.

CHAPTER 5

Reflector Antennas

WILLIAM A. IMBRIALE

5.1 INTRODUCTION

Reflector antennas have existed since the days of Heinrich Hertz (who designed and built the first parabolic cylinder reflector in 1888) and are still one of the best solutions to requirements for cost effective, high gain, high performance antenna systems. Consequently, the majority of ground station antennas are reflector antennas of one type or another as are many spacecraft high gain antennas.

Starting with the classic textbook *Microwave Antenna Theory and Design*, edited by Samuel Silver in 1949 [1], there have been many excellent reference and design handbooks on the subject. Henry Jasik was the editor of the first edition of the *Antenna Engineering Handbook*, published by McGraw-Hill in 1961 [2]. The book has had several revisions by Richard C. Johnson (1992) [3]. An extremely informative book on reflector antennas, *Analysis of Reflector Antennas*, by Rusch and Potter, appeared in 1970 [4]. There are also the *Handbook of Antenna Design* by Rudge, Milne, Olver, and Knight (1982) [5]; the *Antenna Handbook*, by Lo and Lee (1993) [6]; and *Modern Methods of Reflector Antenna Analysis and Design*, by Scott (1993) [7]. In addition, there are two books devoted primarily to reflector antennas: *Large Antennas of the Deep Space Network*, published in 2003 [8] and *Spaceborne Antennas for Planetary Exploration*, published in 2006 [9].

As noted previously, reflector antenna engineering is more than a century old. What then could yet another chapter on the subject contribute to the field? The answer is simply, my own personal view. Over the last four decades, I have designed many reflector systems, both for space and ground applications. Furthermore, I feel that it is the responsibility of the more senior members of the field to share their insights, mentor, and maybe even share a few of their "secrets" with the upcoming generation. It is to these ends that I devote the rest of this chapter.

Modern Antenna Handbook. Edited by Constantine A. Balanis

5.2 RADIATION PATTERN ANALYSIS

By far, the most important analytical tool for radiation pattern analysis is physical optics (PO), which is used to calculate the scattered field from a metallic reflecting surface—in this case, a reflector antenna. Electrical currents, which excite the scattered field, are induced on the conducting surface by an incident wave assumed to be of a known amplitude, phase, and polarization everywhere in space (e.g., from a feed or other reflecting surface). The PO approximations to the induced surface currents are valid when the reflector is smooth and the transverse dimensions are large in terms of wavelengths. The closed reflecting surface is divided into a region, S_1, which is illuminated by direct rays from the source ("illuminated region") and a region, S_2, which is geometrically shadowed ("shadowed region") from direct rays from the source (Figure 5.1). The PO approximations for the induced surface current distribution are

$$\begin{aligned} \mathbf{J}_s &= 2(\hat{\mathbf{n}} \times \mathbf{H}_{\text{inc}}) \qquad \text{on } S_1 \\ \mathbf{J}_s &= 0 \qquad\qquad\qquad\quad \text{on } S_2 \end{aligned} \tag{5.1}$$

where $\hat{\mathbf{n}}$ is the surface normal and $\mathbf{H}_{\text{inc}}$ the incident field. The expressions are then inserted into the radiation integral to compute the scattered field. The far-field secondary pattern for the reflector system is given by

$$\mathbf{E}(P) = \frac{-j\omega\mu}{4\pi}\frac{\exp(-jkR)}{R}\int_{\text{surface}} [\mathbf{J}_s - (\mathbf{J}_s \cdot \mathbf{a}_R)\mathbf{a}_R] \cdot \exp(jk\boldsymbol{\rho} \cdot \mathbf{a}_R)\, dS$$

where $k = 2\pi/\lambda$, $\mathbf{a}_R$ is the unit position vector of the far-field point, dS is the incremental surface area, and the quantities R, ρ, and P are as defined in Figure 5.2.

Before the advent of high speed digital computers, a great deal of effort went into simplifying calculations and not using complete radiation integrals. For example, Refs. 1 and 10 used a scalar formulation of the radiation integral. In addition, Ref. 11 provided a clear explanation of the assumptions that were used in reducing the complete vector formulation to a scalar formulation. Then Ref. 12 used a Jacobi–Bessel expansion to cast the two-dimensional (2D) integral into a sum of one-dimensional (1D) integrals. While these calculations still sometimes provide insight into reflector characteristics, they are no longer required because of the tremendous speed and storage capacity of modern computers.

One of the simplest possible reflector antenna computer programs that makes use of the complete vector formulation is based on a discrete approximation of the radiation integral.

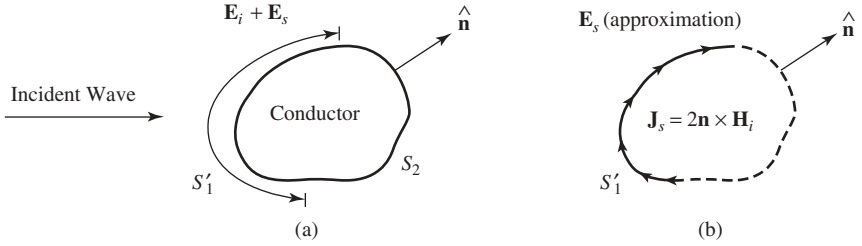


Figure 5.1 The physical optics approximation: (a) original problem and (b) approximation.

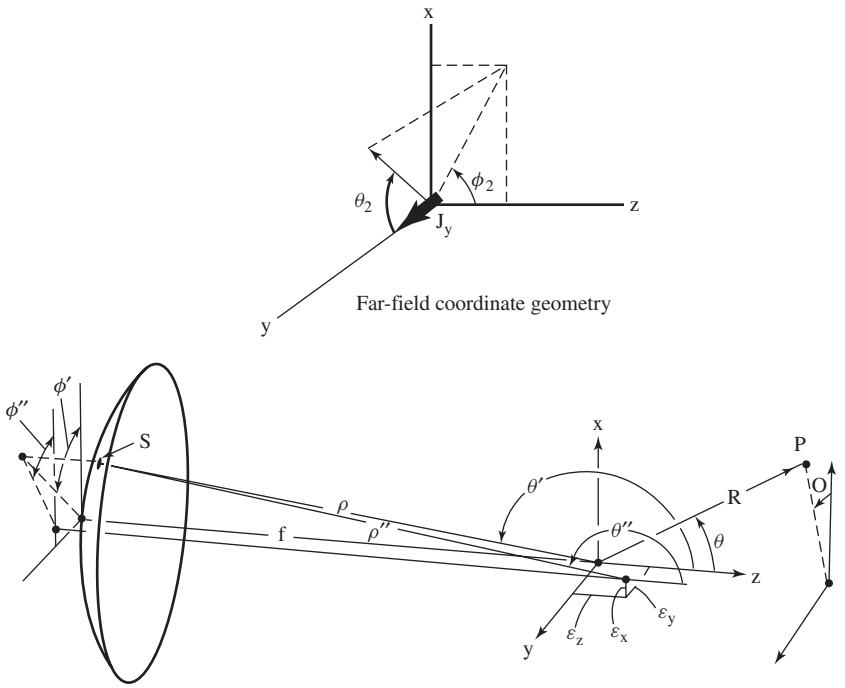


Figure 5.2 Reflector coordinate system.

This calculation replaces the actual reflector surface measurements with a triangular facet representation so that the reflector resembles a geodesic dome. The PO current is assumed to be constant in magnitude and phase over each facet, so the radiation integral is reduced to a simple summation. This program was originally developed in 1970 (see Refs. 13 and 14) and has proved to be surprisingly robust and useful, particularly when the near field is desired and surface derivatives are not known.

An important improvement to the program described above was the development of a more sophisticated approximation of PO surface current; this improvement permitted the use of larger facets by using a linear-phase approximation of the surface current. Within each triangular region, the resulting integral is the 2D Fourier transform of the projected triangle. This triangular-shape function integral can be computed in closed form. The complete PO integral is then a summation of these transforms. The complete formulation is given in Refs. 13 and 14 and is summarized next.

The PO radiation integral over the reflector surface, Σ, can be expressed as

$$\mathbf{H}(\mathbf{r}) = -\frac{1}{4\pi}\int_{\Sigma}\left(jk+\frac{1}{R}\right)\hat{\mathbf{R}}\times\mathbf{J}_s(\mathbf{r}')\frac{e^{-jkR}}{R}ds' \tag{5.2}$$

in which $\mathbf{r}$ designates the field point, $\mathbf{r}'$ the source point, $R = |\mathbf{r}-\mathbf{r}'|$ is the distance between them, and $\hat{\mathbf{R}} = (\mathbf{r}-\mathbf{r}')/R$ is a unit vector.

For the purpose of analysis, the true surface, Σ, is replaced by a contiguous set of triangular facets. These facets, denoted Δ_i, are chosen to be roughly equal in size with their vertices on the surface, Σ. Figure 5.3 shows a typical facet and its projection onto

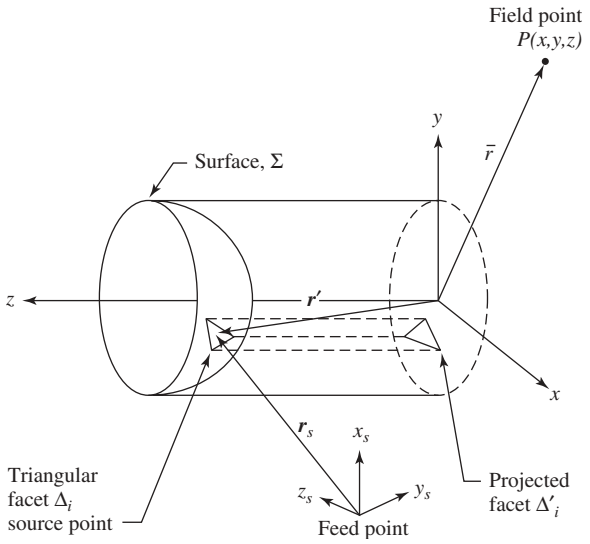


Figure 5.3 Reflector analysis coordinate systems and a typical triangular facet.

the $x-y$ plane. Let (x_i, y_i, z_i) represent the centroid of each triangle, where the subscript $i = 1,\ldots,N$ is associated with a triangle. Then the field obtained by replacing the true surface, Σ, with the triangular facet approximation is

$$\mathbf{H}(\mathbf{r}) = -\frac{1}{4\pi}\sum_{i=1}^{N}\int_{\Delta_i}\left(jk+\frac{1}{R}\right)\hat{\mathbf{R}}\times\mathbf{J}(\mathbf{r}')\frac{e^{-jkR}}{R}ds' \tag{5.3}$$

In Eq. (5.3), $\mathbf{J}$ is now the equivalent surface current evaluated on the triangular facets. Since the triangles are small, it is expected that $\hat{\mathbf{R}}$ and R do not vary appreciably over the area of a given facet. Thus let $\hat{\mathbf{R}}_i$ and R_i be the values obtained at the centroid (x_i, y_i, z_i) of each facet and approximate Eq. (5.3) by

$$\mathbf{H}(\mathbf{r}) = \frac{1}{4\pi}\sum_{i=1}^{N}\left(jk+\frac{1}{R_i}\right)\hat{\mathbf{R}}_i\times\mathbf{T}_i(\mathbf{r}) \tag{5.4}$$

$$\mathbf{T}_i(\mathbf{r}) = \int_{\Delta_i}\mathbf{J}_i(\mathbf{r}')\frac{e^{-jkR}}{R_i}ds' \tag{5.5}$$

Assume that the necessary transformations have been performed so that the incident field, $\mathbf{H}_s$, is given in terms of the reflector coordinate system. Then

$$\mathbf{J}_i(\mathbf{r}') = 2\hat{\mathbf{n}}_i\times\mathbf{H}_s(\mathbf{r}') \tag{5.6}$$

Next, assume that the incident field can be represented by a function of the form

$$\mathbf{H}_s = \mathbf{h}_s(\mathbf{r}_i)\frac{e^{jkr_s}}{4\pi r_{si}} \tag{5.7}$$

where r_s is the distance to the source point and r_{si} is the distance from the triangle centroid to the source point. Then, Eq. (5.5) can be written

$$\mathbf{T}_i(\mathbf{r}) = \frac{\hat{\mathbf{n}}_i \times \mathbf{h}_s(\mathbf{r}_i)}{2\pi R_i r_{si}} \int_{\Delta_i} e^{-jk(R+r_s)} ds' \tag{5.8}$$

Making use of the Jacobian and approximating

$$R(x, y) + r_s(x, y) = \frac{1}{k}(a_i - u_i x - v_i y) \tag{5.9}$$

in which a_i, u_i, and v_i are constants, the expression can be rewritten

$$\mathbf{T}_i(\mathbf{r}) = \frac{\hat{\mathbf{n}}_i \times \mathbf{h}_s(\mathbf{r}_i)}{2\pi R_i r_{si}} J_{\Delta_i} e^{-ja_i} \int_{\Delta_i'} e^{j(u_i x' + v_i y')} dx' dy' \tag{5.10}$$

where the surface normal is

$$\mathbf{N}_i = -\hat{\mathbf{x}} f_{xi} - \hat{\mathbf{y}} f_{yi} + \hat{\mathbf{z}} \tag{5.11}$$

and the Jacobian is

$$J_{\Delta_i} = |\mathbf{N}_i| = \left[f_{xi}^2 + f_{yi}^2 + 1\right]^{1/2} \tag{5.12}$$

It may now be observed that this integral is the 2D Fourier transform of the ith projected triangle Δ_i', expressed as

$$S(u, v) = \int_{\Delta_i'} e^{j(ux' + vy')} dx' dy' \tag{5.13}$$

which can be computed in closed form as described in Ref. 15. The full radiation integral is then the sum of all the transforms of the individual triangles.

5.2.1 Dual-Reflector Systems

The PO integration methodology is incorporated into a sequential fashion for analyzing the dual-reflector antenna system. Initially, the feed illuminates the subreflector, and the currents on the subreflector surface are determined. Subsequently, the near fields scattered from the subreflector are used to illuminate the main reflector, and its induced currents are determined. The main reflector's scattered fields are then determined by integrating these currents.

Many coordinate systems are required to allow flexibility in locating and orienting the feed, subreflector, main reflector, and output-pattern generation. The relation among these coordinate systems is depicted in Figure 5.4.

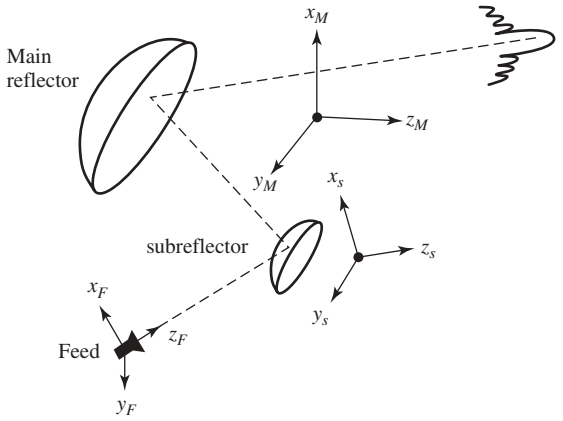


Figure 5.4 Dual-reflector coordinate systems.

5.2.2 Useful Coordinate Transformations

The transformation equations that permit coordinates and vectors described in one coordinate system to be expressed in terms of some other coordinate system must be known. Furthermore, the transformation may require both translation and rotation. The required transformations are given in detail by Rahmat-samii [16] and are summarized below. They are the Cartesian-to-spherical transformation and coordinate rotations using Eulerian angles.

The Cartesian-to-spherical transformation is conveniently summarized in matrix form. With the Cartesian components of a vector, **H**, denoted (H_x, H_y, H_z) and the spherical components (H_γ, H_θ, H_ϕ), the transformation is

$$\begin{bmatrix} H_r \\ H_\theta \\ H_\phi \end{bmatrix} = \begin{bmatrix} \sin\theta\cos\phi & \sin\theta\sin\phi & \cos\theta \\ \cos\theta\cos\phi & \cos\theta\sin\phi & -\sin\theta \\ -\sin\phi & \cos\phi & 0 \end{bmatrix} \begin{bmatrix} H_x \\ H_y \\ H_z \end{bmatrix} \tag{5.14}$$

The inverse transformation is just the transpose of the above matrix.

Rotations are handled by the use of the Eulerian angles (α, β, γ). These angles describe three successive rotations that bring one Cartesian system into alignment with another. Let the two systems be denoted (x_1,y_1,z_1) and (x_2,y_2,z_2). As illustrated in Figure 5.5, the angles are defined as follows:

- α describes a positive rotation about the z_1-axis, which brings the x_1-axis into the x'-axis aligned with the *line of nodes*, that is, the line of intersection between the (x_1,y_1) and (x_2,y_2) planes.
- β describes a positive rotation about the line of nodes (the x'-axis) that brings the z_1-axis to the z_2-axis.
- γ describes a positive rotation about the z_2-axis, which brings the x'-axis to the x_2-axis.

The phrase "positive rotation" means the direction of increasing angular measure as defined by the right-hand rule with respect to the axis about which the rotation occurs.

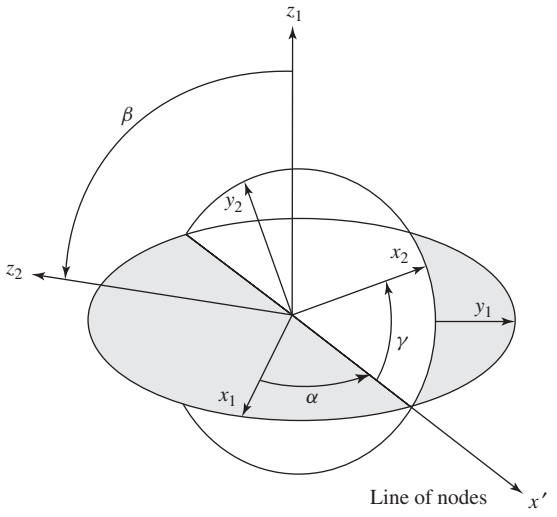


Figure 5.5 Euler-angle definitions.

Each of the rotations just described is performed using the standard rotation of coordinate formulas of plane analytic geometry.

When these expressions are written in matrix form and applied successively as described above, one obtains the following matrix equation that represents a general 3D rotation of coordinates:

$$\begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} \tag{5.15}$$

where the individual matrix elements are

$$\begin{aligned}
A_{11} &= \cos\gamma\,\cos\alpha - \sin\gamma\,\cos\beta\,\sin\alpha \\
A_{12} &= \cos\gamma\,\sin\alpha + \sin\gamma\,\cos\beta\,\cos\alpha \\
A_{13} &= \sin\gamma\,\sin\beta \\
A_{21} &= -\sin\gamma\,\cos\alpha - \cos\gamma\,\cos\beta\,\sin\alpha \\
A_{22} &= -\sin\gamma\,\cos\alpha + \cos\gamma\,\cos\beta\,\cos\alpha \\
A_{23} &= \cos\gamma\,\sin\beta \\
A_{31} &= \sin\beta\,\cos\gamma \\
A_{32} &= -\sin\beta\,\cos\alpha \\
A_{33} &= \cos\beta
\end{aligned}$$

Although the formulas are presented in terms of coordinate transformations, the transformation matrix is equally valid for the Cartesian components of a vector.

5.3 FUNDAMENTAL DESIGN PRINCIPLES

This section describes the fundamental principles for designing reflector antennas.

5.3.1 Aperture Illumination and Spillover

The efficiency of a reflector antenna is primarily determined by (1) the ability of the feed system to illuminate only the reflector while minimizing the energy that radiates elsewhere and (2) the ability of the feed to uniformly illuminate the parabola. The first item above is termed "spillover efficiency" and the second "illumination efficiency." The illumination efficiency is 100% when the energy density on the entire main reflector aperture is a constant.

Since feed patterns gradually taper from their central maxima to nulls, a compromise between aperture illumination and spillover must be arrived at in order to maximize on-axis gain.

A convenient representation for a symmetric feed pattern with peak gain in the z-direction is

$$\begin{aligned} |E(\theta, \psi)| &= \cos^n(\theta), \quad 0 \leqslant \theta \leqslant \pi/2 \\ |E(\theta, \psi)| &= 0, \qquad\quad \theta > \pi/2 \end{aligned} \tag{5.16}$$

where θ and ψ are shown in Figure 5.6.

It can be shown that the gain of the feed is $G = 2(2n + 1)$. With the given feed pattern placed at the focal point of a symmetric reflector, curves of efficiency (or gain) versus n with F/D as a parameter can be generated as shown in Figure 5.7. As can be observed, for each F/D there is a value for n that maximizes gain.

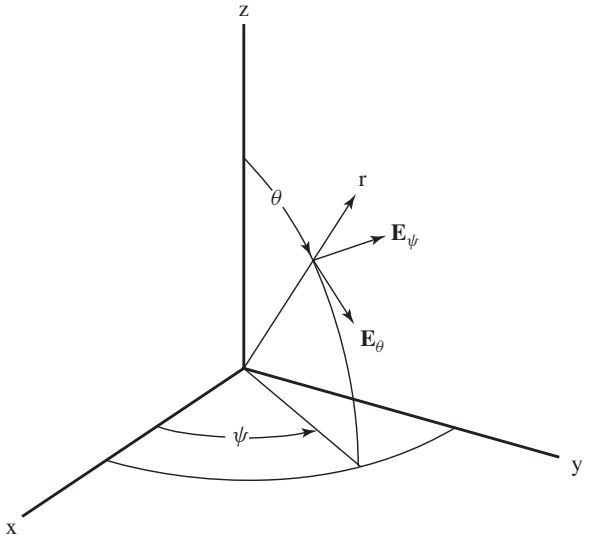


Figure 5.6 Coordinate system for feed patterns.

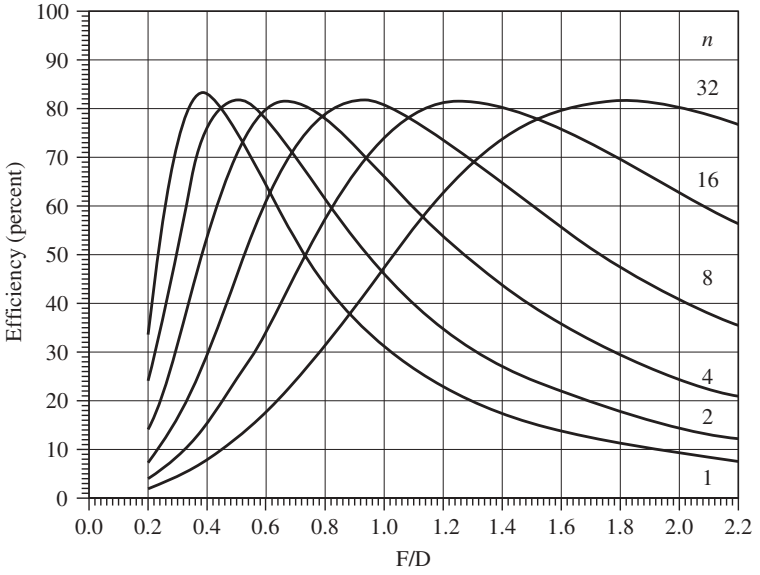


Figure 5.7 Efficiency versus F/D with n as a parameter.

5.3.2 Axial Defocusing

For a surface of revolution, the pattern function for scalar aperture theory is given by (see Figure 5.8)

$$g(u,\phi) = \frac{a^2}{2F}\int_0^{2\pi}\int_0^1 (1+\cos\theta)G(\theta,\psi)e^{jur\cos(\phi-\psi)}e^{-j2\pi Z\cos\theta}r\,dr\,d\psi \tag{5.17}$$

where Z is the defocusing in wavelengths from the focus F and is negative toward the vertex of the parabola, a is the radius in wavelengths, $G(\theta, \psi)$ is the primary voltage pattern, and $u = (2\pi a/\lambda)\sin\theta'$. Observe that θ is a function of r.

If we use the primary feed representation shown in Eq. (5.16) and expand

$$e^{jur\cos(\phi-\psi)} = \sum_{m=-\infty}^{\infty} j^m J_m(ur)e^{jm(\phi-\psi)} \tag{5.18}$$

and integrate over the variable ψ, there results

$$g(u,\phi) = \frac{\pi a^2}{F}\int_0^1 (1+\cos\theta)J_0(ur)f(\theta,\phi)e^{-j2\pi Z\cos\theta}r\,dr \tag{5.19}$$

with

$$\sqrt{D_3}f(\theta,\phi) = \frac{\cos^n\theta+\cos^m\theta}{2} + \frac{1}{2}\frac{J_2(ur)}{J_0(ur)}\cos 2\phi\left\{\cos^n\theta-\cos^m\theta\right\}$$

Observe that the axial defocusing Z represents a phase error over the aperture and consequently the gain is peaked when $Z = 0$.

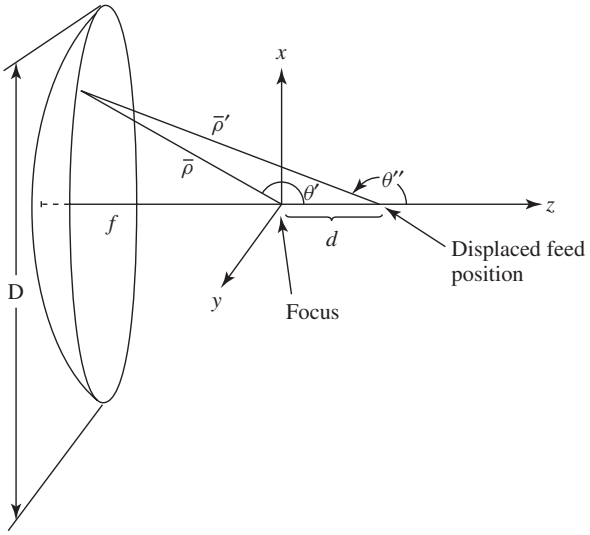


Figure 5.8 Geometry of a parabola with an axially defocused feed.

If the source is an infinitesimal electric dipole (which produces a very broad feed pattern), the linear approximations permit the radiation field to be integrated in closed form, and the gain loss as a function of defocusing can be shown to be proportional to $(\sin x)/x$, where (see also Figure 5.8 and Ref. 17)

$$x = \frac{2\pi(d/\lambda)}{1 + (4f/D)^2} \tag{5.20}$$

Consequently, there are nulls in the on-axis gain at distances given by

$$\frac{d}{\lambda} = \pm\frac{m}{2}\left[1 + \left(\frac{4f}{D}\right)^2\right] \tag{5.21}$$

Near the on-axis nulls, the main beam is widened and, in any plane, the field on either side of the axis is greater than the on-axis value so that the beam appears bifurcated. In Ref. 11 there is a discussion of the difference between the linear approximation and the complete radiation integral solution. For most cases the difference is small for distances not far from the focal point.

Defocusing curves for several feed functions are shown in Figure 5.9. The curves are normalized by the total feed power so that spillover is included. Observe that, for example, as shown in Figure 5.7, $\cos(\theta)$ is optimum for $F/D = 0.4$. The less tapered feeds exhibit deep minima. As the taper is increased, the minima become less pronounced. Figures 5.10 and 5.11 show the angular radiation patterns as a function of defocusing for two illuminations. Observe that the less-tapered feed pattern exhibits deep minima in the patterns while the more tapered feed does not. Consequently, highly tapered feeds can be used in beam-broadening applications while less-tapered feeds are generally unsuitable.

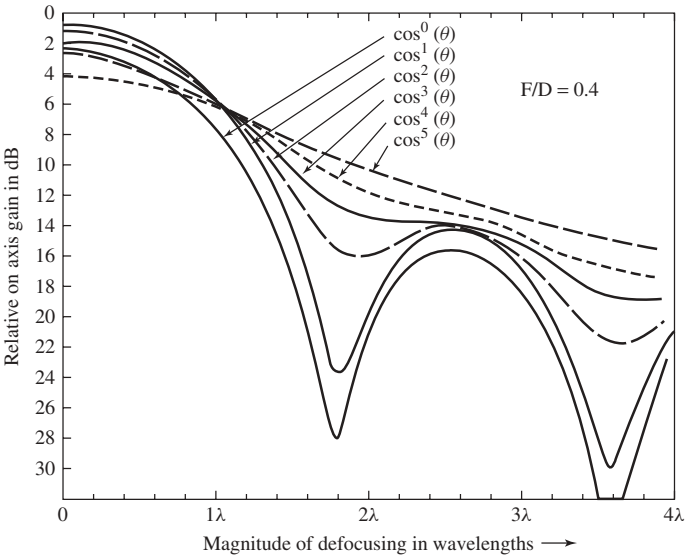


Figure 5.9 Relative boresight gain as a function of defocusing for $\cos^n(\theta)$ illumination.

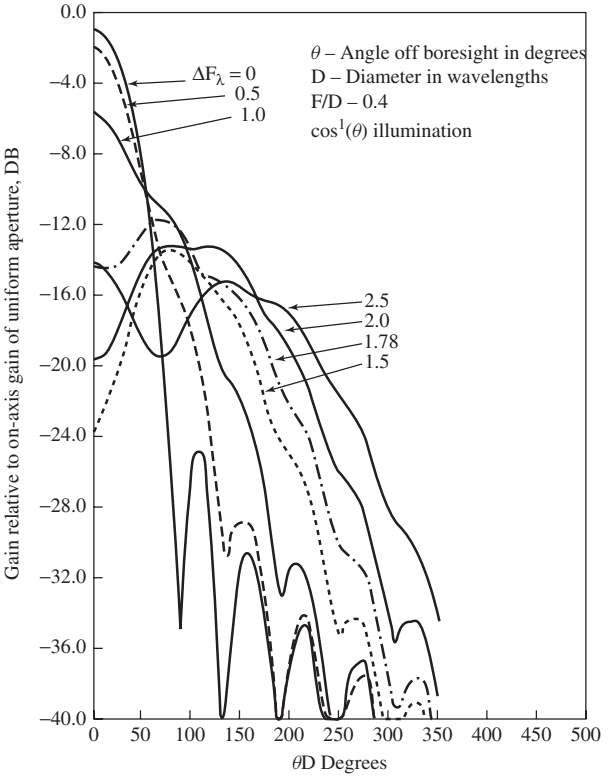


Figure 5.10 Angular patterns for $\cos(\theta)$ illumination with variable defocusing.

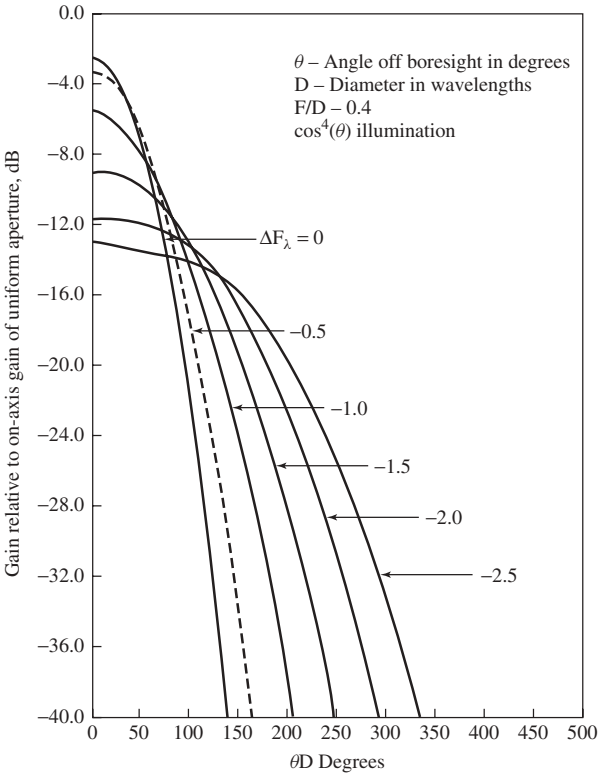


Figure 5.11 Angular patterns for $\cos^4(\theta)$ illumination with variable defocusing.

5.3.3 Lateral Defocusing

When a feed is laterally displaced from the focal point, there is a loss in peak gain, and the beam is scanned from the boresite an amount equal to BDF $\times \theta$, where θ is the specular reflection direction, and BDF is the beam deviation factor (see Figure 5.12). Generally, BDF is in the range of 0.8–0.95, depending on the F/D of the reflector and the feed illumination function. The two classic papers on the subject are Refs. 10 and 11, where the equations for computing the pattern of a laterally scanned feed are given. In Ref. 11 the complete vector formulation is given along with the approximations that are used to derive the scalar formulation given in Ref. 10.

In Ref. 11, vector and scalar analysis were used to examine reflector performance under the conditions of large lateral feed displacements. Results from the two formulations were compared to experimental data obtained from a precision $F/D = 0.4$ reflector with minimum blockage. Figure 5.13 shows the measured secondary patterns as a function of lateral primary-feed displacement. Observe the very significant sidelobe (called a coma lobe because of its similarity to optical aberrations) when the beam is scanned. Figure 5.14 shows the secondary beam squint and scan loss and Figure 5.15 the coma-lobe gain level relative to peak gain. Several pertinent observations can be made. The beam-peak angle position is accurately predicted by both the vector and scalar theories. The peak gain is accurately predicted by the vector theory, but the scalar theory is several decibels

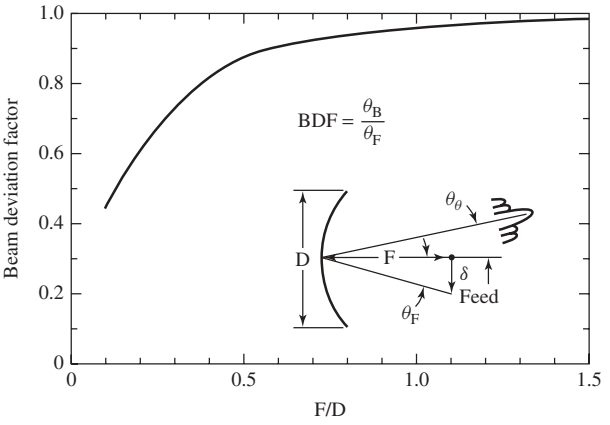


Figure 5.12 Lateral scan geometry.

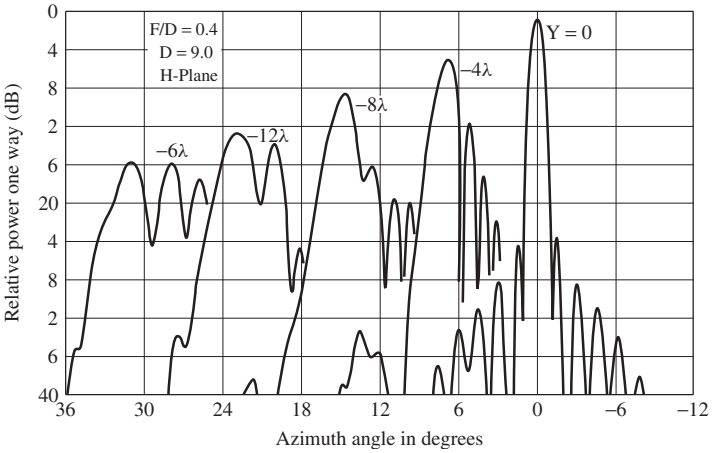


Figure 5.13 Measured secondary patterns as a function of lateral primary feed displacement.

in error for large scan angles. The scalar theory can be greatly in error for predicting the coma-lobe peak. A comparison of the peak-gain difference between the two theories indicates that the accuracy of the scalar theory is more dependent on the actual amount of angular scan rather than on the scan in terms of half-power beamwidths. The scalar analysis indicates that the highest gain level for a given lateral displacement occurs with the feed in the focal plane, whereas the vector formulation correctly predicts scan-plane fields.

5.3.4 Root-Mean-Square Surface Error

The classic article on antenna tolerance is by Ruze [18], wherein he derives the often quoted equation for computing the gain loss in an antenna caused by random surface errors:

$$G = G_0 e^{-(4\pi\varepsilon/\lambda)2} \tag{5.22}$$

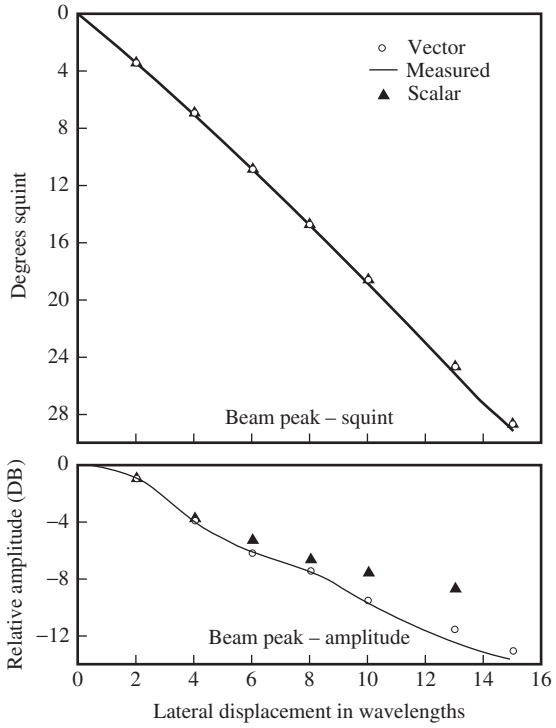


Figure 5.14 Secondary beam squint and beam scan loss as a function of lateral displacement of the primary feed in the focal plane.

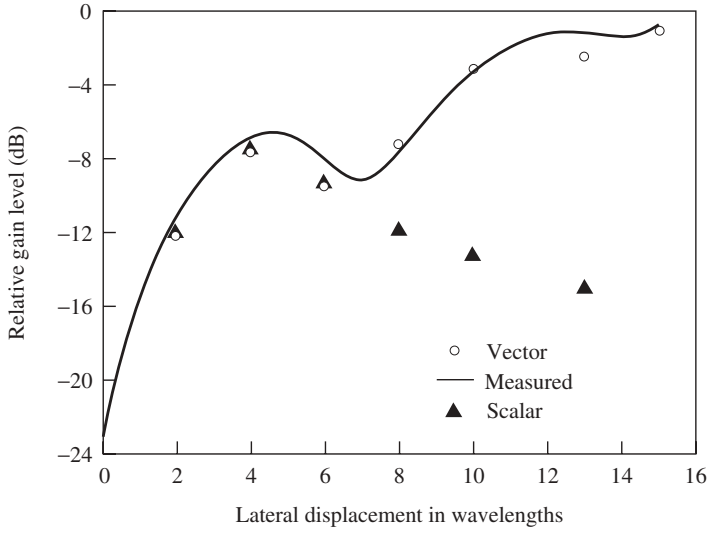


Figure 5.15 Coma-lobe level relative to beam peak as a function of lateral displacement of the primary feed in the focal plane.

where G_0 is the undistorted gain and ε is the root-mean-square (RMS) surface error in the same units as λ. Even though Ruze gives specific conditions for the formula to be valid, experience has shown it to be useful in almost all practical applications where the RMS error is not too extreme. The gain loss due to reflector surface error for several frequencies is shown in Figure 5.16. The curve is shown here because it is incorrect in Ref. 8 due to a mix-up in metric versus English units of measure.

Levy [19] developed the equations to compute the "best fit" RMS for use in the Ruze equation. His method consists of taking points on the antenna surface and computing the half-pathlength error (the microwave equivalent to the RMS surface error) from an alternative surface that best fits the deformed surface. Since the important effect on the antenna performance (specifically gain) is the phase-error distribution over the surface, if the original parabolic surface deforms into another parabolic surface, all rays from the second surface will have the same new overall pathlength. Since these rays will arrive at the feed with a constant phase, there will be no adverse performance effect.

The alternative surface is defined in terms of five parameters that constitute a rigid-body motion and an additional parameter related to a change in the original focal length. However, it is necessary for the position of the subreflector in a dual-reflector system and the position of the feed in a focal-feed reflector to be movable. Mobility allows compatible variations in the microwave-path geometry established by the fitting parameters. Typical antennas actually have provisions for providing these necessary motions. However, if the antenna system is unable to compensate for these parameters, the overall performance can be estimated by combining the deformation error loss with the performance loss caused by the equivalent axial and lateral offsets due to the best-fit parameters.

The five parameters are indicated in Figure 5.17. They consist of three translations, U_0, V_0, and W_0, parallel to the x-, y-, and z-coordinate axes, and rotations θ_x and θ_y about their respective axes. One more parameter is related to the new focal length, F_0. Reference 19 contains a MATLAB software application for computing the best-fit surface and residual pathlength error. In Ref. 20 the six parameters are called "homology

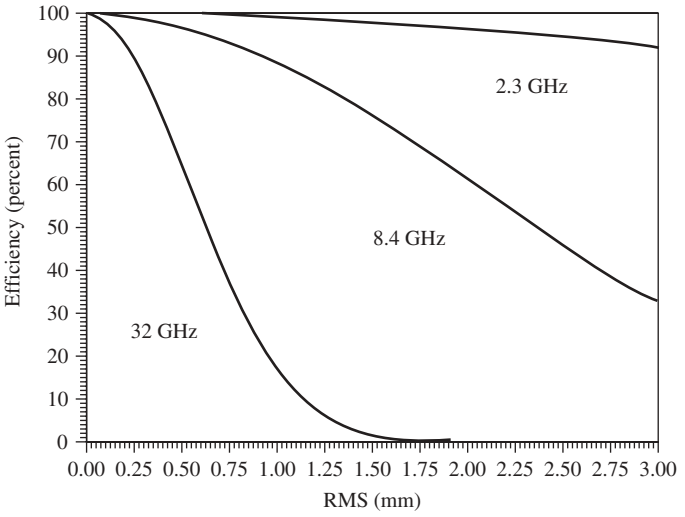


Figure 5.16 Gain loss due to reflector surface error.

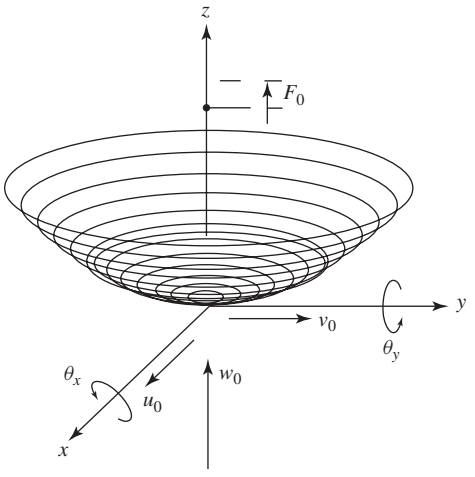


Figure 5.17 The fitting parameters.

parameters" because they represent a transformation from the original parabolic surface to an alternative parabolic surface.

5.3.5 Phase Error

In many respects, phase errors are similar in effect to surface deformation errors; however, they may be due to other causes as, for example, feed phase errors. If the phase errors are symmetric, there is generally a gain loss and an effect on the sidelobe structure. If the errors are asymmetric over the structure, there may also be a beam squint. The errors can be analyzed as distortion effects, or the phase function can be included as part of the feed phase function and included in the determination of the PO radiation integral.

5.3.6 Blockage

In a symmetric parabolic antenna, opaque (or partially opaque) devices are interposed between the feed system and the reflector and between the reflector and the far-field observation point. The simplest and most often used approximations for computing the effect of aperture blockage on gain and radiation patterns are based on optical considerations. The basic concept is to null out or set the PO currents to zero on the portion of the reflector that is shadowed by the feed support or other objects that block the feed energy from getting to the reflector or the far field. Figure 5.18 shows a face-on view with the geometrically blocked area shaded. Blockage of the central region is generally caused by a subreflector or the feed itself. The plane-wave region is a projection of the feed support structure as seen from the far-field observation point, and the spherical-wave shadow results from the part of the feed support structure, which is geometrically between the focal point and the reflector surface. Both Refs. 4 and 19 give formulas for computing the blocked shadow region from typically assumed feed supports or struts.

The major effect on the radiation patterns is to reduce the on-axis gain and modify the sidelobe structure.

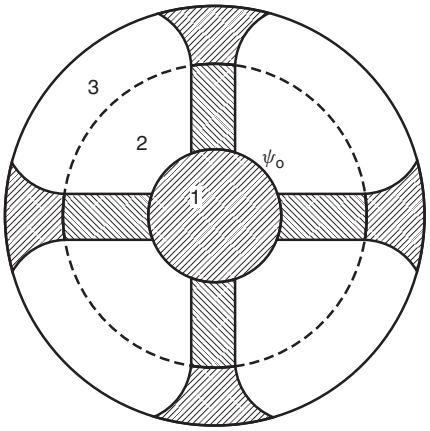


Figure 5.18 Plane- and spherical-wave shadows.

5.3.7 Sidelobes

Aperture distributions necessary to low sidelobes have been known for many years, but being able to achieve these distributions in reflector antenna designs has proved difficult. Figure 5.19 (from Ref. 21) illustrates a basic axisymmetric parabolic reflector antenna with a point-source feed at the focus. The feed illuminates the reflector to produce amplitude and phase distribution in the aperture plane, which then Fourier-transforms into a secondary pattern in the far field. The feed typically produces a monotonically shaped amplitude distribution as shown by the solid curve, whereas a low sidelobe distribution generally requires an inflected shape as shown by the dashed curve. The distribution shape and the edge taper determine the near-in sidelobe levels. However, in addition to the sidelobes caused by the aperture distribution, there are also sidelobes caused by the primary illumination that misses the reflector (spillover) and diffraction effects at the reflector edge (which produce wide-angle and rearward lobes). There is

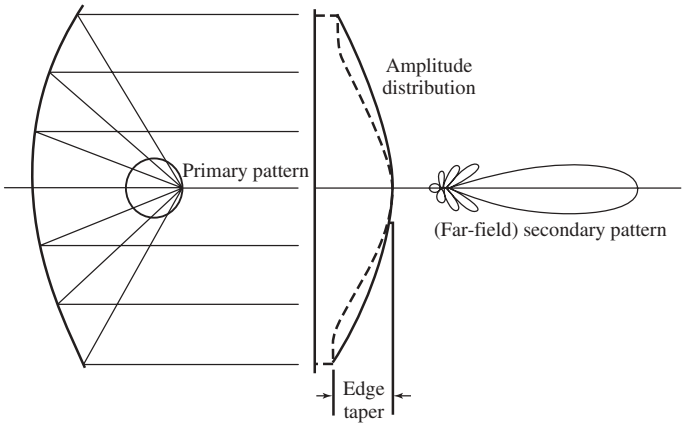


Figure 5.19 Basic axisymmetric reflector showing typical amplitude distribution from a simple feed compared with desired low sidelobe distribution (dashed).

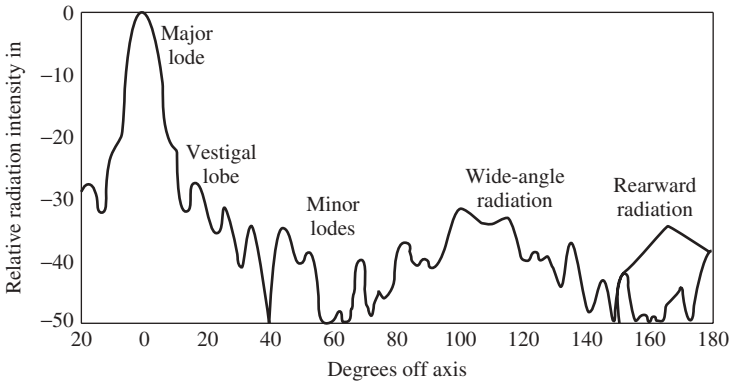


Figure 5.20 Typical pattern for focal-fed axisymmetric paraboloid. (From Ref. 22.)

also degradation in the sidelobe levels caused by aperture blockage and reflector surface accuracy. Figure 5.20 (from an old classic [22]) illustrates a typical pattern produced by early focal-fed axisymmetric paraboloids. Near-in sidelobe levels of −20 to −25 dB were typical, with −30 dB or lower difficult to achieve. Front-to-back lobe ratios were generally about 40–50 dB, depending on the antenna size (gain) and the edge illumination (taper).

Over the years numerous techniques have been used to lower sidelobes such as edge treatments, aperture shaping, horn parabola, and polarization twist (see Ref. 21 for more details). But probably the best methods use offset feeding to avoid aperture blockage and dual shaped designs to control the aperture distribution. It is also important to have a very good (low sidelobe) primary feed. The design for the WMAP (Wilkinson Microwave Anisotropy Probe, a NASA satellite launched in 2001 to survey the sky to measure the temperature of the radiant heat left over from the Big Bang) spacecraft [23] used this technique to produce a design that had greater than 90-dB sidelobes everywhere in space.

5.3.8 Polarization

Any monochromatic wave is elliptically polarized; that is, the end point of its electric vector at each point in space periodically traces out an ellipse or one of its special forms, namely, a circle or straight line. The orthogonal components of the incident radiation (wave traveling along the positive z-axis) at any point in space can be written

$$\begin{aligned} E_x &= E_1 \sin(\omega t - \delta) \\ E_y &= E_2 \sin(\omega t) \end{aligned} \tag{5.23}$$

where δ is the phase difference between E_x and E_y, and E_1 and E_2 are real amplitudes. If ωt is eliminated from the equations, the following result is readily obtained (e.g., see Ref. 24):

$$\begin{aligned} &aE_x^2 - bE_xE_y + cE_y^2 = 1 \\ &a = \frac{1}{E_1^2 \sin^2 \delta} \qquad b = \frac{2\cos\delta}{E_1E_2 \sin^2 \delta} \qquad c = \frac{1}{E_2^2 \sin^2 \delta} \end{aligned} \tag{5.24}$$

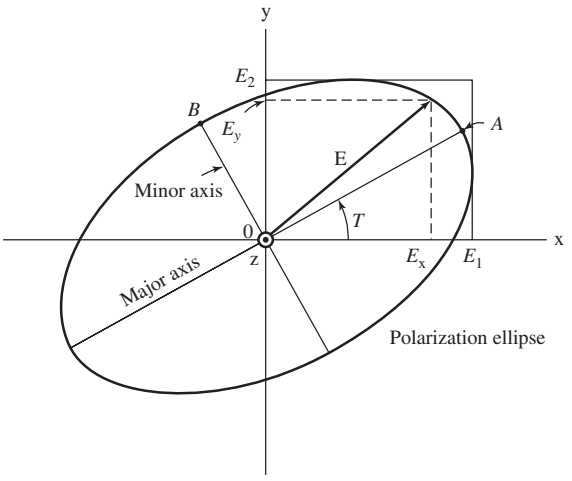


Figure 5.21 Polarization ellipse at tilt angle τ.

Equation (5.24) describes a (polarization) ellipse, as in Figure 5.21. The line segment OA is the semimajor axis, and the line segment OB is the semiminor axis. The tilt angle of the ellipse is τ. The axial ratio is

$$\text{AR} = OA/OB \quad (1 < \text{AR} < \infty) \tag{5.25}$$

If $E_1 = 0$, the wave is linearly polarized in the y-direction. If $E_2 = 0$, the wave is linearly polarized in the x-direction. If $\delta = 0$ and $E_1 = E_2$, the wave is also linearly polarized but in a plane at an angle of 45° with respect to the x-axis ($\tau = 45^\circ$).

If $E_1 = E_2$ and $\delta = \pm 90^\circ$, the wave is circularly polarized. When $\delta = +90^\circ$, the wave is *left circularly polarized*, and when $\delta = -90^\circ$, the wave is *right circularly polarized*.

An antenna receiving a plane wave from a given direction will pick up the maximum signal when the polarization ellipse of the incident electric field has the same axial ratio, the same tilt angle, and the same sense of polarization as those of the receiving antenna in the given direction. If the polarization of the incident wave does not match the polarization of the receiving antenna, a polarization loss occurs due to the mismatch. The amount of mismatch is given by the "polarization efficiency," which is defined as the ratio of the power actually received by the antenna divided by the power that would be received if a wave came from the same direction and with the correct polarization incident on the antenna.

The "cross polarization" of a source is also an important concept. It is defined as the polarization orthogonal to a reference polarization. The "copolar" field of a source is the component of the field that is parallel to the field of the reference source and the "cross-polar" field is the orthogonal component.

This definition, however, is not unique for linear and elliptically polarized fields because it does not define the direction of the reference polarization for all observation angles. Reference 25 identified three possibilities for the case of a linearly polarized source (Figure 5.22).

These have become known as the Ludwig first, second, and third definitions. The first definition says that the reference field is a plane wave in a rectangular coordinate

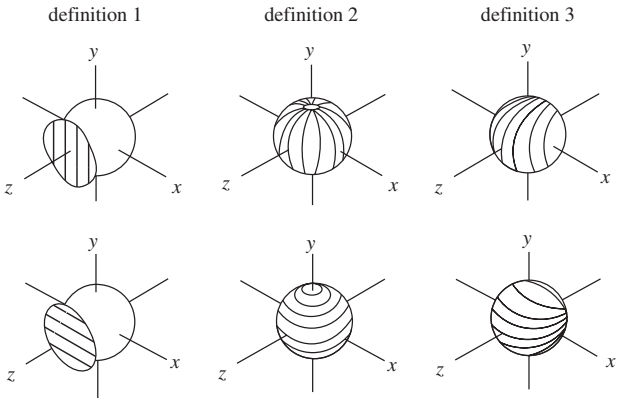


Figure 5.22 Definitions of cross polarization.

system. The second definition says that the reference polarization is that of an electric dipole lying in the aperture plane; the cross polarization is then the polarization of a magnetic dipole with the same axes as the electric dipole. The third definition says that the reference polarization is that of a Huygens source (i.e., electric and magnetic dipoles with orthogonal axes lying in the aperture plane and radiating equal fields in phase along the z-axis); the cross polarization is then the polarization of a similar source rotated 90° in the aperture plane.

The copolar $E_p(\theta, \phi)$ and cross-polar $E_c(\theta, \phi)$ fields using definition 3 are given by

$$\begin{aligned} E_p(\theta, \phi) &= E_\theta(\theta, \phi) \sin\phi \, a_\theta + E_\phi(\theta, \phi) \cos\phi \, a_\phi \\ E_q(\theta, \phi) &= E_\theta(\theta, \phi) \sin\phi \, a_\theta - E_\phi(\theta, \phi) \cos\phi \, a_\phi \end{aligned} \tag{5.26}$$

with the principal vector aligned with the y-axis.

Ludwig recommended the use of the third definition because it most closely corresponds to what is normally measured when an antenna radiation pattern is recorded and, when interchanging the copolar and cross-polar fields as measured in any direction, it corresponds to a 90° rotation of the reference source. This definition has become widely accepted because it is the only one that does not lead to confusion when an elliptically polarized source is considered. Note that in the case of an elliptical wave, it is not sufficient to rotate the reference source by 90° about the direction of propagation. The rotation must be accompanied by a change in the sense of polarization.

5.4 ADVANCED ANALYSIS TECHNIQUES FOR DESIGNING REFLECTOR ANTENNA SYSTEMS

Portions of the material in the following sections are derived from Refs. 8 and 9.

5.4.1 Feeds

Selection of the feed is a critical part of the reflector system design. For the most part, the type of feed used is determined by system requirements. If a very wide frequency band is required, a frequency-independent feed such as a log-periodic dipole antenna could be

used as a feed (as described in the next section). For some narrowband low frequency systems, a simple feed like a dipole could be used. For some circular polarized applications, a helix antenna could be used. Open-ended waveguides, short-backfire antennas, a disk on rod, as well as many other feed types have been used. However, the predominant type of feed is the horn because it generally provides the most efficient antenna system for both single- and dual-reflector antennas. Horns have very low insertion loss and often a Gaussian shape pattern that has minimal spillover loss.

Two types of horn antenna possessing equal E- and H-plane patterns are commonly used. The first is the dual-mode horn [26], and the second is the corrugated horn [27]. In the dual-mode horn, a dominant-mode circular waveguide is connected to another guide of slightly larger diameter, where modes up to TM_{11} may propagate, via a step transition. The step size is chosen to generate the precise amount of TM_{11} mode from the TE_{11} mode so that when the two modes travel through the flared horn section that follows, the E- and H-plane patterns are equalized. The bandwidth of this horn is limited since the two modes must arrive at the horn aperture in phase even though they have phase velocities that vary differently with frequency.

In the corrugated horn, the single-mode smooth-wall waveguide is connected to a corrugated waveguide that supports only the HE_{11} mode. Some matching between the waveguides is provided by gradually changing from $\lambda/2$ slot depth to $\lambda/4$ slot depth in a short transition region. Throughout the transition region, only the HE_{11} corrugated waveguide mode may propagate, and the E- and H-plane radiation patterns of this mode become nearly equal when a balanced condition is reached (slot depth $\sim\lambda/4$). The bandwidth of this horn is larger than that of the dual-mode horn since the transverse electric-field patterns and hence the radiation pattern of the HE_{11} mode are relatively insensitive to small changes in slot depth around the balanced condition (slot depth $\sim\lambda/4$). After the HE_{11} mode is established in the singe-mode corrugated waveguide, the guide is gradually flared, without changing the slot depth, to the required aperture size.

In analyzing reflector systems, it is necessary to accurately compute the radiation pattern of the feed. Other chapters of this book provide more details on the design of feeds, so only a very brief mention of a commonly used method of corrugated horn analysis is given here.

The corrugated section can be analyzed using a computer code developed in Ref. 28–30. The analysis follows the method of James [31], expanding the fields inside each fin and slot in terms of circular-waveguide modes, and matching the fields at each slot–fin boundary. All of the possible propagating modes, as well as a sufficient number of evanescent modes, are matched at each edge, with results for successive edges and waveguide lengths being cascaded as the calculations proceed through the device. In this way, the interaction between the fields of nonadjacent as well as adjacent slots is taken into account. The result of the calculation is a matrix equation relating the reflected and aperture modes to the input modes. More detailed analysis of corrugated horns can be found elsewhere in this book.The agreement between the modal analysis and the actual feed pattern from a well fabricated corrugated horn is excellent; in fact, if the measured and calculated patterns do not agree, there is usually something wrong with the fabrication of the device.

5.4.2 Spherical-Wave Analysis

The spherical-wave expansion coefficients are normally used as one part of a sequence to analyze an antenna system. The basic role of the technique is to transform far-field feed

patterns to near-field H-field so that PO may be used for reflectors in the near field of their illumination source. Using only the far-field pattern for a feed in analyzing reflector antenna performance can lead to erroneous results, as shown in Ref. 32.

The theory of spherical waves is described in Ref. 33 and will only be briefly summarized here. Any electromagnetic field in a source-free region may be represented by a spherical-wave expansion. In general, the expansion must include both incoming and outgoing waves. If the field satisfies the radiation condition, only outgoing waves will be present, and the expansion will be valid outside the smallest sphere enclosing all sources (the sphere center must be at the coordinate origin used for the expansion). The radial dependence of the spherical waves is then given by the spherical Hankel function, $h_n^2(kR)$.

Typically, the input data, which is used to specify the field, is the tangential E-field on the surface of a sphere. For far-field data, the data sphere radius may be thought to be infinite. The maximum value of the Hankel function index that is needed to closely approximate the field is roughly equal to ka ($ka + 10$ is typical, but in some cases a lower limit will work), where a is the radius of the sphere enclosing all sources.

The computer program most often used for spherical wave analysis is patterned after Ref. 34. Input data is specified on a grid of points defined by the intersection of constant contours of θ and ϕ. The amplitude and phase of E_θ and E_ϕ are given at each point. The minimum number of θ values is roughly 1.2 times the maximum value of n.

The azimuthal dependence of spherical waves is given by $\sin(m\phi)$ and $\cos(m\phi)$. In general, m runs from zero to the maximum value of n. Frequently, symmetry can be used to reduce the number of azimuthal terms. In addition, a conical feed radiates only $m = 1$ modes, and reflection from a body of revolution will maintain this behavior. Also, there is even and odd ϕ dependence, and quite often, only odd or even will be present. For the even case, E_θ can be expanded in only $\sin(m\phi)$ terms and E_ϕ in $\cos(m\phi)$ terms. For the odd case, this is reversed. The minimum number of ϕ values for the data sphere is, in general, $2M + 1$, where M is the maximum value of m. Symmetry can be used to reduce this value appropriately. The output of the software application used to compute these terms is the set of spherical-wave expansion coefficients. These coefficients may then be used to compute the field anywhere within the region of validity. Therefore the essential utility of the program is its ability to take data consisting of the tangential E-field on a sphere (whose radius may be infinite) and provide the means to compute the field—all three components of E and H—at any other point in the region of validity.

5.4.3 Quasioptical Techniques

Multiple reflector systems are typically analyzed by using PO, Gaussian beams, or ray-tracing techniques [35]. Physical optics offers high accuracy at the expense of computation time. This trade-off becomes particularly apparent in multiple-reflector antennas, such as beam waveguides (BWGs), where PO is used to compute the current on each subsequent reflector from the current on the previous reflector. At the other end of the spectrum is ray tracing, which uses methods that ignore diffraction effects entirely. These methods are fast but sacrifice the ability to predict some effects accurately.

An intermediate approach (between using PO and ray tracing) is to use an appropriate set of expansion functions to model the field between the reflectors [36]. If the set is chosen wisely, only a few coefficients need to be determined from each reflector current. The field is then computed at the next reflector using the expansion functions

and their coefficients rather than the previous reflector current. For a BWG system with no enclosing tubes, an excellent set of expansion functions is the Gaussian-beam-mode set. In many cases, a preliminary design, which includes the effects on diffraction, may be obtained by considering only the fundamental mode and a thin-lens model for the reflectors. Higher order modes, which include asymmetric distortion of the beam, cross polarization, and beam truncation, are used to model the effects of the curved reflector.

Each analysis technique is briefly described next, and an example comparing the techniques is shown.

5.4.4 Ray Optics Algorithm

A bundle of rays is launched from the feed point to the first reflector. The ray distribution in angular space is proportional to the power pattern of the input feed. The rays are then traced through the multiple-reflector system. A flat plate is placed a large distance from the final mirror, perpendicular to the central ray. The rays from the final mirror are traced to the flat plate. The density distribution of the rays on the flat plate can be processed to yield the far-field pattern scattered from the last mirror.

5.4.5 Gaussian-Beam Algorithm

A solution to the paraxial wave equation in cylindrical coordinates, with the Laguerre polynomial having indices p and m, is generally called the pm Gaussian-beam mode, and the normalized electric-field distribution is given by

$$\begin{aligned} E_{pm}(r,\varphi,z) = &\left[\frac{2p!}{\pi(p+m)!}\right]^{0.5} \frac{1}{w(z)} \left[\frac{\sqrt{2}r}{w(z)}\right]^{m} L_{pm}\left(\frac{2r^2}{w^2(z)}\right) \\ &\cdot \exp\left(\frac{-r^2}{w^2(z)} - jkz - \frac{j\pi r^2}{\lambda R(z)} + j(2p+m+1)\,\phi_0(z)\right) \\ &\cdot \exp(jm\varphi) \end{aligned} \tag{5.27}$$

where the beam radius w, the radius of curvature R, and the phase shift ϕ are exactly the same as for the fundamental Gaussian-beam mode [37]. Aside from the angular dependence and the more complex radial dependence, the only significant difference in the electric-field distribution is that the phase shift is greater than for the fundamental mode by an amount that depends on mode parameters. Consequently, the fields between the reflectors will be composed of a sum of Gaussian-beam modes.

The steps involved in computing the radiation pattern of the reflector system are as follows:

1. Compute the current on the first reflector using PO. The incident magnetic field is provided either by a feed model or by an incident set of Gaussian-beam modes.
2. Compute the direction of propagation for the reflected Gaussian-beam set using ray tracing. Using a gut ray in the input direction specified by the feed coordinate system or by the input Gaussian-beam set propagation direction and the reflector surface description, compute the gut-ray direction for the output Gaussian-beam set.
3. Determine the waist size and location for the output-beam set.

4. Determine the amplitudes of the individual modes in the output-mode set.
5. Repeat Steps 1–4 for each additional reflector in the chain. In each of these cases, the previous Gaussian-beam set provides the input field for the current calculation.
6. Use PO to compute the far-field pattern radiated by the final mirror.

One of the key steps in computing scattering of Gaussian-beam-mode sets by an arbitrary reflector is determining the best choice for the output-beam set waist and location. The problem is depicted in Figure 5.23, where w_{in} and l_{in} are the input waist and location, and w_{out} and l_{out} are the same parameters for the output-beam set. Once these parameters are determined, computation of the mode amplitudes from the current on the reflector follows easily. The steps involved are as follows:

1. Compute a waist at the reflector, denoted as w_{match}. When the input to the reflector is a Gaussian-beam-mode set, compute the input waist at the point of impact on the reflector. When the input to the reflector is a feedhorn, determine an estimate of the waist from the reflector current. The output-beam set is required to produce a waist equal to w_{match} at the point of impact. This computation provides one of the two equations needed to compute the output-beam-set parameters.
2. Next, select a suitable set of points on the reflector to use for deriving field matching or pathlength matching. These points are generally chosen to be within the waist at the point of impact.
3. Compute a phase center location for the input fields. For a Gaussian-beam-mode set, this is determined from the input radius of curvature at the point of impact. For a feed input, the phase center is assumed to be at the origin of the feed coordinate system.
4. Use the set of points on the reflector to compute the set of pathlengths from the input phase center at each point, as shown in Figure 5.24.
5. Search for the output phase center (output radius of curvature), sweeping along the direction of the output ray. The goal is to determine the minimum in the spread

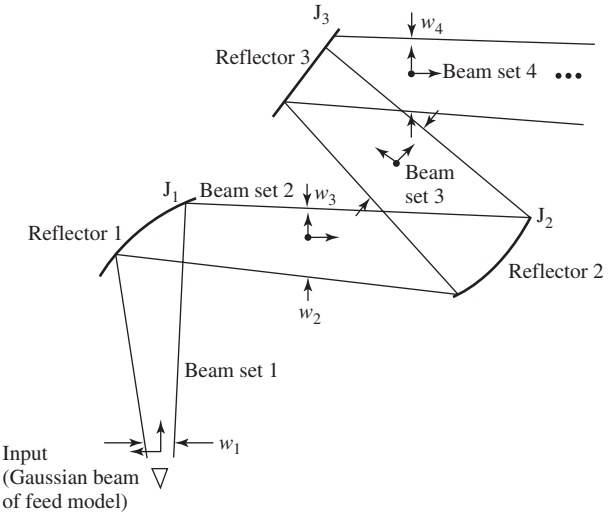


Figure 5.23 Gaussian-beam geometry.

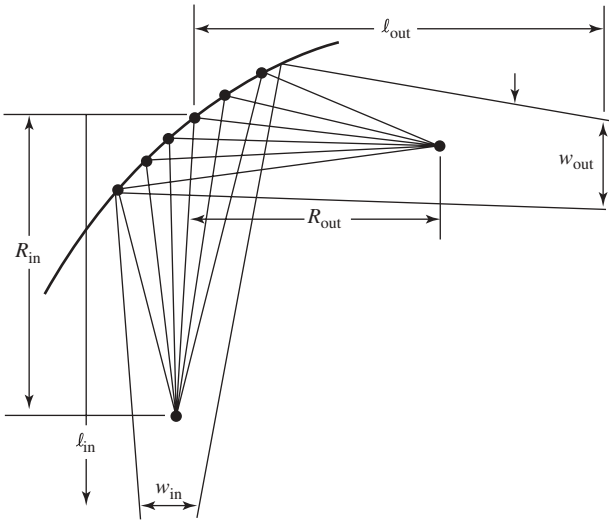


Figure 5.24 Ray-optics approach for determining output beam parameters.

of the total pathlengths from input phase center to reflector and then to output phase center. Generally, a single minimum is found either in front of or behind the reflector.

6. Using the output radius of curvature and the determined waist at the reflector, compute the waist size and location of the output-beam set.

To determine the amplitudes of the Gaussian-beam modes directly from the reflector current, two approximations are necessary: (1) use the reflector current to approximate the PO current—a good approximation for large reflectors with low edge illumination; and (2) use the Gaussian-beam modes as approximations for the solutions to Maxwell's equations in free space—never true but a good approximation if the mode fields are required only in the paraxial region. The Gaussian-beam modes used are given in terms of Laguerre polynomials, as described earlier. Each mode has a polarization, either x or y; a radial index, p; and an azimuthal index, m.

For two arbitrary fields and their associated sources, denoted by A and B, the reciprocity theorem when applied to an arbitrary volume V and its enclosing surface S may be stated as follows:

$$\oiint_S (\mathbf{E}_A \times \mathbf{H}_B - \mathbf{E}_B \times \mathbf{H}_A)\, ds = \iiint_V (\mathbf{E}_B \cdot \mathbf{J}_A - \mathbf{B}_B \cdot \mathbf{M}_A - \mathbf{E}_A \cdot \mathbf{J}_B + \mathbf{B}_A \cdot \mathbf{M}_B) dv \tag{5.28}$$

A half-space completely enclosing the reflector is chosen as V, with the surface S perpendicular to the direction of propagation for the output-beam set. For this particular application, choose the output Gaussian-mode set, with unknown mode amplitudes, as the A field with the reflector current inside the volume being its source. As the B field, choose a test field, the conjugate of the ith Gaussian-beam mode now propagating toward the reflector. The source for this field is chosen to be outside V. The fields are then

given by

$$\mathbf{E}_A = \sum\nolimits_j a_j \mathbf{e}_j, \quad \mathbf{H}_A = \sum\nolimits_j a_j \frac{\hat{\mathbf{z}}_{\text{out}} \times \mathbf{e}_j}{\eta_0}, \quad \mathbf{E}_B = \mathbf{e}_i^*, \quad \mathbf{H}_B = -\frac{\hat{\mathbf{z}}_{\text{out}} \times \mathbf{e}_i^*}{\eta_o} \tag{5.29}$$

Using the reciprocity theorem and the orthogonality condition for the Gaussian-beam modes on the infinite surface S, the desired equation for the unknown coefficients is:

$$a_i = -\frac{1}{2} \iint_{S_{\text{reflector}}} \mathbf{e}_i^* \cdot \mathbf{J}\, ds \tag{5.30}$$

5.4.6 Physical Optics Algorithm

The PO algorithm uses the classical PO technique. The currents on the first mirror are computed from the incident field of the feed. The currents on the subsequent mirrors are derived from the fields radiated from the currents on the previous mirror. The currents from the last mirror radiate to produce the output far-field pattern. It is assumed that only the subsequent mirror can see currents from the immediately proceeding mirror; that is currents from earlier mirrors are blocked from it.

5.4.7 Example Comparing the Three Algorithms

As an example, consider the three-curved-mirror BWG system used in the NASA/Jet Propulsion Laboratory 34-m research and development antenna described in Chapter 7 of Ref. 8. The BWG consists of a beam-magnifier ellipse followed by an imaging pair of parabolas. The system is designed to operate from 2 to 35 GHz. Figures 5.25 and 5.26 compare the three methods at 2.3 and 32 GHz. Of course, the ray-optics technique is frequency independent.

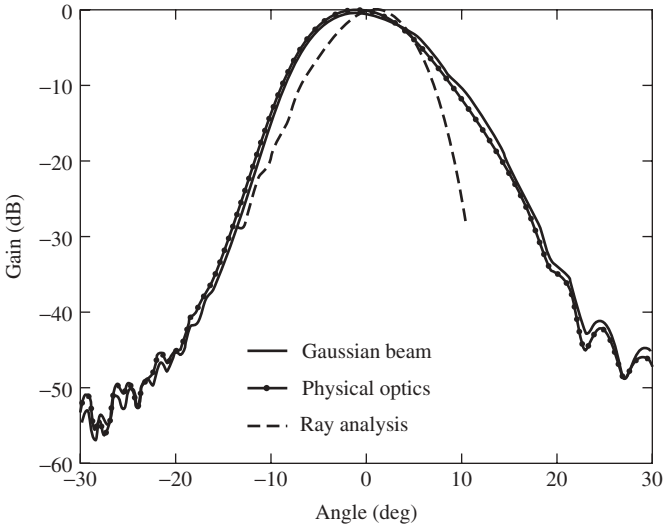


Figure 5.25 Comparison of three methods at 2.3 GHz.

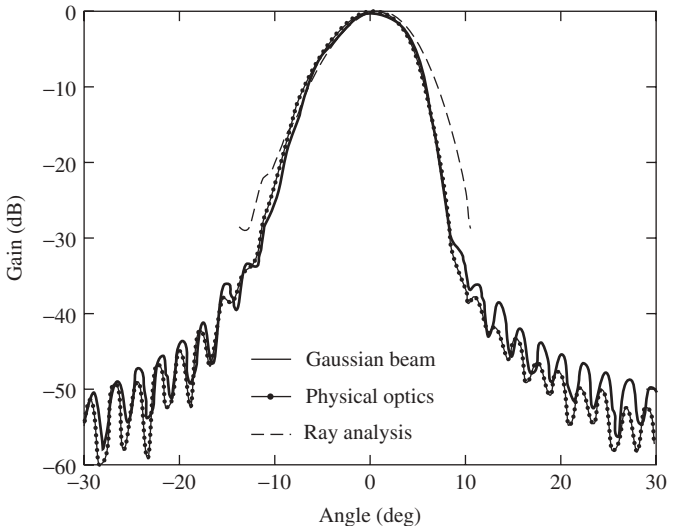


Figure 5.26 Comparison of three methods at 32 GHz.

As can be seen from the figures, the PO technique is the most accurate. However, it is also the most time consuming. On the other hand, for a BWG system, higher order Gaussian modes provide a very good estimate of performance and require minimal computer time. Ray optics is generally not adequate for diffraction pattern calculations.

5.4.8 Dichroic Surfaces

The ability to transmit and receive simultaneously at multiple frequency bands is an important requirement for many antenna systems. Achieving the requirement is usually accomplished by using either a multiband feedhorn or a separate feedhorn along with a frequency-selective surface (FSS) on the antenna—typically referred to as a dichroic reflector. Dichroic reflectors are important components of both ground and spacecraft antennas. The most frequently used type of surface configuration for ground antennas consists of holes in a flat metal plate; the holes allow the higher frequency to pass through the surface but prevent the lower frequency from doing so, reflecting that frequency from the surface (see Ref. 8 for the analysis of frequency-selective surfaces). However, spacecraft antenna designs typically require dichroic subreflectors that reflect the *higher* frequencies and pass the *lower* frequencies; this type of dichroic surface usually consists of a periodic array of thin conductors on a dielectric sheet. Examples are the Voyager, Galileo, and Cassini [9] high gain antennas.

One of the earliest (∼1970) uses of a dichroic surface on a ground antenna was as part of the reflex–dichroic feed on NASA's Deep Space Network 64-m antenna, as shown in Figure 5.27. Its dichroic surface is a half-wave-thick plate of nearly circular (actually Pyle-shaped) holes. The plate reflects S-band (∼2.3 GHz) and passes X-band (∼8.4 GHz) (Figure 5.28). An early dichroic *subreflector* designed for the Tracking and Data Relay Satellite System (TDRSS) was designed to pass S-band (∼2.3 GHz) and reflect Ku-band (11.7–15.121 GHz); see Figure 5.29, which appears in Ref. 38.

A very thorough discussion of the analytical techniques as well as more examples of the use of dichroic surfaces can be found in Ref. 39.

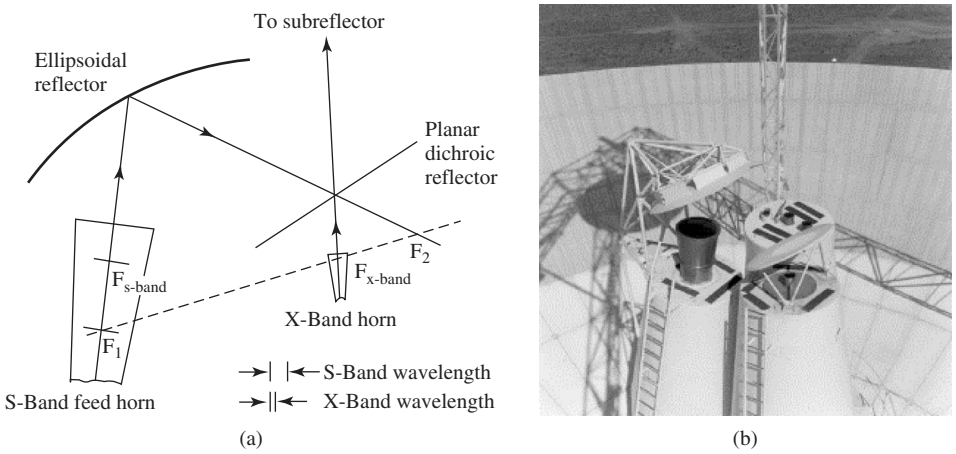


Figure 5.27 Reflex-dichroic feed system: (a) geometry and (b) installed on the 64-m antenna.

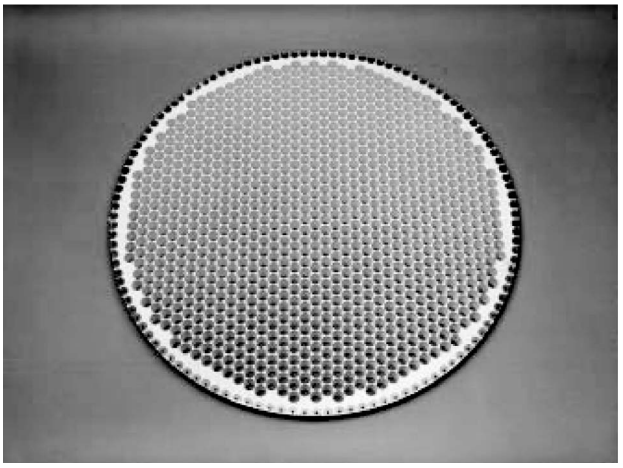


Figure 5.28 Half-wave-thick dichroic filter.

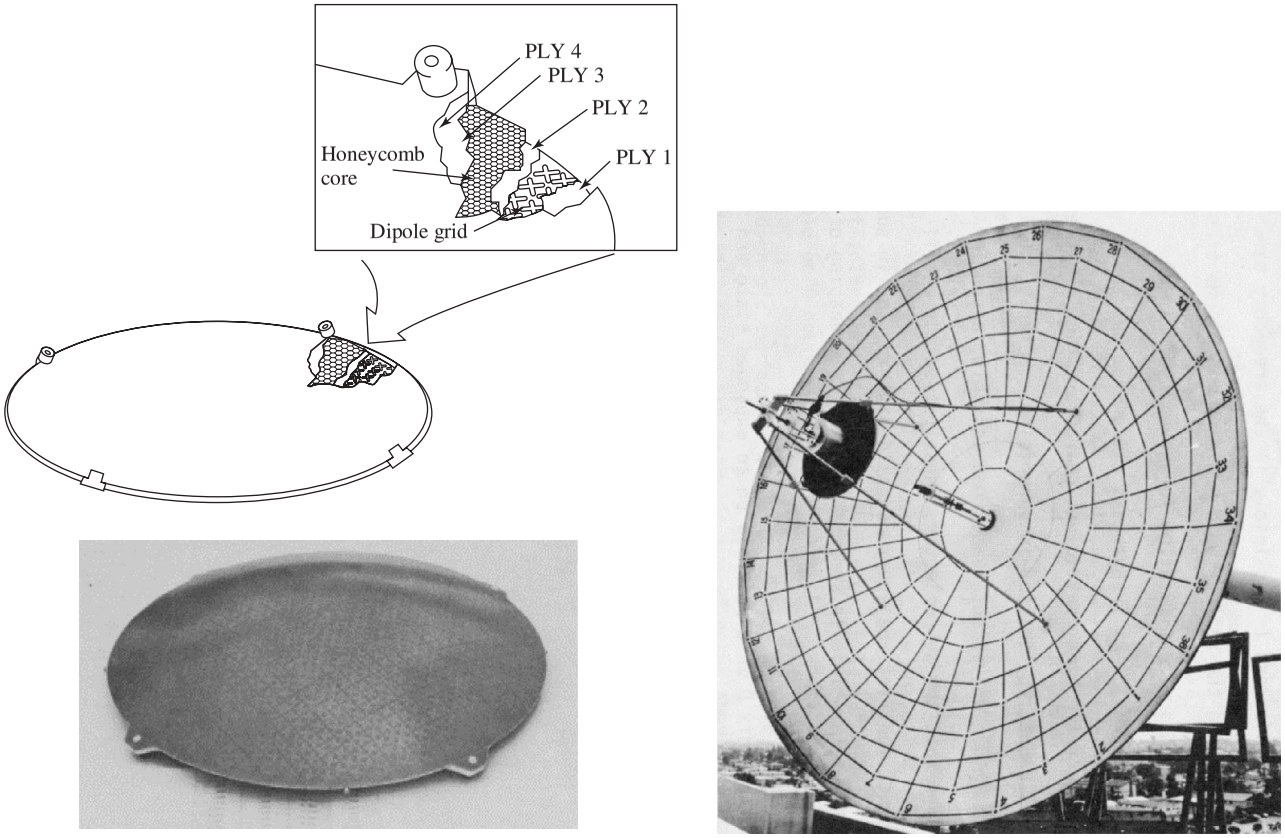
5.4.9 Antenna Noise-Temperature Determination

The system noise temperature of a receiving system is the sum of the receiver noise temperature and the antenna temperature. Individual contributions to the antenna temperature include cosmic radio noise, atmospheric absorption, thermal radiation from the ground, and transmission-line loss.

The noise temperature of an antenna is defined as

$$T_a = \frac{P_A}{kB} \quad \text{(K)} \tag{5.31}$$

where P_A is the noise power delivered by the antenna, over a bandwidth B, into a matched termination, and k is Boltzmann's constant equal to 1.380×10^{-23} J/K.



(a) (b)

Figure 5.29 Dichroic subreflector: (a) geometry and (b) installed on a 2.74-m dish.

Also, the effective antenna noise temperature can be expressed as the radiation pattern-weighted environmental black body temperature, $T(\Omega)$, averaged over all directions; thus

$$T_a = \frac{1}{4\pi}\int_{4\pi} T(\Omega)G(\Omega)\,d\Omega = \frac{\int_{4\pi} T(\Omega)P(\Omega)\,d\Omega}{\int_{4\pi} P(\Omega)\,d\Omega} \tag{5.32}$$

where $G(\Omega)$ is the antenna gain in the direction Ω and $P(\Omega)$ is the normalized antenna pattern gain in direction Ω or

$$P(\Omega) = \frac{G(\Omega)}{G_m} \tag{5.33}$$

with G_m the antenna maximum gain.

Evaluation of Eq. (5.32) may range from rough approximations or rules of thumb to a very painstaking evaluation of the functions involved using numerical integration.

5.5 SINGLE-BEAM ANTENNAS

Single-beam antenna types are (1) paraboloid reflectors with focal-point feeds, (2) dual-reflector antennas such as the Cassegrain, Gregorian, and those with dual-shaped configurations, (3) BWG antennas, (4) offset-fed paraboloid antennas, and (5) offset-fed multiple-reflector antennas. Each of these antenna types has its own unique characteristics, and the advantages and disadvantages of each have to be considered when choosing one for a particular application.

5.5.1 Prime-Focus Fed Antennas

The material in Section 5.3 of this chapter provides comprehensive information about the design of prime-focus-fed antennas for single-frequency or narrowband configurations. More interesting, due to recent developments in wideband amplifiers, is the case for very wideband feeds. An example of a wideband system is the Allen Telescope Array (California) [40] that uses a log-periodic feed to cover the 0.5–11-GHz frequency range. The Square Kilometer Array (yet to be built) [41] wants to cover the 100-MHz to 25-GHz range, and the Goldstone Apple Valley Radio Telescope (California) [42] the 2–14-GHz range.

An interesting example of optimum design curves of a log-periodic dipole antenna (LPDA) used as a prime-focus feed is presented in Ref. 43. LPDA geometry (see Figure 5.30) readily lends itself to an accurate computer solution of its gain, voltage standing wave ratio (VSWR), and radiation patterns, and once the primary patterns of the LPDAs are calculated, it is a straightforward procedure to incorporate these patterns into the calculation of the reflector secondary radiation patterns and gain.

The LPDA is attractive as a broadband feed for parabolic reflectors despite its tendency to cause defocusing because of axial phase-center movement with frequency. The gain loss due to defocusing in a parabolic reflector is only a function of the defocusing in terms of wavelengths and is not dependent on the diameter of the reflector. Hence antennas such as the LPDA that have a frequency-independent phase center location in terms of wavelengths from some geometrical point (virtual apex) will experience a

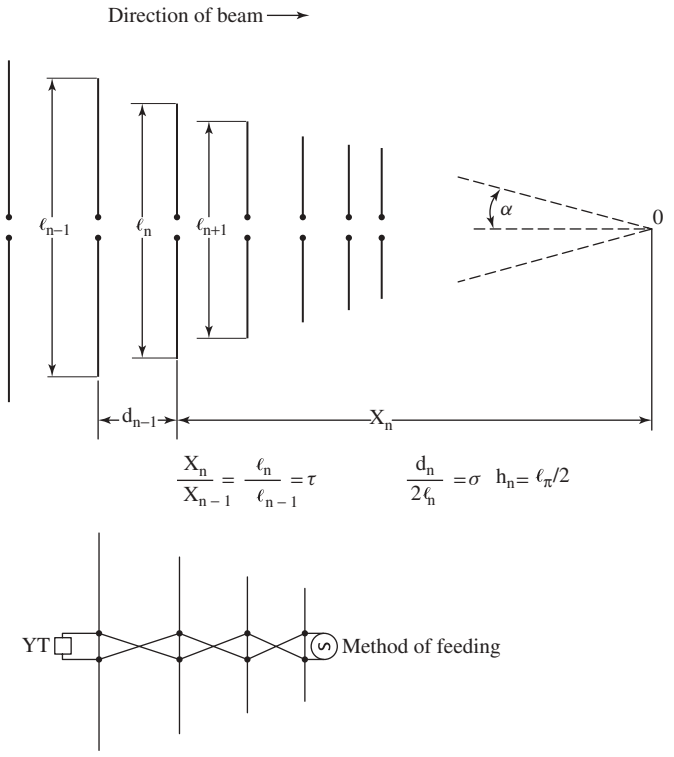


Figure 5.30 Log-periodic dipole antenna geometry.

frequency-independent defocusing loss over their bandwidth if the virtual apex is placed at the focal point. The optimum feed design for maximum gain will change with varying reflector F/D (focal length-to-diameter ratio) because spillover, aperture illumination, and defocusing loss are functions of F/D; however, primary pattern beamwidth and symmetry, gain, and VSWR are functions of the feed parameters. If the LPDA is to be used for narrowband operation, the phase center is placed at the focal point, and the defocusing loss is eliminated. If it is used for broadband operations, the apex is placed at the focal point. Thus the feed designs that optimize gain will be different for narrowband and broadband cases.

For a center-fed parabola, the radiation pattern function for scalar aperture theory is given by (see Figure 5.8)

$$g(u, \phi) = \frac{a^2}{2F} \int_0^{2\pi} \int_0^1 (1 + \cos\theta) G(\theta, \psi)\, e^{jur\cos(\phi - \psi)}\, e^{-j2\pi Z \cos\theta} r\, dr\, d\psi \qquad (5.34)$$

where Z is the defocusing in wavelengths from the focus F and is negative toward the vertex of the parabola, a is the radius in wavelengths, $G(\theta, \psi)$ is the primary voltage pattern, and $u = (2\pi a/\lambda)\sin\theta'$.

In general, the E- and H-plane patterns of the LPDA are different. Assume a primary feed pattern of the form

$$G(\theta, \psi) = \frac{\cos^n \theta \sin^2 \psi + \cos^m \theta \cos^2 \psi}{\sqrt{D_3}} \tag{5.35}$$

This representation is termed the "integral approximation" since the integration in ψ upon substituting Eq. (5.35) into Eq. (5.34) can be done in closed form. The result is

$$g(u, \phi) = \frac{\pi a^2}{F} \int_0^1 (1 + \cos\theta) J_0(ur) f(\theta, \phi) e^{-j2\pi Z \cos\theta} r \, dr \tag{5.36}$$

with

$$\sqrt{D_3}\, f(\theta, \phi) = \frac{\cos^n \theta + \cos^m \theta}{2} + \frac{1}{2} \frac{J_2(ur)}{J_o(ur)} \cos 2\phi \left\{ \cos^n \theta - \cos^m \theta \right\}$$

Determining the E- and H-plane patterns using the techniques described in Ref. 43 results in a very simple representation of the radiation pattern and gain of the LPDA feed in a center-fed parabolic reflector.

5.5.2 Design Considerations

The objective of this section is to demonstrate the use of numerical solutions to develop design data for optimizing the gain of parabolic reflectors fed with LPDAs operating over broad or narrow bandwidths.

In general, the gain of LPDAs depends on the parameters τ, σ, and α tending toward higher gains, with large τ, and large σ, and small α. Recall that the gain-to-loss ratio due to defocusing in a parabolic reflector for a given F/D is only a function of the defocusing in terms of wavelength and is not dependent on the diameter of the reflector. Hence antennas that have a frequency-independent phase-center location in terms of wavelengths from some geometrical point (virtual apex that represents the extension of the structure to an infinitely high frequency) will experience a constant frequency-independent loss over its bandwidth if this point is placed at the focus.

Defocusing loss increases with distance. A typical value for a 0.5λ defocusing and an F/D of 0.4 would be about 1.0 dB. However, the smaller the α angle, the larger the displacement of the phase center from the virtual apex and therefore the greater the defocusing loss. Since loss due to defocusing with small α angles is more severe than aperture distribution losses, antennas with larger α angles are favored for broadband design. For generating design curves, however, two separate cases will be examined: (1) infinite bandwidth, where the virtual apex is placed at the focal point and the defocusing loss is included, and (2) zero bandwidth, where a single frequency is considered and the feed is positioned so that there is no defocusing loss.

It is well known that for a given primary illumination function, there is an F/D for which the gain is a maximum. Also, the effect of defocusing is different for varying F/D. Hence it is to be expected that there is a different LPDA parameter selection that will optimize the reflector gain for each F/D.

The design procedure is straightforward. A technique such as described in Ref. 43 is used to compute primary feed patterns. These primary patterns are then used as inputs for the parabolic scattering analysis to compute the secondary far-field radiation.

The parameters of an LPDA were studied as a function of secondary on-axis gain. Because the LPDA has cyclic gain perturbations as well as front- and rear-end truncation effects, a log-cycle in the center of the antenna bandwidth was chosen for the parameter study. The number of dipoles was also chosen large enough to include elements larger than 3/2 wavelengths such that the effects of energy leaking past the first active region would be included.

It is desirable to derive design curves that provide stable performance over the whole bandwidth. However, while some parameters actually will produce a higher gain over a portion of each log-period, they may have gain dips over some other portion of each period and are thus not acceptable.

Several results of this analysis are worth particular mention. First, the half-power beamwidth variations of an LPDA are not necessarily reliable indicators of secondary gain stability. That is, while for some parameters the half-power primary beamwidth showed very large excursions, the secondary gain variations were no larger than for other parameter ranges where the primary pattern beamwidth was more stable. A possible explanation is that a large part of the LPDA beamwidth variation is due to energy that has leaked past the normal active region and excited the 3/2 mode elements. These elements, however, show a much larger defocusing loss and hence their contribution to the secondary gain is markedly reduced.

While it is desirable to have a more stable primary pattern, it is much harder to obtain $Z_0 = 300\ \Omega$ than $Z_0 = 150\ \Omega$, so construction techniques tend to make lower values more attractive. However, if the value of Z_0 is too low, say, less than 100 Ω for large α angles and low τ, serious gain anomalies occur at some portion of the log-periodic frequency. A characteristic feed line impedance of $Z_0 = 150\ \Omega$ is generally sufficient to overcome this problem.

5.5.3 Optimum Design

Of primary importance for achieving optimum design is finding the major design parameters σ, and τ. Only two of the parameters are independent, the third being related to the other two by $\tan \alpha = (1-\tau)/4\sigma$.

For a given F/D, curves were plotted for secondary gain (relative to uniform aperture gain) versus α, with τ as a parameter. For values of $\sigma > 0.03$, there was a value of α, independent of τ, that gave the highest gain. These curves were made for each F/D, and the optimum value of α versus F/D was obtained, as shown in Figure 5.31. What is needed next is to optimize the τ or σ choice. In general, for a fixed α, there is a small increase in directivity with increasing τ. However, for fixed α, increasing τ gives a smaller σ, and the VSWR with respect to the mean impedance increases with decreasing σ. Therefore, if the VSWR is to be kept small, there is a lower bound on acceptable σ. For reasonable parameter ranges of τ $(0.8 < \tau < 0.95)$, this lower bound is considered to be $\sigma = 0.4$. Below this value of σ, the VSWR is generally > 2:1, and the pattern performance is unstable. The acceptable lower bound is also a function of the characteristic impedance Z_0.The best value of σ or τ can be obtained from Figure 5.32, where σ versus τ, with α as a parameter, is shown. For the infinite bandwidth case the optimum values of σ are from $\sigma = 0.04$ to 0.06, with the best choice of σ depending on the exact h/a and Z_0 desired. For the optimum α, σ changes in this range cause only small changes in the secondary gain, with the gain decreasing as σ increases. In general, the exact selection of σ will depend on Z_0. Since low values of Z_0 (< 150 Ω) and σ tend toward unstable designs, higher σ values should be selected for lower Z_0.

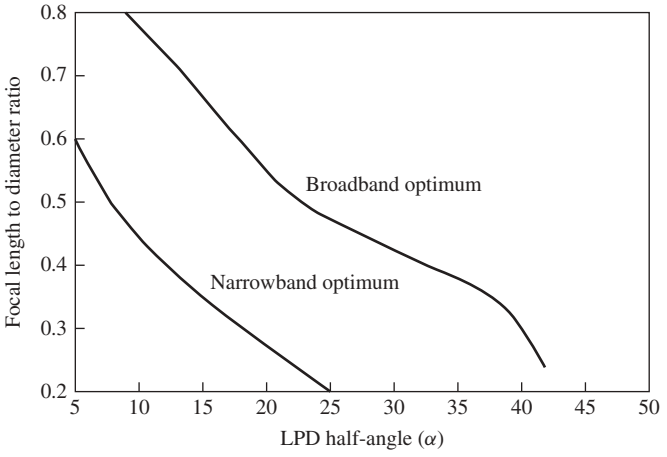


Figure 5.31 Optimum LPD half-angle α versus F/D.

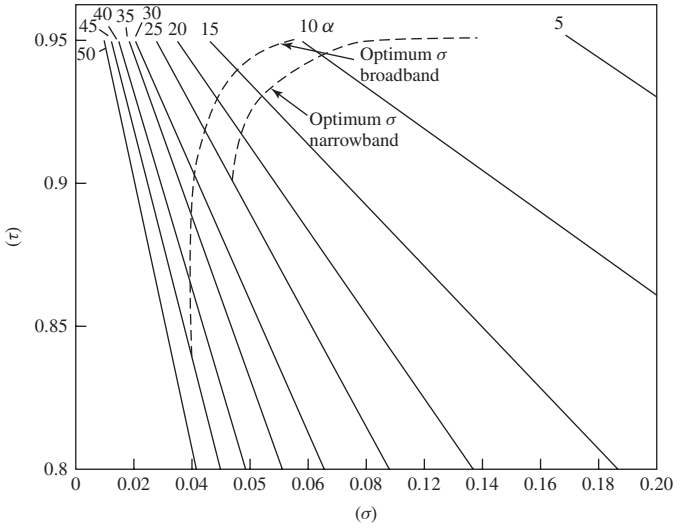


Figure 5.32 Optimum values of α for maximum secondary gain.

Once the major design parameters σ, τ, and α are selected, the other factors influencing the design of the LPDA are selected. Some example cases are shown in Ref. 43.

5.5.4 Offset-Fed Parabolic Reflector Antennas

The geometry of an offset-fed parabolic reflector antenna is shown in Figure 5.33. The major advantage of this type of reflector is that the feed does not block the radiating aperture. This is important for systems that use large multiple-feed systems or that require low sidelobes (no feed support blockage). The major disadvantage is that there is larger cross polarization for linear polarization, resulting in a beam squint for circular polarization.

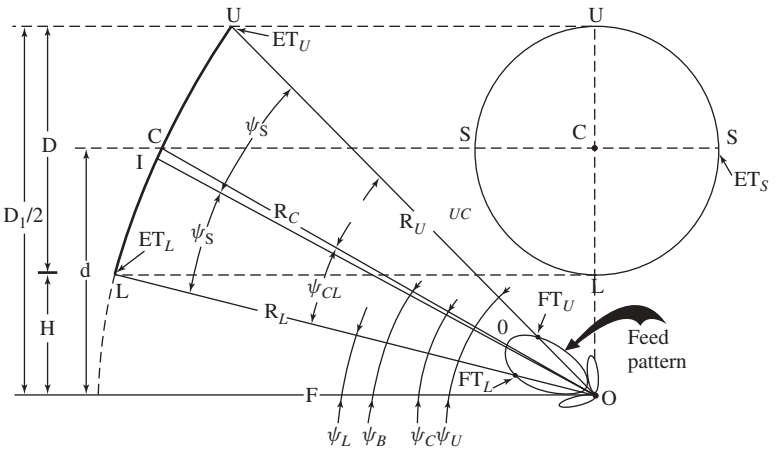


Figure 5.33 Offset-fed parabolic reflector antenna.

The reflector can be defined either by the linear parameters, $F =$ focal length, $D =$ reflector diameter and $H =$ offset height; or by the angular parameters, F, $\psi_B =$ offset angle, and $\psi_S =$ half-subtended angle. The angular parameters can be obtained from the linear parameters

$$\begin{aligned} \psi_B &= \tan^{-1}\frac{D+H}{2F} + \tan^{-1}\frac{H}{2F} \\ \psi_S &= \tan^{-1}\frac{D+H}{2F} + \tan^{-1}\frac{H}{F} \end{aligned} \tag{5.37}$$

and the linear from the angular

$$\begin{aligned} D &= \frac{4F\sin\psi_S}{\cos\psi_B + \cos\psi_S} \\ H &= \frac{2F(\sin\psi_B - \sin\psi_S)}{\cos\psi_B + \cos\psi_S} \end{aligned} \tag{5.38}$$

From a design viewpoint, one generally chooses D for gain, H/D for clearance, and F/D or F/D_p for scan performance.

Feed and edge tapers are important parameters for characterizing the effects of feed patterns on reflector patterns. Referring to Figure 5.34, the feed taper is given by

$$FT = 20\log\frac{C(\angle\mathbf{R}_U - \angle\mathbf{R}_C)}{C(0^\circ)} \tag{5.39}$$

and the edge taper by

$$ET = FT + \underbrace{20\log\frac{R_C}{R_U}}_{\text{path loss term}} \tag{5.40}$$

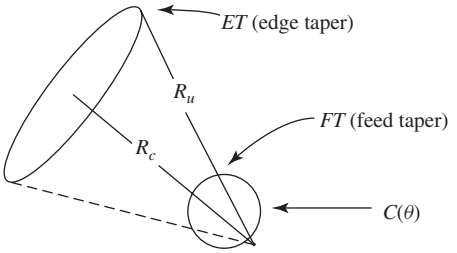


Figure 5.34 Feed and edge taper geometry.

The edge taper directly controls the reflector's amplitude taper and therefore has a more dominant effect on performance than the feed taper. Offset antennas that have the same edge taper as symmetric antennas have radiation patterns similar to those of symmetric antennas.

As stated earlier, the linear polarization term is larger for offset antennas than for symmetric antennas, resulting in a beam squint for circular polarization. A plot of the maximum cross polarization of an offset parabolic reflector illuminated by a balanced linearly polarized feed is shown in Figure 5.35. Chu and Turrin first published computed graphical data showing the effect of the reflector parameters ψ_B and ψ_S on the magnitude of the beam squint [44]. Adatia and Rudge have derived an approximate formula [45]

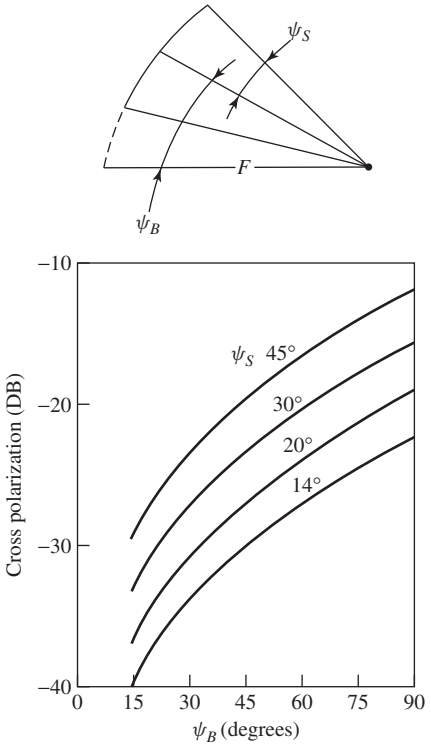


Figure 5.35 Maximum cross polarization of offset parabolic antenna.

that gives the beam squint angle θ_S simply as

$$\sin\theta_S = \mp\frac{\sin\psi_B}{4\pi(F/\lambda)} \tag{5.41}$$

This formula has been tested against computed and measured data and has been found, in all cases, to be accurate within 1.0% of the antenna half-power beamwidth. An example of the squint is shown in Figure 5.36.

As stated earlier, the major advantage of an offset configuration is that the feed system does not block the aperture. Therefore offset configurations are most often used with multiple feeds that can either be used to generate a shaped amplitude radiation pattern or multiple beams. This subject is covered later in the chapter.

5.5.5 Axisymmetric Dual-Reflector Antennas

Classical Cassegrain-antenna geometry [46, 47] employs a parabolic contour for the main reflector and a hyperbolic contour for the subreflector (Figure 5.37). The parabolic reflector is a point-focus device with a diameter D_m and a focal length F_m. The hyperbolic subreflector has two foci. For proper operation, one of the two foci is the real focal point of the system and is located coincident with the phase center of the feed; the other focus, the virtual focal point, is located coincident with the focal point of the main reflector. On the other hand, classical Gregorian antenna design uses an elliptical contour for the subreflector with the feed at one focus of the ellipse and the main reflector focal point at the other. The major difference between the two systems is that the subreflector is between the feed and the main reflector focal point for the Cassegrain design, which

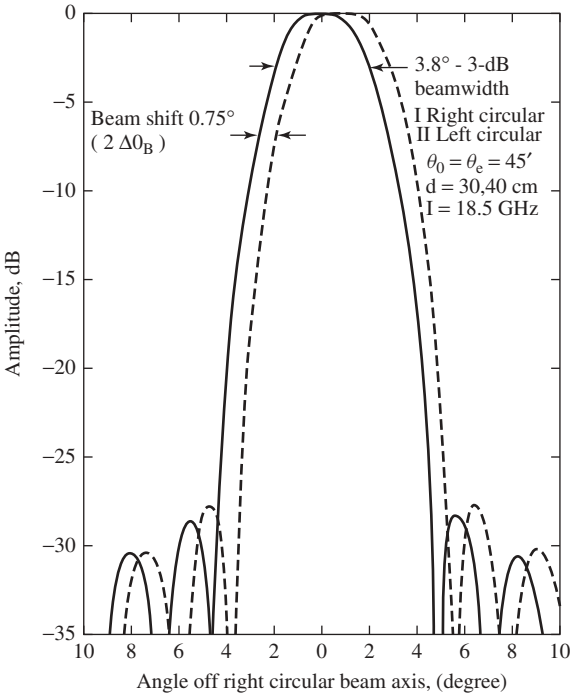


Figure 5.36 Example of beam squint in offset reflectors.

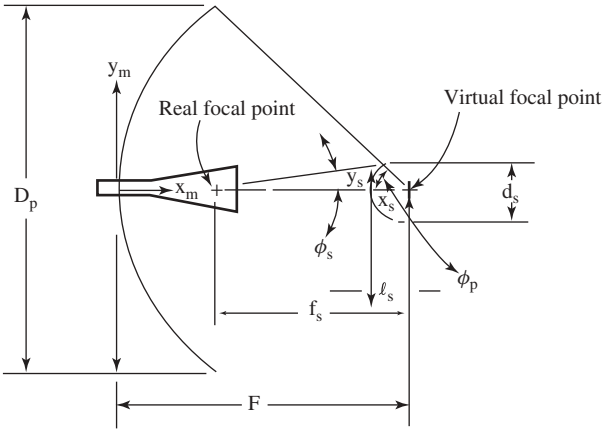


Figure 5.37 Geometry of the Cassegrain antenna system.

is not the case for the Gregorian design. Consequently, the Gregorian system can be operated in the prime focus mode without removing the subreflector. There are several examples of this, the largest being the 100-m radio telescope in Bonn, Germany.

Granet has presented formulas for designing axially symmetric Cassegrain or Gregorian dual-reflector antennas from combinations of prescribed geometric patterns; see Ref. 48 if there is a need to design such a system.

5.5.6 Dual-Reflector Shaping

The efficiency of classic dual reflectors is primarily determined by the ability of the feed system to illuminate only the reflectors while minimizing the energy that radiates elsewhere and the ability of the feed to uniformly illuminate the parabola. The former is termed "spillover efficiency" and the latter "illumination efficiency." Illumination efficiency is 100% when the energy density on the entire main reflector aperture is a constant.

Feed patterns gradually taper from their central maxima to nulls. If all this energy is intercepted by the reflector (for maximum spillover efficiency), the illumination is far from uniform, and the illumination efficiency is very poor. Consequently, any attempt to obtain nearly uniform illumination will result in a great loss of energy in spillover. Ultimately, a compromise must be made. A common choice for both a prime focus system and the Cassegrain system is a 10-dB taper of the illumination pattern at the parabolic edge. This selection results in a combination of spillover and illumination efficiency of about 75–80%.

It is possible, however, to change the shape of the two reflectors to alter the illumination function and improve efficiency. This methodology is termed dual-reflector shaping and was first introduced by Galindo, who demonstrated that one could design a dual-reflector antenna system to provide an arbitrary phase and amplitude distribution in the aperture of the main reflector [49]. Thus if one chose uniform amplitude and constant phase, 100% illumination efficiency could be achieved. When the feed pattern is given, the appropriate choice for subreflector illumination angle would be to have very small spillover.

5.5.7 Theoretical Solution

The complete solution to reflector shaping can be found in Refs. 49–51; only the uniform aperture case is summarized here.

The geometry of the symmetric dual-reflector system is shown in Figure 5.38. Due to circular symmetry, the synthesis reduces to the determination of the intersection curve of the surface with the plane through the axis of symmetry (i.e., x-y plane).

The synthesis method utilizes the analytical expressions of GO principles together with the geometry for the reflectors to develop a pair of first-order nonlinear ordinary differential equations of the form

$$\frac{dy}{dx} = f(x, y) \tag{5.42}$$

This leads to cross sections of each reflector when subject to boundary conditions such as

$$y(x = x_{\max}) = 0 \tag{5.43}$$

The optical principles used to develop the required equations are (1) Snell's law, (2) conservation of energy flow along ray trajectories, and (3) surfaces of constant phase that form normal surfaces to the ray trajectories.

The feed is assumed to have a spherical phase function, that is a phase center and a power radiation pattern, F/θ. For uniform phase in the aperture, the pathlength $r + r' + r''$ must remain constant for all θ. Also, the amplitude function in the aperture $I(x)$ must also be equal to a prescribed distribution or a constant for maximum peak gain.

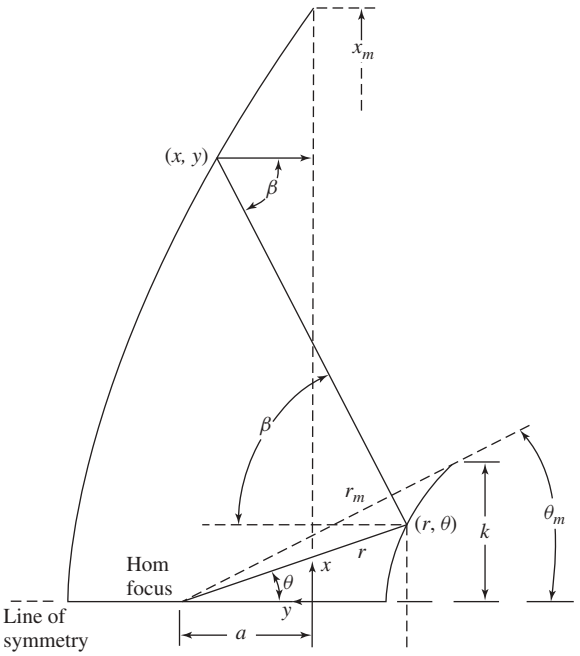


Figure 5.38 Coordinate system for shaping.

The equation for equal pathlengths in the phase front is obtained from trigonometry:

$$r + y + \frac{x - r\sin\theta}{\sin\beta} = C(\text{constant}) \tag{5.44}$$

where (x, y) and (r, θ) are the coordinates of points on the main reflector and subreflector, respectively.

The application of Snell's law to the two surfaces defines a relationship between the angles shown and the first derivates (slopes) of the surfaces. These are

$$\frac{1}{r}\frac{dr}{d\theta} = \tan\frac{\theta + \beta}{2} \tag{5.45}$$

$$\frac{-dy}{dx} = \tan\frac{\beta}{2} \tag{5.46}$$

Since the dual-reflector system is symmetrical about the y-axis, the total power within the increment $d\theta$ of the pattern $F(\theta)$ will be $F(\theta)2\pi\sin\theta\ d\theta$. Similarly, the total power within the increment dx of the main antenna aperture is $I(x)2\pi\ dx$, where $I(x)$ is the illumination function of the antenna aperture. Making $I(x)$ constant and equating the total power from $\theta = 0^\circ$ to angle θ to that within x, and normalizing by the total power, one obtains

$$x^2 = x_{\max}^2 \frac{\int_0^{\theta} F(\theta)\sin\theta\, d\theta}{\int_0^{\theta_{\max}} F(\theta)\sin\theta\, d\theta} \tag{5.47}$$

These four equations now have five dependent variables: x, y, r, θ, and β and can be solved to provide equations for the surfaces. The designs can be "Cassegrainian" (i.e., rays do not cross) or "Gregorian" (i.e., rays do cross).

The first antenna in the Deep Space Network (DSN) to use dual-reflector shaping was the High-Efficiency (HEF) Antenna, whose design is discussed in Chapter 6 of Ref. 8. Although the above solution yields an optimum-gain antenna, the figure of merit of a DSN antenna is gain over noise temperature (G/T). Dual-reflector shaping is based on geometric optics (GO) but due to diffraction, the feed energy scattered from the subreflector does not fall abruptly to zero at the edge of the main reflector. By choosing the design parameters such that the geometric ray from the subreflector edge does not go all the way to the outer edge of the main reflector, it is possible to have the feed energy at a very low level relative to the central region of the main reflector, and the resulting rear spillover noise contribution becomes acceptably small. Applying this approach results in slightly lower illumination efficiency, but the significant reduction in noise from spillover results in an optimum G/T ratio. The part of the main reflector outside the edge geometric ray from the subreflector is sometimes referred to as a noise shield.

5.5.8 Offset-Fed Dual-Reflector Antennas

For the same advantage (no feed blockage), a dual-reflector system can be designed with offset optics. The designs can also be Cassegrain, Gregorian, or dual-shaped.

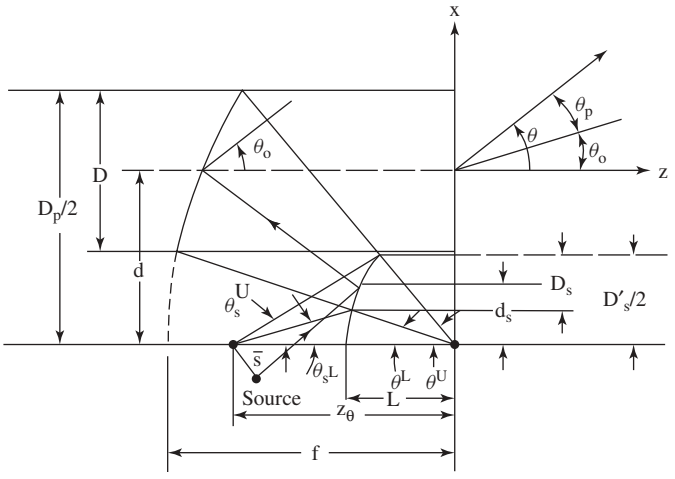


Figure 5.39 Offset Cassegrain reflector.

A typical offset Cassegrain reflector is shown in Figure 5.39. The geometrical parameters for the main reflector are

$$\theta^B = (\theta^U + \theta^L)/2 = \text{bisecting angle}$$
$$\theta^C = (\theta^U - \theta^L)/2 = \text{semi-angle}$$
$$D = \frac{4F \sin\theta^C}{\cos\theta^B + \cos\theta^C}$$
$$d = \frac{2F \sin\theta^B}{\cos\theta^B + \cos\theta^C}$$

where θ^B is the bisecting angle, θ^C is the semi-angle, D is the diameter, F is the focal length, and d is the offset height of the center.

The geometrical parameters for the subreflector are

$$e = \frac{\sin\frac{1}{2}(\theta^U + \theta_S^U)}{\sin\frac{1}{2}(\theta^U - \theta_S^U)}$$
$$M = \frac{e+1}{e-1} \Rightarrow e = \frac{M+1}{M-1}$$
$$\frac{Z_0}{D'_S/2} = \frac{1}{\tan\theta^U} + \frac{1}{\tan\theta_S^U}$$
$$\frac{Z_0}{D'_S/2 - D_S} = \frac{1}{\tan\theta^L} + \frac{1}{\tan\theta_S^L}$$
$$2\frac{L}{Z_0} = 1 - \frac{\sin\frac{1}{2}(\theta^U - \theta_S^U)}{\sin\frac{1}{2}(\theta^U + \theta_S^U)}$$

where e is the eccentricity, M is the magnification factor, D_S is the diameter of the subreflector, z_0 is the interfocal distance, and L is the distance from the ellipse intersection with the axis to the main reflector focal point.

An interesting concept for understanding optics performance is the equivalent parabola (e.g., see Ref. 52); however, since computers are so fast today, it is not often used to compute performance.

5.5.9 Reduction of Cross Polarization

It is well known [53–56] that offset dual-reflector antennas can be designed with perfect rotational symmetry and linear polarization (i.e., with no cross-polarization component) to within the GO approximation. This can be accomplished with a paraboloid main reflector and either an ellipsoid or hyperboloid subreflector—provided that the eccentricity of the subreflector and the relative orientations of the axes of the feed, subreflector, and main reflector satisfy what is termed the basic design equation. Mizuguchi derived one form of the basic design equation using a somewhat lengthy argument based on obtaining an equation expressing the image on the main reflector aperture of an arbitrary ray from the feed center [54], while an equivalent form was obtained by Dragone using a conformal mapping argument [55]. Shore derived the equation using a simple geometric argument [56]. Consider the convex hyperboloidal geometry shown in Figure 5.40. The basic equation for this case is

$$\tan\left(\frac{\alpha}{2}\right) = \frac{e+1}{|e-1|}\tan\left(\frac{\beta}{2}\right)$$

The basic design equation is similar for the other geometries of a concave hyperboloidal system and a convex or concave ellipsoidal system.

5.5.10 Offset Dual-Reflector Shaping

The offset geometry will have higher efficiency than the symmetric geometry because the central blockage due to the subreflector can be eliminated. However, the equations for dual-reflector shaping are significantly more complicated than those for the symmetric designs. Nonetheless, the exact solution has been developed for offset geometries

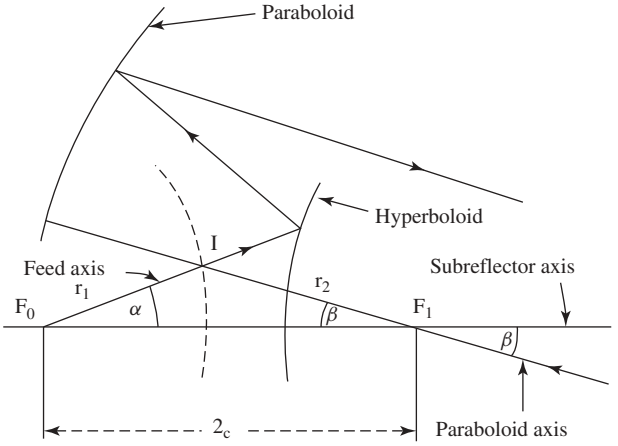


Figure 5.40 Geometry of convex hyperboloidal system.

[57, 58]. An offset geometry was designed and built that had an efficiency of 84.5%, the highest ever recorded [59].

5.5.11 Beam-Waveguide Feed Systems

A BWG feed system is composed of one or more feedhorns with a series of flat and curved mirrors arranged so that power can be propagated from the horn through the mirrors with minimum losses. Horns and equipment can then be located in a large, stable enclosure at an accessible location.

Feeding a large, low noise ground antenna via a BWG system has several advantages over placing the feed directly at the focal point of a dual-reflector antenna. For example, significant simplifications are possible in the design of high power, water-cooled transmitters and low noise cryogenic amplifiers. These systems do not have to rotate as in a normally fed dual reflector. Also, since a BWG system can transmit power over considerable distances at very low losses, BWGs are useful in the design of very high frequency feed systems. Several designs of BWG systems for ground station antennas are shown in Ref. 8; the design of spacecraft antennas using high frequency BWG systems is shown in Ref. 9.

This section addresses three techniques [60] for designing BWG systems: (1) GO analysis, (2) a technique based on a conjugate phase-matching focal plane method, and (3) Gaussian-beam analysis. However, in all cases, the final performance is determined by using physical optics analysis. This section outlines each technique.

5.5.12 Geometric Optics Analysis

For a BWG system where the mirrors are large in terms of wavelengths, GO criteria (introduced by Mizusawa and Kitsuregawa [61, 62]), which guarantee a perfect image from a reflector pair, can be used. Mizusawa's criteria can briefly be stated as follows. For a circularly symmetric input beam, the conditions on a conic reflector pair necessary to produce an identical output beam are: (1) the four loci (two of which may be coincident) associated with the two curved reflectors must be arranged on a straight line; and (2) the eccentricity of the second reflector must be equal to the eccentricity or the reciprocal of the eccentricity of the first reflector.

Figure 5.41 shows some curved reflector pair orientations that satisfy Mizusawa'a criteria.

This technique was used in the design of NASA's first DSN BWG antenna [62]. The design of the center-fed BWG consisted of a beam-magnifier ellipse in a pedestal room located below ground level that would transform a 22-dB gain feedhorn into a high gain 29-dB pattern for input to a Mizusawa four-mirror (two flat and two paraboloid case) BWG system (Figure 5.42). The system was initially designed for operation at 8.45 GHz (X-band) and 32 GHz (Ka-band) and had less than a 0.2-dB loss at X-band (determined by comparing the gain of a 29-dB gain horn feeding the dual-shaped reflector system with that obtained using the BWG system).

Since the upper BWG has a 2.44-m (8-ft) diameter projected aperture, the mirrors are 70 wavelengths at X-band and 260 wavelengths at Ka-band. These mirrors are certainly large enough to use the GO design criteria.

Even though the design was based on GO, all performance analysis was done using PO computer programs.

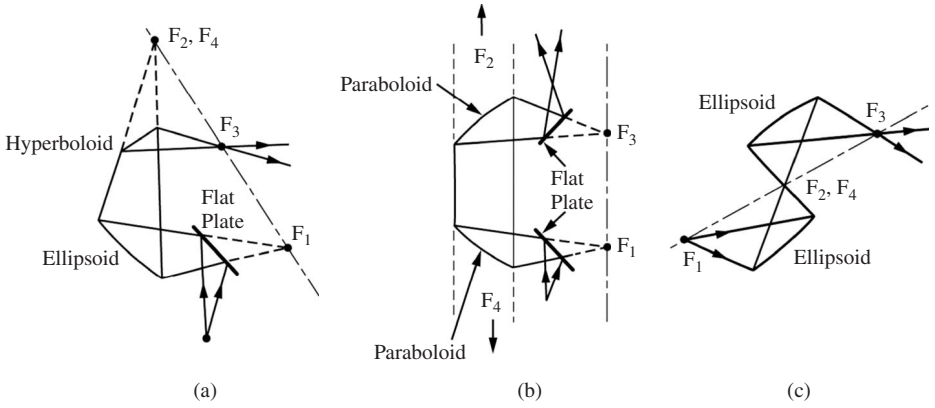


Figure 5.41 Examples of two curved reflector BWG configurations.

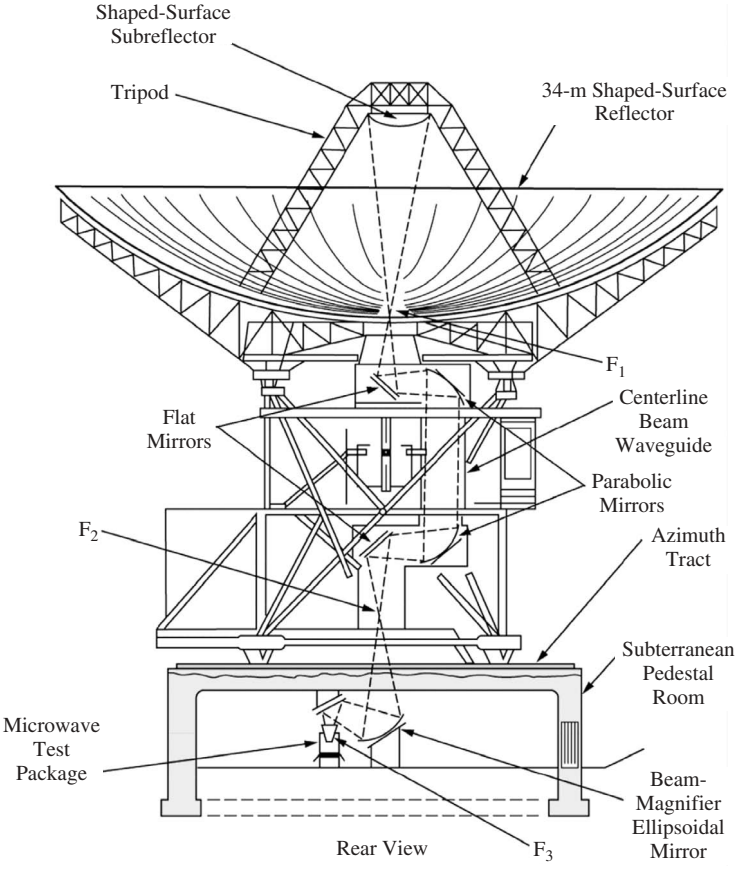


Figure 5.42 DSS-13 highpass BWG design.

5.5.13 Focal-Plane-Matching Technique

The above-mentioned DSN BWG system was initially designed for operation at 8.45 GHz (X-band) and 32 GHz (Ka-band), where the GO design technique was valid. There was also a desire to add a feed system for 2.3 GHz (S-band). However, at S-band, the mirror diameter was only 19 wavelengths, clearly outside the original GO design criteria.

If the same gain horn (22 dB) had been placed at the input focus of the ellipse, the BWG loss would have been greater than 1.5 dB. This is primarily because, for low frequencies, the diffraction phase centers are far from the GO mirror focus, resulting in a substantial spillover and defocusing loss. The defocusing is a problem for the magnifier ellipse especially, where the S-band phase center at the output of the ellipse is 3 m from the GO focus.

A potential solution is to redesign the horn to provide an optimum solution for S-band. The question is how to determine the appropriate gain and location for this feed.

A straightforward design by analysis would prove cumbersome because of the large number of scattering surfaces required for the computation. Rather, a unique application was made of the conjugate phase-matching techniques to obtain the desired solution [63]. A plane wave was used to illuminate the main reflector and the fields from the currents induced on the subreflector propagated through the BWG to a plane centered on the input focal point. By taking the complex conjugate of the currents induced on the plane and applying the radiation integral, the far-field pattern was obtained for a theoretical horn that maximizes the antenna gain.

There is no a priori guarantee that the patterns produced by this method would be easily realized by a practical feedhorn. However, in this case, the pattern was nearly circularly symmetric and the theoretical horn was matched fairly well by a circular corrugated horn.

The corrugated horn performance was only 0.22 dB lower than the optimum theoretical horn and about 1.4 dB better than utilizing the 22-dB feedhorn. A system employing the corrugated horn was built, tested, and installed in the 34-m BWG antenna as part of a simultaneous S/X-band receiving system [64].

5.5.14 Gaussian Beam

While GO is useful for designing systems with electrically large mirrors (>50 wavelengths in diameter with -20-dB edge taper), some BWGs may be operated at low frequencies where the mirrors may be as small as 20 wavelengths in diameter. Due to diffraction effects, the characteristics of a field propagated between small BWG mirrors (<30 wavelengths in diameter) will be substantially different from the GO solution. For these cases, the Gaussian-beam technique can be utilized.

The Gaussian beam is an approximate solution of a wave equation describing a beam that is unguided but effectively confined near an axis. The zero-order mode is normally used in the design. A major advantage of the Gaussian technique is the simplicity of the Gaussian formula, which is easy to implement and has negligible computation time.

Because of the negligible computation time, a Gaussian solution can be incorporated with an optimization routine to provide a convenient method to search the design parameters for a specified frequency range, mirror sizes and locations, and horn parameters.

Goubau gave the first mathematical expression of Gaussian modes derived from the solution of Maxwell's equations described by a continuous spectrum of cylindrical waves [65]. Chu developed the Fresnel zone imaging principle of the Gaussian beam to design

a pseudo-frequency-independent BWG feed [66]. Betsudan, Katagi, and Urasaki used a similar imaging technique to design large ground-based BWG antennas [67]. McEwan and Goldsmith developed a simple design procedure based on the Gaussian-beam theory for illumination of reflector antennas where the reflector is electrically small or in the near field of a feed [68].

Although Gaussian-beam analysis is fast and simple, it is less accurate than the PPO solution for smaller mirrors (<30 wavelengths in diameter). However, by designing with Gaussian-beam analysis, then checking and adjusting using PO analysis, an accurate and efficient tool can be fashioned. Veruttipong, Chen, and Bathker [69] developed such a tool for designing a second 34-m BWG antenna for the Deep Space Network. The goal was to provide good performance over the range of 2–32 GHz.

The design is similar to the DSS-13 antenna (see Figure 5.42) in that it uses three curved mirrors (one in the basement room and two rotating in azimuth) and a 34-m dual-shaped reflector antenna (Figure 5.43). Multiple-frequency operation is provided by the use of dichroic mirrors. The desire is to have the radius of curvature and −18-dB beam diameter of the Gaussian beam at the subreflector be the same at all frequencies. The size and locations of the mirrors are relatively fixed because of the basic structure geometry,

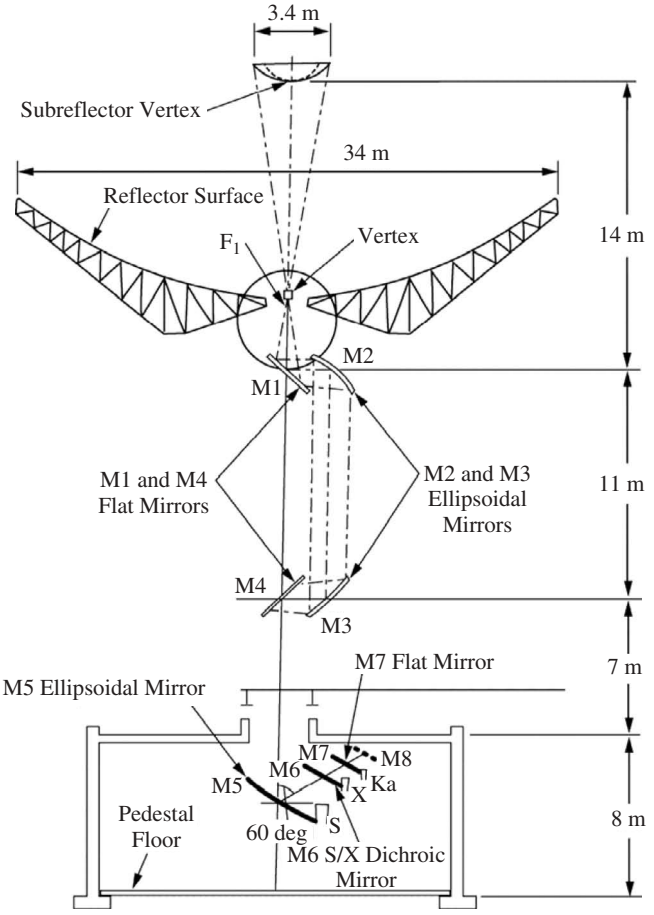


Figure 5.43 Operational BWG antenna.

so the pertinent variables are the horn diameters, horn positions, and mirror curvatures. Approximating the mirrors by a thin lens formula and utilizing Gaussian-mode analysis to iterate the various design parameters, a design is achieved that meets the initial design constraint of identical patterns at the subreflector. The 34-m BWG antenna was built and measured at S-, X-, and Ka-bands (see Ref. 8 for details).

As demonstrated by design and measurements, the use of the GO technique in conjunction with the focal plane matching technique produced an antenna with virtually the same characteristics as the Gaussian-mode technique.

5.6 MULTIPLE-BEAM ANTENNAS WITH EXAMPLES

There is a need for reflector systems with multiple-beam and wide-scan capabilities for both satellite and ground system applications. For satellites, there is the increasing demand for multiple- and contour-beam applications for both commercial and military applications. For ground applications, there is a considerable savings if one antenna can service multiple satellites. Applicable reflector concepts can best be summarized by the diagram shown in Figure 5.44, which demonstrates the interrelationship among concepts with different branches in the diagram.

Important parameters for the antenna systems are the need for high gain (40–55 dB), low sidelobes (below 28–40 dB), low cross polarization (below 30–35 dB), and adequate beam isolation (24–33 dB). A suitable range of component beamwidths for satellite antennas, θ_c, is $0.2^\circ < \theta < 1^\circ$ corresponding to antenna diameters of $65 < D/\lambda < 325$ (e.g., up to 24 meters at 4 GHz). The field of view (FOV) from synchronous altitude for regional coverage such as CONUS is $\pm 4^\circ$; note, however, that INTELSAT systems might require the full Earth FOV of $\pm 9^\circ$. These values imply beams scanned up to 45 beamwidths.

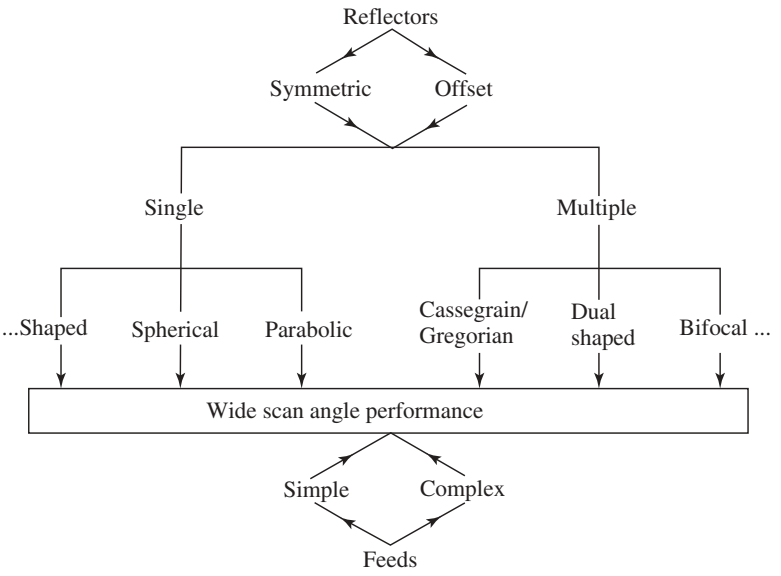


Figure 5.44 Reflector concepts for multiple beams and wide angle scan.

The following section describes the various reflector concepts that have been used to achieve the required characteristics.

5.6.1 Wide Angle Scan

There are many cases where a large number of pencil beams and/or contoured beams are required from a single aperture. For example, anti-ballistic-missile ground stations require either multiple or rapid-scanning pencil beams, while direct broadcast satellites require both contoured and pencil beams. When cost and/or weight requirements prohibit the use of a phased array, it is possible to utilize a reflector aperture instead. Classic reflector geometry, however, is relatively limited with respect to the number of beamwidths of scan off boresight that is possible. This is because the secondary pattern of a reflector antenna deteriorates significantly when the pencil beam is scanned beyond a few beamwidths. Since many applications require large numbers of beamwidth scan, this subject has been investigated intensely.

Several methods have been used to enhance the beamwidth scan capability of reflector antennas.

One method has involved reshaping the reflector surfaces from the classical conic-sections paraboloid or paraboloid and hyperboloid shapes to reflector shapes that specifically enhance the beam scanning capability. Generally, this approach has required two reflector surfaces to obtain the degrees of freedom required in the design synthesis. Two techniques have been explored:

1. The first technique is the bifocal dual-reflector design. This approach—first suggested by Rao [70] for the symmetric design and Galindo-Israel, Mittra, Sheshader, and Cha [71] for the offset design—not only requires two reflectors but is also limited by the difficulty of extending the results to a noncylindrical design. The objectives of this approach are analogous to those of the dual-tuned circuit.
2. A second technique, used by Jorgensen and Balling [72], is to optimize the analytic surface description for wide angle scanning. Although this can be a very time-consuming process, depending on the specifics of the approach and objectives, good results have been obtained.

A second method has been to utilize a more complex multihorn feed system with a simple single reflector (usually a paraboloid). Very often, an offset geometry is required in this approach because the multihorn feed might become large. Note that both methods—two shaped reflectors and a complex feed—can be combined. The complex-feed approach generally attempts to combine a number of adjacent (perhaps distorted) pencil beams in phase and amplitude coherently into one well-formed beam. One of the early successful applications of this approach was by Amitay and Zucker [73] for the design of a phase-correcting feed for a spherical reflector.

The same idea was later applied to a paraboloid by Galindo-Israel, Lee, and Mittra [74]. They used a triangular lattice of feeds, usually in a seven-element, potentially overlapping, cluster to improve the pattern of a beam that was scanned by several beamwidths. The principal object in this type of approach is to choose the excitation coefficients of each radiating element in the cluster so that the coherently summed resultant pencil beam has an "optimum" shape. The definition of this optimum is an important parameter. Galindo-Israel, Lee, and Mittra defined "optimum" as the minimum mean-square

difference between the actual power pattern and the ideal unscanned power pattern. The excitation coefficients were chosen so as to minimize the above difference.

Determination of the optimum set of complex radiating element coefficients can be a very time-consuming task, even on today's largest and fastest computers. Four significant elements enter into the time required for the search:

1. The search (or optimization) algorithm.
2. The initial point in space.
3. The object function: in general, a function of $2N$ real variables when there are N feeds. The object function is usually set up to be minimized (or maximized) in the 2N (or larger) dimensional space.
4. The integration time for the radiated fields.

An example illustrates how all four of the above factors can be dealt with efficiently so that clusters of feeds can be designed for as many as 50 beamwidths of scan in a reflector with nominally an F/D of 0.4 or 0.5.

The definition of the object function to be minimized is critical. Several object functions are considered that make use of a mini–max algorithm. Consider the depiction in Figure 5.45, wherein there is an area where maximum gain is required and/or an area where sidelobe suppression is required. Possible object functions are: (1) maximize the minimum gain in area A1, (2) maximize the difference between the minimum gain in area A1 and the maximum gain in area A2, or (3) maximize the difference between the minimum gain in area A1 and the RMS of the sidelobes in area A2. The set of excitation coefficients that accomplish this is our optimum. In addition, a variety of constraints on the excitation coefficients could be employed. For example, we may allow (1) only a quantized value of amplitude and/or phase for each coefficient or (2) only the phase term of the excitation to vary.

Very significant is the initial set of coefficients used. Since the minimum of our object function in $2N$-dimensional space is likely to be multiple, it is extremely difficult to find, or to even know if it has found the global solution. For example, suppose the far field

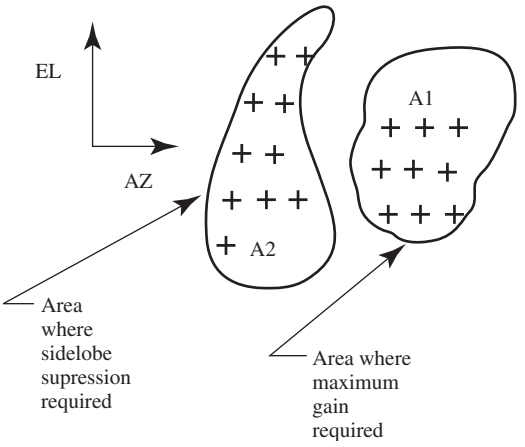


Figure 5.45 Points where radiation field is evaluated.

of N feeds exciting a paraboloid is given by

$$\mathbf{E}_T(\theta, \phi) = \sum_{n=1}^{N} c_n \mathbf{f}_n(\theta, \phi) \tag{5.48}$$

where $\mathbf{f}_n$ (θ, ϕ) are the radiation patterns of each feed excited independently. The gain function at a set of (θ, ϕ) or $(u = \sin\theta \sin\phi, v = \sin\theta \sin\phi)$ is then

$$G(\theta, \phi) = \frac{4\pi |\mathbf{E}_T|^2}{\eta P_{\text{IN}}} \tag{5.49}$$

where P_{IN} is the total power radiated by the composite feeds (all feeds radiating). Now the set of complex C_n that gives the maximum value of $G(\theta_0, \phi_0)$ at one point under condition of $P_{\text{IN}} = \text{constant}$ is readily shown to be

$$C_n = f_n^*(\theta_0, \phi_0) \tag{5.50}$$

This is known as the conjugate-matched field. That is, it is the relative field strength at the feed when the antenna has a plane wave incident from the (θ_0, ϕ_0) direction (or it is the receiving excitation of the dominant mode of the radiating element waveguide under the same conditions). Many applications [75, 76] require only this set of values for the coefficients C_n. We use this set of C_n as our initial condition for more complicated object functions wherein sidelobe level, cross polarization, contour areas, and so on might be involved.

If the entire optimization procedure is directed toward obtaining only the maximum gain at a single point $(\theta_0, \phi_0) = (u_0, v_0)$ in space, then we have seen (Eq. (5.50)) that the radiated field of the nth feed element and reflector (in the presence of mutual coupling, in general) need to be computed at only that one observation point. However, it is often the case that the field needs to be known at many points of observation for use in the object function to be optimized.

We select a number of points in space (AZ, EL) relative to the antenna axis as the working points for imposing desired constraints. The number (field) of points is not necessarily related to the number of excitation coefficients available. The cost functions described above are to maximize either

$$\text{MINI–MAX:} \quad \text{MIN } G_{\text{A1}} - \text{MAX } G_{\text{A2}}$$

$$\text{MINI–RMS:} \quad \text{MIN } G_{\text{A1}} - \sqrt{\frac{\sum}{\text{A2}} |E_{T_i}|^2}$$

where in the above expressions Al and A2 are the regions where maximum gain and suppressed sidelobes, respectively, are required. E_T is the total field at the ith location in area A2. There are a variety of optimization algorithms that have been used. The following example uses the technique described in Ref. 77.

A simple example illustrates the improvements of this mini–max optimization technique over the plane-wave conjugate-matching technique. An offset reflector of 166.75 wavelengths in diameter scanned 6° from boresite is used. The patterns and scanned beam

gain were determined for feed clusters consisting of 1, 7, 19, and 37 feed horns. Comparisons are made between the conjugate-match solution, the mini–max optimization, and the mini–RMS optimization.

An important aspect of the optimization problem is the secondary grid points over which the optimization is to be performed. As indicated in Figure 5.45, A1 is the region where the maximum gain is required and A2 is the region where suppressed sidelobes are required. For the particular examples considered, area A1 consists of only one point; the direction in which the gain is to be maximized that is 6° from boresite. Area A2 consists of equispaced points over the region from $\pm 3^\circ$ in elevation and 3° to 9° in azimuth, with the desired main beam area excluded.

The three cost functions examined are the following:

1. The conjugate-match function, which only considers the direction of the main beam ($AZ = 6^\circ$, $EL = 0^\circ$) and produces the maximum gain solution.
2. Mini–max, which maximizes the difference between the gain and maximum sidelobe level and at the respective grid points.
3. Mini–RMS, which maximizes the gain between the gain and the RMS value of all the sidelobe grid-level points.

The geometry is an offset reflector with a diameter of 166.75λ, an offset height of 35λ, a focal length of 189λ, and an offset angle of 35°. The feed spacing is 1.6λ. Table 5.1 shows the comparison between the various optimization techniques for 19 feeds. Note the spectacular improvement in sidelobe level given by the mini–RMS technique. Table 5.2 shows the gain and sidelobe level versus the number of feeds for the mini–max design. Note that the mini–RMS design at 19 feeds has lower sidelobes than the mini–max design with 37 feeds although the gain is lower by approximately 2 dB.

The basic conclusions of the study are that the optimization techniques can be successfully used to control the sidelobe levels for large angles of scan in a parabolic reflector system. The conjugate match gives the highest gain and the poorest sidelobe performance; the mini–RMS technique produces the best sidelobe performance with the lowest gain.

5.6.2 Contour Beam Antennas

A multibeam antenna can also be used to provide a contour beam over a sector of the Earth as seen from a satellite antenna. To design this type of system it is necessary to

TABLE 5.1 Gain and Sidelobe Level Versus Optimization Criteria for 19 Feeds

Technique	Conjugate Match	Mini–Max	Min–RMS
Gain	51.05	50.62	49.08
Sidelobe level	−22.82	−25.96	−34.16
	Focused on-axis gain 52.01 dB		

TABLE 5.2 Gain and Sidelobe Level Versus Number of Feeds for Mini–Max Optimization

Technique	1	7	19	37
Gain	45.18	49.32	50.62	51.22
Sidelobe level	−8.64	−22.33	−25.96	−28.45
	Focused on-axis gain 52.01 dB			

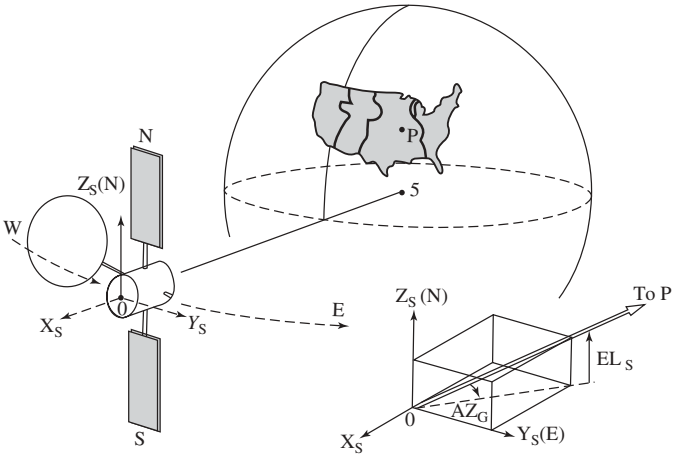


Figure 5.46 The Earth as seen from a satellite.

relate positions on the Earth to directions in reflector coordinates. Consider the Earth as seen in Figure 5.46. The satellite coordinates are ($\mathbf{X}_S$, $\mathbf{Y}_S$, $\mathbf{Z}_S$), where $\mathbf{X}_S$ is a unit vector from Earth center, $\mathbf{Y}_S$ is a unit vector parallel to the East direction, and $\mathbf{Z}_S$ is a unit vector parallel to the North direction. The antenna boresight is pointed at P located at $AZ_S = \alpha$ and $EL_S = \beta$ in satellite coordinates. Antenna coordinates (X, Y, Z) are related to satellite coordinates by the following (see Figure 5.47)

$$\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \\ \mathbf{Z} \end{bmatrix} = \begin{bmatrix} & -\sin\alpha & -\cos\alpha & 0 \\ \sin\beta & \cos\alpha & -\sin\beta\sin\alpha & \cos\beta \\ -\cos\beta & \cos\alpha & \cos\beta\sin\alpha & \sin\beta \end{bmatrix} \begin{bmatrix} \mathbf{X}_S \\ \mathbf{Y}_S \\ \mathbf{Z}_S \end{bmatrix}$$

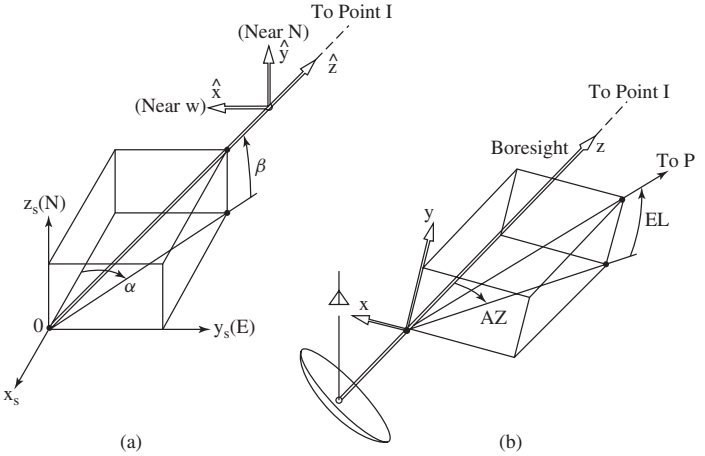


Figure 5.47 Antenna coordinates related to satellite coordinates. Antenna boresight **Z** is in the direction ($AZ_s = \alpha$, $EL_s = \beta$) in the satellite coordinates shown in (a). An observation point P on Earth is described by (AZ, EL) in the antenna coordinates in (b).

An observation point P on Earth is described by (AZ, EL) in antenna coordinates. For the small angular region about the boresight it can be shown that

$$\mathrm{AZ} \approx \mathrm{AZ}_S - \alpha$$

$$\mathrm{EL} \approx \mathrm{EL}_S - \beta$$

Maps are generated by converting latitude and longitude of points on Earth to (AZ, EL) in antenna coordinates. It is a function of the longitude of the satellite as well as the antenna boresight.

The contoured beam antenna design steps are as follows:

1. Plot a coverage map as viewed from a synchronous orbit spacecraft, or a composite map including the pointing error and views from different orbit positions.
2. Select a set of contiguous squares or circles that represent a −3-dB or −4-dB beamwidth for each feed in the offset reflector system (this in part determines or is determined by the reflector diameter).
3. Arrange these squares or circles in a rectangular, triangular, or triangular lattice structure with appropriate scale that best fits the desired coverage area on the map.
4. If sidelobe suppression is desired, add one extra square or circle along the exterior of the desired coverage area.
5. Determine the optimum reflector parameters: the projected reflector diameter D, focal length F, spacing between feeds or feed diameter d, and offset angle Ω. For the sidelobe suppression case, these are related by the appropriate formula:

$$d = \frac{1.06}{D}\left(\frac{2F}{1+\cos\Omega}\right)$$

6. Determine the optimum amplitude and phase distribution for each individual horn by numerical optimization techniques (see previous section).
7. Design and implement the beam-forming network. The procedure is illustrated in the following example from Ref. 78. The Eastern and Central Time Zones with 4-dB coverage circles as seen from longitude 101° west is shown in Figure 5.48. The antenna uses the 2.238-m reflector geometry from INTELSAT V, which is approximately 101λ at the center frequency of 10.95 GHz. Using the 17 feeds that cover the Eastern Time Zone and optimizing the gains over a set of 24 Eastern cities results in the coverage shown in Figure 5.49.

It is also possible to provide a contour beam by using a single feed and shaping the reflector surface. This offers a number of advantages compared to a multibeam antenna regarding feed network losses and reduced spillover.

The optimum shape of the surface is found by optimizing the phase of the aperture field. The phase is given as an expansion in orthogonal polynomials in which the coefficients of the terms are the variables to be optimized. The far field associated to the aperture distribution is calculated at points that represent the desired contour. Thus we have a nonlinear set of equations in which the gain levels at the stations are functions of the coefficients in the phase expansion. By means of an optimization procedure, the coefficients are determined to maximize the minimum gain. The reflector surface corresponding to the selected phase distribution is finally determined by an optical ray-tracing procedure.

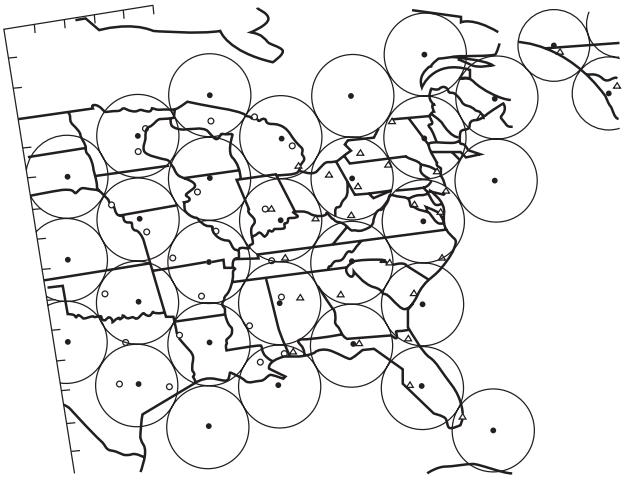


Figure 5.48 Eastern and Central Time Zones from longitude 101° west.

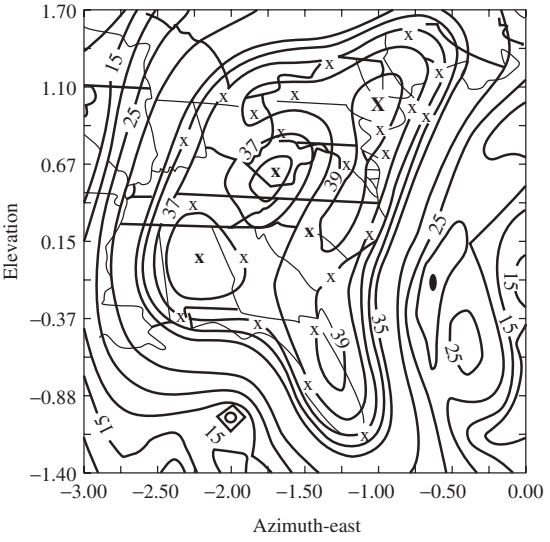


Figure 5.49 Eastern coverage with 17 feeds.

Investigations with the method have been carried out and compared to similar multibeam designs; the concept has been found competitive regarding minimum gain and sidelobe level.

5.6.3 Low Cross-Polarization Systems

Use of orthogonal polarization is often required in communication systems to double transmission capacity. In addition, other antenna applications (e.g., remote sensing, radar) also require antennas capable of providing good polarization discrimination along

with the requirement to provide high gain over a wide field of view. Although good performance over a moderate field of view can be achieved with conventional classical offset dual-reflector geometries, some applications require performances unattainable with these antennas. For such cases, the classical offset Dragonian antenna might provide the required improved performance [79]. The classical Dragonian antenna employs a paraboloidal main reflector illuminated by a concave hyperboloidal subreflector. The relatively large offset distance and focal length of the main reflector and the avoidance of caustics between the two reflector surfaces (a consequence of the concave subreflector) yield relatively flat reflectors with excellent polarization properties and a wide field of view capability. The basic geometry, which is shown in Figure 5.50, can be used in a front-fed or side-fed configuration. Although a higher gain feed is required for the side-fed configuration, when combined with a very good corrugated horn design this geometry can provide better performance than the front-fed configuration. Reference 80 provides a very good design procedure for such systems. A very interesting application is shown in Ref. 81, wherein a 1.4-meter telescope with 91 feeds in the focal plane is used to study the polarization properties of the cosmic microwave background (CMB) radiation. The 1.4-meter configuration at 90 GHz has a 12.6-arc-min beamwidth and can scan out to 4° with only a 0.3-dB loss and a cross-polarization level of -40 dB.

5.6.4 Spherical Reflectors

The properties, practical applications, and aberrations of the spherical reflector are well known [82]. The popularity of this reflector is primarily due to the large angle through which the radiated beam can be scanned by translation and orientation of the primary feed. This wide angle property results from the symmetry of the surface. Multiple-beam operation is realized by placing multiple feeds along the focal surface. In the conventional use of the reflector surface, the minimum angular separation between adjacent beams is determined by the feed-aperture size. The maximum number of beams is determined by the percentage of the total sphere covered by the reflector.

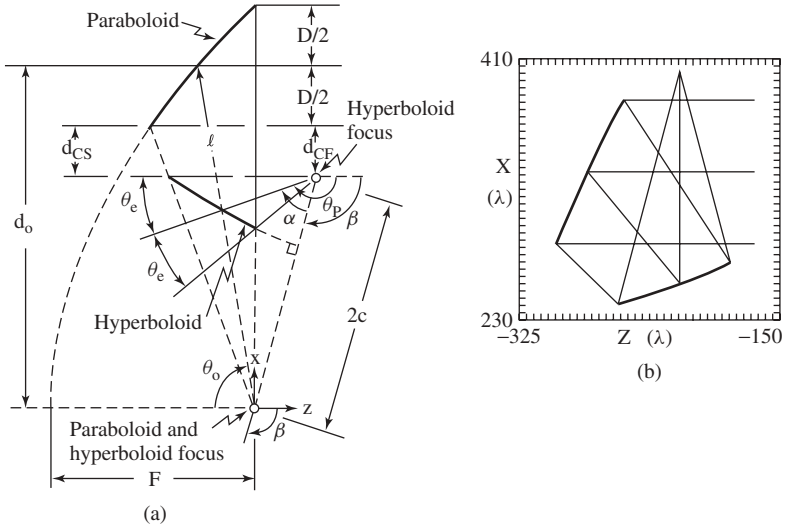


Figure 5.50 Classical offset Dragonian antenna: (a) front-fed geometry and (b) side-fed geometry.

Probably the best known spherical reflector is the Arecibo Observatory (Puerto Rico), which is one of the world's largest radio telescopes. Its main reflector is a spherical surface 305 meters in diameter and nearly 51 meters deep. The Arecibo feed system consists of both narrowband line feeds and Gregorian optics [83, 84] to correct for spherical aberrations, as shown in Figure 5.51. Either the Gregorian optics or line feeds can be positioned up to 20° from the vertical.

Spherical antennas have also been used for beam-steerable multibeam Earth station antennas [85]. The design also uses trireflector geometry similar to the Arecibo Gregorian optics, although in this case, the spherical reflector is offset to provide a blockage-free aperture. Referring to Figure 5.52, the auxiliary reflectors and subreflectors are shaped so that combined with the main reflector they provide a uniform phase distribution and a specified amplitude distribution in the aperture of the main reflector. The details of the optimization procedure to design the reflectors are given in Ref. 85. Because of the symmetry of the main reflector, the same dual-shaped reflector optics are used for all the beams. Since any straight line that passes through the spherical center can be used as an axis of rotation, if the spherical reflector is offset fed with oblique illumination angle β, conical scan without excessive area is possible when the axis of conical rotation coincides with the axis of the reflector. For the design in Ref. 85, which is used at the Osaka teleport, the conical rotation angle $\Delta\gamma$ of 13° is selected with the oblique illumination angle β of 16°. For the parameters chosen, the four beams use an aperture that is only slightly larger than required for a single beam. The calculated directive gains of the four beams cover satellite longitude from 150° to 162° E with only a 0.3-dB variation in performance.

5.6.5 Torus Antenna

The torus antenna is a dual-curvature reflector also capable of multibeam operation when it is fed with multiple feeds. The feed-scan plane can be inclined to be in the orbital-arc plane, allowing the use of a fixed reflector to view geosynchronous satellites. The geometry is shown in Figure 5.53 from Ref. 86. In the figure, the section M is a parabola, V

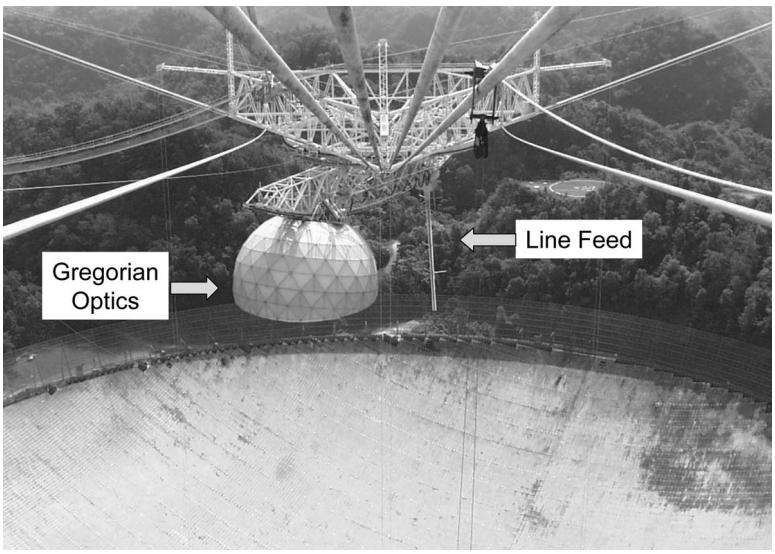


Figure 5.51 Arecibo feed system.

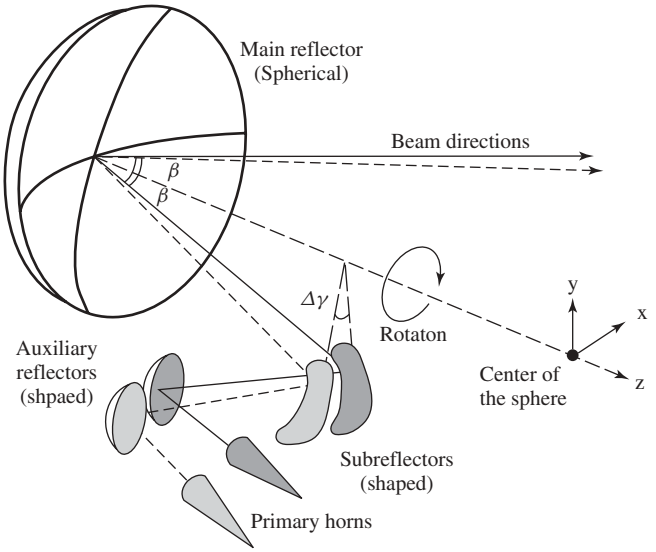


Figure 5.52 Offset-fed spherical antenna geometry.

is its vertex, F is its focus, z is the direction of its axis, R is the radius of rotation at the vertex V, d is the offset of the reflector, and the x-y plane is a plane perpendicular to its axis. The parabolic section produces a beam parallel to the z-axis. The angle between the axis of rotation (the z-axis) and the beam direction is α. The parabola M is rotated about the z-axis to form a torus. COMSAT has shown that for $\alpha = 95.5^\circ$ the reflector is a parabolic torus whose beam directions closely approximate the direction of the geostationary arc. COMSAT has successfully demonstrated its use as an Earth station antenna

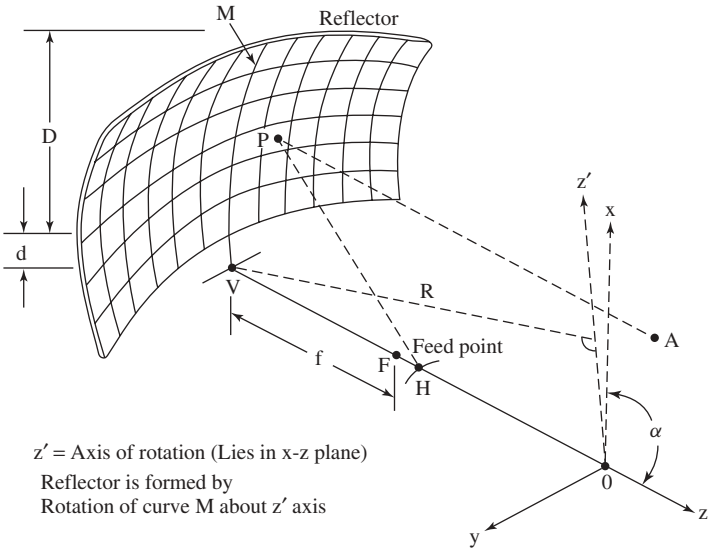


Figure 5.53 Torus antenna geometry.

whose radiation patterns meet a (29−25 log θ)-dBi envelope. The offset-fed geometry gives rise to low wide angle sidelobes as well as eliminating the blockage of the multiple feeds.

The torus antenna has less phase aberration than the spherical antenna because of the focusing in the parabolic plane. Because of the circular symmetry, feeds placed anywhere on the feed arc form identical beams. Therefore no performance degradation is incurred when multiple beams are placed on the focal arc. Point-focus feeds may be used to feed the torus up to aperture diameters of approximately 200 wavelengths. For larger apertures, it is recommended that aberration-correcting feeds be used.

The scanning or multibeam operation of a torus requires an oversized aperture to accommodate the scanning. For example, a reflector surface area of approximately 214 m^2 will allow a field of view (i.e., orbital arc) of 30° with a gain of approximately 50.5 dB at 4 GHz (equivalent to the gain of a 9.65-m reflector antenna). This surface area is equivalent to approximately three 9.65-m antennas.

REFERENCES

1. S. Silver (Ed.), *Microwave Antenna Theory and Design*, Radiation Lab Series, Vol. 12, McGraw-Hill, New York, 1949, pp. 336–338.
2. H. Jasik (Ed.), *Antenna Engineering Handbook*, McGraw-Hill, New York, 1961.
3. R. C. Johnson (Ed.), *Antenna Engineering Handbook*, 3rd ed. McGraw-Hill, New York, 1992.
4. W. V. T. Rusch and P. Potter, *Analysis of Reflector Antennas*, Academic Press, New York, 1970.
5. A. W. Rudge, K. Milne, A. D. Olver, and P. Knight, *The Handbook of Antenna Design*, Vol. I, Peter Peregrinus Ltd., London, 1982.
6. Y. T. Lo and S. W. Lee, *Antenna Handbook*, Springer, New York, 1993.
7. Scott, *Modern Methods of Reflector Antenna Analysis and Design*, Artech House, Norwood, MA, 1993.
8. W. A. Imbriale, *Large Antennas of the Deep Space Network*, John Wiley & Sons, Hoboken, NJ, 2003.
9. W. A. Imbriale, *Spaceborne Antennas for Planetary Exploration*, John Wiley & Sons, Hoboken, NJ, 2006.
10. J. Ruze, Lateral-feed displacement in a paraboloid, *IEEE Trans. Antennas Propag*., Vol. 13, pp. 660–665, September 1965.
11. W. A. Imbriale, P. G. Ingerson, and W. C. Wong, Large lateral feed displacements in a parabolic reflector, *IEEE Trans. Antennas Propag*., Vol. 22, pp. 742–745, November 1974.
12. V. Galindo and R. Mittra, A new series representation for the radiation integral with application to reflector antennas, *IEEE Trans. Antennas Propag*., Vol. 25, pp. 631–641, September 1977.
13. W. A. Imbriale and R. E. Hodges, Linear-phase approximation in the triangular facet near-field physical optics computer program, TDA PR 42-102, April–June 1990, pp. 47–56, August 15, 1990.
14. W. A. Imbriale and R. E. Hodges, The linear phase triangular facet approximation in physical optics analysis of reflector antennas, *Appl. Computational Electromagn. Soc.*, Vol. 6, No. 2, pp. 74–85, Winter 1991.
15. S. W. Lee and R. Mittra, Fourier transform of a polygonal shape function and its application in electromagnetics, *IEEE Trans. Antennas Propag.*, Vol. 31, No. 1, pp. 99–103, January 1983.
16. Y. Rahmat-Samii, Useful coordinate transformations for antenna applications, *IEEE Trans. Antennas Propag.*, Vol. 27, pp. 571–574, July 1979.

17. P. G. Ingerson and W. V. T. Rusch, Studies of an axially defocused paraboloid, *IEEE Int., Antennas Propag. Symp. Dig.*, Vol. 7, pp. 62–68, December 1969.
18. J. Ruze, Antenna tolerance theory—a review, *Proc. IEEE*, Vol. 54, pp. 633–640, 1966.
19. R. Levy, *Structural Engineering of Microwave Antennas*, IEEE Press, Piscataway, NJ, 1996, pp. 289–296.
20. S. von Horner, Homologous deformations of tiltable telescopes, *Proc. ASCE J. Struct. Div*., Vol. 93, ST5, October 1967.
21. H. E. Schrank, Low sidelobe reflector antennas, *IEEE Antennas Propag. Soc. Newslett.*, April 1985.
22. C. C. Cutler, A. P. King, and W. E. Kock, Microwave antenna measurements, *Proc. IRE*, Vol. 35, pp. 1462–1471, December 1947.
23. Wilkinson Microwave Anisotropy Probe, http://map.gsfc.nasa.gov/. Accessed January 2007.
24. J. D. Kraus, *Electromagnetics*, McGraw-Hill, New York, 1984.
25. A. C. Ludwig, The definition of cross polarization, *IEEE Trans. Antennas Propag.*, Vol. 21, pp. 116–119, January 1973.
26. P. D. Potter, A new horn antenna with suppressed sidelobes and equal beamwidths, *Microwave J.*, pp. 71–78, June 1963.
27. S. A. Brunstein, A new wideband feed horn with equal E- and H-plane beamwidths and suppressed sidelobes, in the Deep Space Network, *Space Programs Summary 37-58*, Vol. II, pp. 61–64, Pasadena, California, Jet Propulsion Laboratory, July 1969.
28. D. Hoppe, Scattering matrix program for circular waveguide junctions, in *Cosmic Software Catalog*, 1987 edition, NASA-CR-179669, NTO-17245, NASA's Computer Software Management and Information Center, Georgia, 1987.
29. D. Hoppe, Modal analysis applied to circular, rectangular and coaxial waveguides, *TDA Progress Report 42-95*, Jet Propulsion Laboratory, Pasadena, California, July–September 1988, pp. 89–96.
30. D. Hoppe, W. Imbriale, and A. Bhanji, The effects of mode impurity on Ka-band system performance, *TDA Progress Report 42-80*, Jet Propulsion Laboratory, Pasadena, CA, October–December 1984, pp. 12–23.
31. G. L. James, Analysis and design of TE_{11} and HE_{11}, corrugated cylindrical waveguide mode converters, *IEEE Trans. Microwave Theory Tech*., Vol. 29, pp. 1059–1066, October 1981.
32. Y. Rahmat-Samii and W. A. Imbriale, Anomalous results from PO applied to reflector antennas: the importance of near field computations, *IEEE Int. Antennas Propag. Symp. Dig.*, Vol. 36, pp. 816–819, June 1998.
33. A. C. Ludwig, Spherical wave theory, in *Handbook of Antenna Design*, A. W. Rudge et al (Eds.), Peter Pereginus Ltd., London, 1982, Sec. 2.3.
34. A. C. Ludwig, Calculation of scattered patterns from asymmetrical reflectors, Ph.D. dissertation, University of Southern California, Los Angeles, 1969. Also Jet Propulsion Laboratory, California Institute of Technology, Pasadena, *Tech. Rep. 32-1430*, February 1970.
35. W. A. Imbriale and D. J. Hoppe, Computational techniques for beam waveguide systems, in *2000 IEEE Antennas and Propagation International Symposium*, Salt Lake City, Utah, 16–21 July 2000, pp. 1894–1897.
36. W. A. Imbriale and D. J. Hoppe, Recent trends in the analysis of quasioptical system, in *AP2000 Millennium Conference on Antennas and Propagation*, Davos, Switzerland, 9–14 April 2000.
37. P. F. Goldsmith, *Quasioptical Systems*, IEEE Press, Piscataway, NJ, 1998, pp. 26–29.
38. V. D. Agrawal and W. A. Imbriale, Design of a dichroic Cassegrain subreflector, *IEEE Trans. Antennas Propag.*, Vol. 27, No. 4, pp. 466–473, July 1979.

39. W. A. Imbriale, Dichroic surfaces for spacecraft and ground systems antennas, in *Periodic Structure 2006s*, M. Bossi and L. Perregrini, (Eds.), Research Signpost, Trivandrum, India, 2006, Chap. 4.
40. The Allen Telescope Array Website, http://www.seti.org/ata. Accessed January 2007.
41. The Square Kilometre Array Website, http://www.skatelescope.org/. Accessed January 2007.
42. W. A. Imbriale, S. Weinreb, and M. Handi, Design of a wideband radio telescope, IEEE Aerospace Conference, Big Sky, Montana, March 3–10, 2007.
43. W. A. Imbriale, Applications of the method of moments to thin-wire elements and arrays, in *Topics in Applied Physics, Vol. 3, Numerical and Asymptotic Techniques in Electromagnetics*, R. Mittra (Ed.), Springer-Verlag, Berlin, 1975, Chap. 2.
44. T. Chu and R. H. Turrin, Depolarization properties of offset reflector antennas, *IEEE Trans. Antennas Propag.*, Vol. 21, pp. 339–345, May 1973.
45. N. A. Adatia and A. W. Rudge, Beam-squint in circularly-polarized offset reflector antennas, *Electron. Lett.*, pp. 513–515, October 1975.
46. P. W. Hannan, Microwave antennas derived from the Cassegrain telescope, *IRE Trans. Antennas Propag.*, Vol. 9, pp. 140–153, March 1961.
47. P. A. Jensen, Designing Cassegrain antennas, *Microwave J.*, pp. 10–16, December 1962.
48. C. Granet, Designing axially symmetric Cassegrain or Gregorian dual-reflector antennas from combinations of prescribed geometric parameters, *IEEE Antennas Propag.*, Vol. 40, pp. 76–82, April 1998.
49. V. Galindo, Design of dual reflector antenna with arbitrary phase and amplitude distributions, in *PTGAP International Symposium*, Boulder, July 1963.
50. V. Galindo, Synthesis of dual reflector antennas, *Elec. Res. Lab. Rep. 64-22*, University of California, Berkeley, July 30, 1964.
51. W. F. Williams, High efficiency antenna reflector, *Microwave J.*, pp. 79–82, July 1965.
52. W. C. Wong, On the equivalent parabola technique to predict the performance characteristics of a Cassegrainian system with an offset feed, *IEEE Trans. Antennas Propag.*, Vol. 21, pp. 335–339, May 1973.
53. H. Tanaka and M. Mizusawa, Elimination of crosspolarization in offset dual-reflector antennas, *Elec. Commun. Japan*, Vol. 58, pp. 71–78, 1975.
54. Y. Mizuguchi, M. Akagawa, and H. Yokoi, Offset dual reflector antenna, *Dig. 1976 AP-S Int. Symp. Antennas Propag.*, October 1976.
55. C. Dragone, Offset multireflector antennas with perfect pattern symmetry and polarization discrimination, *Bell Syst. Tech. J.*, Vol. 57, pp. 2663–2684, September 1978.
56. R. A. Shore, A simple derivation of the basic design equation for offset dual reflector antennas with rotational symmetry and zero cross polarization, *IEEE Trans. Antennas Propag.*, Vol. 33, No. 1, January 1985.
57. V. Galindo-Israel, W. A. Imbriale, and R. Mittra, On the theory and synthesis of single and dual offset shaped reflector antennas, *IEEE Trans. Antennas Propag.*, Vol. 35, No. 8, pp. 887–896, August 1987.
58. V. Galindo-Israel, W. A. Imbriale, R. Mittra, and K. Shogen, On the theory of the synthesis of offset dual-shaped reflectors—case examples, *IEEE Trans. Antennas Propag.*, Vol. 39, No. 5, pp. 620–626, May 1991.
59. A. G. Cha, The JPL 1.5-meter clear aperture antenna with 84.5 percent efficiency, *TDA Progress Report 42-73*, Jet Propulsion Laboratory, Pasadena, CA, 15 May 1983, pp. 1–14.
60. W. A. Imbriale, Design and applications of beam waveguide systems, *1997 IEEE Aerospace Conference*, Snowmass, Colorado, Vol. 3, pp. 121–134, 1997.
61. M. Mizusawa and T. Kitsuregawa, A beam-waveguide feed having a symmetric beam for Cassegrain antennas, *IEEE Trans. Antennas Propag.*, Vol. 21, pp. 844–846, November 1973.

62. T. Veruttipong, J. R. Withington, V. Galindo-Israel, W. A. Imbriale, and D. Bathker, Design considerations for beam-waveguide in the NASA Deep Space Network, *IEEE Trans. Antennas Propag.*, Vol. 36, pp. 1779–1787, December 1988.
63. W. A. Imbriale and J. S. Esquival, A novel design technique for beam-waveguide antennas, in *1996 IEEE Aerospace Applications Conference*, Aspen, Colorado, February 1996.
64. W. A. Imbrale, M. S. Esquivel, and F. Manshadi, Novel solutions to low-frequency probelms with geometrically designed beam-waveguide systems, *IEEE Trans. Antennas Propag.*, Vol. 46, pp. 1790–1796, December 1998.
65. G. Goubau and F. Schwering, On the guide propagation of electromagnetic wave beams, *IRE Trans. Antennas Propag.*, Vol. 9, No. 5, pp. 248–256, May 1961.
66. T. S. Chu, An imaging beam waveguide feed, *IEEE Trans. Antennas Propag.*, Vol. 31, No. 4, pp. 614–619, July 1983.
67. S. Betsudan, T. Katagi, and S. Urasaki, Design method of four reflector-type beam waveguide feeds, *Japanese Electron. Commun. Soc. J.*, Vol. J67-B, No. 6, pp. 622–629, June 1984.
68. N. J. McEwan and P. F. Goldsmith, Gaussian beam techniques for illuminating reflector antenna, *IEEE Trans. Antennas Propag.*, Vol. 37, No. 3, pp. 297–304, March 1989.
69. W. Veruttiong, J. C. Chen, and D. A. Bathker, Gaussian beam and physical optics iteration technique for wideband beam waveguide feed design, *Telecommunications and Data Acquisition Progress Report 42-105*, Jet Propulsion Laboratory, California Institute of Technology, pp. 128–135, 15 May 1991.
70. B. L. J. Rao, Bifocal dual reflector antenna, *IEEE Trans. Antennas Propag.*, Vol. 22, No. 5, September 1974.
71. V. Galindo-Israel, R. Mittra, M. Sheshadri, and A. Cha, Offset bifocal dual reflector antennas with profile optimization, in *International Union of Radio Science, USIIC/URSI Meeting*, Boulder, Colorado, November 1979.
72. R. Jorgensen and P. Balling, Contoured-beam synthesis by adjusting the aperture phase in a reflector antenna, *Int. Symp. Dig. URSI*, 1983.
73. N. Amitay and H. Zucker, Compensation of spherical reflector aberrations by planar array feeds, *IEEE Trans. Antennas Propag.*, Vol. 20, pp. 40–56, January 1972.
74. V. S. Galindo-Israel, W. Lee, and R. Mittra, Synthesis of a laterally displaced cluster feed for a reflector antenna with application to multiple beams and contoured patterns, *IEEE Trans. Antennas Propag.*, Vol. 26, No. 2, March 1976.
75. O. M. Bucci, G. Franceschetti, and R. Pierri, Reflector antenna fields—an exact aperture-like approach, *IEEE Trans. Antennas Propag.*, Vol. 29, No. 4, July 1981.
76. C. C. Hung and R. Mittra, Wide angle scanning for reflector antennas, URSI, Houston, 1983.
77. H. H. Rosenbrock, An automatic method for finding the greatest or least value of a Function, *Comput. J.*, October 1960.
78. V. Galindo-Israel, Y. Rahmat-Samii, W. Imbriale, H. Cohen, and R. Cagnon, A bonus time zone DBS antenna system with single polarization, *IEEE Int. Antennas Propag. Symp. Dig.*, Vol. 22, pp. 143–146, June 1984.
79. C. Dragone, Offset multireflector antennas with perfect pattern symmetry and polarization discrimination, *AT&T Tech. J.*, Vol. 57, pp. 2663–2684, 1978.
80. S. Chang and A. Prata, Jr., A design procedure for classical offset Dragonian antennas with circular apertures, in *IEEE International Antennas and Propagation Symposium Digest*, Orlando, Florida, 11–16 July 1999.
81. W. A. Imbriale, The 1.4 meter telescope design for the Q/U imaging experiment, in *URSI National Radio Science Meeting*, Boulder, Colorado, January 2006.
82. T. Li, A study of spherical reflectors as wide-angle scanning antennas, *IRE Trans. Antennas Propag.*, Vol. 7, pp. 223–226, July 1959.

83. A. W. Love, Spherical reflecting antennas with corrected line sources, *IRE Trans. Antennas Propag.*, Vol. 10, pp. 529–537, September 1962.

84. P. Kildal and P. Skyttemyr, Diffraction analysis of a proposed dual-reflector feed for the spherical reflector antenna of the Arecibo Observatory, *Radio Sci.*, Vol. 24, No. 5, pp. 601–617, September–October 1989.

85. S. Nomoto, F. Watanabe, Y. Mizuguchi, A. Abe, T. Ino, and S. Ohno, A Ku-band 4-beam Earth station antenna with a 6.5-m fixed spherical reflector, in *14th AIAA International Communication Satllite System Conference*, Washington, DC, 22–26 March 1992.

86. G. Hyde, R. W. Kreutel, and L. V. Smith, The unattended Earth terminal multiple-beam torus antenna, *Comsat Tech. Rev.*, Vol. 4, No. 2, Fall 1974.

CHAPTER 6

Frequency-Independent Antennas: Spirals and Log-periodics

HISAMATSU NAKANO

6.1 INTRODUCTION

A resonant antenna, such as a dipole antenna, has an outgoing current (flowing from the feed point toward the antenna ends) and an incoming current (the current that is reflected at the arm ends and proceeds toward the feed point). These currents form a standing-wave current distribution on the antenna. The incoming current that reaches the feed point changes the total current at the input terminals. This, in turn, changes the input impedance.

Practically speaking, the incoming current on the antenna conductor is not only the current reflected from the conductor ends but includes reflections from the conductor edges. Note that the total incoming current that reaches the feed point depends on the operating frequency and the conductor shape. If the total current at the input terminals, which is the combination of all incoming currents and the outgoing current, remains constant irrespective of frequency, the input impedance remains constant. Then the question arises: What structures bring about an invariant input impedance characteristic? And what kind of radiation pattern is obtained in such antennas? This chapter, consisting of six sections, presents answers to these questions, with a number of examples.

Section 6.2 presents a relationship between the structure of an antenna and the input impedance. Using Babinet's principle in electromagnetic fields (Section 6.2.1), the input impedance of an antenna with a self-complementary structure (Section 6.2.2) is proved to be independent of frequency (Mushiake's relation). In addition, on the basis of the principle of similitude for electromagnetic fields (Section 6.2.3), the input impedance characteristic of an antenna with a self-similar structure (Section 6.2.4) is discussed. Furthermore, the transformation of the radial distance to angles (Section 6.2.5) is derived for frequency independent antennas, referring to Rumsey's work. Note that an antenna whose input impedance is constant and whose radiation characteristics are also constant is referred to as a *frequency-independent antenna*.

The theory of the self-complementary and self-similar antennas in Section 6.2 is derived under the condition that the antenna structures are of infinite extent. However,

Modern Antenna Handbook. Edited by Constantine A. Balanis

practical antennas have finite structures. From this point of view, Section 6.3 briefly describes analysis methods for handling antennas of finite size. These analysis methods are the method of moments (Section 6.3.1) and the finite-difference time-domain method (Section 6.3.2). The calculation techniques for the input impedance, radiation pattern, and gain, based on these methods, are briefly summarized.

If an antenna structure is specified only in terms of angles, the impedance and radiation characteristics of such an antenna will be frequency independent, as described in Section 6.2.5. Section 6.4 is developed on the basis of this idea. First, an *equiangular spiral antenna* is investigated (Section 6.4.1). The equiangular spiral radiates a circularly polarized bidirectional beam. Next, a technique for transforming this bidirectional beam into a unidirectional beam is discussed (Section 6.4.2). Note that a unidirectional beam can be realized without a cavity by using a *conical spiral* structure (Section 6.4.3).

It is of interest to investigate a planar spiral antenna whose arms are wound with a constant arm width, not with the expanding width of an equiangular spiral. Such a spiral is realized using the Archimedean function (Section 6.4.4). The spacing between neighboring arms in this Section is chosen to be the same as the arm width, so that the Archimedean spiral antenna (composed of two arms) becomes a self-complementary antenna. The radiation mechanism is explained and an Archimedean spiral antenna backed by a cavity is analyzed. Note that the Archimedean function differs from the function derived by Rumsey for a frequency-independent antenna.

Section 6.5 starts with a discussion of two antennas with log-periodic shape (a *toothed log-periodic antenna* and a *trapezoidal log-periodic antenna*), followed by a discussion of a *log-periodic dipole array antenna*. The arms of the toothed log-periodic antenna (Section 6.5.1) have a finite length in the radial direction. This toothed log-periodic antenna is a truncated version of a self-complementary antenna. The trapezoidal log-periodic antenna (Section 6.5.2) is a modified toothed log-periodic antenna. The log-periodic dipole array (Section 6.5.3) is obtained by folding one arm of the trapezoidal log-periodic antenna onto the other arm and replacing the teeth with dipoles, and hence it is a derivative of the log-periodic self-complementary antenna. The active region for each of these antennas is analyzed and discussed.

Finally, some brief comments are presented in Section 6.6.

So far, most of the findings on finite-sized spirals and log-periodics have been obtained through experimental work. Against this historical background, this chapter tries to obtain theoretical antenna characteristics, taking into account the finite size of the antennas. Throughout this chapter, the results obtained on the basis of either the method of moments or the finite-difference time-domain method are checked using a commercially available solver for electromagnetic wave problems (CST Microwave Studio, Version 5.1.2, finite integration method). The results labeled as EMS in the figures of this chapter are those obtained using this electromagnetic wave solver.

6.2 THEORY

An antenna whose input impedance is independent of frequency is called a *constant impedance antenna*. Some constant impedance antennas have radiation characteristics that are independent of frequency. These antennas, having both constant input impedance and constant radiation characteristics, are called *frequency-independent antennas*.

The input impedance is related to the antenna structure. From this point of view, two antenna structures are investigated in this section: self-complementary structures and self-similar structures [1–4].

We start with a derivation of *Babinet's principle in electromagnetic fields* to gain a better understanding of the self-complementary structure and then discuss *Mushiake's relationship* for the input impedance of a self-complementary antenna. In addition, we focus on the self-similar structure and reveal the input impedance of a self-similar antenna. Finally, an equation derived by Rumsey for a frequency-independent antenna is presented and discussed.

6.2.1 Babinet's Principle in Electromagnetic Fields

The relationship between the input impedances of an *arbitrarily shaped slot antenna* and its complementary plate antenna (see Figure 6.1a, c) was first derived by Mushiake in 1948 [3]:

$$Z_{\text{slot}} = \frac{Z_\sigma^2}{4Z_{\text{plate}}} \tag{6.1}$$

where Z_{slot} and Z_{plate} are the input impedances of the slot antenna and the complementary plate antenna, respectively, and Z_σ is the intrinsic impedance of the medium surrounding both antennas. When the medium is air, the intrinsic impedance is $Z_\sigma = 120\pi$ ohms $\equiv Z_0$.

For the derivation of Eq. (6.1) we use Figure 6.2, where the conducting plate in Figure 6.2a and the slot in Figure 6.2b have the same shape [1–4]. Note that the slot is cut in a conducting sheet of infinite extent and the area $A + \overline{A}$ covers the entire conducting sheet; that is, A and $\overline{A}$ are *complementary*. Also, note that $\mathbf{J}$ and $\mathbf{J}'$ in Figure 6.2a are symmetric electric currents with respect to the x-y plane and $\mathbf{M}$ and $-\mathbf{M}'$ in Figure 6.2b are antisymmetric magnetic currents. These currents generate the electric and magnetic fields denoted as ($\mathbf{E}_1$, $\mathbf{H}_1$) in Figure 6.2a and ($\mathbf{E}_2$, $\mathbf{H}_2$) in Figure 6.2b. Maxwell's equations for these two cases are [1–4]

$$\nabla \times \mathbf{E}_1 + \mathrm{j}\omega\mu\mathbf{H}_1 = 0 \tag{6.2}$$

$$\nabla \times \mathbf{H}_1 - (\mathrm{j}\omega\varepsilon + \sigma)\mathbf{E}_1 = \mathbf{J} \quad (z > 0) \tag{6.3a}$$

$$\nabla \times \mathbf{H}_1 - (\mathrm{j}\omega\varepsilon + \sigma)\mathbf{E}_1 = \mathbf{J}' \quad (z < 0) \tag{6.3b}$$

$$\nabla \times \mathbf{E}_2 + j\omega\mu\mathbf{H}_2 = -\mathbf{M} \quad (z > 0) \tag{6.4a}$$

$$\nabla \times \mathbf{E}_2 + \mathrm{j}\omega\mu\mathbf{H}_2 = \mathbf{M}' \quad (z < 0) \tag{6.4b}$$

$$\nabla \times \mathbf{H}_2 - (j\omega\varepsilon + \sigma)\mathbf{E}_2 = 0 \tag{6.5}$$

where sinusoidal time variation ($e^{j\omega t}$) is assumed with ω being the angular frequency ($\omega = 2\pi\mathrm{f}$).

The boundary conditions for the two cases are expressed as

$$\mathbf{E}_1 \times \hat{\mathbf{n}} = \mathbf{H}_1 \bullet \hat{\mathbf{n}} = 0 \quad \text{on } A \tag{6.6a}$$

$$\mathbf{E}_1 \bullet \hat{\mathbf{n}} = \mathbf{H}_1 \times \hat{\mathbf{n}} = 0 \quad \text{on } \overline{A} \tag{6.6b}$$

$$\mathbf{E}_2 \times \hat{\mathbf{n}} = \mathbf{H}_2 \bullet \hat{\mathbf{n}} = 0 \quad \text{on } \overline{A} \tag{6.7a}$$

$$\mathbf{E}_2 \bullet \hat{\mathbf{n}} = \mathbf{H}_2 \times \hat{\mathbf{n}} = 0 \quad \text{on } A \tag{6.7b}$$

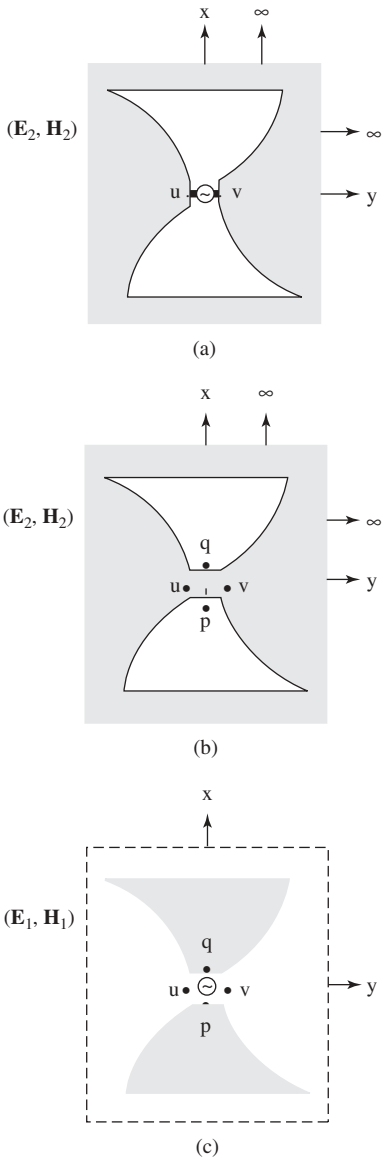


Figure 6.1 Slot antenna and its complementary plate antenna: (a) slot antenna with a voltage source, (b) slot antenna with a magnetic current source, and (c) complementary plate antenna with a voltage source.

where $\hat{\boldsymbol{n}}$ is the unit vector normal to surfaces A and $\overline{A}$. Assuming that $(\mathbf{M}, \mathbf{M}') = (\mathbf{J}, \mathbf{J}')$, we transform Eqs. (6.4a), (6.4b), and (6.5) to

$$\nabla \times (-\mathbf{E}_2) - (j\omega\varepsilon + \sigma)(Z_\sigma^2 \mathbf{H}_2) = \mathbf{J} \quad (z > 0) \tag{6.8a}$$

$$\nabla \times \mathbf{E}_2 + (j\omega\varepsilon + \sigma)(Z_\sigma^2 \mathbf{H}_2) = \mathbf{J}' \quad (z < 0) \tag{6.8b}$$

$$\nabla \times (Z_\sigma^2 \mathbf{H}_2) - j\omega\mu \mathbf{E}_2 = 0 \tag{6.9}$$

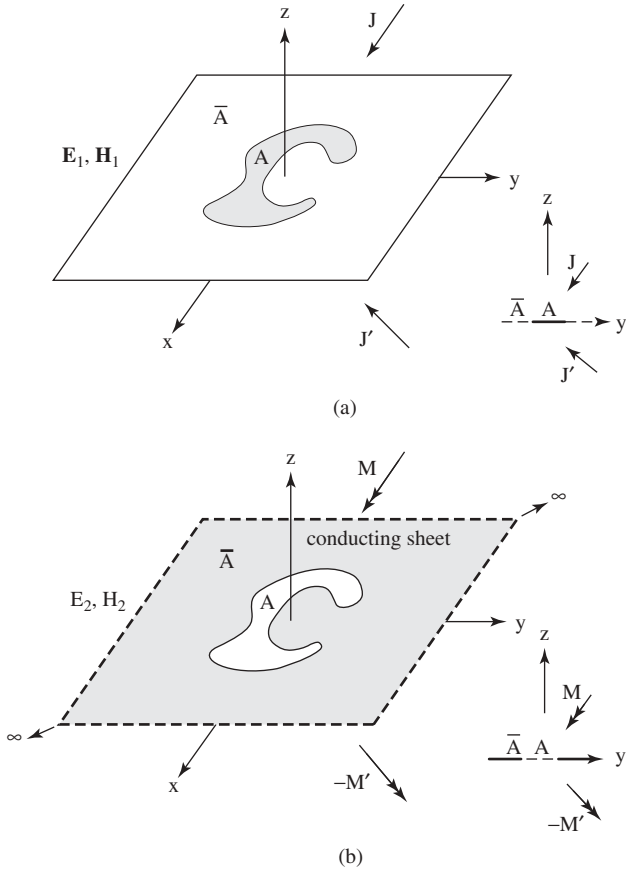


Figure 6.2 Complementary structure: (a) plate and (b) slot cut in an infinite sheet.

where Z_σ is defined as

$$Z_\sigma = \left(\frac{j\omega\mu}{j\omega\varepsilon + \sigma} \right)^{1/2} \tag{6.10}$$

Comparing Eqs. (6.2), (6.3a), and (6.3b) to Eqs. (6.9), (6.8a), and (6.8b), respectively, and taking into account the boundary conditions, Eqs. (6.6) and (6.7), we have

$$\mathbf{H}_1 = \mp \mathbf{E}_2 \quad (z \gtrless 0) \tag{6.11a}$$

$$\mathbf{E}_1 = \pm Z_\sigma^2 \mathbf{H}_2 \quad (z \gtrless 0) \tag{6.11b}$$

This means that, once the electric and magnetic fields are obtained for one of the structures of Figure 6.2, the fields for the remaining structure can be calculated using Eq. (6.11). Such a principle is referred to as *Babinet's principle in electromagnetic fields* (or *Babinet's principle*, for simplicity).

Paralleling Mushiake's original work [1], we focus on the arbitrarily shaped slot antenna shown in Figure 6.1a and denote the voltage and the current at the feed point as V_{slot} and I_{slot}, respectively. The voltage source in Figure 6.1a can be replaced with

an equivalent magnetic current source, as shown in Figure 6.1b, where the magnetic current source surrounds a conducting plate section connecting points u and v. The direction of the magnetic current viewed from the front is opposite to that of the magnetic current viewed from the back and hence the situation shown in Figure 6.1b corresponds to the situation shown in Figure 6.2b. Then, as discussed earlier, the slot structure in Figure 6.1b can be handled using its complementary plate structure shown in Figure 6.1c, where the voltage and the current at the feed point are denoted by V_{plate} and I_{plate}, respectively.

Using Babinet's principle in electromagnetic fields, we relate V_{slot} and V_{plate} to I_{plate} and I_{slot}, respectively:

$$\begin{aligned} V_{\text{slot}} &= -\int_u^v \mathbf{E}_2 \bullet d\mathbf{l} \\ &= \int_u^v \mathbf{H}_1 \bullet d\mathbf{l} \\ &= \tfrac{1}{2} I_{\text{plate}} \end{aligned} \tag{6.12}$$

$$\begin{aligned} V_{\text{plate}} &= -\int_p^q \mathbf{E}_1 \bullet d\mathbf{l} \\ &= -\int_p^q (Z_\sigma^2 \mathbf{H}_2) \bullet d\mathbf{l} \\ &= \frac{Z_\sigma^2}{2} I_{\text{slot}} \end{aligned} \tag{6.13}$$

where $d\mathbf{l}$ is a line-element vector. The impedances for the slot and plate are defined as

$$Z_{\text{slot}} = \frac{V_{\text{slot}}}{I_{\text{slot}}} \tag{6.14}$$

$$Z_{\text{plate}} = \frac{V_{\text{plate}}}{I_{\text{plate}}} \tag{6.15}$$

Then, the product of Z_{slot} and Z_{plate} is proved to be Eq. (6.1) using Eqs. (6.12) and (6.13).

6.2.2 Self-Complementary Structures

A structure whose conducting areas are made of infinitesimally thin, perfectly conducting plates of infinite extent and are congruent to the nonconducting areas (free space) is called a *self-complementary* structure [1–4]. Figure 6.3 shows antennas that have a self-complementary structure, where area A is congruent to area A' in Figure 6.3a, and areas A and B in Figure 6.3b are congruent to areas A' and B', respectively. The structure in Figure 6.3a is point symmetric with respect to the origin (center of the structure); that is, the conducting areas are made to coincide with the nonconducting areas through a 90° rotation around the z-axis. The conducting and nonconducting sections in Figure 6.3b are symmetric with respect to the x-axis.

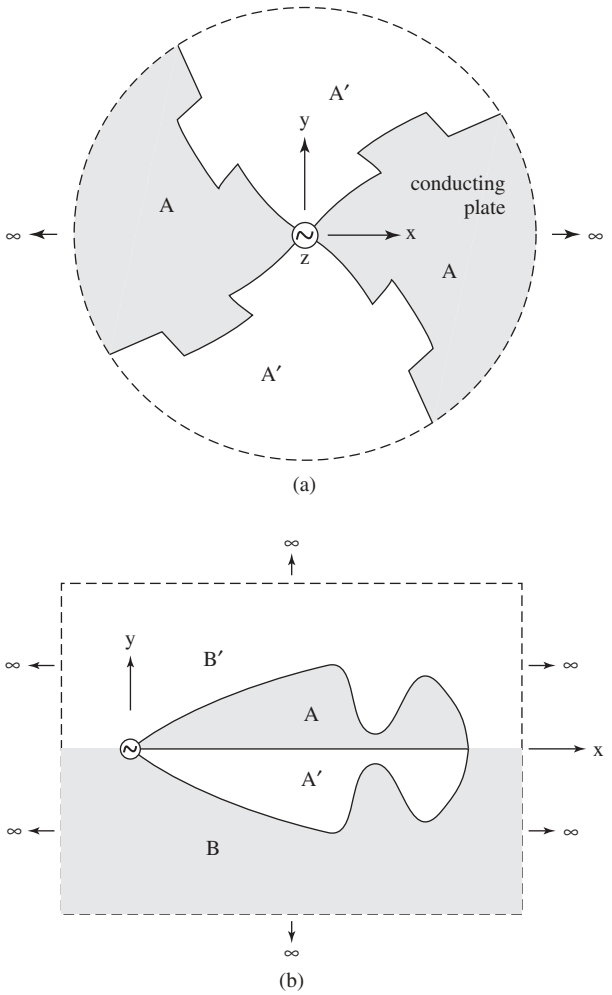


Figure 6.3 Self-complementary antennas: (a) point-symmetric self-complementary antenna and (b) line-symmetric self-complementary antenna.

The input impedance for a self-complementary antenna can be calculated from Eq. (6.1). A relationship of $Z_{\text{slot}} = Z_{\text{plate}}$ exists and hence Eq. (6.1) is

$$Z_{\text{plate}} = \frac{Z_0}{2} = 60\pi \approx 188 \text{ ohms} \tag{6.16}$$

This value is constant, with no dependence on frequency. Equation (6.16) is called *Mushiake's relationship* [1–4].

The input impedance given by Eq. (6.16) is the impedance of a two-port self-complementary antenna. Next, we investigate the input impedance of a four-port self-complementary antenna [3]. Figure 6.4a shows the feed system of a four-port self-complementary plate antenna, where four star-connected sources excite the plate arms. Figure 6.4b shows the slot antenna that is complementary to the antenna of

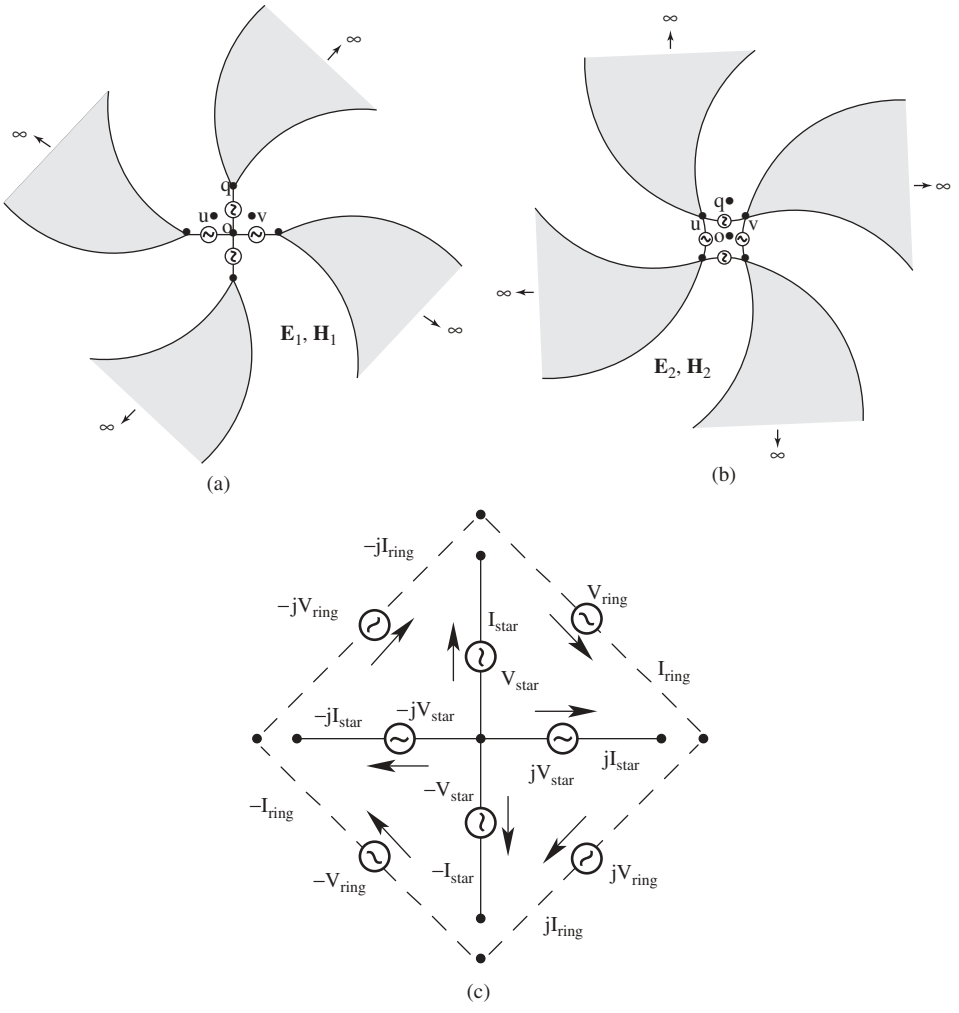


Figure 6.4 Self complementary antenna with four ports: (a) star-connected sources, (b) ring-connected sources, and (c) star (solid line) and ring (dotted line) connections [Ref. 3].

Figure 6.4a, where four ring-connected sources excite the slot arms. Corresponding to Eqs. (6.14) and (6.15), Z_{plate} and Z_{slot} for the four-port case are written

$$Z_{\text{plate}} = -\int_0^q \mathbf{E}_1 \bullet d\mathbf{l} \Big/ 2\int_u^v \mathbf{H}_1 \bullet d\mathbf{l} \tag{6.17}$$

$$Z_{\text{slot}} = -\int_u^v \mathbf{E}_2 \bullet d\mathbf{l} \Big/ 2\int_q^0 \mathbf{H}_2 \bullet d\mathbf{l} \tag{6.18}$$

Using Eqs. (6.11a) and (6.11b), we obtain the product $Z_{\text{plate}}\, Z_{\text{slot}}$:

$$Z_{\text{plate}} Z_{\text{slot}} = (Z_0/2)^2 \tag{6.19}$$

Interestingly, the product $Z_{\text{plate}}Z_{\text{slot}}$ for the four-port self-complementary antenna is the same as that for the two-port self-complementary antenna (Eq. (6.1)). However, Mushiake finds that Z_{plate} (and Z_{slot}) for the four-port case differs from that for the two-port case [3]:

$$Z_{\text{plate}} = Z_0/2\sqrt{2} \quad \text{(four-port case)} \tag{6.20}$$

This is proved by using n-polyphase alternating current theory, which represents the relationships between the voltages and currents of the star and ring connections (see Figure 6.4c):

$$V_{\text{ring}} = -V_{\text{star}} + jV_{\text{star}} \tag{6.21}$$

$$I_{\text{ring}} = -I_{\text{star}} - jI_{\text{ring}} \tag{6.22}$$

Using these equations, the slot impedance Z_{slot} is expressed as

$$\begin{aligned} Z_{\text{slot}} &= \frac{V_{\text{ring}}}{I_{\text{ring}}} \\ &= 2\frac{V_{\text{star}}}{I_{\text{star}}} \\ &= 2Z_{\text{plate}} \end{aligned} \tag{6.23}$$

Substituting Eq. (6.23) into Eq. (6.19), we obtain Eq. (6.20).

As seen above, n-polyphase alternating current theory is used in the antenna impedance calculations. Note that self-complementary antennas having more than four ports are also handled using n-polyphase alternating current theory. Mushiake derives the input impedance of n-port self-complementary antennas as [3]

$$Z_{\text{plate}} = Z_0/4\,\sin(m\pi/n) \tag{6.24}$$

where the voltage sources in the star connection are expressed as V_{star}, δV_{star}, $\delta^2 V_{\text{star}}$, $\delta^3 V_{\text{star}}, \ldots,$ and $\delta^{n-1}V_{\text{star}}$ with $\delta = \exp(j2m\pi/n)$, where m is called the *mode number*. The antenna shown in Figure 6.3 corresponds to a case of $(n, m) = (4, 1)$.

6.2.3 Principle of Similitude for Electromagnetic Fields

Figure 6.5 shows two antenna systems, 1 and 2, operating at frequencies f_1 and f_2, respectively, with a current $\mathbf{J}_0$ flowing on each antenna. Antenna system 1 is characterized by permittivity ε, permeability μ, conductivity σ_1, and length ℓ_1, while antenna system 2 is characterized by ε, μ, σ_2, and ℓ_2. Note that antenna 2 is geometrically similar to antenna 1, with $\ell_2 \equiv K\ell_1$. We denote the electric and magnetic fields from antenna systems 1 and 2 as $(\mathbf{E}_1, \mathbf{H}_1)$ and $(\mathbf{E}_2, \mathbf{H}_2)$, respectively, and assume that $\mathbf{J}_0$, $(\mathbf{E}_1, \mathbf{H}_1)$, and $(\mathbf{E}_2, \mathbf{H}_2)$ vary with a sinusoidal time variation $(e^{j\omega t})$.

Maxwell's equations for antenna system 1 are

$$\nabla \times \mathbf{E}_1 = -j\omega_1\mu\mathbf{H}_1 \tag{6.25}$$

$$\nabla \times \mathbf{H}_1 = \mathbf{J}_0 + (\sigma_1 + j\omega_1\varepsilon)\mathbf{E}_1 \tag{6.26}$$

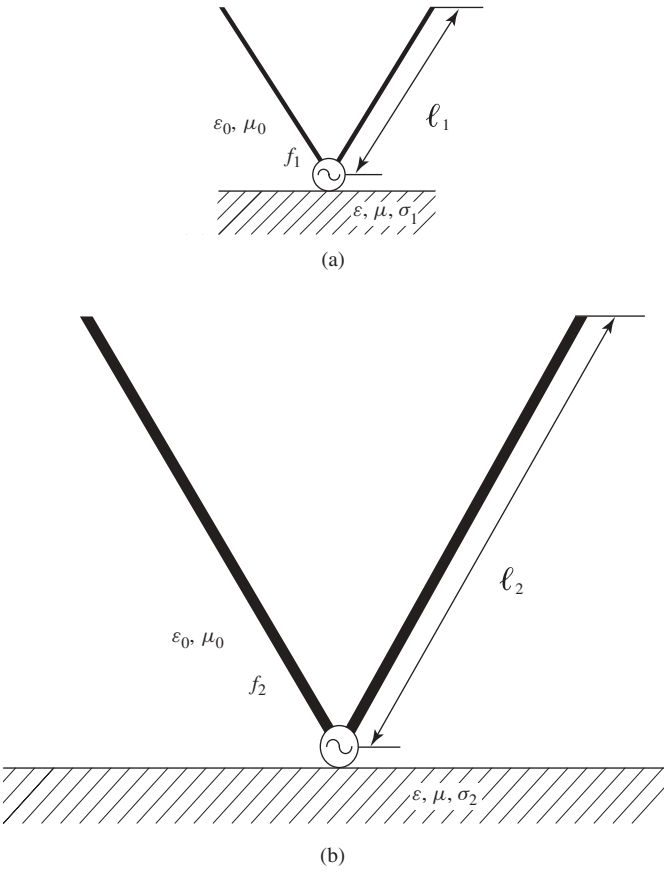


Figure 6.5 Similar antennas: (a) antenna system 1 operating at f_1 and (b) antenna system 2 operating at frequency f_2.

and those for antenna system 2 are

$$\nabla \times \mathbf{E}_2 = -j\omega_2\mu\mathbf{H}_2 \tag{6.27}$$

$$\nabla \times \mathbf{H}_2 = \mathbf{J}_0 + (\sigma_2 + j\omega_2\varepsilon)\mathbf{E}_2 \tag{6.28}$$

For antenna system 2, we adopt new coordinates (x', y', z') that are K times larger than the original coordinates (x, y, z). Under the new coordinates, we denote the electric and magnetic fields for antenna system 2 as $\mathbf{E}'$ and $\mathbf{H}'$, respectively. Then, Maxwell's equations with a new operational element ∇' are [1, 4]

$$\frac{1}{K}\nabla' \times (K\mathbf{E}') = -j\omega_2\mu(K\mathbf{H}') \tag{6.29}$$

$$\frac{1}{K}\nabla' \times (K\mathbf{H}') = \mathbf{J}_0 + (\sigma_2 + j\omega_2\varepsilon)(K\mathbf{E}') \tag{6.30}$$

If the following conditions are satisfied,

$$K\omega_2 = \omega_1 \tag{6.31}$$

$$K\sigma_2 = \sigma_1 \tag{6.32}$$

Eqs. (6.29) and (6.30) can be written

$$\nabla' \times \mathbf{E}' = -j\omega_1\mu\mathbf{H}' \tag{6.33}$$

$$\nabla' \times \mathbf{H}' = \mathbf{J}_0 + (\sigma_1 + j\omega_1\varepsilon)\mathbf{E}' \tag{6.34}$$

Equations (6.33) and (6.34) for antenna system 2 are of the same form as Eqs. (6.25) and (6.26) for antenna system 1, respectively. This means that the electric and magnetic fields for antenna systems 1 and 2 are similar to each other. This is called the *principle of similitude for electromagnetic fields*.

Equations (6.31) and (6.32) are written

$$\frac{\ell_2}{\lambda_2} = \frac{\ell_1}{\lambda_1} \tag{6.35}$$

$$\sigma_2\lambda_2 = \sigma_1\lambda_1 \tag{6.36}$$

where λ_1 and λ_2 are the wavelengths at frequencies f_1 and f_2, respectively. When both conductivities are zero ($\sigma_1 = \sigma_2 = 0$) or infinite ($\sigma_1 = \sigma_2 = \infty$), the principle of similitude is specified with only Eq. (6.35). This leads to the fact that, if all *electrical* lengths for the antenna structure are the same at frequencies f_1 and f_2, the antenna characteristics at these frequencies are the same.

6.2.4 Self-Similar Structures

Consider an antenna whose size is increased (or decreased) by a factor K. If the resulting structure is the same as the original structure, the original antenna structure is called a *self-similar structure*. An antenna that has a self-similar structure is called a *self-similar antenna* [4].

Generally, a self-similar antenna can be produced using powers of K, that is, K^s ($s = \pm 1, \pm 2, \ldots$) [4]. In this case, as seen from the results in Section 6.2.3, the input impedance and radiation characteristics observed at a frequency f_0 repeatedly appear at frequencies $\mathrm{K}^{-s} f_0$; that is, the antenna characteristics are periodic with respect to the logarithm of the frequency. It should be emphasized that the impedance and radiation pattern of this antenna do not necessarily remain constant at the frequencies within the period [3, 4, 35–36].

A biconical antenna, whose conductors are assumed to be of infinite extent, is a special case of a self-similar antenna. The biconical structure can be reproduced using any nonzero factor K (K is continuous). It follows that the biconical antenna shows constant input impedance and constant radiation characteristics irrespective of the frequency. In other words, the antenna acts as a frequency-independent antenna.

6.2.5 Transformation of Radial Distance to Angles

This section summarizes an equation for frequency-independent antennas, paralleling Rumsey's original work [5] and Elliot's later work [6].

We investigate an antenna whose feed terminals are infinitesimally close to the origin of a spherical coordinate system (r, θ, ϕ). The conductors of which the antenna is comprised are symmetrically located with respect to the z-axis. These antenna conductors are assumed to be perfectly conducting and surrounded by an infinite medium that is homogeneous and isotropic. We start with expressing the surface of this antenna at frequency f_1 as

$$r = F(\theta, \phi) \tag{6.37}$$

The antenna specified by Eq. (6.37) is designated as the original antenna.

If we wish to scale the original antenna to a lower frequency f_2 $(=f_1/K)$, the new antenna for f_2 must be made K times larger, as described in Section 6.2.3.

$$r' = Kr = KF(\theta, \phi) \tag{6.38}$$

The new antenna is, by definition, similar to the original antenna. In addition, it can be made congruent, if the antenna conductors are infinite. For congruence,

$$KF(\theta, \phi) = F(\theta, \phi + C) \tag{6.39}$$

where C is the rotation angle around the z-axis.

To obtain the function $F(\theta, \phi)$ that satisfies Eq. (6.39), first, we differentiate Eq. (6.39) with respect to C, and then with respect to ϕ,

$$\begin{aligned} \frac{\partial K}{\partial C} F(\theta, \phi) &= \frac{\partial F(\theta, \phi + C)}{\partial C} \\ &= \frac{\partial F(\theta, \phi + C)}{\partial (\phi + C)} \end{aligned} \tag{6.40}$$

$$\begin{aligned} K \frac{\partial F(\theta, \phi)}{\partial \phi} &= \frac{\partial F(\theta, \phi + C)}{\partial \phi} \\ &= \frac{\partial F(\theta, \phi + C)}{\partial (\phi + C)} \end{aligned} \tag{6.41}$$

From Eqs. (6.40) and (6.41), we have

$$\frac{\partial K}{\partial C} F(\theta, \phi) = K \frac{\partial F(\theta, \phi)}{\partial \phi} \tag{6.42}$$

Using Eq. (6.37), Eq. (6.42) is written

$$\frac{1}{K} \frac{\partial K}{\partial C} = \frac{1}{r} \frac{\partial r}{\partial \phi} \tag{6.43}$$

or

$$a = \frac{1}{r}\frac{\partial r}{\partial \phi} \tag{6.44}$$

where

$$a = \frac{1}{K}\frac{\partial K}{\partial C} \tag{6.45}$$

Equation (6.45) is independent of θ and ϕ.

A general solution to Eq. (6.44) is obtained as

$$\begin{aligned} r &= F(\theta, \phi) \\ &= e^{a\phi} f(\theta) \end{aligned} \tag{6.46}$$

where $f(\theta)$ is an arbitrary function.

Note that the right-hand side of Eq. (6.46) is specified only in terms of angle parameters θ and ϕ. It does not include the parameters for length and frequency. This fact implies that the radiation pattern and input impedance remain unchanged irrespective of the frequency; that is, the antenna is frequency independent.

6.3 ANALYSIS TECHNIQUES

The constant impedance of Eq. (6.24) is obtained under the assumption that the antenna structure is of *infinite* extent. Note that the impedance changes when the infinite structure is made *finite* (i.e., when the structure is truncated). The truncation of the infinite structure also causes the radiation pattern and gain to deviate from those obtained for the infinite structure.

Generally, it is difficult to analytically determine the effects of the finite structure on the impedance, radiation pattern, and gain. However, this difficulty is overcome using numerical methods. In this section, two analysis techniques, one based on the method of moments (MoM) and one based on the finite-difference time-domain method (FDTDM), are briefly summarized. The MoM is suitable for analyzing an arbitrarily shaped finite-length wire antenna (Sections 6.4.3 and 6.5.3), where a free-space Green's function is used to obtain the current distribution, while the FDTDM is powerful for analyzing a finite-size plate antenna with and without a conducting cavity (Sections 6.4.1, 6.4.2, 6.4.4, 6.5.1, and 6.5.2) for which it is difficult to obtain a Green's function. Detailed information on these methods can be found in the literature [7–11].

6.3.1 Analysis Based on Method of Moments

Figure 6.6 shows a curved wire (from s_S to s_E) in free space, where the wire is assumed to be perfectly conducting, and the wire radius ρ is very thin relative to the operating wavelength λ. This thin wire assumption implies that the current flows only in the wire axis direction (there is no circumferential component).

We denote the distance measured along the wire from the origin (that can arbitrarily be chosen on the wire) to a point on the wire as s', and the current at this point as $I(s')$.

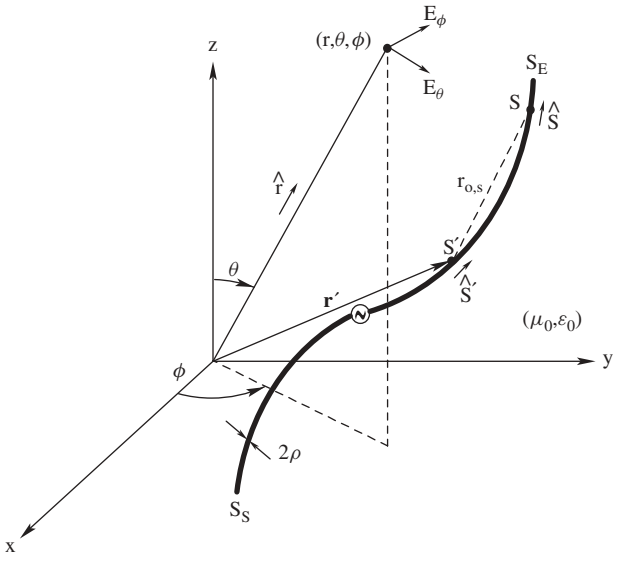


Figure 6.6 Arbitrarily shaped thin wire antenna.

If the current $I(s')$ is a known function, the θ and ϕ components of the radiation field $\mathbf{E}_{\text{rad}}(r, \theta, \phi)$ generated from $I(s')$ can be represented as

$$E_\theta(r,\theta,\phi) = -\frac{j\omega\mu_0}{4\pi}\frac{e^{-jk_0r}}{r}\hat{\boldsymbol{\theta}} \bullet \int_{s_S}^{s_E} \hat{\mathbf{s}}' I(s') e^{jk_0\hat{\mathbf{r}}\bullet\mathbf{r}'} ds' \tag{6.47}$$

$$E_\phi(r,\theta,\phi) = -\frac{j\omega\mu_0}{4\pi}\frac{e^{-jk_0r}}{r}\hat{\boldsymbol{\phi}} \bullet \int_{s_S}^{s_E} \hat{\mathbf{s}}' I(s') e^{jk_0\hat{\mathbf{r}}\bullet\mathbf{r}'} ds' \tag{6.48}$$

where (r, θ, ϕ) and $(\hat{\mathbf{r}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}})$ are the spherical coordinates and their unit vectors, respectively; $j^2 = -1$; ω is the angular frequency ($= 2\pi f$, where f is the frequency); μ_0 is the permeability of free space; k_0 ($k_0^2 = \omega^2\mu_0\varepsilon_0$, where ε_0 is the permittivity of free space) is the phase constant or wavenumber ($k_0 = 2\pi/\lambda$); $\hat{\mathbf{s}}'$ is the unit vector tangential to the wire axis at the source point; and the vector $\mathbf{r}'$ is the position vector from the coordinate origin to the source point.

The radiation field $\mathbf{E}_{\text{rad}}(r, \theta, \phi) = E_\theta(r, \theta, \phi)\hat{\boldsymbol{\theta}} + E_\phi(r, \theta, \phi)\hat{\boldsymbol{\phi}}$, which is a form suitable for a linearly polarized antenna, can be transformed to a form suitable for a circularly polarized (CP) antenna,

$$\mathbf{E}_{\text{rad}}(r,\theta,\phi) = E_{\text{R}}(r,\theta,\phi)(\hat{\boldsymbol{\theta}} - j\hat{\boldsymbol{\phi}}) + E_{\text{L}}(r,\theta,\phi)(\hat{\boldsymbol{\theta}} + j\hat{\boldsymbol{\phi}}) \tag{6.49}$$

where the first term represents a right-hand CP wave component and the second represents a left-hand CP wave component. Using these two components, the axial ratio is given as $\text{AR} = \{|E_{\text{R}}| + |E_{\text{L}}|\}/\{\|E_{\text{R}}| - |E_{\text{L}}\|\}$.

The gain relative to a circularly polarized isotropic antenna (absolute gain) is defined as $G_a = (\{|E_\theta|^2 + |E_\phi|^2)/2\}r^2/30P_{\text{in}}$, where P_{in} is the power observed at the antenna input terminals (not at the starting point of the feed line, where the power source

is connected). Using $|E_\theta|^2 + |E_\phi|^2 = 2(|E_R|^2 + |E_L|^2)$, the absolute gain is written $G_a = (|E_R|^2 + |E_L|^2)\ r^2/30P_{in}$. Therefore the absolute gains relative to right-hand and left-hand circularly polarized isotropic antennas can be written $G_{a(R)} = |E_R(r, \theta, \phi)|^2\ r^2/30P_{in}$ and $G_{a(L)} = |E_L\ (r, \theta, \phi)|^2\ r^2/30P_{in}$, respectively.

For $|E_R| > |E_L|$, as the axial ratio AR decreases, $G_{a(R)}$ increases and $G_{a(L)}$ decreases, according to $G_{a(R)} = G_a/[1 + \{(AR-1)/(AR+1)\}^2]$ and $G_{a(L)} = G_a/[1 + \{(AR+1)/(AR-1)\}^2]$. Note that $G_{a(R)} = G_a$ and $G_{a(L)} = 0$ when AR = 1 (i.e., the polarization is perfectly circular with a right-hand sense).

The input impedance is given as $Z_{in} = V_{in}/I_{in}$, where V_{in} is the voltage across the antenna input terminals and I_{in} is the current flowing at the antenna input terminals. Generally, V_{in} is a known function, and hence I_{in} must be obtained to calculate the input impedance.

As seen from the above, we must obtain the current distribution to calculate the antenna characteristics, including the radiation pattern, axial ratio, gain, and input impedance.

For obtaining the current, we derive the electric field generated from the current and use the boundary condition that the tangential component of the electric field on the perfectly conducting wire surface is zero. The boundary condition for the antenna shown in Figure 6.6 is expressed as the following integral equation:

$$\frac{1}{j4\pi\omega\varepsilon_0}\int_{s_S}^{s_E} I(s')\left(-\frac{\partial^2}{\partial s\,\partial s'} + k_0^2\hat{\mathbf{s}}\bullet\hat{\mathbf{s}}'\right)\frac{e^{-jk_0r_{o,s}(s,s')}}{r_{o,s}(s,s')}ds' = -E_s^i(s) \tag{6.50}$$

where $E_s^i(s)$is the tangential component of an incident electric field on the wire; s is the distance measured along the wire from the origin to an observation point on the wire; $\hat{\mathbf{s}}$ is the unit vector tangential to the wire axis at the observation point; and $r_{o,s}(s, s')$ is the distance between the observation and source points. Note that the explanation of the other notations is given after Eq. (6.48).

To obtain the current $I(s')$ in Eq. (6.50), the MoM is employed with $I(s') = \Sigma\, I_n J_n(s')$, where $J_n(s')$ and I_n $(n = 1, 2, \ldots, N)$ are the *expansion functions* and *unknown coefficients of the expansion functions*, respectively. Then, Eq. (6.50) becomes

$$\sum_{n=1}^{N} I_n e_n(s) = -E_s^i(s) \tag{6.51}$$

where

$$e_n(s) = \frac{1}{j4\pi\omega\varepsilon_0}\int_{s_S}^{s_E} J_n(s')\left(-\frac{\partial^2}{\partial s\,\partial s'} + k_0^2\hat{\mathbf{s}}\bullet\hat{\mathbf{s}}'\right)\frac{e^{-jk_0r_{o,s}(s,s')}}{r_{o,s}(s,s')}ds' \tag{6.52}$$

Multiplying both sides of Eq. (6.52) by weighting functions W_m(s) $(m = 1, 2, \ldots, N)$ and integrating the multiplied results over the wire length from s_S to s_E, we have the matrix equation

$$[Z_{mn}][I_n] = [V_m] \tag{6.53}$$

where $[Z_{mn}]$, $[I_n]$, and $[V_m]$ are called the *impedance, current*, and *voltage matrices*, respectively.

$$Z_{mn} = \int_{s_S}^{s_E} e_n(s) W_m(s) ds \tag{6.54}$$

$$V_m = -\int_{s_S}^{s_E} E_s^i(s) W_m(s) ds \tag{6.55}$$

From Eq. (6.53), we have $[I_n] = [Z_{mn}]^{-1}[V_m]$. This means that the unknown coefficients of the expansion functions (I_n) can be found and hence the current $I(s')$ in Eq. (6.50) can be obtained. Note that the analysis of the wire antennas in Sections 6.4.3 and 6.5.3 adopts piecewise sinusoidal functions [9] for expansion functions $J_n(s')$ and weighting functions W_m(s).

6.3.2 Analysis Based on Finite-Difference Time-Domain Method

In Section 6.3.1, the current distributed along the antenna is used for calculating the antenna characteristics. In this section, the FDTDM is used to obtain the electric and magnetic fields within an analysis space that surrounds the antenna. These fields are in turn used to calculate the antenna characteristics.

Figure 6.7a shows the analysis space, whose volume is expressed as $X \times Y \times Z$. This analysis space is subdivided into numerous cells, each having a volume of $\Delta x \times \Delta y \times \Delta z$. For applying Maxwell's equations to the analysis space, we assign the field components (E_x, E_y, E_z) and (H_x, H_y, H_z) to each cell, as shown in Figure 6.7b, where the Cartesian coordinates are expressed as $(x, y, z) = (i\,\Delta x, j\,\Delta y, k\,\Delta z)$, with i, j, and k being integers. Note that absorbing conditions (ABC) [11] are used to terminate the outer surface of this analysis space.

We express the time coordinate t as $n\ \Delta t$ with n being an integer, and denote a function $g(x, y, z, t) = g(i\ \Delta x, j\ \Delta y, k\ \Delta z, n\ \Delta t)$ as $g^n(i, j, k)$. Then, the derivative of $g(x, y, z, t)$ with respect to x is expressed using a central finite difference as

$$\frac{\partial g^n(i, j, k)}{\partial x} \approx \frac{g^n\left(i + \frac{1}{2}, j, k\right) - g^n\left(i - \frac{1}{2}, j, k\right)}{\Delta x} \tag{6.56}$$

Similarly, the derivatives of the electric and magnetic fields with respect to time t are expressed as

$$\left.\frac{\partial \mathbf{E}}{\partial t}\right|_{t=(n-1/2)\Delta t} \approx \frac{\mathbf{E}^n - \mathbf{E}^{n-1}}{\Delta t} \tag{6.57}$$

$$\left.\frac{\partial \mathbf{H}}{\partial t}\right|_{t=n\,\Delta t} \approx \frac{\mathbf{H}^{n+1/2} - \mathbf{H}^{n-1/2}}{\Delta t} \tag{6.58}$$

Using Eqs. (6.57) and (6.58), Maxwell's curl $\mathbf{E}$ equation at time $t = n\,\Delta t$ and curl $\mathbf{H}$ equation at $t = (n - \frac{1}{2})\Delta t$,

$$\nabla \times \mathbf{E}|_{t=n\Delta t} = -\mu \left.\frac{\partial \mathbf{H}}{\partial t}\right|_{t=n\Delta t} \tag{6.59}$$

$$\nabla \times \mathbf{H}|_{t=(n-1/2)\Delta t} = \sigma \mathbf{E}|_{t=(n-1/2)\Delta t} + \varepsilon \left.\frac{\partial \mathbf{E}}{\partial t}\right|_{t=(n-1/2)\Delta t} \tag{6.60}$$

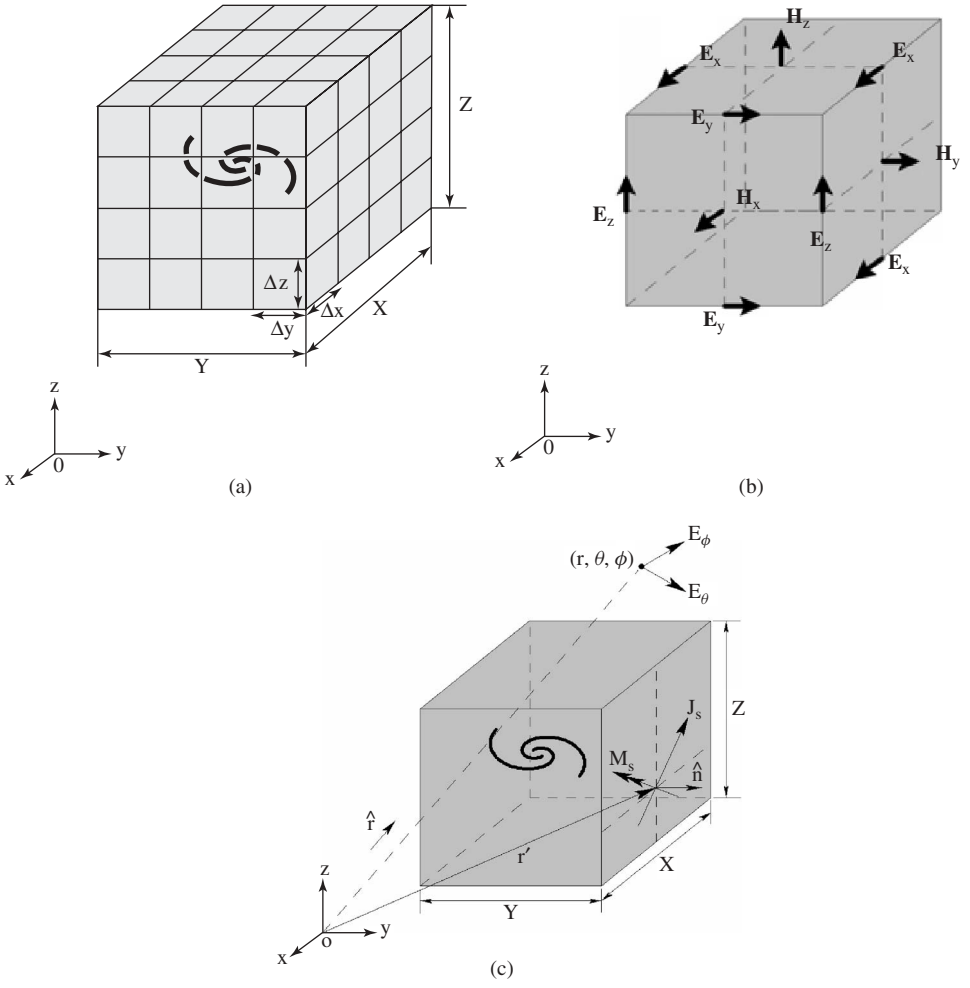


Figure 6.7 Analysis space: (a) subdivision into numerous cells, (b) assignment of field components to a cell, and (c) equivalence theorem.

are transformed to

$$\mathbf{H}^{n+1/2} = \mathbf{H}^{n-1/2} - \frac{\Delta t}{\mu}\nabla \times \mathbf{E}^n \tag{6.61}$$

$$\mathbf{E}^n = \frac{2\varepsilon - \sigma\,\Delta t}{2\varepsilon + \sigma\,\Delta t}\mathbf{E}^{n-1} + \frac{2\,\Delta t}{2\varepsilon + \sigma\,\Delta t}\nabla \times \mathbf{H}^{n-1/2} \tag{6.62}$$

where ε, μ, and σ are the permittivity, permeability, and conductivity, respectively. Note that $\sigma\mathbf{E}$ in Eq. (6.60) is evaluated at $t = (n - \frac{1}{2})\Delta t$. This term is approximated by the average of $\sigma\mathbf{E}^n$ and $\sigma\mathbf{E}^{n-1}$ to obtain Eq. (6.62).

Using Eq. (6.56), the finite-difference form for the x-component of Eq. (6.62) is written

$$\begin{aligned} E_x^n\left(i+\tfrac{1}{2}, j, k\right) = {} & e_{x/0}\left(i+\tfrac{1}{2}, j, k\right) E_x^{n-1}\left(i+\tfrac{1}{2}, j, k\right) \\ & + e_{x/y}\left(i+\tfrac{1}{2}, j, k\right)\left[H_z^{n-1/2}\left(i+\tfrac{1}{2}, j+\tfrac{1}{2}, k\right)\right. \\ & \left. - H_z^{n-1/2}\left(i+\tfrac{1}{2}, j-\tfrac{1}{2}, k\right)\right] \\ & - e_{x/z}\left(i+\tfrac{1}{2}, j, k\right)\left[H_y^{n-1/2}\left(i+\tfrac{1}{2}, j, k+\tfrac{1}{2}\right)\right. \\ & \left. - H_y^{n-1/2}\left(i+\tfrac{1}{2}, j, k-\tfrac{1}{2}\right)\right] \end{aligned} \tag{6.63}$$

where

$$e_{x/0}\left(i+\tfrac{1}{2}, j, k\right) = \frac{2\varepsilon\left(i+\tfrac{1}{2}, j, k\right) - \sigma\left(i+\tfrac{1}{2}, j, k\right)\Delta t}{2\varepsilon\left(i+\tfrac{1}{2}, j, k\right) + \sigma\left(i+\tfrac{1}{2}, j, k\right)\Delta t} \tag{6.64}$$

$$e_{x/y}\left(i+\tfrac{1}{2}, j, k\right) = \frac{2\,\Delta t}{2\varepsilon\left(i+\tfrac{1}{2}, j, k\right) + \sigma\left(i+\tfrac{1}{2}, j, k\right)\Delta t}\frac{1}{\Delta y} \tag{6.65}$$

$$e_{x/z}\left(i+\tfrac{1}{2}, j, k\right) = \frac{2\,\Delta t}{2\varepsilon\left(i+\tfrac{1}{2}, j, k\right) + \sigma\left(i+\tfrac{1}{2}, j, k\right)\Delta t}\frac{1}{\Delta z} \tag{6.66}$$

The two other components of the electric field $\mathbf{E}^n$ and the three components of the magnetic field $\mathbf{H}^{n+1/2}$ are similarly formulated. The solutions to Maxwell's equations, $\mathbf{E}$ and $\mathbf{H}$, are obtained by iterating Eqs. (6.61) and (6.62) with $n = 1, 2, \ldots$ until these fields become constant (converge). This method is called the finite-difference time-domain method (FDTDM).

The current along the antenna conductor, $I(t)$, can be calculated by integrating the magnetic field obtained with the FDTDM around the antenna conductor (Ampere's law). Also, the voltage between two points, $V(t)$, can be calculated by taking the product of the electric field obtained with the FDTDM and the distance between the two points. Based on these techniques, the input impedance $Z_{\text{in}} = \text{R}_{\text{in}} + jX_{\text{in}}$ is obtained using $\mathscr{F}[V_{\text{in}}(t)]/\mathscr{F}[I_{\text{in}}(t)]$, where $\mathscr{F}[V_{\text{in}}(t)]$ and $\mathscr{F}[I_{\text{in}}(t)]$ are the Fourier-transformed antenna input voltage and current, respectively.

Based on the $\mathbf{E}$ and $\mathbf{H}$ obtained with the FDTDM, the radiation field is calculated using the equivalence theorem [12]. The θ and ϕ components of the radiation field (in the frequency domain) are given as

$$E_\theta(\omega) = -\frac{jk_0}{4\pi}\frac{e^{-jk_0 r}}{r}[Z_0\mathbf{N}(\omega)\bullet\hat{\boldsymbol{\theta}} + \mathbf{L}(\omega)\bullet\hat{\boldsymbol{\phi}}] \tag{6.67}$$

$$E_\phi(\omega) = -\frac{jk_0}{4\pi}\frac{e^{-jk_0 r}}{r}[Z_0\mathbf{N}(\omega)\bullet\hat{\boldsymbol{\phi}} - \mathbf{L}(\omega)\bullet\hat{\boldsymbol{\theta}}] \tag{6.68}$$

where Z_0 is the intrinsic impedance ($120\pi\ \Omega$), k_0 is the wavenumber, and

$$\mathbf{N}(\omega) = \int_{\text{closed surface}} \mathbf{J}_s(\omega, \mathbf{r}')e^{jk_0\,\hat{\mathbf{r}}\bullet\mathbf{r}'}dS' \tag{6.69}$$

$$\mathbf{L}(\omega) = \int_{\text{closed surface}} \mathbf{M}_s(\omega, \mathbf{r}')e^{jk_0\,\hat{\mathbf{r}}\bullet\mathbf{r}'}dS' \tag{6.70}$$

Note that the $\mathbf{J}_s(\omega, \mathbf{r}')$ and $\mathbf{M}_s(\omega, \mathbf{r}')$ are, respectively, the Fourier transforms of the time-domain electric current density $\mathbf{J}_s(t, \mathbf{r}') = \hat{\mathbf{n}} \times \mathbf{H}(t, \mathbf{r}')$ and magnetic current density $\mathbf{M}_s(t, \mathbf{r}') = \mathbf{E}(t, \mathbf{r}') \times \hat{\mathbf{n}}$, where $\mathbf{r}'$ is the position vector from the coordinate origin to the point located on the closed surface of the analysis space, and $\hat{\boldsymbol{n}}$ is the outward unit vector normal to the closed surface (see Figure 6.7c).

Thus the input impedance and the radiation field are calculated, as described above. The remaining characteristics to be calculated are the axial ratio and absolute gain. These can be calculated using the same techniques described in Section 6.3.1.

6.4 SPIRAL ANTENNAS

This section presents two types of spiral antennas: a planar type (2D type) spiral antenna and a conical type (3D type) spiral antenna. The former is called an *equiangular* spiral antenna (Section 6.4.1) and the latter is called a *conical equiangular* spiral antenna (Section 6.4.3). The radiation from an equiangular spiral antenna located in free space is bidirectional. In Section 6.4.2, a technique for transforming a bidirectional beam to a unidirectional beam is discussed. Topics related to the planar spiral antenna are presented in Section 6.4.4.

The first practical equiangular and conical spiral antennas were constructed by J. D. Dyson in 1958, implementing Rumsey's transformation for a frequency-independent antenna (see Section 6.2.5). The experimental results for these antennas are found in Refs. 13–15.

6.4.1 Equiangular Spiral Antenna

We focus on Eq. (6.46) and let a derivative of $f(\theta)$ with respect to angle θ be $df/d\theta = r_0 e^{-a\phi_0}\delta(\theta - \pi/2)$. Then, r is

$$r = r_0 e^{a(\phi - \phi_0)}, \quad \text{if } \theta = \pi/2 \tag{6.71}$$

$$r = 0, \quad \text{if } \theta \neq \pi/2 \tag{6.72}$$

where r_0 and ϕ_0 are arbitrary constants. Figure 6.8 illustrates the curve defined by Eq. (6.71), where the angle made between the radial line and the tangent, Ψ, is given as

$$\Psi = \tan^{-1}(1/a) \tag{6.73}$$

Since angle Ψ is constant, the curve in Figure 6.8 is called an *equiangular spiral*.

A two-arm spiral with $\phi_0 = 0$ and π is shown in Figure 6.9a, and a four-arm spiral with $\phi_0 = 0, \pi/2, \pi$, and $3\pi/2$ is shown in Figure 6.9b. The winding angle ϕ of each spiral must range from $-\infty$ to $+\infty$ for the antenna to be frequency independent. However, in reality, the spiral arms cannot be made infinitely long (corresponding to $\phi = +\infty$). Also, the spiral arms cannot start from the center point (corresponding to $\phi = -\infty$), because space is required for the feed system.

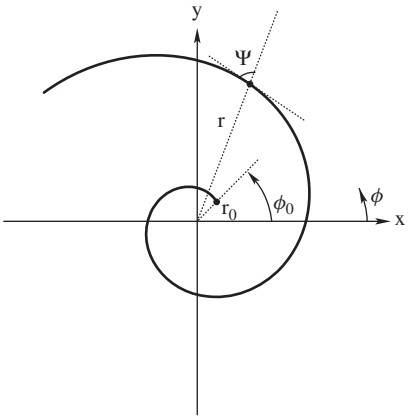


Figure 6.8 Equiangular spiral curve.

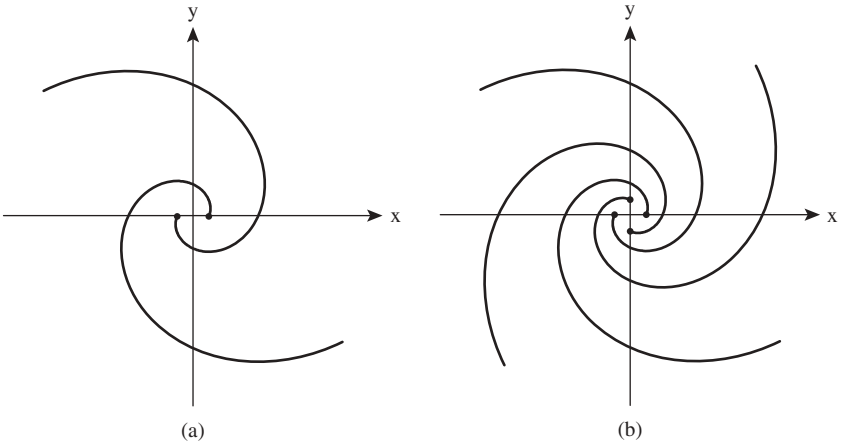


Figure 6.9 Equiangular spirals: (a) two-arm spiral and (b) four-arm spiral.

Figure 6.10 shows a two-arm equiangular slot spiral cut in a conducting sheet (perfect electric conductor—PEC). The sheet is finite and round with diameter D. The edges of the slots are defined as

$$r_1 = r_0 e^{a\phi_1}, \quad \phi(Q_1) \le \phi_1 \le \phi(P_1) \tag{6.74}$$

$$r_2 = r_0 e^{a(\phi_2 - \delta)}, \quad \phi(Q_2) = \phi(Q_1) + \delta \le \phi_2 \le \phi(P_2) \tag{6.75}$$

$$r_3 = r_0 e^{a(\phi_3 - \pi)}, \quad \phi(Q_3) = \phi(Q_1) + \pi \le \phi_3 \le \phi(P_3) = \phi(P_1) + \pi \tag{6.76}$$

$$r_4 = r_0 e^{a(\phi_4 - \delta - \pi)}, \quad \phi(Q_4) = \phi(Q_2) + \pi \le \phi_4 \le \phi(P_4) = \phi(P_2) + \pi \tag{6.77}$$

where the spherical coordinate ϕ at point Q_i is denoted as $\phi(Q_i)$ $(i = 1, 2, 3, 4)$, and ϕ at point P_i is denoted as $\phi(P_i)$ $(i = 1, 2, 3, 4)$. Note that δ in this structure is chosen

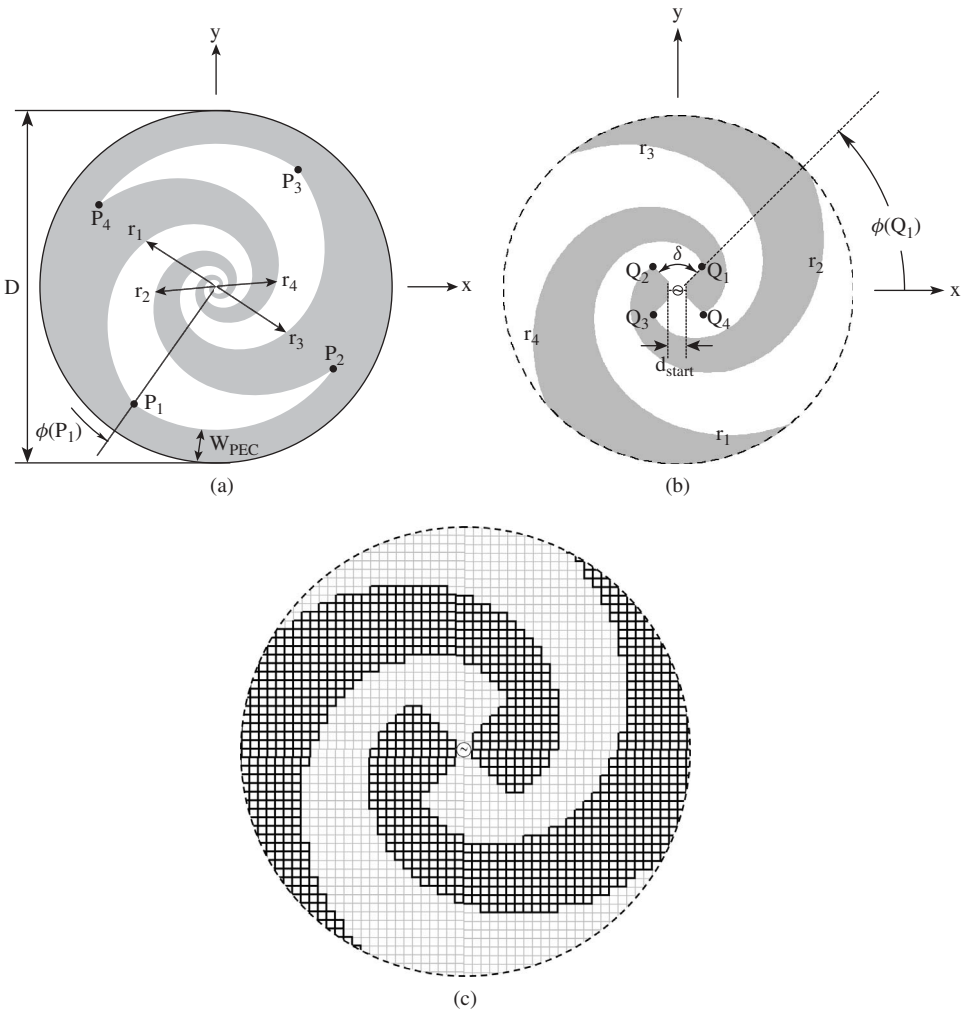


Figure 6.10 Equiangular slot spiral antenna: (a) slots cut in a conducting sheet, (b) magnified view of the input region, and (c) staircase approximation.

to be $\pi/2$ (rad) and a segment of a circle of radius $r = r_0 e^{a\phi(P_1)} \equiv r(P_1)$ is used for expressing edges P_1P_2 and $P_3P_4 (= P_1P_2)$.

We analyze this equiangular spiral antenna using the FDTDM, where a staircase approximation is used for the antenna structure (see Figure 6.10c). The configuration parameters are shown in Table 6.1.

Figure 6.11 shows the frequency response for the input impedance Z_{in} $(= R_{\text{in}} + jX_{\text{in}})$, where three values of sheet diameter D are used: $D = 140$ mm $= 2.8\lambda_6$ (case A), $D =$ 120 mm $= 2.4\lambda_6$ (case B), and $D = 113.7$ mm $= 2.274\lambda_6$ (case C), where λ_6 is the free-space wavelength at 6 GHz. Note that the distance from the sheet edge to the slot edge for case C is $W_{\text{PEC}} = 0$. It is found that the input impedance is constant over a broad frequency range for these three values of D. The input impedance is approximately

TABLE 6.1 Configuration Parameters of an Equiangular Spiral Antenna

Symbol	Value	Unit
d_{start}	1.0	mm
r_0	1.5	mm
a	0.35	rad^{-1}
δ	0.5π	rad
$\phi(Q_1)$	0.25π	rad
$\phi(P_1)$	3.306π	rad
$\phi(P_2)$	3.806π	rad

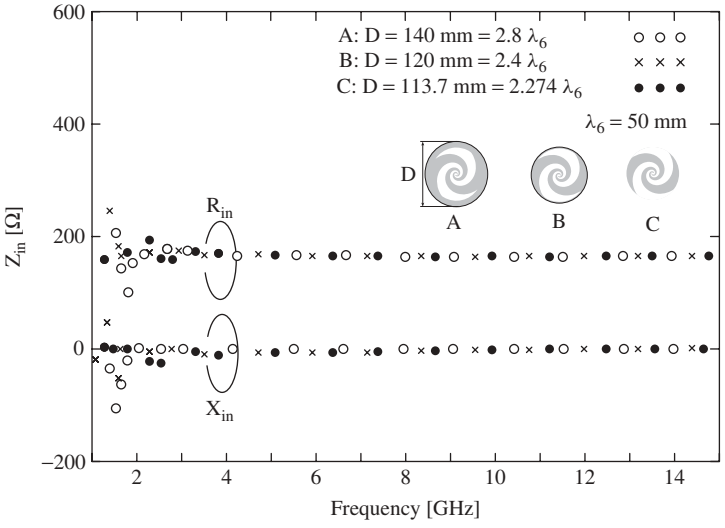


Figure 6.11 Frequency response of the input impedance (calculated result). (Courtesy of K. Kikkawa, Ref. 16.)

160 ohms, which is close to the theoretical value (60π ohms) for a self-complementary antenna. The deviation from 60π ohms is attributed to the following facts: (1) the curved sections of the theoretical structure are approximated using staircases; (2) the spacing between the feed terminals in the present structure is finite, whereas the spacing in the theoretical structure is infinitesimally small (delta-gap); (3) the spiral arms in the present structure are of finite length, whereas the arms in the theoretical structure are of infinite length; and (4) the conducting sheet in the present structure is finite, whereas the sheet in the theoretical structure is of infinite extent.

Figure 6.12 shows the electric current density $\mathbf{J}_s$ (amplitude $|\mathbf{J}_s|$) on the surface of a conducting sheet for case A (sheet diameter $D = 140$ mm). It is revealed that, as the frequency increases, the region where the current contributes most to the radiation (active region) moves to the inner region. Conversely, as the frequency decreases, the active region moves outward. This implies that the low end of the operating frequency

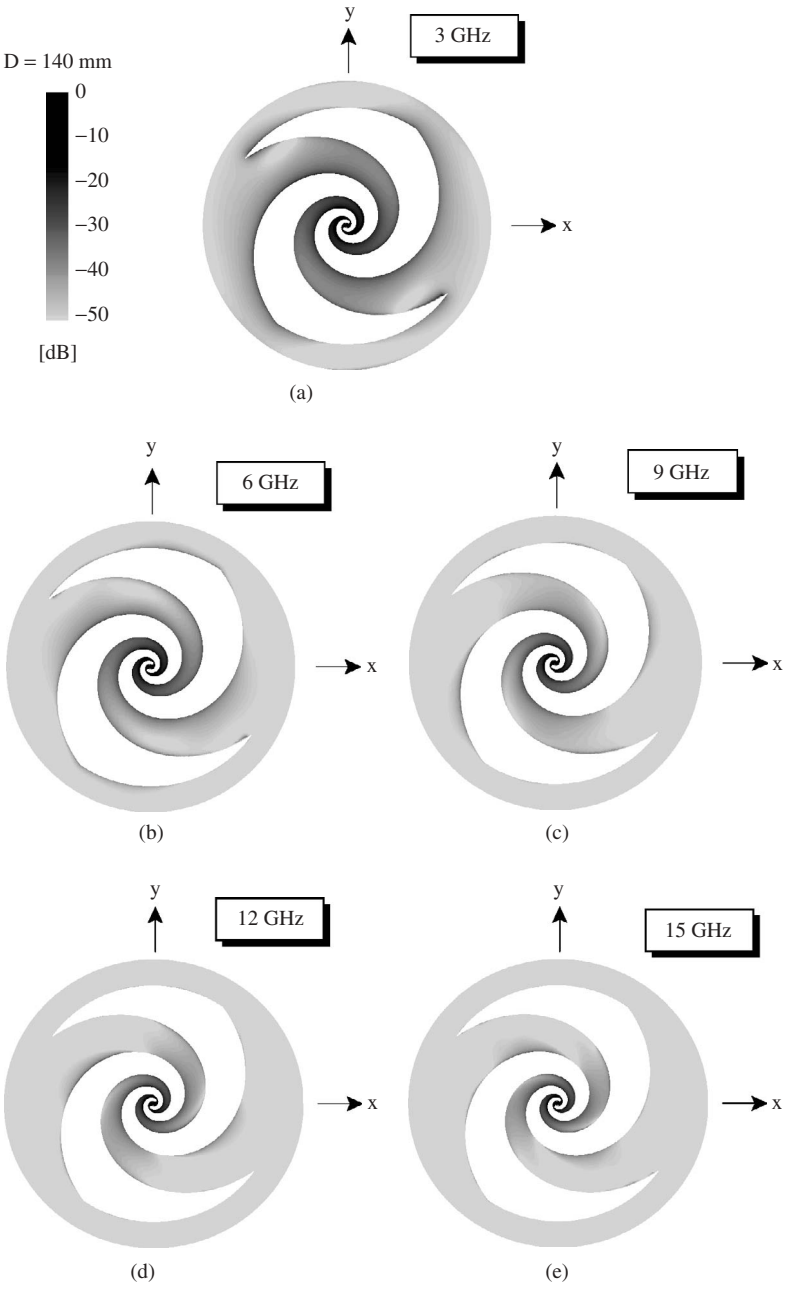


Figure 6.12 Electric current density $|\mathbf{J}_s|$ for $D = 140$ mm (calculated result): (a) 3 GHz, (b) 6 GHz, (c) 9 GHz, (d) 12 GHz, and (e) 15 GHz [Ref. 16].

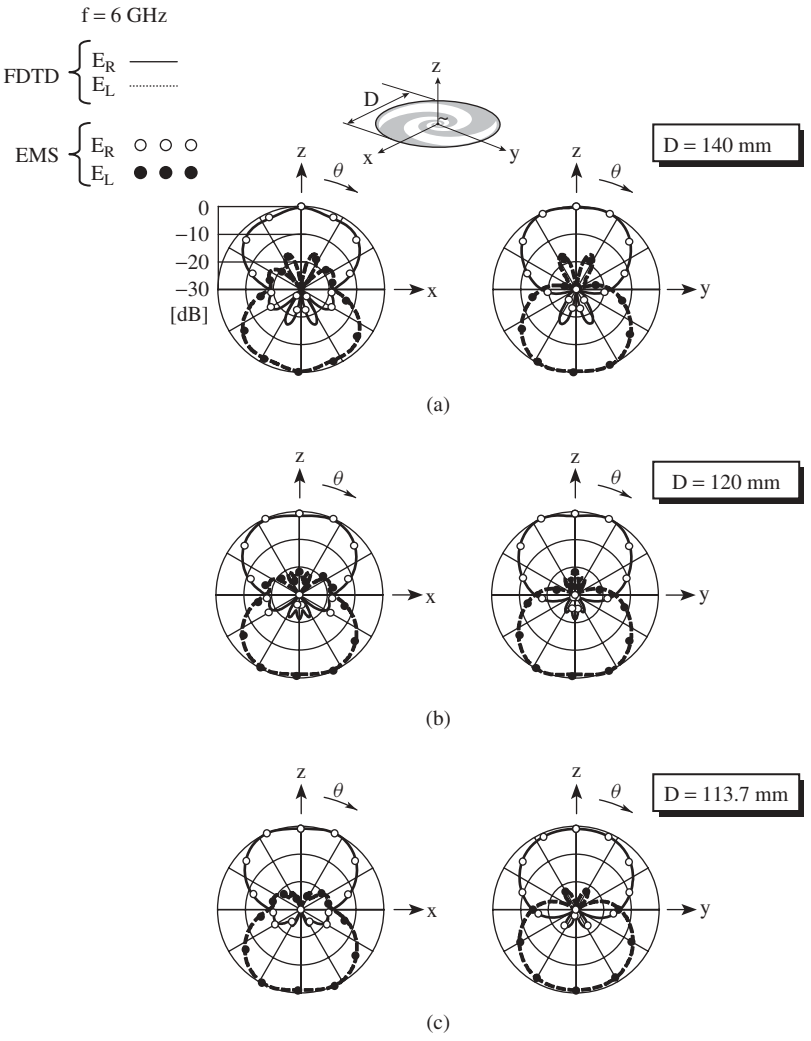


Figure 6.13 Radiation patterns at 6 GHz (calculated result): (a) $D = 140$ mm, (b) $D = 120$ mm, and (c) $D = 113.7$ mm [Ref. 16].

band depends on the outmost radius $r(P_1)$, and the high end depends on the structure of the feed terminals.

The radiation from the spiral is bidirectional. Figure 6.13 shows the radiation patterns for cases A, B, and C, where the radiation field is decomposed into two components: a right-hand circularly polarized wave component E_{R} and a left-hand circularly polarized wave component E_{L}. The main radiation in the positive z-direction is circularly polarized with a right-hand sense, which corresponds to the rotational sense of the current flow. Figure 6.14 shows the axial ratios in the z-direction for cases A, B, and C. Within an analysis range of 3–15 GHz, the axial ratio is less than 3 dB for all three cases.

As can be seen in Figure 6.15 the gains (absolute gains) for cases A, B, and C exhibit similar behavior (wave-like variation). As the frequency decreases, the variation

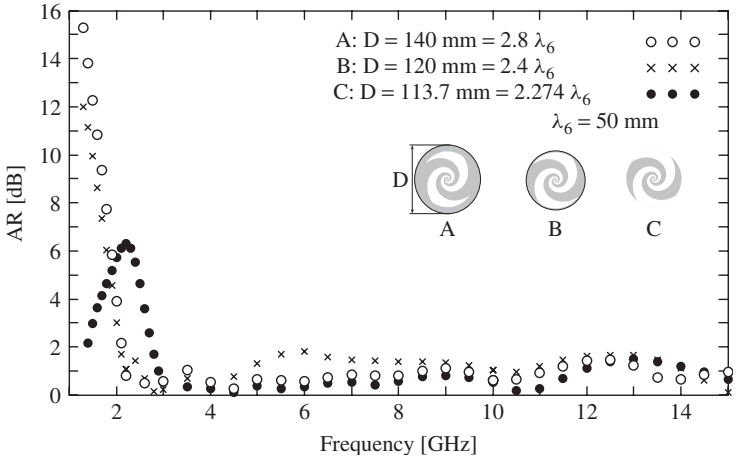


Figure 6.14 Axial ratio as a function of frequency (calculated result.) [Ref. 16].

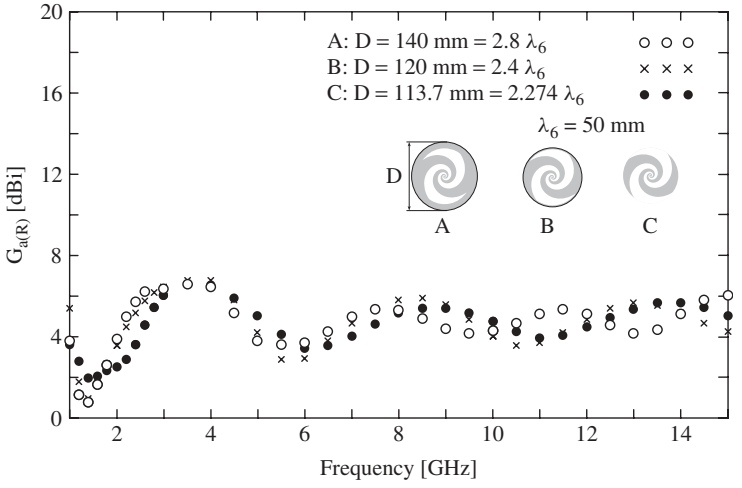


Figure 6.15 Gain as a function of frequency (calculated result.) [Ref. 16].

increases. It is found that the gain for case A (sheet diameter $D = 140$ mm) is approximately 5.0 dBi with a variation of ± 2.0 dB, over a broad frequency range of 2–15 GHz.

6.4.2 Cavity-Backed Equiangular Spiral Antenna

As can be seen in Figure 6.13, the radiation from the spiral in Section 6.4.1 is bidirectional. For unidirectional communications, the bidirectional beam must be transformed into a unidirectional beam. This section deals with a technique for transforming this bidirectional beam into a unidirectional beam using a cavity.

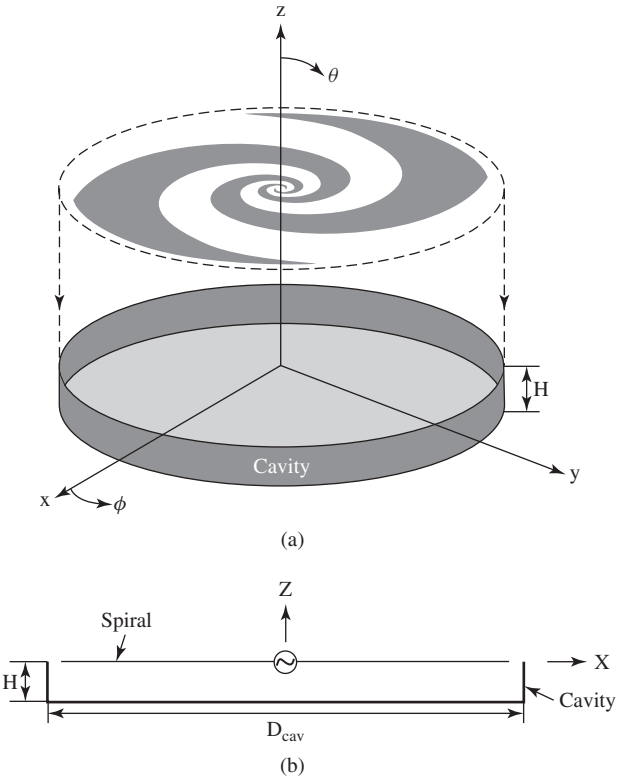


Figure 6.16 Equiangular spiral backed by a cavity, designated as a *spiral-cv*: (a) exploded view and (b) side view.

Figure 6.16 shows the spiral of case C (discussed in Section 6.4.1; see the inset of Figure 6.11: $D = 113.7$ mm and $W_{\text{PEC}} = 0$) backed by a conducting cavity. This is referred to as a *spiral-cv* throughout this section. The configuration parameters for the arms in Figure 6.16 are shown in Table 6.1. Note that the cavity is specified by height H and diameter D_{cav}; these are fixed to be $H = 7$ mm and $D_{\text{cav}} = 120$ mm in the following analysis.

The wave radiated in the negative z-direction from the spiral arms is reflected at the bottom of the cavity and radiates out of the cavity (free space), combining with the wave radiated directly from the spiral arms toward free space. Thus the radiation from the spiral-cv becomes unidirectional, as desired. However, the use of the cavity deteriorates the constant input impedance and axial ratio characteristics that are inherent to the spiral. The dotted lines in Figures 6.17 and 6.18 clearly show the deterioration in the input impedance and axial ratio, of the spiral-cv, respectively. This deterioration is attributed to the fact that the **E** and **H** fields inside the cavity affect the current distribution on the conducting arms; that is, the current flowing in the absence of a cavity is disrupted.

To restore the constant input impedance and radiation characteristics, an absorbing strip (ABS) is attached to the vertical wall of the cavity, as shown in Figure 6.19 where

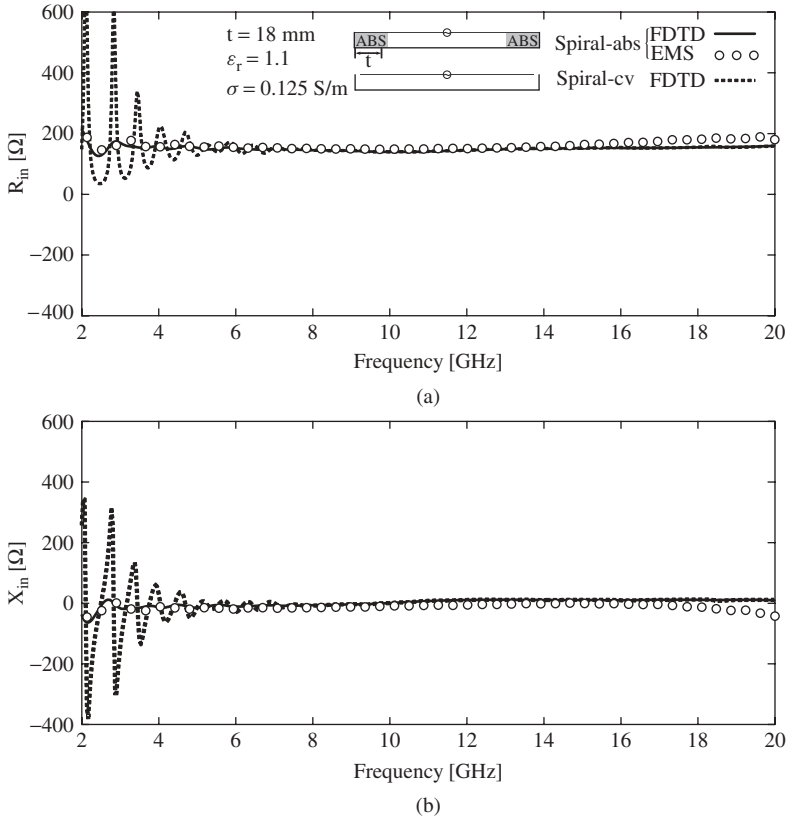


Figure 6.17 Input impedance (calculated result). Dotted line is for the spiral-cv and solid line is for the spiral-abs. (Courtesy of K. Kikkawa, Ref. 16.)

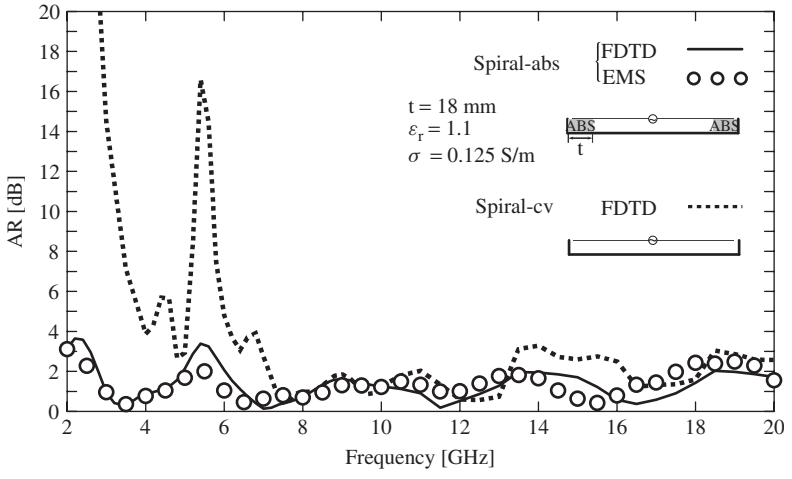


Figure 6.18 Axial ratio (calculated result). Dotted line is for the spiral-cv and solid line is for the spiral-abs. (Courtesy of K. Kikkawa, Ref. 16.)

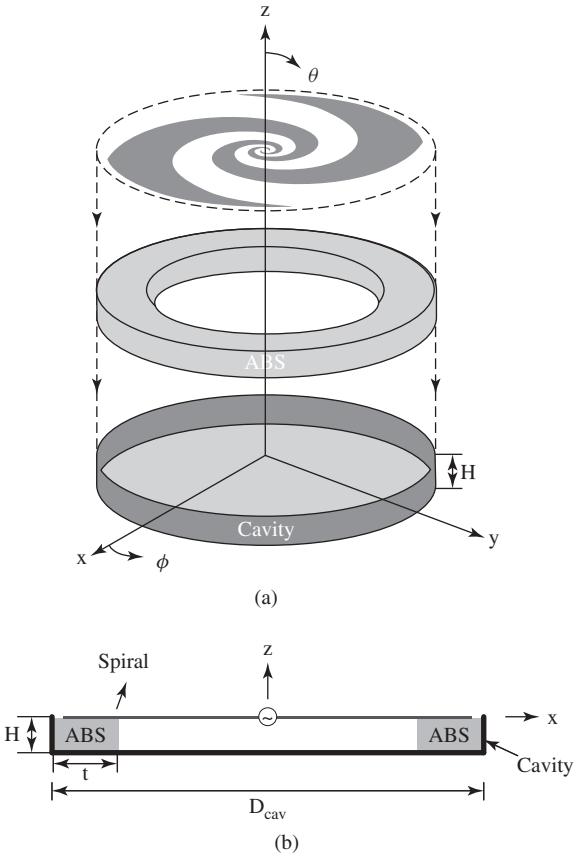


Figure 6.19 Equiangular spiral antenna backed by a cavity with an absorbing strip. The antenna is designated as a *spiral-abs*: (a) Exploded view and (b) side view.

the ABS is specified by thickness t, relative permittivity ε_r, and conductivity σ. This antenna is designated as a *spiral-abs* to distinguish it from the spiral-cv.

The solid lines in Figures 6.17 and 6.18 show the input impedance and axial ratio of the spiral-abs, respectively, where parameters t, ε_r, and σ are shown inside Figures 6.17 and 6.18. It can be seen that the input impedance and the axial ratio are restored to constant values, compared with those of the spiral-cv.

Figure 6.20 shows the radiation patterns of the spiral-abs as a function of frequency. The spiral-abs radiates a unidirectional beam with a low cross-polarization component E_{L}. It is found that the unidirectional beam is circularly polarized over a wide region around the z-axis.

As can be seen in Figure 6.19, the cavity is not completely filled with absorbing material. If we fill the cavity with absorbing material, the power radiated from the conducting arms toward the cavity is absorbed. This causes a loss of 3 dB in the gain. The present cavity is only partially filled with a strip absorber of thickness t, so it follows that the loss in gain is small. Figure 6.21 shows the gain of the spiral-abs as a function of frequency. A gain (absolute gain) of more than 5 dBi is obtained within a broad frequency range of 3–16 GHz.

The power input to the spiral-abs is not completely transformed into the power radiated into free space. Some of the input power is absorbed in the absorbing strip, resulting in a deterioration in the radiation efficiency η_{rad}, where η_{rad} is defined as the ratio of the power radiated into free space (P_{rad}) to the power input to the antenna terminals (P_{in}). Figure 6.22 shows the radiation efficiency as a function of frequency. The increasing radiation efficiency with frequency is due to the fact that the active region moves to the inner region of the spiral arms and is less affected by the absorbing strip.

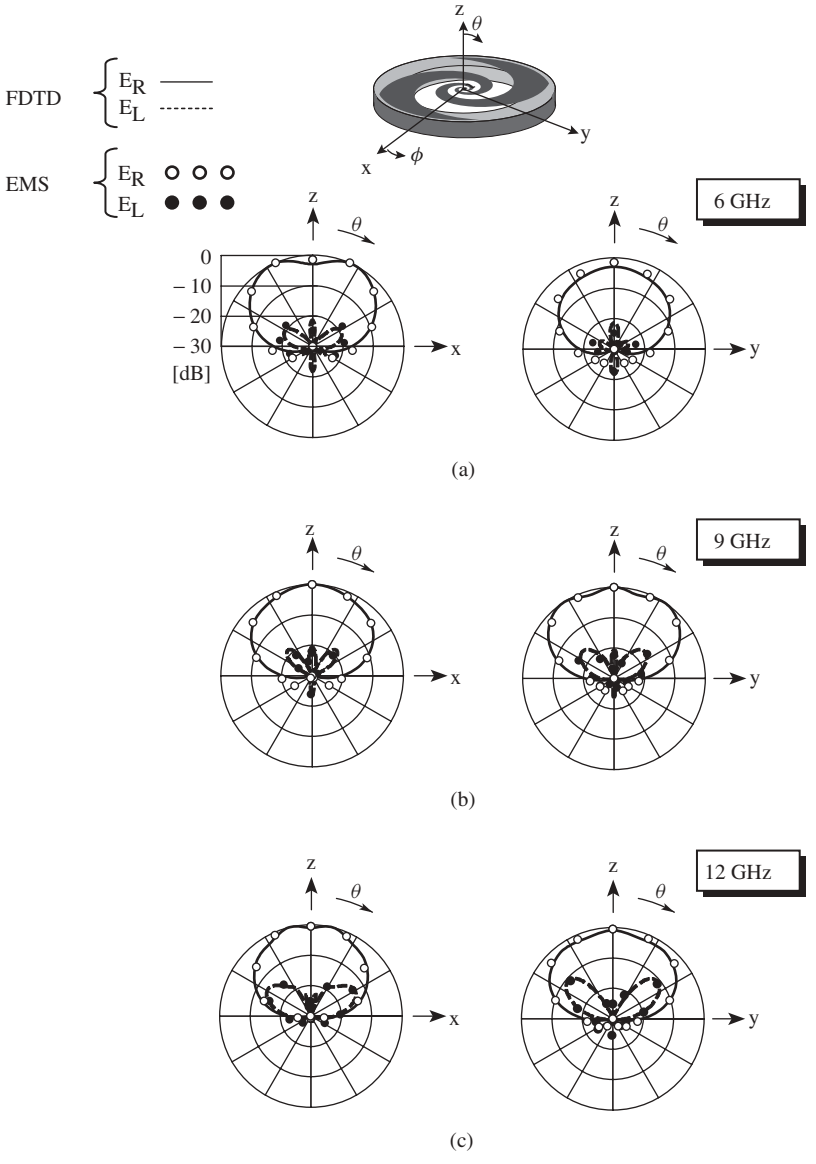
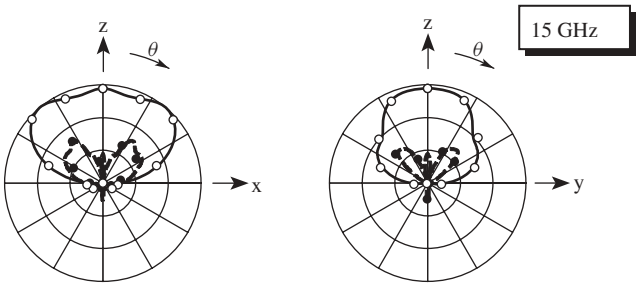
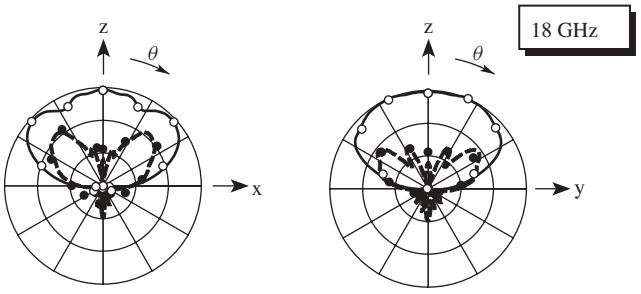


Figure 6.20 Radiation patterns of the spiral-abs as a function of frequency (calculated result): (a) 6 GHz, (b) 9 GHz, (c) 12 GHz, (d) 15 GHz, and (e) 18 GHz [Ref. 16].



(d)



(e)

Figure 6.20 (*Continued*)

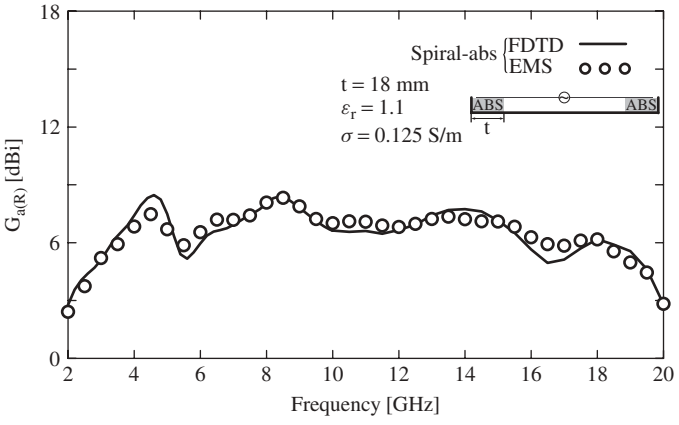


Figure 6.21 Gain of the spiral-abs as a function of frequency (calculated result) [Ref. 16].

6.4.3 Conical Equiangular Spiral Antenna

Figure 6.23 shows a conical equiangular spiral antenna made of two wire arms, where $2\theta_0$ is the cone angle. The arms are defined as

$$r_1 = r_0 e^{(a \sin\theta_0)\phi_1}, \qquad \phi(Q_1) \le \phi_1 \le \phi(P_1) \tag{6.78}$$

$$r_2 = r_0 e^{(a \sin\theta_0)(\phi_2 - \pi)}, \quad \phi(Q_2) = \phi(Q_1) + \pi \le \phi_2 \le \phi(P_2) = \phi(P_1) + \pi \tag{6.79}$$

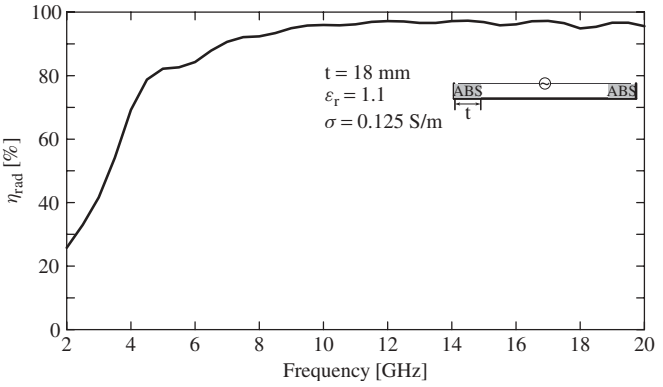


Figure 6.22 Radiation efficiency of the spiral-abs as a function of frequency (calculated result) [Ref. 16].

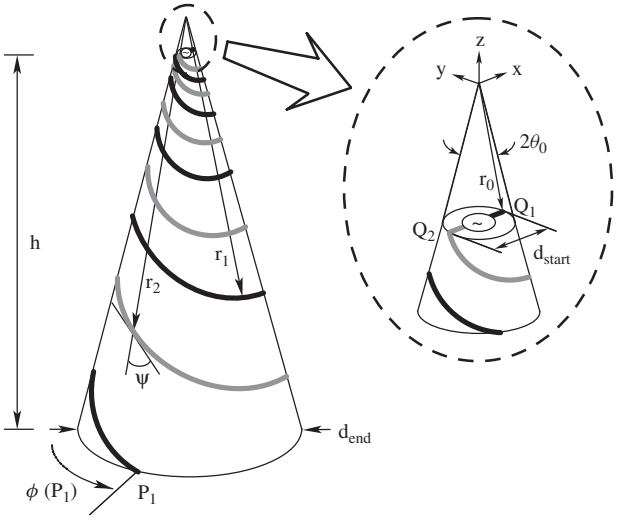


Figure 6.23 Conical equiangular spiral antenna.

where r_0 and a are constants; ϕ (Q_1) and ϕ (Q_2) are the angles at the arm starting points Q_1 and Q_2, respectively; and ϕ (P_1) and ϕ (P_2) are the angles at the arm end points P_1 and P_2, respectively. As in the planar equiangular spiral, the angle between the tangent to the antenna arm and the mother line is

$$\tan\psi = 1/a \tag{6.80}$$

The configuration parameters of the conical equiangular spiral to be analyzed here are shown in Table 6.2, where 2ρ is the diameter of the antenna arm wire, d_{start} and d_{end}, respectively, are the diameters of the top and bottom of the conical structure. The analysis is performed using the MoM.

Figure 6.24 shows the current distributions at representative frequencies, where $|I|$ is the amplitude of the current I ($I = I_r + jI_i$). It is found that the current attenuates

TABLE 6.2 Configuration Parameters of a Conical Equiangular Spiral Antenna

Symbol	Value	Unit
$2\theta_0$	30	degrees
a	0.364	rad^{-1}
$\phi(Q_1)$	0	rad
$\phi(Q_2)$	3.14	rad
$\phi(P_1)$	27.93	rad
$\phi(P_2)$	31.07	rad
d_{start}	1.5	mm
d_{end}	20.9	mm
2ρ	0.125	mm
h	36.1	mm
r_0	2.90	mm

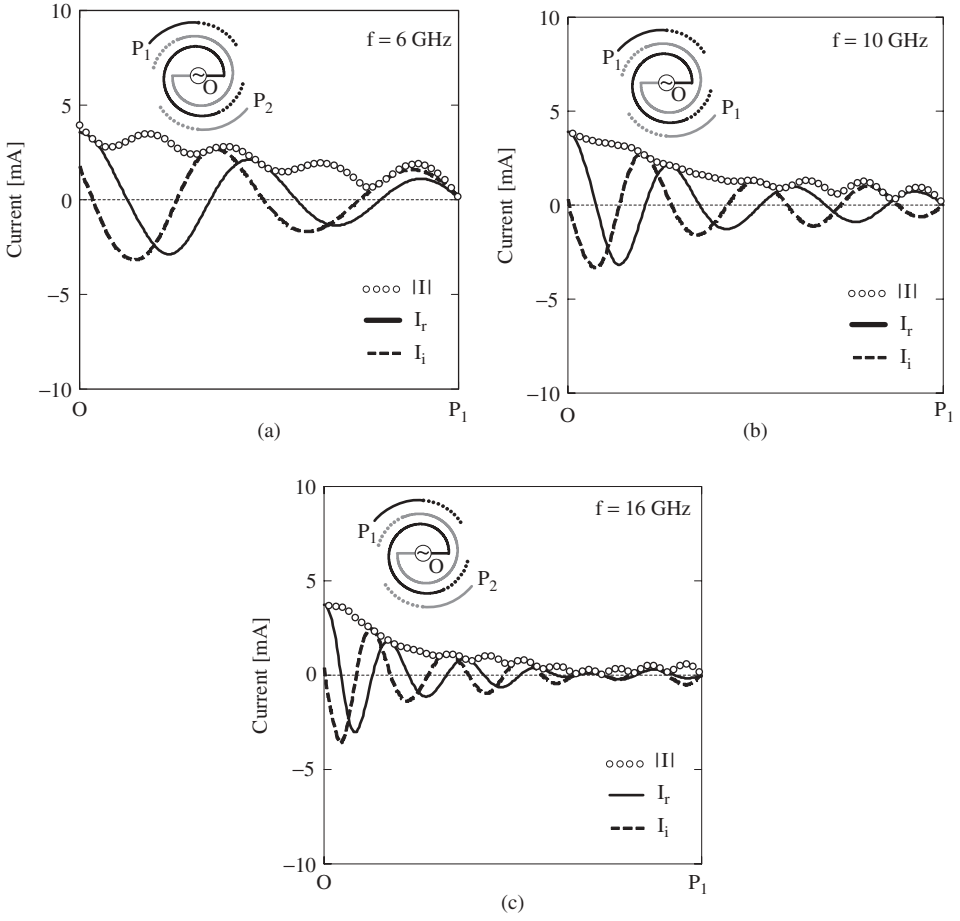


Figure 6.24 Current distributions of the conical equiangular spiral antenna (calculated result): (a) 6 GHz, (b) 10 GHz, and (c) 16 GHz, (Courtesy of H. Osada and G. Tsutsumi [Ref. 17].)

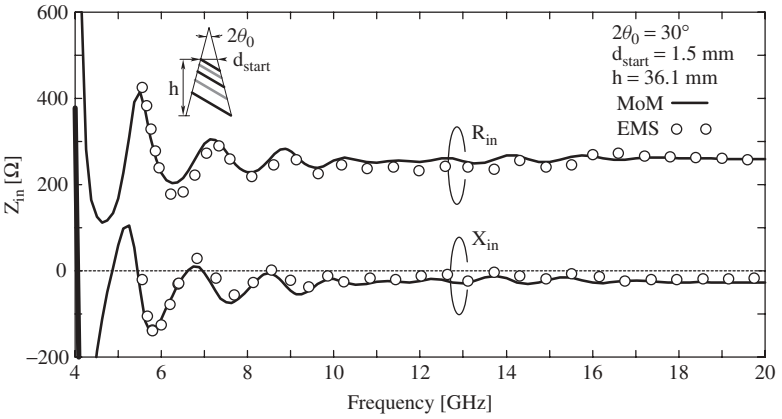


Figure 6.25 Input impedance of the conical equiangular spiral antenna (calculated result) [Ref. 17].

smoothly from the feed point (o) to the arm ends. The smooth attenuation means that the current reflected back to the feed point from the arm ends is small and does not disturb the current at the feed point (the antenna input terminals). This leads to a constant input impedance characteristic and a low cross-polarization characteristic, as observed in Figures 6.25 and 6.26, respectively. Note that the input impedance at low frequencies varies due to the fact that the arms are of finite length and the reflected current increases as the frequency decreases. Also, in contrast to the radiation from the planar equiangular spiral, the radiation from the conical equiangular spiral is unidirectional.

Figure 6.27 shows the axial ratio (AR) as a function of frequency. The deterioration of the axial ratio observed at low frequencies is due to the reflected current, as mentioned earlier. Figure 6.28 shows the frequency response of the gain relative to a circularly polarized isotropic antenna. It is found that the variation in the gain is small over a broad frequency range.

So far, the analysis has been focused on a conical equiangular spiral antenna whose cone angle is $2\theta_0 = 30^\circ$, as shown in Table 6.2. Further analysis reveals that, as the cone angle is decreased from 30°, the radiation in the negative z direction decreases, as desired. For reference, Figure 6.29 shows the radiation patterns for $2\theta_0 = 20^\circ$ and 10°, where the arm length for these two cases is the same as that for $2\theta_0 = 30^\circ$. Note that these radiation patterns are for 16 GHz.

The conical equiangular spiral arms in this section consist of conducting wires of equal diameter. Figure 6.30 shows a modification of the conical spiral, where conducting strip arms are used instead of wire arms. These two strip arms are fed from a coaxial cable, where the inner conductor of the cable is bonded to one of the arms and the outer conductor is bonded to the other arm. This structure, which does not require a balun circuit for feeding the spiral, is found in the experimental work of Dyson [13, 14].

6.4.4 Related Topic: Archimedean Spiral Antennas

This section presents a spiral antenna whose arms are defined by the Archimedean function, abbreviated as an ARSPL. Actually, the ARSPL can be a self-complementary antenna [3, 35], because the spacing between neighboring antenna arms can be selected

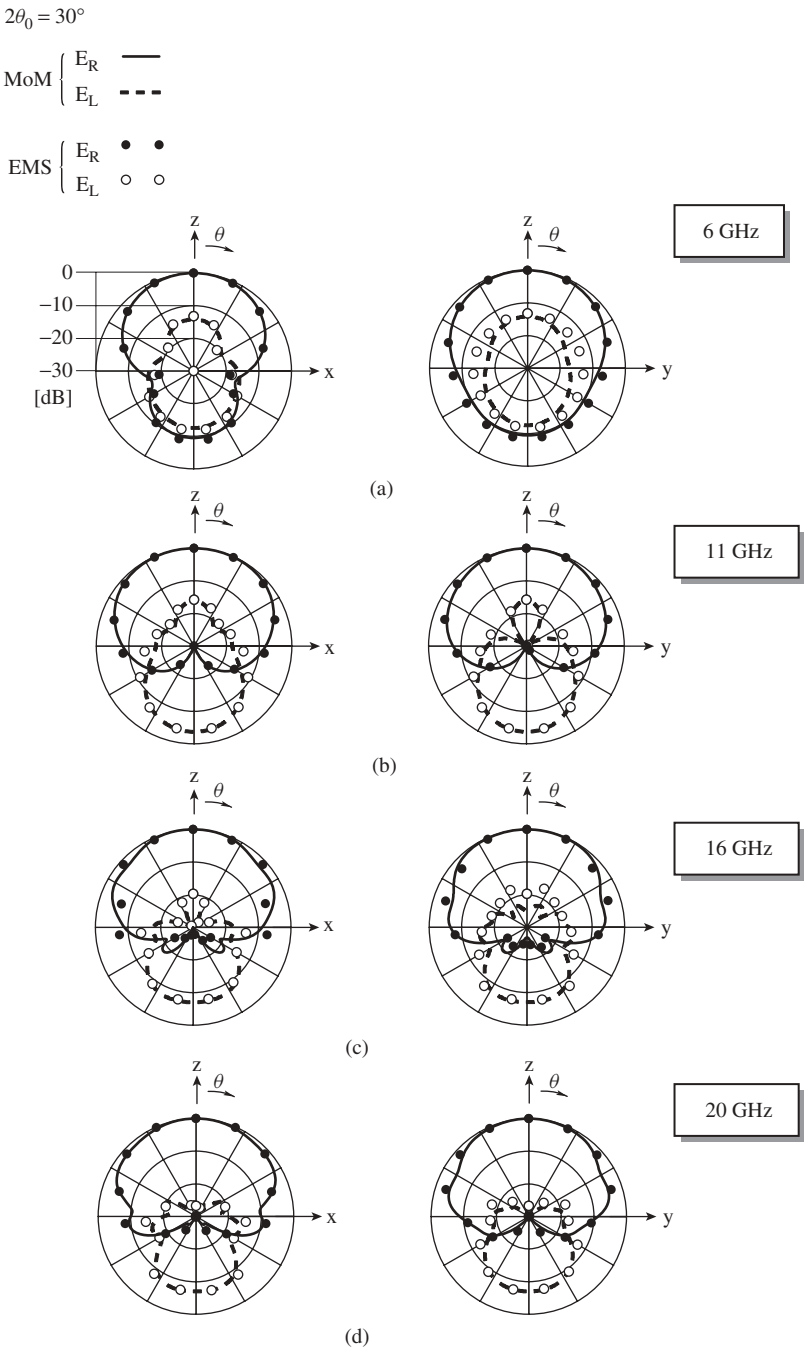


Figure 6.26 Radiation patterns of the conical equiangular spiral antenna (calculated result): (a) 6 GHz, (b) 11 GHz, (c) 16 GHz, and (d) 20 GHz [Ref. 17].

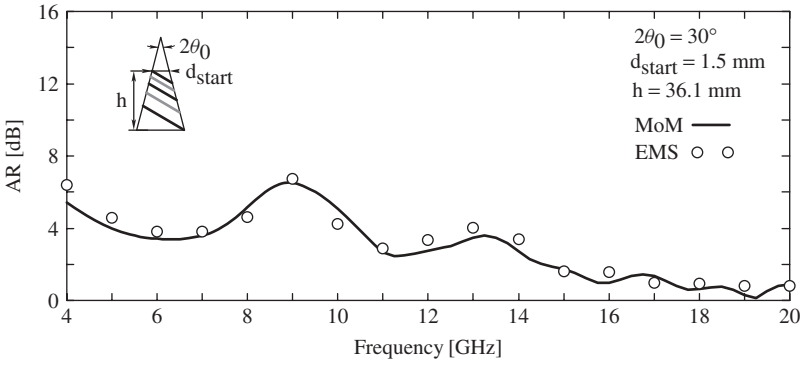


Figure 6.27 Axial ratio of the conical equiangular spiral antenna (calculated result) [Ref. 17].

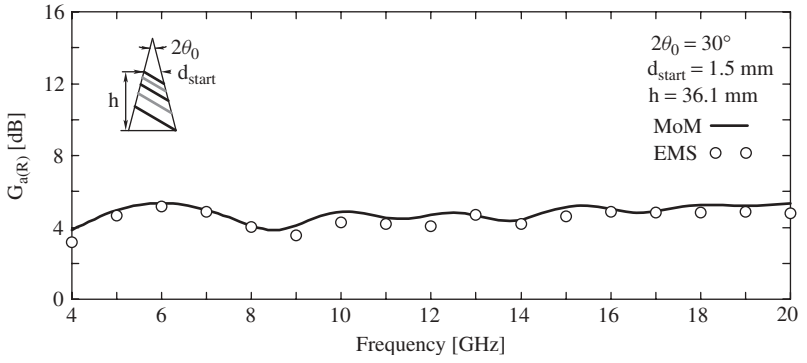


Figure 6.28 Gain of the conical equiangular spiral antenna (calculated result) [Ref. 17].

to be the same as the arm width. The radiation mechanism and antenna characteristics are discussed in Ref. 18. Note that the Archimedean function is not the function of Eq. (6.46) derived by Rumsey for a frequency-independent antenna (see Section 6.2.5).

Figure 6.31 shows a two-arm ARSPL antenna. The two arms are wound symmetrically with respect to the center point (the feed point), where the radial distance from the center point to a point on each arm is defined by the Archimedean function

$$r = a_{\mathrm{AR}}\phi, \qquad \phi_{\mathrm{st}} \le \phi \le \phi_{\mathrm{end}} \tag{6.81a}$$

$$r = a_{\mathrm{AR}}(\phi - \pi), \quad \phi_{\mathrm{st}} + \pi \le \phi \le \phi_{\mathrm{end}} + \pi \tag{6.81b}$$

where a_{AR} is the spiral constant and ϕ is the winding angle.

The two arms, assumed to be infinitely long (winding angle $\phi = \infty$) and tightly wound with a small spiral constant a_{AR}, support currents that travel from the feed point toward the arm ends. When the currents reach a region where the radial distance from the center point is $r \approx \lambda/2\pi$ (λ is the operating wavelength), the currents on neighboring arms of this region (called a ring region) are nearly in phase (see Figure 6.31b). As the currents travel further and reach subsequent regions of $r \approx n\lambda/2\pi$ ($n = 3, 5, \ldots$), in-phase currents

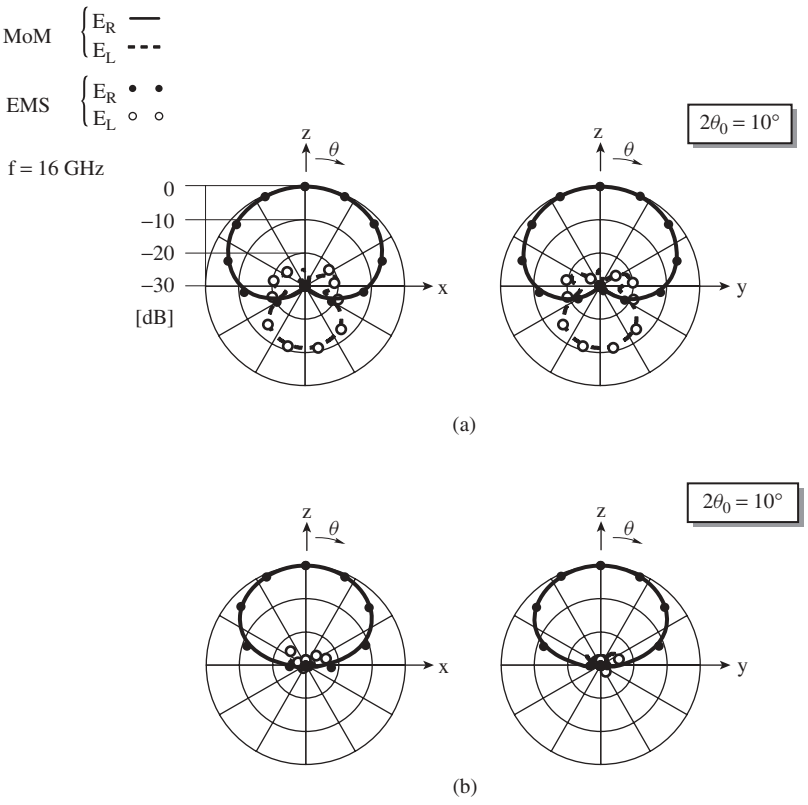


Figure 6.29 Effects of the cone angle on the radiation pattern (calculated result): (a) $2\theta_0 = 20^\circ$ and (b) $2\theta_0 = 10^\circ$ [Ref. 17].

are again obtained at these ring regions. These in-phase currents play a role in forming the radiation pattern.

The currents traveling toward the arm ends gradually decrease, due to radiation into free space. Therefore the radiation from the currents in the first ring region ($n = 1$) is the strongest among all the radiation from the in-phase currents [18]. In other words, the first ring region is the major active region responsible for the radiation pattern. Note that the first ring region has a circumference of approximately one wavelength, and hence the currents traveling (rotating) on this ring region generate a circularly polarized wave in the direction normal to the spiral plane ($\pm z$-directions).

As the frequency decreases, the $n = 1$ active region shifts outwardly in the spiral plane. This means that, if the two arms of the ARSPL are of finite length (which is true in reality), the performance at the low end of the operating frequency band depends on this finite arm length. Conversely, as the frequency increases, the $n = 1$ active region shifts inwardly. It follows that the performance at the high end of the operating frequency band depends on the structure near the feed point.

The ARSPL radiates in the $\pm z$-directions, that is, it radiates a bidirectional beam. The bidirectional beam can be transformed into a unidirectional beam by using the same technique discussed in Section 6.4.2. Figure 6.32 shows the technique using a conducting cavity, where an absorbing strip (ABS) is attached to the vertical wall of the cavity

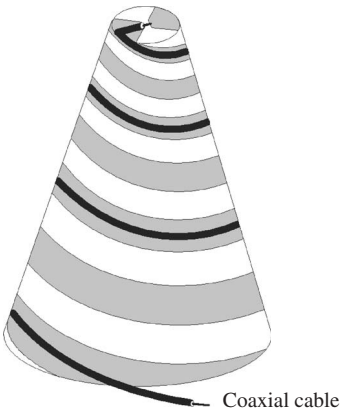


Figure 6.30 Conical equiangular spiral composed of conducting strip arms.

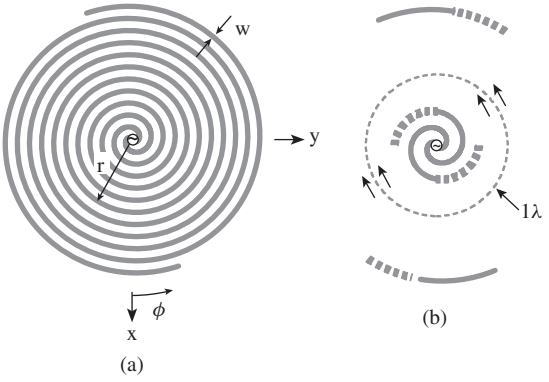


Figure 6.31 Two-arm Archimedean spiral antenna: (a) top view and (b) active region ($n = 1$)

[19, 20]. This antenna is designated as the *ARSPL-abs*. Note that the case where the ABS is removed from the cavity is designated as the *ARSPL-cv*.

The configuration parameters for the ARSPL-abs to be analyzed in this section are shown in Table 6.3, where the spiral arm width w and the spacing between neighboring spiral arms are selected to be the same, so that the structure of the ARSPL is self-complementary.

Figure 6.33 shows the input impedance of the ARSPL-abs, together with that of the ARSPL-cv. A comparison between the input impedances of these two ARSPLs reveals that the ABS in the cavity contributes to making the variation in the input impedance small (note the low frequency region), as observed in the equiangular spiral antenna of Section 6.4.2. The other antenna characteristics of the ARSPL-abs, including the axial ratio, radiation pattern, and gain, are shown in Figures 6.34–6.36, together with those of the ARSPL-cv. These figures reveal that the ARSPL-abs acts as a circularly polarized antenna over a broad frequency band, without large variation in the input impedance and gain, as desired.

The above discussion focuses on an ARSPL composed of two arms. Figure 6.37 shows a modified ARSPL antenna. The ARSPL in Figure 6.37a is composed of a single

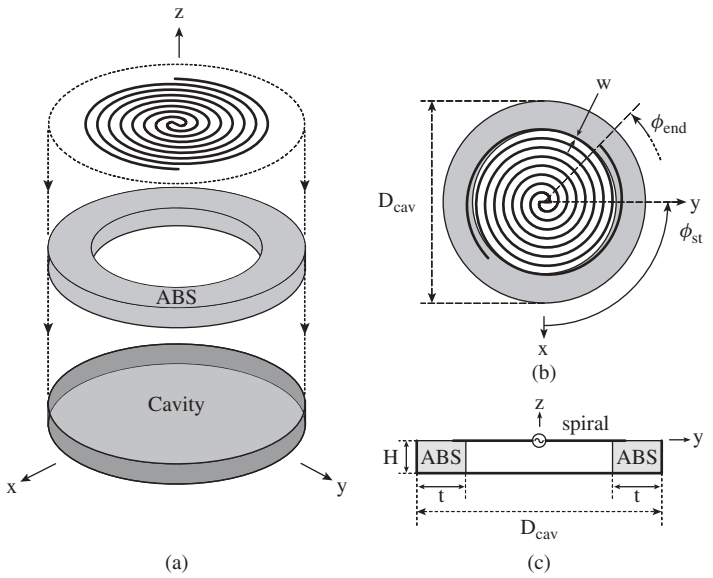


Figure 6.32 Archimedean spiral antenna backed by a cavity: (a) exploded view, (b) top view, and (c) side view.

TABLE 6.3 Configuration Parameters for an Archimedean Spiral Antenna Backed by a Cavity

Symbol	Value	Unit
a_{AR}	1.273	mm/rad
ϕ_{st}	0.5π	rad
ϕ_{end}	8.5π	rad
D_{cav}	94	mm
H	7	mm
w	2	mm

arm, obtained by removing one arm from the two-arm ARSPL shown in Figure 6.31. The single-arm antennas shown in Figures 6.37b–d, which have a small number of spiral turns, are often called *curl antennas*, with the maximum radial distance from the center point to the arm end set to be less than λ/π. Note that the ring-like structure in Figure 6.37d is obtained by setting the spiral constant a_{AR} to be extremely small. These single-arm ARSPL antennas have the advantage that they are fed from a coaxial line without a balun circuit (required for feeding two-arm spirals). Detailed investigations of these single-arm ARSPLs are found in Refs. 22–24.

6.5 LOGPERIODICS

Two antennas with log-periodic shape are presented: a toothed log-periodic antenna and a trapezoidal log-periodic antenna. In addition, a log-periodic dipole array antenna

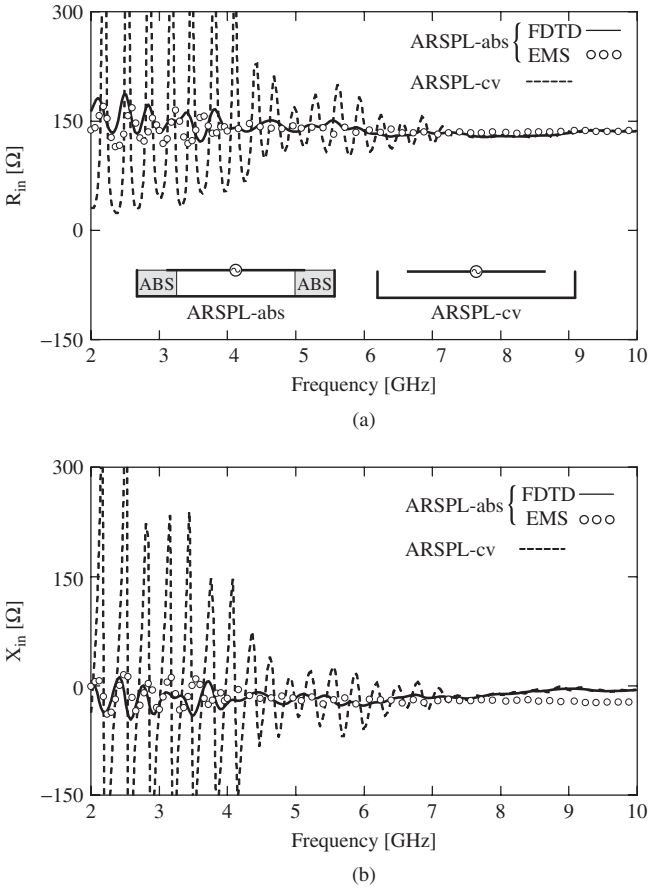


Figure 6.33 Input impedance of the Archimedean spiral antenna backed by a cavity with and without an absorbing strip (calculated result). (Courtesy of R. Kobayashi [Ref. 21].)

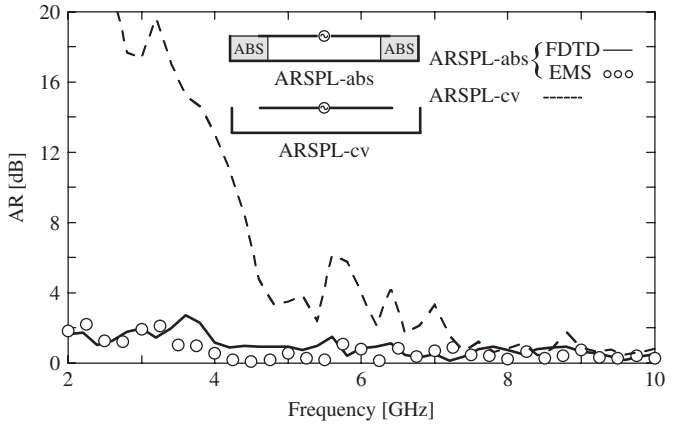


Figure 6.34 Axial ratio of the Archimedean spiral antenna backed by a cavity with and without an absorbing strip (calculated result) [Ref. 21].

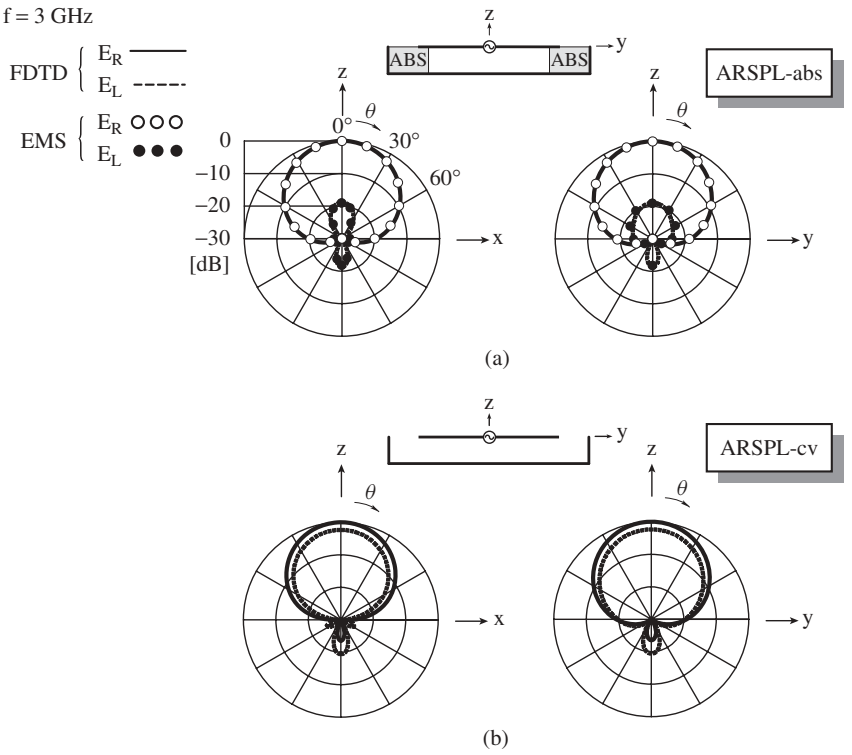


Figure 6.35 Radiation pattern at 3 GHz of the Archimedean spiral antenna backed by a cavity (calculated result): (a) With an absorbing strip and (b) without an absorbing strip [Ref. 21].

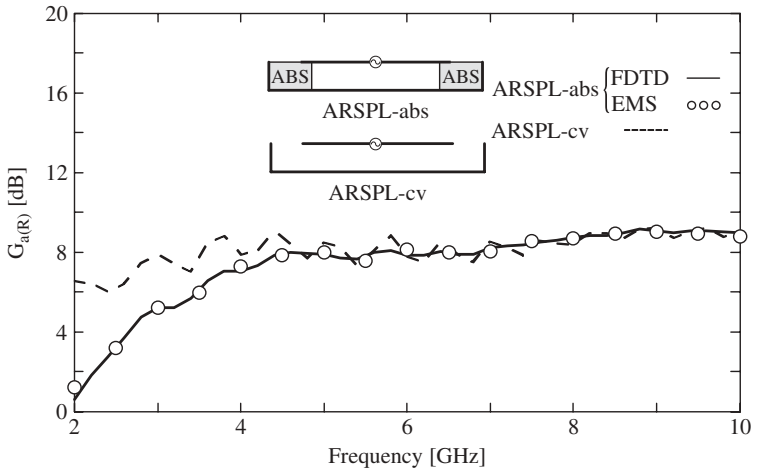


Figure 6.36 Gain of the Archimedean spiral antenna backed by a cavity with and without an absorbing strip (calculated result) [Ref. 21].

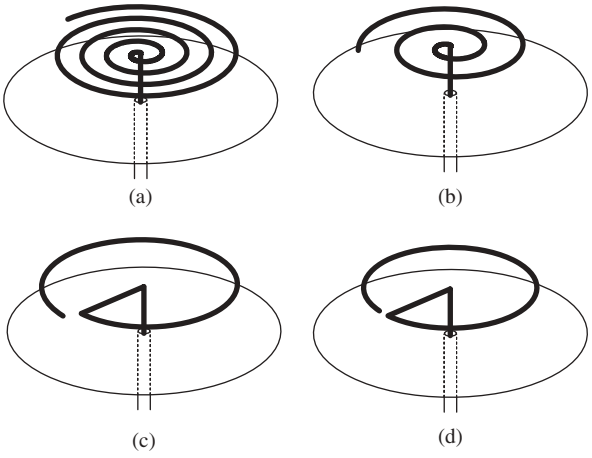


Figure 6.37 Modified Archimedean spiral antennas. The antennas shown in (b), (c), and (d) are called *curl* antennas.

(a derivative of the log-periodic self-complementary antenna) is discussed. These antenna structures are finite and hence the antenna characteristics are affected by their finite structures. Note that the radiation from each of these antennas is linearly polarized.

6.5.1 Toothed Log-Periodic Antenna

Figure 6.38a shows a self-complementary antenna, whose two arms (of infinite length) are point symmetric with respect to the coordinate origin [3, 35]. This antenna is called a *toothed log-periodic antenna* (T-LPA). The antenna shown in Figure 6.38b is a truncated version of this antenna. The truncation effects must be reduced to maintain the self-complementary characteristics as much as possible. Note that the T-LPA design was first proposed by DuHamel and Isbel and their experimental work is reported in Refs. 25 and 26.

Figure 6.39 shows notations for a T-LPA. The antenna is specified with expansion parameter ξ, tooth-width parameter τ, inner angle Φ_{in}, and outer angle Φ_{out}, where

$$\xi = \frac{R_{n+1}}{R_n} \tag{6.82}$$

$$\tau = \frac{r_n}{R_n} \tag{6.83}$$

The configuration parameters to be used in this section are shown in Table 6.4, where the maximum number of teeth and the sum of angles Φ_{in} and Φ_{out} are set to be $N = 6$ and $\Phi_{\text{in}} + \Phi_{\text{out}} = 90^\circ$, respectively. The analysis is performed using the FDTDM, for which the antenna structure is approximated using staircases. A detailed view of the staircase approximation around the feed point is shown in Figure 6.39b.

Figure 6.40 shows the input impedance as a function of frequency. The resistive component of the input impedance at each of frequencies f_1, f_2, f_3, and f_4, marked by arrows, shows a local maximum and the period made by the logarithms of these

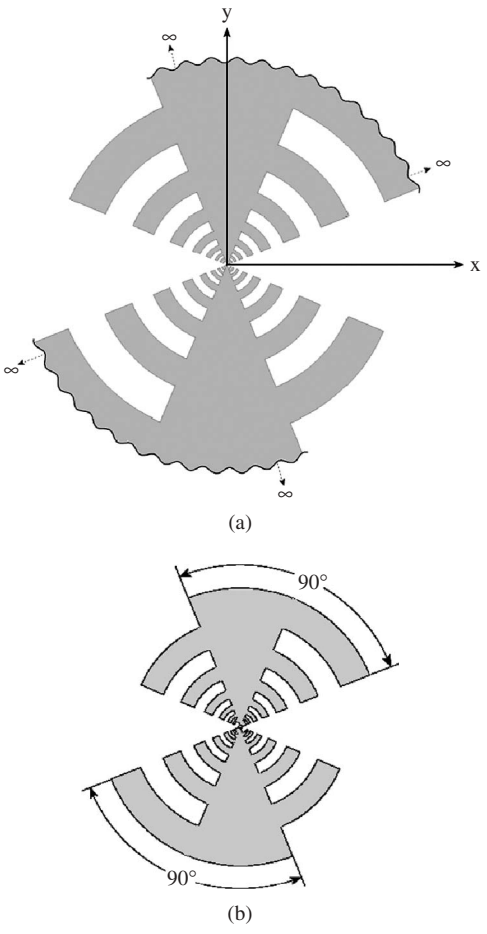


Figure 6.38 A toothed log-periodic antenna: (a) arms of infinite length and (b) arms of finite length.

frequencies ($\log f_{n+1} - \log f_n$ for $n = 1, 2, 3$) approximates $\log \xi$ ($= 0.222$). Note that $L_5/2$ is approximately 0.25 wavelength at 518 MHz, below which the input impedance varies significantly due to the finite arm length.

Figure 6.41 shows the electric current density $\mathbf{J}_s$ (amplitude $|\mathbf{J}_s|$) at the frequencies marked in Figure 6.40 (f_1, f_2, f_3, and f_4). As the frequency increases, the active region moves toward the feed point region. This behavior of the active region with frequency is similar to that observed in the equiangular spiral antenna discussed in Section 6.4.1 and the conical equiangular spiral antenna discussed in Section 6.4.3.

The T-LPA radiates a bidirectional beam. Figure 6.42 shows the radiation pattern as a function of frequency. The polarization is linear. The copolarization component is generated from the currents flowing on the teeth in the azimuthal direction. The cross-polarization component is due to the currents in the radial direction. From the fact that the radiation has similar patterns over much of the analysis range, it is expected that the variation in the gain of the T-LPA will be small. This is confirmed in Figure 6.43. A gain of 5.3 ± 1.1 dBi is obtained over a frequency range of 0.6–3.0 GHz.

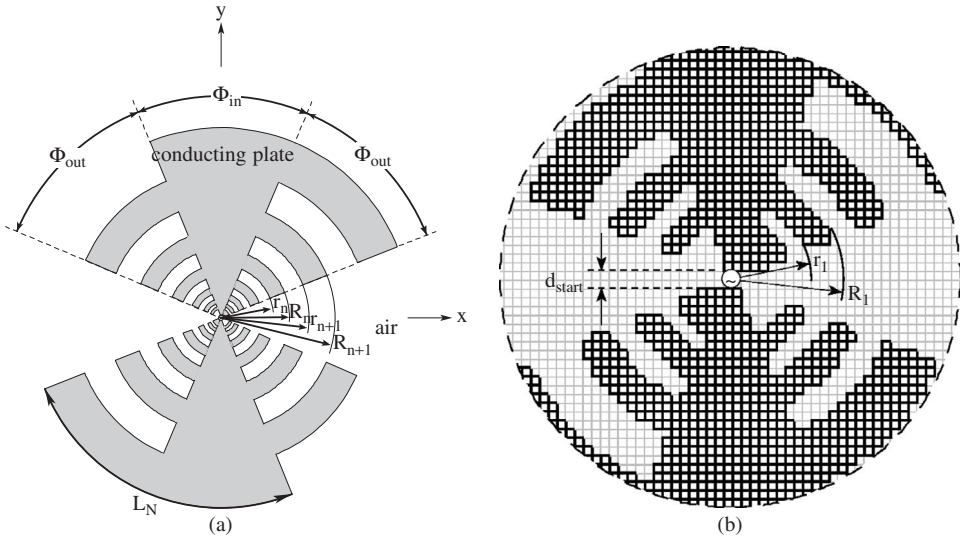


Figure 6.39 Toothed log-periodic antenna: (a) top view and (b) staircase approximation around the feed point.

TABLE 6.4 Configuration Parameters for a Toothed Log-Periodic Antenna

Symbol	Value	Unit
Φ_{in}	45	degrees
Φ_{out}	45	degrees
$1/\xi$	0.6	—
$\tau = 1/\sqrt{\xi}$	0.7746	—
R_6	300	mm
R_1	23.3	mm
L_6	471.2	mm
d_{start}	4	mm
N	6	—

As shown in Figure 6.42, the radiation is bidirectional. Figure 6.44 shows a technique for obtaining a unidirectional beam (endfire beam), where arm B is folded toward arm A. Figure 6.45 shows the radiation pattern when the angle between the two arms is $\theta_{FLD} = \pi/2$. It is found that the radiation is unidirectional in the direction of $\theta = \pi + \theta_{FLD}/2$.

6.5.2 Trapezoidal Log-Periodic Antenna

This section discusses T-LPAs (toothed log-periodic antennas) with modified structures. Figure 6.46 shows one modified structure, where the curved teeth of the T-LPA are transformed into linear teeth [26]. This antenna is designated as a *trapezoidal log-periodic antenna* and abbreviated as a *Trpz-LPA*.

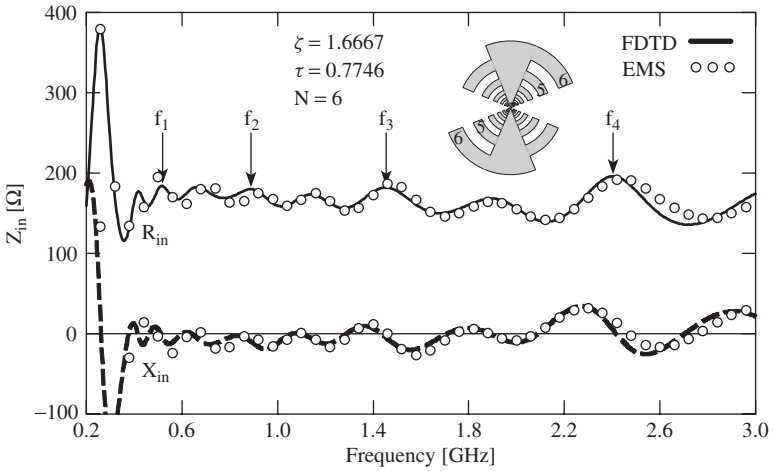


Figure 6.40 Input impedance of the toothed log-periodic antenna (calculated result). (Courtesy of H. Iwaoka [Ref. 27].)

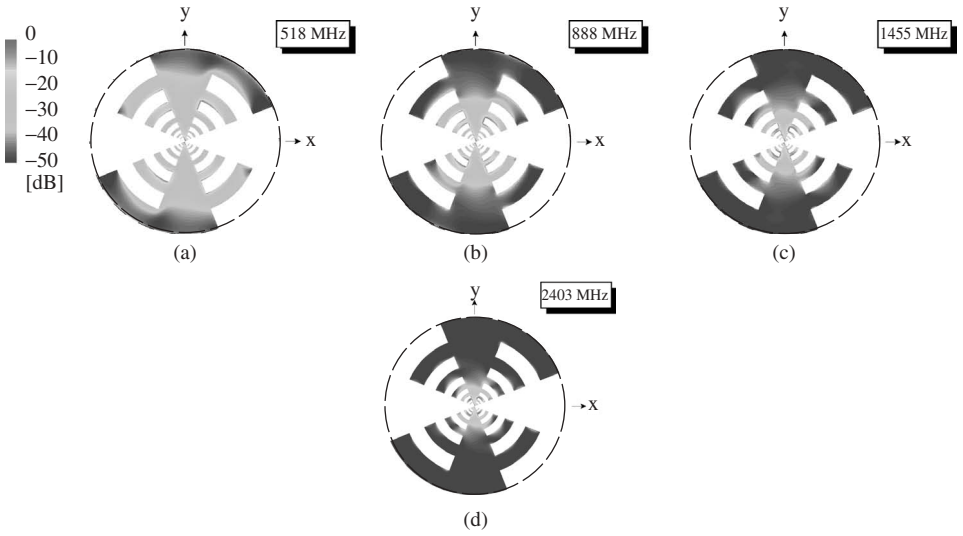


Figure 6.41 Electric current density $|\mathbf{J}_s|$ on the toothed log-periodic antenna (calculated result): (a) 518 MHz, (b) 888 MHz, (c) 1455 MHz, and (d) 2403 MHz [Ref. 27].

The configuration parameters to be used in this section are summarized in Table 6.5. For comparison, the outermost length L_N for the Trpz-LPA to be used here is chosen to be the same as that for the T-LPA whose configuration parameters are shown in Table 6.4. The analysis is performed using the FDTDM, where a staircase approximation is used for the antenna structure, as shown in Figure 6.46b.

Figures 6.47–6.50 show the antenna characteristics of the Trpz-LPA, including the input impedance, radiation pattern, current distribution, and gain. The main findings

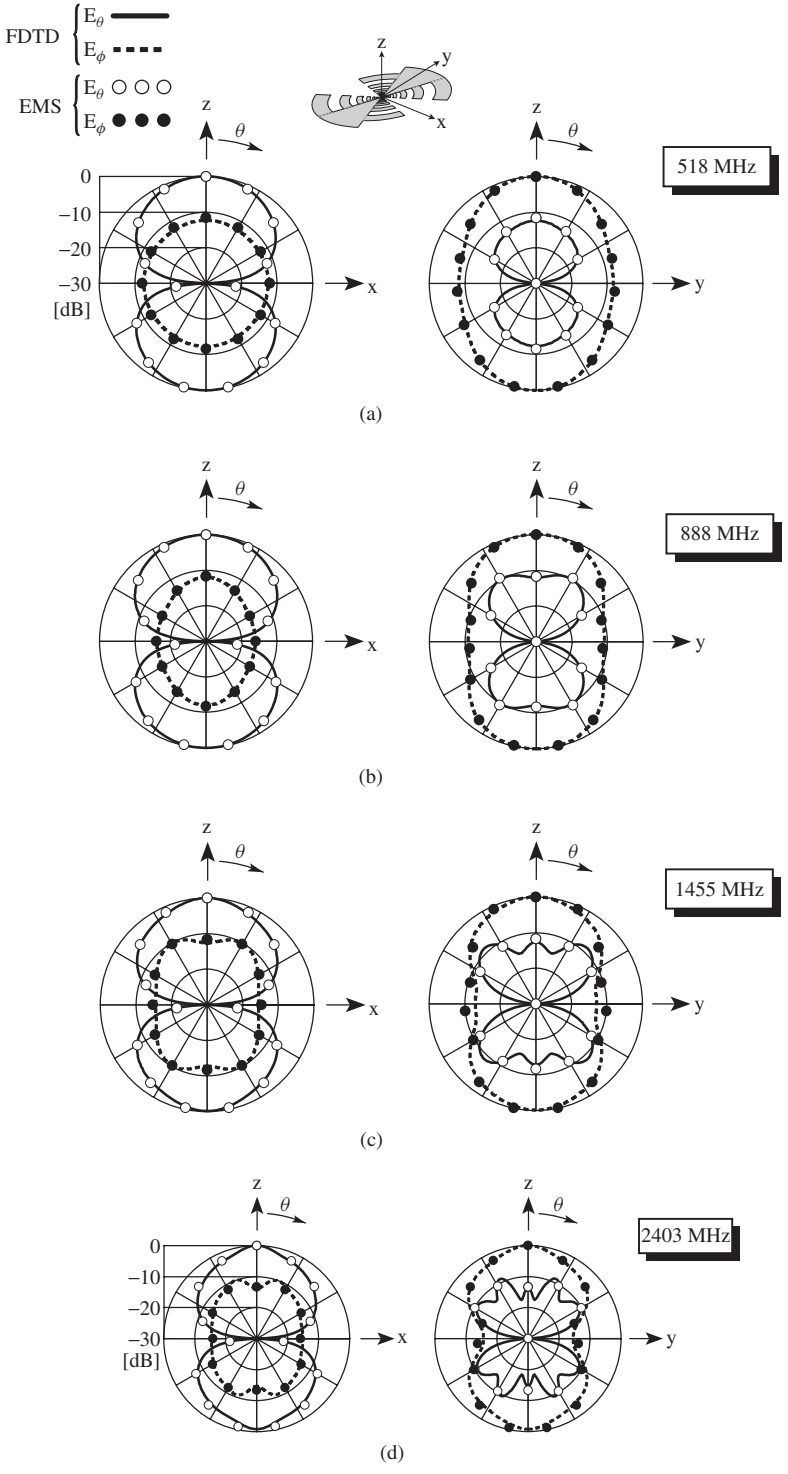


Figure 6.42 Radiation patterns of the toothed log-periodic antenna (calculated result): (a) 518 MHz, (b) 888 MHz, (c) 1455 MHz, and (d) 2403 MHz [Ref. 27].

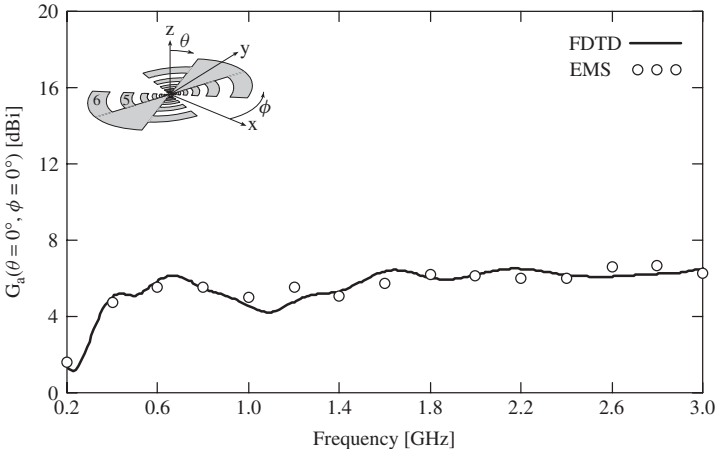


Figure 6.43 Gain of the toothed log-periodic antenna (calculated result) [Ref. 27].

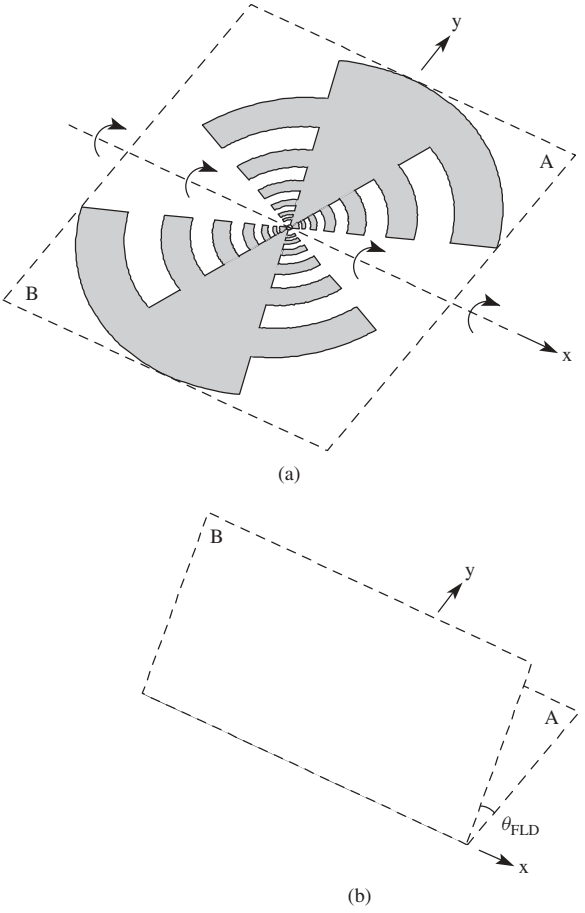


Figure 6.44 Folded toothed log-periodic antenna for obtaining a unidirectional beam: (a) planar structure and (b) folded structure.

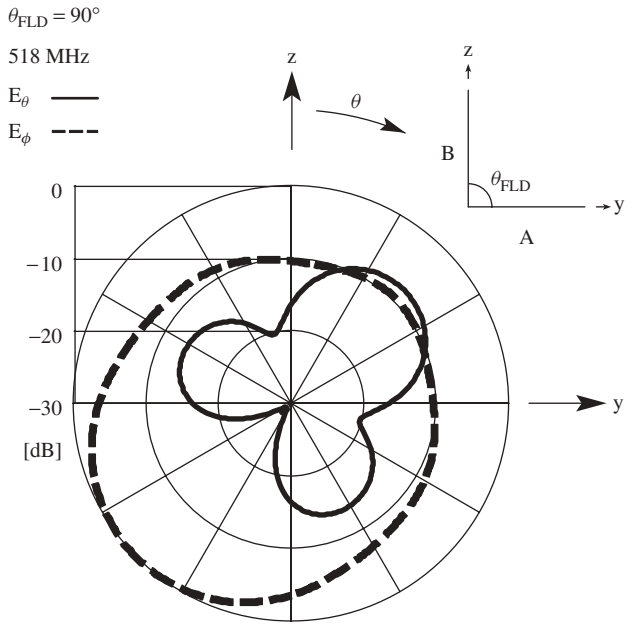


Figure 6.45 Radiation pattern of a folded toothed log-periodic antenna with $\theta_{FLD} = \pi/2$ (calculated result) [Ref. 27].

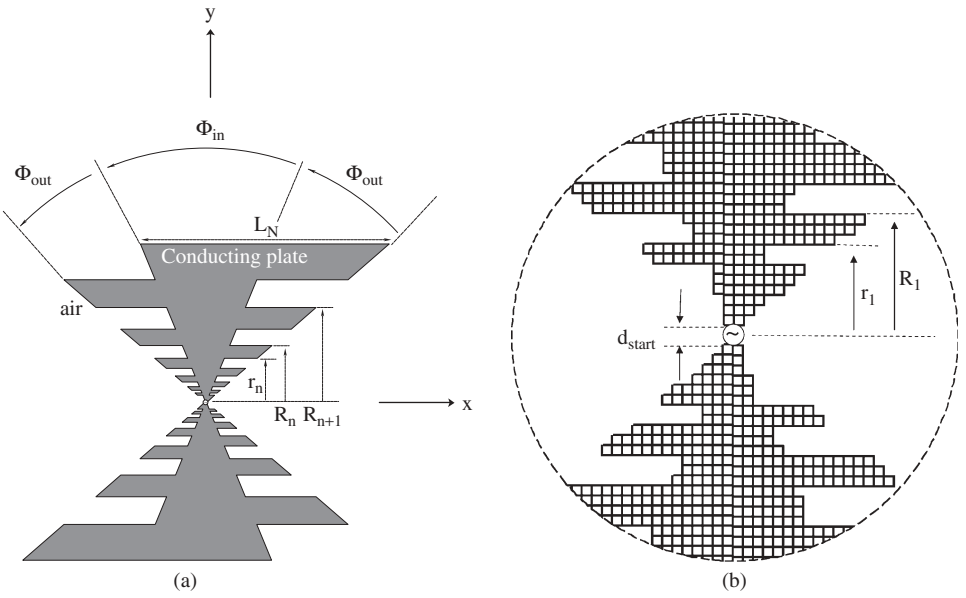


Figure 6.46 Trapezoidal log-periodic antenna: (a) top view and (b) magnified view of the staircase approximation for the feed region.

TABLE 6.5 Configuration Parameters for a Trapezoidal Log-Periodic Antenna

Symbol	Value	Unit
Φ_{in}	45	degrees
Φ_{out}	26.65	degrees
ξ	1/0.6	—
$\tau = 1/\sqrt{\xi}$	0.7746	—
N	6	—
L_N	471.2	mm
R_N	300	mm
R_1	23.32	mm
d_{start}	4	mm

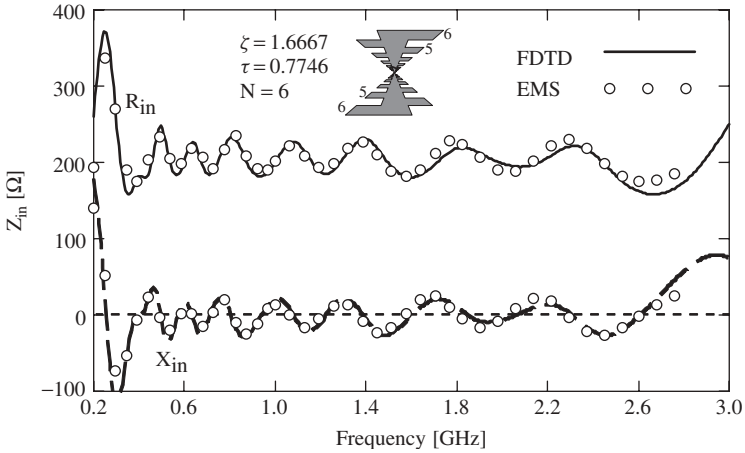


Figure 6.47 Input impedance of the trapezoidal log-periodic antenna (calculated result). (Courtesy of K. Morishita [Ref. 28].)

are as follows: (1) the Trpz-LPA has slightly larger variation in the input impedance (observe the variation in the low frequency region) than the T-LPA; (2) the Trpz-LPA has a slightly smaller gain than the T-LPA; (3) the Trpz-LPA has a cross-polarization component, similar to that for the T-LPA; and (4) the behavior of the current distribution for the Trpz-LPA as a function of frequency is similar to that of the T-LPA.

Figure 6.51 shows further modification of the T-LPA structure, where the two arms (the upper and lower sections) of each structure are point symmetric with respect to the arm center point (if not so, broad frequency characteristics are not obtained [3, 35, 36]). It can be said that the structures in Figure 6.51e,f are physically suitable for use at low frequencies, where the antenna is large, because the antenna is light and the wind resistance is small relative to the structure in Figure 6.51a. The antenna in Figure 6.51f is discussed in the following section.

6.5.3 Log-Periodic Dipole Array Antenna

A dipole is a radiation element that has a narrow input impedance characteristic. However, if dipoles are arrayed as shown in Figure 6.52, this array has a relatively constant input impedance over a broad frequency range. Note that this array, called a *log-periodic dipole array (LPDA)*, was first reported by D. Isbell in 1960 [29].

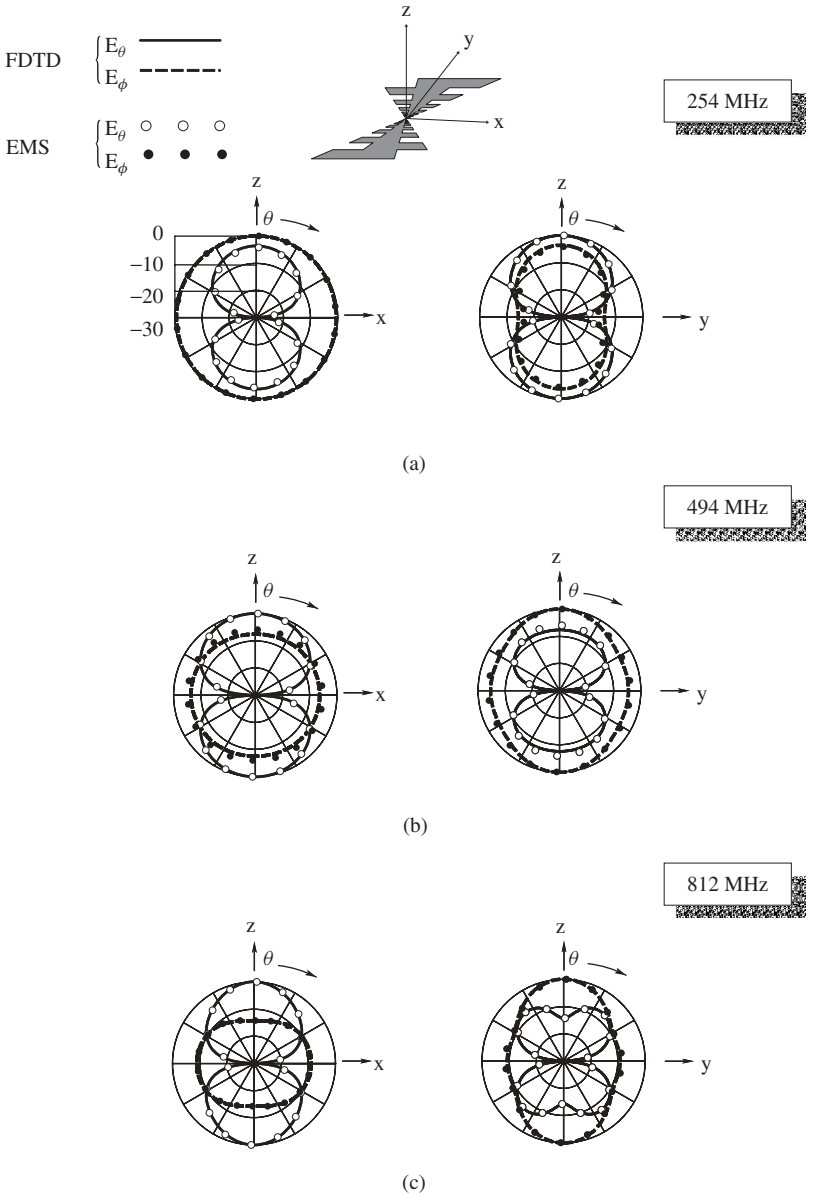
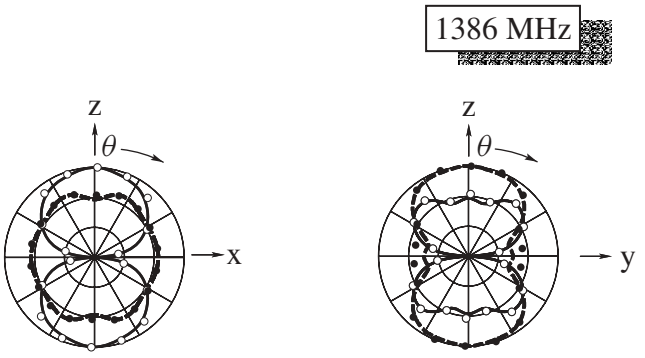
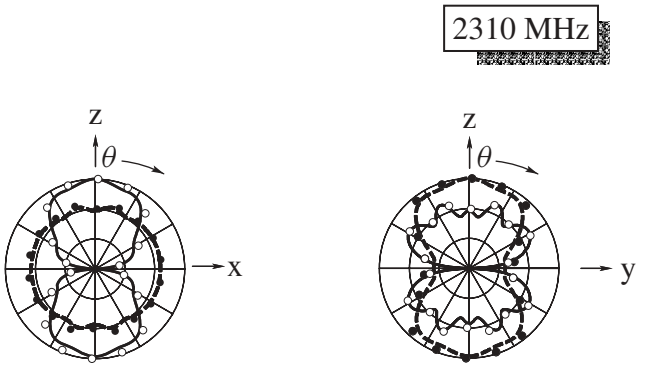


Figure 6.48 Radiation patterns of the trapezoidal log-periodic antenna (calculated result): (a) 254 MHz, (b) 494 MHz, (c) 812 MHz, (d) 1386 MHz, and (e) 2310 MHz [Ref. 28].



(d)



(e)

Figure 6.48 (*Continued*)

The nth dipole of the LPDA composed of N dipoles is specified by location R_n, length ℓ_n, radius ρ_n, and feed spacing Δ_n.

We define the neighboring dipoles by a constant K_1

$$\frac{R_{n+1}}{R_n} = \frac{\ell_{n+1}}{\ell_n} = \frac{\rho_{n+1}}{\rho_n} = \frac{\Delta_{n+1}}{\Delta_n} = K_1, \quad n = 1, 2, \ldots, N-1 \tag{6.84}$$

In addition, we define the relative spacing by a constant K_2

$$(R_{n+1} - R_n)/2\ell_{n+1} = K_2, \quad n = 1, 2, \ldots, N-1 \tag{6.85}$$

The dipoles in the LPDA in Figure 6.52 are fed by crossed transmission lines [3, 35, 36]. However, for practicality (avoiding a complicated feed system), this feed system is replaced with the transmission lines shown in Figure 6.53, where a coaxial line and a solid conductor play the role of the crossed transmission lines. The inner conductor of the coaxial line is connected to an upper solid conductor (whose diameter is the same as that of the outer conductor of the coaxial line). Half of the dipoles are connected to

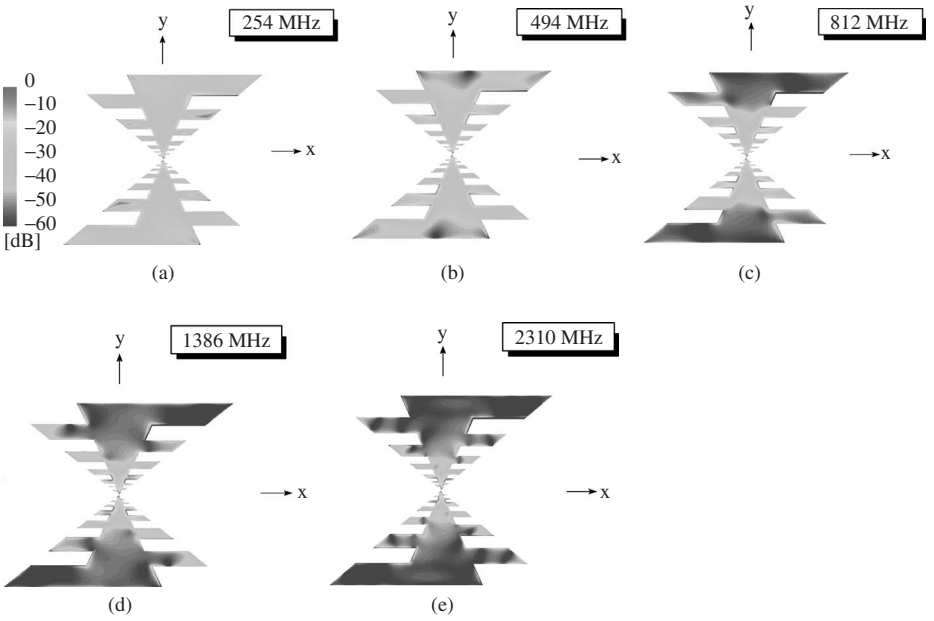


Figure 6.49 Electric current density $|\mathbf{J}_x|$ on the trapezoidal log-periodic antenna (calculated result): (a) 254 MHz, (b) 494 MHz, (c) 812 MHz, (d) 1386 MHz, and (e) 2310 MHz [Ref. 28].

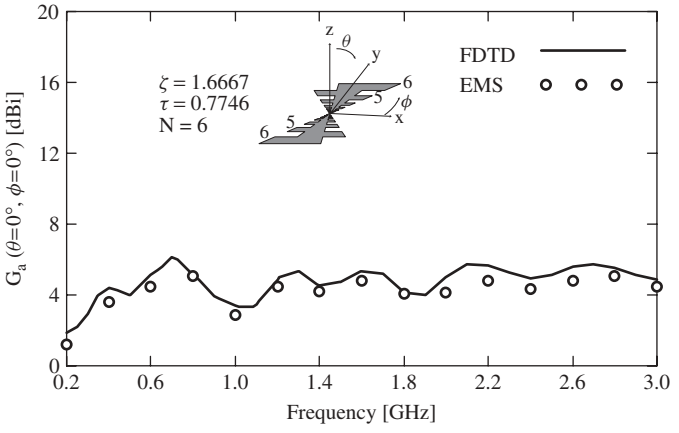


Figure 6.50 Gain of the trapezoidal log-periodic antenna (calculated result) [Ref. 28].

the solid conductor and the remaining half are connected to the outer conductor of the coaxial line.

The LPDA in Figure 6.53 can be handled using the equivalent structure shown in Figure 6.54. (This corresponds to the structure of Figure 6.51f, where arm A is folded onto arm B around the feed point.) In this section, this equivalent structure is analyzed using the MoM. The configuration parameters for the MoM analysis are shown in Table 6.6,

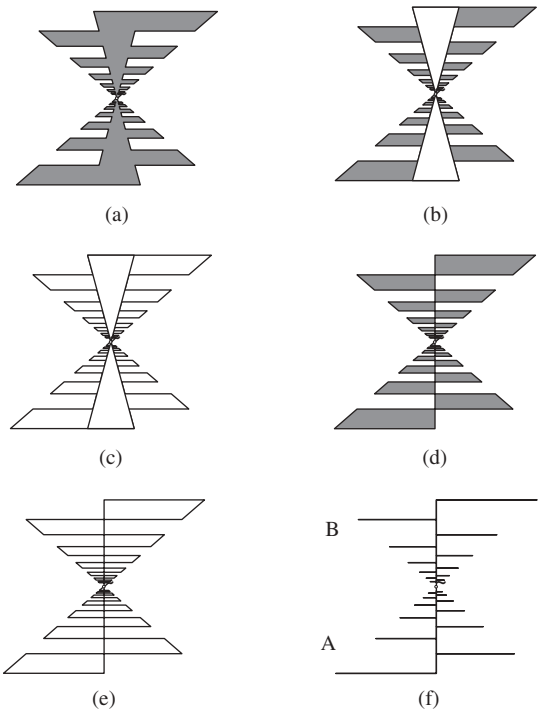


Figure 6.51 Further modification of the toothed log-periodic antenna.

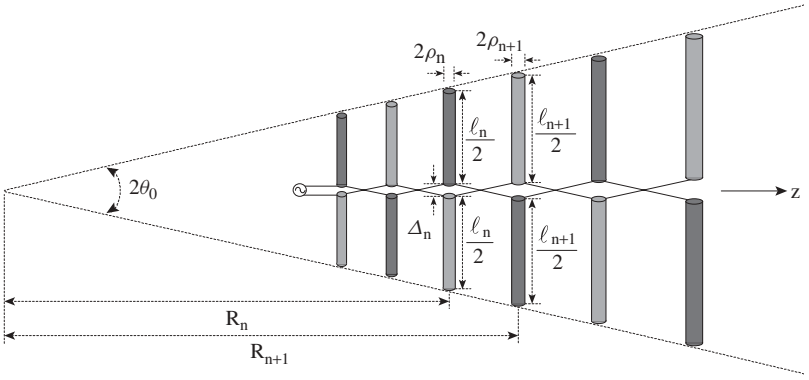


Figure 6.52 Log-periodic dipole array.

where the dipole radii ρ_n ($n = 1, 2, \ldots, N$) are given a small constant value of ρ (which does not satisfy Eq. (6.84)) and the feed spacing of each dipole Δ_n is chosen to be the same (zero) (which also does not satisfy Eq. 6.84)). Note that these unsatisfactory ρ_n and Δ_n do not remarkably deteriorate the antenna characteristics, as will be seen later.

The input terminals of the LPDA, where a voltage source is attached, are located near the shortest dipole, as shown in Figure 6.54. Figure 6.55 shows the input impedance of the LPDA as a function of frequency. It is emphasized that the input impedance is

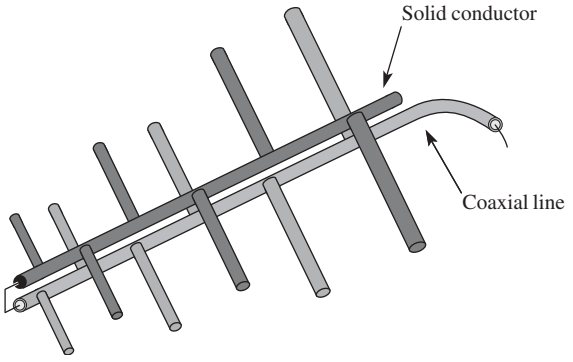


Figure 6.53 Practical feed system using a coaxial line and a solid conductor.

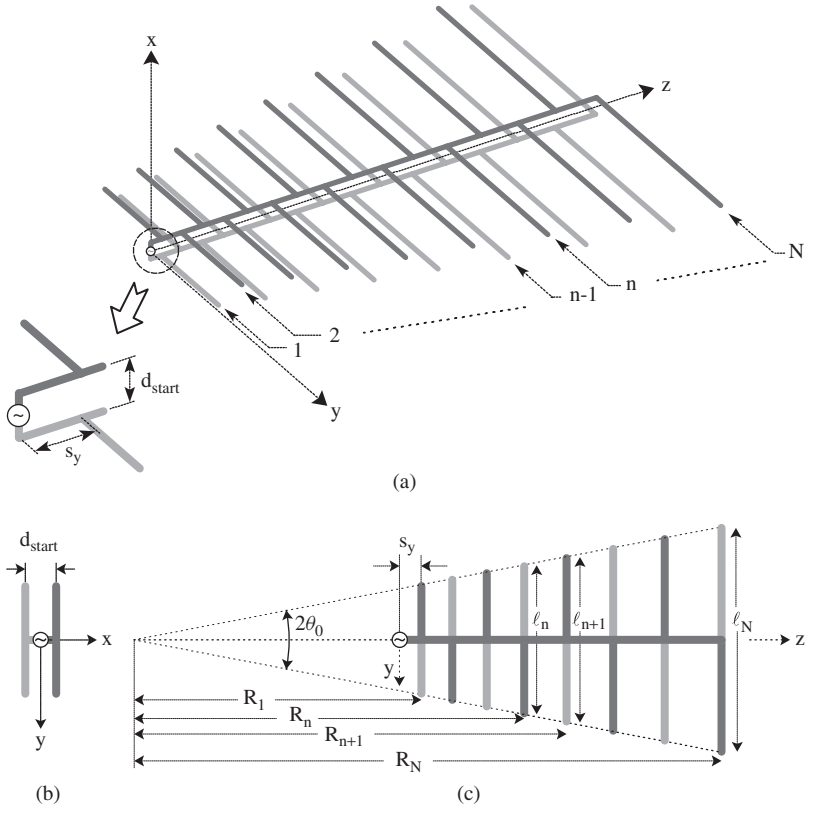


Figure 6.54 Equivalent structure for the structure shown in Figure 6.53: (a) perspective view, (b) side view of the feed region, and (c) top view.

calculated taking into account all the mutual effects, namely, effects between the dipoles and effects between the dipoles and the feed lines. It is found that there exists a broad operating frequency band where the input impedance shows relatively small variation. The low end of the operating frequency band depends on the longest dipole length, and the high end depends on the shortest dipole length.

TABLE 6.6 Configuration Parameters for the Equivalent Structure

Symbol	Value	Unit
$2\theta_0$	21.85	degrees
$2\rho_n = 2\rho$	1.6	mm
Δ_n	0	mm
s_y	6	mm
K_1	1/0.95	—
N	16	—
ℓ_N	583	mm
R_N	1511	mm
R_1	700	mm
d_{start}	4	mm

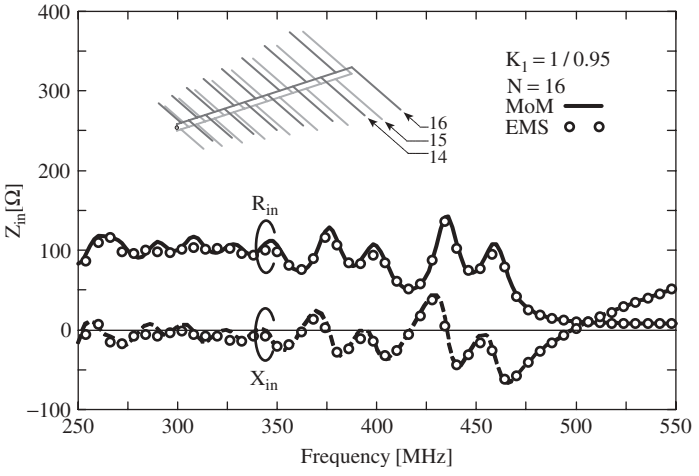


Figure 6.55 Input impedance of the log-periodic dipole array (calculated result). (Courtesy of H. Honma [Ref. 30].)

Figure 6.56 reveals the behavior of the current when the frequency is varied. The ordinate shows the amplitude of the current $|I|$ (where $I = I_r + jI_i$) observed at the input terminals of each dipole and the abscissa shows the element number. As the frequency is increased, the active region (where large currents flow along some dipoles relative to the currents along the remaining dipoles) moves to the left (negative z-direction). Note that the dipoles in the active region are approximately one-half wavelength long, as Carrel describes [31].

The LPDA has a unidirectional radiation beam, as is found in the case of the folded toothed log-periodic antenna shown in Figure 6.45. Figure 6.57 shows the radiation pattern when the frequency is varied. It is found that the radiation pattern remains relatively unchanged within the region where the variation in the input impedance is small. This makes the variation in the gain small, as shown by a solid line in Figure 6.58. The gain shows 7.7 ± 1.5 dBi within the frequency range of 250–550 MHz.

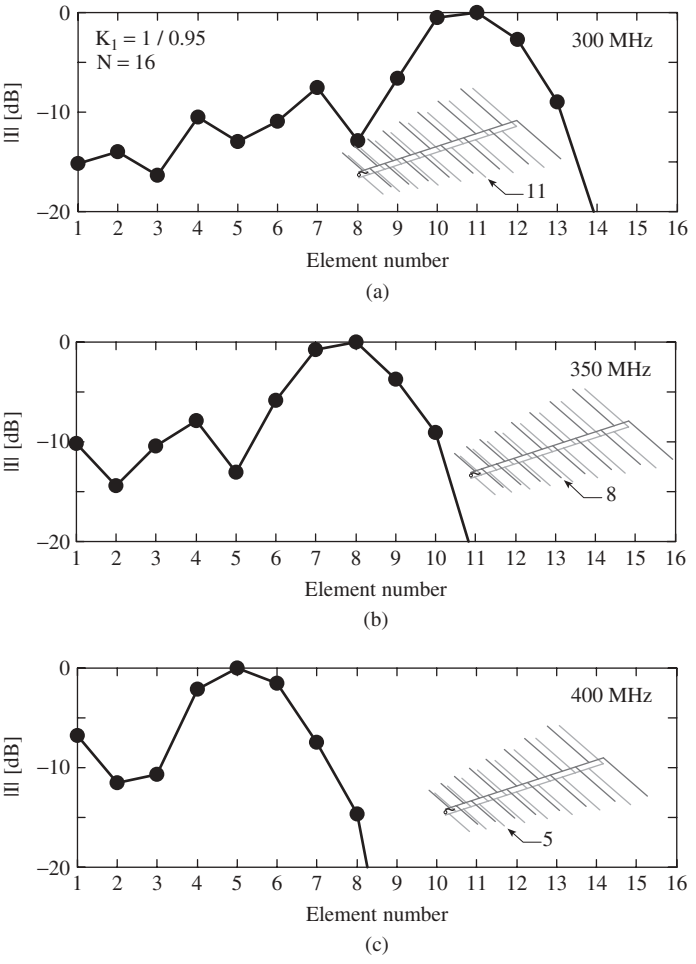


Figure 6.56 Normalized amplitude of the currents on the dipoles (calculated result): (a) 300 MHz, (b) 350 MHz, and (c) 400 MHz [Ref. 30].

So far, the analysis has been performed holding the number of dipoles constant at $N = 16$. To show the effect of changing the number of dipoles on the gain, three and six dipoles are removed from the original structure ($N = 16$). The circles and crosses in Figure 6.58 show the gains for $N = 13$ (dipoles 14 through 16 are removed) and $N = 10$ (dipoles 11 through 16 are removed), respectively. It is found that the removal of these long dipoles affects the gain at low frequencies, while having almost no effect on the gain at high frequencies.

The above LPDA is analyzed for a frequency range of 250–550 MHz (1:2.2). Note that a design example for an LPDA antenna that operates from 54 to 216 MHz (1:4) is given in Refs. 32 and 37. This design is summarized as follows. (1) Using Figure 11 of Ref. 37, the scale factor τ $(= 1/K_1)$ and relative spacing σ $(= K_2)$ are determined for a required gain. (2) From these τ and σ, the end angle $2\theta_0$ is obtained as $2\theta_0 = 2\tan^{-1}[(1-\tau)/4\sigma]$. (3) The active region bandwidth B_{ar} is calculated using a semiempirical

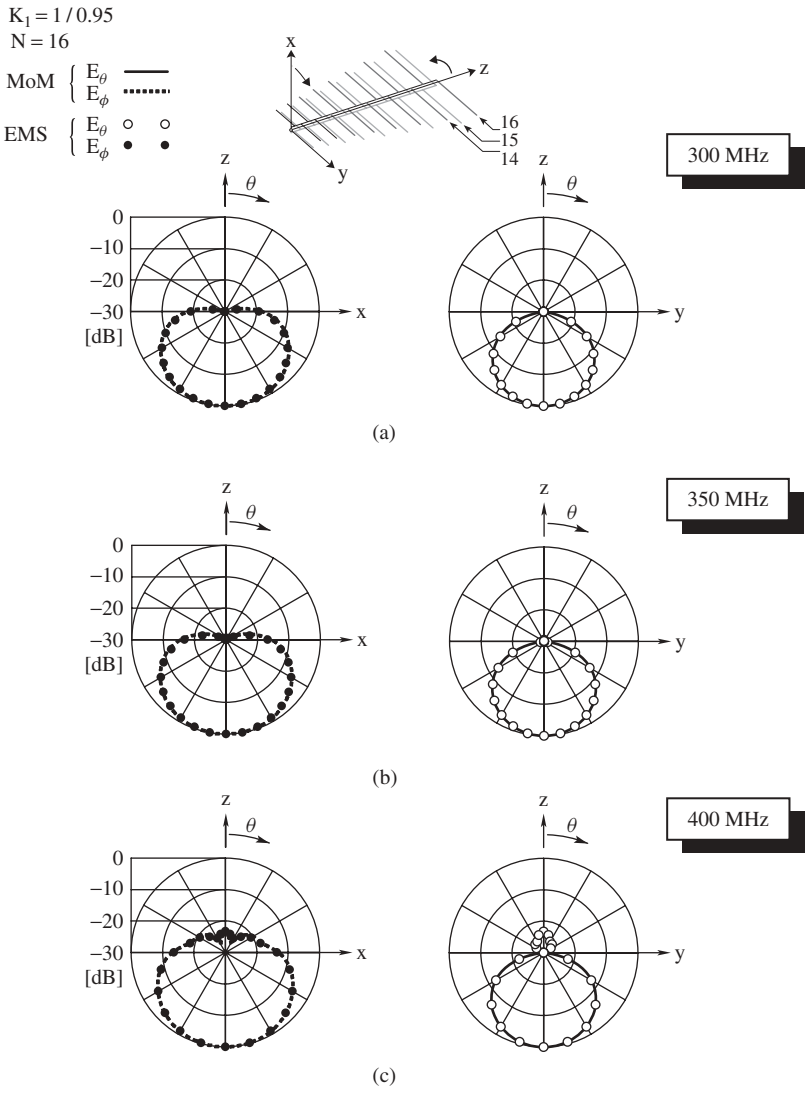


Figure 6.57 Radiation patterns of the log-periodic dipole array (calculated result): (a) 300 MHz, (b) 350 MHz, and (c) 400 MHz [Ref. 30].

equation $B_{ar} = 1.1 + 7.7(1-\tau)^2 \cot \theta_0$ and the bandwidth of the structure B_s (designed bandwidth) is calculated using $B_s = BB_{ar}$, where B is the required operating bandwidth ($B = 216$ MHz/54 MHz). (4) The number of dipole elements is determined on the basis of $N = 1 + [\ln (B_s)/\ln (K_1)] = 1 + [\ln(B_s)/\ln (1/\tau)]$. (5)The longest dipole length is chosen to be $\ell_N = \lambda_{max}/2$, where λ_{max} is the wavelength at the lowest frequency. Then, $\ell_n/2\rho_n$ (the same for all dipole elements) is calculated for a given dipole diameter $2\rho_n$. (6) An average characteristic impedance of the dipole elements, Z_a, is approximated as $Z_a = 120\ [\ln (\ell_n/2\rho_n) - 2.25]$, which determines Z_a/R_0, where R_0 is chosen to be the input impedance value (of the feed line) we want, for example, $R_0 = 50$ ohms. (7) Using

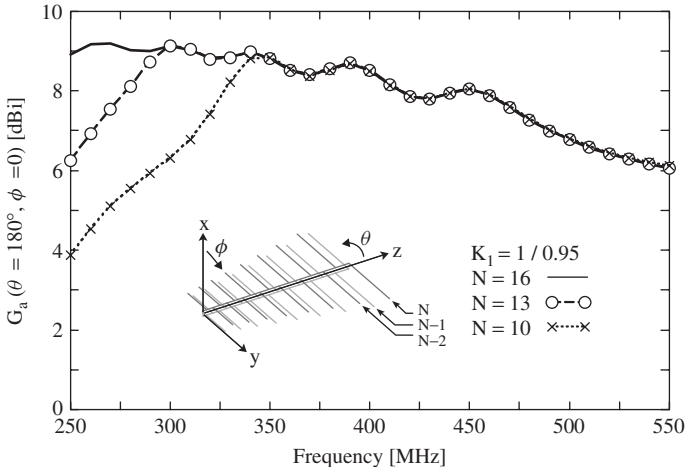


Figure 6.58 Gain of the log-periodic dipole array as a function of frequency (calculated result) [Ref. 30].

Figure 16 of Ref. 37, which is a chart for (Z_0/R_0) versus (Z_a/R_0) for $\sigma' = \sigma/\sqrt{\tau}$, the normalized characteristic impedance of the feed line Z_0/R_0 is determined. (8) The spacing of the feed line made of two rods, s, is calculated as $s = d \cosh (Z_0/120)$, where d is a given rod diameter.

6.6 SUMMARY

A frequency-independent antenna loses its invariant antenna characteristics when materials, such as a conducting material, are placed near the antenna for a certain purpose. Use of a cavity for transforming a bidirectional beam into a unidirectional beam is one example where deterioration of the frequency-independent characteristics occurs. However, the deterioration in the antenna characteristics can be minimized by attaching an absorbing strip to the vertical wall of the cavity, as shown in Section 6.4.2.

Recent study reveals that a reflector having an electromagnetic band gap (EBG) can transform the bidirectional beam of a planar spiral antenna into a unidirectional beam [33]. The EBG reflector differs from a conventional perfect electric conductor (PEC) in that it acts as a perfect magnetic conductor (PMC), having a reflection coefficient of +1 at its surface [34]. By virtue of this reflection coefficient, the antenna height above the surface of the EBG reflector can be made extremely small (e.g., 0.06 wavelength; See Figure 6.59).

The PMC characteristic of the EBG reflector is only sustained over a small frequency range. Therefore it cannot cover the broad operating frequency range of the planar spiral antenna. However, the EBG reflector is still useful for realizing a unidirectional, narrowband (10–16% frequency range) spiral antenna, because the antenna height can be made extremely small with a radiation efficiency of nearly 100%. Recall that use of an absorbing material decreases the radiation efficiency, as illustrated in Figure 6.22.

Finally, it is emphasized that an antenna must have a self-complementary structure in order to have broadband characteristics—the log-periodic shape itself is not a necessary

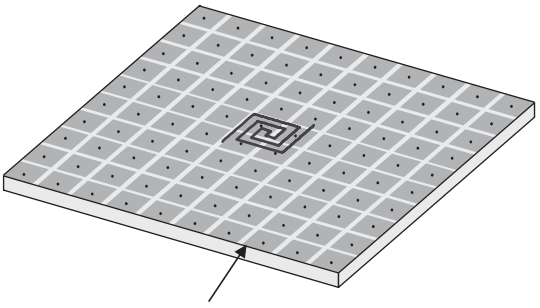


Figure 6.59 Spiral antenna above an EBG reflector.

condition for realizing broadband characteristics [3, 35, 36] (refer to the comments on Figure 6.51). This means that new advancement in broadband antenna technology should be based on the self-complementary concept, and not simply on the use of the log-periodic shape.

ACKNOWLEDGMENTS

The author thanks Y. Mushiake (Professor Emeritus, Tohoku University) for his invaluable discussion on self-complementary antennas. Also, thanks are extended to J. Yamauchi (Professor, Hosei University), V. Shkawrytko (K.K. Studio Victor), H. Mimaki (Hosei University), and M. Chiba (Hosei University) for their assistance in the preparation of this chapter.

REFERENCES

1. Y. Mushiake, *Antennas and Radio Propagation*, Chapter 7, Corona, Tokyo, 2001.
2. Y. Mushiake, Self-complementary antennas, *IEEE Antennas Propag. Mag.* Vol. 34, No. 6, pp. 23–29, December 1992.
3. Y. Mushiake, *Self-Complementary Antennas*, Springer, London, 1996.
4. Y. Mushiake and S. Adachi, *Fundamental Electromagnetic Wave Engineering*, Chapter 9, Kyoritsu Shuppan, 1973.
5. V. H. Rumsey, *Frequency Independent Antennas*, Academic Press, NewYork, 1966.
6. R. S. Elliot, *Antenna Theory and Design*, Chapter 8, Prentice Hall, Englewood Cliffs, NJ, 1981.
7. R. F. Harrington, *Field Computation by Moment Methods*. Macmillan, NewYork, 1968.
8. J. Moore and R. Pizer, *Moment Methods in Electromagnetics*, Research Studies Press, Series No. 4, John Wiley & Sons, NewYork, 1984.
9. E. Yamashita, Ed., *Analysis Methods for Electromagnetic Wave Problems*, Vol. 2, Chapter 3 Artech House, Norwood, MA, 1996.
10. K. S. Yee, Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media, *IEEE Trans. Antennas and Propagation*, Vol. 14, No. 3, pp. 302–307, May 1966.

11. A. Taflove, *Computational Electrodynamics*, Artech House, Norwood, MA, 1995.
12. C. A. Balanis, *Antenna Theory*, Third Edition, Chapter 12, John Wiley & Sons, Hoboken, NJ, 2005.
13. J. Dyson, Equiangular spiral antenna, *IRE Trans. Antennas and Propag*. Vol. 7, No. 2, pp. 181–187, April 1959.
14. J. Dyson, The unidirectional equiangular spiral antenna, *IRE Trans. Antennas Propag.* Vol. 7, No. 4, pp. 329–334, October 1959.
15. J. Dyson, New circularly-polarized frequency-independent antennas with conical beam or omnidirectional patterns, *IRE Trans. Antennas Propag*. Vol. 9, No. 4, pp. 334–342, July 1961.
16. K. Kikkawa, J. Yamauchi, and H. Nakano, Equiangular spiral antennas, Bio-Communication Research Center Report, Graduate School of Engineering, Hosei University, No. 1, 2006.
17. H. Osada, G. Tsutsumi, J. Yamauchi, and H. Nakano, Conical equiangular spiral antenna, Bio-Communication Research Center Report, Graduate School of Engineering, Hosei University, No. 2, 2006.
18. J. A. Kaiser, The Archimedean two-wire spiral antenna, *IRE Trans. Antennas Propag*. Vol. 8, No. 3, pp. 312–323, May 1960.
19. Y. Okabe, J. Yamauchi, and H. Nakano, A strip spiral antenna with absorbing layers inside a cavity, *Proc. 2001 Communications Society Conference of IEICE*, B-1-107, Chofu, Tokyo, Japan, September 2001.
20. H. Nakano, Y. Okabe, Y. Nakayama, and J. Yamauchi, Numerical analysis of a low-profile rectangular spiral antenna, *Int. J. of Applied Electromag. Mechanics*, Vol. 16, pp. 129–137, 2002.
21. R. Kobayashi, J. Yamauchi, and H. Nakano, Archimedean spiral antennas, Bio-Communication Research Center Report, Graduate School of Engineering, Hosei University, No. 3, 2006.
22. H. Nakano, S. Okuzawa, K. Ohishi, H. Mimaki, and J. Yamauchi, A curl antenna, *IEEE Trans. Antennas Propag.* Vol. 41, No. 11, pp. 1570–1575, November 1993.
23. K. Hirose and H. Nakano, Dual-spiral slot antenna, *IEE Proceedings-H*, Vol. 138, No. 1, pp. 32–36, February 1991.
24. H. Nakano, H. Yoshida, and Y. Wu, C-figured antennas, *Electron. Lett*. Vol. 31, No. 9, pp. 693–694, April 1995.
25. R. DuHamel and D. Isbell, Broadband logarithmically periodic antenna structures, *IRE Int. Conv. Rec*. Part 1, Vol. 5, pp. 119–128, March 1957.
26. R. DuHamel and F. Ore, Logarithmically periodic antenna designs, *IRE Int. Conv. Rec*. Part 1, Vol. 6, pp. 139–151, March 1958.
27. H. Iwaoka, J. Yamauchi, and H. Nakano, Tooth log-periodic antennas, Bio-Communication Research Center Report, Graduate School of Engineering, Hosei University, No. 4, 2006.
28. K. Morishita, J. Yamauchi, and H. Nakano, Trapezoidal log-periodic antennas, Bio-Communication Research Center Report, Graduate School of Engineering, Hosei University, No. 5, 2006.
29. D. Isbell, Log periodic dipole arrays, *IRE Trans. Antennas Propag*. Vol. 8, No. 6, pp. 260–267, May 1960.
30. H. Honma, J. Yamauchi, and H. Nakano, Log-periodic dipole array antennas, Bio-Communication Research Center Report, Graduate School of Engineering, Hosei University, No. 6, 2006.
31. R. L. Carrel, Analysis and design of the log-periodic dipole antenna, Ph.D. Dissertation, Elec. Eng. Dept., University of Illinois, 1961.
32. C. A. Balanis, *Antenna Theory*, Third Edition, pp. 634–635, John Wiley & Sons, Hoboken, NJ, 2005.

33. H. Nakano, K. Hitosugi, N. Tatsuzawa, D. Togashi, H. Mimaki, and J. Yamauchi, Effects on the radiation characteristics of using a corrugated reflector with a helical antenna and an electromagnetic band-gap reflector with a spiral antenna, *IEEE Trans. Antennas Propag*. Vol. 53, No. 1, pp. 191–199, January 2005.

34. D. Sievenpiper, L. Zhang, R. F. J. Broas, N. G. Alexopoulos, and E. Yablonovitch, High-impedance electromagnetic surfaces with a forbidden frequency band, *IEEE Trans. Microwave Theory Tech.* Vol. 47, No. 11, pp. 2059–2074, November 1999.

35. Y. Mushiake, http://www.sm.rim.or.jp/~ymushiake/sub.en.3.htm

36. H. Nakano, Recent progress in broadband antennas, *Proc. Int. Symp. Antennas Propag*. Singapore, CD ROM a367 November 2006.

37. R. Carrel, The design of log-periodic dipole antennas, *IRE Int. Conv. Rec*. Vol. 1, pp. 61–75, 1961.

ADDITIONAL READING

The following publications provide additional information related to the topics found in this chapter.

R. G. Corzine and J. A. Mosko, *Four-Arm Spiral Antennas*, Artech House, Norwood, MA, 1990.

D. S. Filipovic and J. L. Volakis, Novel slot spiral antenna designs for dual-band/multiband operation, *IEEE Trans. Antennas Propag.* Vol. 51, No. 3, pp. 430–440, March 2003.

D. S. Filipovic and J. L. Volakis, A flush-mounted multifunctional slot aperture (combo-antenna) for automotive applications, *IEEE Trans. Antennas Propag.* Vol. 52, No. 2, pp. 563–571, February 2005.

T. W. Hertel and G. S. Smith, The conical spiral antenna over the ground, *IEEE Trans. Antennas Propag.* Vol. 50, No. 12, pp. 1668–1675, December 2002.

J. M. Laheurte, Dual-frequency circularly polarized antennas based on stacked monofilar square spirals, *IEEE Trans. Antennas Propag.* Vol. 51, No. 3, pp. 488–492, March 2003.

R. L. Li and H. Nakano, Numerical analysis of arbitrarily shaped probe-excited single-arm printed wire antennas, *IEEE Trans. Antennas Propag.* Vol. 46, No. 9, pp. 1307–1317, September 1998.

S. K. Khamas and G. G. Cook, Moment-method analysis of printed wire spirals using curved piecewise sinusoidal subdomain basis and testing functions, *IEEE Trans. Antennas Propag.* Vol. 45, No. 6, pp. 1016–1022, June 1997.

J. D. Kraus and R. J. Marhefka, *Antennas for All Applications*, Third Edition, Chapters 8 and 11, McGraw-Hill, NewYork, 2002.

H. Nakano, *Helical and Spiral Antennas*, Research Studies Press series, John Wiley & Sons, NewYork, 1987.

H. Nakano, (Ed.), *Special Issue on Spiral Antennas, Electromagnetics*, Vol. 14, No. 3–4, pp. 259–426, July-December 1994.

H. Nakano, (Ed.), *Special Issue on Spiral Antennas*, Electromagnetics, Vol. 20, No. 4, pp. 296–386, July 2000.

H. Nakano, S. R. Kerner, and N. G. Alexopoulos, The moment method solution for printed wire antennas of arbitrary configuration, *IEEE Trans. Antennas Propag.*, Vol. 36, No. 12, pp. 1667–1674, December 1988.

H. Nakano, K. Hirose, T. Suzuki, S. R. Kerner, and N. G. Alexopoulos, Numerical analysis of printed line antennas, *IEE Proceedings, Pt. H*, Vol. 136, No. 2, pp. 98–104, 1989.

H. Nakano, K. Hirose, I. Ohshima, and J. Yamauchi, An integral equation and its application on spiral antennas on semi-infinite dielectric materials, *IEEE Trans. Antennas Propag.*, Vol. 46, No. 2, pp. 267–274, February 1998.

H. Nakano and K. Nakayama, A curved spiral antenna above a conducting cylinder, *IEEE Trans. Antennas Propag.* Vol. 47, No. 1, pp. 3–8, January 1999.

H. Nakano, J. Eto, Y. Okabe, and J. Yamauchi, Tilted- and axial-beam formation by a single-arm rectangular spiral antenna with compact dielectric substrate and conducting plane, *IEEE Trans. Antennas Propag.* Vol. 50, No. 1, pp. 17–24, January 2002.

H. Nakano, H. Yasui, and J. Yamauchi, Numerical analysis of two-arm spiral antennas printed on a finite-size dielectric substrate, *IEEE Trans. Antennas Propag.* Vol. 50, No. 3, pp. 362–370, March 2002.

H. Nakano, Y. Okabe, H. Mimaki, and J. Yamauchi, A monofilar spiral antenna excited through a helical wire, *IEEE Trans. Antennas Propag.* Vol. 51, No. 3, pp. 661–664, March 2003.

H. Nakano, M. Ikeda, K. Hitosugi, and J. Yamauchi, A spiral antenna sandwiched by dielectric layers, *IEEE Trans. Antennas Propag.* Vol. 52, No. 6, pp. 1417–1423, June 2004.

J. J. H. Wang and V. K. Tripp, Design of multioctave spiral-mode microstrip antennas, *IEEE Trans. Antennas Propag.* Vol. 39, No. 3, pp. 332–335, March 1991.

CHAPTER 7

Leaky-Wave Antennas

DAVID R. JACKSON and ARTHUR A. OLINER

7.1 INTRODUCTION

Leaky-wave antennas are a class of antennas that use a traveling wave on a guiding structure as the main radiating mechanism [1–4]. These antennas are capable of producing narrow beams, with the beamwidth limited by the size of the structure. They are often planar or nearly planar, with only a modest depth requirement. They possess the advantage of simplicity, since no complicated feed network is required as, for example, in a planar array antenna. Because of this simplicity, they are often attractive for higher microwave and millimeter-wave frequencies. Most leaky-wave antennas have the inherent property that the beam scans with frequency. For scanning applications this is an important advantage, while for point-to-point communications this is usually a disadvantage, since it limits the pattern bandwidth of the antenna (e.g., defined from the frequency range over which the gain in a fixed observation direction is within 3 dB of that at the optimum frequency). Because of this, the pattern bandwidth of the leaky-wave antenna for fixed-point communications will usually decrease along with the beamwidth.

Leaky-wave antennas support a fast wave on the guiding structure, where the phase constant β is less than the free-space wavenumber k_0. The leaky wave is therefore fundamentally a radiating type of wave, which radiates or "leaks" power continuously as it propagates on the guiding structure (and hence the name of the antenna). The operation is therefore quite different from a slow-wave or surface-wave type of antenna, where radiation mainly takes place at the end of the antenna [5]. Because of the leakage of power, the propagation wavenumber $k_z = \beta - j\alpha$ on the guiding structure is complex, consisting of both a phase constant β and an attenuation constant α (even if the structure is lossless).

Highly directive beams at some angle may be achieved with this type of antenna, with a low sidelobe level (this usually requires a tapered aperture illumination in practical situations, depending on the aperture efficiency that one is trying to obtain). As a general rule, the phase constant of the leaky wave controls the beam angle, while the attenuation constant controls the beamwidth. The aperture distribution may be tapered to control the sidelobe level or the beam shape.

Modern Antenna Handbook. Edited by Constantine A. Balanis

7.2 HISTORY

The first known leaky-wave antenna was the slitted rectangular waveguide, introduced by W. W. Hansen in 1940 [6]. The structure is shown in Figure 7.1. The area of leaky-wave antennas did not see a lot of development until the 1950s, however. Starting in the 1950s a variety of different types of leaky-wave antennas were introduced, and methods were developed for their analysis. An excellent summary of these structures and methods may be found in Ref. 4.

Most of the initial leaky-wave antennas were based on closed waveguides, where the leakage was obtained by introducing long uniform slits into the waveguides to allow them to radiate. However, these slits cut across the current lines in the closed waveguides, producing a strong perturbation on the fields, so that it was difficult to produce leaky waveguides with low leakage per unit length, and therefore narrow beams. One method for overcoming this difficulty was proposed by Hines and Upson [7], in which the long slit was replaced by a series of closely spaced holes, thereby avoiding cutting the current lines. This structure was called a "holey waveguide," and it permitted the antenna to radiate much narrower beams. The structure is shown in Figure 7.2, when the waveguide is air-filled and the holes are closely spaced, so that they act in a quasi-uniform manner.

Other novel structures were introduced by Rotman and colleagues. One of them was the "sandwich wire antenna" [8], which conceptually was the earliest example of a one-dimensional periodic leaky-wave antenna, but it was not practical. Another basic novel approach was Rotman's recognition that one can take an open guiding structure that did not radiate due to its inherent symmetry, but then became leaky through the introduction of some asymmetry in the design. He chose a symmetric trough waveguide as the open waveguide that did not radiate and then added a simple form of asymmetry to produce the radiation in a leaky-wave fashion. With different colleagues, he took measurements of its behavior as a leaky-wave antenna and developed a successful simple theory for its performance [9]. A description of the structure and a summary of the theory are included in Ref. 4. The antenna was very successful, and it has been widely used in various frequency ranges. This same basic principle was widely applied to various other leaky-wave antenna structures decades later.

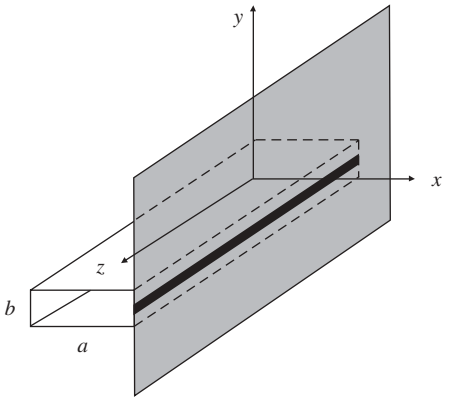


Figure 7.1 A leaky-wave antenna consisting of a rectangular waveguide with a long longitudinal slot in the narrow wall of the waveguide. An infinite ground plane is shown surrounding the slot.

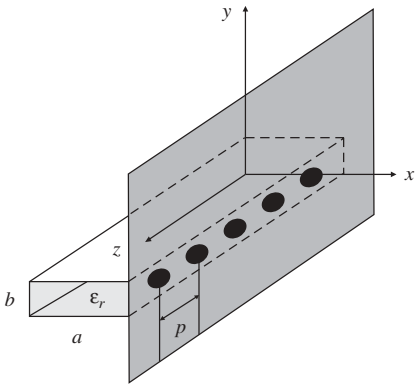


Figure 7.2 A periodic leaky-wave antenna consisting of a rectangular waveguide that is filled with a dielectric material and loaded with a periodic array of holes (apertures) in the narrow wall of the waveguide. An infinite ground plane is shown surrounding the holes.

Much later, an investigation of one-dimensional (1D) *arrays* of 1D leaky-wave antennas was performed by Oliner and co-workers, principally Lampariello, Shigesawa, and Peng, starting in the 1980s, and this work is summarized in a comprehensive report [10].

Also introduced in the 1950s, by von Trentini [11], was the concept of a *two-dimensional* leaky-wave antenna using a periodic partially reflective screen over a ground plane in order to obtain directive pencil beams at broadside. This novel structure allowed for directive beams to be obtained when a simple waveguide aperture feed was used as the source. This pioneering work also laid the foundation for further work in the area of two-dimensional (2D) leaky-wave antennas. However, it was not appreciated at the time that this type of structure was actually a quasi-uniform leaky-wave antenna.

Investigations of a uniform 2D leaky-wave antenna consisting of a dielectric superstrate layer over a substrate layer were conducted by Alexopoulos and Jackson in the 1980s [12, 13], although this structure was also not initially recognized as a leaky-wave antenna. The analysis of this type of structure as a leaky-wave antenna was done by Jackson and Oliner in the late 1980s and early 1990s [14, 15]. Later, 2D leaky-wave antennas using periodic partially reflecting screens (as was used in the original von Trentini antenna) were examined in more detail by Feresidis and Vardaxoglou [16] using quasi-uniform partially reflective screens consisting of various elements including metal dipoles and rectangular patches, circular patches, circular loops, square loops, crossed dipoles, and their complements. An investigation of the radiation characteristics of quasi-uniform 2D leaky-wave antennas using metal dipoles or slots was also discussed in some detail in Refs. 17 and 18.

More recently, the interest that has developed in metamaterial structures has led to an investigation of novel 2D structures for obtaining narrow-beam patterns at broadside. Unfortunately, much of this work has not recognized that the structures that were introduced are in fact 2D leaky-wave antennas. Many of the narrow-beam radiation effects were explained in terms of Fabry–Perot resonances or electromagnetic bandgaps. These alternative points of view provide useful insights, although an explanation in terms of leaky waves provides the most physically fundamental way to explain their operation. Research in this area is evolving, and some examples of metamaterial leaky-wave

antennas are given later in this chapter. Also evolving is the development of novel periodic 1D leaky-wave antennas that have a greatly reduced or even absent stopband at broadside, allowing the antennas to scan continuously through broadside. (As discussed later, the stopband feature is usually a serious limitation of 1D periodic leaky-wave antennas.) Some of these antennas are based on metamaterial concepts (such as the "composite right/left-handed antenna") while others are not.

The various types of leaky-wave antennas that have been mentioned in this historical overview are discussed in more detail in Section 7.3, while the fundamental properties of these different types of leaky-wave antennas, and design formulas for them, are examined in greater detail in the remaining sections of this chapter. Because many specific leaky-wave antennas are described in detail in Ref. 4, the emphasis in the present chapter is on the fundamental properties of leaky waves and how these properties affect the basic performance behavior of the different antenna types discussed here.

7.3 CLASSIFICATION OF LEAKY-WAVE ANTENNAS

Leaky-wave antennas can be divided into different categories, depending on the geometry and the principle of operation. The first distinction that can be made is between a *one-dimensional* (1D) leaky-wave antenna and a *two-dimensional* (2D) leaky-wave antenna. Another classification is whether the structure is *uniform*, *quasi-uniform*, or *periodic*. These different types of leaky-wave antennas are overviewed in the subsections below, and then discussed in more detail in later sections.

7.3.1 One-Dimensional Uniform Leaky-Wave Antennas

In the terminology used here, a 1D leaky-wave antenna is one where the guiding structure is basically one dimensional; that is, the structure supports a wave traveling in a fixed direction. An example would be a rectangular waveguide that has a long slit to allow power to leak continuously from the waveguide [19], as shown in Figure 7.1. This is also an example of a *uniform* structure, since the geometry does not change in the longitudinal (z) direction. (The slot may be gradually tapered to allow control of the beam shape, but for classification purposes, the antenna is still considered as a uniform one.) Reference 19 also presents practical solutions for several other uniform leaky-wave antennas based on rectangular waveguide.

If an infinite ground plane is used to form a baffle as shown in Figure 7.1, and the slot is narrow, the structure shown in Figure 7.1 is equivalent, by image theory, to an infinite magnetic line current flowing in the z-direction in free space. For a leaky-wave antenna of finite length, the line source will be of finite length. The corresponding radiation pattern will then be a beam in the form of a cone, with the axis of the cone being the z-axis. The radiation from this type of leaky-wave antenna is limited to the forward quadrant ($z > 0$). As the angle of the cone from the z-axis approaches 90°, we approach a broadside beam, where the pattern is a narrow-beam donut type of pattern. For such a uniform leaky-wave antenna fed at one end (and therefore supporting a wave propagating in a single direction on the guiding structure) it is difficult to obtain a beam exactly at broadside, since this corresponds to operation of the waveguide at cutoff. A broadside beam may be achieved, however, by operating close to (but slightly above) cutoff and feeding the antenna at both ends, or with a source in the middle of the structure, so

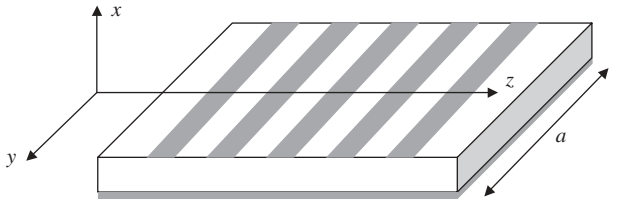


Figure 7.3 The geometry of the leaky-wave antenna studied by Honey [20], consisting of a closely spaced array of metal strips over a ground plane. The transverse width a of the structure is large relative to a wavelength and is comparable to the length. The electric field is parallel to the strips.

that two beams, each pointing close to but on opposite sides of broadside, are created. Through proper design, these two beams may merge to form a single broadside beam.

A pencil (or spot) beam can be obtained by using a one-dimensional array of 1D leaky-wave antennas [4, 10] (e.g., a 1D array of slitted waveguides). This creates a two-dimensional radiating aperture that produces a pencil beam, with the pointing direction determined by the beam angle of the individual leaky-wave antennas and the phasing between adjacent leaky-wave antennas [4, 10]. Therefore the beam angle may be controlled in the elevation plane (x-z plane) by frequency, and in the azimuth plane (x-y plane) by changing the phase shift between adjacent waveguides.

Another variation of the 1D leaky-wave antenna is a structure where the propagation is in one direction, but the width of the antenna (perpendicular to the direction of propagation) is large, so that a two-dimensional radiating aperture is created. Such a structure was examined by Honey [20] using a wire grating structure to form a wide leaky parallel-plate waveguide. Such a structure is illustrated in Figure 7.3, where the electric field is in the y-direction. This structure may produce a fan beam or a pencil beam (depending on the width of the structure), but the beam is scannable only in one plane, and the beam scan is limited to the forward quadrant (unless the structure is fed from both ends).

7.3.2 One-Dimensional Periodic Leaky-Wave Antennas

A 1D *periodic* leaky-wave antenna structure is one that consists of a uniform structure that supports a slow (nonradiating) wave with $\beta > k_0$, which has been periodically modulated in some fashion in the longitudinal (z) direction. An example would be the structure shown in Figure 7.2, consisting of a rectangular waveguide that is filled with a dielectric material so that the fundamental waveguide mode is a slow wave in the frequency range of operation, and which then has a periodic set of small holes or slots introduced into the narrow wall of the waveguide. Since a slow wave radiates only at discontinuities, the periodic modulations (discontinuities) allow the wave to radiate continuously along the length of the structure. From another point of view, the periodic modulation creates a guided wave that consists of an infinite number of space harmonics (Floquet waves) [2, 3]. The nth Floquet wave has a wavenumber given by $k_{zn} = k_{z0} + 2\pi n/p$, where p is the period and $k_{z0} = \beta - j\alpha$ is the wavenumber of the fundamental Floquet wave. Although the main ($n = 0$) space harmonic is a slow wave, one of the space harmonics (usually $n = -1$) is designed to be a fast wave, so that $-k_0 < \beta_{-1} < k_0$, and hence this space harmonic is a radiating wave.

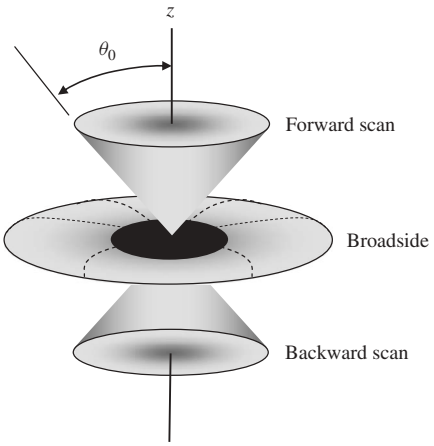


Figure 7.4 An illustration of the type of beams that can be produced by a 1D periodic leaky-wave antenna, showing a beam that scans from the backward region to broadside and then to the forward region.

An advantage of a periodic leaky-wave antenna is that the beam can point in either the backward or the forward direction, corresponding to $\beta_{-1} < 0$ or $\beta_{-1} > 0$, respectively. Furthermore, the beam will scan with frequency, so that a beam that scans from the backward quadrant to the forward quadrant as the frequency increases can easily be obtained. The beam will then be in the form of a conical beam that opens along the negative z-axis at an angle θ_0 that increases with frequency, approaching a donut-shaped pattern at broadside. As the frequency continues to increase past the broadside point the beam becomes a conical beam that now opens about the positive z-axis, with an angle θ_0 that decreases with increasing frequency. This scan behavior is illustrated in Figure 7.4. It is often difficult to obtain a good beam exactly at broadside, however, since this corresponds to an "open stopband" point for the periodic structure [2, 3]. At the stopband point, the periodic traveling-wave antenna turns into a standing-wave antenna, where the attenuation constant drops to zero. At this point the input match to the antenna will typically become quite poor, and the beamwidth of the antenna will change dramatically. (If the structure were infinitely long, there would be a perfect mismatch at the input, and the beamwidth would drop to zero.) Further details about the stopband, and how it may be overcome through the use of novel designs, are discussed in more detail later.

If operation only at broadside is required, then a standing-wave antenna design may be used, where the end of the waveguiding structure is terminated with a reactive load such as a short circuit instead of an absorbing load. In this case the standing wave will ensure that all of the radiating elements are fed in phase if they are spaced one guide wavelength apart. Many waveguide-fed slot arrays are based on this type of principle, where radiating slot elements are spaced one-half of a guide wavelength apart along the waveguide and arranged in an alternating fashion about the centerline of the waveguide, in order to obtain in-phase excitation [21]. The reader is referred to Ref. 21 to obtain further details about waveguide-fed slot arrays. These types of antennas are well suited for broadside operation but not for frequency-controlled beam scanning.

As with the 1D uniform leaky-wave antenna, the 1D periodic leaky-wave antenna may be used as an element of an array in order to obtain pencil beams. The pencil beams may be scanned over all quadrants of the hemispherical space, except that care must be taken when scanning near broadside, as mentioned earlier.

7.3.3 One-Dimensional Quasi-Uniform Leaky-Wave Antennas

This type of antenna is similar to the 1D uniform leaky-wave antenna, except that a periodic structure is used in the geometry of the waveguiding structure. However, unlike the 1D periodic leaky-wave antenna, the fundamental waveguide mode is still a fast wave, and radiation occurs via the fundamental waveguide mode, and not one of the space harmonics. An example would be the "holey waveguide" antenna introduced in Ref. 7, which is the structure shown in Figure 7.2 when the waveguide is air filled. The fundamental TE_{10} mode is therefore a fast wave. Although there would still be an infinite set of space harmonics, the period would now be chosen small enough so that only the fundamental ($n = 0$) space harmonic, corresponding to the fundamental waveguide mode (which is perturbed by the holes), would be a fast wave and would therefore radiate. For modeling purposes, then, the wall of the waveguide containing the holes could be replaced with an equivalent uniform surface impedance, resulting in a perfectly uniform structure. Another example of a quasi-uniform structure is the array of closely-spaced metal strips introduced by Honey [20] that is shown in Figure 7.3.

7.3.4 Two-Dimensional Leaky-Wave Antennas

A 2D leaky-wave antenna is one where the guiding structure itself has a *two-dimensional* surface and the leaky wave is a *cylindrical leaky wave* that propagates outward in the radial direction from the source or feed. This type of leaky-wave antenna is capable of producing a pencil beam at broadside, or a conical beam with the cone axis perpendicular to the aperture plane, as illustrated in Figures 7.5a and b. One example is a two-layer substrate/superstrate structure consisting of a substrate layer with a high permittivity superstrate layer on top, as shown in Figure 7.6, where a simple horizontal electric dipole source is also shown (a further discussion of this type of antenna is given later). A 2D leaky-wave antenna provides a simple means to obtain a directive beam at broadside using only a simple source. The pattern bandwidth of the beam will be limited, however, since the beam will evolve into a conical beam as the frequency increases, resulting in a loss of gain in the broadside direction.

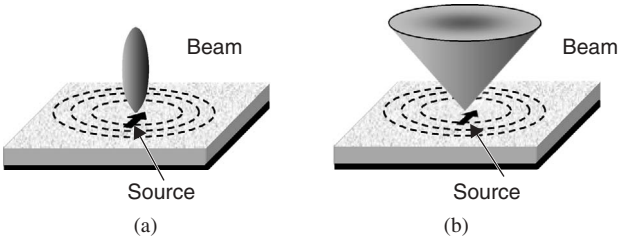


Figure 7.5 An illustration of the type of beams that can be produced by a 2D leaky-wave antenna: (a) a pencil beam at broadside and (b) a conical beam at an angle θ_0.

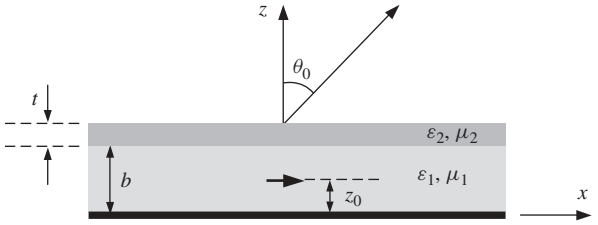


Figure 7.6 The substrate/superstrate structure, shown excited by a horizontal electric dipole at height z_0 above the ground plane.

For most 2D leaky-wave antennas the guiding structure is either uniform or quasi-uniform, and the radially propagating mode is a fast wave. Therefore the discussion in this chapter is limited to these types of 2D leaky-wave antennas, which are essentially leaky parallel-plate waveguiding structures, where the top plate of the parallel-plate waveguide has been replaced with a partially reflecting surface of some sort, which allows leakage to occur. Such structures are discussed in more detail later in this chapter.

7.4 PHYSICS OF LEAKY WAVES

7.4.1 Field Behavior of Leaky Waves

As mentioned previously, a leaky-wave antenna supports a fast wave with $|\beta| < k_0$. To illustrate some of the important radiation features of a leaky wave, consider the simple case of an aperture as shown in Figure 7.7, which has an electric field $E_y(x, z)$ on the aperture ($x = 0$) that has the form of a leaky wave,

$$E_y(0, z) = Ae^{-jk_z z} \tag{7.1}$$

where the complex wavenumber of the leaky wave is given by

$$k_z = \beta - j\alpha \tag{7.2}$$

where β is the phase constant and α is the attenuation constant, and A is a constant. (The guiding structure that is used to support such a wave may take a variety of forms, including, for example, the layered structure shown in Figure 7.6.) In a leaky-wave antenna the attenuation constant corresponds to a loss due to the leakage of power along the structure as the wave propagates. (If the structure supporting the leaky wave has

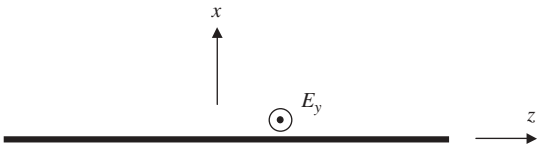


Figure 7.7 An infinite aperture that is used in the discussion of radiation from a leaky wave. The aperture is at $x = 0$.

conductor and/or dielectric loss, then the attenuation constant accounts for these losses as well.) The field in the air region above the aperture ($x > 0$) is given by

$$E_y(x, z) = Ae^{-jk_z z}e^{-jk_x x} \tag{7.3}$$

where the vertical wavenumber is

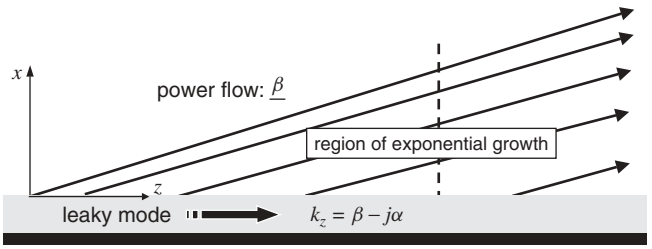
$$k_x = \left(k_0^2 - (k_z)^2\right)^{1/2} \tag{7.4}$$

with k_0 being the wavenumber of free space. Denoting $k_x = \beta_x - j\alpha_x$, we have, after squaring both sides of Eq. (7.4) and equating the imaginary parts, that

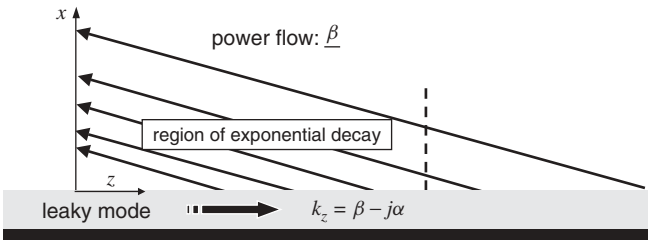
$$\beta\alpha = -\beta_x\alpha_x \tag{7.5}$$

If we assume that the wave is a forward wave with $\beta > 0$ and $\alpha > 0$, and is propagating vertically away from the structure, then from Eq. (7.5) it immediately follows that $\alpha_x < 0$. This means that the wave in the air region must be *exponentially increasing*. Interestingly, any forward wave that decays in the longitudinal (z) direction due to leakage loss must increase exponentially in the surrounding air region. The leaky wave is often described as being "improper" or "nonspectral," meaning exponentially increasing in the air region.

At first glance this behavior might seem rather unnatural and nonphysical. It might even be thought that a leaky wave is a nonphysical type of wave because of this behavior, since it violates the radiation condition at infinity in the transverse direction. However, this behavior makes sense from a physical point of view and can be explained by a simple ray diagram [19, 22, 23]. Figure 7.8a shows rays that indicate the direction of power flow in the air region, which for an inhomogeneous plane-wave field (the type of



(a)



(b)

Figure 7.8 Ray pictures that are used in the physical interpretation of a leaky-wave field: (a) ray picture for a forward leaky wave and (b) ray picture for a backward leaky wave.

field that describes the field of the leaky wave in the air region) is in the direction of the phase vector that is the real part of the complex wavenumber vector,

$$\boldsymbol{\beta} = \hat{\mathbf{x}}\beta_x + \hat{\mathbf{z}}\beta_z = \text{Re}\,(\mathbf{k}) = \text{Re}\,(\hat{\mathbf{x}}k_x + \hat{\mathbf{z}}k_z) \tag{7.6}$$

where $\beta_z = \beta$. The phase vector makes an angle θ_0 with respect to the z-axis, defined by

$$\tan\theta_0 = \frac{\beta_x}{\beta_z} \tag{7.7}$$

If the attenuation constant α is small, the angle θ_0 is given to a good approximation by

$$\cos\theta_0 = \frac{\beta}{k_0} \tag{7.8}$$

Equation (7.8) is very useful for determining the beam angle of most practical leaky-wave antennas. The separation between the rays in Figure 7.8a indicates the strength of the field (a closer separation denotes a stronger field level), with the field being stronger for the rays that emanate closer to the source. If an observer moves vertically away from the aperture along a vertical line (e.g., the dashed line shown in the figure), it is seen that the field level increases. This simple ray picture is thus consistent with the conclusion drawn from Eq. (7.5).

In Figure 7.8b a ray picture is provided for the case where the leaky wave is a "backward wave," meaning that the phase and group velocities are in opposite directions. It is assumed here that the group velocity (which is usually the direction of power flow) is in the positive z-direction, while the phase velocity is in the negative z-direction, and therefore $\beta < 0$. The wave field is still assumed to be an outgoing field (carrying power away from the aperture) and therefore $\beta_x > 0$. Note that from Eq. (7.5) we then have that $\alpha_x > 0$. The fields of the leaky wave are therefore "proper," meaning that they decay exponentially in the vertical direction away from the aperture. This conclusion is consistent with the ray picture shown in Figure 7.8b for this type of wave.

Therefore, for the field behavior of a leaky wave, the conclusion is that the wave is improper if the wave is a forward wave, and proper if it is a backward wave. A uniform leaky-wave antenna normally only supports a forward type of wave, while a periodic leaky-wave antenna can support either type. For example, consider a periodic leaky-wave antenna supporting a mode that is carrying power in the positive z-direction, and assume that radiation is occurring from the $n = -1$ space harmonic. If this space harmonic is radiating into the backward quadrant ($z < 0$), the field of the -1 space harmonic will be proper. If the frequency is increased so that the beam now scans into the forward quadrant ($z > 0$) the field of this space harmonic will become improper. These field considerations can be important when analyzing leaky-wave antennas, in order to make sure that the field behavior is that which is expected on physical grounds, so that the leaky mode is a physical one.

The previous discussion has assumed an infinite leaky wave propagating from $z = -\infty$ to $z = \infty$. In a practical leaky-wave antenna, the leaky wave will never exist over this entire range, since it will be excited, or launched, by a feed or source at some point on the structure. For example, consider a leaky wave that is launched by a source at $z = 0$. Assuming that the wave is launched equally in both directions, the field on the aperture would then have the form

$$E_y\,(0, z) = Ae^{-jk_z|z|} \tag{7.9}$$

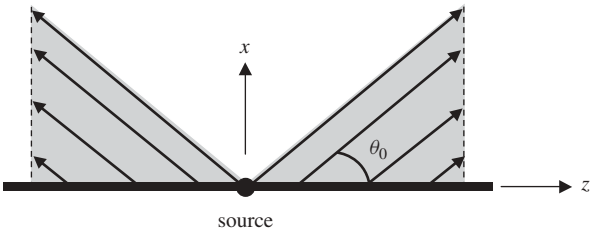


Figure 7.9 Ray picture for a leaky wave that is excited by a line source at $z = 0$.

In this case the wave decays exponentially in both directions, $z = \pm\infty$. Assuming that the wave is a forward wave, the ray picture for this case is shown in Figure 7.9. The picture is symmetric about the vertical x-axis, so for simplicity the following discussion is limited to the region $z > 0$. Note that there now exists a leakage "shadow boundary" at the angle $\theta = \theta_0$ from the z-axis. Within a wedge-shaped region defined by $\theta < \theta_0$ the field is similar to that of the infinite leaky wave shown in Figure 7.8a. Within the shadow region defined by $\theta_0 < \theta < \pi/2$ the field is very weak. (The simple approximate ray picture predicts a zero field in this region, but in actuality there will be some field in the shadow region.) The simple ray model predicts that, as an observer moves vertically away from the aperture, the field level will increase exponentially up to the leakage shadow boundary and will then decrease very quickly above this boundary. Therefore the field from this bidirectional leaky wave will not increase indefinitely in the vertical direction and will not violate the radiation condition at infinity.

The exact field $E_y(x, z)$ due to the aperture field given in Eq. (7.9) may be calculated by using a simple Fourier transform approach. The result is

$$E_y(x, z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{E}_y\left(0, k_z'\right) e^{-jk_x' x} e^{-jk_z' z} dk_z' \tag{7.10}$$

where the Fourier transform of the aperture field is

$$\tilde{E}_y\left(0, k_z'\right) = A\left[\frac{2jk_z}{\left(k_z'^2 - (k_z)^2\right)}\right] \tag{7.11}$$

and the vertical wavenumber is given by

$$k_x' = \left(k_0^2 - k_z'^2\right)^{1/2} \tag{7.12}$$

Note that in Eq. (7.12) the vertical wavenumber is chosen as a positive real number or a negative imaginary number, in order to satisfy the radiation condition at infinity.

Figure 7.10 shows a plot of the field level for a typical leaky wave having a wavenumber corresponding to $\beta/k_0 = \sqrt{3}/2$ and $\alpha/k_0 = 0.02$. This corresponds to a leakage angle of $\theta_0 = 30^\circ$ according to Eq. (7.8). It is seen that the field level does increase vertically up to the angle predicted (30°) and decreases vertically above this boundary. The exact field changes smoothly across the leakage shadow boundary, as it must. However, the smaller the value of the attenuation constant, the more closely the exact field will resemble that predicted by the simple ray model shown in Figure 7.9.

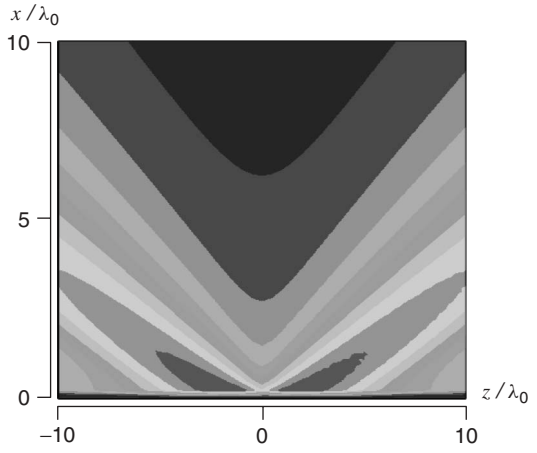


Figure 7.10 A contour plot of the magnitude of the electric field produced by a leaky wave on an aperture. The leaky wave has the parameters $\beta/k_0 = \sqrt{3}/2$, $\alpha/k_0 = 0.02$.

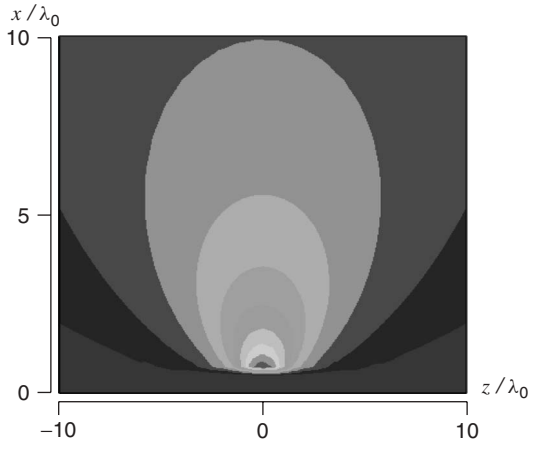


Figure 7.11 The same type of plot as in Figure 7.10, but for a slow wave that has $\beta/k_0 = 3/2$ and $\alpha/k_0 = 0.02$.

As an interesting contrast, Figure 7.11 shows the same type of result as in Figure 7.10, but for a slow wave, having $\beta/k_0 = 3/2$ and $\alpha/k_0 = 0.02$. Equation (7.8) predicts that there will be no leakage angle in visible space (a real angle θ_0). The consequence of this is that the field does not exhibit any noticeable leakage at a particular angle; instead, the field is a rather diffuse near field that decays rapidly away from the origin. Because this wave field is a slow wave, radiation occurs only at discontinuities. (If the aperture field were an infinite one without discontinuities, as given by Eq. (7.1), the field above the aperture would correspond to a surface-wave type of field that would decay exponentially vertically, and there would be no radiation.) The only discontinuity here is that corresponding to the source, located at $z = 0$, and hence the radiation field is seen to emanate from the source.

A slow wave may correspond to a surface-wave type of mode (which only radiates at discontinuities) or a leaky wave that is slow, so that $|\beta| > k_0$. It is certainly possible to find leaky modes that exist on various structures, where the wavenumber is complex but the wave is slow. Such leaky modes are generally regarded as "nonphysical."

7.4.2 Mathematical Description of the Physical Significance of Leaky Waves

The previous section illustrated the basic field behavior of a leaky wave and qualitatively discussed the issue of the physical significance of leaky waves, by using simple ray pictures. Based on this, it was postulated that leaky waves may be physically significant if they are fast waves, while they are not likely to be physically significant if they are slow waves (waves that do not radiate except at discontinuities). From a practical point of view, a leaky wave may be regarded as being physically significant if it is possible to measure fields in a region of space (e.g., the wedge-shaped region shown in Figure 7.9) that closely resemble the fields of the leaky mode, when a guiding structure is excited by a finite source. This point of view also allows for the development of a mathematical theory to explain the physical significance of leaky waves, by analyzing a guiding structure that is excited by a finite source. This is perhaps the most sophisticated means to examine the physical significance of leaky waves. This type of mathematical development is presented in this section, where it is assumed that the reader has some familiarity with the theory of sources near layered media and the steepest-descent method of asymptotic evaluation. The reader is referred to references such as Ref. 24 for a discussion of these topics.

To obtain further insight into the issue of physical versus nonphysical leaky modes, it is illustrative to consider the simple case of a layered structure consisting of one or more lossless dielectric layers over a ground plane. As discussed later, such structures, when designed properly, can make very effective leaky-wave antennas (Figure 7.6 shows one such example). The leaky modes that exist on such a structure evolve from the well-known surface-wave modes that exist on the structure, as the frequency is lowered. The behavior of a typical mode as the frequency changes, illustrated in Figure 7.12, is described next. It is assumed that we are discussing a surface-wave mode that has a nonzero cutoff frequency, which evolves into a complex leaky mode at a sufficiently low frequency. This means any mode other than the fundamental TM_0 surface-wave mode (which has a zero cutoff frequency) and the TE_1 surface-wave mode (which remains an improper real surface-wave mode below its cutoff frequency, and therefore never becomes a complex leaky mode).

Above the cutoff frequency of the surface-wave mode, the mode is a slow wave, so that $\beta > k_0$. The field of the mode is decaying exponentially away from the structure in the vertical direction in the air region, with a typical dispersion behavior that is illustrated by the solid line in Figure 7.12a. As the frequency is lowered, the normalized wavenumber β/k_0 decreases, and at the cutoff frequency f_c of the surface-wave mode shown in Figure 7.12a, $\beta/k_0 = 1$. As the frequency is lowered below the cutoff frequency, the surface-wave mode becomes an "improper real" surface-wave mode, where the wavenumber is real (no attenuation constant) and the wave is slow, with $\beta > k_0$. In this frequency region the wavenumber of the mode is denoted with a dashed line in Figure 7.12a. As the frequency is lowered further below cutoff, the value of the phase constant continues to increase. At a sufficiently low frequency, a "splitting-point" frequency f_s is encountered, where the improper real surface-wave mode merges with another improper real

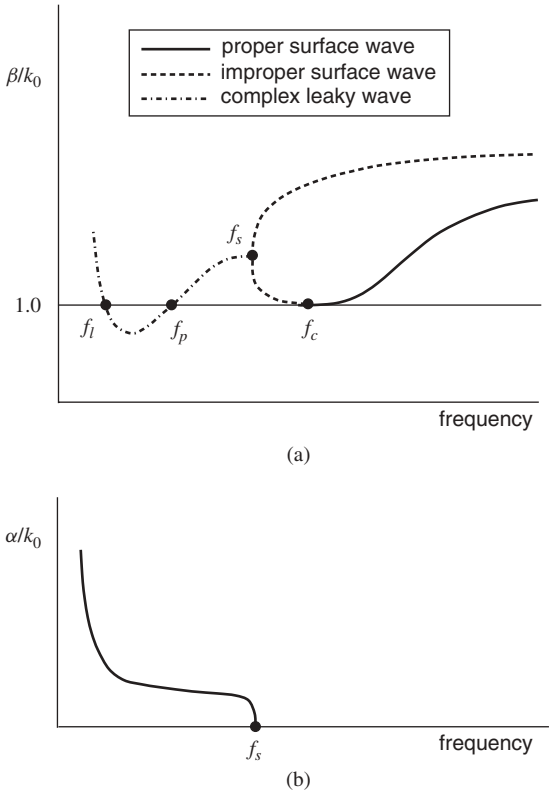


Figure 7.12 (a) An illustration of typical behavior for the normalized phase constant β/k_0 versus frequency for a guided mode on a grounded lossless dielectric slab. Above the cutoff frequency f_c, the mode is a proper nonradiating surface-wave mode. Below the cutoff frequency but above the splitting-point frequency f_s, the mode is a nonphysical improper surface-wave mode with a real wavenumber. Below the splitting-point frequency, the mode is a complex improper leaky-wave mode. The leaky-wave mode becomes physical below the frequency f_p, where it enters the fast-wave region, and it remains physical below this frequency until the mode becomes a slow wave at the frequency f_l. (b) An illustration of typical behavior for the normalized attenuation constant α/k_0 versus frequency for the guided mode.

surface-wave mode (shown by the upper dashed curve) that is never physical, and which does not evolve from a proper surface-wave solution. The two wavenumbers split apart into two complex wavenumbers as the frequency is then further lowered. The first of these two wavenumbers has the form $k_z = \beta - j\alpha$, where the phase and attenuation constants are both positive, and the second one has a wavenumber that is the complex conjugate of the first one. The phase constants for these two modes are the same, and the wavenumber is denoted with a dotted–dashed line in Figure 7.12a. The second complex solution is always regarded as being nonphysical, since this wavenumber corresponds to a mode that grows exponentially in the direction of propagation z. If we track the first wavenumber as the frequency is further lowered, the phase constant will typically continue to decrease until we reach the frequency f_p, where $\beta = k_0$. A further lowering of the frequency will then result in $\beta < k_0$. The leaky mode is then said to have entered into the "physical"

region (more will be said about this later). Below a still lower frequency f_l the leaky mode again becomes a nonphysical slow wave. Within the physical region between f_p and f_l the total attenuation consists partly of power leakage of the type described in Figure 7.8a, and partly of below-cutoff field behavior. The power leakage occurs primarily in the portion nearest to f_p, for which β/k_0 decreases as frequency decreases (and α/k_0 increases), and the below-cutoff behavior occurs nearer to f_l, for which β/k_0 increases as frequency decreases further (and α/k_0 is found to increase strongly). A sketch of the normalized attenuation constant is shown in Fig. 7.12b. The attenuation constant is zero for frequencies above the splitting-point frequency f_s. The normalized attenuation constant typically increases as the frequency is lowered, and becomes large rather quickly for frequencies below the quasi-cutoff frequency where $\beta = \alpha$.

The behavior of the wavenumber in the complex plane for frequencies below the splitting-point frequency is illustrated in Figure 7.13, which shows the trajectory of the wavenumber for a typical mode in the complex k_z plane once the mode has become leaky (below the splitting-point frequency). Also shown is the usual Sommerfeld (hyperbolic) branch cut for the wavenumber $k_x = \left(k_0^2 - k_z^2\right)^{1/2}$ that emanates from the branch point at $k_z = k_0$. (There is also a symmetrically located branch cut from the branch point at $k_z = -k_0$, but this is not shown.) The surface-wave mode starts on the real axis on the top (proper) sheet of the two-sheeted Riemann surface for the wavenumber k_x. At cutoff the wavenumber k_z moves to the branch point at $k_z = k_0$ and then below cutoff it emerges from the branch point and moves upward along the real axis, on the bottom (improper) sheet. At the splitting point the improper real surface-wave mode merges with another one that is moving down along the real axis. The two solutions then become complex and leave the real axis in opposite vertical directions at the splitting point (the nonphysical growing mode, which departs vertically upward from the real axis, is not shown in Figure 7.13). As the frequency is lowered further the leaky mode crosses the vertical line $\mathrm{Re}\,(k_z) = k_0$ and enters into the physical region.

The frequency region between the cutoff frequency of the surface-wave mode (on the high end) and the frequency for which the leaky mode has $\beta = k_0$ (on the low end) is referred to as the "spectral-gap region" [25], a term coined by Oliner. This is the part

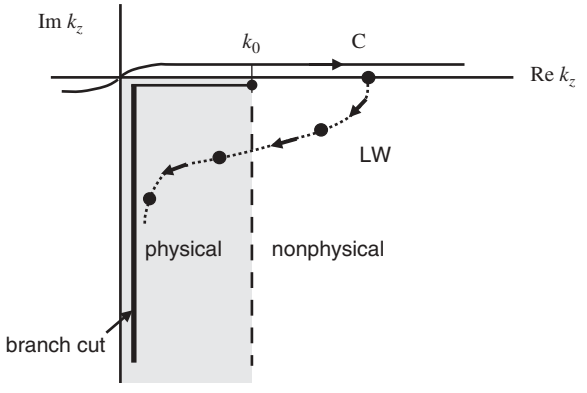


Figure 7.13 A sketch showing the trajectory of a typical leaky-mode pole in the complex wavenumber (k_z) plane below the splitting-point frequency. Also shown is the path of integration C and the branch cut from the branch point at $k_z = k_0$. The physical fast-wave region and the nonphysical slow-wave region are also labeled.

of the frequency spectrum for which the mode is nonphysical. Above the spectral-gap region the mode is a physical surface-wave mode, which carries power but does not radiate. Below the spectral-gap region the mode is a physical radiating leaky mode with $\beta < k_0$, and the attenuation constant mainly corresponds to leakage loss (assuming a lossless structure), for frequencies such that $\beta > \alpha$. As the frequency is lowered, the trajectory of the curve in the physical region typically bends down and the mode goes through a quasi-cutoff at roughly the frequency where $\beta = \alpha$. At this frequency the wavenumber is located on a diagonal line tilting 45° down from the real k_z axis. As the frequency is lowered still further the curve continues to move downward as shown in Figure 7.13, and the attenuation constant continues to increase so that $\alpha > \beta$, but the attenuation in this region mainly corresponds to reactive attenuation instead of leakage. The mode is improper in the region of physical radiation, but this is expected on physical grounds, as explained earlier. Within the spectral-gap region the mode is improper, but it is a slow wave (either an improper surface-wave mode or a leaky mode, corresponding to frequencies above or below the splitting-point frequency). This type of wave field is inconsistent with what is expected from the simple ray picture shown in Figure 7.8a, where the improper nature of the field is associated with leakage at some physical angle in space, corresponding to a *fast* wave. The mode is therefore regarded as being nonphysical within the spectral-gap region.

Further insight into the physical significance (or lack of it) for a leaky mode may be obtained by considering the problem of an infinite electric line source of I_0 amperes at a radian frequency ω located at the top interface of (but infinitesimally above) a layered structure, as shown in Figure 7.14. (A practical leaky-wave antenna would usually be excited by a source within the structure, as shown in Figure 7.6, but a line source on top of a layered structure is assumed here for simplicity in the analysis.) The exact electric field above the structure ($x > 0$) may be written in the form

$$E_x(x, z) = -j\omega \left(\frac{\mu_0 I_0}{4\pi j}\right) \int_{-\infty}^{+\infty} \frac{1}{k_x'} \left[1 + \Gamma^{\mathrm{TE}}\left(k_z'\right)\right] e^{-jk_x' x} e^{-jk_z' z} dk_z' \tag{7.13}$$

where $k_x' = \left(k_0^2 - k_z'^2\right)^{1/2}$ and $\Gamma^{\mathrm{TE}}\left(k_z'\right)$ is the reflection coefficient at the top of the structure for a plane wave having a wavenumber k_z' in the z-direction. The path of integration is along a path C as shown in Figure 7.13, which stays on the top sheet and detours around the branch points and the poles of the function $\Gamma^{\mathrm{TE}}\left(k_z'\right)$ on the real axis, with the poles corresponding to the wavenumbers of the surface-wave modes of the structure.

Consider a path C that lies above the positive real axis as shown in Figure 7.13, by a small amount. If a leaky-wave pole is in the slow-wave region $\beta > k_0$, the pole is not close to the path C on the Riemann surface, since it is on the opposite sheet (the path is on the top sheet while the pole is on the bottom sheet). However, if the pole lies within

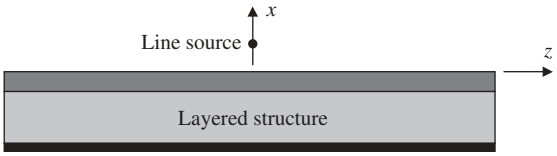


Figure 7.14 An infinite line source over a layered structure. The layered structure consists of an arbitrary number of infinite dielectric layers over a ground plane.

the fast-wave region $\beta < k_0$, the pole will be close to the path on the Riemann surface if the pole is geometrically close to the real axis (the attenuation constant of the mode is small). Therefore, a pole that lies within the fast-wave region and has a small attenuation constant is one that will have a strong influence on the path. Such a mode is said to be "physical." A leaky mode having a pole located in the fast-wave region is therefore said to be a physical mode, while a mode located in the slow-wave region is said to be "nonphysical."

To first order, a physical mode is one that can be measured in some region of space (roughly defined by the wedge-shaped leakage region shown in Figure 7.9) if one probes the field. The measured field will resemble that of the leaky mode in this region if the amplitude of excitation of the leaky mode is strong enough and the observation point is not too far away from the source (depending on the attenuation constant of the leaky mode). The larger the attenuation constant of the mode, the less sharp will be the distinction between a "physical" and a "nonphysical" mode.

From a more mathematical point of view, the use of the *steepest-descent method* of asymptotic evaluation [24] provides a convenient means to recognize the physical significance of a leaky-wave pole. To simplify the analysis it is convenient to first use the steepest-descent transformation, which is represented by the pair of equations

$$k_z' = k_0 \sin\zeta \tag{7.14}$$

$$k_x' = k_0 \cos\zeta \tag{7.15}$$

In the complex ζ plane the path is transformed into the path labeled as Γ in Figure 7.15, which maps into the real axis of the k_z plane and also maintains the correct branch for the wavenumber k_x along the path (either a positive real number or a negative imaginary number). There are no branch points in the ζ plane since $\cos\zeta$ is an analytic function. The field is represented in cylindrical coordinates as

$$E_x(\rho,\theta) = -j\omega\left(\frac{\mu_0 I_0}{4\pi j}\right)\int_\Gamma \left[1+\Gamma^{\mathrm{TE}}(k_0\sin\zeta)\right]e^{-j(k_0\rho)\cos(\zeta-\varphi)}d\zeta \tag{7.16}$$

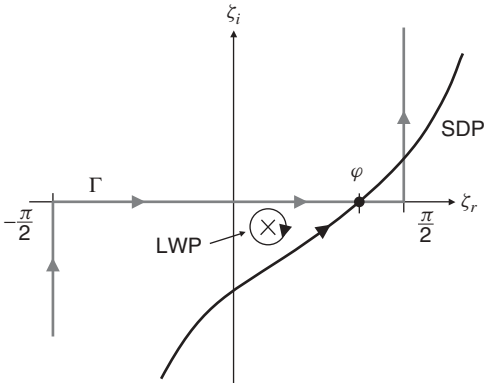


Figure 7.15 The path of integration Γ in the steepest-descent plane. Also shown is the steepest-descent path (SDP) and a typical leaky-wave pole (LWP). The leaky-wave pole that is shown here has been captured by the deformation of the original path Γ to the SDP.

The path is then deformed to the steepest-descent path (SDP) shown in Figure 7.15 that passes through the saddle point at $\zeta = \zeta_0 = \varphi$, where $\varphi = \pi/2 - \theta$ is the observation angle with respect to the x axis (with θ measured from the z-axis). The SDP is defined by the equation

$$\cos(\zeta_r - \varphi)\cosh\zeta_i = 1 \tag{7.17}$$

where $\zeta = \zeta_r + j\zeta_i$. Leaky-wave poles at $\zeta_p = \zeta_{rp} + j\zeta_{ip}$ corresponding to modes that have a positive β and α are located in the region defined by $0 < \zeta_r < \pi/2$ and $\zeta_i < 0$. If the pole lies to the left of the SDP, then the total field consists of the integration along the SDP plus the residue contribution from the pole. If the pole is to the right of the SDP, there is no residue contribution to the field.

Consider next an idealized situation where the leaky-wave pole has no attenuation ($\alpha \to 0$) and therefore lies on the real axis of the ζ plane at $\zeta_p = \zeta_{rp}$. If $\varphi > \zeta_{rp}$ the pole is captured and contributes to the field representation; otherwise it does not. Applying the steepest-descent method to asymptotically evaluate the integral along the SDP for large $k_0\rho$, the total field is (the function F denotes the total integrand)

$$E_x(\rho,\varphi) \sim \begin{cases} -2\pi j \mathrm{Res} F\left(\zeta_p\right) e^{-j(k_0\rho)\cos(\zeta_p-\varphi)} \\ \quad + \sqrt{\dfrac{2\pi}{k_0\rho}} e^{j\pi/4} F(\varphi)\, e^{-j(k_0\rho)\cos(\zeta_p-\varphi)}, \quad \varphi > \zeta_{rp} \\ \sqrt{\dfrac{2\pi}{k_0\rho}} e^{j\pi/4} F(\varphi)\, e^{-j(k_0\rho)\cos(\zeta_p-\varphi)}, \quad \varphi < \zeta_{rp} \end{cases} \tag{7.18}$$

This result shows that, in this idealized limit, the asymptotic evaluation of the field is discontinuous across the angle $\varphi = \zeta_{rp}$. For $\varphi > \zeta_{rp}$ (inside the leakage region shown in Figure 7.9) the total field contains a leaky-wave contribution, and this becomes the dominant field as $k_0\rho \to \infty$ due to the $1/\sqrt{\rho}$ decay of the "space-wave" term (the second term of (Eq. 7.18)) arising from the saddle-point contribution. For $\varphi < \zeta_{rp}$ (outside the leakage region shown in Figure 7.9) the total field does not contain a leaky-wave contribution, and the total field is equal to the space-wave field and has the usual form of a cylindrical wave, decaying as $1/\sqrt{\rho}$. As the attenuation constant increases from zero, the boundary between the region of existence of the leaky-wave field and the "shadow region" becomes less well defined (see Figure 7.10 for a typical case). This simple analysis shows, however, that to first order if a leaky-wave pole is captured it will contribute to the field, while if it is not captured it will not contribute to the field. In the idealized limit of $\alpha \to 0$ the leaky-wave field is the dominant one on the aperture (interface) at $x = 0$ as $\rho \to \infty$. For a nonzero value of α the leaky-wave field will decay exponentially with distance ρ from the source on the aperture, so that the space-wave field will always be the dominant one for sufficiently large distances away from the source. However, for a small attenuation constant the leaky-wave field may be dominant on the aperture out to a considerable distance, beyond which the field level is very small. In this case the far-field radiation pattern (which comes from a Fourier transform of the aperture field) will be dominated by that of the leaky-wave field.

As $\varphi \to \pi/2$ the observation point approaches the interface and the saddle point moves to $\zeta_0 = \varphi = \pi/2$. The SDP is now called the "extreme" SDP (ESDP). If a leaky-wave pole is not captured by the ESDP, it will never be captured for any observation angle. A useful feature of the ESDP is that it separates the fast-wave region $\beta < k_0$ (to the left of the ESDP) from the slow-wave region $\beta > k_0$ (to the right of the ESDP). Therefore

the preceding analysis shows that if a leaky-wave pole lies within the fast-wave region, it will contribute to the field representation (with the corresponding fields existing in a wedge-shaped region of space), while if it lies in the slow-wave region it will not make a significant contribution to the total field. This is consistent with the previous discussion based on the "closeness" of the pole to the path of integration C on the Riemann surface.

The above discussion in this section, although simple, serves to illustrate the basic physical features of leaky waves. For further in-depth information about the physical properties of leaky waves, the reader is referred to the comprehensive pair of papers by Tamir and Oliner [22, 23] in which the physical and mathematical properties of leaky waves are discussed in much greater detail.

In the next two sections the radiation properties of the various types of leaky-wave antennas are discussed in more detail.

7.5 RADIATION PROPERTIES OF ONE-DIMENSIONAL LEAKY-WAVE ANTENNAS

In this section the basic radiation properties of the different types of 1D leaky-wave antennas that were reviewed earlier are discussed in more detail.

7.5.1 Uniform Structures

A typical example of a uniform leaky-wave antenna is the air-filled rectangular waveguide with a longitudinal slot, as shown in Figure 7.1. The fundamental TE_{10} waveguide mode is a fast wave, with $\beta < k_0$. In particular, to a good approximation (assuming that the slot does not affect the phase constant significantly),

$$\beta = \sqrt{k_0^2 - k_c^2} \tag{7.19}$$

where $k_c = \pi/a$. The fast-wave property of the aperture distribution in the slot causes the antenna to radiate a conical beam about the z-axis at an angle given approximately by Eq. (7.8). The radiation causes the wavenumber k_z of the propagating mode within the (now open) waveguide structure to become complex, so that $k_z = \beta - j\alpha$. From the equivalence principle and image theory [26, 27], the radiation from this structure in the region $x > 0$ is approximately that due to a z-directed magnetic line current flowing along the z-axis of the form

$$K(z) = A \exp(-jk_z z) \tag{7.20}$$

radiating in free space (where A is a constant). For a semi-infinite aperture that starts at $z = 0$, the line source will exist for $0 < z < \infty$. In the far field the electric field will be polarized in the ϕ direction and this component is given in terms of the Fourier transform $\tilde{K}(k_z)$ of the line source as [26, 27]

$$E_\phi = -jk_0 \sin\theta \psi(r) \tilde{K}(k_0 \cos\theta) \tag{7.21}$$

where

$$\psi(r) = \frac{e^{-jk_0 r}}{4\pi r} \tag{7.22}$$

and θ and ϕ denote the usual spherical coordinate angles with respect to the (x, y, z) axes. Taking the transform of the current in Eq. (7.20), the result is

$$E_\phi(r,\theta) = A\psi(r)\left(\frac{\sin\theta}{\cos\theta - k_z/k_0}\right) \tag{7.23}$$

The magnitude of the pattern is

$$\left|E_\phi(r,\theta)\right| = |A|\,|\psi(r)|\left(\frac{\sin^2\theta}{(\cos\theta - \beta/k_0)^2 + (\alpha/k_0)^2}\right)^{1/2} \tag{7.24}$$

The beam is a conical one, with a maximum at an angle $0 < \theta_0 < \pi/2$ that is given approximately by Eq. (7.8). The attenuation constant controls the beamwidth of the pattern. An approximate formula for the beamwidth, measured between half-power points, is

$$BW = 2\csc(\theta_0)(\alpha/k_0) \tag{7.25}$$

As is typical for a uniform leaky-wave antenna (LWA), the beam cannot be scanned too close to broadside ($\theta_0 = \pi/2$), since this corresponds to the cutoff frequency of the waveguide. (As mentioned previously, a broadside beam may be obtained by feeding the waveguide in the center or from both ends, and thereby creating two beams that point near to broadside but on opposite sides of it, which merge together. For a structure excited in the center, an optimum broadside beam occurs when the phase and attenuation constants of the leaky wave are equal, which will occur close to the cutoff of the waveguide. In this case, the above beamwidth formula is modified by the addition of an extra factor of $\sqrt{2}$. See also Eqs. (7.57) and (7.58), given later.) Also, the beam cannot be scanned too close to endfire ($\theta_0 = 0^\circ$) for an air-filled waveguide since this requires operation at frequencies significantly above cutoff, where higher-order modes can propagate. The $\csc\theta_0$ term in Eq. (7.25) also limits the endfire scanning. (This term arises from the element pattern of a z-directed magnetic dipole, which cannot radiate at endfire.)

H-plane patterns (in the x-z plane) for the case $\beta/k_0 = 0.7071$ and $\alpha/k_0 = 0.1$ and 0.01 are shown in Figure 7.16. This particular value of β corresponds to a beam angle

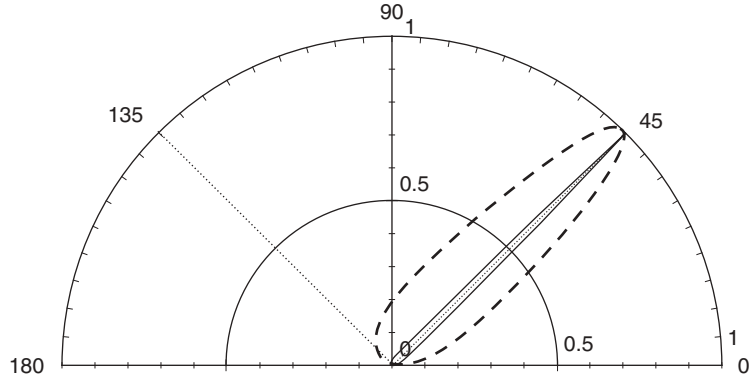


Figure 7.16 H-plane (x-z plane) patterns for a leaky mode having $\beta/k_0 = 0.7071$ and two different values of α/k_0: 0.1 (dashed line) and 0.01 (solid line).

of 45°. It is seen that, in accordance with Eq. (7.25), the pattern corresponding to the much smaller α value has a much smaller beamwidth. It is also observed in Figure 7.16 that the pattern of the semi-infinite aperture has no sidelobes.

If the aperture has a finite length L, extending from $z = 0$ to $z = L$, and an ideal absorbing load is placed at the end of the aperture to avoid a reflected wave, the pattern becomes

$$E_\phi(r, \theta) = A\psi(r)\left(\frac{\sin\theta}{\cos\theta - k_z/k_0}\right)\left(1 - e^{-jk_zL}e^{jk_0L\cos\theta}\right) \tag{7.26}$$

The magnitude of the pattern is then

$$\left|E_\phi(r, \theta)\right| = |A|\,|\psi(r)|\left(\frac{\sin^2\theta}{(\cos\theta - \beta/k_0)^2 + (\alpha/k_0)^2}\right)^{1/2}\left|1 - e^{-\alpha L}e^{jk_0L(\cos\theta - \beta/k_0)}\right| \tag{7.27}$$

The last term in Eq. (7.27) results in the presence of sidelobes [4]. The sidelobe level depends on the length of the aperture. As the aperture becomes very long so that $\alpha L \to \infty$, the sidelobes disappear. In the other extreme, as $\alpha L \to 0$, the aperture assumes a uniform illumination and the sidelobe level approaches approximately -13.5 dB.

The radiation efficiency e_r (power radiated into space divided by total power into the antenna) is less than 100% due to the power absorbed at the end of the structure (where the load is) and is related to the attenuation constant as

$$e_r = 1 - e^{-2\alpha L} \tag{7.28}$$

(This equation ignores dissipative losses in the structure.) One interesting feature can be established for uniform leaky-wave antennas based on air-filled waveguides (so that Eq. (7.19) holds). If we assume that the aperture length L is fixed, and assume that the attenuation constant is chosen so that αL, and therefore the radiation efficiency, is also fixed (i.e., it has the same value regardless of the scan angle the antenna is designed for), then the beamwidth of the pattern is independent of the scan angle θ_0 for which the antenna is designed [4]. (However, for a given structure, the beamwidth will typically change as the beam scans with frequency, since the value of α typically changes with frequency.)

A structure with a large aperture length has both a high radiation efficiency and small sidelobes, but the aperture efficiency is very poor in that the directivity will be much lower than a uniformly illuminated aperture of the same length. Therefore, in practice, the aperture is usually tapered so that the attenuation constant α is a function of z, in order to realize a specified aperture illumination (e.g., a Taylor distribution) that has a given sidelobe level [28].

Suppose that it is desired to achieve an amplitude taper $A(z) = |K(z)|$ in the line source amplitude $K(z)$, while maintaining a value of the phase constant β that does not change with z. The constraint on β is necessary to assure that all parts of the aperture will radiate at the same angle. Approximately, the power radiated (leaked) per unit length $P_l(z)$ is proportional to $A(z)^2$. The attenuation constant is related to $P_l(z)$ and to the power $P(z)$ flowing down the waveguide as [1, 4]

$$\alpha(z) = \frac{P_l(z)}{2P(z)} = -\frac{1}{2P(z)}\frac{dP(z)}{dz} \tag{7.29}$$

Consider a finite length of radiating aperture, extending from $z = 0$ to $z = L$, with a terminating load at $z = L$ that absorbs all remaining power. After some manipulations, the formula for $\alpha(z)$ can be cast into a form involving the desired aperture function $A(z)$. The derivation appears in several places; for example, it is given in Refs. 1 and 4. In terms of the radiation efficiency e_r, the result may be phrased as

$$\alpha(z) = \frac{\frac{1}{2}A^2(z)}{\frac{1}{e_r}\int_0^L A^2(z)\,dz - \int_0^z A^2(z)\,dz} \tag{7.30}$$

A typical design would call for a 90% radiation efficiency ($e_r = 0.9$). If the radiation efficiency is chosen to be too close to unity, the denominator in Eq. (7.30) becomes very small as z approaches L (near the load end of the structure), and the value of α becomes impractically large there. In a practical design, the loading would typically also be tapered to zero at the input (or feed) end to ensure that there is a gradual transition from the nonleaky to the leaky section of the waveguide.

If the leaky-wave antenna has a significant width in the y-direction, the beam that will be produced may be a fan beam or even a pencil beam (if the width becomes comparable to the length). The antenna shown in Figure 7.3 is an example of this [20]. The beam is scannable in elevation (the x-z plane) within the forward quadrant, by changing the frequency.

7.5.2 Periodic Structures

This type of leaky-wave antenna consists of a slow-wave structure that has been modified by periodically modulating the structure in some fashion. A typical example is a rectangular waveguide that is loaded with a dielectric material (so that the TE_{10} mode is a slow wave) and then modulated with a periodic set of holes or slots, as shown in Figure 7.2. Many of the features common to 1D periodic leaky-wave antennas may be appreciated by consideration of this simple (but practical) structure.

It is assumed here that the relative permittivity of the filling material is sufficiently high so that the TE_{10} mode is a slow wave over the frequency region of interest. This will be the case provided

$$\varepsilon_r > 1 + \left(\frac{\pi}{k_0 a}\right)^2 \tag{7.31}$$

over the frequency range of interest. The fundamental waveguide mode is thus a nonradiating wave. However, because of the periodicity, the modal field of the periodically loaded waveguide is now in the form of a Floquet-wave expansion [2, 3, 29] and may be written as

$$\mathbf{E}(x, y, z) = \sum_{n=-\infty}^{\infty} \mathbf{A}_n(x, y)\, e^{-jk_{zn}z} \tag{7.32}$$

where

$$k_{zn} = k_{z0} + \frac{2\pi n}{p} \tag{7.33}$$

is the wavenumber of the nth Floquet mode (or space harmonic), and p is the period. The wavenumber of the zero Floquet wave, $k_{z0} = \beta_0 - j\alpha$, is usually defined to be

the wavenumber of the Floquet wave that approaches the wavenumber of the closed waveguide when the loading (e.g., the hole size in Figure 7.2) tends to zero. It is then customary to denote $\beta = \beta_0$.

Leakage (radiation per unit length of the structure) will occur provided one of the space harmonics (usually the $n = -1$ space harmonic) is a fast wave, so that $-k_0 < \beta_{-1} < k_0$, where $\beta_{-1} = \beta_0 - 2\pi/p$. By choosing the period p appropriately, the beam can be aimed from backward endfire to forward endfire. The beam will automatically scan as the frequency changes, and the structure can be designed to scan from backward endfire to forward endfire. If one wishes to have single-beam scanning over the entire range, the $n = -2$ space harmonic must remain a slow backward wave ($\beta_{-2} < -k_0$) while the fundamental space harmonic must remain a slow forward wave ($\beta_0 > k_0$) as the -1 space harmonic is scanned from backward to forward endfire. This requires the constraint that $p/\lambda_0 < \frac{1}{2}$ at the highest (forward endfire) frequency of the scan range, and also results in the design constraint that [30]

$$\varepsilon_r > 9 + \left(\frac{p}{a}\right)^2 \tag{7.34}$$

where a is the larger waveguide dimension. Note that single-beam operation from backward to forward endfire requires a relative permittivity that is always greater than 9.

If a periodic leaky-wave antenna uses a waveguiding structure that is based on a quasi-TEM mode instead of the fundamental rectangular waveguide mode, the constraint is different. Assuming that the effective relative permittivity of the TEM mode is $\varepsilon_r^{\text{eff}}$, then the constraint is

$$\varepsilon_r^{\text{eff}} > 9 \tag{7.35}$$

Again, it is required that $p/\lambda_0 < \frac{1}{2}$ at the highest (forward endfire) frequency.

One difficulty encountered in the scanning of periodic leaky-wave antennas is that the beam shape degrades as the beam is scanned through broadside. This is because the broadside point $\beta_{-1} = 0$ corresponds to $\beta_0 p = 2\pi$. This is called an "open stopband" of the periodic structure, because it corresponds to a stopband on an open structure, where one of the space harmonics ($n = -1$) is radiating. At the open stopband point all reflections from the radiating hole discontinuities in Figure 7.2 add in phase back to the source [2]. At this point a perfect standing wave is set up within each unit cell of the structure, and the attenuation constant drops to zero.

To understand the open stopband physically, consider the equivalent circuit of this periodic waveguide structure, which is shown in Figure 7.17. The transmission line represents the equivalent circuit for the fundamental TE_{10} mode of the waveguide, chosen so that the phase constant β_{TL} of the transmission line is equal to the wavenumber of the TE_{10} mode of the waveguide. The shunt impedances Z_L represent the active impedances

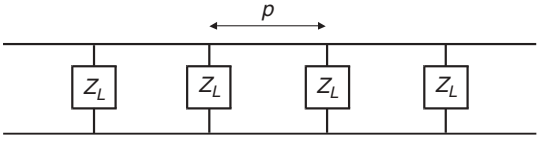


Figure 7.17 A simple approximate transmission-line model for the periodic leaky-wave antenna in Figure 7.2. The impedance Z_L represents the active impedance of each radiating hole (aperture).

of the holes (the active impedance is the impedance of a hole when radiating in the periodic environment). When the broadside point is reached, all of the holes are excited in phase, so that $\beta_{-1} = 0$, or equivalently $\beta_0 p = 2\pi$. This corresponds to $\beta_{\mathrm{TL}} p = 2\pi$. This may be established by realizing that for $\beta_{\mathrm{TL}} p = 2\pi$ all of the loads are a half guide-wavelength apart on the transmission line, and therefore the admittances $Y_L = 1/Z_L$ all add together in phase. The result is a short-circuit condition at the location of each impedance. The field within the unit cell between adjacent impedances thus becomes a perfect standing wave, and not a traveling wave. There is no radiation at this point, since the voltage at each impedance drops to zero. The attenuation constant of the leaky mode drops to zero when this point is reached. For an infinitely long structure, the input impedance would be purely reactive, and hence no power could be delivered to the antenna, since the incident power from the source would all be reflected back. In practice, for a finite-length structure, some power would reach the load but the input match and VSWR would degrade as the length of the structure increases.

The topic of eliminating or at least reducing the open stopband effect is a very important one, since this is the main limitation for achieving a continuous scanning from the backward to the forward quadrants with a single antenna. One simple method that minimizes the open stopband effect is to introduce two radiating elements per unit cell, spaced a distance $p/4$ apart within each cell [31, 32]. At the open stopband point where $\beta_0 p = 2\pi$, the electrical distance between the adjacent elements within the unit cell will be $\pi/2$. The round-trip phase delay between the two elements will then be 180°, which tends to minimize the effects of the reflection from the pair of elements. As shown in Ref. 32, this technique greatly reduces, but does not completely eliminate, the open stopband effect. However, recent work has shown that by adjusting the parameters of the structure, and optimizing the distance between the elements in the pair, it may be possible to almost eliminate the open stopband completely [33].

Other recent work has been directed at overcoming the open stopband problem by using a novel "composite right/left-handed (CRLH) metamaterial" leaky-wave antenna. This structure is based on a microstrip line, with series capacitance and parallel inductance per unit length added to make an artificial transmission line [34, 35]. It was first shown in [34] that such a structure may be used to achieve a continuous scanning from the backward to the forward region, passing through broadside. One interesting feature of this design is that the fundamental quasi-TEM mode of the microstrip line becomes a fast wave due to the loading, so that radiation occurs from the fundamental $n = 0$ space harmonic. The $n = -1$ space harmonic is a slow wave (nonradiating wave) for this design. If the period is small enough, the structure acts as a quasi-uniform 1D structure, which supports a fast wave. The microstrip line by itself has a natural "right-handed" series inductance and parallel capacitance per unit length denoted as L_{R} and C_{R}. The added series capacitance and parallel inductance per unit length are termed the "left-handed" values and are denoted as C_{L} and L_{L}. The open stopband is greatly minimized when a "balanced condition" [36, 37] is satisfied, namely,

$$\sqrt{\frac{L_{\mathrm{L}}}{C_{\mathrm{L}}}} = \sqrt{\frac{L_{\mathrm{R}}}{C_{\mathrm{R}}}} \tag{7.36}$$

This condition is approximate, since it is based on transmission-line theory, which cannot account for all full-wave effects. Therefore, when a design is based on Eq. (7.36), there will typically be a small residual stopband effect that remains [38]. However, it has

recently been shown that it is possible to optimize the structure using a full-wave solution to obtain a complete elimination of the open stopband [39].

Other designs have recently been proposed to eliminate the open stopband. For example, the design shown in Figure 7.18 has been studied, which consists of a microstrip line that is loaded with series capacitors and parallel U-shaped stubs that are connected across the capacitive gaps [40]. A full-wave analysis of this structure reveals that when it is optimized there is no open stopband (the attenuation constant does not drop sharply as the beam is scanned through broadside), and the beam may be scanned through broadside with an almost constant beamwidth [40]. A result is given in Figure 7.19, which shows that the beam shape remains almost constant as the beam scans through broadside. The characteristic impedance of the structure, viewed as an artificial transmission line, remains almost constant as the beam scans through broadside. Another design that eliminates the open stopband uses simple radiating elements (microstrip stubs) but employs impedance transformers within the unit cell [41]. Using an optimized design, the characteristic impedance remains almost perfectly constant as the beam is scanned through broadside. The beamwidth also remains nearly constant as the beam is scanned through broadside.

When designing, analyzing, and interpreting periodic leaky-wave antennas, a useful tool is the Brillouin diagram [2, 10]. This is a plot of $k_0 p$ versus $\beta_n p$, plotted on the type of diagram that is shown in Figure 7.20. Note that on this diagram each space harmonic n will have a plot that is shifted from the adjacent ones, $n-1$ and $n+1$, by 2π. The thick solid lines on the diagram indicate boundaries where a space harmonic (usually the $n=-1$ space harmonic for a periodic leaky-wave antenna) will be radiating at backward endfire and forward endfire. Between these two lines is the radiation region, where the space harmonic will be a fast wave, and hence a radiating wave. The shaded regions (the regions inside the lower triangles) are the bound-wave (nonradiating) regions [2, 10, 42]. If any space harmonic lies within one of these triangles, all of the space harmonics will be within these triangles, and hence they will all be slow (nonradiating) waves. The overall mode is then a bound mode that does not radiate. For a point outside the bound-wave triangles, there must be at least one space harmonic that is a radiating fast wave, and hence the overall mode is a leaky mode. One immediate consequence of the Brillouin diagram is that if the frequency is high enough so that $k_0 p > \pi$, the mode must

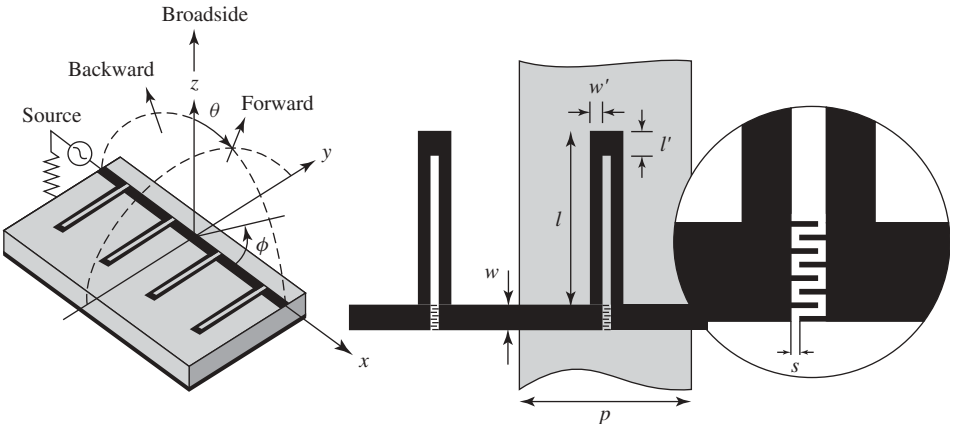


Figure 7.18 Geometry of a novel periodic leaky-wave antenna that has no stopband at broadside.

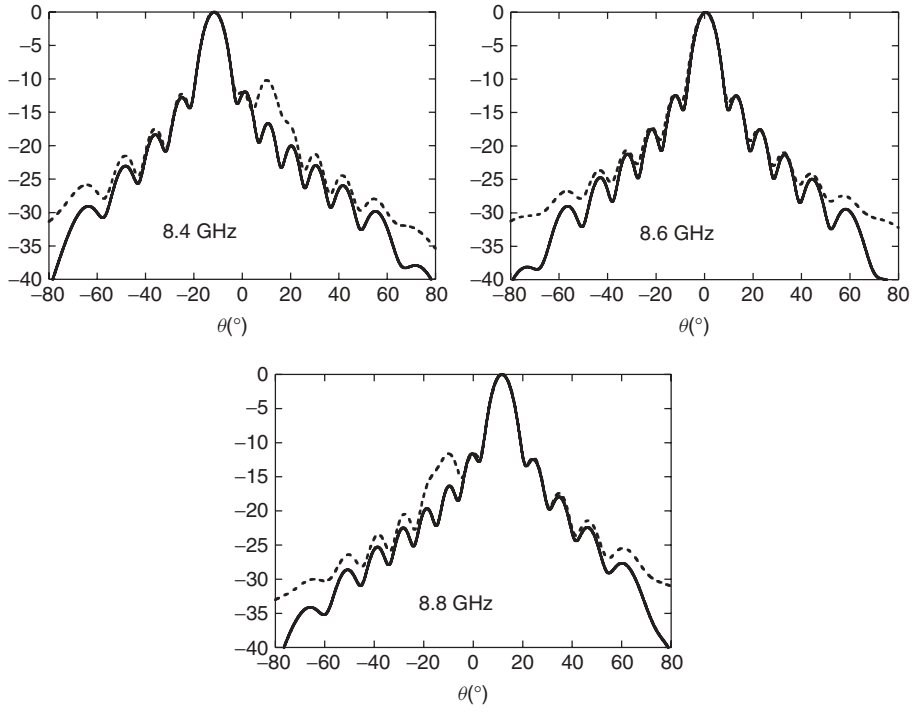


Figure 7.19 Results for the structure shown in Figure 7.18, showing the shape of the radiation patterns in the x-z plane as the beam scans through broadside. The solid curve shows results predicted from a leaky-wave model that employs a full-wave simulation, while the dashed curve is obtained from simulating a large finite-size structure (48 unit cells) that is excited with a voltage source at one end, using a moment-method simulator (Ansoft Designer).

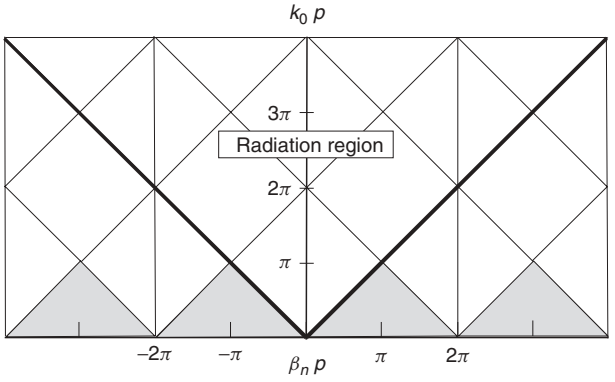


Figure 7.20 The Brillouin diagram, which is very useful for interpreting the properties of a periodic leaky-wave antenna. This is a plot of $k_0 p$ versus $\beta_n p$. The plot for each different value of n is shifted by 2π from the plot for the adjacent value of n. The region between the dark solid lines is the radiation region, where a space harmonic will leak into free space. The shaded region inside the triangles is the bound region, where the solution is purely bound, and none of the space harmonics leak.

be a leaky mode. The Brillouin diagram shown in Figure 7.20 applies to any periodic structure that can radiate into free space, including a quasi-uniform structure such as the holey waveguide of Figure 7.2 (which uses an air-filled waveguide and closely-spaced holes) or a periodic leaky-wave antenna, such as the one shown in Figure 7.2 where the waveguide is filled with dielectric and the holes are further apart (so that radiation occurs from the $n = -1$ space harmonic). Another class of periodic leaky-wave antennas consists of *printed* periodic structures, where a periodic metallization is etched on a grounded dielectric substrate. For this class of structures, leakage may occur into space and also into the fundamental TM_0 surface-wave mode of the grounded substrate [43]. For these structures a "generalized" Brillouin diagram may be constructed [44], which shows separate regions where there is either no leakage, leakage into only the surface wave, or leakage into both the surface wave and into space. (It is not physical to have leakage into only space and not into the surface wave for this class of structures, since the surface wave is always a slow wave with respect to free space.) These different leakage mechanisms are not considered further here, however.

A sketch showing typical behavior for the fundamental ($n = 0$) space harmonic is given in Figure 7.21. The open stopband occurs at $\beta_{-1} = 0$, so that $\beta_0 p = 2\pi$. Near this point the plot exhibits a sharp transition, becoming nearly vertical. The attenuation constant (not shown) varies rapidly near the open stopband region and drops to zero exactly at the point $\beta_0 p = 2\pi$, which is the "stopband null point."

When the $n = -1$ space harmonic lies within the radiation region it is a proper wave when $\beta_{-1} < 0$, and an improper wave when $\beta_{-1} > 0$, in accordance with the earlier discussion regarding the proper/improper nature of radiating waves.

The point $\beta_0 p = \pi$ is also a stopband, but at this point the solution usually lies within the bound-mode triangle (if the fundamental space harmonic is a slow wave), and hence the overall mode is a nonleaky mode with the fields of all space harmonics decaying transversely away from the structure. The behavior of this stopband is thus the same as if the structure were closed (perfectly shielded), since no radiation takes place. Hence this region is a "closed stopband." For a lossless structure, the behavior at the closed stopband is like that of a closed structure, so that throughout the stopband frequency region the value of the phase constant remains exactly constant at $\beta_0 p = \pi$, while $\alpha \neq 0$ in this region. The attenuation in this region, however, corresponds not to leakage radiation

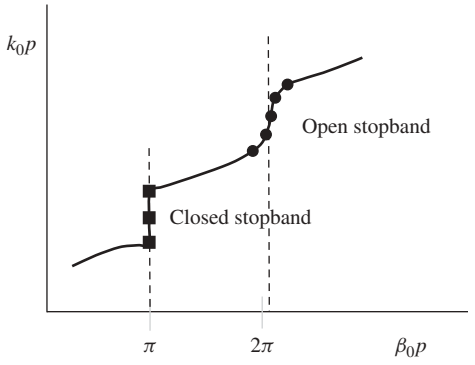


Figure 7.21 A sketch showing typical behavior for the plot of the fundamental ($n = 0$) space harmonic on the Brillouin diagram. The closed stopband occurs exactly at $\beta_0 p = \pi$ while the open stopband occurs in the vicinity of $\beta_0 p = 2\pi$.

but to reactive evanescent decay of the fields (similar to a waveguide mode that is operating below cutoff). In the open stopband region, there is already some leakage loss present when the mode enters the stopband region. Consequently, in contrast to the closed stopband, the phase constant is not perfectly constant in the open stopband region, and the fields are not completely reactive, but largely so. However, as mentioned earlier, the attenuation constant drops exactly to zero at the stopband null point, where $\beta_0 p = 2\pi$.

As a consequence of the open stopband, the attenuation constant, and hence the beam shape, will change dramatically in the open stopband region, making scanning through broadside very difficult. However, as also mentioned earlier, novel designs have been introduced recently to overcome this problem. For example, for the structure shown in Figure 7.18, a plot of the wavenumber and attenuation constant near the open stopband region (not shown here) reveals that a continuous scan through broadside is possible with little or no noticeable stopband effects. This is consistent with the continuous scanning behavior shown in Figure 7.19.

7.6 RADIATION PROPERTIES OF TWO-DIMENSIONAL LEAKY-WAVE ANTENNAS

This class of leaky-wave antennas is based on planar technology, and it offers a simple way to obtain a highly directive beam with a small source such as a dipole or waveguide-fed slot. The beam may be a pencil beam at broadside, or a conical beam pointing at an angle θ_0 (see Fig. 7.5). This category of leaky-wave antennas was first introduced by von Trentini in 1956, although it has become much more active in recent years. These antennas bear some resemblance to the classical Fabry–Perot structure in the optics field, and in fact some of the antennas in this category have been explained in these terms. Often antennas that are in this category were not originally recognized as being leaky-wave antennas and were originally explained in other terms, such as Fabry–Perot resonances or electromagnetic bandgap (EBG) effects.

7.6.1 Introduction

The antenna structures in this section are different from the 1D antennas discussed previously in this chapter, but they are also leaky-wave antennas, and their performance obeys the fundamental relations described in the earlier sections of this chapter. These antennas are planar in nature, and most of them consist of a metal ground plane with a dielectric layer on it (which may be air) that is covered with a *partially reflective surface or screen* (PRS) on top of the dielectric layer. The partially reflective (or partially transparent) screen can take various forms, such as a stack of one or more dielectric layers, or a metal screen consisting of a periodic array of slots or metal patches or other elements, or an array of parallel wires or strips. Several examples of such structures are shown in Figure 7.22 (where an x-directed electric dipole source is also shown). Note that for the discussion of these structures the z-axis is normal to the interface, which is taken as $z = 0$.

The structure may be excited by a simple source inside the dielectric layer using, for example, a horizontal electric dipole in the middle of the layer or a magnetic dipole on the ground plane. Such sources may be realized in practice in a variety of ways, with some examples being a printed dipole inside the layer, a microstrip patch on the ground

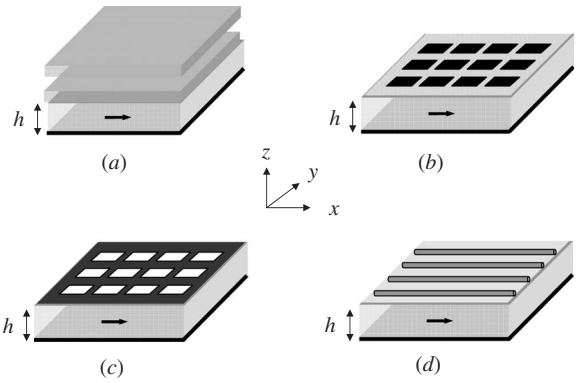


Figure 7.22 Examples of different types of partially reflecting screens (PRSs) that may be used to make a 2D leaky-wave antenna. The structures consist of a grounded substrate layer of thickness h covered by a PRS. The structures are shown excited by an x-directed horizontal electric dipole source. (a) The PRS consists of a stack of dielectric layers. (b) The PRS consists of a 2D periodic array of rectangular metal patches. (c) The PRS consists of a 2D periodic array of rectangular slots in a thin metal plate. (d) The PRS consists of an array of wires or strips.

plane, a slot in the ground plane, or a waveguide-fed aperture in the ground plane. Since the fundamental nature of the beam forming is through leaky-wave radiation, the source merely acts as a launcher for the leaky waves that propagate on the structure. Depending on the type of source, and the design of the structure, either a pencil beam at broadside or a conical beam at an angle θ_0 about the z-axis may be produced, as shown in Figure 7.5.

In addition to the PRS design, it is also possible to make a leaky-wave antenna by using metamaterials, which are (usually) periodic structures that exhibit unusual properties not typically found in nature [45–48]. In particular, a leaky-wave antenna can be realized by using a grounded metamaterial slab of low-permittivity material, with $0 < \epsilon_r \ll 1$ [49, 50]. (In Ref. 49 a 1D leaky-wave antenna with a line-source excitation is investigated, while in Ref. 50 a 2D leaky wave antenna excited by a dipole source is investigated.) Such a low-permittivity material could be realized in practice by an arrangement of parallel and horizontal conducting wires in free space (or in a host material). In this antenna design, the significant discontinuity between the high impedance of the metamaterial region and the moderate impedance of the air region above creates a large reflection at the interface, confining the guided wave mainly in the grounded slab region, which acts as a parallel-plane waveguide with a perfect conducting ground below and an imperfect (impedance) surface on top. The difference in impedances across the interface thus achieves the same wave-trapping effect as does the PRS in the designs shown in Figure 7.22. One disadvantage of the metamaterial design is the larger slab thickness that is required. For example, as discussed in the section below, the optimum slab thickness for a broadside beam is about one-half of a wavelength in the slab material [see Eq. (7.61)]. Hence, a low-permittivity slab will be physically much thicker.

Another interesting metamaterial-based design that has been recently introduced uses a 2D composite right/left-handed (CRLH) metamaterial structure to obtain conical beams [51, 52]. In this design a 2D version of the CRLH discussed in Section 7.5 is used to create a planar CRLH "metasurface." This surface may consist of a planar 2D periodic arrangement of conducting square patches on a grounded substrate (which have a naturally

occurring series capacitance between them) loaded with parallel inductive vias at the centers of the patches [51, 52]. Other configurations may also be used [51, 52].

Another method for using leaky wave antennas to produce a pencil beam, either at broadside or at any set of angles (θ, ϕ), is to use a 1D phased array, with each element of the phased array being a 1D periodic leaky-wave antenna. Each leaky-wave antenna may be one based on a "conventional" design (where radiation occurs through the $n = -1$ space harmonic), or a metamaterial design, such as the CRLH design, where radiation occurs through the $n = 0$ space harmonic. A discussion of the former type may be found in [4], while a discussion of the latter type may be found in [52, 53]. A phase shift between the leaky-wave antennas controls the beam angle in one plane, while the frequency controls the beam angle in the other plane. An electronic control may also be used to steer the beam of the leaky-wave antennas, but this requires an integration of electronically tunable elements into the leaky-wave antennas [53].

7.6.2 Basic Principles of Operation

One of the main differences between this type of structure and the ones considered previously is that the leaky wave on this structure is a *two-dimensional* (2D) cylindrical wave, which propagates outward *radially* from the source along the interface [54]. The leaky wave then furnishes a large 2D aperture that in turn produces the narrow radiated beam. As is true for all leaky-wave antennas, the beam angle is frequency sensitive.

A vertical electric or magnetic dipole source may be used to produce a conical beam, but not a broadside beam (the pattern will always have a null at broadside). This type of source launches only a TM_z or TE_z leaky wave, respectively, which has no ϕ variation. (A $\mathrm{TM}_z/\mathrm{TE}_z$ wave has only an electric/magnetic field component in the z-direction. A TM_z wave is described by the magnetic vector potential component A_z, while a TE_z wave is described by an electric vector potential component F_z [26].) The magnetic vector potential A_z or the electric vector potential F_z in the region above the aperture $(z = 0)$ has the respective form

$$A_z(\rho, z) = \tfrac{1}{2} H_0^{(2)}\left(k_\rho \rho\right) e^{-jk_z z} \tag{7.37}$$

or

$$F_z(\rho, z) = \frac{1}{2}\left(\frac{\omega\mu_0}{k_z}\right) H_0^{(2)}\left(k_\rho \rho\right) e^{-jk_z z} \tag{7.38}$$

where $k_\rho = \beta - j\alpha$ is the complex wavenumber of the leaky wave (which represents either k_ρ^{TM} or k_ρ^{TE}) and $k_z^2 = k_0^2 - k_\rho^2$. (The factors $\frac{1}{2}$ and $\frac{1}{2}\omega\mu_0/k_z$ have been added for convenience.) The Hankel function has order $n = 0$ due to the azimuthal symmetry.

The corresponding aperture fields from the electric or magnetic dipole source result in an omnidirectional (in azimuth) conical beam [54] that is polarized with the electric field in either the θ or ϕ direction, respectively. In either case the beam angle is given approximately by

$$\sin\theta_0 = \beta/k_0 \tag{7.39}$$

(which is the same as Eq. (7.8) except for a change in the definition of the angle θ_0) while the beamwidth (between the -3-dB points) is given by

$$\Delta\theta = \frac{2\alpha/k_0}{\cos\theta_0} \tag{7.40}$$

The radiation patterns for the TM_z and TE_z cases are

$$E_\theta(r,\theta) = R(r)\,P_0(\theta) \tag{7.41}$$

and

$$E_\phi(r,\theta) = -R(r)\,P_0(\theta) \tag{7.42}$$

respectively, where

$$P_0(\theta) = -4\frac{\sin\theta}{\left(k_\rho/k_0\right)^2 - \sin^2\theta} \tag{7.43}$$

and

$$R(r) = -\frac{j\omega\mu_0}{4\pi r}e^{-jk_0 r} \tag{7.44}$$

A horizontal electric or magnetic dipole source launches a pair of leaky waves, one TM_z and one TE_z. The TM_z leaky wave determines the E-plane pattern while the TE_z leaky wave determines the H-plane pattern [54]. For a horizontal electric dipole source, the forms assumed for the magnetic and electric vector potentials above the aperture for the normalized TM_z and TE_z leaky waves are

$$A_z(\rho,\phi,z) = \tfrac{1}{2}\cos\phi H_1^{(2)}\left(k_\rho^{\mathrm{TM}}\rho\right)e^{-jk_z^{\mathrm{TM}}z} \tag{7.45}$$

and

$$F_z(\rho,\phi,z) = \frac{1}{2}\left(\frac{\omega\mu_0}{k_z^{\mathrm{TE}}}\right)\sin\phi H_1^{(2)}\left(k_\rho^{\mathrm{TE}}\rho\right)e^{-jk_z^{\mathrm{TE}}z} \tag{7.46}$$

respectively. The corresponding radiation pattern for the TM_z wave is

$$E_\theta(r,\theta,\phi) = R(r)\cos\theta\cos\phi P_1(\theta) \tag{7.47}$$

$$E_\phi(r,\theta,\phi) = -R(r)\sin\phi C(\theta) \tag{7.48}$$

while the pattern of the TE_z wave is

$$E_\phi(r,\theta,\phi) = -R(r)\sin\phi P_1(\theta) \tag{7.49}$$

$$E_\theta(r,\theta,\phi) = R(r)\cos\theta\cos\phi C(\theta) \tag{7.50}$$

where, for either case,

$$P_1(\theta) = \frac{2j}{k_\rho} - \frac{4jk_\rho}{k_\rho^2 - k_0^2\sin^2\theta} \tag{7.51}$$

and

$$C(\theta) = \frac{-2j}{k_\rho} \tag{7.52}$$

In these expressions k_ρ denotes either k_ρ^{TM} or k_ρ^{TE}. The function $C(\theta)$ is actually a constant that does not depend on θ.

In order to have a narrow pencil beam at broadside or a narrow conical beam, a necessary condition is that [54]

$$\frac{A}{k_\rho^{\mathrm{TM}}} = \frac{B}{k_\rho^{\mathrm{TE}}} \tag{7.53}$$

where A and B are the (complex) amplitudes of the TM_z and TE_z leaky waves, respectively (with respect to the normalization used in Eqs. (7.45) and (7.46)). The beamwidths in the E- and H-planes are then given by the same expression as in Eq. (7.40) for the $n = 0$ leaky wave, so that

$$\Delta\theta_E = \frac{2\alpha_{\mathrm{TM}}/k_0}{\cos\theta_0} \tag{7.54}$$

$$\Delta\theta_H = \frac{2\alpha_{\mathrm{TE}}/k_0}{\cos\theta_0} \tag{7.55}$$

If we wish to have a narrow broadside beam, the optimum condition is that [54]

$$\alpha_{\mathrm{TM}} = \beta_{\mathrm{TM}} \quad \text{and} \quad \alpha_{\mathrm{TE}} = \beta_{\mathrm{TE}} \tag{7.56}$$

The beamwidths in the E-plane ($\phi = 0^\circ$) and H-plane ($\phi = 90^\circ$) will then be [54]

$$\Delta\theta_E = 2\sqrt{2}\alpha_{\mathrm{TM}} \tag{7.57}$$

$$\Delta\theta_H = 2\sqrt{2}\alpha_{\mathrm{TE}} \tag{7.58}$$

In order to have a symmetric pencil beam at broadside (equal beamwidths in the E- and H-planes),

$$k_\rho^{\mathrm{TM}} \approx k_\rho^{\mathrm{TE}} \tag{7.59}$$

Thus for an omnidirectional pencil beam at broadside, the two leaky waves should have very nearly the same phase and attenuation constants, so that $\alpha^{\mathrm{TM}} \approx \beta^{\mathrm{TM}} \approx \alpha^{\mathrm{TE}} \approx \beta^{\mathrm{TE}}$, and also (from Eq. (7.53)), they should have approximately the same amplitudes as well.

Fortunately, when a 2D leaky-wave antenna is constructed using a partially reflective surface or screen (PRS), and the structure is optimized for a broadside beam by adjusting the substrate thickness, the conditions in Eqs. (7.53), (7.56), and (7.59) are automatically satisfied. Interestingly, this is true even if the PRS is not similar in the E- and H-plane directions. For example, the PRS may consist of a periodic array of slots in a metal plate, with the slots being long in the x-direction and narrow in the y-direction, having very different periodicities in the two directions. Even in this case, an omnidirectional pencil beam at broadside is obtained when the substrate thickness is properly chosen [18].

As the scan angle θ_0 increases from zero, the pencil beam turns into a conical beam, similar to how the petals on a flower unfold. As the scan angle increases (by increasing the frequency or the substrate thickness) the wavenumbers of the two leaky waves typically begin to differ, and this explains why the beamwidths often become different in the principal planes. (The exact nature of the beamwidth variation with scan angle in the two principal planes depends on the particular type of PRS.)

From a physical point of view, the PRS is used to create a leaky parallel-plate waveguide region, and the leaky waves are leaky (radiating) versions of the parallel-plate

waveguide modes that would be excited by the source in an ideal parallel-plate waveguide (which results if the PRS is replaced by a perfectly conducting metal plate). This point of view allows for a simple design formula for the thickness of the dielectric layer in order to obtain a beam at a desired angle θ_0 (either a broadside or a conical beam). The parallel-plate waveguide modes are described by $m = 1$, meaning that there is one half-wavelength variation vertically inside the parallel-plate waveguide. (Although larger integer values of m could be used, this would result in a design that has a thicker substrate layer.) The radial wavenumber of the TM_z and TE_z parallel-plate waveguide modes for an ideal waveguide would be

$$\beta = \beta^{\mathrm{TM}} = \beta^{\mathrm{TE}} = \sqrt{k_1^2 - \left(\frac{\pi}{h}\right)^2} \tag{7.60}$$

where h is the thickness of the substrate layer and k_1 is the wavenumber of the layer, which may also be expressed as $k_1 = k_0 n_1$, where n_1 is the refractive index of the layer. Using the simple approximate relation $\beta = k_0 \sin\theta_0$ that is valid for any leaky wave (with θ_0 measured here from the vertical z-axis), we obtain the result

$$k_0 h = \frac{\pi}{\sqrt{n_1^2 - \sin^2\theta_0}} \tag{7.61}$$

The vertical location of the source usually has little effect on the pattern shape, since this is dictated by the leaky-wave phase and attenuation constants. The phase constant is determined primarily by the thickness of the dielectric layer (see Eq. (7.60)), while the attenuation constant is determined primarily by the properties of the PRS. However, the *power density* at the peak of the beam will be maximized when a horizontal electric dipole source is placed in the middle of the substrate layer, or a horizontal magnetic dipole is placed on the ground plane. A vertical electric dipole source has a maximum peak power density when it is placed on the ground plane, while a vertical magnetic dipole source does so when it is placed in the middle of the layer. Changing the peak power density, and hence the overall power radiated by the source, directly affects the input resistance of the source.

The PRS may be either uniform (e.g., one or more dielectric layers) or periodic in one or two dimensions (e.g., an array of closely spaced wires or strips, or a two-dimensional array of slots in a metal plate). However, it is important to note that, for the periodic PRS structures discussed here, the radiation still occurs via the fundamental parallel-plate waveguide modes and not a space harmonic of these modes. That is, the PRS acts as a quasi-uniform reflective surface, where the reflection coefficients of the fundamental parallel-plate waves determine the characteristics of the antenna. The physical principle of operation is thus as a quasi-uniform leaky-wave antenna, and not a periodic leaky-wave antenna that radiates from a space harmonic.

As mentioned, the first example of this type of antenna was introduced by von Trentini [11]. A more recent and improved version of the von Trentini antenna has been developed by Feresidis and Vardaxoglou [16]. These authors employed a feed consisting of a rectangular waveguide built into the ground plane. The authors make the interesting observation that the antenna would have a greater pattern bandwidth if the phase of its PRS were to linearly increase with frequency, so as to compensate for the change in the electrical thickness of the substrate layer—see Eq. (7.61). With this aim, they investigated PRSs loaded with several different elements, such as crossed dipoles, patches,

rings, and square loops. They found that dipoles, or square or circular patches (or their complementary structures), particularly with close packing of the elements in the array, produced less of a variation of the beam with frequency. This slower variation was not found for crossed dipoles, square loops, and rings, even for close packing of these elements. They therefore chose to use arrays of closely spaced dipoles in their PRS structure.

Another method for increasing the pattern bandwidth is to design the PRS using multiple layers of periodic elements [55], so that each layer becomes the main PRS at different parts of the frequency band, thereby creating a structure in which the effective substrate thickness varies with frequency.

7.6.3 General Design Formulas

The far-field pattern of a PRS leaky-wave antenna structure may be calculated by reciprocity, in which the far field is determined by illuminating the structure with an incident plane wave and calculating the field at the source dipole location [56]. In the following discussion the source dipole is assumed to be a horizontal electric dipole in the x-direction. The plane-wave calculation may be carried out by using a simple transverse equivalent network (TEN) model, which is a transmission-line model that represents the vertical field behavior in the plane-wave problem. If the PRS is assumed to be lossless and infinitesimally thin in the vertical direction, it may be represented as a shunt susceptance B_s in the TEN model. Based on this simple model, an accurate closed-form calculation of the far-field pattern may be obtained, and, from this, approximate closed-form expressions for the properties of the antenna may be obtained.

One formula of importance is that for the optimum thickness of the substrate layer h for procuring a narrow beam at angle θ_0 (either a pencil beam at broadside or a conical beam), which is [56]

$$h = h^{\text{pp}} \left(1 + \frac{\overline{Y}_1}{\pi \overline{B}_s} \right) \tag{7.62}$$

where h^{pp} is the optimum substrate thickness based on the ideal closed parallel-plate waveguide model (given by Eq. (7.61)), and $\overline{Y}_1 = Y_1 \eta_0$ is the normalized characteristic admittance of the transmission line that models the substrate region in the TEN model. The term η_0 is the intrinsic impedance of free space. The admittance $\overline{Y}_1$ is taken as either $\overline{Y}_1^{\text{TM}}$ or $\overline{Y}_1^{\text{TE}}$ for optimizing the beam angle in either the E- or H-plane, respectively, where

$$\overline{Y}_1^{\text{TM}} = \frac{\varepsilon_r}{\sqrt{n_1^2 - \sin^2 \theta_0}} \tag{7.63}$$

$$\overline{Y}_1^{\text{TE}} = \frac{1}{\mu_r} \sqrt{n_1^2 - \sin^2 \theta_0} \tag{7.64}$$

and $n_1 = \sqrt{\varepsilon_r \mu_r}$ is the index of refraction of the substrate layer. The value of $\overline{B}_s = B_s \eta_0$ is calculated at the beam angle θ_0 in either the E- or H-plane.

Note that $\overline{Y}_1^{\text{TM}} = \overline{Y}_1^{\text{TE}}$ for a broadside beam, but not a conical beam ($\theta_0 > 0$). Furthermore, $\overline{B}_s$ has the same value at broadside ($\theta_0 = 0$) when calculating E_θ in the E-plane

($\phi = 0^\circ$) or E_ϕ in the H-plane ($\phi = 90^\circ$), but it will in general be different in the E- and H-planes for a beam angle $\theta_0 > 0$. Therefore a unique substrate thickness exists that optimizes the structure for a broadside beam. However, for a conical beam, it may not be possible to optimize the structure in both the E- and H-planes simultaneously. The result is that the beam angle θ_0 will actually be slightly different for the E- and H-planes, and this will limit the useful range of scan angles. The beam angle is also limited by the fact that the beamwidths in the two principal planes usually become different as the beam angle increases, which is discussed later. An interesting exception to this rule is seen when using a PRS in the form of a metal strip grating, and this is discussed in more detail later.

Formulas may also be derived for the *beamwidth* and *pattern bandwidth* in the E- and H-planes, in terms of B_s [56]. It is important to keep in mind that the value of B_s will be a function of the beam angle θ_0, even though the physical PRS structure is fixed. However, the manner in which B_s changes with beam angle is dependent on the particular type of PRS.

Table 7.1 shows the beamwidth (defined from the -3-dB points) in the E- and H-planes for three separate cases: (1) a pencil beam at broadside ($\theta_0 = 0^\circ$), (2) a conical beam with a general beam angle $0 < \theta_0 < 90^\circ$, and (3) a conical beam at endfire ($\theta_0 = 90^\circ$). The formulas are expressed in terms of the normalized shunt susceptance $\overline{B}_s = B_s \eta_0$. It is seen that the beamwidths are equal at broadside, which implies that the beam is nearly omnidirectional. However, as the beam angle increases, the H-plane pattern becomes narrower while the E-plane pattern becomes broader (for the same value of B_s). A narrow beam can be obtained at the horizon in the H-plane (at least in theory) but not in the E-plane.

The beamwidth is inversely proportional to B_s^2 for a conical beam ($\theta_0 > 0$), but inversely proportional to B_s for a broadside beam. This means that the PRS must be much closer to a perfectly reflecting surface in order to obtain a very narrow beam at broadside, compared to what is necessary for a conical beam.

Table 7.2 shows the pattern bandwidth in the E- and H-planes for the same three cases as in Table 7.1. The pattern bandwidth is defined as $(f_2 - f_1)/f_0$, where f_0 is the design frequency, for which a maximum power density is radiated at angle θ_0, and the frequencies f_1 and f_2 are the two frequencies on either side of the design frequency for which the power density radiated at the angle θ_0 is down by -3 dB from the maximum

TABLE 7.1 Expressions for Beamwidth

	E-Plane	H-Plane
Broadside	$2\sqrt{\dfrac{2n_1^3}{\pi \overline{B}_s^2}}$	$2\sqrt{\dfrac{2n_1^3}{\pi \overline{B}_s^2}}$
General beam angle	$\dfrac{2n_1^2\sqrt{n_1^2 - \sin^2\theta_0}}{\pi \overline{B}_s^2 \sin\theta_0 \cos^2\theta_0}$	$\dfrac{2\left(\sqrt{n_1^2 - \sin^2\theta_0}\right)^3}{\pi \overline{B}_s^2 \sin\theta_0}$
Endfire	Narrow beam not possible	$\dfrac{2\left(\sqrt{n_1^2 - 1}\right)^3}{\pi \overline{B}_s^2}$

TABLE 7.2 Expressions for Bandwidth

	E-Plane	H-Plane
Broadside	$\dfrac{2n_1}{\pi \overline{B}_s^2}$	$\dfrac{2n_1}{\pi \overline{B}_s^2}$
General beam angle	$\dfrac{2n_1^2 \sec\theta_0}{\pi \overline{B}_s^2 \sqrt{n_1^2 - \sin^2\theta_0}}$	$\dfrac{2\sqrt{n_1^2 - \sin^2\theta_0}}{\pi \overline{B}_s^2 \sec\theta_0}$
Endfire	Narrow beam not possible	0

value obtained at the frequency f_0. Table 7.2 shows that the pattern bandwidth is inversely proportional to B_s^2 for both a broadside beam and a conical beam. This means that for the same beamwidth (which requires a much larger value of B_s in the broadside case), a broadside pattern will have a much smaller pattern bandwidth than a conical-beam pattern will.

In the following sections a more detailed discussion is presented of the radiation properties for different types of PRS structures.

7.6.4 Dielectric-Layer PRS Structures

Printed-circuit antennas, such as microstrip dipoles and patches, have many advantages but they have the major disadvantage of low gain. During 1984 and 1985, Alexopoulos and Jackson published a pair of papers [12, 13] that showed that the gain could be enhanced significantly by placing a dielectric superstrate, or cover layer, over the original dielectric substrate layer. By choosing the layer thicknesses and the dielectric constants appropriately, a large gain can be obtained for radiation at any desired angle, in the form of a pencil beam at broadside or a conical beam pointing at an angle θ_0. The first of these two papers showed that a properly designed two-layer (substrate/superstrate) structure could produce such beams. The later paper [13] by Jackson and Alexopoulos then examined in greater detail the radiation properties of the two-layer structure. Neither paper recognized the gain enhancement effect as due to leaky modes, however. The substrate/superstrate geometry of the structure, identifying the various parameters, is shown in Figure 7.6, where the bottom layer of thickness b acts as the substrate layer ($h = b$ in Eq. (7.61)). A horizontal electric dipole source is shown located in the substrate.

The superstrate layer will act as an optimum reflecting surface when the thickness is chosen so that it is an odd multiple of one-quarter of a wavelength in the vertical direction. This corresponds to the condition

$$t = \frac{(2m-1)\,\lambda_0}{4\sqrt{n_2^2 - \sin^2\theta_0}} \tag{7.65}$$

where m is a positive integer and n_2 is the refractive index of the superstrate. Using $m = 1$ gives the thinnest superstrate.

An important step was to recognize that the two-layer structure can support *leaky waves*, and that the structure can be analyzed as a *leaky-wave antenna*. A detailed analysis appeared in 1988 in a paper by Jackson and Oliner [14], in which it was established that

the directive beams obtained in the two-layer substrate/superstrate structure are due to the excitation of TM_z and TE_z leaky modes by the horizontal dipole within the substrate layer. The leaky modes travel radially outward from the source and are supported by the two-layer structure. In particular, it was verified that a TM_z leaky mode determines the E-plane pattern while a TE_z leaky mode determines the H-plane pattern, in accordance with the discussion given earlier in this chapter. The trajectory of these two leaky modes in the steepest-descent plane illuminated very clearly how the radiated beam, due to the combination of these two leaky modes, behaves as a function of frequency. Knowledge of the leaky modes is not necessary to calculate the far-field pattern of the infinite structure (which, as mentioned previously, is conveniently done by using reciprocity). However, the leaky-mode approach, which yields the phase and attenuation constants of the leaky waves, directly indicates how large the ground plane and layer structure must be in order to have a beam that approaches the limiting case of an infinite structure. (For example, if the structure is of finite size with a circular shape, the radius of the structure in the horizontal direction might be chosen so that 90% of the power in the leaky waves is radiated by the time the waves reach the periphery of the structure, where an absorber would be placed.) This is a significant advantage of the leaky-wave approach.

The directivity increases as the permittivity of the superstrate layer increases relative to that of the substrate layer, since the superstrate PRS then acts as a more reflective surface. Another way to increase the reflection from the PRS was proposed by Yang and Alexopoulos [57] in 1987, in which the single superstrate is replaced by a *periodic array of such superstrates*, as shown in Figure 7.23. In this structure the PRS consists of a stack of multiple superstrate layers, where the high permittivity superstrate layers with parameters (ε_2, μ_2) are separated by low permittivity spacer layers with parameters (ε_1, μ_1). The high permittivity layers each have thickness t chosen from Eq. (7.65), while the low permittivity spacer layers have one-half the thickness b of the bottom dielectric layer ($h = b$ in Eq. (7.61)). By using the multiple superstrates the reflection of the parallel-plate waveguide modes from the PRS (the stack of superstrate layers) is increased, and thus the radiated beam becomes even narrower than for a single superstrate. The directivity increases geometrically with the number of superstrate layers, and thus very directive beams may be obtained using modest values of superstrate permittivity, provided that several superstrate layers are used. A leaky-wave explanation for the multiple-layer dielectric structure proposed by Yang and Alexopoulos was presented in

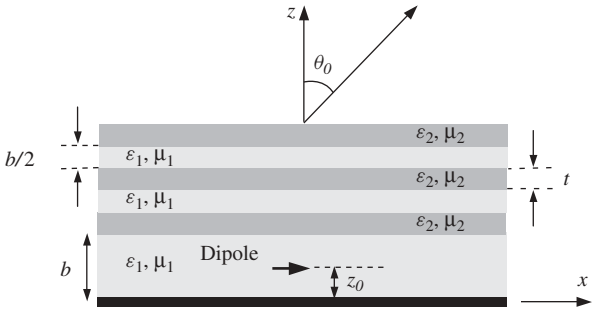


Figure 7.23 A dielectric-layer leaky-wave antenna consisting of multiple superstrate layers over a grounded substrate layer of thickness b. The structure is shown excited by a horizontal electric dipole source in the middle of the substrate.

1993 by Jackson, Oliner, and Ip [15]. Formulas were derived for the leakage constant as a function of the layer parameters and the number of superstrate layers.

A fundamentally different approach to analyzing the multiple-superstrate structure was published in 1999 by Thevenot, Cheype, Reineix, and Jecko [58]. The structure is the same as the one seen in Figure 7.23, but for only two superstrate layers. The source is also a patch antenna rather than a dipole, and the low permittivity region is air. The approach and associated terminology used to describe the operation of the antenna were completely different from the leaky-wave antenna point of view. In the adopted approach [58] the low/high permittivity superstrate layers, which serve as the cover for the bottom dielectric layer, were viewed as a finite section of a periodic *electromagnetic bandgap* (EBG) or *photonic bandgap* (PBG) structure [45], meaning that it is a periodic structure operating in a stopband. From this point of view the substrate of thickness b acts as a defect in the periodic EBG structure.

7.6.5 Periodic PRS Structures

The PRS may consist of a periodic structure that is periodic in one direction, such as an array of closely spaced wires or strips (or more generally, a stack of such wire layers). The PRS may also consist of a two-dimensional periodic array of elements. The original work of von Trentini [11] and the early work of Honey [20] consisted of such PRSs. In Honey's work, however, the structure was fed from one end, making Honey's structure a type of 1D leaky-wave antenna instead of the 2D leaky-wave antennas discussed here. Here, we first examine the radiation properties for two particular PRSs, one consisting of a periodic array of metal patches [17] and the other its complement, namely, a periodic array of rectangular apertures or slots in a thin metal plate [18].

The metal-patch PRS structure is shown in Figure 7.22b, while the slot-PRS structure is shown in Figure 7.22c. Both PRSs may be used to produce a broadside pencil beam when air is used as the substrate dielectric. Figure 7.24 shows a typical set of E-plane

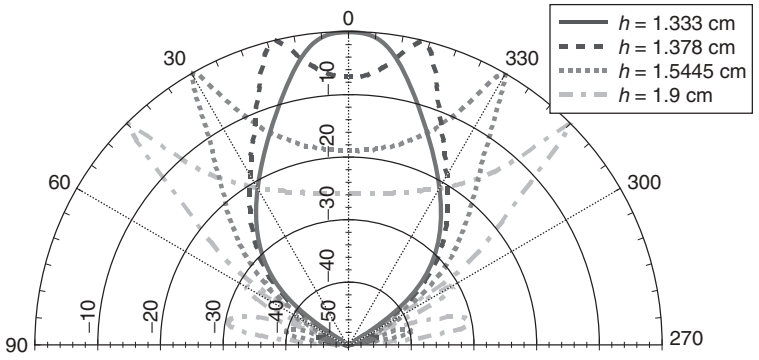


Figure 7.24 E-plane radiation patterns (in dB) at 12 GHz for a broadside beam and three different scan angles (15°, 30°, and 45°), for the metal-patch PRS structure of Figure 7.22b, using an air substrate. The patches have lengths (x-dimension) of 1.25 cm and widths (y-dimension) of 0.1 cm, with periodicities of 1.35 cm in the x-direction and 0.3 cm in the y-direction. The structure is excited by an x-directed infinitesimal horizontal electric dipole in the middle of the layer, directly below one of the patches. For the broadside beam, the substrate thickness is $h = 1.333$ cm. For the conical beams, the substrate thickness is 1.378 cm, 1.545 cm, and 1.900 cm, corresponding to beam angles of 15°, 30°, and 45°, respectively. Each pattern is normalized to 0 dB at its maximum.

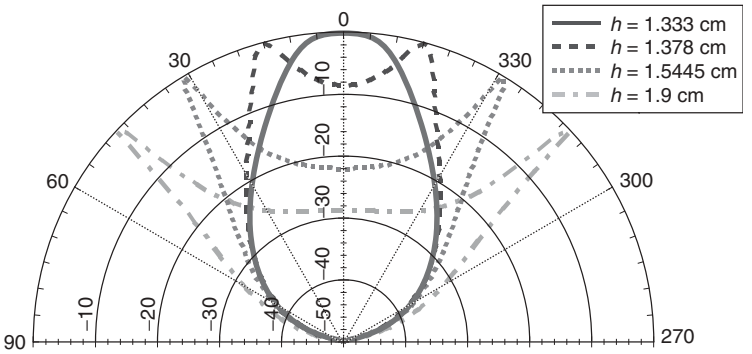


Figure 7.25 H-plane radiation patterns for the same structure as in Figure 7.24. Each pattern is normalized at 0 dB at its maximum.

($\phi = 0$) patterns obtained with the patch-PRS structure, showing the E-plane pattern for a broadside beam and a conical beam at $\theta_0 = 15^\circ$, 30°, and 45°. Figure 7.25 shows the H-plane ($\phi = 90^\circ$) patterns, which are similar. However, it can be observed that the H-plane pattern becomes narrower than the E-plane pattern as the beam angle increases, consistent with the results shown in Table 7.1.

For a conical beam, the beam angle θ_0 is limited to about 45°. This is because the thickness h of the air layer increases without limit as the angle θ_0 increases toward 90°, as seen from Eq. (7.61). As the substrate thickness increases beyond one wavelength, the next higher order set of TM_z and TE_z parallel-plate modes ($m = 2$) begin to propagate, and radiation from this set of leaky modes results in an undesirable secondary beam. Also, beyond about 45° the difference in the E- and H-plane patterns starts to become severe.

The limitation on the beam angle due to the higher-order parallel-plate modes may be overcome by using a dielectric layer with $\varepsilon_r > 4/3$, which allows for the beam angle θ_0 to approach 90° before radiation from a secondary beam occurs [17]. However, when using a dielectric substrate layer with a metal-patch PRS, it is observed that undesirable secondary beams arise from another source, namely, the $n = -1$ space harmonic in the x-direction of the (perturbed) TM_0 surface wave that is supported by the grounded substrate layer (which is perturbed by the metal patches) [17]. It is found that when using the slot-PRS structure of Figure 7.22c this problem is avoided, since the guiding structure is now a perturbed parallel-plate waveguide (which is perturbed by the apertures) rather than a perturbed grounded slab. A beam approaching 90° may be realized (in principle) by using a slot-PRS structure with a dielectric substrate layer [18]. However, as the beam angle increases, the E- and H-plane patterns become increasingly different, with the E-plane pattern broadening and the H-plane pattern narrowing, consistent with the results shown in Table 7.1.

7.6.6 Strip-Grating PRS

Recent work has shown that it is possible to overcome the problem of different beamwidths in the E- and H-planes by the use of a strip-grating PRS (an array of closely spaced conducting strips or wires) together with a dielectric layer that is air or a low permittivity material, as shown in Figure 7.22d [59]. This interesting property is obtained because the characteristics of the strip grating (i.e., the equivalent shunt

susceptance B_s) change with the angle θ_0 in a manner so as to exactly compensate for the natural change in the different principal-plane patterns that occurs with a fixed value of B_s.

An investigation of this effect is provided in Ref. 59, where it is shown that a single leaky mode that is a TM_x polarized wave is responsible for the narrow-beam effect. (The leaky mode is a TM_x wave since the current that flows on the strips of the PRS is purely in the x-direction for narrow strips.) This single leaky mode is a merging of the two leaky modes, one TM_z and the other TE_z, which are normally responsible for the conical beam. Furthermore, it is shown in Ref. 59 that because of the spatial dispersion of the strip grating, the complex propagation wavenumber of the TM_x leaky mode is independent of the azimuth angle of propagation ϕ of the leaky mode on the structure. This unique property is responsible for the attribute that the pattern has equal beamwidths in the E- and H-planes, regardless of the beam angle θ_0.

Unfortunately, the peak power densities radiated at angle θ_0 are not exactly equal in the E- and H-planes for the strip-grating PRS structure, so the conical beam is not perfectly omnidirectional in azimuth (even though the beamwidth and the beam angle remain constant with respect to the angle ϕ). However, they are nearly equal for a considerable range of angles.

7.7 CONCLUSIONS

In this chapter various types of leaky-wave antennas have been examined. These include one-dimensional leaky-wave antennas based on a guiding structure that supports wave propagation in one direction (e.g., a rectangular waveguide or a microstrip line). These one-dimensional structures form the most common type and may be uniform, quasi-uniform, or periodic. This class produces either a fan beam or a pencil beam, depending on the width of the structure.

Another type of leaky-wave antenna that has been discussed is the two-dimensional leaky-wave antenna, where a simple small source, such as a dipole or slot, excites a radially expanding cylindrical wave, which in turn produces a narrow radiating beam in the form of a pencil beam at broadside or a conical beam at a scan angle. This class of leaky-wave antennas is not yet widely recognized, but its use has been increasing in recent years.

Recent developments, including the introduction of metamaterial-based leaky-wave antennas, have also been overviewed here.

The fundamental role of leaky waves in explaining the performance of a leaky-wave antenna has been discussed in some detail. In particular, the physical nature of the leaky waves and the radiation from them have been described for both one-dimensional and two-dimensional types of structures.

The emphasis in this chapter has been on the fundamental physical properties of leaky-wave antennas. More specific design information for various types of leaky-wave antennas can be found in Ref. 4; furthermore, many additional specific leaky-wave antenna structures, and practical design information for them, are presented in Ref. 4.

ACKNOWLEDGMENT

The authors would like to thank Varada Rajan Komanduri for carefully proofreading the manuscript.

REFERENCES

1. C. H. Walter, *Traveling Wave Antennas*, McGraw-Hill, New York, 1965.
2. A. Hessel, General characteristics of traveling-wave antennas, in *Antenna Theory, Part 2*, R. E. Collin and F. J. Zucker (Eds.), McGraw-Hill, New York, 1969, Chap. 19.
3. T. Tamir, Leaky-wave antennas, in *Antenna Theory, Part 2*, R. E. Collin and F. J. Zucker (Eds.), McGraw-Hill, New York, 1969, Chap. 20.
4. A. A. Oliner and D. R. Jackson, Leaky-wave antennas, in *Antenna Engineering Handbook*, 4th ed., J. L. Volakis (Ed.), McGraw-Hill, New York, 2007, Chap. 11.
5. W. Croswell and F. J. Zucker, Surface-wave antennas, in *Antenna Engineering Handbook*, 4th ed., J. L. Volakis (Ed.), McGraw-Hill, New York, 2007, Chap. 10.
6. W. W. Hansen, Radiating electromagnetic waveguide, U.S. Patent No. 2,402,622, 1940.
7. J. N. Hines and J. R. Upson, A wide aperture tapered-depth scanning antenna, *Ohio State University Research Foundation*, Report 667–7, Columbus, OH, December 1957.
8. W. Rotman and N. Karas, The sandwich wire antenna: a new type of microwave line source radiator, *IRE Conv. Rec.,* Part 1, p. 166, 1957.
9. W. Rotman and A. A. Oliner, Asymmetrical trough waveguide antenna, *IRE Trans. Antennas Propag.*, Vol. 7, pp. 153–162, April 1959.
10. A. A. Oliner, Scannable millimeter wave arrays, Final Report on RADC Contract No. F19628–84-K-0025, Polytechnic University, 30 September 1988.
11. G. von Trentini, Partially reflecting sheet arrays, *IEEE Trans. Antennas Propag.*, Vol. 4, pp. 666–671, October 1956.
12. N. G. Alexopoulos and D. R. Jackson, Fundamental superstrate (cover) effects on printed circuit antennas, *IEEE Trans. Antennas Propag.*, Vol. 32, pp. 807–816, August 1984.
13. D. R. Jackson and N. G. Alexopoulos, Gain enhancement methods for printed circuit antennas, *IEEE Trans. Antennas Propag*., Vol. 33, pp. 976–987, September 1985.
14. D. R. Jackson and A. A. Oliner, A leaky-wave analysis of the high-gain printed antenna configuration, *IEEE Trans. Antennas Propag*., Vol. 36, pp. 905–910, July 1988.
15. D. R. Jackson, A. A. Oliner, and A. Ip, Leaky-wave propagation and radiation for a narrow-beam multiple-layer dielectric structure, *IEEE Trans. Antennas Propag.*, Vol. 41, pp. 344–348, March 1993.
16. P. Feresidis and J. C. Vardaxoglou, High gain planar antenna using optimised partially reflective surfaces, *IEE Proc. Microwaves Antennas Propag*., Vol. 148, pp. 345–350, December 2001.
17. T. Zhao, D. R. Jackson, J. T. Williams, H. Y. Yang, and A. A. Oliner, 2-D periodic leaky-wave antennas—Part I: Metal patch design, *IEEE Trans. Antennas Propag.*, Vol. 53, pp. 3505–3514, November 2005.
18. T. Zhao, D. R. Jackson, and J. T. Williams, 2-D periodic leaky-wave antennas—Part II: Slot design, *IEEE Trans. Antennas Propag.*, Vol. 53, pp. 3515–3524, November 2005.
19. L. O. Goldstone and A. A. Oliner, Leaky-wave antennas—Part I: Rectangular waveguides, *IRE Trans. Antennas Propag*., Vol. 7, pp. 307–319, October 1959.
20. R. C. Honey, A flush-mounted leaky wave antenna with predictable patterns, *IRE Trans. Antennas Propag.,* Vol. 7, pp. 320–329, October 1959.
21. R. Gilbert, Slot antenna arrays, in *Antenna Engineering Handbook*, 4th ed., J. L. Volakis (Ed.), McGraw-Hill, New York, 2007, Chap. 9.
22. T. Tamir and A. A. Oliner, Guided complex waves, Part I: Fields at an interface, *Proc. IEE,* Vol. 110, pp. 310–324, February 1963.
23. T. Tamir and A. A. Oliner, Guided complex waves, Part II: Relation to radiation patterns, *Proc. IEE*., Vol. 110, pp. 325–334, February 1963.

24. W. C. Chew, *Waves and Fields in Inhomogeneous Media*, Wiley/IEEE Press, Piscataway, NJ, 1995.

25. H. Shigesawa, M. Tsuji, and A. A. Oliner, The nature of the spectral-gap between bound and leaky solutions when dielectric loss is present in printed-circuit lines, *Radio Sci.*, Vol. 28, No. 6, pp. 1235–1243, November–December 1993.

26. R. F. Harrington, *Time Harmonic Electromagnetic Fields*, Wiley/IEEE Press, Piscataway, NJ, 2001.

27. C. A. Balanis, *Antenna Theory: Analysis and Design*, 3rd ed., John Wiley & Sons, Hoboken, NJ, 2005.

28. R. S. Elliott, *Antenna Theory and Design*, Prentice Hall, Englewood Cliffs, NJ, 1981.

29. R. E. Collin, *Field Theory of Guided Waves*, Wiley/IEEE Press, Piscataway, NJ, 1991.

30. M. Guglielmi and G. Boccalone, A novel theory for dielectric-inset waveguide leaky-wave antennas, *IEEE Trans. Antennas Propag.*, Vol. 39, pp. 497–504, April 1991.

31. J. R. James and P. S. Hall, Microstrip antennas and arrays, Part 2: New array-design technique, *IEE J. Microwaves Optics Antennas*, Vol. 1, pp. 175–181, 1977.

32. M. Guglielmi and D. R. Jackson, Broadside radiation from periodic leaky-wave antennas, *IEEE Trans. Antennas Propag.*, Vol. 41, pp. 31–37, January 1993.

33. S. Paulotto, P. Baccarelli, F. Frezza, and D. R. Jackson, A novel technique to eliminate the open stopband in 1-D periodic printed leaky-wave antennas, in *Proceedings of EuCAP 2007*, 11–16 November, Edinburgh, Scotland.

34. L. Liu, C. Caloz, and T. Itoh, Dominant mode leaky-wave antenna with backfire-to-endfire scanning capability, *Electron. Lett.*, Vol. 38, pp. 1414–1416, November 2002.

35. A. Grbic and G. V. Eleftheriades, A backward-wave antenna based on negative refractive index L–C networks, in *Proceedings of the IEEE International Symposium on Antennas and Propagation*, San Antonio, TX, 16–21 June 2002, Vol. 4, pp. 340–343.

36. G. V. Eleftheriades, A. K. Iyer, and P. C. Kremer, Planar negative refractive index media using periodically loaded transmission lines, *IEEE Trans. Microwave Theory Tech.*, Vol. 50, No. 12, pp. 2702–2712, December 2002.

37. A. Sanada, C. Caloz, and T. Itoh, Characteristics of the composite right/left-handed transmission lines, *IEEE Microwave Wireless Components Lett.*, Vol. 14, No. 2, pp. 68–70, February 2004.

38. T. Kokkinos, C. D. Sarris, and G. V. Eleftheriades, Periodic FDTD analysis of leaky-wave structures and applications to the analysis of negative-refractive-index leaky-wave antennas, *IEEE Trans. Microwave Theory Tech*., Vol. 54, pp. 1619–1630, April 2006.

39. S. Paulotto, P. Baccarelli, F. Frezza, and D. R. Jackson, Full-wave dispersion analysis and broadside optimization for the microstrip CRLH leaky-wave antenna, in *Proceedings of Metamaterials 2007*, 22–26 October, Rome, Italy.

40. S. Paulotto, P. Baccarelli, F. Frezza, and D. R. Jackson, A microstrip periodic leaky-wave antenna optimized for broadside scanning, in *Digest of the IEEE AP-S International Symposium*, Hawaii, June 2007, pp. 5789–5792.

41. S. Paulotto, P. Baccarelli, F. Frezza, and D. R. Jackson, A novel technique to eliminate the open stopband in 1-D periodic printed leaky-wave antennas, *Proceedings 2nd European Conf. on Antennas and Propagation (EuCAP)*, 11–16 November, 2007.

42. A. A. Oliner, Radiating periodic structures: analysis in terms of k vs. β diagrams, Report SC-5-63 distributed at a seminar given at the Polytechnic Institute of Brooklyn (now Polytechnic University), June 4, 1963, as part of the short-course series "Microwave Field and Network Techniques."

43. T. Rozzi and M. Mongiardo, *Open Electromagnetic Waveguides*, IEE Electromagnetic Waves Series, No. 43, Institution of Electrical Engineers, 1997.

44. P. Baccarelli, S. Paulotto, D. R. Jackson and A. A. Oliner, A new Brillouin dispersion diagram for one-dimensional periodic printed structures, *IEEE Trans. Microwave Theory Tech.*, Vol. 55, pp. 1484–1495, July 2007.
45. N. Engheta and R. W. Ziolkowski, Eds., *Electromagnetic Metamaterials: Physics and Engineering Explorations*, Wiley-IEEE Press, 2006.
46. C. Caloz and T. Itoh, *Electromagnetic Metamaterials: Transmission Line Theory and Microwave Applications*, Wiley, 2006.
47. G. V. Eleftheriades and K. G. Balmain, *Negative-Refraction Metamaterials*. Wiley Hoboken, NJ, 2005.
48. S. Zouhdi, A. Sihvola, and M. Arsalane, Eds., *Advances in Electromagnetics of Complex Media and Metamaterials*, Kluwer Academic Publishers, 2003.
49. G. Lovat, P. Burghignoli, F. Capolino, D. R. Jackson, and D. R. Wilton, Analysis of directive radiation from a line source in a metamaterial slab with low permittivity, *IEEE Trans. Antennas and Propagation*, Vol. 54, pp. 1017–1030, March 2006.
50. G. Lovat, P. Burghignoli, F. Capolino, D. R. Jackson, On the combinations of low/high permittivity and/or permeability substrates for highly directive planar metamaterial antennas, *IET Microwaves, Antennas & Propagation*, Vol. 1, No. 1, February, 2007.
51. C. A. Allen, K. M. K. H. Leong, C. Caloz, and T. Itoh, A two-dimensional edge excited metamaterial-based leaky-wave antenna, in *Proceedings IEEE AP-S International Symposium*, Washington, DC, June 2005.
52. C. Caloz, T. Itoh, and A. Rennings, CRLH metamaterial leaky-wave and resonant antennas, *IEEE AP-S Magazine* (to appear).
53. H. V. Nguyen, S. Abielmona, A. Rennings, and C. Caloz, Pencil-beam, 2D scanning leaky-wave antenna array, *International Symposium on Signals, Systems and Electronics (ISSSE) Digest*, Montréal, Canada, pp. 139–142, July-Aug. 2007.
54. A. Ip and D. R. Jackson, Radiation from cylindrical leaky waves, *IEEE Trans. Antennas Propag.*, Vol. 38, pp. 482–488, April 1990.
55. F. Bichelot, R. Loison, and L. Le Coq, FSS-EBG antenna with improved directivity bandwidth: theory, design and measurements, in *Digest of the IEEE AP-S International Symposium*, Hawaii, June 2007, pp. 5423–5426.
56. T. Zhao, D. R. Jackson, and J. T. Williams, General formulas for 2D leaky wave antennas, *IEEE Trans. Antennas Propag.*, Vol. 53, pp. 3525–3533, November 2005.
57. H. Y. Yang and N. G. Alexopoulos, Gain enhancement methods for printed circuit antennas through multiple superstrates, *IEEE Trans. Antennas Propag*., Vol. 35, pp. 860–863, July 1987.
58. M. Thevenot, C. Cheype, A. Reineix, and B. Jecko, Directive photonic-bandgap antennas, *IEEE Trans. Microwave Theory Tech*., Vol. 47, pp. 2115–2122, November 1999.
59. G. Lovat, P. Burghignoli, F. Capolino, D. R. Jackson, and D. R. Wilton, High-gain omnidirectional radiation patterns from a metal strip grating leaky-wave antenna, in *Digest of the IEEE AP-S International Symposium*, Hawaii, June 2007, pp. 5797–5800.

CHAPTER 8

Reconfigurable Antennas

GREGORY H. HUFF and JENNIFER T. BERNHARD

8.1 INTRODUCTION

Reconfigurable antennas have existed for many years and enjoy a very rich and diverse history of innovation and design. Reconfiguration of an antenna is achieved through an intentional redistribution of the currents or, equivalently, the electromagnetic fields of the antenna's effective aperture, resulting in reversible changes in the antenna impedance and/or radiation properties. These changes are enabled through various mechanisms such as switching, material tuning, and structural modifications. System control can then be applied to result in desired antenna performance. Under this definition, reconfiguration does not include the use of peripheral matching circuits or the effects from phase shifters, bias/control lines, or any other device/component that does not interact directly with the radiating mechanism. Thus a reconfigurable antenna provides a trade-off in increased complexity for enhanced performance. From a practical perspective, the additional performance provided by a reconfigurable antenna should therefore offset the costs and complexity of reconfiguration.

Application areas that drive the development of reconfigurable antennas include multifunction wireless devices, multiple-input multiple-output (MIMO) and ultrawideband (UWB) systems, and antijamming, secure communications. Meeting the needs of these systems presents several challenges. One of the most fundamental occurs when restricting reconfiguration to a specific aspect of the antenna performance, while simultaneously requiring continuity of the remaining antenna parameters throughout the range of reconfigured states. Implementation and control of the reconfiguration mechanism can also prove difficult. Considering all aspects of this process, the design of reconfigurable antennas can be quite demanding, but it also provides an excellent opportunity to combine state-of-the-art technologies with antenna theory in the pursuit of additional degrees of freedom in system performance.

This chapter addresses the fundamental properties of reconfigurable antennas and also provides case studies to illustrate the potential of these radiators. First considered are large aperture reconfigurable antennas, followed by single element and small aperture reconfigurable antennas. The use of antenna reconfiguration in multiple antenna

Modern Antenna Handbook. Edited by Constantine A. Balanis

topologies and arrays (both parasitic and driven) follows this discussion, and the chapter concludes with a brief summary.

8.2 THE RECONFIGURABLE ANTENNA

8.2.1 Initial Considerations

Reconfigurable antennas come in a large variety of different shapes and forms. Their operation can largely be analyzed through existing design principles [1, 2] by utilizing well defined antennas as the base design and a point of reference for the desired operation. By considering the properties of a base design, reconfigurable antennas can be classified according to three categories that describe their operation: (1) the reconfigurable antenna parameters of interest, (2) the proximity of reconfiguration, and (3) the continuity of reconfiguration (e.g., having reconfigurable antenna parameters over a continuous range of values). Reconfigurable antennas are typically described by the first of these categories, including reconfigurable radiation (pattern or polarization) and reconfigurable impedance (frequency or bandwidth). The proximity of reconfiguration describes physical properties inherent to the base antenna design—either direct (alteration of a driven element) or parasitic (alteration of a parasitic component). The continuity of the reconfiguration is defined by the nature and capabilities of the reconfiguration mechanism, either discrete (a finite number of reconfigured states) or continuous (reconfiguration within a range of states). Thus for the reconfigurable antennas discussed in this chapter, these three properties will have the notation (*reconfigurable antenna parameter(s), proximity of reconfiguration, continuity of reconfiguration*) contained in their description.

Example 1: Frequency Reconfigurable Dipole A generic wireless communication link shown in Figure 8.1 illustrates a basic application of reconfigurable antennas. This scenario involves a transmitter T that broadcasts to two sets of wireless receivers R_1 and R_2. These receivers operate at two different frequency bands B_1 and B_2, centered at f_1 and f_2, respectively (with $f_1 < f_2$). The scenario assumes that all receiving antennas are coincidentally polarized, the transmitter broadcasts at the frequency bands centered at f_1 and f_2 at times t_1 and t_2, respectively, and the radios require isolation between the bands such that a dual-band antenna is undesirable. Thus the reconfigurable antenna serves to allow communication with both sets of receivers using a single antenna.

Although many types of antennas could be used to achieve the desired operation—using a single antenna to communicate with both sets of receivers—the dipole

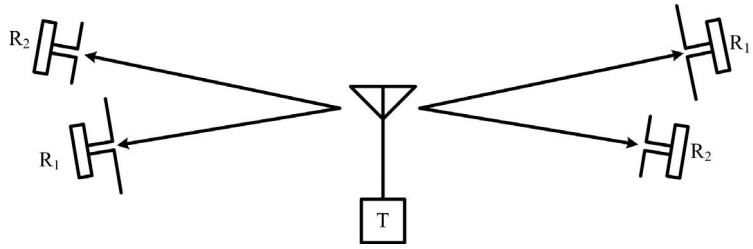


Figure 8.1 Example of a wireless communication link utilizing reconfigurable antennas. The transmitter T utilizes a frequency reconfigurable antenna for broadcasting to the receivers R_1 and R_2, operating at f_1 and f_2, at time t_1 and t_2, respectively.

provides an intuitive and well understood candidate for reconfiguration. Recalling the electrical properties of the dipole at its first resonance, the desired performance can be achieved by adjusting the length of the dipole accordingly. In doing so, the overall system avoids the need for two sets of transmitting antennas by combining their functionality into a single—more complex (e.g., the trade-off)—antenna design (assuming a dual-band antenna is undesirable). Figure 8.2 depicts the process of combining the functionality of two transmitters (T_1 and T_2) and their corresponding dipoles of lengths L_1 and L_2 into a frequency reconfigurable dipole antenna (*frequency, direct, discrete*). For this example, the integration of two ideal switches (placed accordingly) on the longer length L_1 (corresponding to the resonant length for f_1) achieves the desired operation. Figure 8.2 also shows the resulting current distribution in gray for the two configurations (or reconfigured states) of the antenna for the time intervals t_1 and t_2.

Figure 8.2 also includes the control signal (indicated by a control voltage), allowing for a very general description of the antenna's operation. The control signal remains high during the interval t_1 and both switches remain in the ON state. This provides electrical continuity between the respective lengths and results in a total length of $L_1 = \frac{1}{2}\lambda_1$ and a dipole resonance corresponding to f_1 (the lower of the two frequency bands). The switches turn off (OFF state) during the next interval t_2 when the control signal falls below the threshold level, creating discontinuities along the arms of the dipole. This shortens the total length to $L_2 = \frac{1}{2}\lambda_2$ and a dipole resonance at f_2 (the higher of the two frequency bands). Equation (8.1) shows the pattern function $F(\theta)$ for the dipole in

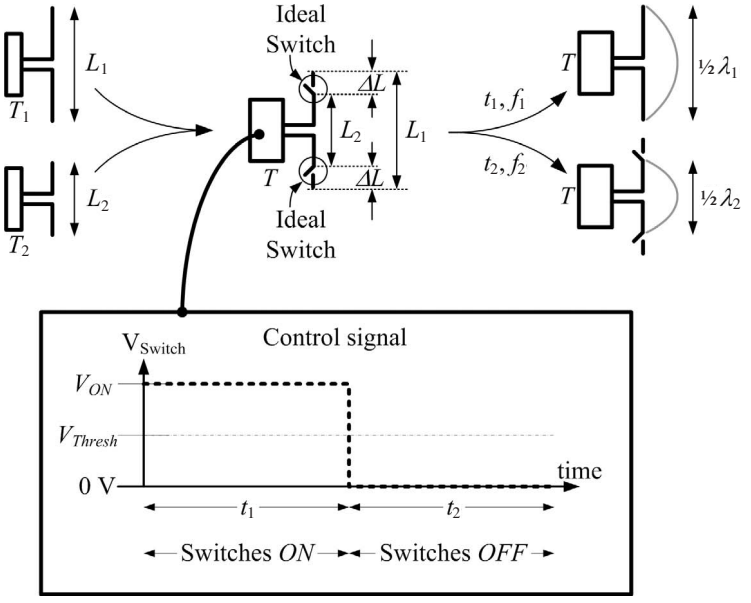


Figure 8.2 By combining the functionality of the two antenna and transmitter pairs (upper left) into a reconfigurable antenna using ideal switches (upper middle), both bands of operation can be achieved (upper right). The basic control signal for this system (lower middle) also shows the switch state underneath the corresponding time interval.

both states [2] (θ assumes alignment of the axis of the dipole in the z-direction) and demonstrates that the radiation patterns are identical for each configuration.

$$\left.\begin{aligned} F_1(\theta) &= \frac{\cos\left(\frac{k_1 L_1}{2}\cos\theta\right) - \cos\left(\frac{k_1 L_1}{2}\right)}{\sin\theta} = \frac{\cos\left(\frac{\pi}{2}\cos\theta\right)}{\sin\theta} \\ F_2(\theta) &= \frac{\cos\left(\frac{k_2 L_2}{2}\cos\theta\right) - \cos\left(\frac{k_2 L_2}{2}\right)}{\sin\theta} = \frac{\cos\left(\frac{\pi}{2}\cos\theta\right)}{\sin\theta} \end{aligned}\right\} \rightarrow F_1(\theta) = F_2(\theta) \qquad (8.1)$$

8.2.2 Practical Considerations

There are numerous practical considerations associated with the implementation of a reconfigurable antenna design. The most challenging issue occurs during the process by which a conceptual design finds realization. A proof-of-concept design (e.g., using hard-wired connections for ideal switches in Example 1) may achieve the desired performance, but its functionality can be severely degraded, or even lost, when implementing the reconfiguration mechanism and its peripheral components (bias, control, etc.). Many important factors should be considered in this process—each specific to the particular type of antenna and reconfiguration mechanism under consideration. Some general effects include parasitic loading from switches and other components, and spurious radiation from bias structures or control networks.

Example 2: Realization of the Frequency Reconfigurable Dipole Figure 8.3 shows a potential topology for the frequency reconfigurable dipole discussed in Example 1, providing an implementation using diodes as the reconfiguration mechanism in place of the ideal switches. Assumptions for this example include the following: (1) the input impedance Z_{in} of the dipole remains adequately matched to the system impedance Z_0

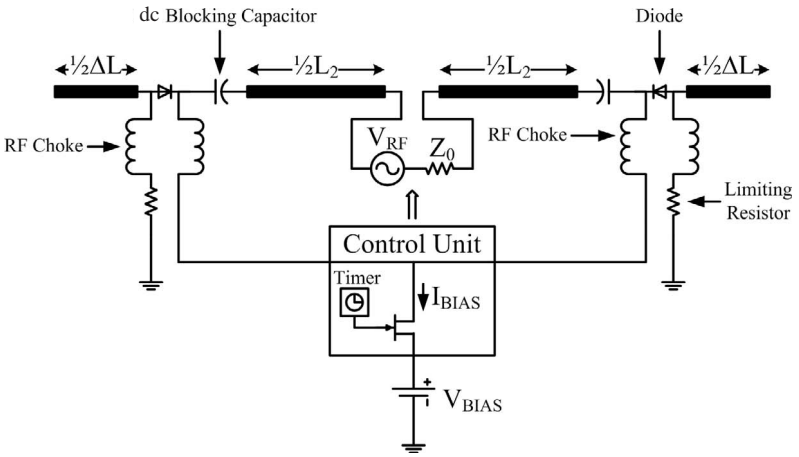


Figure 8.3 Topology of the reconfigurable frequency dipole antenna, including the diode switches, biasing components (dc blocks, RF chokes, and limiting resistors), and the control unit (timer, FET gate, and power supply.)

throughout both configurations, (2) the diodes are ideal (lossless, negligible parasitic reactance, and electrically small dimensions), and (3) the control unit also adjusts the source frequency V_{RF} [e.g., via a voltage-controlled oscillator (VCO)] between f_1 and f_2. Figure 8.3 shows the diodes—appropriately oriented across both discontinuities—that are controlled using V_{BIAS}, routed through a control unit (with an identical control signal from Example 1) consisting of a solid-state switch such as a field-effect transistor (FET) and a timer that dictate the control signal. Inductors or very high impedance transmission lines are used as radio frequency (RF) chokes to mitigate RF leakage into the bias and control lines by presenting a very high impedance to the RF energy at the terminals of the diodes. It should be noted that the improper design of these lines can generate spurious radiation and impedance loading that can affect the performance. Blocking capacitors are used in this design to isolate and protect the RF source from the dc bias, and current limiting resistors are used to avoid damaging the diodes and/or dc voltage source. The basic system shown in Figure 8.3 contains many components common to a reconfigurable antenna system.

At this stage, losses in the antenna and reconfiguration mechanism represent a particularly critical parameter in the design and implementation of reconfigurable antennas, especially when considering the total efficiency across all reconfigured states. These losses can be manifested through detuned impedances and/or can alter radiation patterns but can also be a function of lossy materials (accumulation of material and/or conductor losses), current crowding through small dimensions (often encountered in switched designs), and other less apparent mechanisms. For example, serpentining the bias or control lines (as from the previous example) represents an effective and often used compensating strategy to mitigate spurious radiation losses or loading from these structures. These lines are typically designed to have a high input impedance and must be turned several times to create opposing current directions to cancel spurious radiation. However, the turns in the narrow line will inevitably create parasitic effects and losses from current crowding that may affect performance if improperly designed or placed. Comprehensive design methodologies developed in the future for these antennas must provide the ability to account for these kinds of losses as well as provide strategies to mitigate them while balancing design choices against additional system complexity.

8.2.3 Reconfiguration Mechanism

The choice of reconfiguration mechanism presents trade-offs in functionality and complexity as well as device or system lifetime. For instance, mechanisms that rely on movable parts to facilitate reconfiguration [e.g., RF micro electro mechanical systems (RF MEMS)] may experience mechanical wear or failure from hot-switching. However, they are typically actuated electrostatically and the performance they provide over solid-state designs—low intrinsic losses and power consumption (especially at higher frequencies)—often offsets the potential for a degraded lifetime and failure of the device. As another example, the integration of tunable materials into the antenna design can cause an entirely different set of problems if the material properties or tunability degrade over time. Both of these degradations in the performance are in addition to the complexity and integration of the antenna, reconfiguration mechanism, and supporting structures (bias, control, etc.). Therefore selection of the reconfiguration mechanism should be made carefully, with system lifetime, cost, complexity, and desired performance levels and losses considered from the outset.

8.3 CASE STUDIES IN RECONFIGURABLE LARGE APERTURE ANTENNAS

8.3.1 Continuously Reconfigurable Antennas

8.3.1.1 Traveling-Wave Antennas

Trough Waveguide Antenna The electromechanically scanned trough waveguide antenna (*radiation, direct, continuous*) [3–5] represents one of the first reported examples of a reconfigurable large aperture antenna. The reconfiguration of this antenna represents several milestones in reconfigurable antenna technologies and was notably the first widely published reconfigurable antenna design (although the term *reconfigurable* was not yet used). In addition to its chronological significance, the antenna was also married with many novel technologies to achieve reconfiguration. The authors examined numerous reconfiguration mechanisms, some of which were among the first applications of artificial dielectrics and electromechanical means to operate on the electromagnetic fields of a device. The structure was first considered in the early 1950s by W. Rotman (while at the Air Force Cambridge Research Center) and A. A. Oliner (at the Microwave Research Institute, Polytechnic Institute of Brooklyn) for the United States Air Force (USAF). The intended application was an S-band pulsed-Doppler radar system, where its specific purpose was to feed a parabolic reflector and scan a pencil beam in the horizon (with mechanical scanning in the vertical plane). The significance of this operation resided in its ability to provide fixed-frequency beam steering from a leaky-wave antenna—which normally must be scanned by adjusting the frequency.

The trough waveguide provides the base design for this antenna, and Figure 8.4 [3, 4] shows the cross section of the structure, electric field of the traveling-wave mode, and electric field of the leaky mode. The open side (top) of the antenna represents the effective aperture and the critical parameter (including the associated fields in the guide) that facilitates reconfigurable radiation.

The structure radiates (appreciably) when the two troughs are asymmetric in the transverse dimension, and the transverse resonance method has been used in Ref. 5 to analyze this behavior. By asymmetrically perturbing the field structure of the traveling-wave mode, a portion of the traveling-wave energy converts to a loosely bound (radiating) mode [3]. When the perturbation (or asymmetry between the two troughs) remains fixed along the longitudinal direction, a main beam points in the direction θ and scans in the

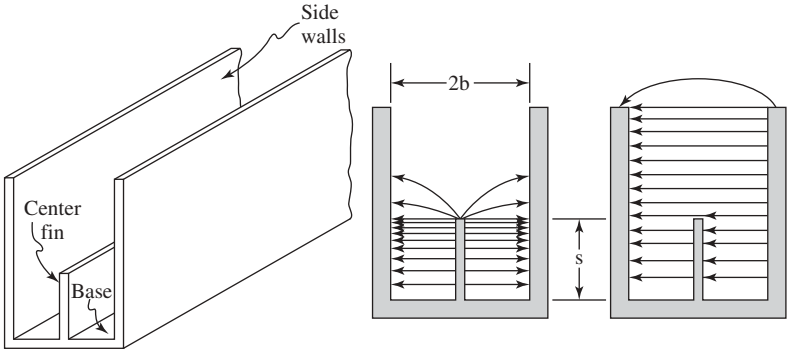


Figure 8.4 Trough waveguide antenna (left) (adapted from Ref. 3), electric field of traveling wave mode (middle), and electric field of leaky-wave mode (right) (adapted from Ref. 3.)

plane defined by the direction normal to the open dimension and the longitudinal axis of the guide (i.e., direction of the traveling wave). The scan angle θ follows Eq. (8.2), determined by the ratio of the guided wavelength λ_g and the free-space wavelength λ_0—a common trait of leaky-wave aperture antennas [1].

$$\sin\theta = \lambda_0/\lambda_g \tag{8.2}$$

A *traveling-wave linear array* can be created from this structure by periodically perturbing (period $2a$ and perturbation length a) the traveling wave in an antipodal (or antisymmetric) fashion about the center fin. Equation (8.3) [3] shows the relationship between θ, a, λ_g, and λ_0 that results from these perturbations, effectively creating a secondary mechanism that influences the scan angle of the main beam. With this mechanism in place, the perturbation length and its periodicity, along with the guided wavelength, represent the critical parameters that operate on the fields of the effective aperture—resulting in reconfigurable radiation (beam steering).

$$\sin\theta = \frac{\lambda_0}{\lambda_g} - \frac{\lambda_g}{2a} \tag{8.3}$$

Three different methods [5] are presented next, which demonstrate the ability to reconfigure the guided wavelength from a wide variety of fixed-length perturbations. They are (1) simultaneously rotating two low-dielectric isocane foam rods implanted with a parallel grid of small metallic filaments to create an artificial dielectric in both troughs, (2) mechanically varying a serrated center fin, and (3) actuating fixed-length metallic blocks at the base of the guide. All of these methods effectively reconfigure the electromagnetic fields on the aperture. Figures 8.5 and 8.6 [3] show the second two reconfiguration mechanisms and the effect of block height on the guided wavelength, respectively. Figure 8.7 [5] shows the reconfigurable radiation patterns for several different block heights (method (3)), actuated by a cam and gear drive to adjust the height of the block and alter the guided wavelength.

Periodic Grating Antenna Tunable materials can also be quite effective in the design of reconfigurable large aperture antennas, and two different permutations of the periodic

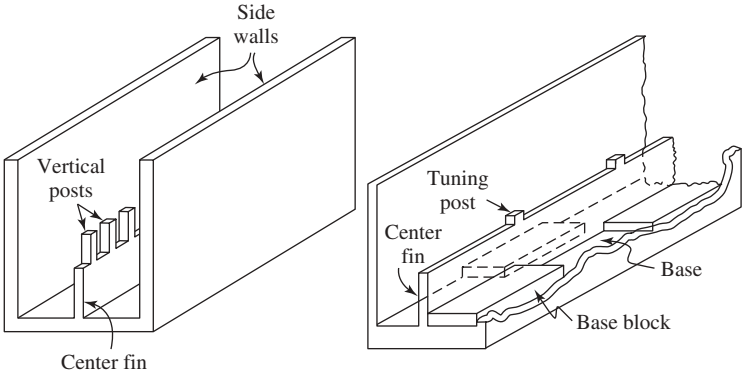


Figure 8.5 The mechanically actuated serrated center fin (left) (from Ref. 3) and actuated metallic blocks at the base of the guide (right). (From Ref. 4.)

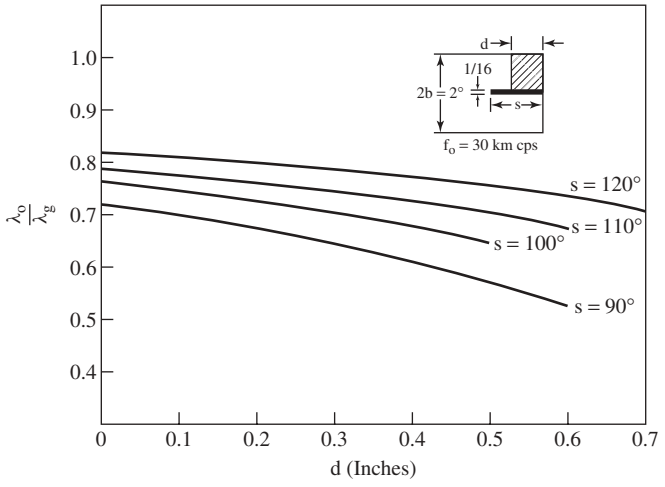


Figure 8.6 Effect of reconfiguring the block height d on the free-space to guided wavelength ratio that determines the angle θ. (From Ref. 5.)

grating antenna [6] are considered here to demonstrate this: (1) altering the permittivity of the substrate and (2) altering the permeability of the substrate. These designs have the advantage of not relying on moving parts (such as the implementation in Ref. 7) but can suffer from substantial substrate losses in the exotic materials used to reconfigure the antenna. Figure 8.8 depicts the structure of the base design, and the guided wavelength determines the direction of the main beam, which scans with the frequency. The radiation pattern can also be reconfigured by altering the guided wavelength (equivalent to the trough waveguide antenna shown by Eq. (8.2))—determined by the physical dimensions (height and width) of the guide and the material properties (permittivity and permeability). The resulting reconfiguration of the radiation can then be expressed by a figure of merit determined by the static fields that create the material variation and the resulting change in the main beam direction.

The first design utilizes a substrate with a tunable permittivity to alter the guided wavelength and reconfigure the direction of the main beam. The relative permittivity ε_r of the ferroelectric material in Ref. 8 (barium strontium titanate (BST)) can typically be tuned over a wide range (depending on the quality and type of material composition) by applying a dc electric field in its paraelectric state. This can be accomplished with the periodic metallic grating antenna by applying the bias through the substrate between the metal gratings and the ground plane. The prototype in Ref. 8 (*radiation, direct, continuous*) was designed for Ku-band operation and was capable of reconfiguring the main beam direction up to 6° (-14° to -8°) at 13.375 GHz using an applied electric field ranging from 0.0 to 9.0 kV/cm (V/m units referring to the field strength in the substrate)—resulting in a figure of merit of 1°/8.33 V. It should be noted that a much greater degree of tuning is possible by using higher quality ferroelectric materials.

The second design utilizes a stratified substrate configuration that incorporates a layer of material with a tunable permeability that can be used to alter the guided wavelength and reconfigure the direction of the main beam. The relative permeability μ_r of the ferrite material in Ref. 9 (polycrystalline yttrium iron garnet (YIG)) can be tuned by applying a dc magnetic field. This material has been corrugated with rectangular grooves to create a

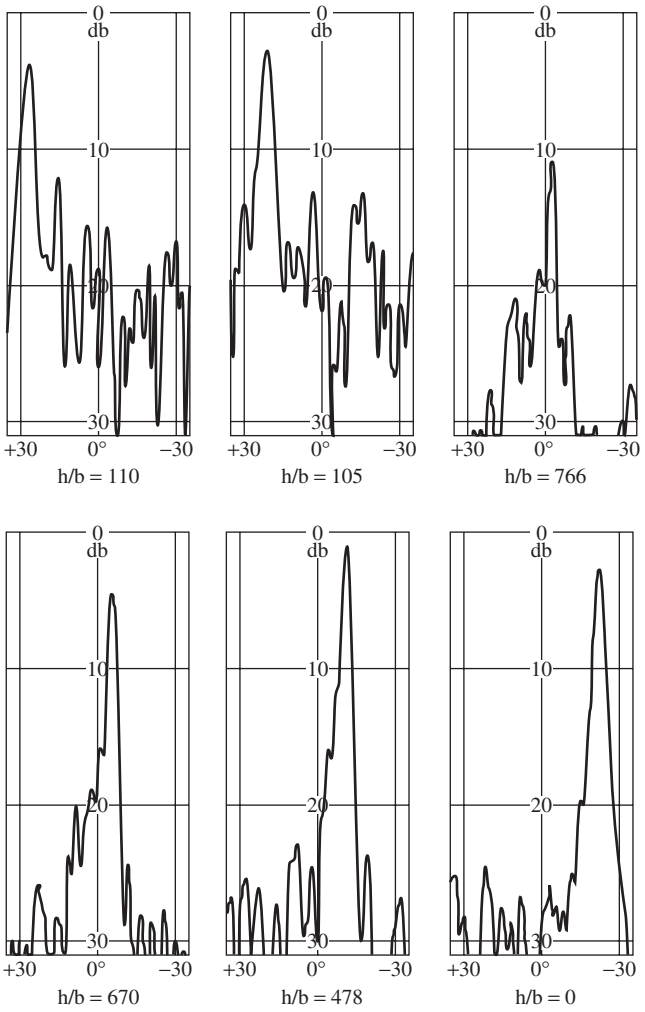


Figure 8.7 Azimuthal radiation patterns of the trough waveguide fed parabolic reflector for different block heights h (b refers to the width of the trough). (From Ref. 5.)

periodic dielectric-grating antenna and resides on the top of the substrate configuration, supported by a Teflon ($\varepsilon_r = 2.1$) slab bonded to a copper plate with wax. The antenna in Ref. 9 (*radiation, direct, continuous*) was designed for U-band operation and was capable of reconfiguring the main beam direction up to 41° (−0.5° to 40.5°) at 46.8 GHz using an applied magnetic field of 0.0 to ∼1.4 T—a figure of merit of 1°/0.036 T. A diagram of the antenna can be seen in Figure 8.9 [9].

8.3.1.2 Reflector Antennas

Contour Reflector The reconfigurable contour reflector antennas in Refs. 10–14 (*radiation, direct, continuous*) represent a very different type of reconfigurable large aperture antenna, with the contoured reflector providing the base design for these

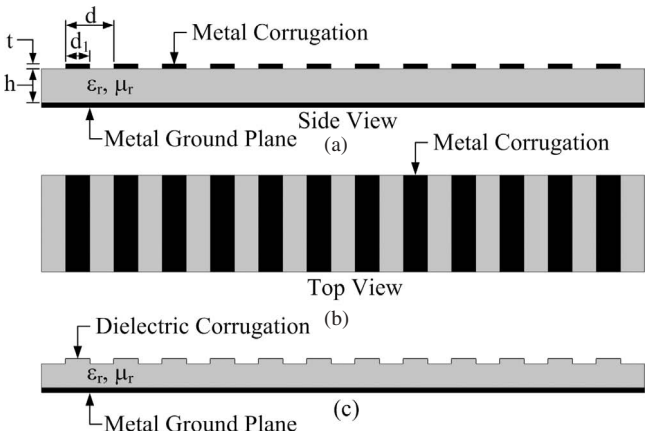


Figure 8.8 (a) Side profile of the periodic grating antenna with metallic corrugations, (b) top view of the periodic grating antenna with metallic corrugations, and (c) side profile of the periodic grating antenna with dielectric corrugations (From Ref. 1.)

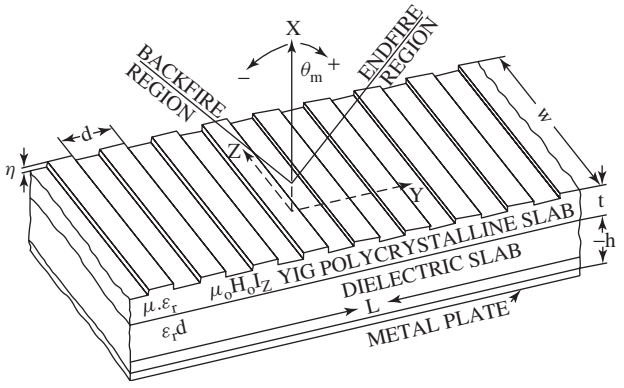


Figure 8.9 Stratified substrate configuration of the periodic grating antenna using magnetic materials. (From Ref. 9.)

structures. The illumination of a deformable reflector in Refs. 10–13, and a deformable subreflector in Ref. 14, provides a mechanism to alter the induced currents on the reflector/subreflector surface. This in turn reconfigures the effective aperture field distribution and reconfigures the shape of the resulting radiation pattern. The reflecting surfaces of these antennas are typically deformed by connecting the reflecting surface to an array or bank of actuators that are operated simultaneously through a control unit to provide the desired illumination of the reflector. Figure 8.10 shows the general arrangement of antennas with the deformable reflector and deformable subreflector surfaces, but many different actuation methods, discretization schemes, and reflector surfaces (or materials) have been examined to reconfigure this antenna.

These antennas can provide very complex radiation patterns, and small changes in the contour can be applied to the surface of the reflector for corrections or alterations of the

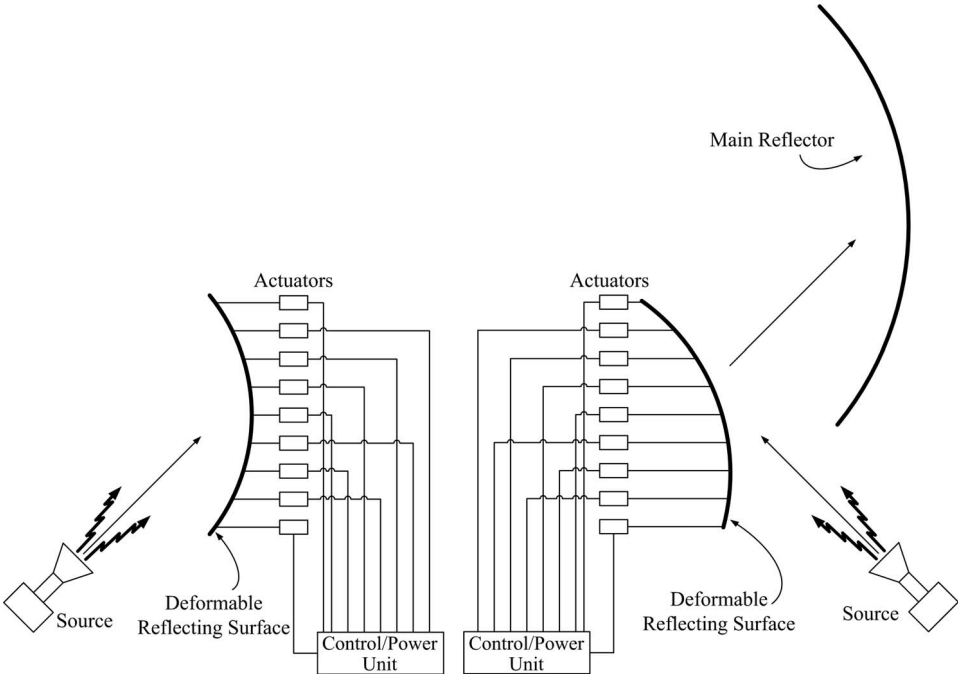


Figure 8.10 Antennas using a reconfigurable reflector or subreflector (From Ref. 2.)

pattern. Genetic algorithms in Refs. 13 and 14 and a finite element mechanical description of the deformable subreflector surface have been included in Ref. 14 to aid in pattern synthesis. In addition to the electromagnetic analysis, the effect of material stiffness and actuation schemes (e.g., number of actuators, placement, and surface deflection) can also be included in the design process using these methods. Figure 8.11 shows an example of the patterns that can be created by the antenna in Ref. 14 (continental United States (CONUS) and Brazil).

Reflectarray The reflectarray represents a slightly different type of reflector antenna, in that it uses individually reconfigurable single elements to alter the resulting large aperture field distribution. However, it can be included with the large aperture antennas since reconfiguration of the individual elements (reconfigurable small aperture antennas) serves the larger goal of reconfiguring the total aperture. In this type of arrangement, the elements must significantly reconfigure one aspect of their performances, while maintaining other parameters that may degrade the overall performance of the array.

The structure in Ref. 15 (*radiation, indirect, continuous*) illustrates this type of this array—using reconfigurable phase (or phase shifting) individual elements (*phase, direct, continuous*). This example uses vertically tuned MEMS-based patch antennas in a reflectarray application to create a large aperture radiation reconfigurable antenna. The elements are placed in a periodic array and are individually actuated to alter their height. The vertical displacement of the element alters the phase of the far-field radiation and can be used to steer the main beam generated by the overall aperture. Figure 8.12 shows the proposed MEMS patch.

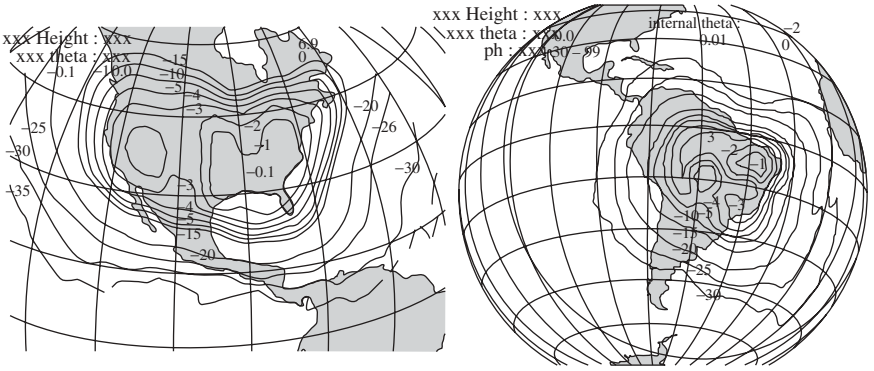


Figure 8.11 Broadcast radiation patterns achieved using contour reflector antennas: the radiation pattern contours for the continental United States (CONUS) (left), and for Brazil (right). (From Ref. 14.)

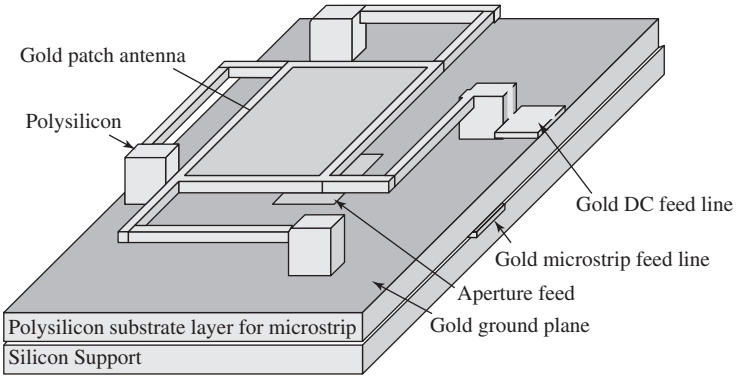


Figure 8.12 Vertically displaced reconfigurable patch element proposed for the reconfigurable reflectarray. (From Ref. 15.)

8.3.2 Discretely Reconfigurable Antennas

8.3.2.1 Reconfigurable Aperture

Switched Grid of Electrically Small Metallic Patches The switched grid of electrically small metallic patches in Refs. 16–18 (*radiation and frequency, direct, discrete*) demonstrates one method to directly reconfigure a large aperture antenna. Figure 8.13 shows the antenna structure for a monopole configuration (determined by a genetic algorithm). Each patch in the grid has a connection to neighboring patches through a switched link, and the individual switches can be opened or closed in a predetermined arrangement to reconfigure the effective aperture (and the resulting radiation pattern). This particular structure can achieve many different reconfigured states, but the arrangement of the patches to obtain the desired radiation pattern and input impedance behavior must be predetermined. Although the physical structure remains quite different from the reconfigurable contoured reflector, both reconfigurable large apertures have very complex field distributions that share the need for predetermined configuration settings to achieve the

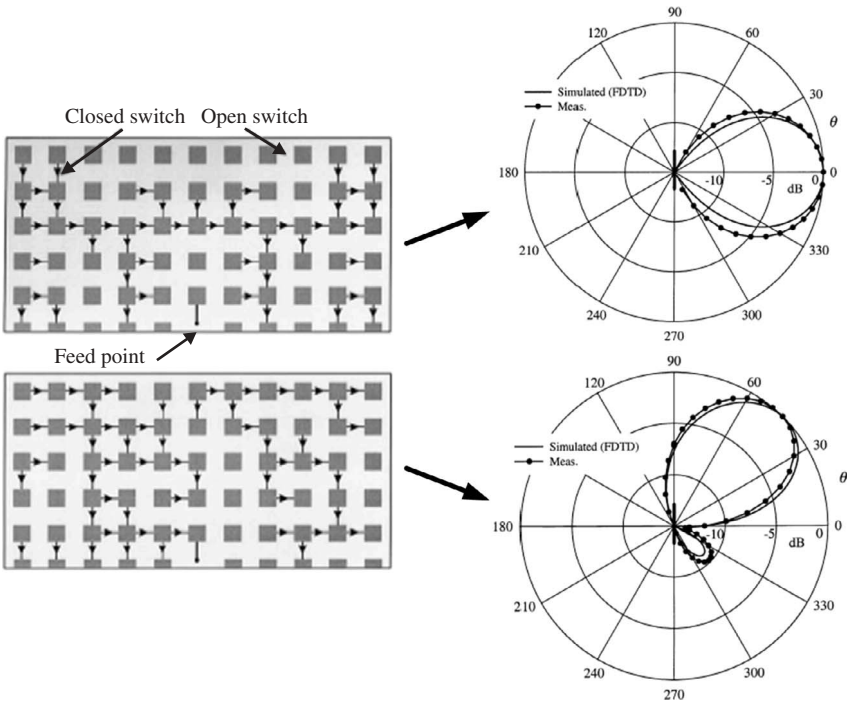


Figure 8.13 Two variations of the switched grid of metallic patches (left side): a narrowband design with broadside radiation (switched configuration at top left, resulting radiation pattern at top right) and a narrowband design with a beam steered to $\theta = 45^{\circ}$ (switched configuration at bottom left, resulting radiation pattern at bottom right). (From Ref. 18.)

desired results. The antenna remains discretely reconfigurable in this configuration since there exists a finite number of switch combinations, but a version of this antenna in Ref. 19 provides a continuously reconfigurable design.

Several different methods to realize the switched links have been examined for this structure and illustrate the many possibilities and potential difficulties that can arise when integrating the reconfiguration mechanism and providing bias and control. One such design uses FET-based electronic switches with an optically coupled control system, implemented in Ref. 18 for a structure operating at UHF (0.85–1.45 GHz). Figure 8.14 illustrates this system.

Switched Array of Crossed Slots The switched array of crossed slots in Ref. 20 (*frequency, discrete, direct*) represents another large aperture design but shares similar properties with the arrays discussed in the continuously reconfigurable section. In this case, however, the elements are designed to reconfigure their frequency and remain closely spaced ($<\lambda/2$) in the array to avoid grating lobes. Crossed slot elements—with dimensions that are altered by the use of switches—provide the repeating elements in this array that reconfigure their frequency. Figure 8.15 demonstrates an array configuration using two sets of elements that are interleaved in the array to provide the desired range of frequencies. The slots and switches reside on the top of a multilayer structure comprising

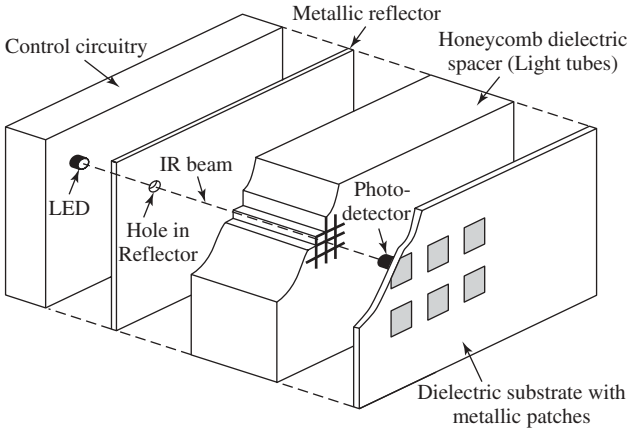


Figure 8.14 The FET-based electronic switches with an optically coupled control system for the grid of electrically small patches. (From Ref. 18.)

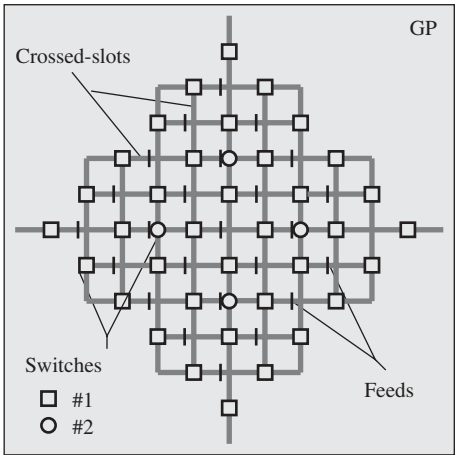


Figure 8.15 The switched, reconfigurable, crossed slot array. (Adapted from Ref. 20.)

the antenna. A series of passive, resistively loaded frequency selective surface elements below the slots and switches act as a broadband ground plane (and another component of the overall design). The broadband nature of the ground plane structure in Ref. 20 facilitates its operation between 0.8 and 3.2 GHz.

8.3.2.2 Traveling-Wave Antenna

Periodic Grating Antenna Solid-state devices are used in Ref. 21 (*radiation, discrete, direct*) to alter the electrical structure of a periodic grating antenna (similar to Refs. 8 and 9). Elongated (distributed) silicon PIN diodes are placed along the sidewall of this antenna to enable reconfiguration, occurring when current is passed through the diodes to excite the valence electrons in the silicon to the conduction band. The resulting

change in conductivity creates an effect similar to the placement of a movable metallic wall. Through field confinement, the guided wavelength can be altered. The resulting antenna demonstrated a beam switching of 9.5° (22.5° to 32.0°) by applying a current of 300 mA (1°/31.6 mA). Figure 8.16 shows the antenna and field structure resulting from the biasing of the diode. Conceptually, this antenna provides a continuous degree of radiation. However, increasing the diode current to a maximum severely deteriorates the performance due to the lossy nature of the PIN diode while transitioning to a conductive state.

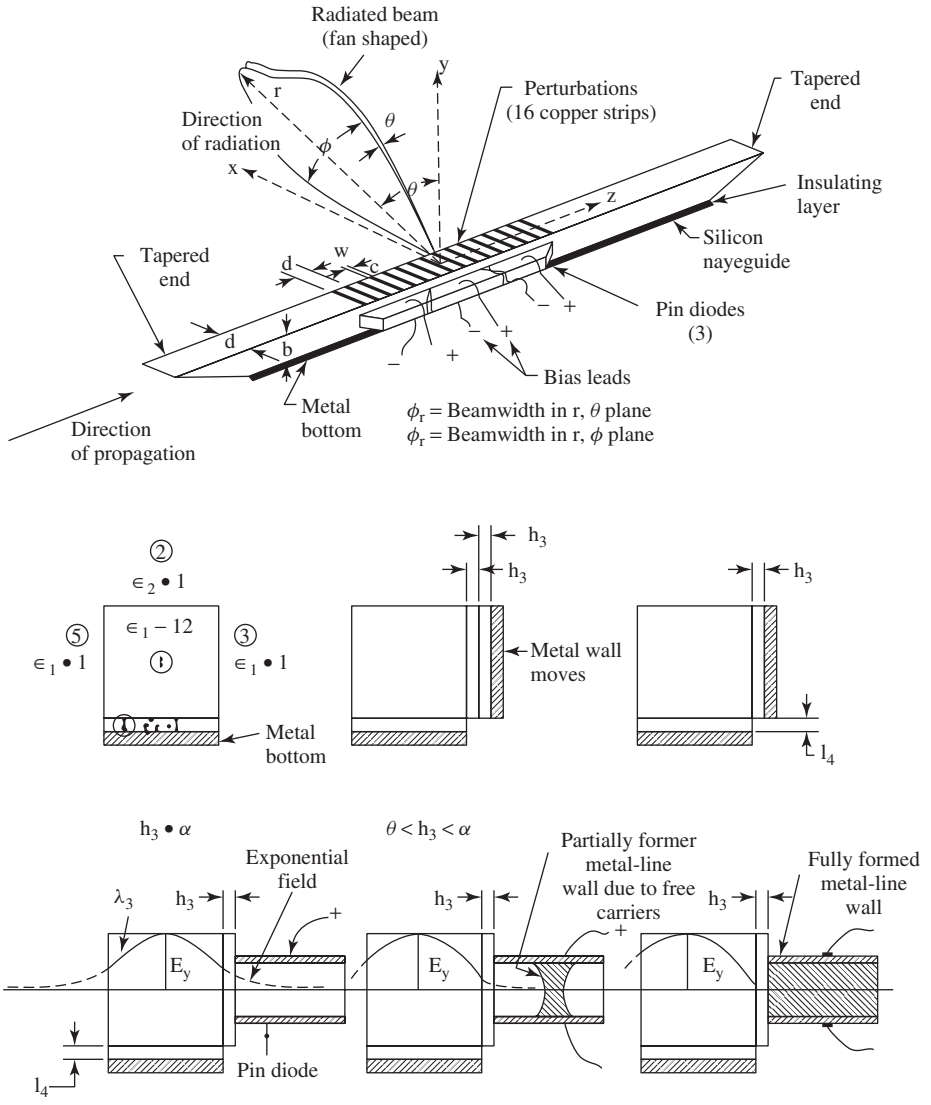


Figure 8.16 Antenna (top), PIN diode configuration (middle), and resulting field structure (bottom) of the grating antenna using elongated PIN diodes. (From Ref. 21.)

8.4 CASE STUDIES IN RECONFIGURABLE SMALL APERTURE ANTENNAS

Reconfigurable small aperture antennas (e.g., single elements) represent a significant portion of reconfigurable antenna designs developed to date. Resonant antennas and their sensitivity to small changes in the physical structure or critical material properties provide several straightforward approaches to reconfiguration. In addition, these antennas meet the needs of mobile communication devices and integration with signal combining techniques. In addition, these elements are commonly considered as candidates for the integration of RF MEMS, tunable materials and devices, and other novel technologies. The reconfiguration mechanism typically plays a direct role in the operation of the device. Advances in MEMS technologies and bulk micromachining processes have significantly enhanced the capabilities and potential of continuously reconfigurable small aperture antenna designs. As with the large aperture antennas, the organization for this section is categorized by the degree of reconfiguration, beginning with continuously reconfigurable designs and ending with discretely reconfigurable ones.

8.4.1 Continuously Reconfigurable Antennas

8.4.1.1 Micromachined Microstrip Patch Antenna The microstrip antenna's overall popularity as a base design for array elements, broadband antennas, multiband antennas, and so on makes it a natural candidate for reconfigurable designs. With a well known theoretical foundation and design equations, changes in a number of antenna dimensions and/or materials can produce antennas that are extremely flexible and diverse [1].

Vertical displacement of a microstrip antenna, changing its effective height above a substrate, is one way to achieve reconfigurability. An increase in height results in an upward shift in resonant frequency that can be demonstrated through an equivalent parallel RLC circuit model [1] (neglecting the feed reactance) of the microstrip patch antenna. This displacement increases the effective height of the antenna above the ground plane, which decreases the total capacitance in the equivalent circuit (i.e., consider the very basic relationship for a parallel-plate capacitor and the result of increasing the separation between the plates). Assuming the radiation resistance and inductance in this equivalent circuit are not significantly altered, the resonant frequency and bandwidth will experience an upward shift in frequency. The antennas presented next utilize this principle in different ways but are both based on rectangular, edge-fed, microstrip patch antennas that use the out-of-plane deflections to reconfigure their resonant frequency.

Magnetostatic Actuation The antenna in Ref. 22 (*frequency, direct, continuous*) uses the out-of-plane principles—similar to a cantilever—to reconfigure its resonant frequency and operating 2:1 voltage standing wave ratio (VSWR) bandwidth. The patch geometry was fabricated over a sacrificial layer residing on the top side of the substrate, with a thin layer of magnetic material (permalloy) electroplated onto the top surface of the antenna. The patch antenna covered in magnetic material was released from the substrate by etching away the sacrificial layer between the antenna and substrate but remains anchored to the substrate by the microstrip feed line that excites the structure. The patch can then be deflected upwards of $\theta = 45^{\circ}$—where θ is defined as the angle the antenna makes with the substrate—using magnetostatic actuation via an applied dc magnetic field and still maintain its operation in a patch mode. (Actuation beyond 45° results in behavior closer to that of a bent monopole or TEM horn antenna.) This actuation decreases the

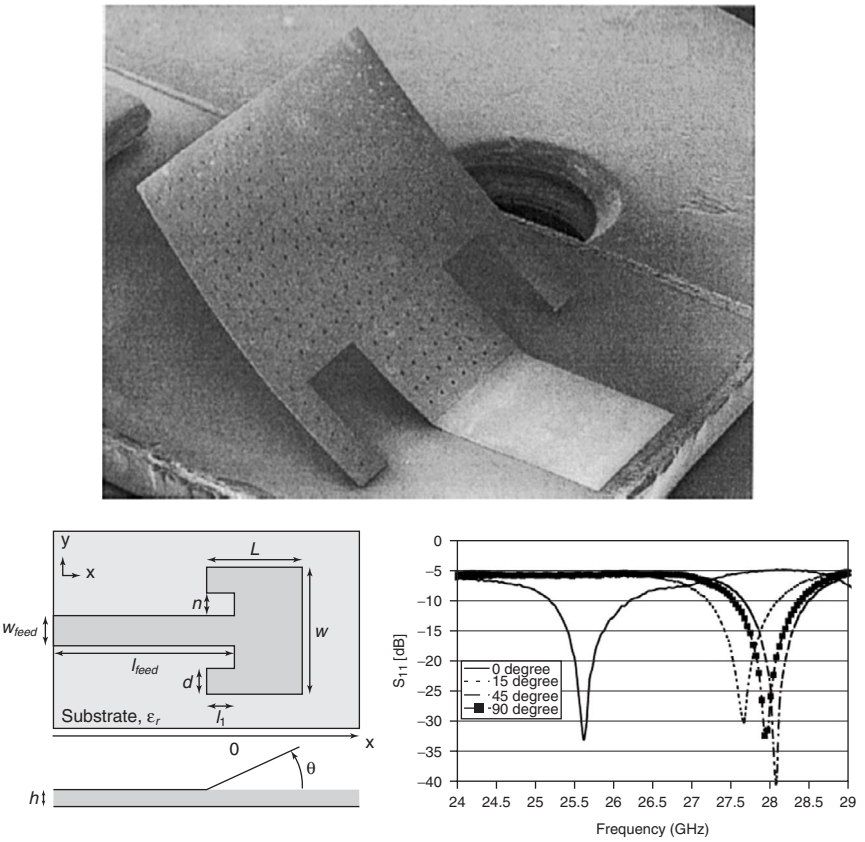


Figure 8.17 Micromachined, deflectable, patch antenna fabricated on a glass substrate using permalloy and magnetic actuation (top), design of patch antenna (bottom left), and operation (bottom right). (From Ref. 22.)

horizontal profile (effective length) of the antenna and decreases the overall capacitance, generating an upward shift in a total shift in resonant frequency of 9.4% (25.6–28.0 GHz). Figure 8.17 shows the fabricated antenna, diagram, and measured results.

Electrostatic Actuation The antenna in Ref. 23 (*frequency, direct, continuous*) utilizes out-of-plane actuation principles—applied to the ground plane of the antenna—to reconfigure its resonant frequency and operating 2:1 VSWR bandwidth. The antenna was fabricated on a thin glass substrate with a deformable metallic ground plane below the patch. The area of deformable ground plane can be electrostatically actuated downward into a cavity (etched below the center of the patch antenna) by applying an appropriate voltage between the patch and ground. This action reconfigures the antenna by increasing the overall distance between the patch and ground plane—decreasing the total capacitance of the antenna and shifting the resonant frequency upward. The actuation of the ground plane results in a maximum downward deflection of 138 μm and a total shift in resonant frequency of 6% (16.8–17.82 GHz). The fabricated antenna, cross section, and results are shown in Figure 8.18.

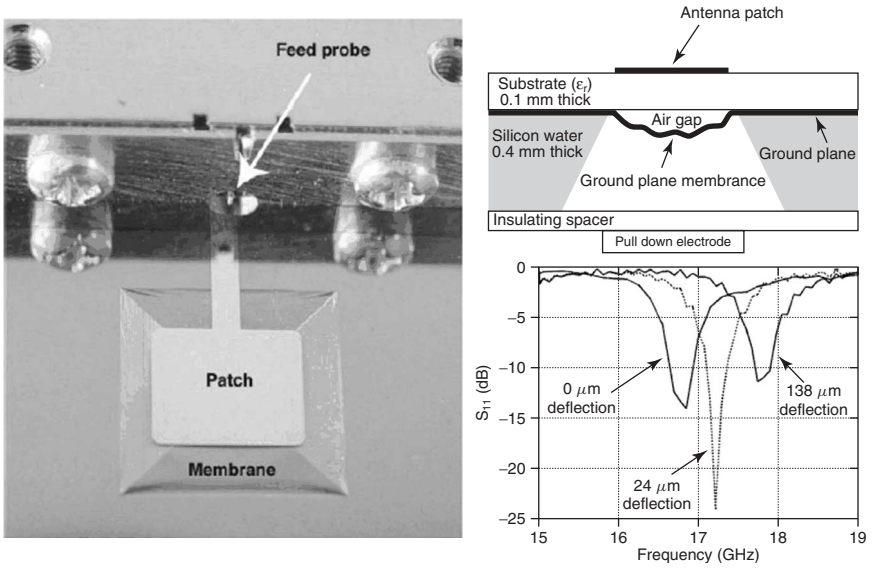


Figure 8.18 Micromachined patch antenna on glass substrate with deflectable ground plane (left), diagram of antenna (top right), and operation (bottom right). (Adapted from Ref. 23.)

8.4.1.2 Micromachined Vee-Antenna The pattern reconfigurable Vee-antenna in Ref. 24 (*radiation, direct, continuous*) (base design from Ref. 25) represents a different kind of structurally reconfigurable antenna using MEMS technologies. The arms of the Vee-antenna [24] are movable through pulling or pushing by actuators. A rotational hinge attached to the substrate holds one end of the antenna arm. This allows the arm to rotate with the hinge as the center of a circle. The antenna arms are pulled or pushed by support bars connected to the actuators with movable rotation hinges on both ends. The antenna arms and the support bars are physically connected but electrically separated with dielectric material. These arms are connected to contact pads through the fixed rotation hinges, and each one can be controlled independently. When both antenna arms move in the same direction with a fixed Vee-angle, the antenna can be used to steer the radiation beam. In addition, when the Vee-angle changes, the radiation beam shape can be adjusted. The fabricated antenna, schematic, and beam-steering results can be seen in Figure 8.19, which has demonstrated overall beam steering of 45°.

8.4.1.3 Microstrip Parasitic Array Solid-state devices are also quite useful as actuation mechanisms in reconfigurable antennas. In particular, varactors and diodes are commonly implemented to facilitate reconfiguration. They use an applied electric bias to alter their impedance, which can be used to reconfigure the frequency or radiation properties of an antenna. Figure 8.20 shows the microstrip parasitic array (*radiation, indirect, continuous*), which proposes the use of two varactors to reconfigurable the radiation pattern (discrete switching has also been demonstrated [27]).

The antenna consists of three parallel strips printed on a grounded dielectric substrate. The spacing between strips is approximately a quarter wavelength in free space. The center strip is approximately one-half of a guided wavelength at the resonant frequency

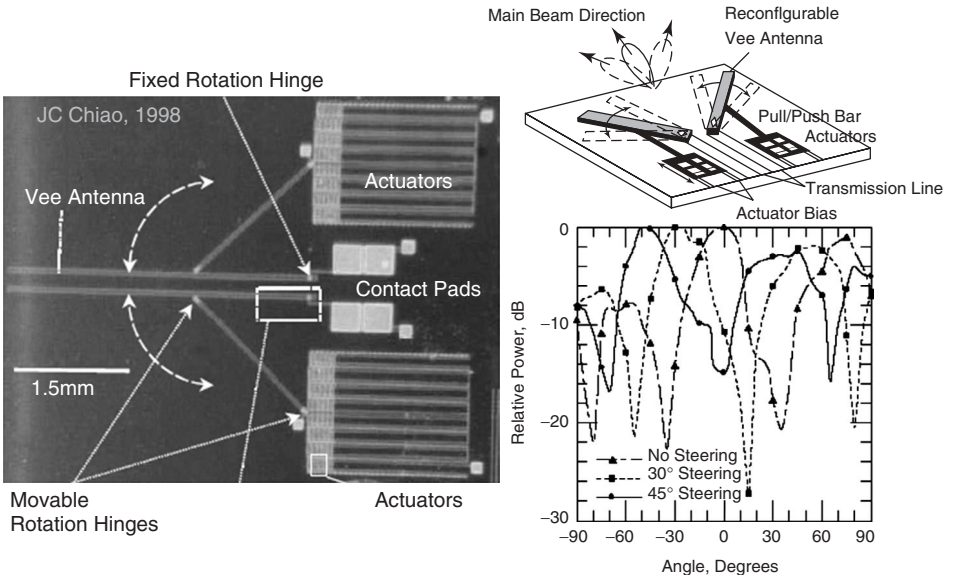


Figure 8.19 Micromachined reconfigurable Vee-antenna with actuated armature (left), diagram (top right), and reconfigured radiation patterns (bottom right). (Adapted from Ref. 24.)

and is fed with a subminiature version A (SMA) probe. The parasitic strips on either side of this center driven element can be lengthened or shortened with respect to the center strip. The antenna behaves much like a Yagi–Uda antenna set on a dielectric substrate backed by a finite ground plane. The lengthened parasitic element is inductive and works as a reflector and the shortened parasitic element is capacitive and works as a director. The effective electrical lengths of the parasitic strips can be continuously tuned by electronically controlling the value of the varactors. This effect reconfigures the radiation pattern while maintaining input impedance characteristics. The measured results for several values of capacitance and the resulting radiation patterns and tuning range are shown in Figure 8.20.

8.4.1.4 Dual-Band Bent Slot Antenna The dual-band frequency reconfigurable antenna discussed in Ref. 28 (*frequency, direct, continuous*) also utilizes two varactors. They are used to independently tune two matched-impedance bands of a slot antenna. The frequency of one of the bands can be fixed at will and the frequency of the other band can continuously be tuned over a wide frequency range. Furthermore, if desired, the frequencies of both bands can simultaneously be tuned over a wide frequency range by changing the bias voltages across the varactors. An equivalent circuit of the antenna in Figure 8.21 has been used to determine the placement of the varactors to achieve the optimum tuning range for a given range of capacitance. By adjusting the bias of the two varactors, the antenna is capable of tuning each band independently or simultaneously, shown in Figure 8.21. The antenna uses a simple feed structure and matching network to achieve simultaneous impedance match at both frequencies of operation and over its entire tuning range. Furthermore, the antenna shows similar radiation patterns and polarization at both bands and over tuning, with low levels of cross-polarized radiation.

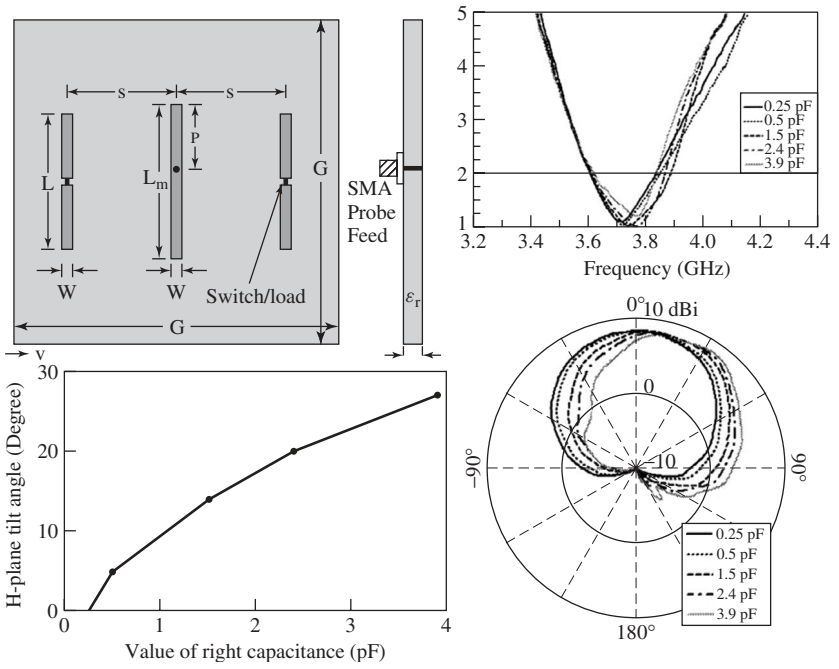


Figure 8.20 The radiation reconfigurable microstrip parasitic array designed for varactors (upper left), measured VSWR for different values of capacitance (upper right), resulting radiation patterns (bottom right), and summary of tilt angles for values of capacitance (bottom left). (From Ref. 26.)

8.4.2 Discretely Reconfigurable Antennas

8.4.2.1 Fractal Antenna A fractal antenna provides many different opportunities to create reconfigurable antennas. The antenna shown in Figure 8.22 uses an iteration of the Sierpinski gasket, with resistive pad control line structures shown in Figure 8.23. The antenna's adjacent triangular patches are connected with four cantilever ohmic contact RF MEMS switches in series to obtain the desired frequency reconfigurability. The introduction of these electrostatic actuators permits the physical connection/disconnection of sections of the antenna's conductive parts to change the antenna's effective length.

A bow-tie mode of operation is created when all switches are in an OFF position. A MEMS-enabled (or fractal) mode of operation is obtained when all switches are in an ON state. Switching between all-OFF and all-ON configurations can reconfigure the operating frequency.

8.4.2.2 Magnetic Line Source Antenna The magnetic line source antenna in Refs. 30, 31 (*radiation, direct, discrete*) uses switched connections to significantly reconfigure the radiation pattern while maintaining a common impedance bandwidth and polarization. The operation of the antenna in Ref. 30 can be synthesized from different microstrip patch antenna modes (shown in Figure 8.24). The design-oriented analysis treats the structure as an equivalent magnetic current distribution (similar to the cavity model of the patch antenna). Commercially available packaged RF MEMS switches are used to connect/disconnect the sections of microstrip line, which change the current

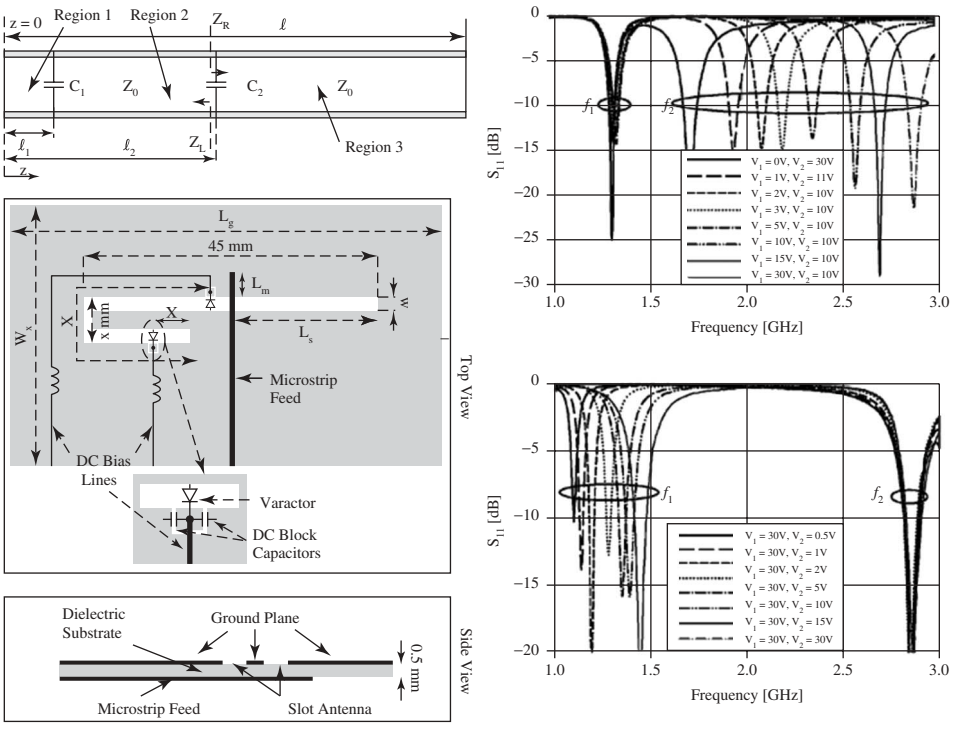


Figure 8.21 Equivalent circuit (top left) and antenna diagram (bottom left) of the dual-band reconfigurable frequency slot antenna with reconfiguration of upper band (top right) and lower band (bottom right). (From Ref. 28.)

distribution and resulting electric field distribution. The resulting broadside and end-fire radiation pattern configurations are achieved over a common impedance bandwidth. Figure 8.25 shows the antenna, biasing structures, and control lines, including lumped elements and distributed components.

8.4.2.3 Microstrip Patch Antenna Solid-state switching mechanisms are commonly used in reconfigurable antenna designs. They can easily be mass produced so their cost is minimal when compared to RF MEMS. Their parasitic packaging effects and leakage current can sometimes degrade the performance at higher operating frequencies, so the choice of a proper device is crucial. One example of a microstrip patch antenna equipped with PIN diodes is demonstrated in Refs. 32 and 33 with the patch antenna with switchable slots (PASS). The antenna's resonant length L [Eq. (8.3)] has been altered in all of these designs. The antenna is capable of reconfiguring its frequency using one diode and its polarization handedness using two diodes. An example of an antenna (including biasing structures) using this type of reconfiguration can be seen in Figure 8.26.

8.4.2.4 Multimode Microstrip Antenna PIN diodes can also be used to create much more elaborate structures based on the operation of the microstrip patches and transmission lines. The base design of the antenna in Ref. 34 is a leaky-wave microstrip antenna (a large aperture antenna), but it has been segmented into several smaller patch

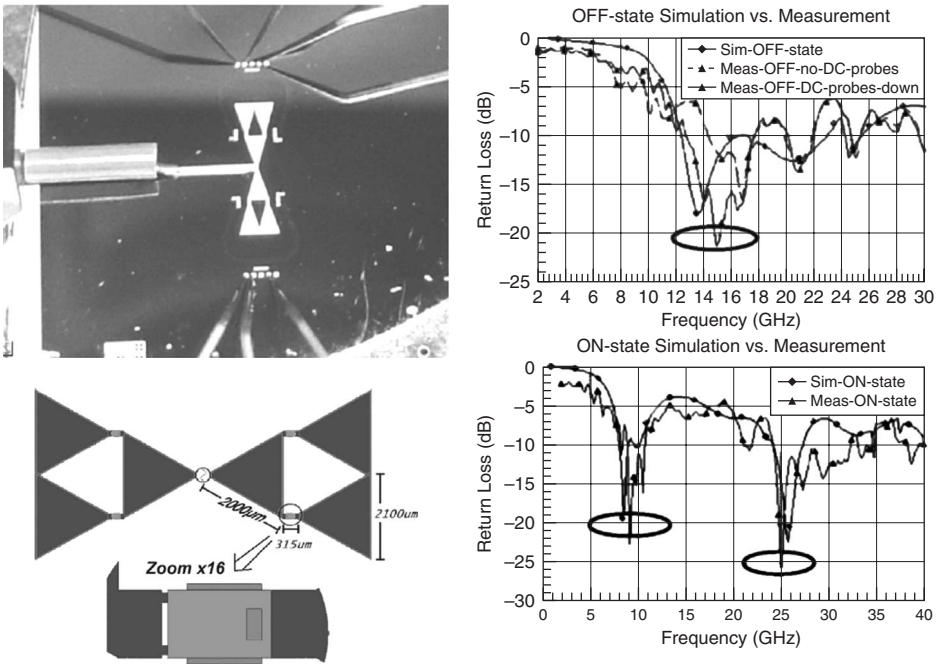


Figure 8.22 Fabricated RF MEMS-based self-similar reconfigurable antenna (top left), antenna and switch diagram (bottom left), and reconfigured frequency for two switch states (top and bottom right). (From Ref. 29.)

Figure 8.23 High impedance bias pads for RF MEMS-based self-similar reconfigurable antenna (left) and interface with the switch (right). (From Ref. 29.)

antennas. The multimode structure can be reconfigured into a patch antenna anywhere along the aperture of the leaky-wave antenna with two degrees of freedom. PIN diode switches are utilized to switch between the different aperture configurations. The structure's unique field profile is utilized to minimize insertion loss in the leaky-wave mode and also to maximize isolation between the different aperture ports. The antenna is shown in Figure 8.27.

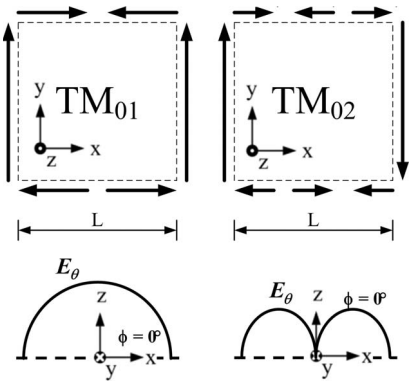


Figure 8.24 Magnetic current distributions for the TM_{01} and TM_{02} modes of a rectangular microstrip patch antenna (resonant length L). (Adapted from Ref. 29.)

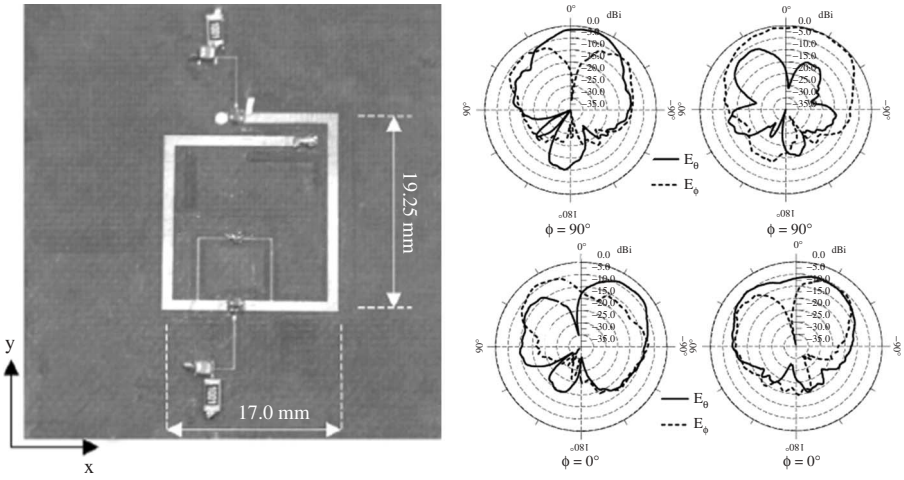


Figure 8.25 Fabricated reconfigurable radiation magnetic line source antenna (left) and broadside operation (top right) and endfire operation (bottom right). The reconfigured radiation pattern is θ-polarized in the $\phi = 0^\circ$ plane. (Adapted from Ref. 31.)

8.4.2.5 Printed Dipole The optically reconfigurable printed dipole antenna in Ref. 35 demonstrates an alternative use of solid-state devices by optically switching a set of high resistivity silicon switches. Its operation is analogous to the examples provided in Sections 8.2.2 and 8.2.3. When optically illuminated, the silicon changes from an insulator state to a near conducting state by creating electron–hole pairs by virtue of the photoelectric effect, and the antenna is configured for its lowest frequency of operation. Removing the light source returns the silicon to its previous state and the antenna to its higher operating frequency. The reliability of the switch is governed only by the properties of the light source and has no moving parts. This method of reconfiguration also provides the benefit of electrical isolation from the driven antenna circuit. Figure 8.28 shows the fabricated antenna and measured results.

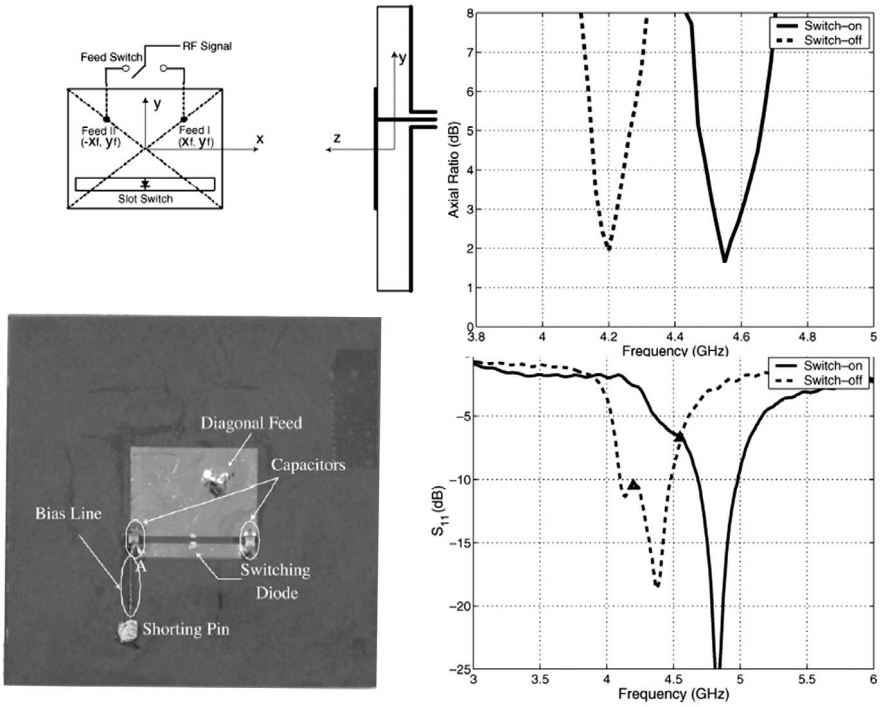


Figure 8.26 PASS diagram (top left), fabricated antenna with reconfigurable frequency and circular polarization using PIN diodes (bottom left), axial ratio of both bands (top right), and return loss of both bands (bottom right). (From Ref. 33.)

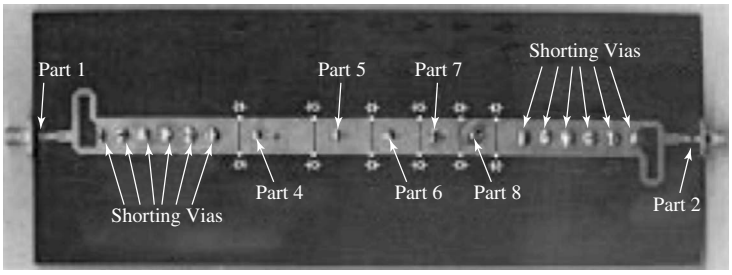


Figure 8.27 Multimode microstrip antenna using PIN diodes. (From Ref. 34.)

8.5 ADVANCED RECONFIGURABLE ANTENNA APPLICATIONS

Reconfigurable antennas are designed to integrate the performance of several antennas into a single radiating structure. Multifunction systems, ad hoc networks, and software-defined radios are three broad applications that can utilize these properties. Example 2 from Section 8.22 represents a very basic multifunction system, but many others with more complicated requirements could be possible through the use of reconfigurable antennas. In the following sections, multiple antenna techniques and array techniques are

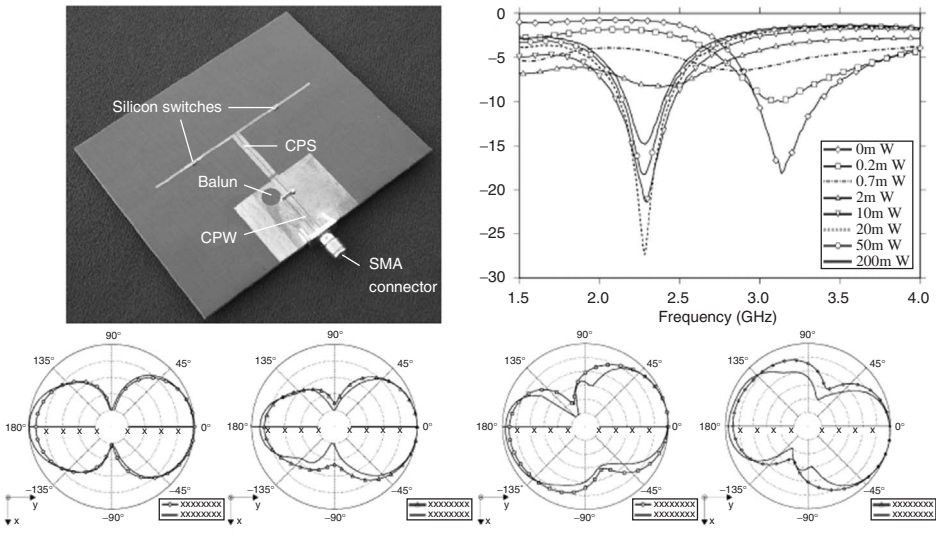


Figure 8.28 Optically switched reconfigurable printed dipole antenna (top left), return loss versus input optical power (top right), and radiation patterns for the two symmetric modes (bottom left) and the two asymmetric modes (bottom right). (Adapted from Ref. 35.)

considered as two examples of advanced systems that could be enabled with reconfigurable antennas.

8.5.1 Multiple Antenna Techniques

Signal combining or diversity techniques [36–39] and multiple-input multiple-output (MIMO) [40–42] systems represent some of the evolving uses of reconfigurable antennas. The additional degrees of freedom and functionality that reconfigurable antennas provide make them excellent candidates for these applications. The ultimate goal of these systems is to raise (or improve) the signal-to-noise ratio, thereby improving the bit error rate, increasing the channel capacity or system throughput, and reducing the power consumption in mobile networks. Polarization and pattern diversity are particularly useful for these applications, which rely on antennas that can alter their spatial properties and maintain their impedance match. While simple diversity schemes with fixed antennas are now prevalent, the ability to reconfigure radiation patterns with single elements can provide additional performance benefits in complex communication environments while keeping the number of antennas on a platform reasonable. The mutual coupling between antennas and superposition of far-field radiation patterns from reconfigurable antennas can also be used for beam forming with a small number of elements.

8.5.2 Array Techniques

Small aperture reconfigurable frequency and reconfigurable radiation antennas can also provide additional degrees of freedom in the design of arrays. Their impact depends on both the reconfigured properties and the array topology, but their use can result in significant enhancements or improvements in the performance of the array. In addition

to the operation of the array, reconfigurable antennas may have the added benefit of reducing the cost and complexity of the overall system supporting the array. For example, the ability of a single element to reconfigure the phase of its radiated fields can eliminate or reduce phase shifting networks—resulting in a lighter weight and lower cost design. Another example is the use of single elements to reconfigure their radiation patterns across an array to complement traditional amplitude and phase tapering to reduce sidelobes and increase gain.

Two classical examples of this behavior—typically reconfigured radiation—were originally demonstrated in Refs. 43 and 44 with the use of reactively controlled antenna arrays. In both of these examples (which are the fundamental basis for many reconfigurable antenna designs using parasitic elements), the fields from excitation of a single element are coupled to adjacent elements with reactive impedance control (e.g., the reconfiguration mechanism). Figure 8.29 shows the array configuration on the left and a set of radiation patterns on the right for the first of these examples, in which parasitic elements (dipoles) with reactive impedance control (denoted as X) at the feed point surround a driven element to form a circular dipole array. The radiation pattern from this configuration can then be steered around the plane of the array by altering the reactive impedance of the parasitic elements. The second example can be seen in Figure 8.30 (configuration on the left and resulting radiation patterns on the right), where energy from a driven waveguide couples into closely spaced adjacent radiators and—through the adjustment of a shorting wall in the parasitic elements—alters the phase so the linear array can steer its beam along the plane of the array from endfire to broadside. In both of these examples, the driven element has been centrally located and the required reactive impedances have been calculated a priori (a variety of means can provide this information). In general, these driven elements do not have to be centered, and a variety of reactive impedance control configurations can be used.

Conversely, other advanced array techniques can be used to maintain the main beam direction when the reconfigured frequency bands are widely spaced. The crossed slot array [20] (previously discussed in Section 8.3.2.1) is specifically designed to avoid grating lobes over its wide bandwidth. It accomplishes this by providing a closely spaced array. Random arrays [45–49], fractal (self-similar) arrays [50–52], and other techniques [53, 54] are methods that can be used to decrease the effects that are associated with

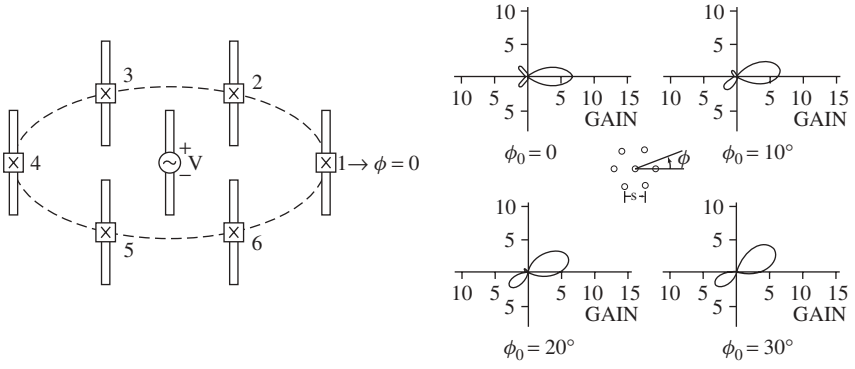


Figure 8.29 Reactively controlled dipole array from Ref. 43. The circular array configuration can be on the left (centrally located driven element) and an example of the radiation pattern control (beam steering) can be seen on the right.

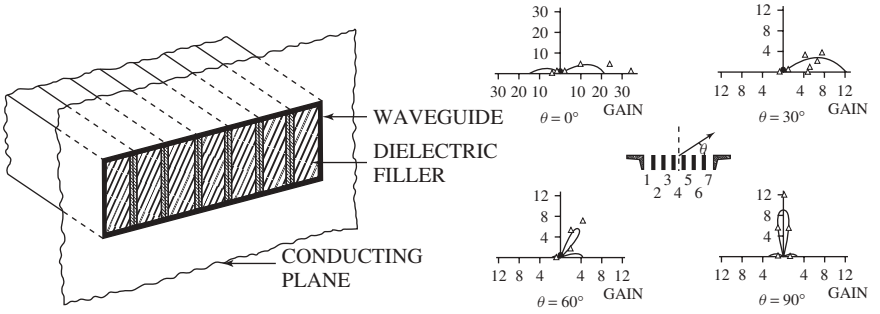


Figure 8.30 Reactively controlled waveguide array from Ref. 44. The linear array configuration can be on the left (centrally located driven element) and an example of the radiation pattern control (beam steering) can be seen on the right.

periodic array behavior (grating lobes, sidelobe level, scanning, etc.). The random array has many benefits and is typically the most robust of these methods over frequency. These arrays are typically less populated than traditional arrays—providing the added benefit of reduced interelement coupling—but the increased functionality comes at the price of reduced gain.

8.6 SUMMARY

This chapter provides a conceptual description of reconfigurable antennas, a discussion on reconfigurable antenna parameters, and practical considerations for their design. Case studies in both reconfigurable large and small aperture antennas serve as examples of the kinds of functionality that can be achieved with a very diverse set of topologies and technologies. Finally, advanced reconfigurable antenna applications for the future are discussed.

Reconfigurable antennas provide additional functionality and flexibility at the expense of increased complexity. Innovative designs can incorporate state-of-the-art technologies to interact with the radiating mechanism and facilitate reversible changes. Reconfigurable antennas can provide new degrees of freedom that will affect how communication, sensing, and radar systems are specified and designed in the future. New approaches may require interdisciplinary solutions to the design of electrical, mechanical, and material components, as well as decision and control circuitry and protocols, for these systems. This new level of integration with system operation has the potential to dramatically increase system performance while providing new incentives to reduce complexity and cost.

REFERENCES

1. S. W. Lee and Y. T. Lo (Eds.), *Antenna Handbook, Theory, Applications, and Design*, Van Nostrand Reinhold, New York, 1988.
2. J. R. Harrington, *Time Harmonic Electromagnetic Fields*, McGraw-Hill, New York, 1961.

3. W. Rotman and N. Karas, Some new microwave antenna designs based on the trough waveguide, *IRE Int. Conv. Rec.*, Vol. 4, pp. 230–235, March 1956.
4. W. Rotman and A. Oliner, Asymmetrical trough waveguide antennas, *IEEE Trans. Antennas Propag.*, Vol. 7, pp. 153–162, April 1959.
5. W. Rotman and A. Maestri, An electromechanically scannable trough waveguide array, *IRE Int. Conv. Rec.*, Vol. 8, pp. 67–83, March 1960.
6. R. Elliott, Spherical surface-wave antennas, *IEEE Trans. Antennas Propag.*, Vol. 4, pp. 422–428, July 1956.
7. R. Honey, A flush-mounted leaky-wave antenna with predictable patterns, *IEEE Trans. Antennas Propag.*, Vol. 7, pp. 320–329, October 1959.
8. V. K. Varadant, V. V. Varadant, K. A. Joset, and J. F. Kelly, Electronically steerable leaky wave antenna using a tunable ferroelectric material, *Smart Materials Structures*, Vol. 10, pp. 470–475, December 1994.
9. H. Maheri, M. Tsutsumi, and N. Kumagai, Experimental studies of magnetically scannable leaky-wave antennas having a corrugated ferrite slab/dielectric layer structure, *IEEE Trans. Antennas Propag.*, Vol. 36, pp. 911–917, July 1988.
10. P. J. B. Clarricoats, Z. Hai, R. C. Brown, G. T. Poulton, and G. E. Crone, A reconfigurable mesh reflector antenna, in *Proceedings of the 1998 IEEE International Symposium on Antennas and Propagation*, San Jose, CA, April 1989, Vol. 1, pp. 112–116.
11. P. J. B. Clarricoats and H. Zhou, Design and performance of a reconfigurable mesh reflector antenna. 2. Antenna design, *Proc. IEE Microwaves Antennas Propag.*, Vol. 138, pp. 485–492, December 1991.
12. P. J. B. Clarricoats and H. Zhou, Design and performance of a reconfigurable mesh reflector antenna. 2. Antenna performance, *Proc. IEE Microwaves Antennas Propag.*, Vol. 138, pp. 493–496, December 1991.
13. R. G. Brown, A reconfigurable reflector using hinged panels, in *Proceedings of the IEE International Conference on Antennas and Propagation (ICAP)*, April 1991, Vol. 1, pp. 15–18.
14. W. H. Theunissen, H. T. Yoon, G. N. Washington, and W. D. Burnside, Mechanical finite element diffraction synthesis of reconfigurable contour beams from dual offset reflector antennas, in *Proceedings of the IEEE International Symposium on Antennas and Propagation*, July 1999, Vol. 4, pp. 2348–2351.
15. J. P. Gianvittorio and Y. Rahmat-Samii, Reconfigurable patch antennas for steerable reflectarray applications, *IEEE Trans. Antennas Propag.*, Vol. 54, pp. 1388–1392, May 2006.
16. L. N. Pringle, P. Friederich, L. Fountain, P. Harms, D. Denison, E. Kuster, S. Blalock, R. Prado, G. Kiesel, G. Smith, M. Allen, K. Kim, J. Maloney, and M. Kesler, Architecture and performance of a reconfigurable aperture, in *Proceedings of the 2001 Antenna Applications Symposium*, Monticello, IL, September 2001, pp. 36–58.
17. J. G. Maloney, M. P. Kesler, L. M. Lust, L. N. Pringle, T. L. Fountain, P. H. Harms, and G. S. Smith, Switched fragmented aperture antennas, in *Proceedings of the IEEE International Symposium on Antennas and Propagation*, Salt Lake City, UT, July 2000. CD-ROM.
18. L. N. Pringle, P. H. Harms, S. P. Blalock, G. N. Kiesel, E. J. Kuster, P. G. Friederich, R. J. Prado, J. M. Morris, and G. S. Smith, A reconfigurable aperture antenna based on switched links between electrically small metallic patches, *IEEE Trans. Antennas Propag.*, Vol. 52, pp. 1434–1445, June 2004.
19. D. Sievenpiper, J. Schaffner, J. J. Lee, and S. Livingston, A steerable leaky-wave antenna using a tunable impedance ground plane, *Antennas Wireless Propag. Lett.*, Vol. 1, pp. 179–182, June 2002.
20. Y. E. Erdemli, R. A. Gilbert, and J. L. Volakis, A reconfigurable slot aperture design over a broad-band substrate/feed structure, *IEEE Trans. Antennas Propag.*, Vol. 52, pp. 2860–2870, November 2004.

21. R. E. Horn, H. Jacobs, K. L. Klohn, and E. Freibergs, Single-frequency electronic-modulated analog line scanning using a dielectric antenna, *IEEE Trans. Microwave Theory Tech.*, Vol. 82, pp. 816–820, May 1982.
22. J.-C. Langer, J. Zou, C. Liu, and J. T. Bernhard, Micromachined reconfigurable out-of-plane microstrip patch antenna using plastic deformation magnetic actuation, *IEEE Microwave Wireless Components Lett.*, Vol. 13, pp. 120–122, March 2003.
23. R. Al-Dahleh, C. Shafai, and L. Shafai, Frequency agile microstrip patch antenna using reconfigurable MEMS ground plane, *Microwave Optical Technol. Lett.*, Vol. 43, pp. 64–67, October 2004.
24. J.-C. Chiao, S.-Y. Cheng, J. J. L. Chang, I. M. Chio, Y. Kang, and Junichi Hayasaka, MEMS reconfigurable antennas, *Int. J. RF Microwave Computer-Aided Eng.*, Vol. 11, pp. 301–309, September 2001.
25. K. Iizuka, The traveling-wave V-antenna and related antennas, *IEEE Trans. Antennas Propag.*, Vol. 15, pp. 236–243, March 1967.
26. S. Zhang, G. H. Huff, G. Cung, and J. T. Bernhard, Three variations of a pattern-reconfigurable microstrip parasitic array, *Microwave Optical Technol. Lett.*, Vol. 45, pp. 369–372, June 2005.
27. S. Zhang, G. H. Huff, J. Feng, and J. T. Bernhard, A pattern reconfigurable microstrip parasitic array, *IEEE Trans. Antennas Propag.*, Vol. 52, pp. 2773–2776, October 2004.
28. N. Behdad and K. Sarabandi, Dual-band reconfigurable antenna with a very wide tunability range, *IEEE Trans. Antennas Propag.*, Vol. 52, pp. 409–416, February 2004.
29. D. E. Anagnostou, Z. Guizhen Zheng, M. T. Chryssomallis, J. C. Lyke, G. E. Ponchak, J. Papapolymerou, and C. G. Christodoulou, Design, fabrication, and measurements of an RF-MEMS-based self-similar reconfigurable antenna, *IEEE Trans. Antennas Propag.*, Vol. 54, pp. 422–432, February 2006.
30. G. H. Huff, A radiation reconfigurable magnetic line source antenna: modeling, integration with RF MEMS, and applications, Ph.D. dissertation, University of Illinois at Urbana-Champaign, 2006.
31. G. H. Huff and J. T. Bernhard, Integration of packaged RF MEMS switches with radiation pattern reconfigurable square spiral microstrip antennas, *IEEE Trans. Antennas Propag.*, Vol. 54, pp. 464–469, February 2006.
32. F. Yang and Y. Rahmat-Samii, Patch antennas with switchable slots (PASS): concept, theory and applications, in *Proceedings of the 2001 Antenna Applications Symposium*, Monticello, IL, September 2001, pp. 84–103.
33. F. Yang and Y. Rahmat-Samii, A novel patch antenna with switchable slot (PASS): dual-frequency operation with reversed circular polarizations, *IEEE Trans. Antennas Propag.*, Vol. 54, pp. 1031–1034, March 2006.
34. J. Sor, C.-C. Chang, Y. Qian, and T. A. Itoh, A reconfigurable leaky-wave/patch microstrip aperture for phased-array applications, *IEEE Trans. Microwave Theory Tech.*, Vol. 50, pp. 1877–1884, August 2002.
35. C. J. Panagamuwa, A. Chauraya, and J. C. Vardaxoglou, A frequency and beam reconfigurable antenna using photoconducting switches, *IEEE Trans. Antennas Propag.*, Vol. 54, pp. 449–454, February 2006.
36. M. T. Oswald, S. C. Hagness, B. D. Van Veen, and Z Popovic, Reconfigurable single-feed antennas for diversity wireless communications, in *Proceedings of the IEEE International Symposium on Antennas and Propagation*, July 2002, Vol. 1, pp. 469–472.
37. A. Grau,, S. Liu, B. A. Cetiner, and F. De Flaviis, Investigation of the influence of antenna array parameters on adaptive MIMO performance, in *Proceedings of the IEEE International Symposium on Antennas and Propagation*, June 2004, Vol. 2, pp. 1704–1707.
38. D. Anderson, E. Fotheringham, D. Popovic, O. Popovic, S. Romish, and P. Smith, Smart and multibeam diversity antenna arrays with high-bandwidth analog signal processing, in

Proceedings of the IEEE Topical Conference on Wireless Communication Technology, October 2003, pp. 86–89.

39. T. L. Roach, G. H. Huff, and J. T. Bernhard, Enabling high performance wireless communication systems using reconfigurable antennas, in *Military Communication Conference Proceedings (MILCOM)*, October 2006.
40. D. Piazza and K. R. Dandekar, Reconfigurable antenna solution for MIMO-OFDM systems, *Electron. Lett.*, Vol. 42, pp. 446–447, April 2006.
41. A. Grau, J. Romeu, L. Jofre, and L. De Flaviis, On the polarization diversity gain using the ORIOL antenna in fading environments, in *Proceedings of the IEEE International Symposium on Antennas and Propagation*, Monterey, CA, July 2005, pp. 14–17.
42. B. A. Cetiner, H. Jafarkhani, J.-Y. Qian, H. J. Yoo, A. Grau, and F. De Flaviis, Multifunctional reconfigurable MEMS integrated antennas for adaptive MIMO systems, *IEEE Commun. Mag.*, Vol. 42, pp. 62–70, December 2004.
43. R. Harrington, Reactively controlled directive arrays, *IEEE Trans. Antennas Propag.*, Vol. 26, pp. 390–395, 1978.
44. J. Luzwick and R. Harrington, A reactively loaded aperture antenna array, *IEEE Trans. Antennas Propag.*, Vol. 26, pp. 543–547, 1978.
45. Y. T. Lo, A probabilistic approach to the design of large antenna arrays, *IEEE Trans. Antennas Propag.*, Vol. 11, pp. 95–96, January 1963.
46. Y. T. Lo, A mathematical theory of antenna arrays with randomly spaced elements, *IEEE Trans. Antennas Propag.*, Vol. 12, pp. 257–268, May 1965.
47. Y. T. Lo and R. Simcoe, An experiment on antenna arrays with randomly spaced elements, *IEEE Trans. Antennas Propag.*, Vol. 15, pp. 231–235, March 1967.
48. V. D. Agrawal and Y. T. Lo, Distribution of sidelobe level in random arrays, *IEEE Trans. Antennas Propag.*, Vol. 57, pp. 1764–1765, October 1960.
49. K. C. Kerby and J. T. Bernhard, Sidelobe level and wideband behavior of arrays of random subarrays, *IEEE Trans. Antennas Propag.*, Vol. 54, pp. 2253–2262, August 2006.
50. C. Puente-Baliarda and R. Pous, Fractal design of multiband and low side-lobe arrays, *IEEE Trans. Antennas Propag.*, Vol. 54, pp. 730–737, May 1996.
51. D. H. Werner and R. L. Haupt, Fractal constructions of linear and planar arrays, in *Proceedings of the IEEE International Symposium on Antennas and Propagation*, Monterey, CA, July 1997, Vol. 3, pp. 1968–1968.
52. D. H. Werner, M. A. Gingrich, and P. L. Werner, A self-similar fractal radiation pattern synthesis technique for reconfigurable multiband arrays, *IEEE Trans. Antennas Propag.*, Vol. 51, pp. 1486–1498, July 2003.
53. R. J. Mailloux, S. G. Santarelli, and T. M. Roberts, Irregular shaped subarrays for time domain delay control of planar subarrays, in *Proceedings of the 2001 Antenna Applications Symposium*, Monticello, IL, September 2001, pp. 280–292.
54. R. J. Mailloux, S. G. Santarelli, and T. M. Roberts, Polyomino shaped subarrays for limited field of view and time delay control of planar arrays, in *Proceedings of the 2005 Antenna Applications Symposium*, Monticello, IL, September 2005, pp. 268–288.

CHAPTER 9

Wideband and Traveling-Wave Antennas

LOTFOLLAH SHAFAI and SIMA NOGHANIAN

9.1 INTRODUCTION

Wideband antennas refer to a category of antennas with a relatively constant performance over a wide frequency band. Historically, this referred to an octave or more. However, this is a general statement as an antenna has several electrical parameters like the input impedance, gain, polarization, sidelobe level, loss, and aperture efficiencies. This is due to the fact that an antenna can have very diverse applications and its desirable parameters can vary significantly. Even the size of its bandwidth can depend on the application and the term broadband can mean a different frequency range for different applications. Similar difficulties can also be experienced in considering a specific antenna type, where the bandwidth can depend on the design goals. For instance, a microstrip antenna can be narrowband in one design and wideband in another. Thus the antenna bandwidth definition, and classification of antennas using the bandwidth, can be confusing and case dependent. Problems also exist in the antenna types considered in this chapter, namely, biconical, helical, and Yagi antennas. Not only will their relative performance in bandwidth be different, but they can also be designed to have many different bandwidths. Therefore the material in this chapter is presented to provide a broad spectrum of these antennas, rather than only those with broadest bandwidths. Adequate examples are also provided to understand the potential bandwidth performance for different designs.

Section 9.2 presents biconical antennas. As an infinite structure, this antenna is truly broadband. However, its size and shape are impractical in real applications, and a truncation of its length limits its bandwidth performance. Modifying its rotationally symmetric shape to planar ones also causes a similar effect. Thus numerous design changes are considered and perfected to improve the modified antenna shapes and maintain the wideband characteristics. Representative examples are presented and discussed. The family of biconical antennas is generally broadband in input impedance. However, their radiation characteristics change with frequency. Therefore the design challenges are in maintaining the radiation performance satisfactory within their impedance bandwidth.

The second antenna type is the helical antenna. A helical antenna is a quasi-broadband antenna, since its geometry is angular dependent, which is the main requirement for frequency-independent antennas. However, it has a finite length that limits its bandwidth.

Modern Antenna Handbook. Edited by Constantine A. Balanis

It is shown that the antenna has two distinct parts. One part is its input end, the first couple of turns, which acts as a transducer and converts the input electrical power to radiated wave. For this reason, this section is known as the launcher section. The remaining turns act as the directors and guide the wave energy. As one might expect, the launcher section primarily influences the antenna input parameters and its impedance bandwidth. The director section mostly controls the radiation characteristics. This knowledge facilitates the antenna design and optimization. The rotational nature of the helix geometry can also be used to generate multifilar helices, which offers additional benefits in both input impedance and radiation characteristics. These antennas are also discussed, and we show that they can be designed for better performance or geometrical simplicity. The special case is the quadrifilar helix design, which can provide diverse performance ranges and thus is used in many applications. It is discussed only briefly, as historically it was not a wideband antenna. Methods for broadening its impedance bandwidth are also presented.

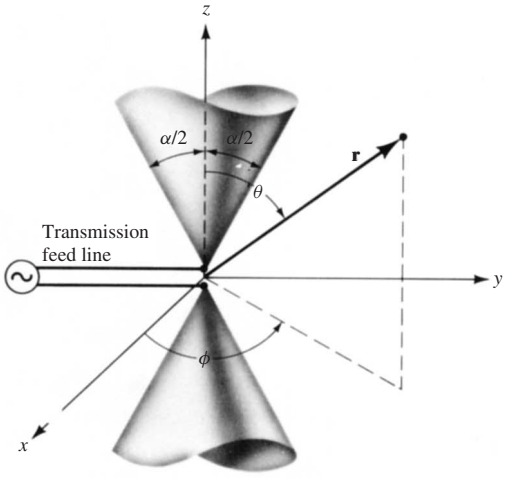
The third antenna discussed is the Yagi–Uda antenna. It is shown that its operation is very similar to the helix antenna and can also be viewed as having two distinct parts, the launcher and the director sections. Because this antenna does not enjoy an angular geometrical character, its bandwidth is not as wide as the previous two antennas, especially in high gain applications. The gain optimization further reduces its effective bandwidth.

Finally, not all antenna types covered in this chapter are wideband or traveling. They are, however, a variation of the original configurations of the biconical, helix, and Yagi–Uda type antennas. They are included to show the results of geometrical or electrical parameter adaptations for specific applications. A good example is the high gain Yagi antenna, where the input impedance bandwidth can be maintained quite large, but the gain bandwidth reduces progressively when increasing the gain.

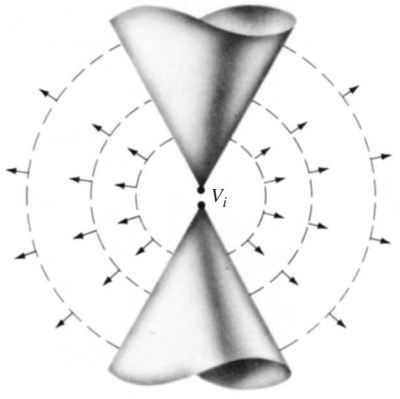
9.2 BICONICAL ANTENNAS

9.2.1 Introduction

The biconical antenna is a simple modification of a dipole antenna, where the conductor thickness linearly increases with the distance from the origin or the antenna center. Mathematically, each arm of the antenna is an infinite conducting cone, and the geometry is a rotationally symmetric structure. In a symmetric biconical antenna the cone angles for the two arms are equal, as shown in Figure 9.1 [1–3]. However, this is not a necessity and each arm can have a separate cone angle. In the latter case, when the cone angle for the lower arm increases to 180°, the geometry modifies to a conical monopole antenna. An interesting property of this antenna is that its geometry is defined and governed by angles. Thus it satisfies the main requirement of frequency-independent antennas. Thus an infinite biconical antenna is a frequency-independent antenna. In practice, it must be truncated in length, which introduces the frequency effect through the arm length and the antenna bandwidth becomes finite. As with other frequency-independent antennas, this arm truncation controls the low frequency limit. The high frequency limit is normally controlled by the feed region size, which in this case may be due to a coaxial probe. Thus the antenna bandwidth becomes a function of the arm length ratio, between its maximum and minimum lengths. Since in practical designs this ratio can be quite large, a truncated biconical antenna can be a significantly large broadband antenna.



(a) Biconical geometry



(b) Spherical waves

Figure 9.1 Geometry of biconical antennas (Ref. 1.): (a) biconical geometry and (b) spherical waves.

The major difficulty with this antenna is its volume and especially its weight, which can be significant at low frequencies. The latter can be overcome by replacing the solid surfaces by wires, similar to other conducting surface antennas, like corner reflectors. Since for the fundamental TEM mode the induced currents on surface of this antenna travel in the radial direction, the conducting wires will be radial and the geometry will modify to the wire cones shown in Figure 9.2. The rotational symmetry of these structures can cause azimuthal mode excitations that can introduce cross polarizations. In practice, the number of cones is preferable to be eight or larger, to minimize the cross-polarization effects. This means that every 45° a cone wire must be located. In the truncated cones, the sharp bends at the cone ends cause severe impedance discontinuities, and thus reflections that cause the bandwidth limitations. These reflection effects can be reduced by tapering the cone ends, rather than truncating them. Each truncated cone therefore becomes a dual back-to-back cone. However, the end cones are normally made much smaller, as shown in Figure 9.3. For the higher order modes, the currents will travel in both azimuthal and

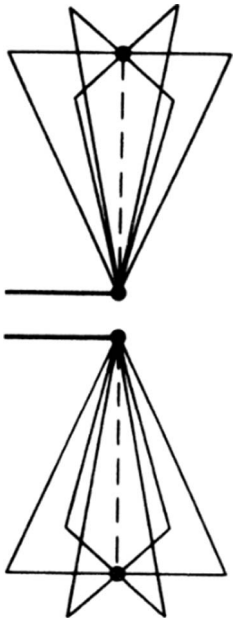


Figure 9.2 Wire mesh approximation to biconical antennas (Ref. 1.)

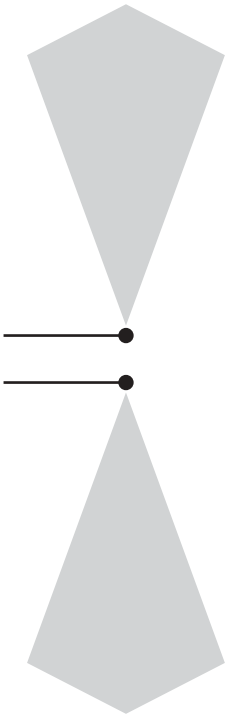


Figure 9.3 Planar double triangle approximation to biconical antennas (Ref. 1.)

radial directions and the solid cones can only be replaced by wire meshes, to allow current flow in all directions.

The problem of large cone volume is also severe and limits the antenna application. It also makes the fabrication very difficult. To remedy, a vertical cross section of the antenna is used in most applications. The cross section is planar, where the three-dimensional cone converts to a planar triangle, or double back-to-back triangles, in the case of end tapering. The geometry of this antenna is shown in Figure 9.4 and is known as the bow-tie antenna. It can be made by a variety of low cost techniques. However, as will be shown later, its bandwidth is significantly lower than the original biconical antenna. This antenna can also be made by wire frames, to reduce its weight in large designs. The conversion of the solid biconical and planar bow-tie antennas to wire grids also simplifies their numerical analysis, which can easily be done using the moment methods.

9.2.2 Solution for Radiation Fields

Once the biconical antenna is excited by a time-varying source, similar to a dipole antenna, the E- and H-field vectors travel outward from its surface. In the region between the conducting cones, these field vectors satisfy source-free wave equations. Since the cone geometry coincides with the constant θ-coordinate of the spherical coordinate system, the wave equation will best be solved in the spherical coordinate system. In addition, because of the finite cone angle, there will be an infinite number of solutions in this direction. The fundamental mode is the TEM mode, travels with the speed of light, and has no angular dependence, in neither the ϕ or θ directions. It also has only the E_θ and H_ϕ components. Thus Maxwell equations can be simplified to help in solution.

$$\nabla \times \mathbf{E} = -j\omega\mu\mathbf{H} \tag{9.1}$$

$$\nabla \times \mathbf{H} = j\omega\varepsilon\mathbf{E} \tag{9.2}$$

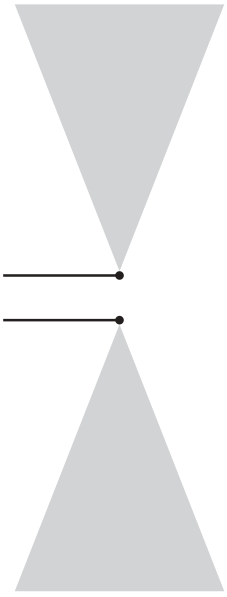


Figure 9.4 Bow-tie approximation to biconical antennas (Ref. 1.)

Equation (9.1) simplifies to

$$\frac{1}{r}\frac{\partial}{\partial r}(rE_\theta) = -j\omega\mu H_\phi \tag{9.3}$$

and Eq. (9.2) provides the following two relationships:

$$\frac{\partial}{\partial\theta}(r\sin\theta H_\phi) = 0 \tag{9.4}$$

$$\frac{1}{r}\frac{\partial}{\partial r}(rH_\phi) = -j\omega\varepsilon E_\theta \tag{9.5}$$

From the latter two equations one can find the differential equation satisfied by the only surviving component of the magnetic field H_ϕ as

$$\frac{\partial^2}{\partial r^2}(rH_\phi) = -\omega^2\mu\varepsilon(rH_\phi) = -k^2(rH_\phi) \tag{9.6}$$

which can be solved to find H_ϕ. A simplification can be obtained using Eq. (9.4), to avoid lengthy boundary condition applications, to find the following two equations:

$$H_\phi = \frac{f(r)}{\sin\theta} \tag{9.7}$$

$$f(r) = H_0\frac{e^{-jkr}}{r} \tag{9.8}$$

These equations can be used to find the required solutions for the components of the radiated electric and magnetic fields for the biconical antenna.

$$E_\theta = \eta H_\phi = \eta\frac{H_0}{\sin\theta}\frac{e^{-jkr}}{r} \tag{9.9}$$

$$H_\phi = \frac{H_0}{\sin\theta}\frac{e^{-jkr}}{r} \tag{9.10}$$

This set of solutions represents a spherical wave propagating in the space between the two cones. Since it is a TEM wave, its electric and magnetic field components are related by η, the intrinsic impedance of free space. They may be compared with the well known solutions for an infinitesimal electric dipole and finite length electric dipole antennas. For the infinitesimal dipole, the far-field components have an angular dependence of $\sin\theta$, which is the reciprocal angular dependence of the infinite biconical antenna field components. In the former case the radiated field is maximum in the horizontal plane, while for the latter case it is minimum.

The electric and magnetic field components can be used to calculate the mode voltage and current. They are given as

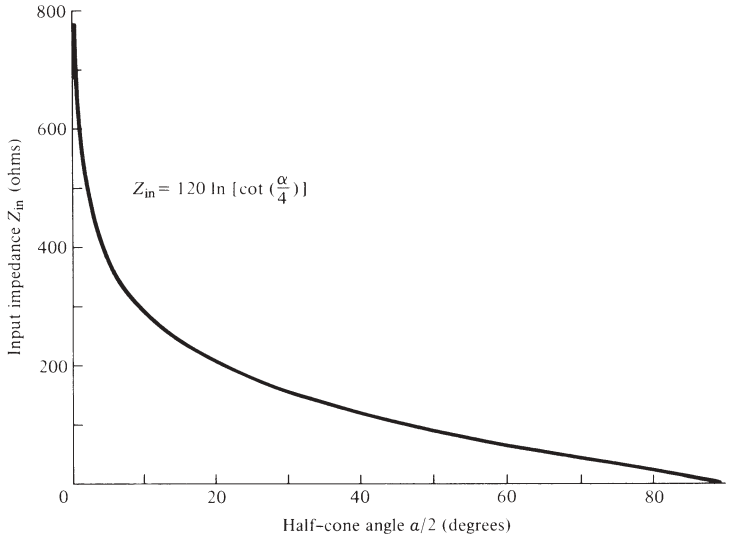
$$V(r) = \int_{\alpha/2}^{\pi-\alpha/2}\mathbf{E}\cdot dl = 2\eta H_0 e^{-jkr}\ln\left[\cot\left(\frac{\alpha}{4}\right)\right] \tag{9.11}$$

$$I(r) = \int_0^{2\pi} H_\phi r\sin\theta\, d\phi = H_0 e^{-jkr}\int_0^{2\pi} d\phi = 2\pi H_0 e^{-jkr} \tag{9.12}$$

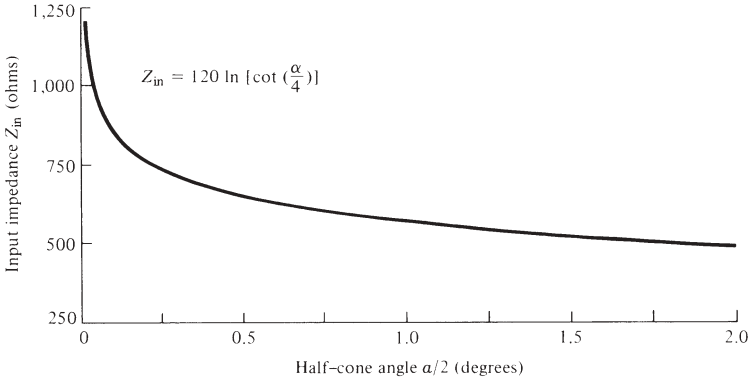
Their ratio gives the characteristic impedance of the mode in the antenna, which is constantly independent of the coordinates. Thus it is also the input impedance of the infinite biconical antenna,

$$Z_c = Z_{\text{in}} = 120 \ln[\cot(\alpha/4)] \tag{9.13}$$

The constancy of the characteristic impedance means that it will also be equal to the radiation resistance of the antenna. Figure 9.5 shows the variation of the input impedance with the cone angle [1]. For small cone angles it is similar to the input resistance of a thin wire antenna and decreases continuously as the cone angle increases.



(a) Input impedance



(b) Input impedance in expanded scale

Figure 9.5 Infinite biconical antennas (Ref. 1.): (a) input impedance and (b) input impedance in expanded scale.

9.2.3 Finite Cones

In a finite biconical antenna, the discontinuity in the radial direction causes the excitation of all higher order radial modes, which are mostly evanescent and store energy in the antenna near field, which gives rise to the antenna input reactance. Also, the reflected fields from the cone's end causes the input impedance to be frequency dependent, reducing the input impedance bandwidth. Sample results for the input resistance and input reactance of narrow angle cones are shown in Figure 9.6. In this figure the parameter L is the cone side length. The trend of the resistance peak is similar to the variation of the infinite biconical antenna resistance. Its peak magnitude decreases rapidly with the cone angle. Since the cone angles are small, the shape of the input resistance and its variation with the antenna length are similar to those of a wire dipole antenna. For a wider range of the cone angles, the measured input resistance and input reactance versus the antenna height ($2\pi a/\lambda$), are shown in Figures 9.7 and 9.8. Their variation diminishes as the cone angle increases. This indicates that the antenna becomes very broadband,

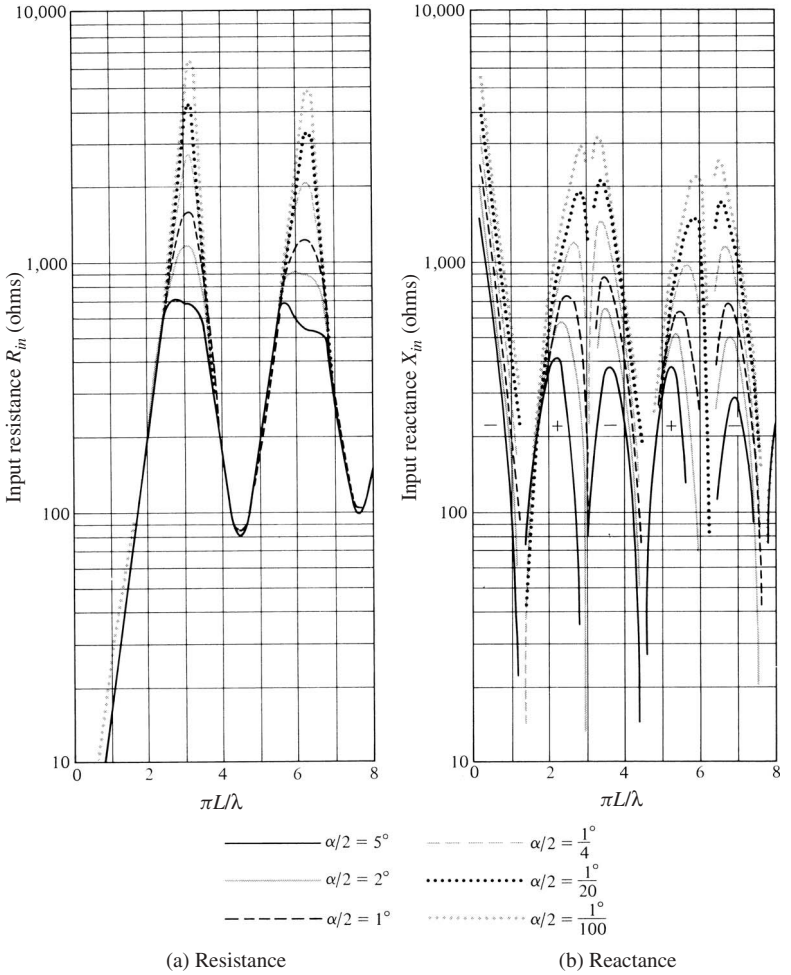


Figure 9.6 Input impedance of finite biconical antenna: (a) resistance and (b) reactance. (Ref. 2.)

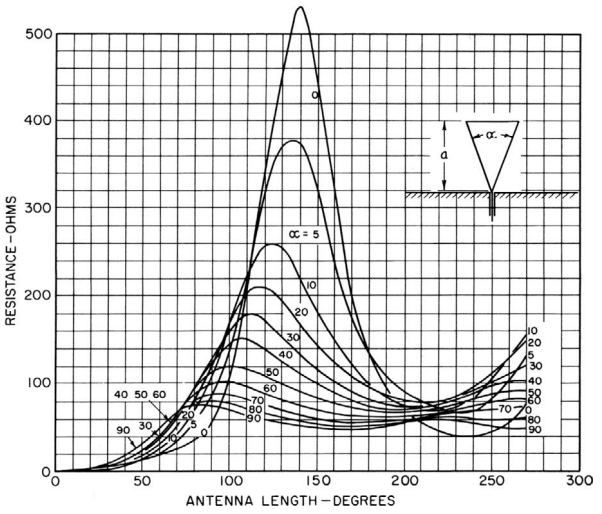


Figure 9.7 Measured input resistance of finite conical monopole for different cone angles. (Ref. 2.)

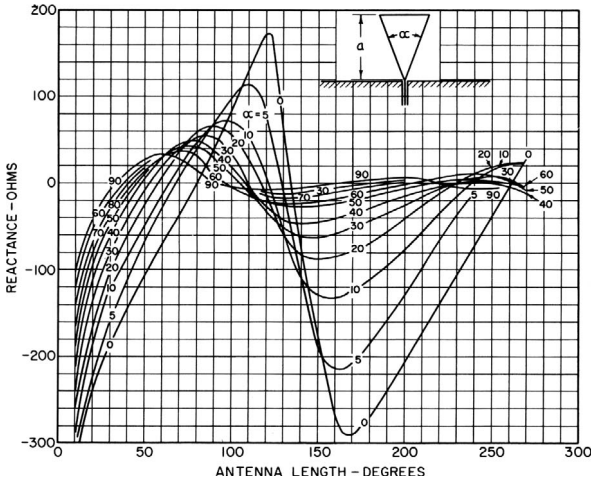


Figure 9.8 Measured input reactance of finite conical monopole for different cone angles. (Ref. 2.)

especially when the input reactance gradually approaches zero. The computed radiation patterns are shown in Figure 9.9. For small cone heights, the results are similar to those of a small dipole and have $\sin\theta$ dependence. They gradually change and approach the radiation patterns of an infinite biconical antenna (i.e., $1/\sin\theta$), being maximum near the truncated cones. In fact, Figures 9.9e and 9.9f clearly show that the peak radiation is along the cone angle.

Figure 9.10 compares the radiation patterns for a finite biconical antenna with different lengths and flare angle of 60°. These patterns are compared with patterns of dipole antennas with different lengths.

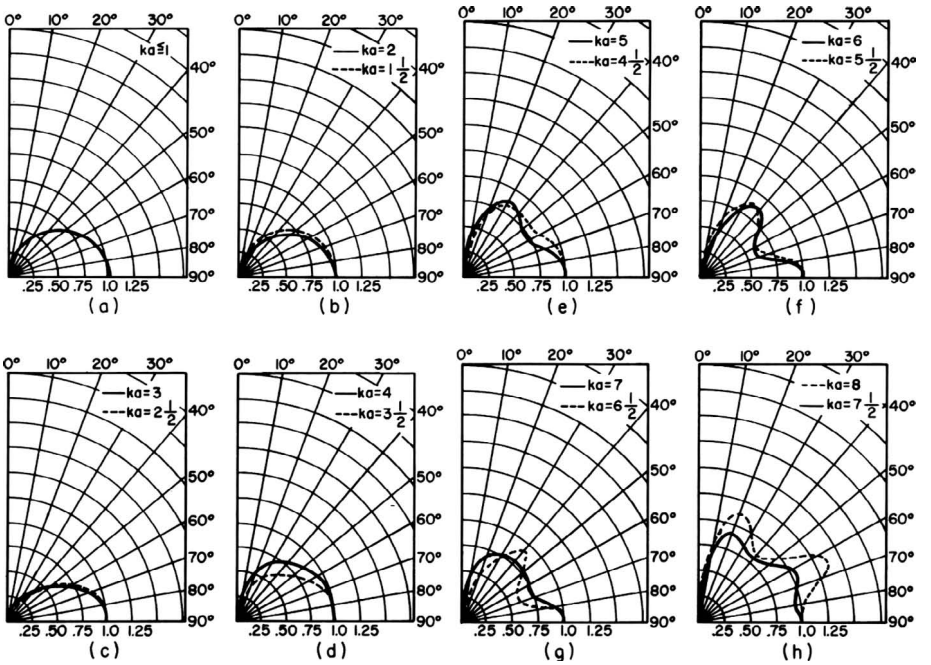


Figure 9.9 Radiation patterns of finite biconical antenna for different lengths, cone angle = 60°. (Ref. 2.)

9.2.4 Discone

A discone antenna is in fact a biconical antenna where one of the cones is replaced by a disk-shaped ground plane. The disk is mounted on top of the cone as shown in Figure 9.11. This structure was first developed by Kandoian in 1945 [4]. Discones have linearly polarized radiation patterns like a monopole. The radiation pattern is near omnidirectional in the azimuth plane for about one octave and the impedance bandwidth can be up to 10:1 [5]. In the elevation plane radiation pattern, it is donut shaped with a typical beamwidth of 20° to 80°, shown in Figure 9.11. Figure 9.12 shows typical E-plane radiation patterns of a discone antenna at different frequencies (02676 Discone Antenna, Benelec Pty Ltd). Figure 9.13 [6] shows the voltage standing wave ratio (VSWR) curves for discones of various angles. Typically peak gains are from 1 to 4 dB.

The discone antenna has three components: the disk, the cone, and the insulator, as shown in Figure 9.14. The diameter of the top of the cone depends on the diameter of the coaxial cable used to feed the antenna. A coaxial line outer conductor is connected to the cone and the inner one is connected to the ground disk. The insulator keeps the disk and cone apart. The antenna's input impedance depends on the cone angle and disk to cone spacing.

The length L of the cone elements should be a quarter-wavelength of the minimum operating frequency. The diameter of the top of the cone depends on the diameter of the coaxial cable and determines the upper frequency limit of the antenna: the smaller the diameter, the higher the frequency. The disk elements should have an overall length D of 0.7 times a quarter-wavelength at the minimum frequency. The slant height of the cone is equal to a full free-space quarter-wavelength at the lowest frequency of interest.

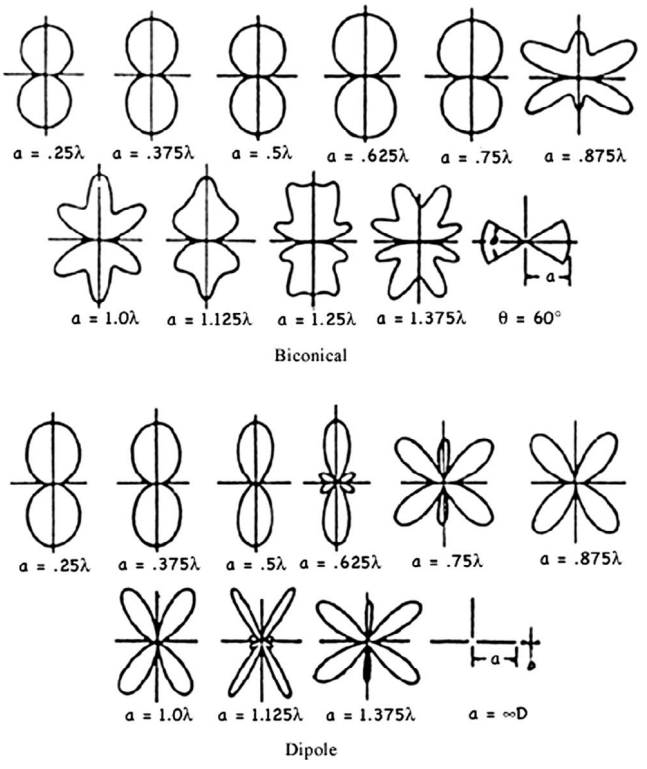


Figure 9.10 Radiation patterns of finite biconical antenna with different lengths and compared to similar dipoles. (Ref. 2.)

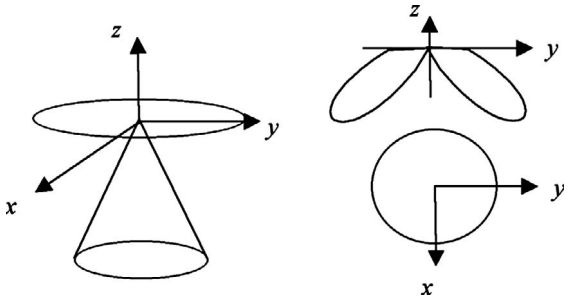


Figure 9.11 Typical radiation patterns of discone antenna.

Discone antennas have been used widely for various applications such as amateur radio and wireless systems. Because of its wideband characteristics it is a suitable choice for ultrawideband (UWB) communication. In UWB communication a pulse transmitting antenna is usually necessary. In Ref. 7 the effects of angular dispersion of discone antennas are studied. The discone antenna was made of a cone with a length of 0.065 cm, half-flare angle (α) of 45°, and ground plane diameter of 20 cm. Figure 9.15 shows how the pulse is distorted at different radiation angles.

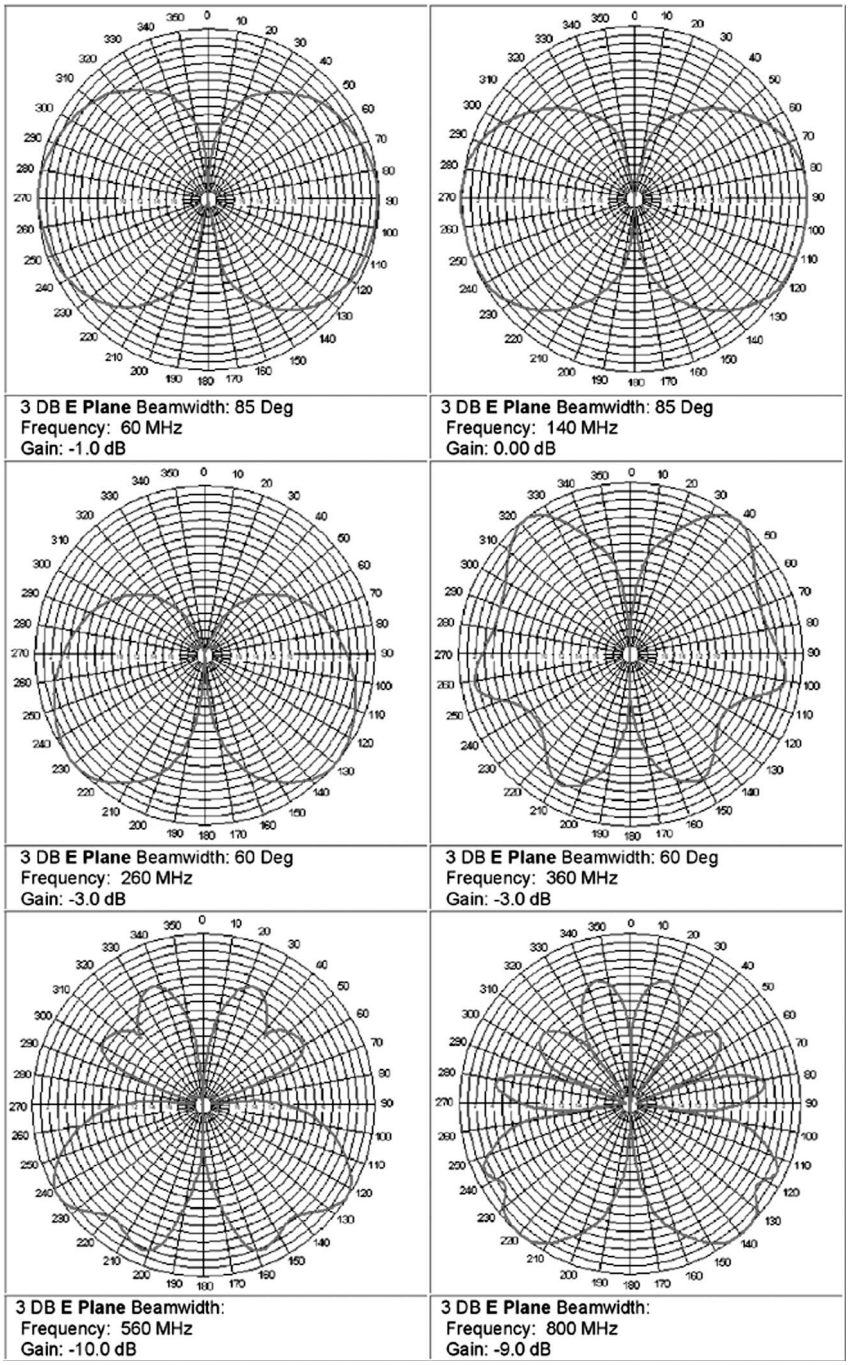


Figure 9.12 Radiation patterns of a discone antenna at different frequencies.

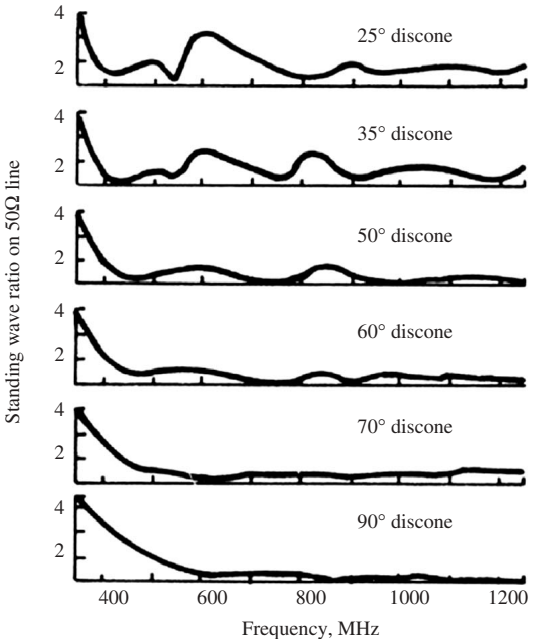


Figure 9.13 VSWR curves for dicones of various half-flare angles. (Ref. 6.)

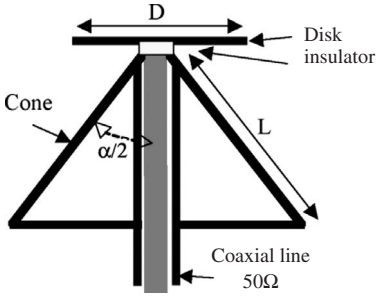


Figure 9.14 Discone antenna elements.

Different variations of discone antennas have been proposed to increase the bandwidth. Three of them are discussed here.

One limiting parameter of discone is the feed point. Since the cone is fed by a coaxial cable, the tip of the cone has to be truncated to the diameter of the coaxial cable. The frequency bandwidth is inversely proportional to the size of truncation. To overcome this problem and increase the bandwidth, the configuration shown in Figure 9.16 is suggested [8]. In this design a discone antenna is designed for a higher frequency band of 2.4–9 GHz and placed upside down with a shared feeding point with the larger discone. The larger discone is designed for a lower frequency band of 300 MHz to 2.4 GHz. The radiation patterns, VSWR, and gain of the optimized antenna are shown in Figures 9.17–9.19, respectively.

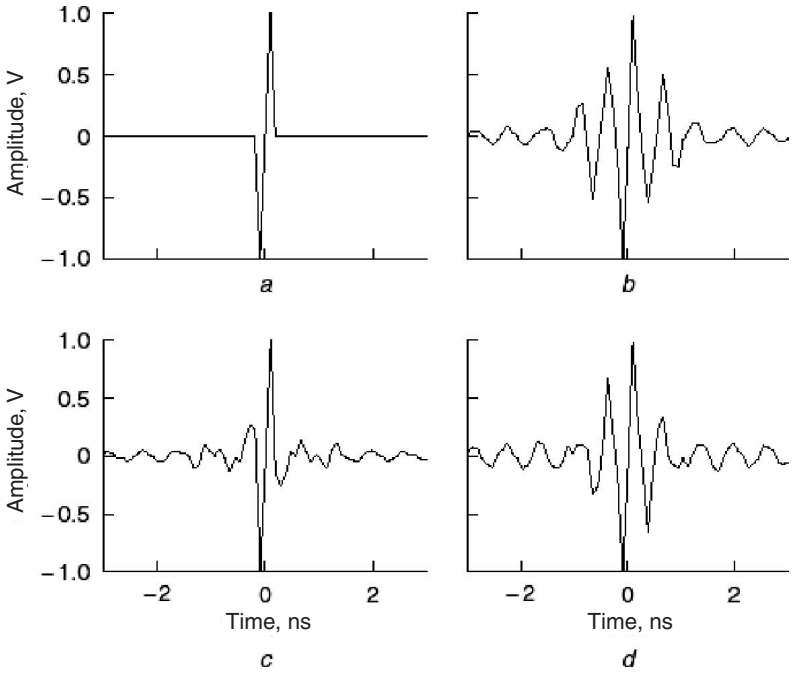
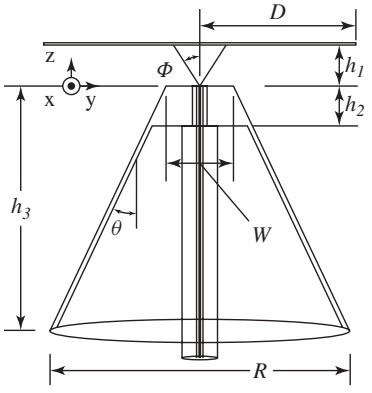


Figure 9.15 Waveform distortion by a discone antenna at various angles: (a) original pulse, (b) $\theta = 30^\circ$, (c) $\theta = 60^\circ$, and (d) $\theta = 90^\circ$. (Ref. 7.)



Φ	55°	θ	30°
h_1	25 mm	W	70 mm
h_2	40 mm	D	145 mm
h_3	230 mm	R	345 mm

Figure 9.16 Double discone antenna. (Ref. 8.)

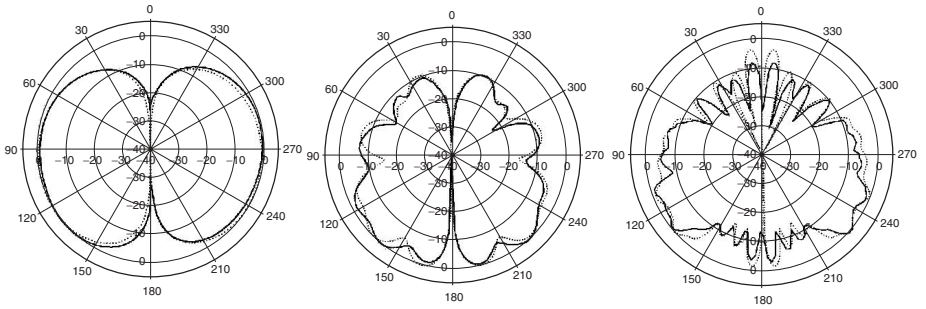


Figure 9.17 Radiation patterns of double discone antenna: (a) 0.5 GHz, (b) 2 GHz, and (c) 5 GHz (_____ measured result; ------ simulated result). (Ref. 8.)

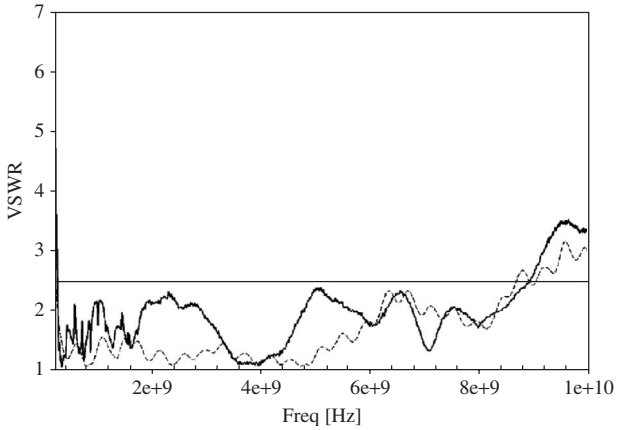


Figure 9.18 VSWR of double discone (_____ measured; ------ simulated VSWR). (Ref. 8.)

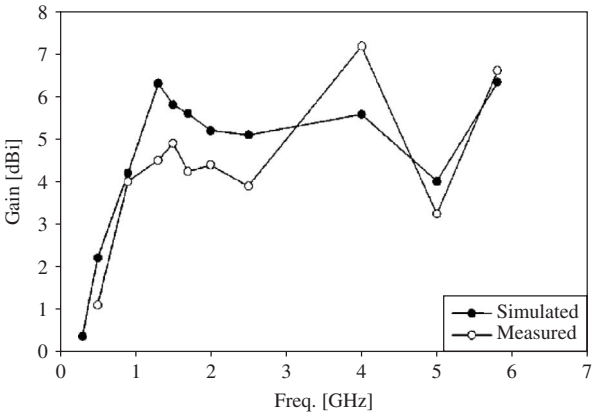
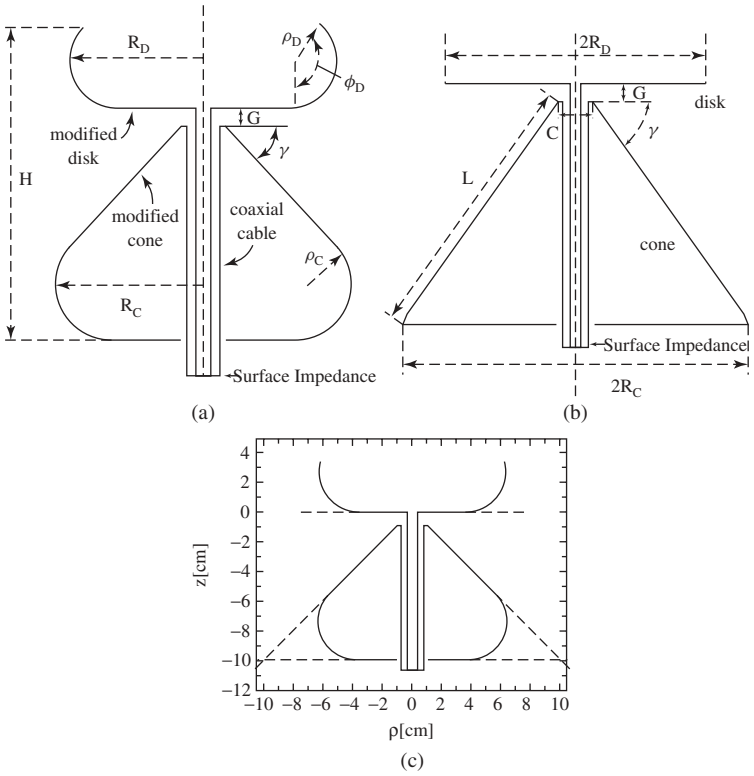


Figure 9.19 Gain of double discone. (Ref. 8.)

A better performance and more compact antenna can be obtained by rolling the edges of both the ground plane and cone as shown in Figure 9.20(a). In Ref. 9 this structure is studied in detail. The rolled edges permit current waves to travel around them smoothly, thus reducing the reflections that degrade the performance at the lower end of the frequency band [9]. By optimizing the radii of curvatures it is possible to obtain a wider frequency bandwidth and lower VSWR by using the rolled edge structure shown in Figure 9.20. This approach allows a reduction in the diameter of the transversal section and shortens the length of the antenna. Figure 9.21 shows the VSWR of the conventional and optimized antennas. The simulations were done using the method of moments (MoM) to solve the electric field integral equation (EFIE).

In the third design to increase the impedance bandwidth the ground plane is replaced by a spiral [10]. The configuration is shown in Figure 9.22. The structure is designed using the MoM and measured. In the frequency band of 650 MHz to 3 GHz this antenna exhibits a VSWR better than 2:1, as shown in Figure 9.23.



Discone Type	R_D [cm]	R_C [cm]	γ	C [cm]	G [cm]	ρ_D [cm]	ρ_C [cm]	H [cm]
Conventional	7	10	45°	0.9	0.8	–	–	10
Modified	6.2	6.2	45°	0.9	0.71	0.36	0.27	14

(d)

Figure 9.20 (a) Rolled edge dicone, (b) conventional discone, (c) comparison of size for conventional and rolled edge antennas, and (d) dimensions of optimized antenna. (Ref. 9.)

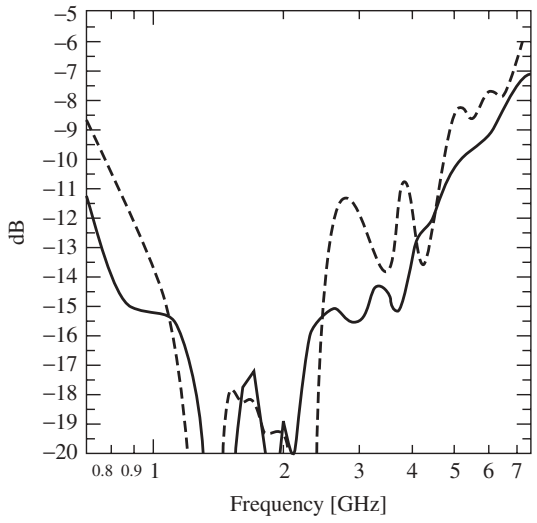


Figure 9.21 Return loss for (_____) rolled edge and (------) conventional antennas. (Ref. 9.)

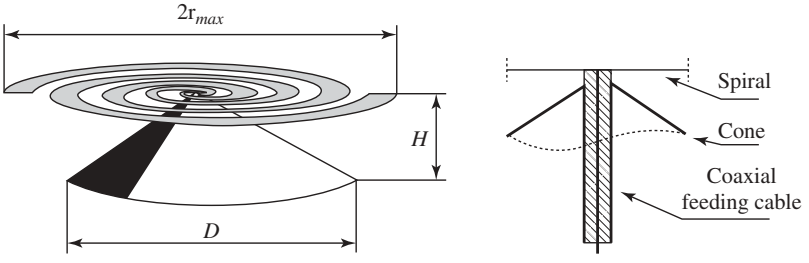


Figure 9.22 Combined spiral–discone antenna. (Ref. 10.)

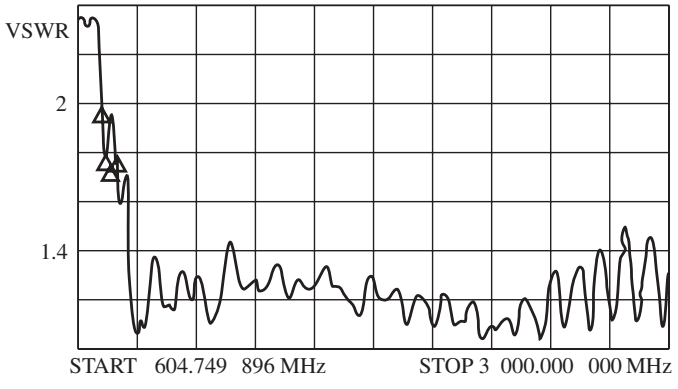


Figure 9.23 Measured VSWR of spiral–discone antenna. (Ref. 10.)

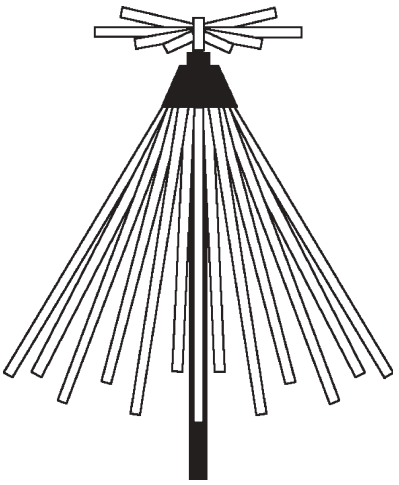


Figure 9.24 Wire discone.

9.2.5 Wire Discone

A discone antenna is a good choice for VHF and UHF applications, however, a solid discone may not be suitable for these ranges of frequency. To reduce the size and weight, a discone can be built using radial wires, as shown in Figure 9.24. These wires act as filters and reduce the frequency band. In Ref. 11, by using a double discone structure with tapered wires and four posts, placed to support the upside–down discone, a bandwidth of 180 MHz to 18 GHz is achieved, as shown in Figure 9.25.

9.2.6 Bow-Tie Antennas

Bow-tie antennas are the planar cross sections of biconical antennas. These antennas may consist of two triangular metal sheets or they can be realized in a planar configuration on a dielectric sheet. The input impedance is significantly affected by the feeding technique. The simplest feeding technique is to use a balun and feed the metal bow-tie antenna like a dipole. Figure 9.26 shows an example of a bow-tie antenna fed by a Roberts balun [12]. This type of feeding is suitable for hand-held devices. The balun bandwidth can limit the antenna bandwidth. In the example shown, the combined VSWR is better than 2:1 for 22.8% frequency bandwidth, centered at 1.575 GHz, obtained by a simulation using an IE3D moment method software package.

The planar bow-tie antenna can be fed by a coplanar waveguide, a microstrip line, and parallel strip lines, when the two parts of the bow-tie antenna are placed on two sides of a dielectric substrate, like a planar Vivaldi antenna. The impedance matching can be done by quarter-wavelength transformers or by tapering. The latter method is used in Ref. 13 and a bandwidth of 19% for VSWR better than 1.5:1 is achieved, as shown in Figure 9.27. In Ref. 14 an exponential tapering is used to match the strip lines to a 50-Ω microstrip line. The antenna has a bandwidth of 68% as shown in Figure 9.28.

Another method of feeding a planar bow-tie antenna is by mircrostrip/coplanar strip lines that include a 180^o phase shifter, as shown in Figure 9.29 [15]. This method has the advantage of having both arms of a bow-tie antenna on one side. This combination has

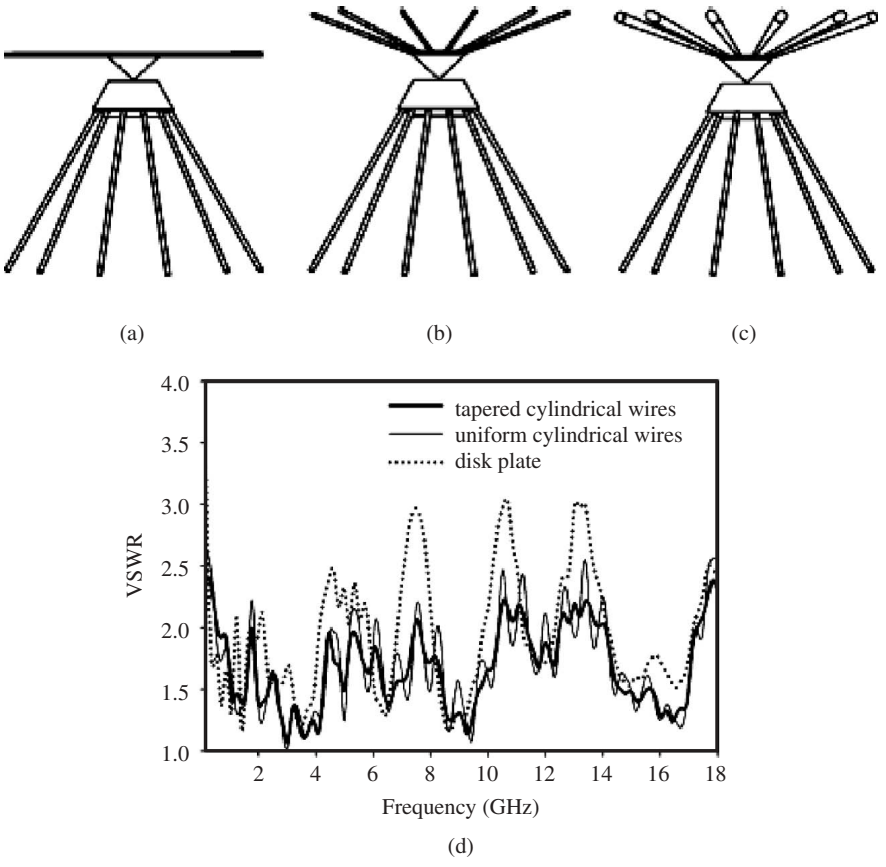


Figure 9.25 Wire double discone antennas: (a) disk plate, (b) wire discone, (c) tapered wire discone, and (d) VSWR. (Ref. 11.)

been used for a quasi-Yagi antenna. However, replacing the Yagi antenna with a bow-tie antenna increases the bandwidth and improves the gain. The bandwidth reported in Ref. 15 is 54.5% by simulation using Ansoft HFSS and 60.1% by measurement. In the array format this combination also shows lower mutual coupling, as shown in Figure 9.29c.

9.2.6.1 Loaded Bow-Tie Antenna Due to its wide bandwidth, a bow-tie antenna is a good candidate as a pulse radiating antenna. One approach to reduce reflections and ringing effects at the feed point is to load the antenna. Continuous resistive loading was first introduced by Wu and King [16] for a dipole antenna. Similar loading was studied numerically by Shalger et al. [17]. Similar to Wu and King, they used continuous loading, where internal resistance per square R of the thin sheet that forms the antenna, has a dependency on ρ/s, given by the following equation:

$$R(\rho/s) = R(1/2)\left(\frac{\rho/s}{1-(\rho/s)}\right), \quad 0 \le \rho/s \le 1 \tag{9.14}$$

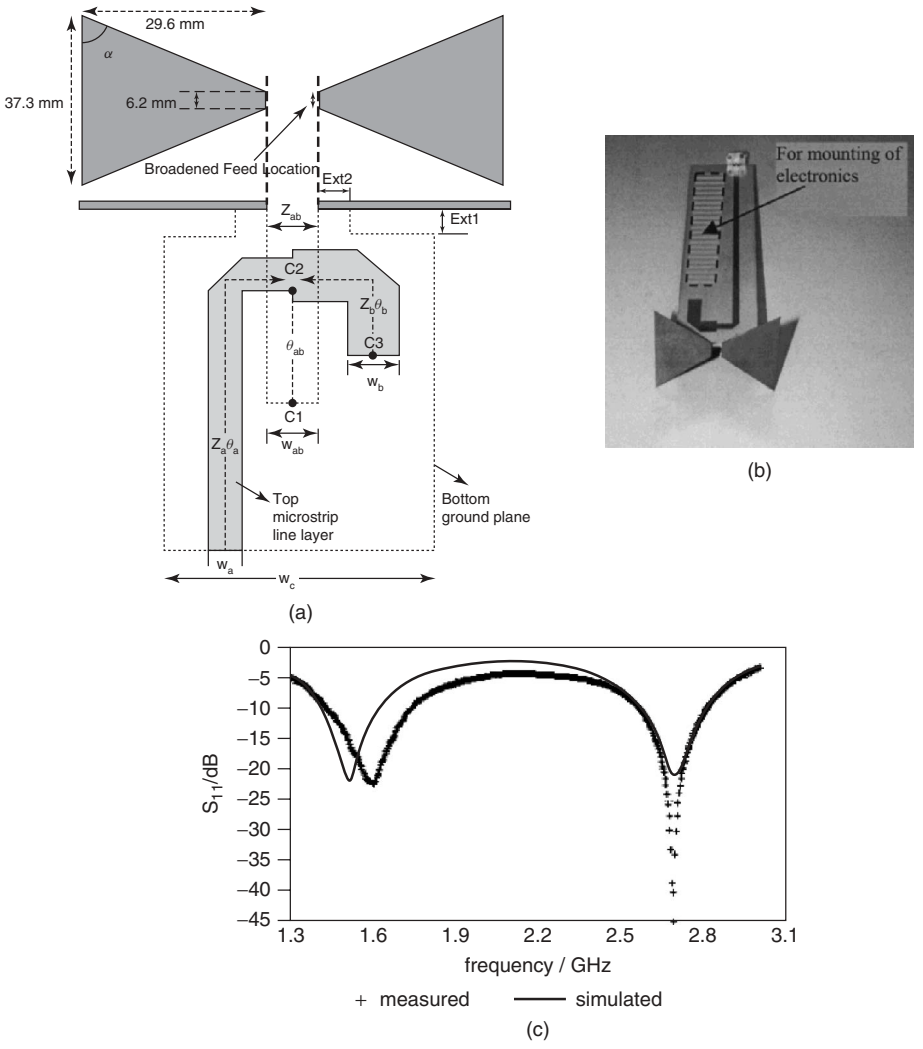


Figure 9.26 Bow-tie antenna fed with a balun: (a) balun geometry, (b) top side of the balun, and (c) antenna return loss. (Ref. 12.)

where ρ is the radial distance from the feed point, s is the distance from the feed to the end of the antenna arm, and $R(1/2)$ is the resistance halfway along the antenna. The value $R(1/2)$ is used as an optimization factor. Figure 9.30 shows the pulse radiation of a loaded antenna calculated by a finite-difference time-domain method [17].

Another method of resistive loading is presented in Ref. 18. In this method lumped resistors consisting of surface mounted (SMT) resistors are used, as shown in Figure 9.31. The resistors are divided into sections to load the structure. A smaller scale of this antenna was constructed on a substrate with $\varepsilon_r = 3.2$. The return loss for this antenna was calculated by CST Microwave Studio software and measured. Figure 9.32 shows clearly how the resistive loading decreases the return loss level.

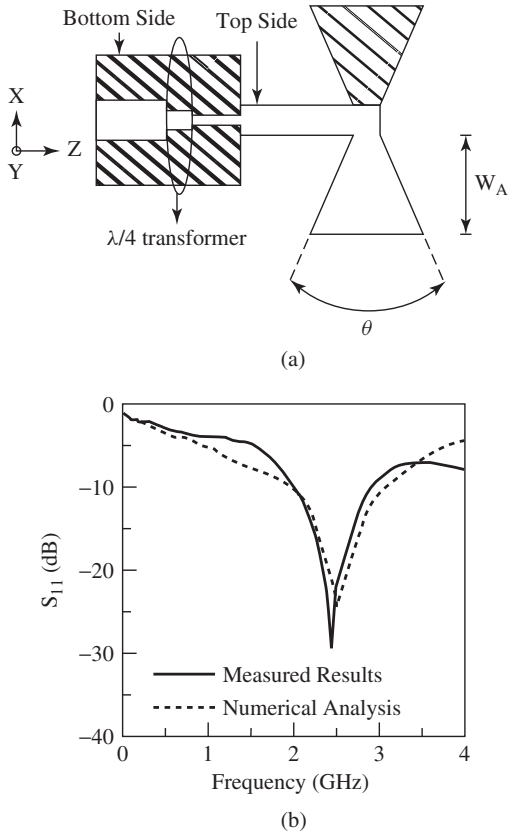


Figure 9.27 (a) Configuration of stripline and quarter-wavelength transformer feed design for bow-tie antenna and (b) return loss of this antenna. (Ref. 13.)

Resistive loading is the most effective way to reduce the ringing effect. However, it reduces the antenna efficiency. On the other hand, the capacitive loading does not reduce the efficiency. To increase the efficiency and also have the optimum ringing effects, the combination of resistive and capacitive loading was proposed by Lestari et al. [19]. Figure 9.33 shows the antenna geometry and a fabricated antenna [19]. A method is presented in this paper to realize capacitive and resistive loadings. A circular end bow-tie antenna with 90° flare angle on an epoxy substrate is considered. The capacitive loading is done by etching concentric slots on the bow-tie surface. To increase the reactance linearly the slot widths are increased toward the ends. The resistive loading was realized by covering the conducting side of the antenna with microwave absorbers. Thus to realize resistive loading with constant profile, the areas of the strips should be made equal along the antenna. The first slot position (l_{fs}) acts as a second source and it has to be located where its radiation adds constructively to the first strip. The value of l_{fs} by which the waveform has the highest peak value in the broadside direction is denoted by $l_{\max}$; its value is given by

$$2\,l_{\max} = \frac{v}{2\,f_c} \tag{9.15}$$

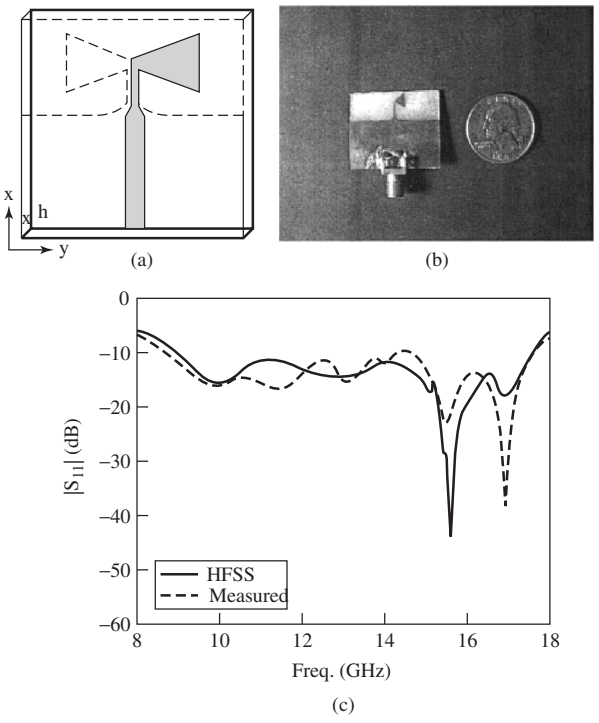


Figure 9.28 (a) Schematic of tapered microstrip/stripling feed for bow-tie antenna, (b) photograph of top layer, and (c) return loss of this antenna. (Ref. 14.)

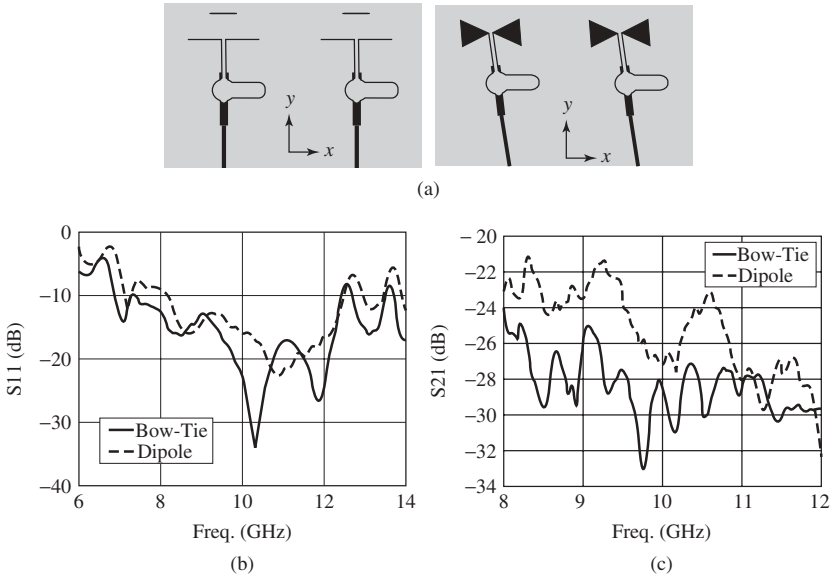


Figure 9.29 (a) Photograph of two-element bow-tie and quasi-Yagi arrays fed by microstrip lines, (b) return loss of a single bow-tie antenna fed by microstrip line, and (c) mutual coupling between two elements of array, where separation between elements is 15 mm. (Ref. 15.)

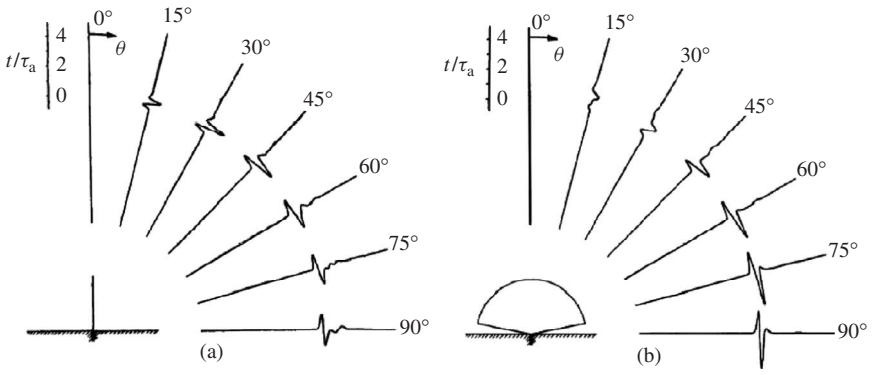


Figure 9.30 Radiation patterns for resistive loaded bow-tie antenna with flare angle = 78.1° and $R(1/2) = 25\ \Omega$/m: (a) broadside radiation and (b) edge-on radiation. (Ref. 17.)

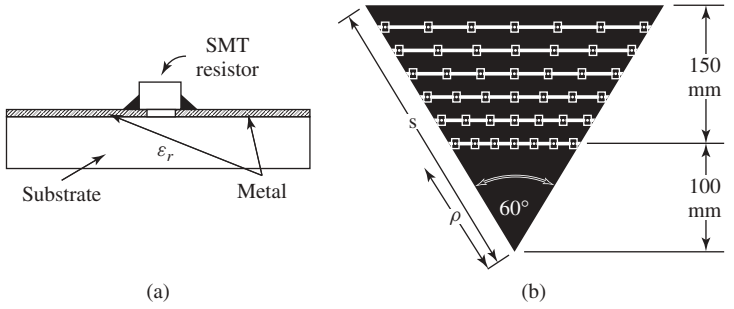


Figure 9.31 Geometry of the resistively loaded bow-tie antenna: (a) cross section of a typical lump-loaded connection and (b) top view of one arm. (Ref. 18.)

where v is the wave velocity in substrate and f_c is the center frequency. Figure 9.34 shows the comparison of ringing reduction for antennas without loading, with resistive loading and with resistive and capacitive loading when $l_{fs} = 6$ cm.

9.2.6.2 Microstrip Bow-Tie Antenna A microstrip bow-tie antenna can be viewed as a planar bow-tie antenna. However, it has different characteristics than a dipole type bow-tie antenna. Since microstrip antennas inherently have limited bandwidth, the frequency band of the microstrip bow-tie antenna is typically around 10%. A microstrip bow-tie antenna may be fed by an electromagnetic coupling or microstrip line. In Figure 9.35 an example of a microstrip line fed microstrip patch is shown [20]. This design provides 10.6% bandwidth in the X-band for frequencies of 6.5–7.8 GHz. The microstrip bow-tie antenna can be arranged in an array configuration. Figure 9.36 shows an array and its return loss [21].

In Figure 9.37a a geometry is shown, where the bow-tie antenna is fed by a coaxial probe and is loaded by an integrated reactive loading to enhance the bandwidth [22]. A rectangular notch is cut in one of the radiating edges. The dimension d has to be optimized for a given flare angle. For the flare angle of 30°, this dimension is 4.2 mm and provides a 4.2% bandwidth as shown in Figure 9.37b.

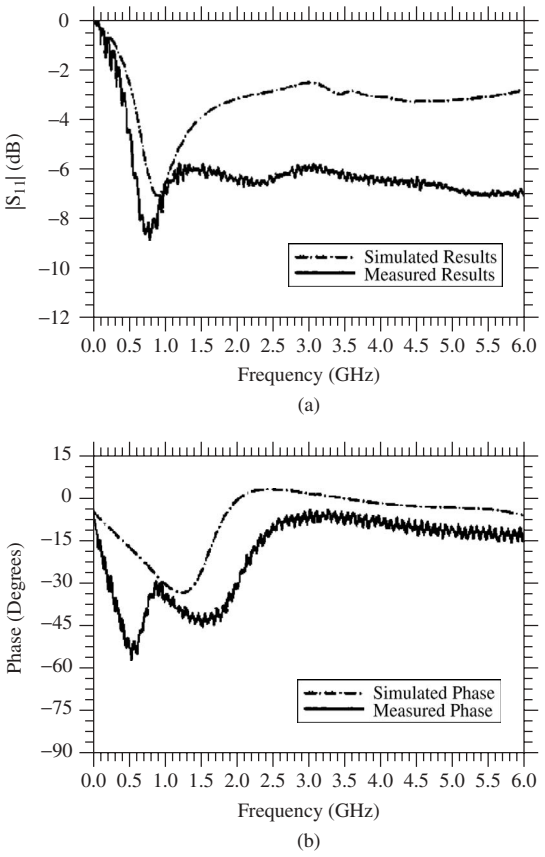


Figure 9.32 (a) Measured magnitude and (b) measured phase responses of the loaded bow-tie antenna relative to a 50-Ω system impedance. (Ref. 18.)

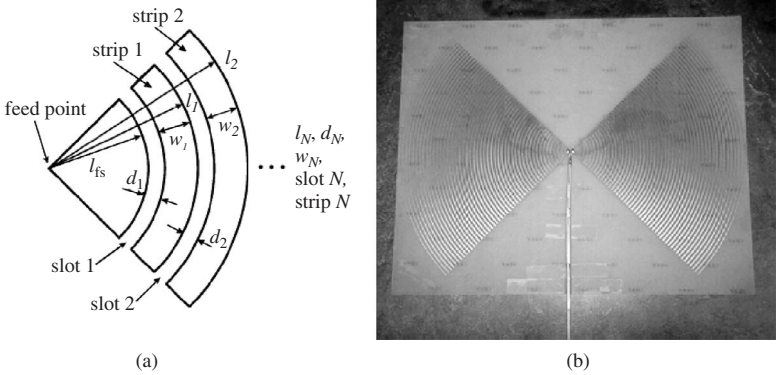


Figure 9.33 (a) Geometry of the circular end bow-tie antenna with concentric slots for realizing a capacitive loading and (b) circular end bow-tie antenna with linear capacitive loading constructed on an epoxy substrate. The flare angle is 90°. (Ref. 19.)

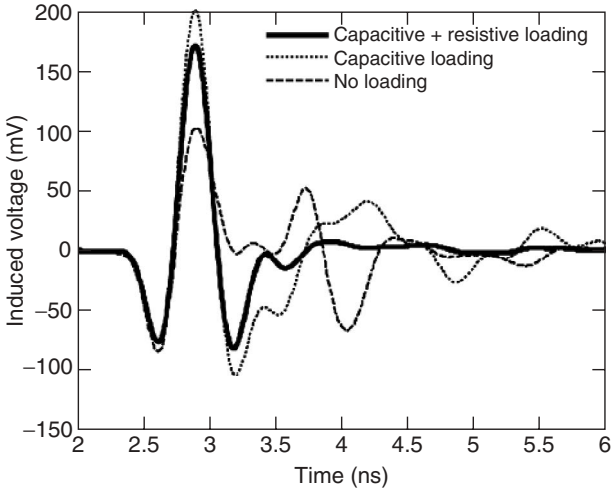
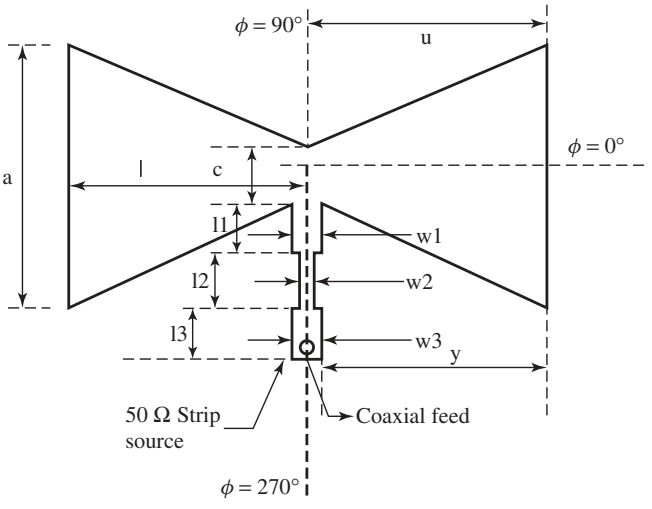


Figure 9.34 Free-space transmit waveforms due to different loading schemes for $l_{fs} = 6$ cm. The observation point is 25 cm from the antenna (in the broadside direction). (Ref. 19.)



Parameter	Dimension (mm)	Parameter	Dimension (mm)
a	23.97	*l*2	6.76
c	5.0	*l*3	5.55
u	20.76	*w*1	1.89
v	19.56	*w*2	0.90
*l*1	4.25	*w*3	2.40

Figure 9.35 Microstrip bow-tie patch antenna fed by 50-Ω microstrip line: $\varepsilon_r = 2.2$ and $h = 0.787$ mm. (Ref. 20.)

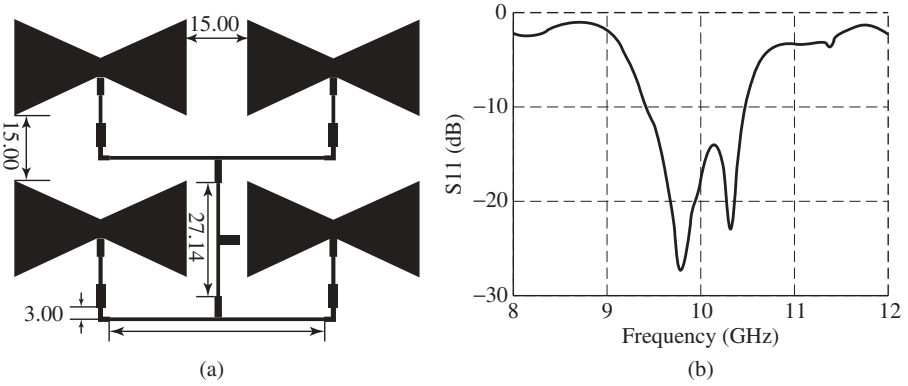


Figure 9.36 (a) Microstrip 2 by 2 bow-tie array with half-wavelength spacing (dimensions in mm) and (b) measured return loss, bandwidth = 10.7%. (Ref. 21.)

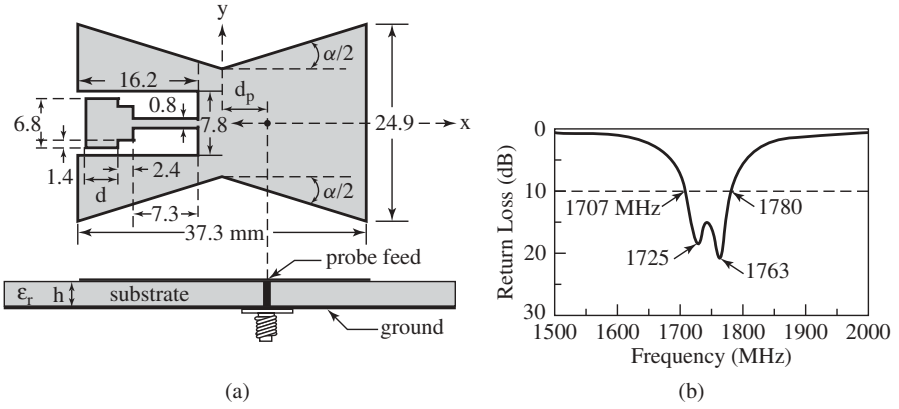
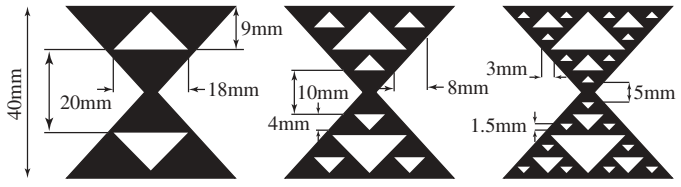


Figure 9.37 (a) Reactive loaded bow-tie geometry and (b) return loss for 30° flare angle. (Ref. 22.)

9.2.6.3 Fractal Bow-Tie Antenna Fractal geometries have been used for different antenna types. Fractals are composite geometries that repeat themselves and are self-similar. The repetition might be in combination with rotations and translations. A fractal bow-tie antenna is formed by using triangular patches. If infinite numbers of iteration are applied to each hand of a bow-tie antenna (one triangle) the resulting shape is called a Sierpinski gasket. Figure 9.38 [23] shows three first iterations of this fractal microstirp bow-tie antenna and compares them with a regular bow-tie antenna in terms of resonant frequency, electrical size, and directivity. Applying this perturbation reduces the electrical size of the antenna and increases the directivity, while it decreases the resonant frequency. This means that by applying fractal shapes, one can miniaturize the microstrip antennas. Figure 9.39 [23] shows the miniaturizing effects clearly. By applying three iterations of slots, the size is reduced to 33% of a square patch, but the frequency bandwidth is also reduced from 1.12% to 0.4%.

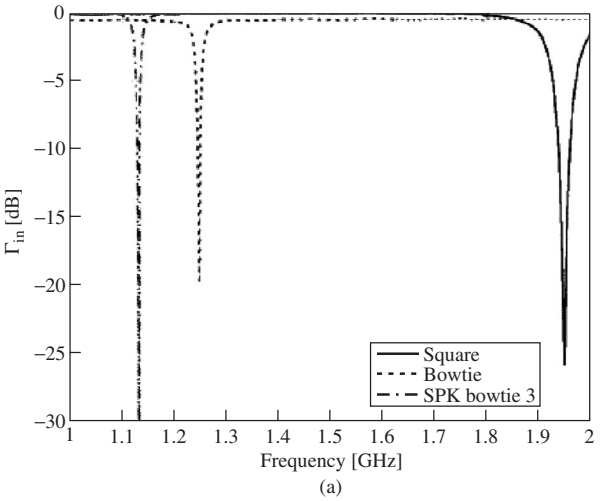


(a)

Simulated Antenna	Bow Tie	SPK Bow Tie 1	SPK Bow Tie 2	SPK Bow Tie 3
Resonant freq. [GHz]	4.684	4.216	4.071	3.866
Electrical size [λ]	0.624	0.562	0.542	0.515
Directivity [dB]	8.2	11.0	11.1	10.9

(b)

Figure 9.38 (a) Geometries of fractal Sierpinski bow-tie patch for the first three iterations and (b) resonant frequencies, electrical size, and directivities of fractal bow-tie antennas. (Ref. 23.)



(a)

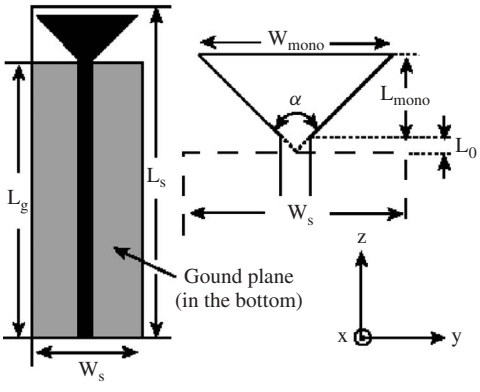
Antenna	Square	Bow Tie 1	SPK Bow Tie 3
Resonant freq. [GHz]	1.947	1.247	1.128
BW [%, SWR = 2]	1.12	0.44	0.40
Electrical size [λ]	0.259	0.166	0.150
Area reduction [A/A_{square}]	1	0.40	0.33

(b)

Figure 9.39 (a) Return loss comparison between square, bow-tie, and Sierpinski (third iteration) patch antennas and (b) comparison of three antennas in terms of resonant frequency, bandwidth (BW), and size. (Ref. 23.)

9.2.7 Monopole

A triangular planar monopole is a planar version of a discone antenna, which has very wideband characteristics. A triangular radiating element on a substrate may be fed by a microstrip line or coplanar waveguide. An example for ultrawideband application is provided in Ref. 24. Figure 9.40 shows the antenna configuration and dimensions. This antenna was simulated using HFSS software and the current distributions for different frequencies are shown in Figure 9.41. Figure 9.42 shows the VSWR of this antenna fabricated on FR4 epoxy substrate, and Figure 9.43 shows the radiation patterns for frequencies of 4–7 GHz. The measured input VSWR is less than 2.5 from 4 to 10 GHz and the radiation pattern is unidirectional in the H-plane, for this range of frequencies.



(a)

ANTENNA dimensions	
W_{mono} (mm)	17
L_{mono} (mm)	7.6
L_0 (mm)	1
α	90°
L_s (mm)	60
W_s (mm)	20
L_g (mm)	50
T (mm)	1
PCB substrate (ε_r, tanδ)	(4.7, 0.018)

(b)

Figure 9.40 Triangular monopole fed by microstrip line on FR4 substrate: (a) antenna configuration and (b) antenna dimensions. (Ref. 24.)

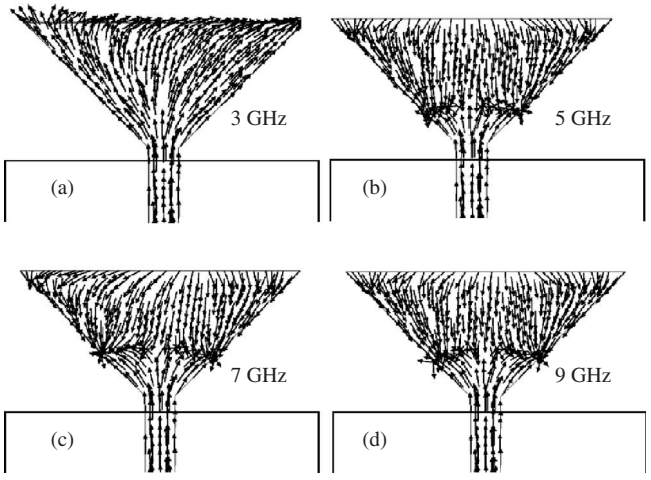


Figure 9.41 Current distribution given by HFSS simulation. (Ref. 24.)

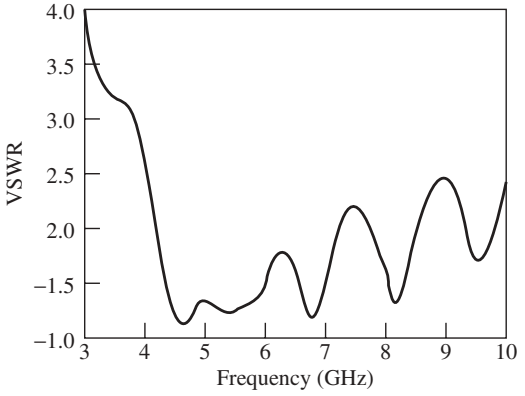


Figure 9.42 Measured VSWR for triangular monopole. (Ref. 24.)

9.2.8 Other Types of Bow-Tie Antennas

Different shape modification has been proposed in the literature to increase the bandwidth of bow-tie antennas. As an example, Figure 9.44 shows a hexagonal planar geometry [25]. This antenna provides a bandwidth of 2.8–10.6 GHz for a VSWR of less than 2 and can be used for ultrawideband applications. Since these applications overlap with other services, a rejection band is sometimes needed. The V-shaped slot is added to create a rejection at a center frequency of 5.25 GHz. Referring to Figure 9.44, the antenna width and length are $a = 22$ mm (0.23λ at 3.1 GHz) and $b = 31$ mm (0.32λ at 3.1 GHz). The coplanar waveguide (CPW) feed geometry has the signal line width $w = 1.5$ mm, and the feed line gap $g = 0.22$ mm, on each side of the signal line to form a 50-Ω transmission line. They are etched on a 1-mm thick FR4 epoxy ($\varepsilon_r = 4.4$) substrate with 0.036-mm copper coating. The half-length of the V-slot L is approximately $\lambda/8$ at 5.25 GHz. The measured and simulated VSWRs (using HFSS) of this antenna, with and without the slot, are shown in Figure 9.45.

Different types of bow-tie antennas with round corners are studied in detail in Ref. 26. Round corners provide better return loss, flatter input impedance, and more stable radiation patterns for the quadratic bow-tie antenna (QBA); see Figure 9.46a. They also decrease the antenna area. Figure 9.46b shows the configuration for the quadratic bow-tie antenna with round corners (QBARC) and Figure 9.46c compares the return loss for antennas with and without round corners. These antennas are fabricated on a substrate with $\varepsilon_r = 4.4$ and thickness = 0.8 mm. In pulse radiating applications the rounded-edge bow-tie antenna (REBA) is usually employed. In the next example, round corners are applied to this antenna, called rounded-edge antenna with round corners (REARC) [26]. Figures 9.47a and 9.47b show the configurations of these two antennas and Figure 9.47c compares their return losses. They were also fabricated on a substrate with $\varepsilon_r = 4.4$ and thickness = 0.8 mm. Similar to the QBA, round corners on the REBA cause better return loss due to flatter input impedance. Applying round corners also provides a flatter gain over the frequency range by improving the gain at high frequencies.

Another bow-tie antenna fed by a coplanar waveguide and with tapered tuning stubs is introduced in Figure 9.48 [27]. It achieves bandwidths in excess of 70%. The effects of various dimensions are studied and some guidelines for designing the antenna are given, starting with dimensions $W1 = 0.77\lambda/f_H$, $L1 = 0.33\lambda_L$, $L2 = 0.03\lambda_L$, $L3 = 0.09\lambda_L$,

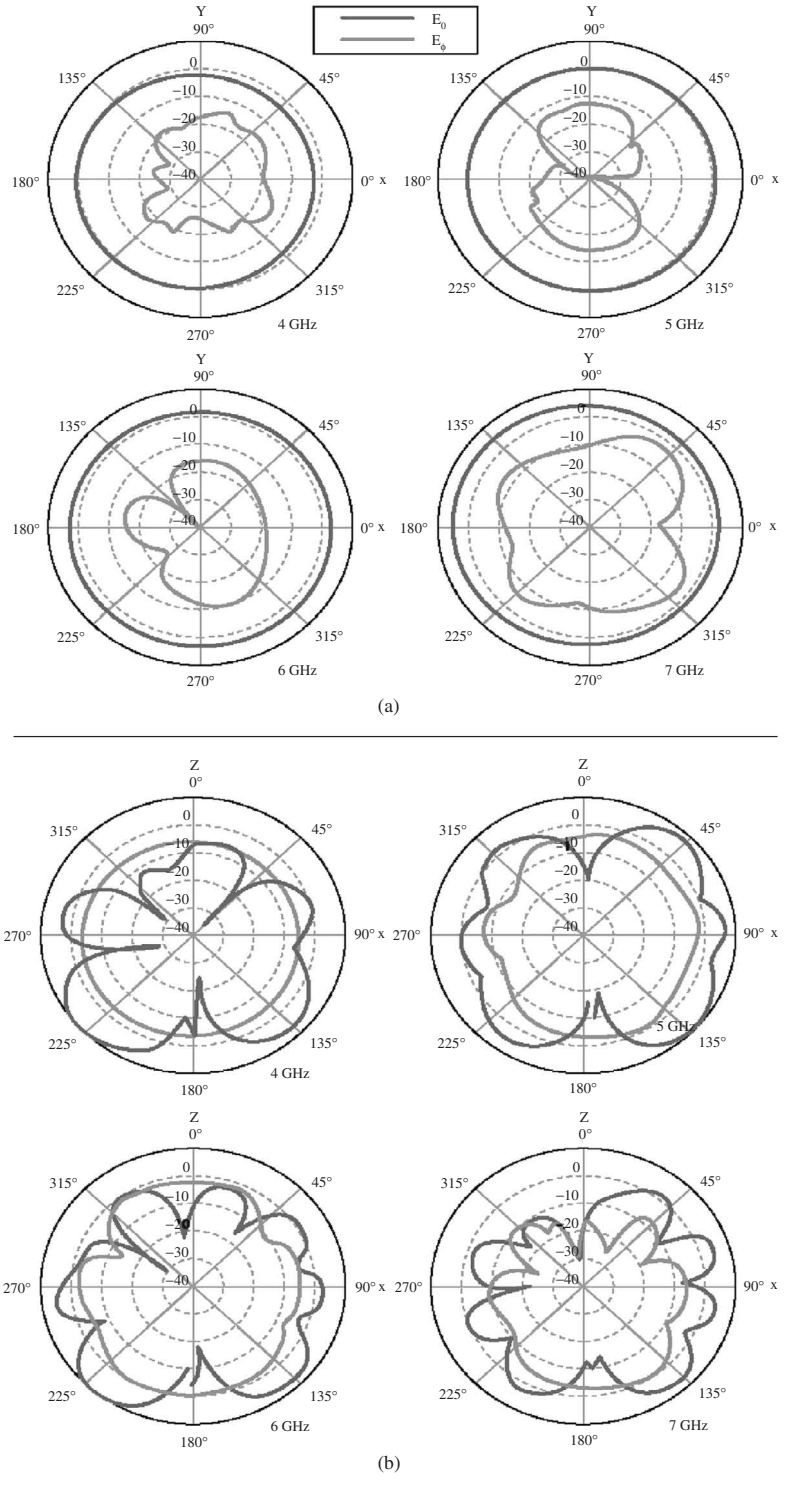


Figure 9.43 Radiation pattern of triangular monopole at 4, 5, 6, and 7 GHz: (a) H-plane and (b) E-plane. (Ref. 24.)

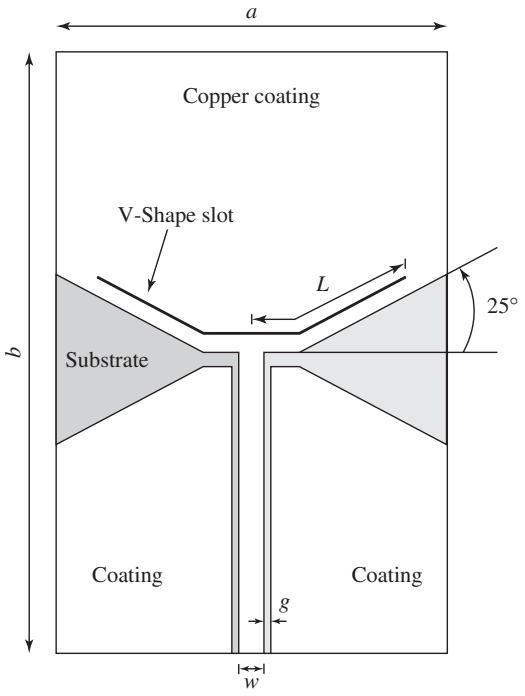


Figure 9.44 CPW fed hexagonal antenna with V-shaped slot. (Ref. 25.)

$L4 = 0.1\lambda_L$, and $W2 = 0.4\lambda_L$, where λ_L and λ_H are wavelengths in free space for the antenna lower and higher frequencies, f_L and f_H, respectively. A sample antenna was designed on a substrate with $\varepsilon_r = 3.2$ and thickness $= 0.76$ mm and fed by a CPW having line and slot widths of 3 mm and 0.16 mm, and dimensions ($W1$, $W2$, $L1$, $L2$, $L3$, $L4$) = (18.85, 14.1, 11.62, 0.875, 2.2, 2.68 mm). The ground plane was truncated 1.8 cm away from the bow-tie slot edge. Its simulated and measured return loses are shown in Figure 9.49, operating from 8 to 16.8 GHz and achieving a bandwidth of 73%.

9.2.9 General Design Guidelines

Infinite biconical antennas are inherently wideband. For finite sizes the bandwidth is affected by the terminations. The lower frequency is limited by the antenna's overall length and the upper frequency by the feed region's accuracy and matching technique. Increasing the flare angle can reduce the reflection from the end and improve the bandwidth. In time-domain radiation it can reduce the ringing effects.

Cones can be reduced to wires or planar patches (bow-tie antennas), but with a reduction in their bandwidth. They can be fed like a dipole. For planar bow-tie antennas on a dielectric, a coplanar waveguide or microstrip line feed can be used, with a 20–70% bandwidth.

A microstrip bow-tie antenna is not wideband. This is due to the microstrip antenna's limitations. A typical bandwidth of 10% can be achieved.

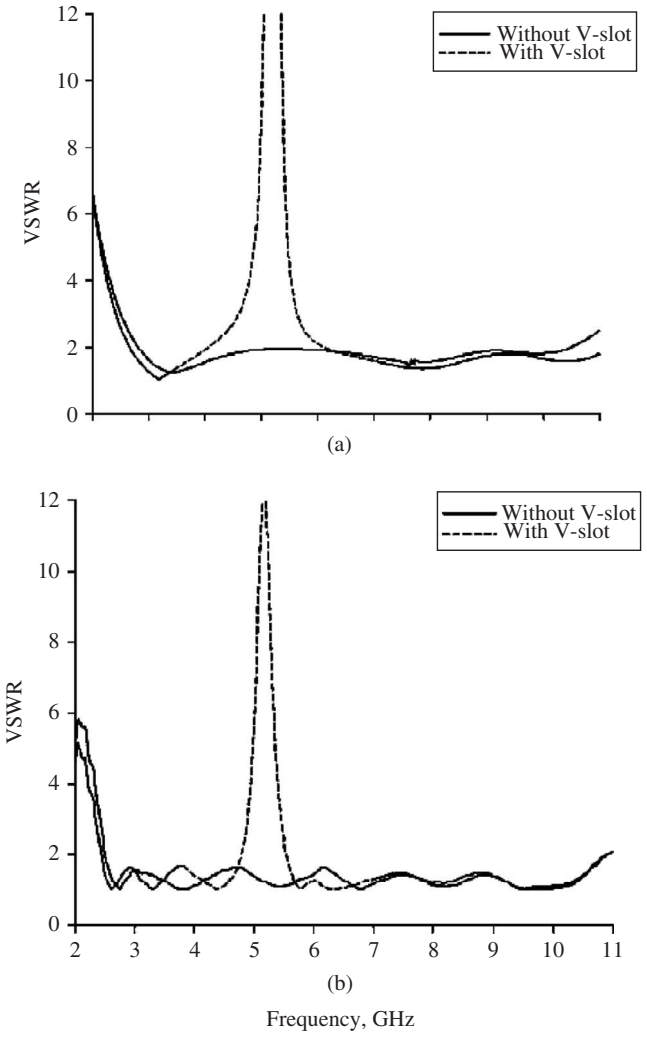


Figure 9.45 VSWRs for hexagonal antenna with and without slot: (a) simulated and (b) measured. (Ref. 25.)

To increase the bandwidth, resistive loading may be used, but it can also reduce the antenna's efficiency to below 50%. A numerical study of loading techniques and their effects on different antennas including bicones are given in Ref. 28.

Fractal shaped bow-tie antennas, for the first few iterations, reduce the physical size of antennas but offer small bandwidths at around 1–2%. Higher order iterations do not usually show significant effects, as the currents tend to concentrate near the large-scale features of the edges and borders.

Monopole antennas offer wide varieties and are good candidates for ultrawideband applications. They can be built planar or nonplanar, with near omnidirectional radiation patterns and around 100% bandwidths. However, their cross-polarization level, especially near the upper frequency end, can be high.

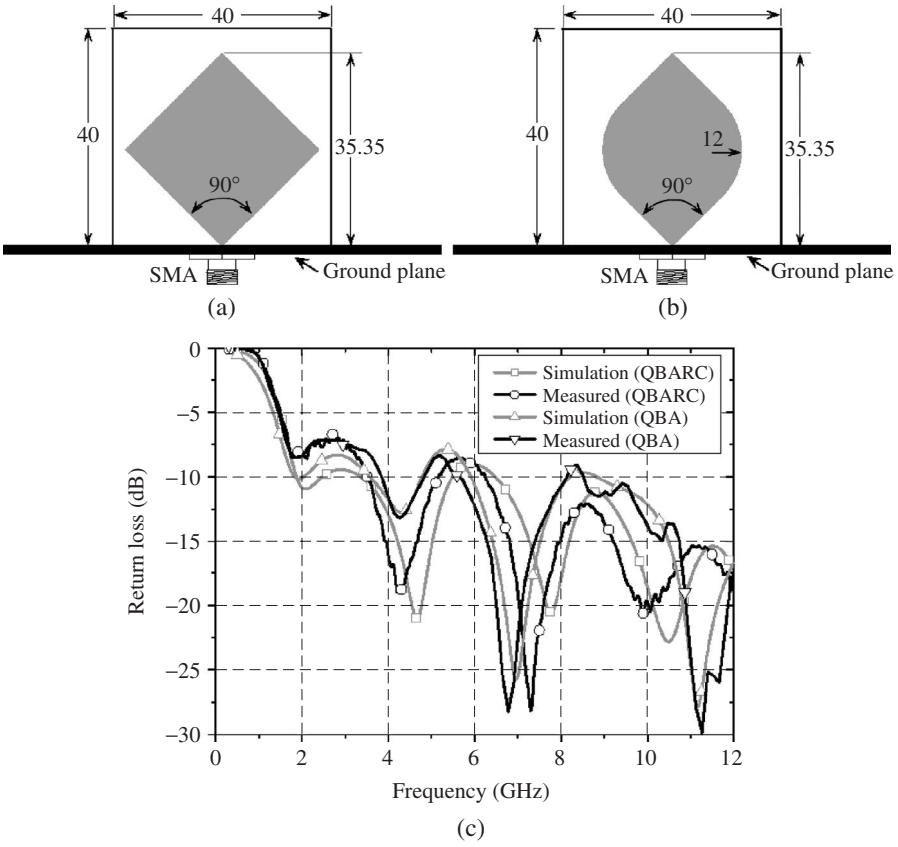


Figure 9.46 (a) Quadratic bow-tie antenna with round corners (QBARC), (b) quadratic bow-tie antenna (QBA) without round corners (dimensions in mm), and (c) return loss for antennas with and without round corners. (Ref. 26.)

9.3 HELICAL ANTENNAS

9.3.1 Introduction

Helical antennas consist of a conducting wire wound into a helix. Its cross section, or view from its axis, can be circular, elliptical, square, or any other shape, but the circular helix is the most common antenna type. Its concept was established experimentally by Kraus [29], who also developed empirical rules for its design [30], also described by others [31, 32]. It is one of the most important circularly polarized antennas and relatively easy to design or fabricate. In addition, the simple geometry of the helix also makes it convenient for numerical investigation and optimization. Consequently, it is extensively investigated and practically utilized.

The geometry of a helical antenna is shown in Figure 9.50. Its various parameters are as follows:

D = diameter of the helix

C = circumference of helix = πD

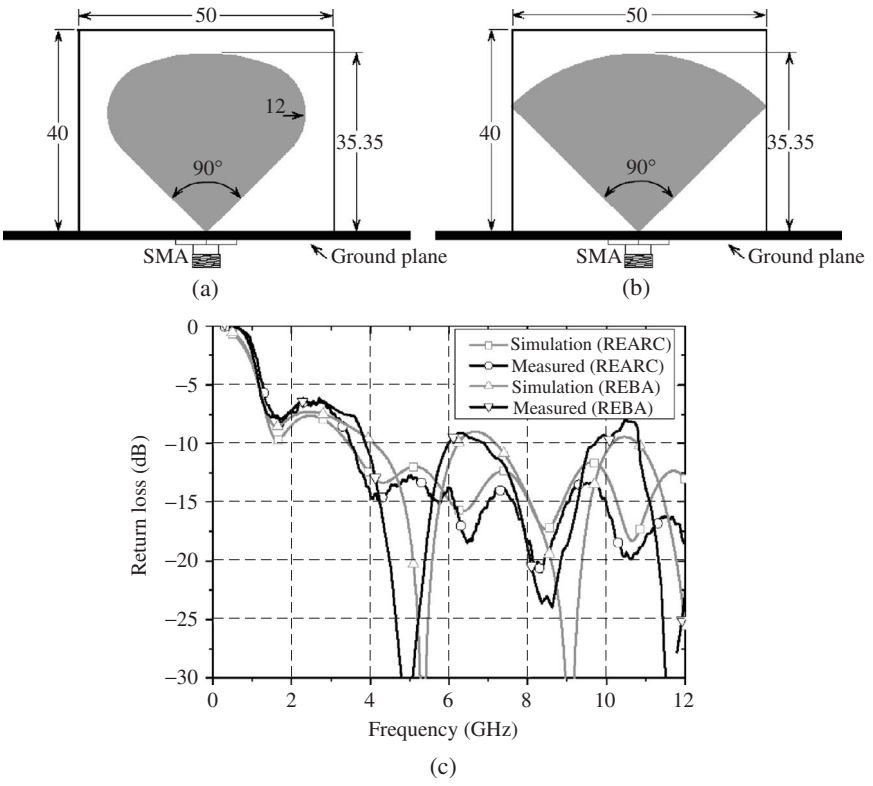


Figure 9.47 (a) Rounded-edge bow-tie antenna (REBA), (b) rounded-edge bow-tie antenna with round corners (REARC) (dimensions in mm), and (c) return loss for antennas with and without round corners. (Ref. 26.)

S = spacing between helix turns
α = pitch angle of helix = $\tan^{-1}(S/\pi D)$
N = number of turns
L = total axial length of helix = NS
l = length of one turn = $\sqrt{(\pi D)^2 + S^2}$
d = diameter of the helix wire

A helix has an interesting geometry and in the limit can be a loop antenna (for $S = 0$) or dipole antenna (for $D = d$). As such, it enjoys their properties but avoids some of their limitations. A circular loop has a rotational symmetry and can support infinite azimuthal modes. However, because of its finite size, it is a resonant structure and has a narrow impedance bandwidth for each mode. The helix, on the other hand, has many turns N and the mode currents can run along its length as they radiate. Thus it behaves more like a traveling-wave antenna than a resonant one and is significantly more wideband. In fact, as will be shown later, it can be designed to have almost constant input impedance for its modes. Another difference between the two antennas is due to the pitch angle or the spacing between turns of the helix. Because of this axial length, the helix current has an

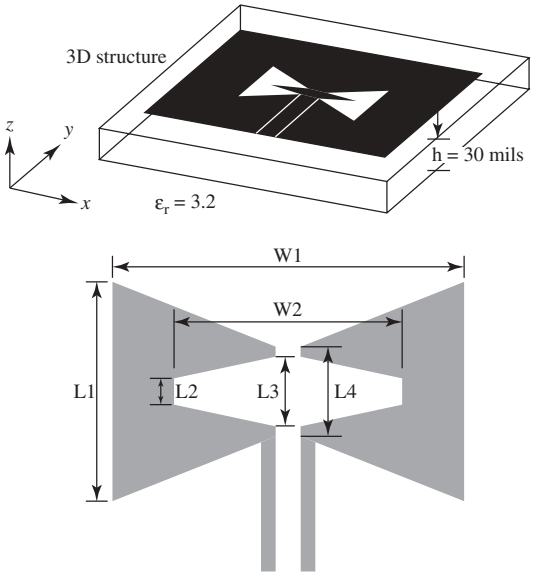


Figure 9.48 Slot bow-tie antenna with tuning stubs. (Ref. 27.)

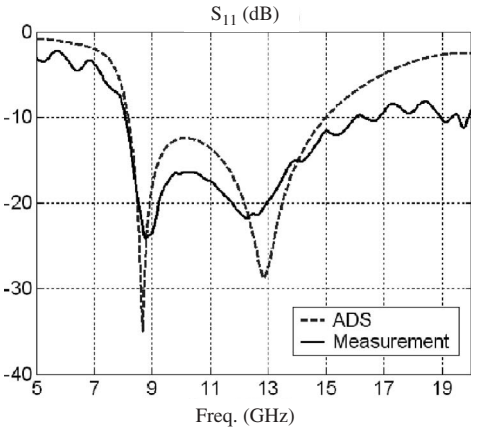


Figure 9.49 Return loss of slot bow-tie antenna designed for X-band. (Ref. 27.)

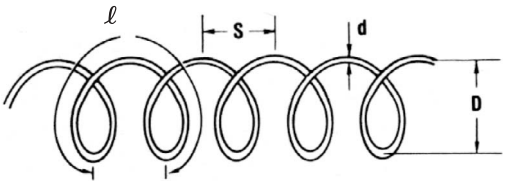


Figure 9.50 Geometry of a helix antenna with its parameters.

axial component and can radiate a wideband circularly polarized wave with a single feed. In this respect, the helix behaves as a combination of loop and dipole antennas, fed in phase quadrature. The interrelationship between the circular and axial components of the current also provides the ability for beam shaping. This property becomes an important tool for designing multifilar helices with shaped conical beams [33–36].

For a uniform helical antenna, the conducting wire is wound over a cylinder of constant diameter and its current distribution has both axial and circumferential components. Thus, in general, its components can be written

$$I(z, \phi)_n = A_n \exp(j\beta_n z) \exp(jn\phi), \quad n = 0, 1, 2, \ldots \tag{9.16}$$

where n is the mode number in the azimuthal direction ϕ, I_n is the nth component of the helix current, β_n is the axial propagation constant, and A_n is the mode excitation constant. The axial propagation constant can be related to the propagation constant along the helix wire by its geometry. As is known in loop antennas, the nth azimuthal mode resonates, when the circumference of the loop becomes about $n\lambda$, where λ is the wavelength of the signal propagating through the helix wire. In reality, the nth mode excites within a bandwidth around $n\lambda$, the size of which depends on the antenna type. Since the loop antenna is generally narrowband, the mode excitation is restricted to a small frequency band around its resonance. These modes are well separated from each other and provide appropriate modal radiation patterns and characteristics. The situation, however, is different for a helical antenna. Its bandwidth is wider, and adjacent modes can be excited simultaneously, which will affect its performance, especially in shaping its radiation pattern and causing large sidelobes.

In both loop and microstrip antennas, only the $n = 1$ mode radiates axially. Other modes generate a boresight null. However, this is not necessarily the case for helical antennas, as multifilar helices can be used effectively to generate conical beams with the $n = 1$ mode, with much smaller diameters. Since they are more compact and simpler in design, they are preferred in communication applications, especially in mobile communications. The geometry of the helix also detrimentally affects the radiation beam of the higher order modes. Its consecutive turns act as an endfire array and force the antenna beam toward the axis, which counters the design goal of generating a conical beam. Thus higher order modes of helix have not seen widespread applications. The zero-order mode, known as the normal mode, is different because it requires small helix dimensions in wavelength and is an ideal antenna for circular polarization.

9.3.2 Normal Mode Helix

For this mode of operation the helix circumference must be smaller than the wavelength, and much smaller preferably. However, decreasing the helix diameter reduces its efficiency, and enlarging beyond $C > 0.5\lambda$ excites the axial mode. A compromise must be made depending on the application. Because of the small size of this antenna, it can easily be designed and studied using a numerical method for wire antennas. However, much can be learned by determining its approximate radiation characteristics, which can also be used as the starting point for the numerical analysis. Since the circumference is small, the helix current can be assumed constant, in both phase and amplitude, and one turn of the helix can be approximated by a combination of loop and dipole antennas,

having a diameter and length of D and S, respectively. Their respective radiation fields for the helix axis along the z axis will be

$$E_L = \left(\frac{\eta\pi I k^2 D^2}{16\pi r}\right)\exp(-jkr)\sin\theta\,\hat{\phi} \tag{9.17}$$

$$E_D = \left(\frac{j\omega\mu I S}{4\pi r}\right)\exp(-jkr)\sin\theta\,\hat{\theta} \tag{9.18}$$

where $E_L = E_\phi$ and $E_D = E_\theta$ are the radiated fields of the loop and dipole, respectively, and k, μ, and η are the propagation constant, permeability, and intrinsic impedance of free space. Since there is a phase shift of 90° between the two field components, the resulting axial ratio becomes

$$|AR| = \frac{|E_\theta|}{|E_\phi|} = \frac{2\lambda S}{(\pi D)^2} \tag{9.19}$$

or, for circular polarization, one must have

$$\pi D = \sqrt{2\lambda S} \tag{9.20}$$

which can be used to find the helix pitch angle as

$$\tan\alpha = \frac{\pi D}{2\lambda} = \sqrt{S/2\lambda} \tag{9.21}$$

Equation (9.21) relates the required pitch angle to the helix diameter D, or the spacing S. As numerical data, one can show that if the length l of a helix's single turn is selected to be 0.1λ, 0.2λ, or 0.5λ, then the required pitch angles become 2.86°, 5.68°, and 13.65°, respectively.

In low frequency applications, where the helix size becomes too small in wavelength, its input impedance can have a small resistive part. In such cases the antenna input impedance can be difficult to match and small bandwidths result. To remedy this, a dual-conductor helix design was proposed and investigated for operation at 760 MHz, using both an equivalent circuit approach and the moment method [37]. The helix wire was replaced by a two-conductor transmission line, fed at its input and short circuited at the end, as shown in Figure 9.51. In this manner the helix conductor acts both as the radiator and a shorted stub, in parallel with its input impedance. The stub reactance helps with input impedance matching and in broadbanding the antenna. In applications with low helix input impedance, the benefit is most significant when the transmission line characteristic impedance is also low, which favors strip lines shown in Figure 9.52. The input impedance and VSWR bandwidth of such a design are compared with those of the single-conductor case in Figure 9.53, where the moment method (MoM) case indicates the numerical data of the single-wire helix. These results show that using a dual-conductor line not only improves the antenna bandwidth (12%) but also significantly raises its input impedance. The latter helps in feeding the antenna using conventional transmission lines. In Figure 9.53 *sp* is the distance of the short point, shown in Figure 9.51, from the helix end.

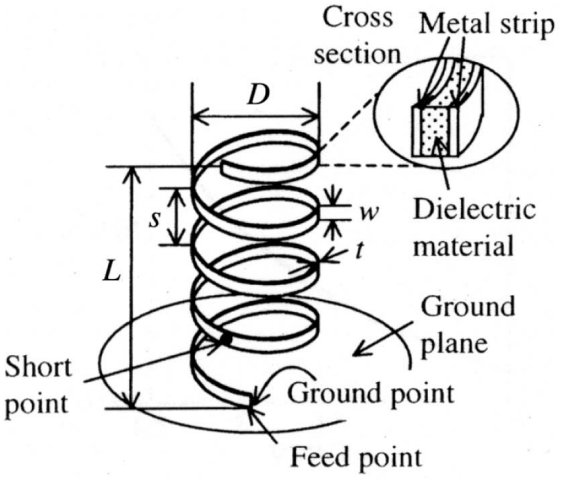


Figure 9.51 Geometry of a dual-conductor helix antenna over a ground plane. One conductor is fed and the second one is shorted to the ground. (Ref. 37.)

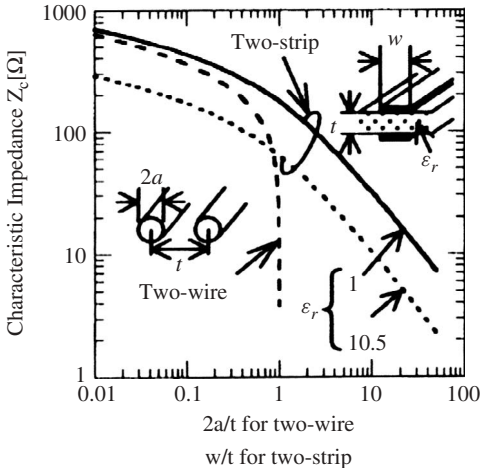


Figure 9.52 Characteristic impedances of two-wire and two-strip transmission lines. (Ref. 37.)

A dual-band operation of a normal mode helix can also be obtained, similar to wideband ones and was studied in Refs. 38 and 39. A parasitic coupling for feeding the helix is another possibility for these applications, which is being used successfully in broadbanding of other antennas. It has another interesting application in areas like high temperature superconductivity, where the antenna cannot be excited directly. A similar parasitically excited normal mode helical antenna has been studied in Ref. 40. The exciter is a loop antenna at normal temperature, with accessible terminals.

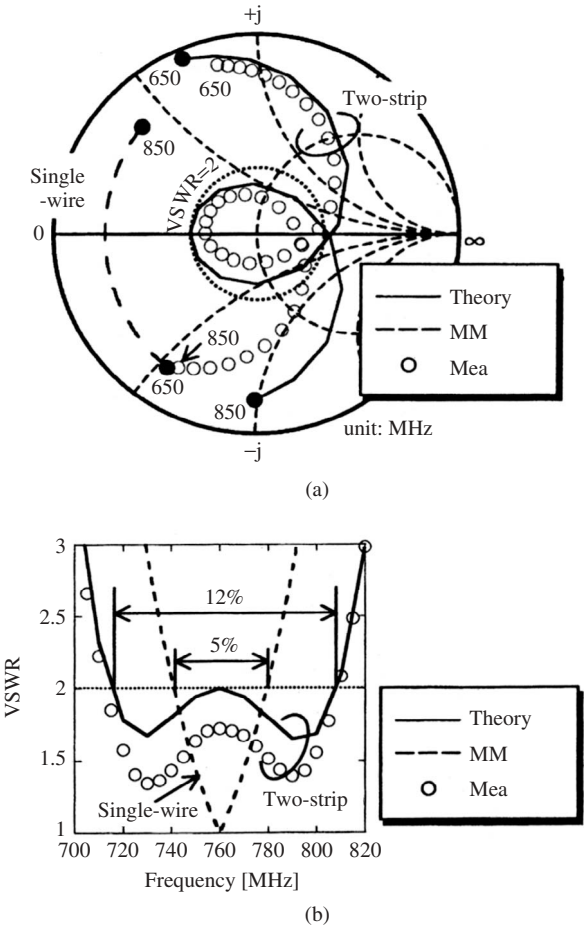


Figure 9.53 Input impedance characteristics of a dual-conductor normal mode helix antenna, using two-strip lines, $D = 10$, $S = 11$, $W = 3$, $t = 0.2$, $L = 44$, $N = 4$, $sp = 30.5$ (dimensions in mm), and $\varepsilon_r = 10.5$: (a) input impedance and (b) VSWR with frequency. (Ref. 37.)

9.3.3 Axial Mode Helix

A monofilar helix, with mode index $n = 1$, radiates in the axial direction, similar to other antennas supporting such a mode like a loop, microstrip, or horn antenna. In its simplest form, it is a constant diameter helix, having a circumference of about one wavelength, $C \approx \lambda$, wound in either right-hand or left-hand, similar to a screw. When placed over a ground plane, it can be fed easily by a coaxial input connector or microstrip line. An axial feeding maintains the symmetry and can facilitate rotation, and microstrip feeding can be useful for impedance matching. In either case, the current induced on the helix conductor radiates as it rotates and progresses along its length, with the polarization being controlled by the direction of its winding.

Historically, the helix was fed by a launcher in the form of a finite ground plane or a cavity [30, 32, 41] with satisfactory results. However, the rationale for such feed systems was not completely studied or understood. This problem was later investigated numerically [42]. The current distribution was computed and it was found that its amplitude decreases rapidly over the first two turns of the helix and then remains almost constant, going to zero at the end. A sample of the computed currents and the corresponding radiation patterns are shown in Figure 9.54. The helix parameters were the circumference $C = 10$ cm, the number of turns $= 5$, pitch angle $\alpha = 12.5^\circ$, and wire radius $= 0.5$ mm. From this figure, it is clear that the helix antenna has two regions—the C region and S region [42]. The C region is the first region and can also be called the exciter region of the helix, where the current decreases monotonously as it radiates and reaches a minimum. Its size is about two turns and is almost independent of the helix length and number of turns. It is interesting to observe in Figure 9.54 that the phase velocity and phase progression in this region are the same as in free space. The second region is the S region, which is the remaining part of the helix and is located in front of the exciter region. This region, which is also called the surface wave region, has an almost constant current. Its phase velocity is less than free space, with a larger phase progression, which indicates it has slow-wave characteristics.

The above phenomena concerning the helix current in the second (S) region were further investigated for other helices with a larger number of turns. The behaviors were similar and the uniformity of the current over the region improved with the number of turns. One can conclude from this study that the helix antenna is very similar to a Yagi antenna. It has an exciter region to launch the wave, and a director region to guide the launched wave. Since the guiding characteristics improve with the number of its turns, the helix antenna's gain or directivity will improve for the axial $n = 1$ mode, by increasing the number of turns. To further understand this guiding characteristic the computed relative phase velocities, p, along the helix in its two regions have been compared with those of free-space, endfire arrays, and the Hansen–Woodyard condition. The results for a 16-turn helix antenna, when its circumference is about one wavelength, are shown in Figure 9.55 [42], normalized with the free-space phase velocity. It is clear that over the

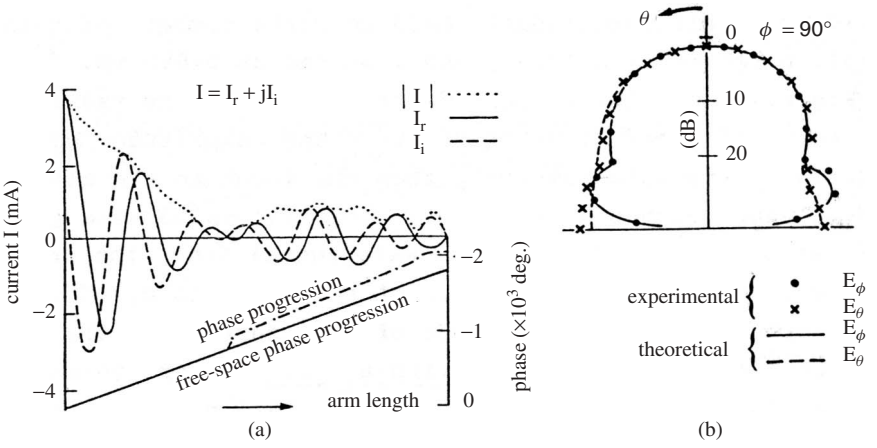


Figure 9.54 Current distribution and radiation patterns of a helix antenna: $C = 10$ cm, $N = 5$, $\alpha = 12.5^\circ$, and wire radius $= 0.5$ mm. (Ref. 42.)

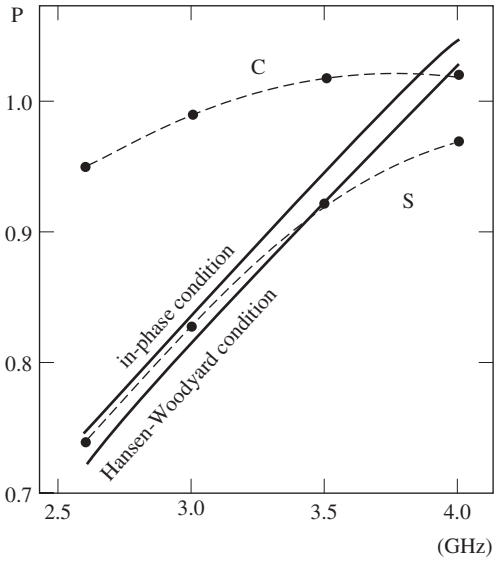


Figure 9.55 Relative phase velocity $p = v/c$ of a helix antenna with $N = 16$, $\alpha = 12.5^\circ$, and $C = 10$ cm. (Ref. 42.)

exciter C region the phase velocity is almost equal to the free-space velocity, and at frequencies above the resonant frequency of 3 GHz. In the S region, where the guided wave is propagating, it is initially equal to the phase velocity of the endfire antennas but gradually decreases to reach that of the Hansen–Woodyard condition. At higher frequencies beyond 3.5 GHz, the excitation of the higher order mode gradually reduces the phase velocity and deteriorates the helix antenna performance and directivity.

The similarity of the helix antenna with a Yagi–Uda array was further investigated for an eight-turn helix, by disconnecting the helix arm after the second turn and placing the remaining six turns parasitically in front of the exciter turns. The configuration was investigated numerically and its performance compared with a corresponding eight-turn helix, when the parasitic one was connected to the exciter. Their performance in current distribution and radiation patterns were practically the same. The parasitically coupled antenna provided a gain of 7.9 dBic with an axial ratio of 1.8, which compared favorably with 8.2 dBic and 1.5, the gain and axial ratio of the original helix [43]. This study demonstrated that the two regions of the helix antenna behave quite independently and need not be connected electrically. The exciter region launches the surface waves and influences the antenna input parameters, and the director region guides the launched wave and controls the radiation characteristics. Their electrical coupling influences the antenna bandwidth and other characteristics.

The guided wave in region S of the helix induces an almost constant current on the helix wire that must go to zero at its end. This causes a reflected current that propagates back toward the input end, affecting the helix input impedance, making it frequency dependent. The same effect can be observed by using the impedance discontinuity between the guided wave in section S and the free-space wave after the wave leaves the helix. Since the reflected wave is traveling in the opposite direction, it also influences the antenna axial ratio and thus the gain. This is shown in Figure 9.56 for a

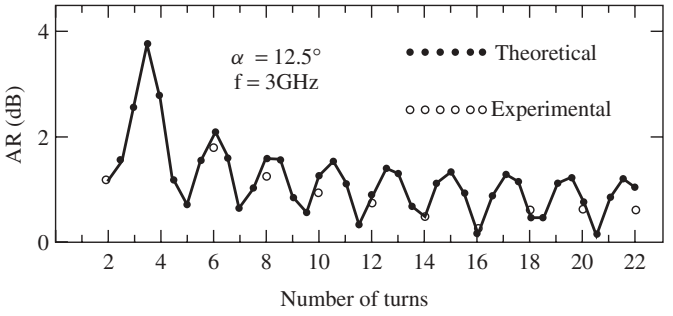


Figure 9.56 Axial ratio with an increase in the number of helical turns, $C = 10$ cm. (Ref. 42.)

helix with circumference $C = 10\ \text{cm} = \lambda$, as a function of the helix number of turns. It is clear that the achieved axial ratio oscillates with the number of turns. In fact a two-turn helix gives a better axial ratio than larger helices. To remedy this, the reflection from the helix end must be reduced. A possible way to reduce this reflection, or the impedance discontinuity at the end point, is to taper the helix at its end, to allow the wave to leave the helix gradually. This has been investigated numerically and experimentally [42–45].

It should also be noted that small-turn helices could also radiate excellent circularly polarized waves, with very small axial ratios [45], as shown in Figure 9.56. In practice, one can optimize small-turn helices for optimum axial ratios, by adjusting their conductor lengths, that is, by selecting fractional turns rather than integer turns, to optimize the axial ratio. Such designs, however, will be optimum over a short frequency range and the antennas will not be broadband, which are outside the scope of this chapter.

9.3.4 Radiation Characteristics

The helix antenna is an endfire antenna and can be viewed as a linear endfire array, made of N identical elements separated by an interelement separation distance of S. In this respect, there are two possibilities for the array performance and behavior: the ordinary endfire array and the Hansen–Woodyard endfire array. From the antenna point of view, the Hansen–Woodyard arrays provide higher directivities, by a factor of 1.789, or 2.5 dB. This is a very significant improvement in the array performance, as far as its directivity is concerned, which naturally comes at the expense of its other parameters. Since the directivity is optimized, it becomes sensitive to the array parameters and frequency. So its bandwidth becomes narrower, in comparison to the ordinary endfire array. Also, since there is an inverse relationship between the antenna's directivity and sidelobes, the first sidelobe of the Hansen–Woodyard array increases by as much as 4 dB over that of the endfire array, which is normally around 12 dB. Thus a Hansen–Woodyard array provides a sidelobe level of around 9–10 dB below its main beam. The beamwidth similarly reduces to account for the increased directivity.

The helix antenna behaves somewhere between the two endfire arrays. Since a helix is normally made of a good conductor, its losses are negligible and its gain is about the same as its directivity, provided its input port is well matched. So for the purpose of discussion in this section, the gain will be used as the helix directivity. The numerical computations of the helix current, and its phase progression along its wire and axis, confirmed the presence of the slow-wave structure and similarity between the helix antenna

and the Hansen–Woodyard arrays. This was shown in Figure 9.55, for a 16-turn helix. It showed that the helix phase progression changes with frequency within its band. Below the frequency f_0, where $C = \lambda$, it is similar to the ordinary endfire array, but gradually the phase progression approaches that of the Hansen–Woodyard array, when the frequency becomes $1.17 f_0$. Experimental investigations have also confirmed this fact for long helices. But for small helices the peak gain performance occurs at even higher frequencies, as the number of turns decreases. For this reason, the operating band of the helix for its axial mode is normally set between $0.8 f_0$ and $1.2 f_0$; that is, $0.8 f_0 < f < 1.2 f_0$. Wider bandwidths can also be defined, if larger gain variations can be tolerated. It should also be noted that the impedance bandwidth of a helix antenna can be wider than its gain bandwidth and their difference will increase with the helix number of turns and more strict gain bandwidth definitions.

The similarities between the helix antenna and the endfire arrays have been used to develop useful relationships for the design and prediction of antenna performance, which have been confirmed and improved by experiments. A summary of these relationships is provided here and discussed. However, one must accept the fact that these are approximate relationships and more accurate performance data must be obtained by numerical investigations.

A uniform helix may be assumed equivalent to a uniform array that has an array factor AF given by

$$AF = \frac{\sin(N\psi/2)}{\sin(\psi/2)} \tag{9.22}$$

where

$$\psi = kS\cos\theta - \delta \tag{9.23}$$

and

$k = 2\pi/\lambda$
$\delta = kl/(v/c)$
$v =$ phase velocity along the helix wire
$c =$ velocity of light in free space
$l =$ length of the helix turn

If the radiation pattern of a single turn of helix is approximated by $\cos\theta$, which is a good approximation for the main beam pattern of the helix's axial $n = 1$ mode, the total radiated field of the helix can be written

$$E(\theta) = AF\cos\theta \tag{9.24}$$

The phase δ varies between that of the endfire array and the Hansen–Woodyard array, provided its pitch angle is in its optimum range of $12^\circ < \alpha < 15^\circ$. With these assumptions, design expressions for the helix gain, axial ratio, and beamwidth have been found and reported in the literature [30, 32, 36]. However, they are too optimistic even though they can serve as good starting points.

More realistic expressions have been found through extensive experimental data and will be provided later. However, again it should be said that these expressions are valid only for the helix parameters used during the experiments. Some of these experimental

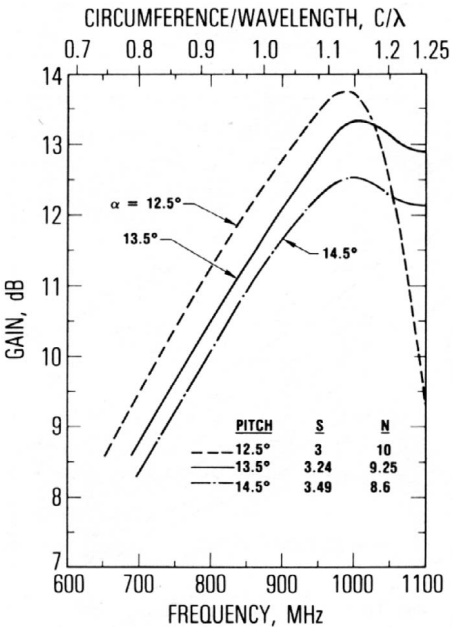


Figure 9.57 Gain versus frequency of a fixed-length, $NS = 30$ inches (762 mm) helix antenna; $\alpha = 12.5^\circ$, 13.5°, and 14.5° and $D = 4.3$ inches (109 mm). (Ref. 47.)

results are provided here for discussion [46, 47]. Figure 9.57 shows the gain variations for three helices, having pitch angles of 12.5°, 13.5°, and 14.5°. The length and diameter are 30 inches (762 mm) and 4.3 inches (109 mm). With these parameters the three helices will have from 8.6 to 10 turns. The gains peak at about 1 GHz, where the helix circumference is about 1.15λ, in close vicinity of the Hansen–Woodyard condition, shown in Figure 9.55. The peak gains change about 1 dB, and the maximum gain is for the pitch angle of 12.5°, which also has more turns. This antenna is near optimum and, as expected, its gain drops rapidly on both sides of the peak frequency. At the low frequency side all three antennas behave similarly with frequency, again in close agreement with the results of Figure 9.55, indicating an ordinary endfire array behavior.

Keeping the helix length and diameter constant at 30 inches and 10 turns, respectively, one must change both the number of turns and pitch angle. The performance of the resulting helices is shown in Figure 9.58. The peak gain variation with frequency is now limited to about 0.5 dB, and all gain curves behave similarly. However, the helix with the largest diameter and smallest pitch angle of 11.5° has now the highest peak gain, but its gain is most sensitive to frequency. This result also indicates that the pitch angle of 12.5°, which provided the highest gain in Figure 9.57, was not the true optimum.

The effect of helix number of turns is shown in Figure 9.59 [46]. These helices have from 5 to 35 turns, keeping other parameters constant at $\alpha = 12.8^\circ$, $S = 3.03$ inches (76.96 mm), and $D = 4.23$ inches (107.4 mm). The peak gain for $N = 5$ occurs at $C \approx 1.15\lambda$ and gradually decreases to $C \approx 1.07\lambda$ for $N = 35$, indicating the move

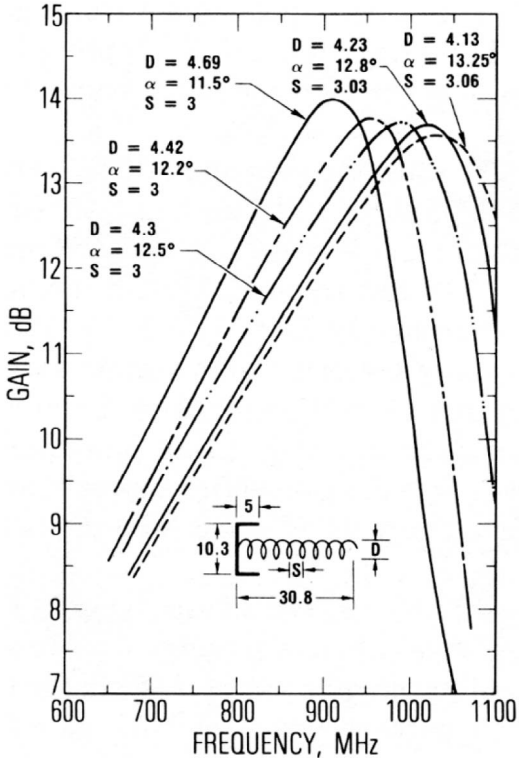


Figure 9.58 Gain of a fixed length $NS = 30$ inches (762 mm) helix antenna with different diameters. (Ref. 47.)

toward $C = \lambda$ for very long helices. This figure also shows the slope of the gain curves with frequency, being $f^{2.5}$ for the 5-turn helix and increasing to f^6 for the 35-turn one. Samples of the measured radiation patterns are shown in Figure 9.60. The axial ratios at the beam peaks, especially for the optimum patterns, are better than 1 dB, and the principal plane patterns have almost equal beamwidths and are within 1° of each other. At frequencies above the peak gain, the sidelobes gradually increase to merge with the main beam. This phenomenon can be due to the excitation of the next higher order mode that will radiate off-axis near the first sidelobe.

The results of the above studies can be summarized in Figure 9.61. It shows the dependence of the gain on the helix number of turns, when its diameter and pitch angles are fixed at $D = 4.23$ inches (107.4 mm) and 12.8°. The helix circumference decreases linearly to approach about 1λ, for very large helix turns. Figure 9.62 shows the dependence of the gain on the helix axial length. With larger helix diameters the peak gain occurs at lower lengths. Reducing the diameter to around one wavelength can further increase the gain. The most interesting results are in Figure 9.63: the gain beamwidth products decrease with the helix number of turns, from about 34,000 for a 5-turn helix to about 22,000 for the 35-turn helix. This trend is within expectation, as the number of

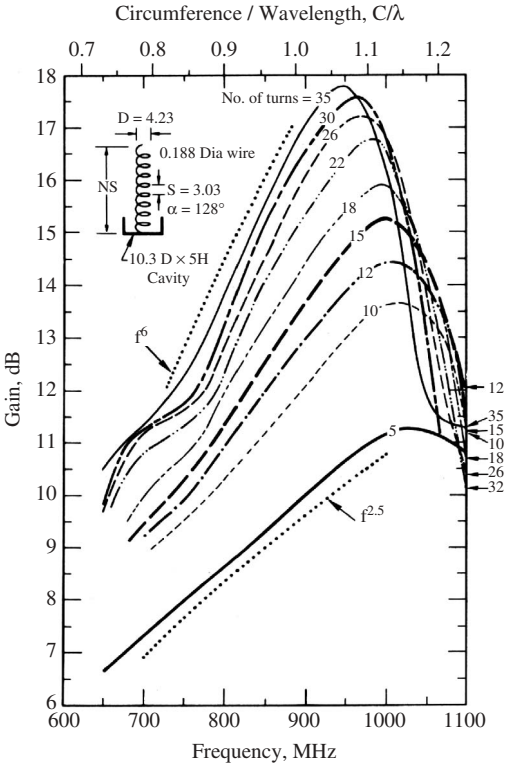


Figure 9.59 Gain versus frequency of a helix antenna with different numbers of turns; $D = 4.23$ inches and $\alpha = 12.8^{\circ}$. (Ref. 47.)

sidelobes increases with the helix turns. The results of this figure can be used in design studies, for selecting the initial helix parameters, to be numerically optimized.

9.3.5 Design Rules

Summarizing the results, the following empirical equations for the gain and half-power beamwidth (HPBW) are determined. Typical gain values fall within the 10–20-dB range, and the side levels are below −10 dB.

$$G_p = 8.3 \left(\frac{\pi D}{\lambda_p}\right)^{\sqrt{N+2}-1} \left(\frac{NS}{\lambda_p}\right)^{0.8} \left(\frac{\tan 12.5^{\circ}}{\tan \alpha}\right)^{\sqrt{N}/2} \tag{9.25}$$

and

$$\text{HPBW} \approx \frac{K_B \left(\dfrac{2N}{N+5}\right)^{0.6}}{\left(\dfrac{\pi D}{\lambda}\right)^{\sqrt{N}/4} \left(\dfrac{NS}{\lambda}\right)^{0.7}} \left(\frac{\tan 12.5^{\circ}}{\tan \alpha}\right)^{\sqrt{N}/4} \tag{9.26}$$

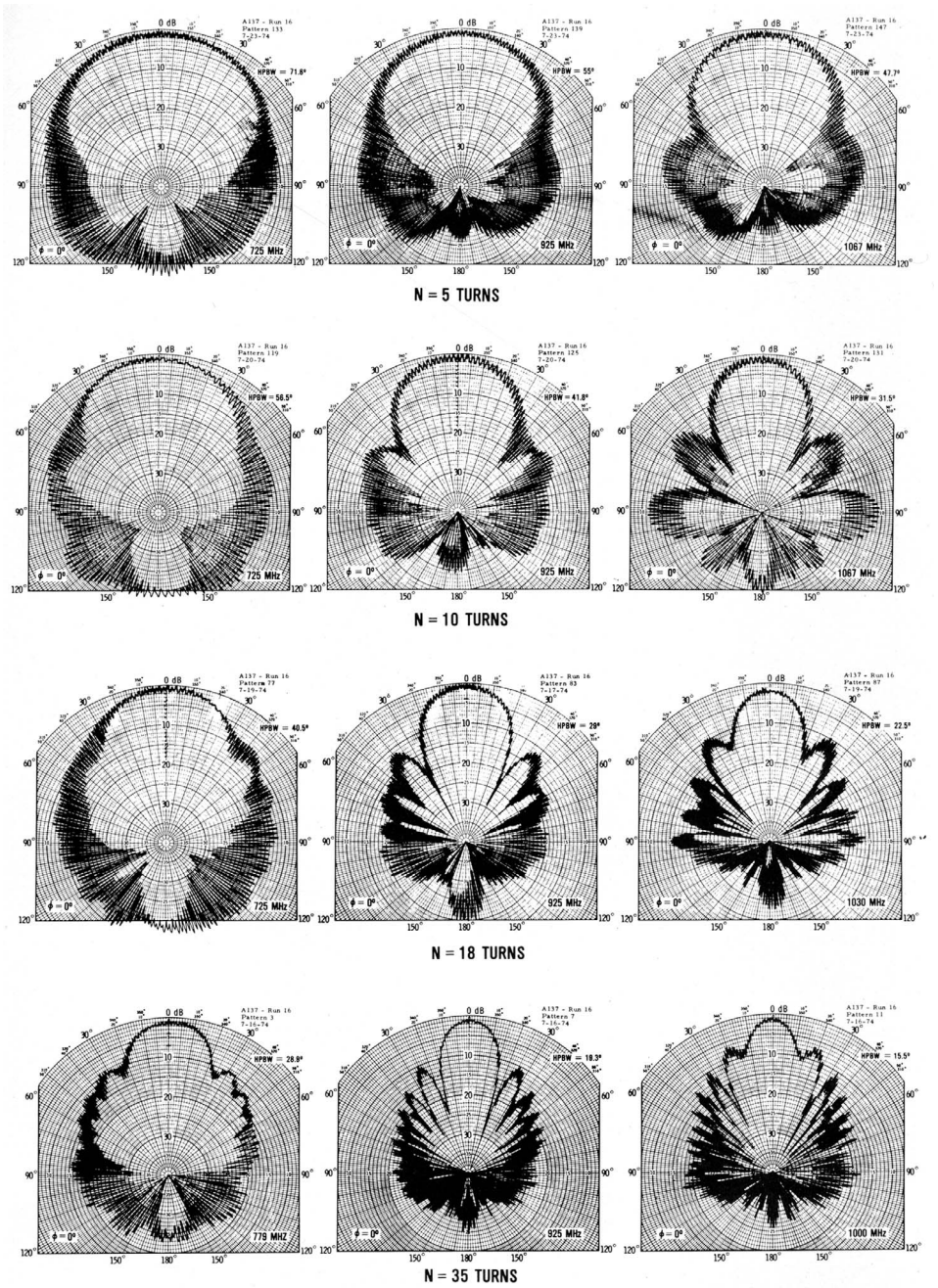


Figure 9.60 Typical radiation patterns of a 5- to 35-turn helix antenna; $\alpha = 12.8^{\circ}$ and $D = 4.23$ inches (107.4 mm). (Ref. 47.)

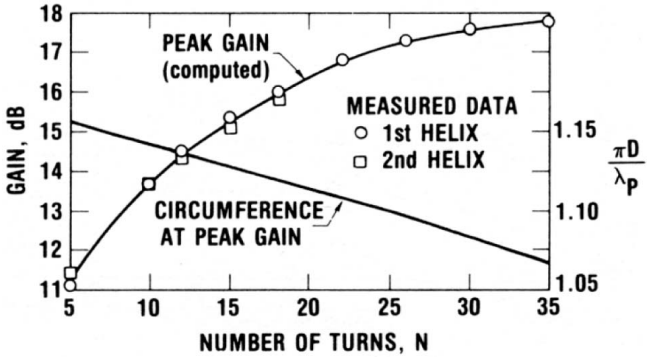


Figure 9.61 Peak gain of a helix antenna versus number of turns N; $D = 4.23$ inches and $\alpha = 12.8^\circ$. (Ref. 47.)

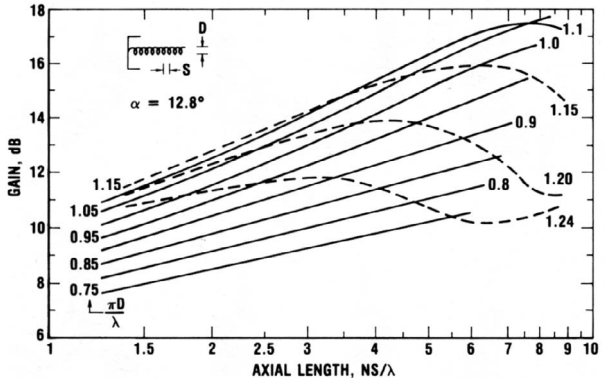


Figure 9.62 Helix antenna gains as a function of axial length with circumference as a parameter. (Ref. 47.)

These equations can be used as guides for selecting the helix size for a given application. Other useful design equations are

$$R_{\text{in}} \approx 140C/\lambda\ \Omega$$

$$0.15 < S/\lambda < 0.3$$

$$0.71 < C/\lambda < 1.20$$

$$12^\circ < \alpha < 14^\circ$$

These equations provide guidelines for the helix design in its conventional form and should not be assumed as hard facts. For instance, the input impedance R_{in} is dependent on the helix diameter at its input and can be adjusted by tapering. It is also dependent on the feed region geometry, feed type such as coaxial or microstrip, and the ground plane structure. Having a numerical analysis tool can facilitate optimization for improved gain, axial ratio, and bandwidth.

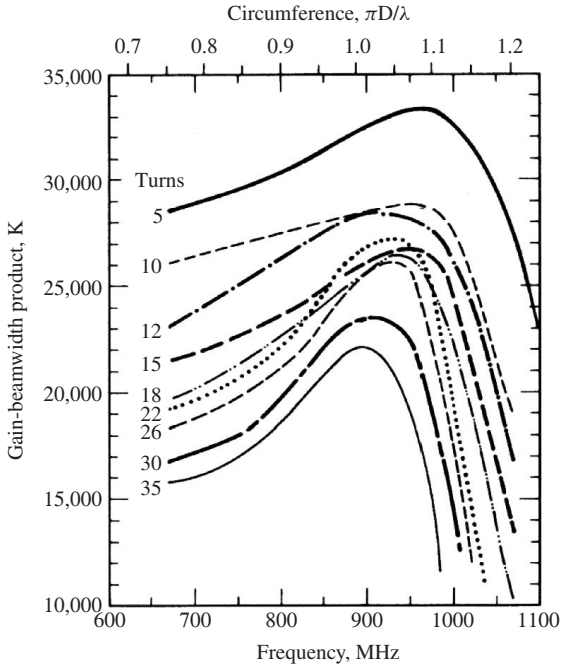


Figure 9.63 Gain beamwidth products of a 5- to 35-turn helix antenna; $\alpha = 12.8^{\circ}$ and $D = 4.23$ inches (107.4 mm). (Ref. 47.)

However, it is important to note that the helix is a traveling-wave antenna and has two sections—the exciter and the director regions. The input impedance is controlled by the exciter region and is stable, except for the end reflections. The helix antenna therefore has a wide impedance bandwidth, even though it is a gain-optimized antenna, similar to Hansen–Woodyard array. The radiation characteristics are controlled by the director region and can change widely, especially the gain and its bandwidth. For a uniform helix the geometrical parameters like the diameter, pitch angle, circumference, and number of turns can be optimized together to achieve a far-field performance goal. It can be designed to have a wide gain bandwidth, in which case the gain will have a moderate value as shown above. Or it can be designed for a maximum gain, beyond the 20-dB range. In this case, the antenna will approach a super gain condition and the gain bandwidth will be small.

9.3.6 Variations and Enhancements

A parametric study of nonuniform axial mode helices was performed numerically [48], which also included loading. It was found that adding the ground plane increases the gain by about 3 dB. The effect of the ground plane size on radiation from a uniform helix was investigated [49]. It was shown that reducing the ground plane size below the helix diameter causes the excitation of the backward mode, which can be used as a feed for reflector antennas. To understand this behavior, Figures 9.64 and 9.65 are included. Figure 9.64 shows the current distribution, phase progression, and radiation patterns of an eight-turn helix, $C = 1\lambda$, with a variable ground plane diameter. These results are typical

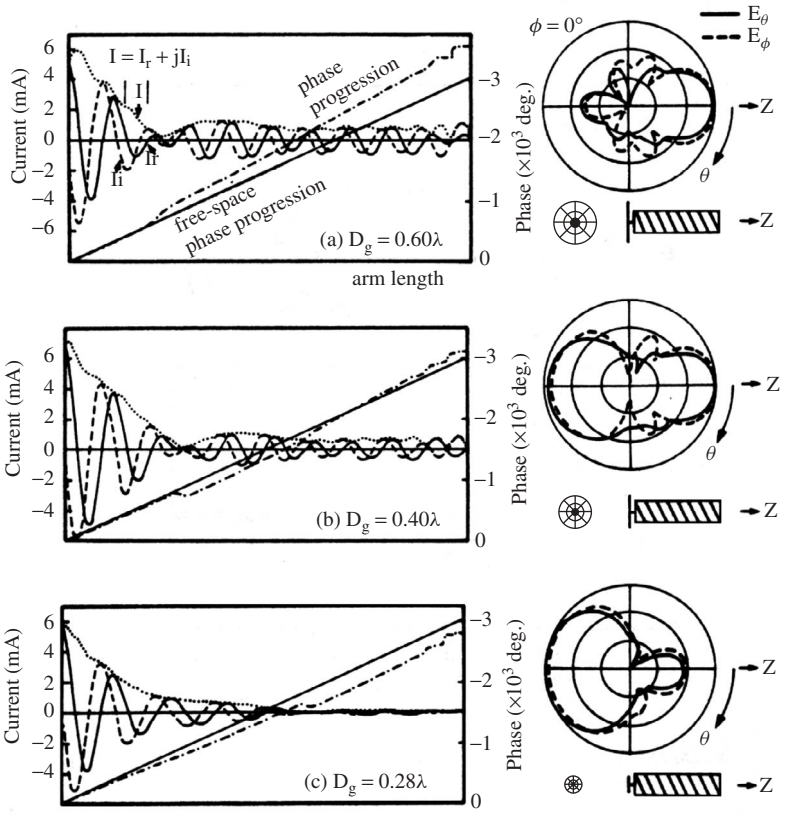


Figure 9.64 Current distribution and radiation patterns of a helix antenna with variable ground diameter D_g, $C = 1\lambda$, $\alpha = 12.5^\circ$, $N = 8$, and wire radius $= 0.005\lambda$: (a) $D_g = 0.6\lambda$, (b) $D_g = 0.4\lambda$, and (c) $D_g = 0.28\lambda$. (Ref. 49.)

of a conventional helix. The current decays in the exciter region and remains almost constant in the remaining director region. The radiation is in the forward axial direction. In Figure 9.65 the helix circumference and ground plane diameters are reduced to $C = 0.746\lambda$ and $D = 0.216\lambda$, respectively. The effects are striking. The current distribution decreases gradually throughout the helix and the phase progression becomes identical to that of free space. The direction of radiation reverses and the antenna radiates backward. Dual shaped helices for application in phased arrays were studied in Ref. 50. The element gain was used to reduce the number of elements in the array.

The enhancement of the gain has also been studied. In Ref. 51 a second parasitic helix was added symmetrical to the active helix, only in the directive S region. It was found that the currents of the parasitic helix resemble those of the active one, as shown in Figure 9.66. The antenna gain increased by about 1 dB and the radiation patterns improved, showing narrower beamwidths. The gain enhancement is not significant, but the method does not affect the overall helix geometry. The effect of the ground plane on the helix gain has been investigated in Ref. 52. It was shown that both cylindrical and conical caps around the ground plane Figure 9.67 can increase the helix gain. A numerical

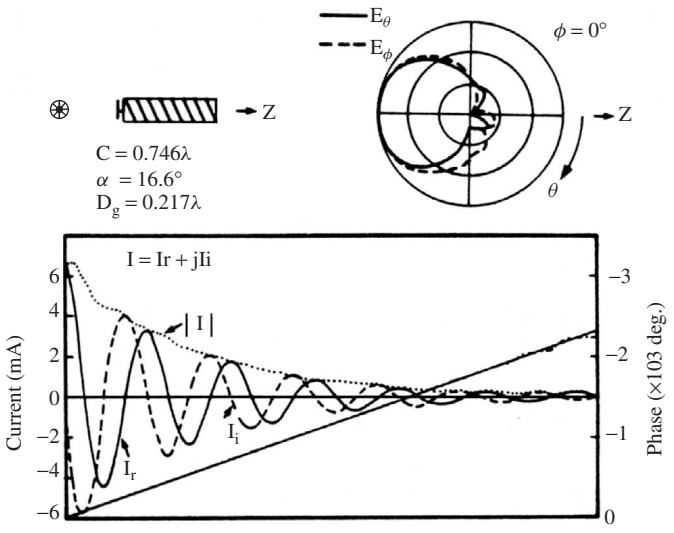


Figure 9.65 Current distribution and radiation patterns of a helix antenna with 0.217λ ground optimized for backfire mode, $C = 0.746\lambda$, $\alpha = 16.6^\circ$, $N = 8$, and wire radius = 0.005λ. (Ref. 49.)

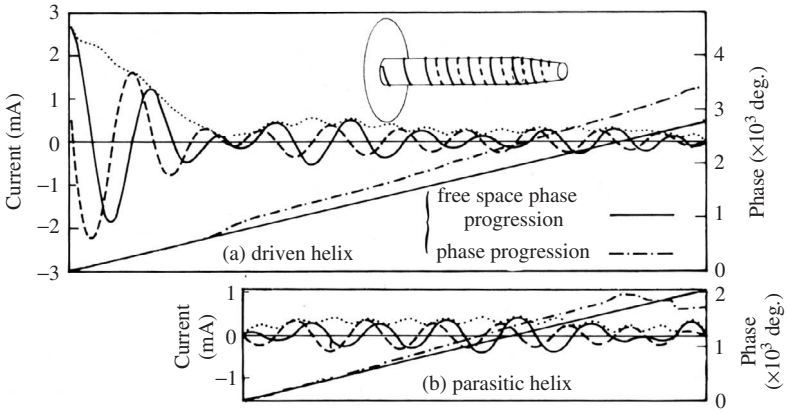


Figure 9.66 Current distribution of a helix antenna with a parasitic element with 0.6λ ground, $C = 1\lambda$, $\alpha = 12.5^\circ$, $N = 8$, and wire radius = 0.005λ. (Ref. 51.)

optimization was used to find the optimum cap shape. Figure 9.68 shows that the truncated cone shape increases the gain by as much as 4 dB.

To shape the radiation patterns, a spherical version of the axial mode helix has been studied [53, 54]. The helix wire is wound around a hemisphere and is placed over a ground plane (Figure 9.69). In comparison with the conventional helix, this antenna is a variation of the end-tapered helix. Its performance was investigated for different numbers of turns and other parameters. Figure 9.70 shows the behavior of the gain and axial ratio on the sphere diameter, in terms of the wavelength. Generally, the axial ratio performance

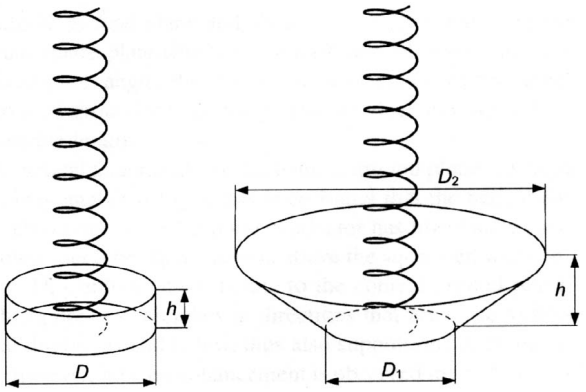


Figure 9.67 Helix antenna with cylindrical and truncated cone feeds, $L = 684$ mm, $D = 28$ mm, wire radius $= 0.3$ mm, $\alpha = 13.5^{\circ}$, and $N = 16.2$. (Ref. 52.)

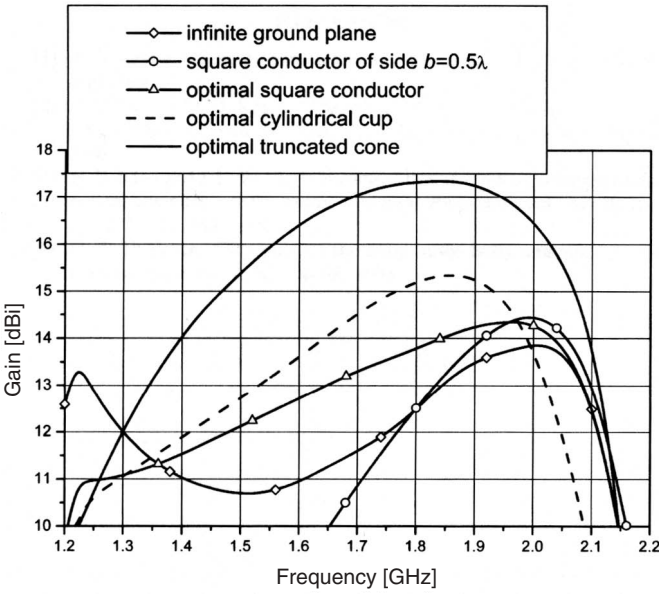


Figure 9.68 Computed gain of the helix antenna: infinite ground plane, square conductor of side $b = 0.5\lambda$, optimal square conductor $b = 1.5\lambda$, optimal cylindrical cup $D = 1\lambda$, $h = 0.25\lambda$ and optimal truncated cone, $D_1 = 0.75\lambda$, $D_2 = 2.5\lambda$, $h = 0.5\lambda$. (Ref. 52.)

for this helix is improved in comparison with short conventional helices. This antenna also provided wider radiation patterns (Figure 9.71).

9.3.7 Multifilar Helices

The helix antenna has another useful property in the rotational direction. It can be duplicated multiple times, by a simple rotation around its axis. If the rotation is 180°, then

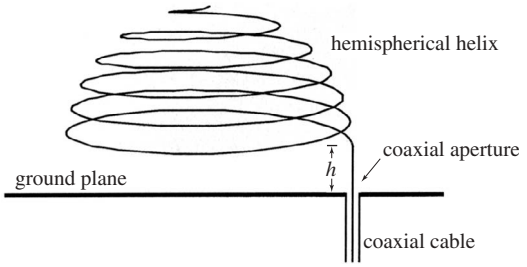


Figure 9.69 Geometry of a spherical helix antenna. (Ref. 53.)

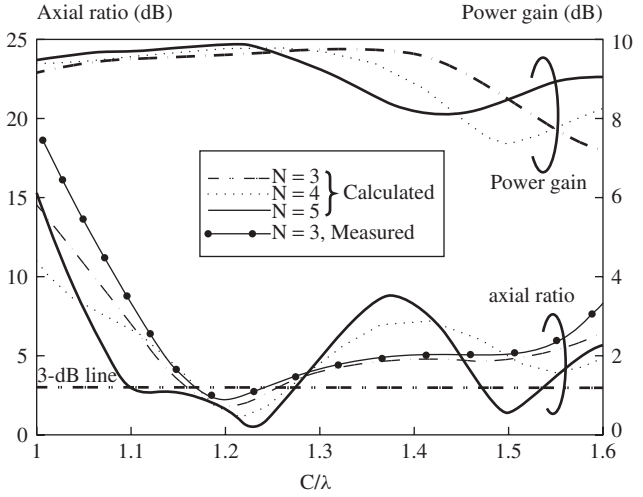


Figure 9.70 Variation of the axial ratio and power gain of the hemispherical helix antenna with C/λ and with number of turns N; $h = 0.5$ cm and wire radius $= 0.05$ cm. (Ref. 53.)

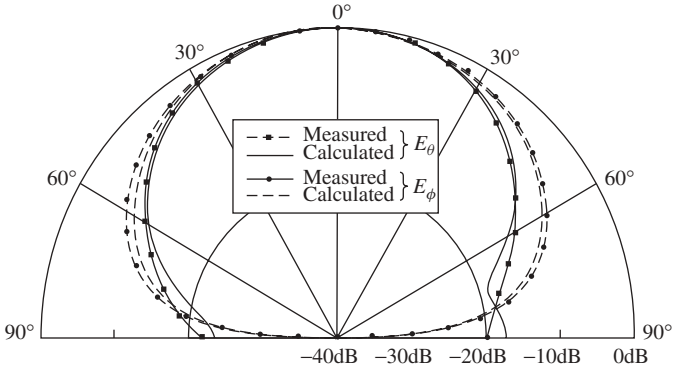


Figure 9.71 Radiation patterns of a five-turn hemispherical helix antenna over the plane with $\phi = 0^\circ$ and 180° and the dimensions $C/\lambda = 1.14$, $h = 0.5$ cm, and wire radius $= 0.05$ cm. (Ref. 53.)

a bifilar helix is generated. A 90° rotation four times generates a quadrifilar helix. Multifilar helices can similarly be generated, but the most useful new configurations have been the bifilar and quadrifilar geometries. These antennas are essentially arrays of a monofilar helix and can possess useful properties. For example, the symmetry of the input excitations can be used to force the excitation of only odd or even modes and generate a backward or forward wave. In particular, the quadrifilar helix can generate a forward or backward wave using the same antenna, simply by controlling the phase progression of its input terminals. This property makes its design very useful, as a single antenna can operate simultaneously in two different modes. Another important property of a quadrifilar helix is its multiple input terminals. By using an input excitation phase progression equal to the angular rotation of the arms, the antenna can generate circular polarization, regardless of its geometry or size. Consequently, the circumference rule of $C/\lambda \approx 1$ can be relaxed and a quadrifilar helix can be miniaturized, or its radiation patterns can be shaped to generate conical beams. For these reasons, the quadrifilar helix has become one of the most used and investigated antennas. The number of papers on the subject is extensive, and designs are developed for widespread applications in communications, remote sensing, and mobiles. Because it has so many different and useful properties, it is essentially a separate antenna from the monofilar helix and needs to be studied separately. Here, it will be discussed only briefly.

9.3.8 Bifilar Helix

A bifilar helix is generated by rotating the monofilar helix 180° about its axis. Because the configuration now has two helices, it can easily be fed by a balanced transmission line at either end. The antenna therefore does not need a ground plane. One of the interesting properties of the bifilar helix is its ability to generate a backward wave [35]. This property removes the need for a ground plane and opens new application areas, such as a reflector feed. To understand the operation principle of this antenna, three cases are numerically studied: a uniform bifilar helix (BH), a bifilar helix with an input taper (TBH), and a conical bifilar helix (CBH). Their geometries are shown in Figure 9.72 [42].

The tapered bifilar helix is shown in Figure 9.72a. The tapered feed section is described by $r = r_0 \exp(a\phi_c)$, with $r_0 = 3.8$ cm, $a = 0.048$/rad, and $\theta_0 = 12.5°$. The circumference of the uniform section is $C = 10$ cm. In both cases the helix wire radius and pitch angle are 0.5 mm and $\alpha = 12.5°$. The tapered helix of Figure 9.72a has $n_t = 2.2$ turns in the tapered section and $n_u = 3$ turns in the uniform section. The uniform helix of Figure 9.72b has five turns, so that the lengths of both helices are the same. The antennas were fed by a voltage source on the axis at the end, where a radial wire was placed between the two arms. The conical helix of Figure 9.72c is an extension of the tapered section and has the same axial length as the other two. Thus all three antennas have the same axial lengths.

The input impedance values are shown in Figure 9.73. The results for the uniform case are frequency dependent. However, they are similar for both the tapered bifilar helix and the conical bifilar helix. In both cases, the real and imaginary parts of the input impedance are independent of frequency, indicating wide impedance bandwidths. Thus the input tapering has improved its impedance characteristics.

The radiation characteristics are also studied and shown in Figures 9.74 and 9.75, respectively, for the gain and beamwidth, and the axial ratio. The surprising results are the gains and beamwidths. Contrary to monofilar helices, where the gain increases with

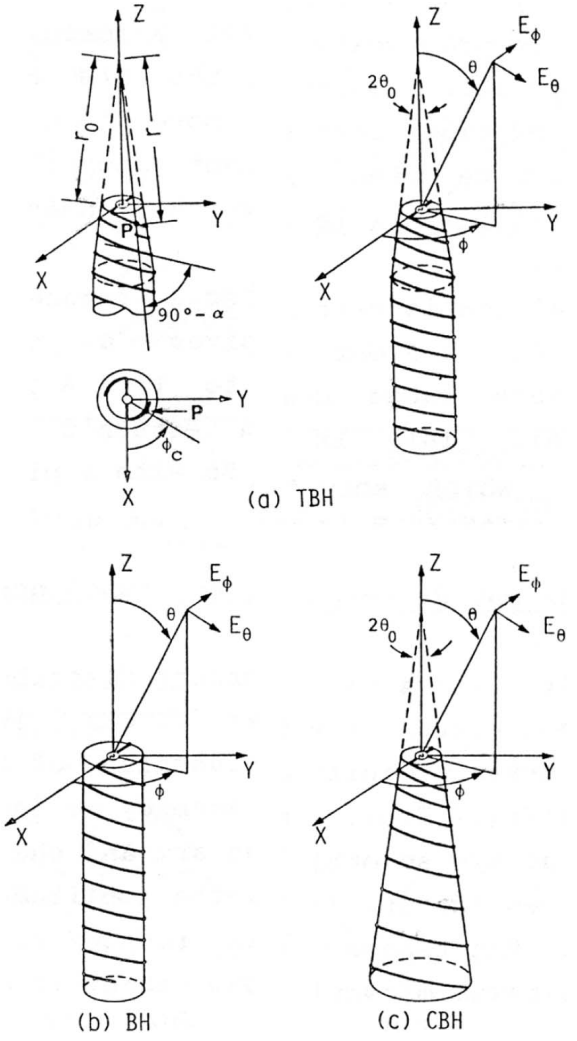


Figure 9.72 Configuration of bifilar helix antennas and coordinate system: (a) tapered feed region, (b) uniform, and (c) conical. (Ref. 42.)

frequency until about $C \approx 1.15\lambda$, the gain of the bifilar helix decreases with frequency. This is a useful property for the antenna miniaturization, where smaller diameters can be utilized. In fact, it has been shown that the bifilar helix radius can be reduced significantly by proper optimization of its parameters. Figure 9.75 shows the axial ratio performance, which is again superior for the tapered bifilar helix.

The backward wave excitation and superior performance of this antenna make it a good candidate as a reflector feed. In this case, the phase properties of the far field, in particular, its phase center location and stability with frequency, are important parameters. The latter also shows the most important region of the antenna far-field radiation. This may be a problem in small helix lengths, where the end reflections may be high and can cause dual phase centers. The minimization of the end reflection will be a desirable

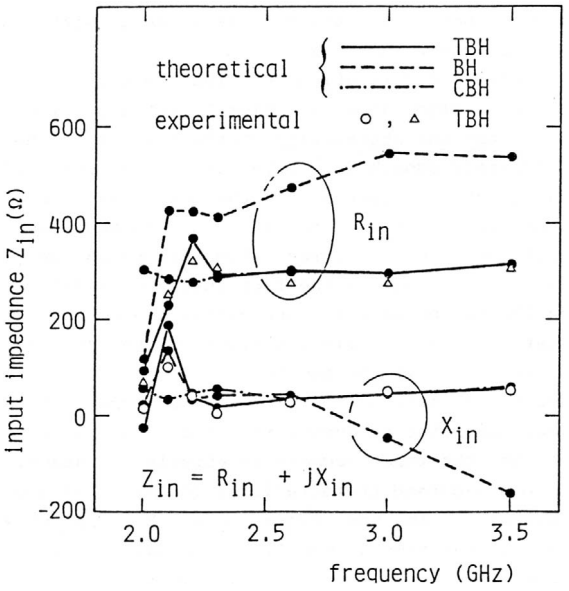


Figure 9.73 Input impedance of bifilar helix antennas. (Ref. 42.)

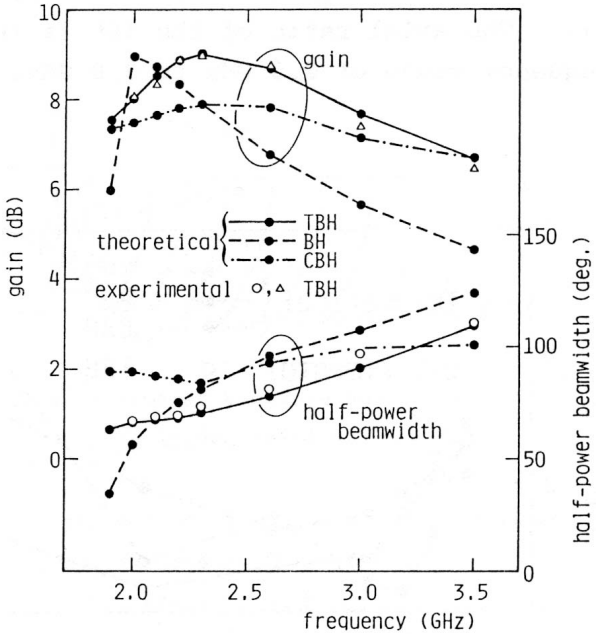


Figure 9.74 Half-power beamwidth and power gain of bifilar helix antennas. (Ref. 42.)

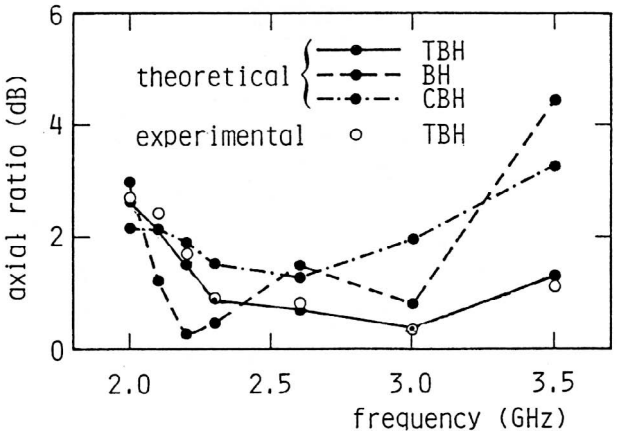


Figure 9.75 Axial ratio of bifilar helix antennas. (Ref. 42.)

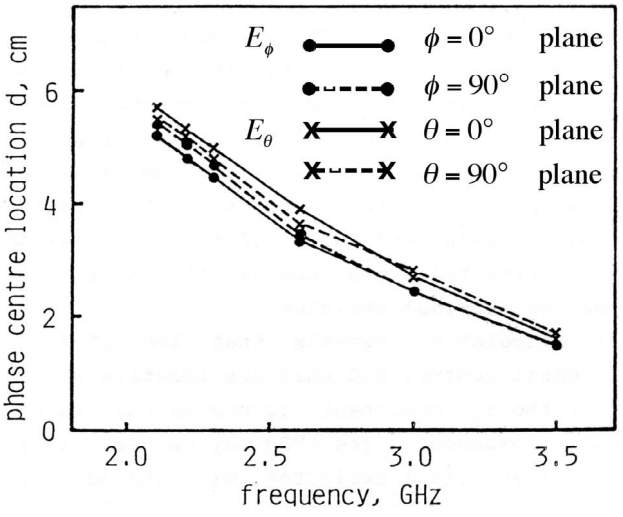


Figure 9.76 Phase center location of bifilar helix antennas. (Ref. 42.)

feature. Figure 9.76 shows the dependence of the phase center location on the frequency. In this figure the distance d is measured from the feed point end of the helix and should not be confused with the helix wire diameter, defined earlier. For most applications the variation of d will be within the reflector tolerances.

9.3.9 Quadrifilar Helix

A quadrifilar helix is formed by triple 90° rotations of a helix about its axis, to generate three additional ones. Thus each helix will be azimuthally separated from the next by 90°, and the antenna has four terminals at each end, as shown in Figure 9.77. The multiplicity of the turns and terminals at both ends add to the parameters of the helix and make this

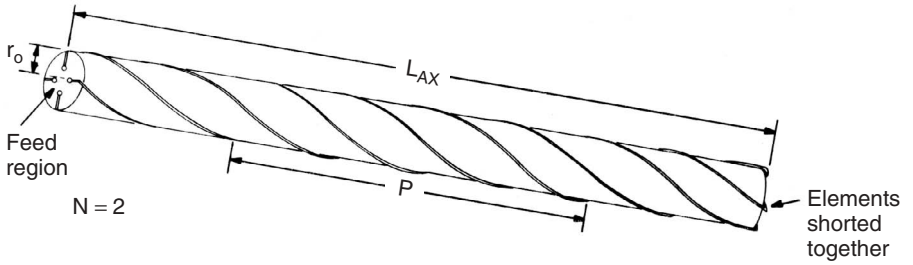


Figure 9.77 Geometry of a quadrifilar helix antenna. (Ref. 34.)

antenna very unique, which can produce an incredible array of radiation patterns, shaped beams, forward and backward waves, and geometrical configurations. As a result, the quadrifilar helix has generated a special class of its own and has become one of the most studied and used antennas in current times. For these reasons, it is not appropriate to lump it with the ordinary helix antenna. It must be studied separately and deserves a special category for itself. Here, some of its properties are briefly described.

Because it has terminals at both ends, a quadrifilar helix can be fed at either end. The remaining terminals at the other end can be shorted together or left open. Since in a helix antenna the direction of wire rotation dictates the sense of circular polarization, the input feed has two possibilities. The terminals can be fed clockwise in phase quadrature like 0°, 90°, 180°, and 270°, or they can be fed counterclockwise. Additionally, in the clockwise feed, the terminals can be fed with a reverse phase progression like 0°, -90°, -180°, and -270°. If the phase progression coincides with the sense of circular polarization, the antenna radiates a forward wave. Otherwise, a backward wave is launched to match the sense of circular polarization to feed the phase progression. Thus a quadrifilar helix antenna can function in both modes simultaneously, a significant advantage. Also, since the total phase progression at the feed terminals is 2π, the radiated field is automatically circularly polarized. This means the antenna radius can be decreased further, without affecting its axial ratio. This property also enables the selection of very diverse configurations, from very thin and tall designs, to very short and broad ones, which can meet very broad application requirements.

The axial mode of a quadrifilar helix can radiate both high gain and shaped conical beams. Since the axial mode characteristics for generating high gains were discussed in relation to the helix, for brevity only the shaped beam capability of the axial mode with a quadrifilar helix will be discussed. The axial $n = 1$ mode excitation is established by the feed phase angles, as described in the previous paragraph. This mode, with small pitch angle α, or pitch distance S, radiates a high gain axial beam. However, as the pitch angle or distance is increased, the beam gradually broadens and changes to a conical beam with its peak moving toward the horizon. Simultaneously, the helix diameter decreases, while its length increases, making the design more slender. Figure 9.78 shows the region of conical beam occurrence [55]. In this figure the parameter $k = r_0/p$ is the ratio of the helix radius to its pitch distance. Normally beam shaping improves with the number of turns.

Printed quadrifilar helices were investigated in Ref. 55, using an integral equation method. Both normal mode and axial mode designs were developed and experimentally verified. For the normal mode design the parameters were $C = 0.315\lambda$, $\alpha = 72^\circ$, $L = 2\lambda$, and $N = 1.8$. The main beam for this antenna was located between 75° and 105° off-axis,

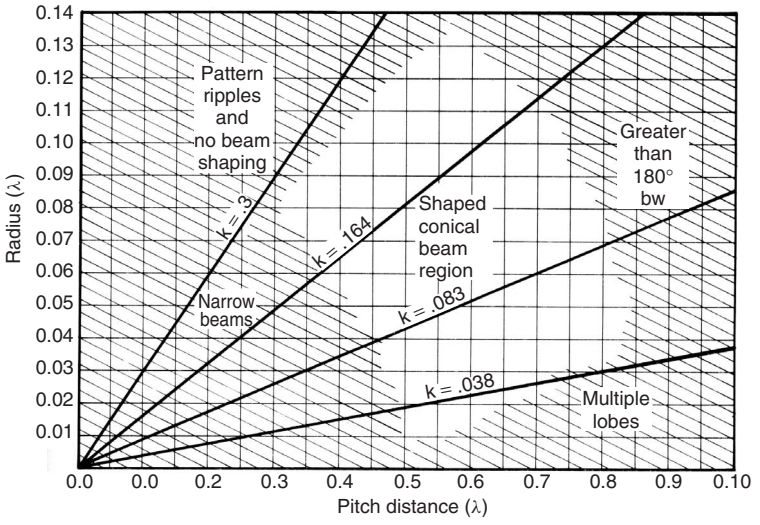


Figure 9.78 Region of conical beam radiation for a quadrifilar helix antenna. (Ref. 34.)

with sidelobes falling below -10 dB. The corresponding parameters for the axial mode design were $C = 0.56\lambda$, $\alpha = 60^\circ$, $L = 0.77\lambda$, and $N = 0.6$. Broadbanding of quadrifilar helices has also been investigated [56, 57]. In Ref. 56 the conducting width of each arm is increased from the input to the output. It is shown that the VSWR bandwidth can be doubled from 7% to 14% in the L-band.

A more effective means for increasing the bandwidth was developed in Ref. 57, by using a dual printed wire for each arm. Only one of the conductors is fed and the second one is parasitic. The geometry of this antenna, a folded parasitic quadrifilar helix antenna (FPQHA), is shown in Figure 9.79. The effects of the dual-arm spacing s, the dual-arm width ratio $r = W_a/W_b$, and the shorted section of dual arms d are shown in Figures 9.80–9.82. It is clear that these parameters have significant impact on the quadrifilar helix bandwidth. The geometry of an optimized antenna is shown in Figure 9.83 with parameters given by $L = 166$ mm, $r = 8$, $\alpha = 50^\circ$, $W_b = 16$ mm, $s = 18$ mm, $d = 0.5$ mm, $h = 0.127$ mm, and $\varepsilon_r = 2.2$, where h and ε_r are the thickness and relative permittivity of the dielectric tube on which the helix is printed. Figure 9.84 shows the input impedance performance of the optimized antenna. A bandwidth of 30% at 1.4 GHz is obtained, a significant improvement from a conventional design, with bandwidths around 6%. Figure 9.85 shows the radiation patterns at the center and edge frequencies, with very small cross polarizations. The measured gain within the band is shown in Figure 9.86. The gain variation within the band is about ± 1 dB.

9.4 YAGI–UDA ANTENNAS

9.4.1 Introduction

The Yagi–Uda antenna is a simple dipole array with a single excited element. The excitation of only one element of the array makes the structure very simple to design, fabricate, and use. It is also a very low cost antenna, as its elements can be fabricated

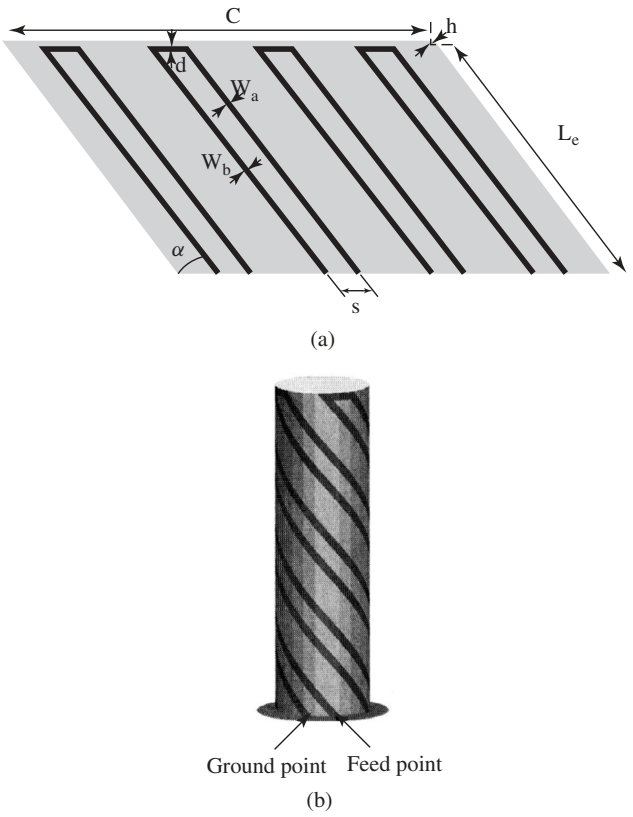


Figure 9.79 Geometry of folded parasitic quadrifilar helix antenna (FPQHA): (a) unwrapped and (b) wrapped. (Ref. 57.)

from low cost metallic rods or printed conductors on a dielectric. Its invention [58, 59] was perhaps one of the most important turning points in antenna engineering, which changed the widely accepted concept that a conductor will reflect the signal rather than guide it. It is also a very practical antenna, because of its simple feed structure, even though it is an array. As a result, it is used widely and in many diverse applications. However, because of the coupling between its elements, it is not a wideband antenna. For this reason, it will only be briefly discussed, as it is a traveling-wave antenna.

The original Yagi–Uda antenna was made of dipole-like wire conductors (Figure 9.87). However, they can be replaced by other resonant elements, like loops (Figure 9.88) [60, 61], bent wires, or disks. The greatest advantage of a dipole-like Yagi is its small volume. It is a planar structure with a very small cross section. The loop Yagi antenna, on the other hand, has a cylindrical shape and is three-dimensional, a disadvantage. However, because the radiation beam is axial, several loop Yagi arrays, having increased frequency of operation, can be placed coaxially, one inside the other. In this manner, dual-band or multiple-band loop Yagi arrays can be designed and can function satisfactorily, an advantage.

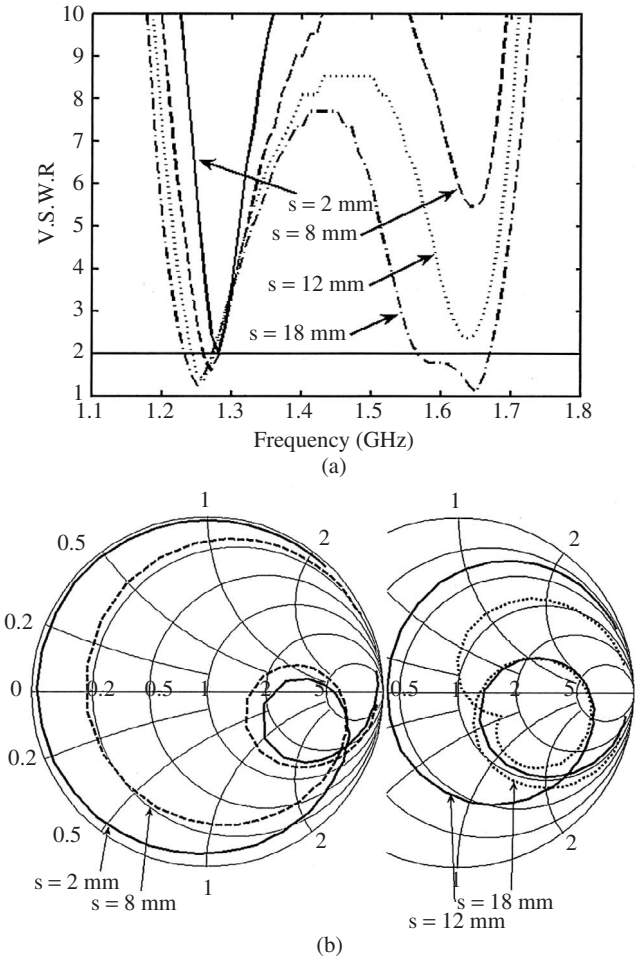


Figure 9.80 (a) FPQHA VSWR and (b) input impedance in Smith chart versus frequency for different values of s. (Ref. 57.)

9.4.2 Operating Principle

Electrically, a Yagi antenna is a traveling-wave type antenna and, as such, has two separate parts—the exciter region and the director region. In the original design, all elements were dipoles and the exciter region consisted of two elements—an excited dipole and a passive reflector. The director region can contain multiple elements, depending on the design or gain requirements. The operation is feasible because of the fact that a conducting wire having a length slightly larger than $\lambda/2$ reflects the incident wave efficiently, and when its length reduces slightly below $\lambda/2$, it guides the wave. The difference in the lengths is only a few percent. This property of the Yagi antenna can be understood from the observation of the element currents. Figure 9.89 shows the current distribution on the elements of a 27-element Yagi array, with 25 directors [62]. It is evident that the

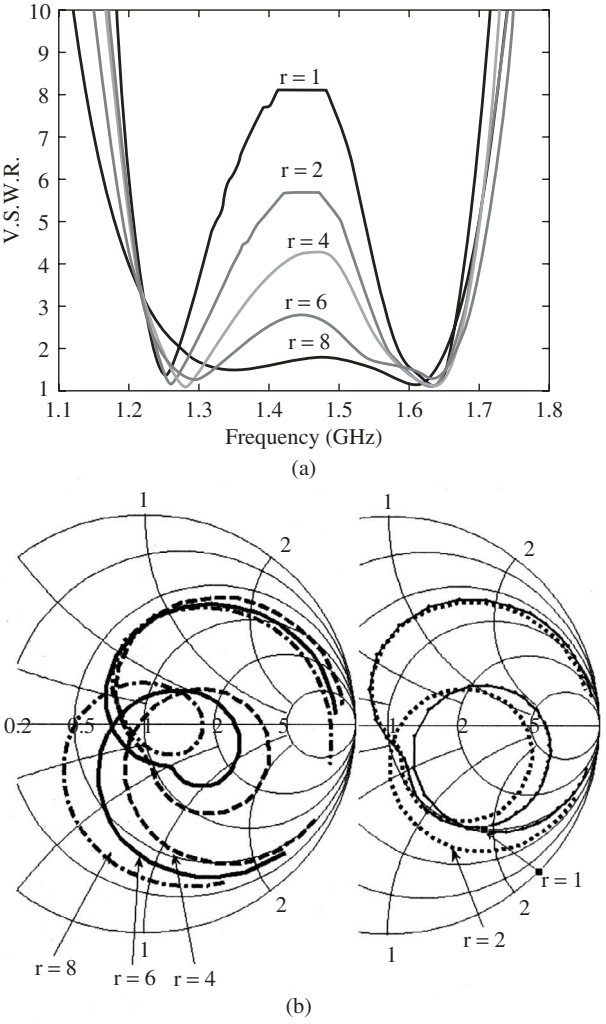


Figure 9.81 (a) FPQHA VSWR and (b) input impedance in Smith chart versus frequency for different values of r. (Ref. 57.)

currents and the reflector on the excited element are high and decrease significantly on the directors. In addition, the magnitude of the current stays relatively constant on all directors. This indicates that there is a traveling wave over the directors. Because of this property, the two regions of a Yagi antenna can be designed separately. In particular, the exciter region can be made of different antenna types. For instance, the exciter region of a dipole Yagi can be replaced by two loops, a microstrip antenna over a ground, or any other endfire antenna like an open-ended waveguide. In particular, the excited element can be selected to be a wider band antenna than a dipole, like a folded or sleeve dipole, or a wideband microstrip antenna. Such designs will improve the input impedance bandwidth, without significantly affecting the radiation characteristics.

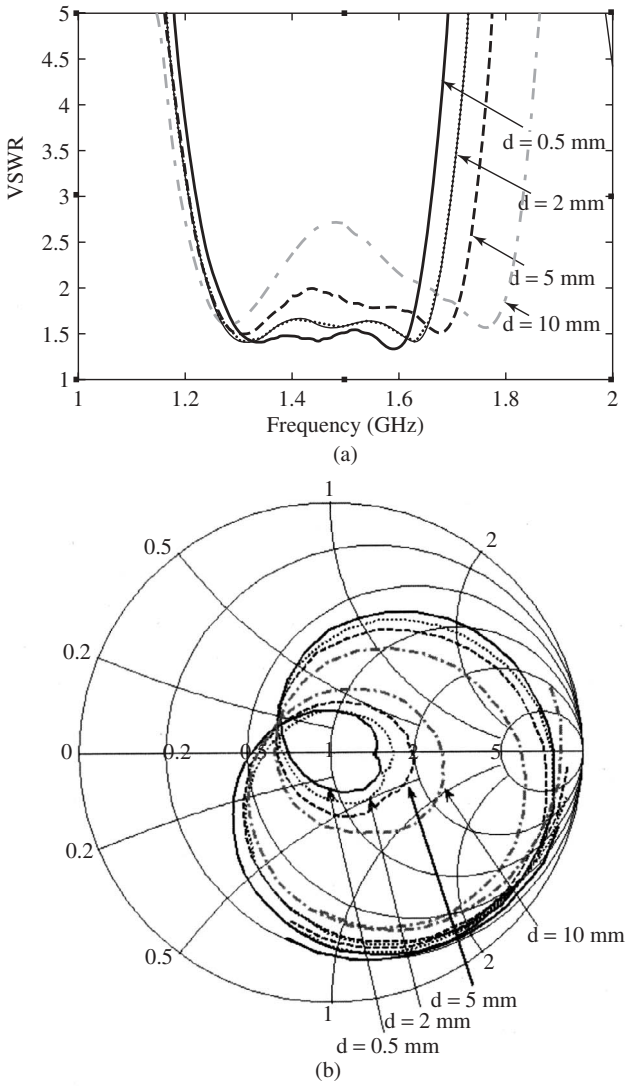


Figure 9.82 (a) FPQHA VSWR and (b) input impedance in Smith chart versus frequency for different values of d. (Ref. 57.)

9.4.3 Design Procedure

Because the Yagi–Uda antenna has a simple geometry, it can readily be designed and optimized by a numerical approach. However, certain design rules will help with initial design or design of general purpose antennas without optimization.

In dipole Yagi antennas, the reflector is about 1–2% larger than the exciter. Once the excited element is selected, the reflector size can be chosen as stated. The reflector size also has a significant effect on the antenna front-to-back ratio and the input impedance.

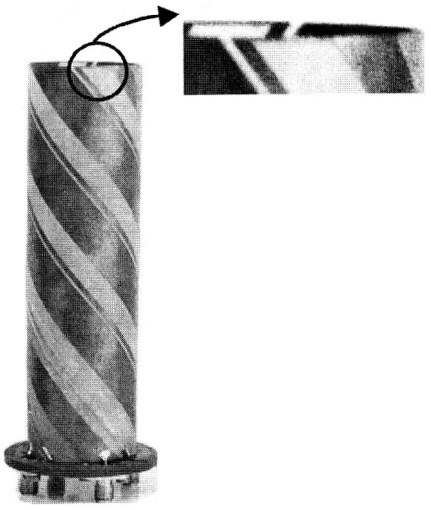


Figure 9.83 Picture of the prototype FPQHA zoomed on the shorted section. (Ref. 57.)

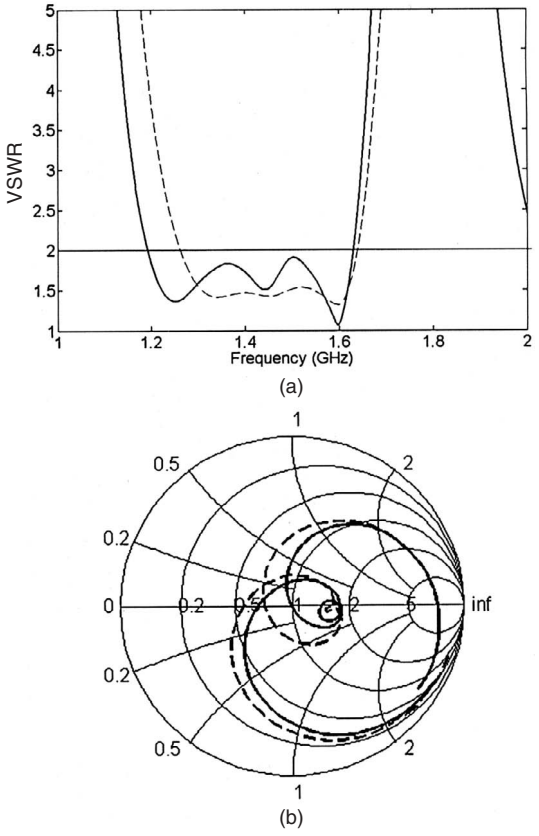


Figure 9.84 (a) VSWR and (b) input impedance in Smith chart of FPQHA (_____ measured results; ------ theoretical results). (Ref. 57.)

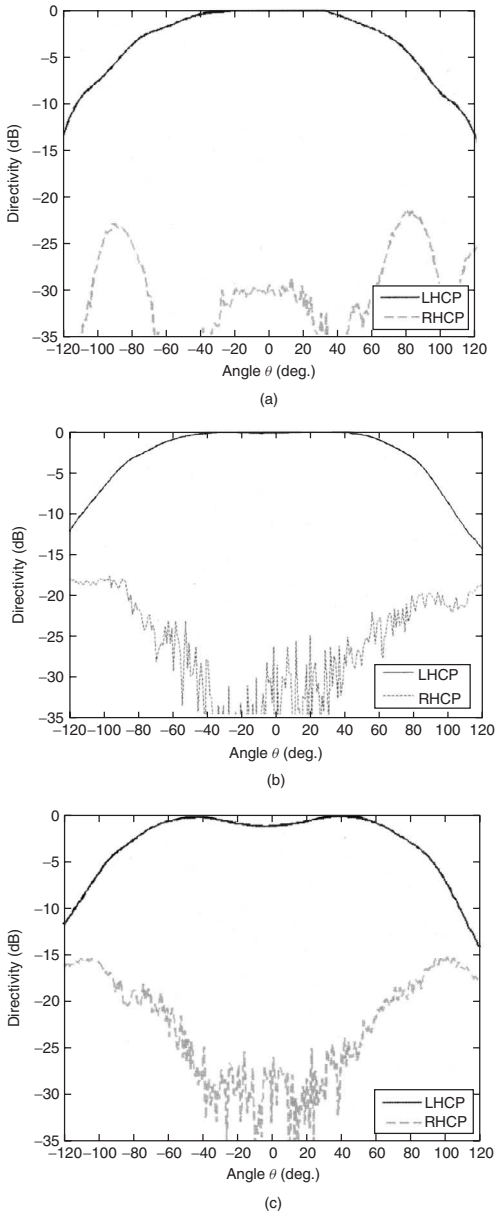


Figure 9.85 Measured radiation pattern in circular polarization of the FPQHA: (a) $f = 1.2$ GHz, (b) $f = 1.4$ GHz, and (c) $f = 1.6$ GHz. (Ref. 57.)

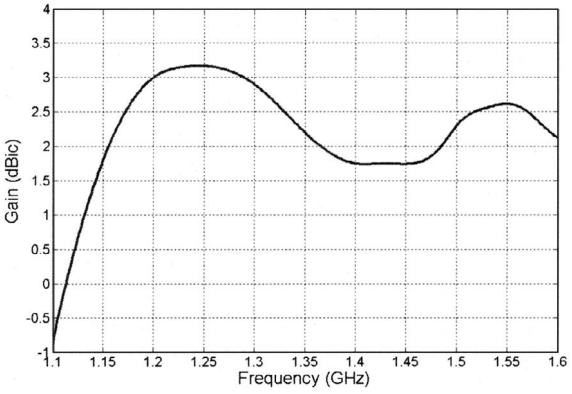


Figure 9.86 Measured gain versus frequency of the FPQHA. (Ref. 57.)

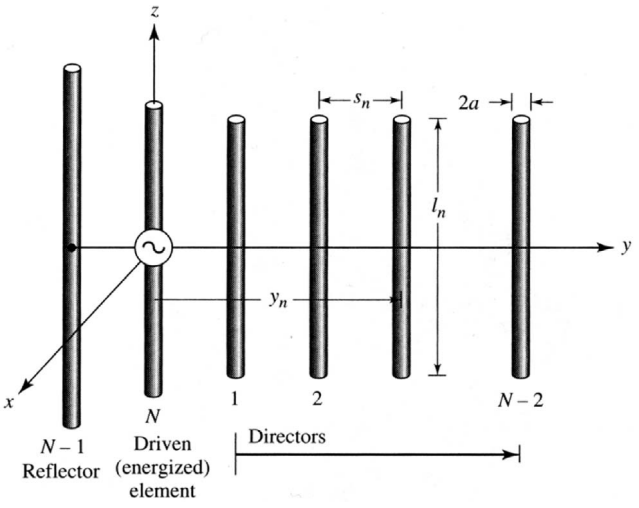


Figure 9.87 Yagi–Uda antenna configuration.

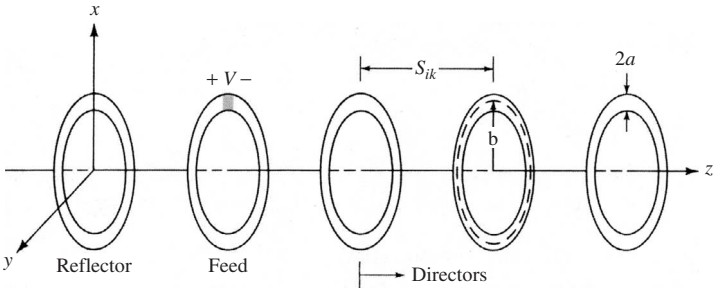


Figure 9.88 Yagi–Uda array of circular loops.

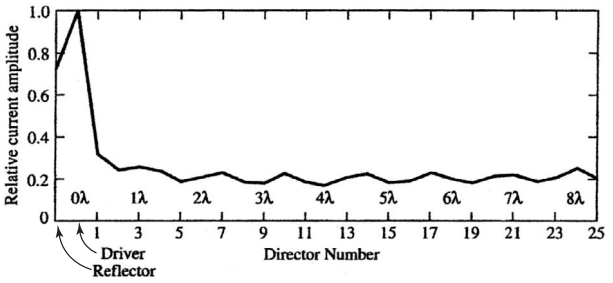


Figure 9.89 Relative current amplitudes for the 27-element Yagi array. (Ref. 62.)

For these reasons its size may be modified to improve these parameters. The effect on the forward gain will be small. The directors must be a few percent smaller than the excited element. There are ample published data to help in selecting the element sizes and their spacing [62, 63]. End tapering may also be used, which will improve the impedance bandwidth, similar to the helix antenna. Typically, Yagi antennas have gains at around 10–15 dBi. However, with proper optimization much higher gains, approaching 20 dBi, can also be achieved [64].

With a loop Yagi array, the design rules are similar and are summarized as follows:

- The circumference of the excited element is chosen as $2\pi b_2 = 1.1\lambda$ (b_2 is its radius), to make its input impedance real.
- The reflector size is selected as $2\pi b_1 = 1.05\lambda$. The size of the reflector affects the front-to-back ratio, making the input impedance stronger than the array gain.
- The spacing of the excited element with the reflector should be about 0.1λ. Its effects are similar to the reflector size.
- The sizes of the directors are given as $2\pi b_n = 0.7\lambda$, where b_n are the radii, and are chosen to be the same for all directors.
- Spacing of the directors is 0.25λ and is uniform.
- The wire radius a is the same for all elements and is calculated from the relationship $2\ln(2\pi b_2/a) = 11$.

9.4.4 Design Optimization

There is a significant volume of work on optimization of Yagi antennas. However, most of these studies deal with improvement of the gain [65–67] or gain bandwidth [68], which naturally comes at the expense of the impedance bandwidth. A joint gain and impedance optimization has also been done [69, 70], which shows that trying to maintain a specified bandwidth limits the gain performance. In recent years, printed Yagi antennas have become popular for high frequency and millimeter-wave frequency applications, and the number of studies is significant. However, they primarily address two different goals: the high gain [64] and broadband [71] performances. These antennas use unconventional feeds, basically different forms of a microstrip line that can be printed on one side of a dielectric substrate. Such a design facilitates antenna fabrication at high microwave and millimeter frequencies.

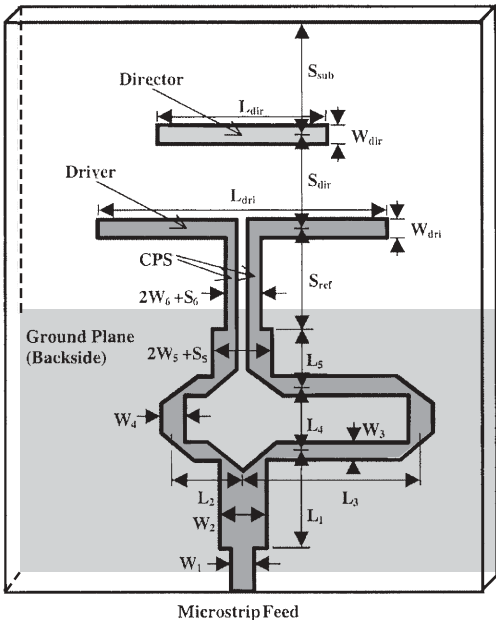


Figure 9.90 Geometry of the printed quasi-Yagi antenna. $W_1 = W_3 = W_4 = W_5 = W_{dri} = W_{dir} = 0.6$, $W_2 = 1.2$, $W_6 = S_5 = S_6 = 0.3$, $L_1 = L_5 = 1.5$, $L_3 = 4.8$, $L_4 = 1.8$, $S_{ref} = 3.9$, $S_{dir} = 3$, $S_{sub} = 1.5$, $L_{dri} = 8.7$, and $L_{dir} = 3.3$ (dimensions in mm). (Ref. 71.)

The broadband geometry is shown in Figure 9.90, where a microstrip line is split into two separate lines for balance feeding of a printed dipole antenna with a 180° phase difference [71]. A single director is used to improve the gain performance. Its input return loss is shown in Figure 9.91 There is a frequency shift between the simulated and measured results, but both bandwidths are large, on the order of 48%. The corresponding gain results are shown in Figure 9.92. Within the impedance bandwidth the simulated

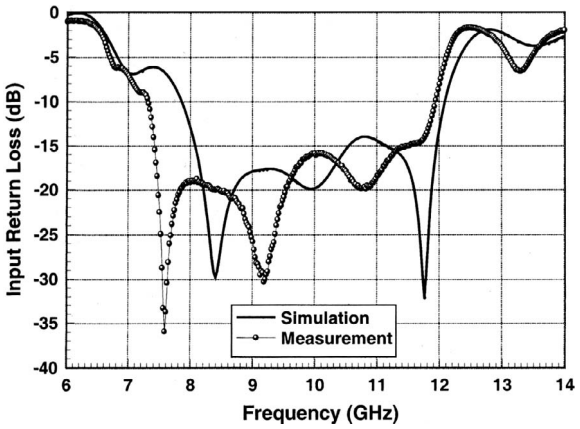


Figure 9.91 Finite-difference time-domain (FDTD) simulation and measured input return loss characteristic of the prototype quasi-Yagi antenna. (Ref. 71.)

gains remain near 5 dBi, but the measured ones are frequency dependent, which is a characteristic of an optimized antenna that becomes sensitive to fabrication and other tolerances.

The high gain design in Ref. 64 has used an undulated crank-type printed line as the exciter, which feeds 30 directors for a computed gain of about 19 dBi at 30 GHz. Figure 9.93 shows its geometry, which is ideal for high frequency and millimeter-wave applications, where the antenna dimensions are physically very small and the crank feed is simply made by undulating the feeding microstrip line. In addition, the feed is a traveling-wave type, which is less sensitive to tolerances, and its impedance bandwidth can be very large. However, this is not the measure of the antenna bandwidth, since its gain bandwidth is much smaller. An experimental model was fabricated for operation at 15 GHz. Its measured input impedance and return loss plots are shown in Figure 9.94,

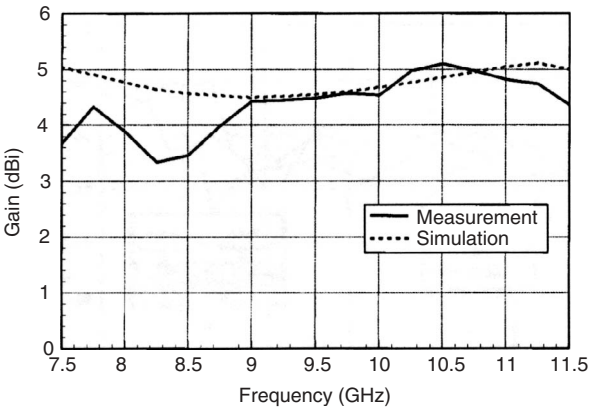


Figure 9.92 Simulated and measured gain of the quasi-Yagi antenna. (Ref. 71.)

Figure 9.93 Geometry of the high Yagi antenna with crank feed. (Ref. 71.)

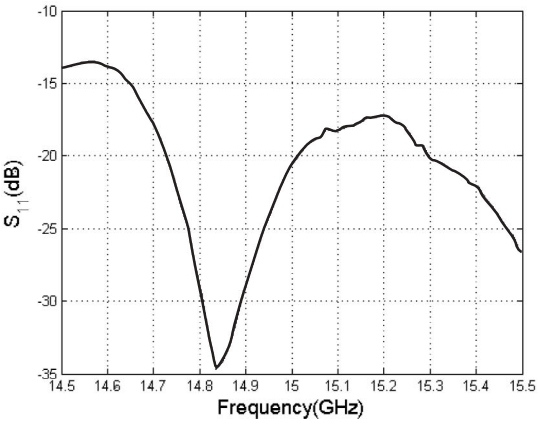


Figure 9.94 Measured return loss of the Yagi antenna with crank feed.

where the return loss is very small, below 20 dB. For the same frequency band the measured gain is shown in Figure 9.95. The feed line loss was about 0.9 dB, making the measured peak antenna gain to be 17.51 dBi. The simulated antenna directivity was about 19 dBi, indicating a discrepancy of 1.5 dB. This is reasonable, since the antenna was designed in free space but was etched on 0.8-mm dielectric with a permittivity of 2.5. Considering the dielectric and conductor losses, etching imperfection, and detuning effects of the dielectric on an optimized high gain antenna, this discrepancy is quite acceptable. The antenna gain rises rapidly but decreases gradually at high frequencies. Samples of its radiation patterns are shown in Figure 9.96, showing almost equal beamwidths in the principal planes.

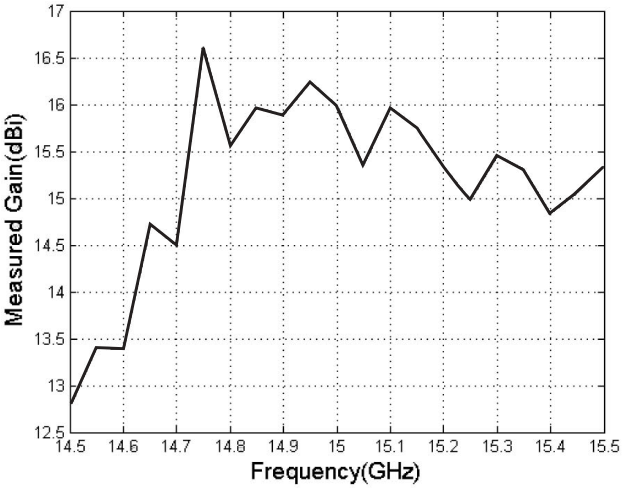


Figure 9.95 Measured gain of the Yagi antenna with crank feed.

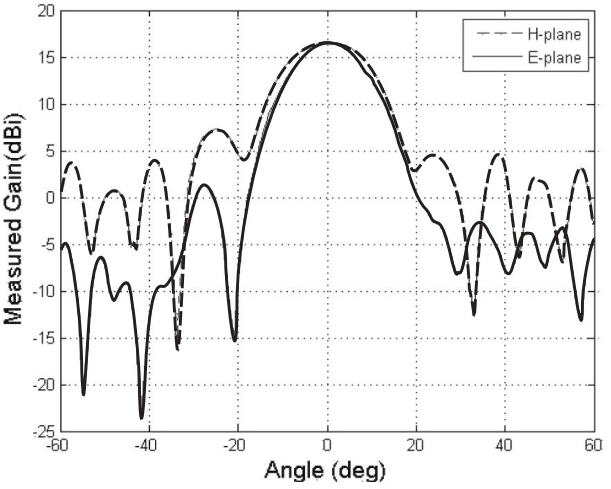


Figure 9.96 Measured radiation patterns of the Yagi antenna with crank feed.

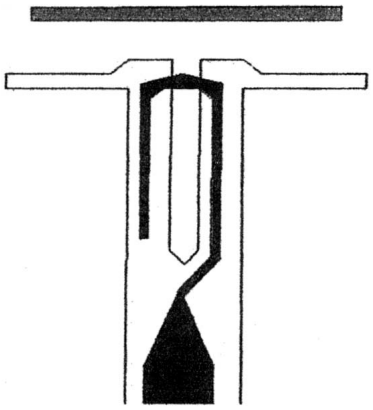


Figure 9.97 Feed geometry of the dipole feed for the 30-element Yagi antenna; only one director is shown. (Ref. 72.)

Another high gain 30-element Yagi antenna was designed using a genetic algorithm to maximize its gain [72]. A wideband printed dipole, shown in Figure 9.97, was used as its feed, where only a single director is shown for brevity. Its simulated directivity performance versus frequency is shown in Figure 9.98. In this design the frequency dependence of the directivity is opposite to the previous case of Figure 9.95. Here, the directivity increases slowly with frequency but drops rapidly after its peak of 18.3 dBi at 10 GHz. Table 9.1 shows the physical lengths of its elements. Figures 9.99 and 9.100 show its near-field amplitude and phase distributions.

The above examples show that a Yagi–Uda array has two distinct parts—the exciter and the directors. They function reasonably independently from each other, and each can be designed separately to optimize the antenna performance. The input impedance

TABLE 9.1 Dimensions in Wavelength of the Optimized 30-Element Yagi Antenna with Dipole Feed and Radius = 0.004λ

Element	Length	Spacing	Element	Length	Spacing
Feed	0.5000	0.0000	Dir.1	0.4620	0.1000
Dir.2	0.3960	0.1460	Dir.3	0.4410	0.1430
Dir.4	0.3870	0.1960	Dir.5	0.4240	0.2980
Dir.6	0.4240	0.3480	Dir.7	0.3420	0.1370
Dir.8	0.3980	0.2460	Dir.9	0.3940	0.1420
Dir.10	0.4270	0.3130	Dir.11	0.4010	0.3500
Dir.12	0.4030	0.3470	Dir.13	0.3940	0.2720
Dir.14	0.3960	0.2650	Dir.15	0.4080	0.3170
Dir.16	0.4100	0.3480	Dir.17	0.2730	0.0540
Dir.18	0.3750	0.2870	Dir.19	0.3460	0.1890
Dir.20	0.4010	0.2630	Dir.21	0.4010	0.3350
Dir.22	0.3090	0.2770	Dir.23	0.3800	0.1660
Dir.24	0.3800	0.2960	Dir.25	0.4030	0.2790
Dir.26	0.4130	0.3120	Dir.27	0.3680	0.3450
Dir.28	0.4150	0.3500	Dir.29	0.4220	0.2750

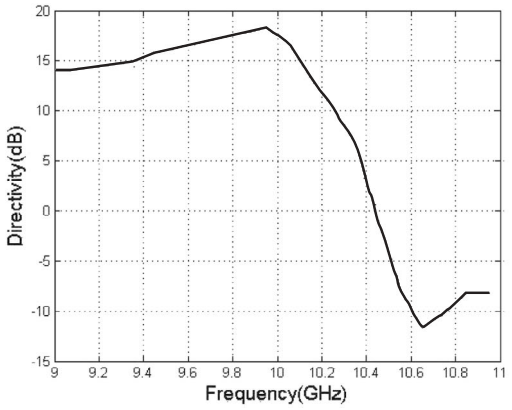


Figure 9.98 Simulated directivity of the Yagi antenna with dipole feed.

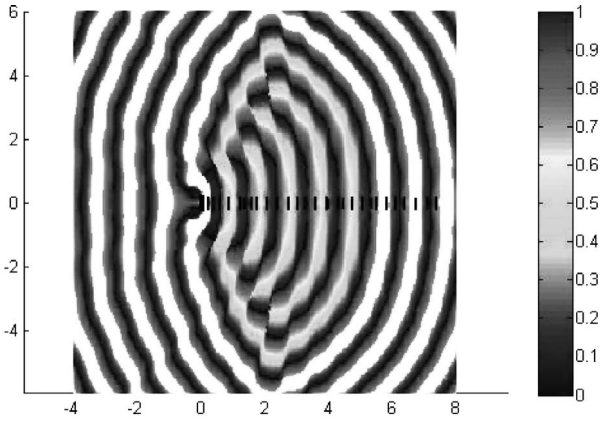


Figure 9.99 Near-field phase distribution for the Yagi antenna with dipole feed.

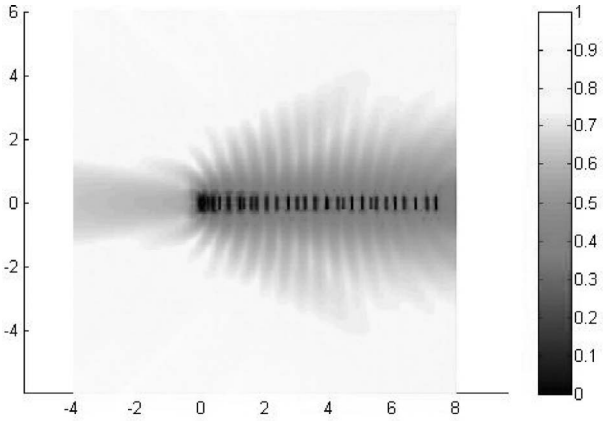


Figure 9.100 Near-field amplitude distribution for the Yagi antenna with dipole feed.

bandwidth is controlled by the exciter region and can be quite wide. The directivity is controlled by the director region and has a constant gain bandwidth product. As the antenna directivity is increased, its related bandwidth decreases accordingly. For high gain designs, the antenna bandwidth is controlled by the bandwidth of its directivity.

REFERENCES

1. C. A. Balanis, *Antenna Theory: Analysis and Design*, 3rd ed., John Wiley & Sons, Hoboken, NJ 2005.
2. C. T. Tai, *Antenna Engineering Handbook*, 2nd ed., McGraw-Hill, New York, 1984, Chap. 4.
3. A. J. Poggio and E. K. Miller *Antenna Handbook, Theory, Applications and Design*, Van Norstrand, New York, 1988, Chap. 3.
4. A. G. Kandoian, Broadband antennas, U.S. Patent 2,368,663, 1945.
5. R. A. Burberry, *VHF and UHF Antennas*, Peter Peregrinus Ltd., London, 1992, Chap. 8.
6. R. A. Burberry, *Handbook of Antenna Design*, Vol. 2, Peter Peregrinus Ltd., London, 1982, Chap. 17.
7. W. Q. Malik, D. J. Edwards, and C. J. Stevens, Angular-spectral antenna effects in ultra wideband communications links, *IEE Proc. Commun.*, Vol. 153, No. 1, pp. 99–106, February 2006.
8. J. Kim and S. Park, Novel ultra-wideband discone antenna, *Microwave Optical Techno. Lett.*, Vol. 42, No. 2, pp. 113–115, July 2004.
9. J. R. Bergmann, On the design of broadband omnidirectional compact antennas, *Microwave Optical Technol. Lett.*, Vol. 39, No. 5, pp. 418–422, December 2003.
10. L. Januszkiewicz and M. Czarnecki, Simulation of a broadband antenna with the method of moments, *Radioengineering*, Vol. 14, No. 4, pp. 43–47, December 2005.
11. K. Kim, J. Kim, and S. Park, An ultrawide-band double discone antenna with the tapered cylindrical wires, *IEEE Trans. Antennas Propag.*, Vol. 53, No. 10, pp. 3403–3406, October 2005.
12. P. T. Teo, K. S. Lee, Y. B. Gan, and C. K. Lee, Development of bow-tie antenna with an orthogonal feed, *Microwave Optical Technol. Lett.*, Vol. 35, No. 4, pp. 255–257, November 2002.
13. Y. Lin and S. Tsai, Analysis and design of broadside-coupled striplines-fed bow-tie antennas, *IEEE Trans. Antennas Propag.*, Vol. 46, No. 3, pp. 459–460, March 1998.
14. G. Zheng, A. A. Kishk, A. W. Glisson, and A. B. Yakovlev, A broadband printed bow-tie antenna with a simplified balanced feed, *Microwave Optical Technol. Lett.*, Vol. 47, No. 6, 534–536, December 2005.
15. A. A. Eldek, A. Z. Elsherbeni, and C. E. Smith, Wideband microstrip-fed printed bow-tie antenna for phased array systems, *Microwave Optical Technol. Lett.*, Vol. 43, No. 2, pp. 123–126, October 2004.
16. T. T. Wu and R. W. P. King, The cylindrical antenna with nonreflecting resistive loading, *IEEE Trans. Antennas Propag.*, Vol. 13, No. 3, 369–373, 1965.
17. K. L. Shlager, G. S. Smith, and J. G. Maloney, Optimization of bow-tie antennas for pulse radiation, *IEEE Trans. Antennas Propag.*, Vol. 42, pp. 975–982, July 1994.
18. T. Amert, J. Wolf, L. Albers, D. Palecek, S. T. Thompson, B. Askildsed, and K. W. Whites, Economical resistive tapering of bowtie antennas, in *IEEE Antennas and Propagation Society Conference*, Monterey, California, June 2004.
19. A. A. Lestari, A. G. Yarovoy, and L. P. Ligthart, RC-loaded bow-tie antenna for improved pulse radiation, *IEEE Trans. Antennas Propag.*, Vol. 52, No. 10, pp. 2555–2563, October 2004.

20. K. W. Loi, S. Uysal, and M. S. Leong, Design of a wideband microstrip bowtie patch antenna, *Proc IEE. Microwave Antennas Propag.*, Vol. 145, No. 2, pp. 137–140, April 1998.
21. S. Uysal, M. Leong, and C. H. Ng, Bowtie patch antennas and simple arrays for wireless indoor communications, *IEEE Trans. Microwave Theory Tech.*, Vol. 47, No. 6, pp. 738–745, June 1999.
22. K. L. Wong and J. S. Kuo, Bandwidth enhancement of bow-tie microstrip antenna using integrated reactive loading, *Microwave Optical Technol. Lett.*, Vol. 22, No. 1, pp. 69–71, July 1999.
23. J. Anguera, C. Puente, C. Borja, R. Montero, and J. Soler, Small and high-directivity bow-tie patch antenna based on the Sierpinski fractal, *Microwave Optical Technol. Lett.*, Vol. 31, No. 3, pp. 239–241, November 2001.
24. H.-R. Chuang, C.-C. Lin, and Y.-C. Kan, A printed UWB triangular monopole antenna, *Microwave J.*, Vol. 49, No. 1, pp. 108–120, January 2006.
25. Y. Kim and D.-H. Kwon, CPW-fed planar ultra wideband antenna having a frequency band notch function, *IEE Electron. Lett.*, Vol. 40, No. 7, pp. 403–405, April 2004.
26. S. W. Qu and C. L. Ruan, Effect of round corners on bowtie antennas, *Prog. Electromagnetics Res.*, Vol. 57, pp. 179–195, 2006.
27. A. A. Eldek, A. Z. Elsherbeni, and C. E. Smith, Characteristics of bow-tie slot antenna with tapered tuning stubs for wideband operation, *Prog. Electromagnetics Res.*, Vol. 49, pp. 53–69, 2004.
28. D. Ghosh, A. De, M. C. Taylor, T. K. Sarkar, M. C. Wicks, and E. L. Mokole, Transmission and reception by ultra-wideband (UWB) antennas, *IEEE Trans. Antennas Propag.*, Vol. 48, No. 5, pp. 67–99, October 2006.
29. J. D. Kraus, Helical beam antennas, *IEE Electron. Lett.*, Vol. 20, pp. 109–111, April 1947.
30. J. D. Kraus, *Antennas*, 2nd ed., McGraw-Hill, New York, 1988, Chap. 7.
31. E. F. Harris, *Antenna Engineering Handbook*, McGraw-Hill, New York, 1961, Chap. 7.
32. H. E. King and J. L. Wong, *Antenna Engineering Handbook*, 2nd ed., McGraw-Hill, New York, 1984, Chap. 13.
33. A. T. Adams, R. K. Greenough, R. K. Wallenberg, A. Mendelovicz, and C. Lumjiak, The quadrifilar helix antenna, *IEEE Trans. Antennas Propag.*, Vol. 22, pp. 173–178, March 1974.
34. C. C. Kilgus, Shaped conical radiation pattern performance of the backfire quadrifilar helix, *IEEE Trans. Antennas Propag.*, Vol. 23, pp. 392–397, May 1975.
35. W. T. Patton, The backfire bifilar helical antenna, Ph.D. Dissertation, University of Illinois, Urbana, October 1963.
36. J. Bach Andersen, *Handbook of Antenna Design*, Vol. 1, Peter Peregrinus, London, 1982, Chap. 7.
37. K. Noguchi, S. I. Betsudan, T. Katagi, and M. Mizusawa, A compact broad-band helical antenna with two-wire helix, *IEEE Trans. Antennas Propag.*, Vol. 51, No. 9, pp. 2176–2180, September 2003.
38. H. Nakano, N. Ikeda, Y. Y. Wu, R. Suzuki, H. Mimali, and J. Yamauchi, Realization of dual-frequency and wideband VSWR performance using normal-mode and inverted-F antennas, *IEEE Trans. Antennas Propag.*, Vol. 46, No. 6, pp. 788–793, June 1998.
39. Y. W. Chow, E. K. N. Yung, and H. T. Hui, Dual frequency monopole–helical antenna by using a parasitic normal mode helix for mobile handset, in *2000 Asia Pacific Microwave Conference*, Sydney, Australia, December 2000, pp. 183–187.
40. R. A. Abd-Alharmeed and P. S. Excell, Analysis of a normal-mode helical antenna including non-uniform wire surface current effects, *IEE Proc. Microwave Antennas Propag.*, Vol. 146, No. 1, pp. 1–5, February 1999.

41. B. A. Munk and L. Peters, A helical launcher for the helical antenna, *IEEE Trans. Antennas Propag.*, Vol. 16, pp. 362–363, May 1968.
42. H. Nakano, *Helical and Spiral Antennas—A Numerical Approach*, Research Studies Press, London, 1987, Chap. 6.
43. H. Nakano and J. Yamauchi, Radiation characteristics of helix antenna with parasitic elements, *Electron. Lett.*, Vol. 16, No. 18, 687–688, 1980.
44. H. Nakano and J. Yamauchi, Characteristics of modified spiral and helical antennas, *IEE Proc.*, Vol. 129, Pt. H, No. 5, pp. 232–237, 1982.
45. H. Nakano, N. Asaka, and J. Yamauchi, Radiation characteristics of short helical antenna and its mutual coupling, *Electron. Lett.*, Vol. 20, pp. 202–204, 1984.
46. H. E. King and J. L. Wong, Gain and pattern characteristics of 1 to 8 wavelength uniform helical antennas, in *IEEE International Symposium on Antennas and Propagation*, Washington DC, May 1978, pp. 69–72.
47. H. E. King and J. L. Wong, Characteristics of 1 to 8 wavelength uniform helical antennas, *IEEE Trans. Antennas Propag.*, Vol. 28, No. 2, pp. 291–296, March 1980.
48. H. M. Elkamchouchi and A. I. A. Salem, Effects of geometrical parameters loading and feeding on non-uniform helical antennas, *Nineteenth National Radio Science Conference*, Alexandria, Egypt, March 2002, pp. 90–100.
49. H. Nakano, J. Yamauchi, and H. Mimaki, Backfire radiation from a monofilar helix with a small ground plane, *IEEE Trans. Antennas Propag.*, Vol. 36, No. 10, pp. 1359–1364, October 1988.
50. V. Volman, Dual helix antenna for limited scan phased arrays, *IEEE Symp. Antennas Propag.*, Vol. 1, pp. 257–260, June 2003.
51. H. Nakano, Y. Samada, and J. Yamauchi, Axial mode helical antennas, *IEEE Trans. Antennas Propag.*, Vol. 34, No. 9, pp. 1143–1148, September 1986.
52. A. R. Djordjevic, A. G. Zajic, and M. M. Ilic, Enhancing the gain of helical antennas by shaping the ground conductor, *IEEE Antennas Propag. Lett.*, Vol. 5, pp. 138–140, 2006.
53. H. T. Hui, K. Y. Chan, and E. K. N. Yung, The low-profile hemispherical helical antenna with circular polarization radiation over a wide angular range, *IEEE Trans. Antennas Propag.*, Vol. 51, No. 6, pp. 1415–1418, June 2003.
54. Y. Zhang and H. T. Hui, A printed hemispherical helical antenna for GPS applications, *IEEE Microwave Optical Comp. Lett.*, Vol. 15, No. 1, pp. 10–12, January 2005.
55. A. Sharaiha and C. Terret, Analysis of quadrifilar resonant printed helical antenna for mobile communications, *Proc. IEE*, Vol. 140, No. 4, Pt. H, pp. 269–273, August 1993.
56. J. C. Louvigné and A. Sharaiha, Broadband tapered printed quadrifilar helical antenna, *Electron. Lett.*, Vol. 37, No. 15, pp. 932–933, July 2001.
57. Y. Letestu and A. Sharaiha, Broadband folded printed quadrifilar helical antenna, *IEEE Trans. Antennas Propag.*, Vol. 54, No. 5, pp. 1600–1604, May 2006.
58. S. Uda, Wireless beam of short electric waves, *J. IEE.* (*Japan*), pp. 273–282, March 1926, and pp. 1209–1219, November 1927.
59. H. Yagi, Beam transmission of ultra short waves, *Proc. IRE*, Vol. 26, pp. 715–741, June 1928; *Proc. IEEE*, Vol. 72, No. 5, pp. 634–645, May 1984.
60. A. Shoamanesh and L. Shafai, Properties of coaxial Yagi loop arrays, *IEEE Trans. Antennas Propag.*, Vol. 26, No. 4, pp. 547–550, July 1978.
61. A. Shoamanesh and L. Shafai, Design data for coaxial Yagi array of circular loops, *IEEE Trans. Antennas Propag.*, Vol. 27, No. 5, pp. 711–713, September 1979.
62. G. A. Thiele, Analysis of Yagi–Uda type antennas, *IEEE Trans. Antennas Propag.*, Vol. 17, pp. 24–31, January 1969.

63. C. A. Balanis, *Antenna Theory, Analysis and Design*, 3rd ed., John Wiley & Sons, Hoboken, NJ, 2005.

64. L. Shafai, Microstrip line fed microstrip end-fire antenna, U.S. Patent No. 5896108, April 1999.

65. D. Chang and C. A. Chen, Optimum element spacing for Yagi–Uda arrays, *IEEE Trans. Antennas Propag.*, Vol. 21, No. 5, pp. 615–623, September 1973.

66. E. A. Jones and W. T. Joines, Design of Yagi–Uda antennas using genetic algorithms, *IEEE Trans. Antennas Propag.*, Vol. 45, No. 9, pp. 1386–1392, September 1997.

67. N. V. Venkatarayalu and T. Ray, Optimum design of Yagi–Uda antennas using computational intelligence, *IEEE Trans. Antennas Propag.*, Vol. 52, No. 7, pp. 1811–1818, July 2004.

68. D. Kajfez, Nonlinear optimization extends the bandwidth of Yagi-antenna, *IEEE Trans. Antennas Propag.*, Vol. 23, No. 2, pp. 287–289, March 1975.

69. Y. Kuwahra, Multiobjective optimization of Yagi–Uda antenna, *IEEE Trans. Antennas Propag.*, Vol. 53, No. 6, pp. 1984–1992, June 2005.

70. D. Correia, A. J. M. Soares, and M. A. B. Terada, Optimization of gain, impedance bandwidth in Yagi–Uda antennas using genetic algorithm, in *Microwave and Optoelectronics Conference*, 1999. SBMO/IEEE MTT-S, APS and LEOS–IMOC '99, Rio de Janeiro, Brazil, Vol. 1, pp. 41–44, August 1999.

71. N. Kaneda, W. R. Deal, Y. Qian, R. Waterhouse, and T. Itho, A new quasi-Yagi antenna for planar active antenna arrays, *IEEE Trans. Antennas Propag.*, Vol. 50, No. 8, pp. 1158–1160, August 2002.

72. M. Modaresi and L. Shafai, A high-gain planar array using parallel Yagi–Uda antenna, *12th International Symposium on Antenna Technology and Applied Electromagnetics and URSI/CNC Conference*, ANTEM/URSI 2006, pp. 531–534, July 2006.

CHAPTER 10

Small and Fractal Antennas

STEVEN R. BEST

10.1 INTRODUCTION

With the continuing size reduction of today's ubiquitous wireless devices, there is an increasing demand for smaller antennas. Often, the demand for a smaller antenna does not come with a relaxation of the antenna's performance requirements. This can be a significant challenge for the antenna engineer since an antenna's performance properties are significantly affected by a reduction in size. Additionally, many of today's wireless devices are required to operate over multiple frequency bands leading to further design challenges for the antenna engineer.

When designing antennas that are small relative to the operating wavelength, it is important to have a good understanding of the fundamental concepts associated with how the size of an antenna relates to its electrical performance properties. In this chapter, the focus is on these fundamental concepts as they provide the foundation necessary to design small antennas for many of today's wireless applications. This chapter is not intended to serve as a review of specific antenna designs, like the widely used planar inverted-F antenna (PIFA), as these are described in the open literature.

The chapter begins with a review of important basic antenna properties, particularly bandwidth and quality factor, how they are related, and why they are particularly significant in the design of a small antenna. The chapter continues with a more detailed discussion on how to impedance match small antennas, describing techniques for achieving self-resonance, a 50-Ω impedance match, high radiation efficiency, and wide bandwidth (as allowable given the small antenna's occupied volume). In discussing these techniques, particular focus is given to how significant the antenna's physical properties (wire length, conductor diameter, height, overall size and volume, etc.) are in establishing the antenna's radiation performance. A number of small antenna design techniques that allow an antenna design to achieve a good impedance match and high radiation efficiency are described.

The chapter concludes with a discussion on fractal antennas, which were introduced to the antenna community in the early 1990s. Fractal antennas have been the subject of much discussion and some controversy. The focus in this chapter is not whether fractal antennas perform as claimed. Fractal antennas perform like all antennas, following the

Modern Antenna Handbook. Edited by Constantine A. Balanis

same fundamental physics. What we examine in this chapter are the aspects of the fractal antenna's electrical performance that can be ascribed to the uniqueness of the fractal geometry. We consider both small and multiband fractal antennas.

10.2 DEFINING ELECTRICALLY SMALL

When considering the performance properties and limitations of small antennas, it is important to distinguish between physically small and electrically small. At frequencies above approximately 2 or 3 GHz, where the operating wavelength, λ, is on the order of 15 cm or less, antennas that have dimensions on the order of $\lambda/2$ or $\lambda/4$ are not electrically small but they are often physically small. At frequencies above 2 or 3 GHz, where physically small antennas can more easily be used to meet performance requirements, the need for an electrically small antenna may not be a consideration. At frequencies below 2 or 3 GHz, where the operating wavelength ranges from 15 cm to several or hundreds of meters (or larger), the demand for an electrically small antenna increases since antennas that have dimensions on the order of $\lambda/2$ or $\lambda/4$ are often physically large and likely not suitable for use in small wireless devices or many other applications.

To distinguish between the physical and electrical size of the antenna, we define the antenna's electrical size in terms of its occupied volume relative to the operating wavelength. An electrically small antenna is one whose overall occupied volume is such that ka is less than or equal to 0.5, where k is the free-space wavenumber $2\pi/\lambda$, and a is the radius of a sphere circumscribing the maximum dimension of the antenna, as illustrated in Figure 10.1. A value of $ka = 0.5$ represents an overall spherical volume equal to $\lambda^3/48\pi^2$.

Other important factors to consider when defining the electrical size of an antenna are the presence of surrounding dielectric material and any ground plane structure. If dielectric material surrounds the antenna element, its dimensions must be included in the definition of the antenna's electrical size (the definition of a), particularly if it extends beyond the radiating, conductive structure of the antenna. If the antenna is located on a ground plane structure, the antenna image must be included in the definition of a. In the case of a very large ground plane, the definition of a encompasses the physical portion of the antenna above the ground plane. If the ground plane is small or if the ground plane is not symmetric with respect to the location of the antenna, the entire ground plane

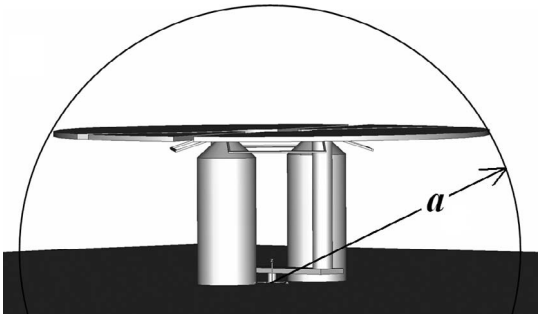


Figure 10.1 Depiction of a, the radius of an imaginary sphere circumscribing the maximum dimension of an antenna. In the case of an antenna mounted at the center of a large ground plane, the value of a circumscribes the "image" of the antenna within the ground plane.

structure or a significant portion of the ground plane structure must be included in the definition of a. Generally, if the impedance of the antenna on the finite ground plane is nearly the same as the impedance of the antenna located at the center of a very large (or infinite) ground plane, the ground plane dimension does not need to be included in the definition of a. If the impedance of the antenna is remarkably different from when the antenna is mounted at the center of a large ground plane, the dimensions of the ground plane do need to be included in the definition of a. Very often with small antennas on small ground planes, the currents on the ground plane structure are the primary source of radiation and are significant in determining the impedance, bandwidth, pattern, and polarization properties of the antenna.

The significance of defining the overall or occupied volume of the antenna in terms of ka lies in the fact that fundamental limits on the antenna's performance can be defined in terms of ka. For a given value of ka, there are certain performance limits that cannot be exceeded. These are discussed in the following sections.

10.3 FUNDAMENTAL PERFORMANCE PROPERTIES

Generally, the most important performance characteristics of interest when designing a small antenna are its impedance, gain, and bandwidth. In many wireless applications, pattern and polarization performance are important but may be of less concern since many wireless devices operate in environments with significant local scattering and numerous multipath fields. In these environments, pattern shape and polarization may be of secondary concern since the multipath fields arrive at the antenna from many different angles and with varying polarization sense and axial ratio. Additionally, the antenna engineer has little or no control as to how the user will hold or orient the wireless device during use.

The feed point impedance of any antenna is defined as

$$Z_A(\omega) = R_A(\omega) + jX_A(\omega) \tag{10.1}$$

where $R_A(\omega)$ is the antenna's frequency-dependent total resistance, comprised of a radiation resistance term, $R_r(\omega)$, and a loss resistance term, $R_l(\omega)$, and $X_A(\omega)$ is the antenna's frequency-dependent total reactance. ω is the radian frequency $2\pi f$, where f is the frequency in hertz(Hz).

The radiation resistance of the small antenna is primarily determined by its overall height or length relative to the operating wavelength. This fact will be important to remember when we discuss the significance of antenna geometry in establishing the performance property of an antenna. The loss resistance of the small antenna is determined by both conductor and dielectric losses. The antenna's total feed point reactance is primarily determined by the self-inductance and self-capacitance within the antenna structure. These concepts will be discussed in greater detail in a subsequent section.

The radiation efficiency of the antenna, $\eta_r(\omega)$, is determined from the ratio of the antenna's radiation resistance to its total resistance as follows:

$$\eta_r(\omega) = \frac{R_r(\omega)}{R_A(\omega)} = \frac{R_r(\omega)}{R_r(\omega) + R_l(\omega)} \tag{10.2}$$

The impedance of the small antenna is generally the first performance property to characterize since its value establishes the amount of power accepted by the antenna from

a transmitter (in the transmit mode) and the amount of power delivered to a receiver (in the receive mode). If the small antenna is 100% efficient and conjugately matched to the transmitter or receiver, it will accept or deliver maximum possible power, respectively. While no antenna is 100% efficient, the small antenna can be designed to exhibit very high efficiency ($\eta_r > 90\%$), even at very small values of ka.

The antenna's voltage standing wave ratio (VSWR) is used to quantify the level of impedance match between the antenna and the transmitter and the connecting transmission line. The frequency-dependent VSWR is given by

$$VSWR(\omega) = \frac{1 + |\Gamma(\omega)|}{1 - |\Gamma(\omega)|} \tag{10.3}$$

where $\Gamma(\omega)$ is the antenna's frequency-dependent reflection coefficient given by

$$\Gamma(\omega) = \frac{Z_A(\omega) - Z_{CH}}{Z_A(\omega) + Z_{CH}} \tag{10.4}$$

where Z_{CH} is the characteristic impedance of the transmission line connecting the antenna and the transmitter. In Eq. (10.4) Z_{CH} is assumed to be real and constant as a function of frequency.

The realized gain, G, of the small antenna is a function of its radiation efficiency, reflection coefficient, and directivity, D, and is given by

$$G(\omega) = \eta_r(\omega)(1 - |\Gamma(\omega)|^2)D(\omega) \tag{10.5}$$

where the term $(1 - |\Gamma(\omega)|^2)$ represents the mismatch loss between the antenna and transmitter when the VSWR is not exactly equal to 1. The first two terms of Eq. (10.5) are often combined to define the overall or realized efficiency of the antenna, $\eta_0(\omega) = \eta_r(\omega)(1 - |\Gamma(\omega)|^2)$. Note that this definition of gain assumes that the antenna is perfectly polarized, otherwise, Eq. (10.5) must include a polarization mismatch loss term. A small antenna expected to be vertically polarized will have less gain than given by Eq. (10.5) if it simultaneously radiates a substantial portion of horizontal polarization.

With decreasing values of ka, the directivity of the small antenna approaches a constant value of 1.5 or approximately 1.8 dB[†]. For small values of ka, the maximum gain that can be achieved with the small antenna is therefore approximately 1.8 dBi. This gain cannot be achieved in practice (assuming $D = 1.8$ dB) because it requires that the small antenna be 100% efficient and perfectly matched to the transmitter or receiver. Achieving high radiation efficiency is a function of the antenna configuration, dielectric losses, and the choice of conductor diameter and thickness. Techniques to achieve high radiation efficiency will be discussed in subsequent sections.

Any small antenna can be impedance matched at any single frequency. Impedance matching can be incorporated within the antenna structure or it can be accomplished with the use of external matching components. Impedance matching very often adds additional

[†]In this case, the small antenna is assumed to operate in a fundamental dipole mode with the usual omnidirectional, figure-eight radiation pattern of a dipole. For a small monopole antenna on an infinite ground plane, the directivity approaches 4.8 dB with decreasing values of ka. If the small antenna operates over a small ground plane, as is typical in many wireless devices, the maximum directivity will be somewhere between 1.8 and 4.8 dB. Typical values of directivity in this case may range between 2 and 4 dB.

loss, reducing the overall efficiency of the antenna. While there is no theoretical limit for the gain of the small antenna (other than its directivity), achieving and maintaining a gain close to 1.8 dBi can be a practical challenge at small values of ka.

More problematic at small values of ka is the practical effect of the antenna's narrow operating bandwidth. At small values of ka, the antenna typically exhibits high quality factor (Q) and a narrow operating bandwidth that may be on the order of a few percent or much less. While a good impedance match and reasonable efficiency can be achieved at small values of ka, the antenna's high Q and narrow operating bandwidth often make the practical realization of reasonable gain at a specific frequency a challenge. In practice, small antennas that exhibit high efficiency and very narrow bandwidth (high Q) are often susceptible to frequency detuning from nearby objects. Having an inherently narrow bandwidth, a slight change in operating frequency through detuning can result in a substantial increase in VSWR at the desired operating frequency, reducing the realized gain of the antenna.

While the small antenna's operating bandwidth is an important performance characteristic, we often quantify the bandwidth properties of the small antenna using its Q. This is done for a number of reasons. First, the definition of bandwidth can be ambiguous. Bandwidth may be defined in terms of half-power bandwidth (VSWR = 5.828), 2:1 VSWR bandwidth with respect to an arbitrary characteristic impedance, −10-dB return loss bandwidth, and so on. The definition of Q is not ambiguous. It is defined exactly in terms of the ratio of reactive energy to accepted power. Second, an approximate inverse relationship between VSWR bandwidth and Q can be established. Third, a fundamental limit on antenna Q can be defined as a function of ka, leading to a fundamental limit on VSWR bandwidth that can be defined as a function of ka.

Both Q and bandwidth are defined for the tuned antenna. The tuned antenna is one whose feed point reactance is tuned to zero at a frequency ω_0 using a single, lossless, series tuning element having a reactance $X_S(\omega_0) = -X_A(\omega_0)$. The tuned antenna's impedance is defined as

$$Z_0(\omega) = R_0(\omega) + jX_0(\omega) = R_A(\omega) + j[X_A(\omega) + X_S(\omega)] \tag{10.6}$$

The Q of the tuned antenna is defined in terms of the ratio of reactive energy, W, and accepted power, P_A, as [1, 2]

$$Q(\omega_0) = \frac{\omega_0 W(\omega_0)}{P_A(\omega_0)} \tag{10.7}$$

Note that the accepted power, P_A, includes both power associated with radiation and loss within the antenna structure. In defining exact Q, the challenge is to define an exact definition for the reactive energy W, such that Q and bandwidth are inversely related over all ranges of frequency. This was done in Ref. 2, where it was shown that the exact Q can be approximated using

$$\begin{aligned} Q(\omega_0) \approx Q_Z(\omega_0) &= \frac{\omega_0}{2R_A(\omega_0)}|Z_0'(\omega_0)| \\ &= \frac{\omega_0}{2R_A(\omega_0)}\sqrt{R_A'(\omega_0)^2 + \left(X_A'(\omega_0) + \frac{|X_A(\omega_0)|}{\omega_0}\right)^2} \end{aligned} \tag{10.8}$$

Having defined Q, it is necessary to define bandwidth such that Q and bandwidth are inversely related over all ranges of frequency. The definition of bandwidth suitable for this purpose is fractional matched VSWR bandwidth, $FBW_V(\omega_0)$, where the VSWR of the tuned antenna is determined using a characteristic impedance, Z_{CH}, equal to the antenna's feed point resistance, $R_A(\omega_0)$. Fractional matched VSWR bandwidth is given by

$$FBW_V(\omega_0) = \frac{\omega_+ - \omega_-}{\omega_0} \tag{10.9}$$

where ω_+ and ω_- are the frequencies above and below ω_0, respectively, where the VSWR is equal to any arbitrary value denoted by s. Fractional matched VSWR bandwidth and Q are related as [2]

$$Q(\omega_0) = \frac{2\sqrt{\beta}}{FBW_V(\omega_0)}, \quad \sqrt{\beta} = \frac{s-1}{2\sqrt{s}} \leq 1 \tag{10.10}$$

The approximations in Eqs. (10.8) and (10.10) were derived in Ref. 2 under the assumptions that the tuned or self-resonant antenna exhibits a single impedance resonance within its defined operating bandwidth and that the half-power matched VSWR bandwidth is not too large. If the small antenna is designed so that it exhibits multiple, closely spaced resonances within its defined VSWR bandwidth, these approximations may not hold [2–4].

The final performance characteristic discussed in this section is the realized effective area of the small receive antenna, which, like realized gain, is a function of the antenna's radiation efficiency, mismatch loss, directivity, and polarization mismatch loss. In defining the antenna's realized gain, it was assumed that the characteristic impedance of the feed transmission line is real and constant versus frequency, which is the case in most practical applications. In the receive mode, the small antenna may be connected to a complex load impedance, in which case the definition for mismatch loss, $(1 - |\Gamma|^2)$, no longer holds.

In the receive mode, the antenna delivers maximum power to a load that is the conjugate of the antenna's feed point impedance. If this load impedance is complex, the mismatch loss is given by [5]

$$ML = \frac{(1 - |\Gamma|^2)(1 - |\Gamma_L|^2)}{|1 - \Gamma\Gamma_L|^2} \tag{10.11}$$

and the realized effective area of the small receive antenna is given by

$$A_{er}(\omega) = \eta_r(\omega) ML(\omega) \frac{\lambda^2 D}{4\pi} \tag{10.12}$$

where Γ is the usual antenna reflection coefficient determined using Eq. (10.4), Γ_L is the usual reflection coefficient determined for the impedance seen looking from the antenna's feed point terminals toward the receiver (the load impedance connected at the antenna's terminal), and Z_{CH} used to determine both Γ and Γ_L can be any arbitrary real impedance. Note that Eq. (10.12) assumes the antenna is polarization matched to the incident electromagnetic wave.

10.4 FUNDAMENTAL PERFORMANCE LIMITATIONS

There is no theoretical limit on the achievable gain of the electrically small antenna other than the fact that its gain cannot exceed its directivity[†]. Practical limits on realized gain are associated with the challenges of (1) impedance matching the small antenna and minimizing losses within any matching network; (2) achieving maximum possible radiation efficiency through appropriate choice of conductor diameter or thickness and minimization of dielectric losses; (3) minimizing, if possible, the effects of detuning when the antenna is placed in its operating environment; and (4) ensuring that the antenna radiates all of the available power into the intended polarization.

There is a theoretical limit on the minimum Q that can be achieved for the electrically small antenna as a function of its size relative to the operating wavelength (ka). Since Q and bandwidth are inversely related, it follows that there is a corresponding upper limit on the maximum bandwidth that can be achieved with electrically small antennas.

Fundamental limits on the performance of small antennas were first addressed by Wheeler in 1947 [6]. Wheeler preferred to characterize fundamental limits on the performance of the small antenna in terms of its radiation power factor [7, 8], which he defined as the inverse of Q. The maximum achievable power factor for the small antenna is defined $(ka)^3$, which leads to a minimum achievable Q, or lower bound of

$$Q_{\text{lb-Wheeler}} = \eta_r \frac{1}{(ka)^3} \tag{10.13}$$

Subsequent work by Chu [1] and McLean [9] has led to the commonly accepted lower bound on Q (the Chu limit) being defined as

$$Q_{\text{lb-Chu}} = \eta_r \left(\frac{1}{(ka)^3} + \frac{1}{ka} \right) \tag{10.14}$$

In Wheeler's and Chu's work, the lower bound on Q is stated for the lossless antenna. The effect of loss on the lower bound is accounted for by including the radiation efficiency term in Eqs. (10.13) and (10.14). Inclusion of the radiation efficiency term illustrates that lossy antennas can achieve a lower Q, and hence increased bandwidth, relative to their lossless counterparts. At small values of ka, the difference between the Chu and Wheeler limits is not significant and likely within any error in the approximate inverse relationship between Q and bandwidth.

Using Eqs. (10.10) and (10.14), an upper bound on the fractional matched VSWR bandwidth can be written

$$FBW_{Vub} = \frac{1}{\eta_r} \frac{(ka)^3}{1 + (ka)^2} \frac{s - 1}{\sqrt{s}} \tag{10.15}$$

A plot of the upper bound on fractional matched VSWR bandwidth ($s = 2$ and $s = 5.828$) versus ka for the lossless antenna is presented in Figure 10.2. At $ka = 0.5$,

[†]It is assumed here that for values of $ka \leq 0.5$, the directivity of the electrically small antenna is limited to the directivity associated with the fundamental dipole or monopole modes, having values of 1.8 and 4.8 dB, respectively.

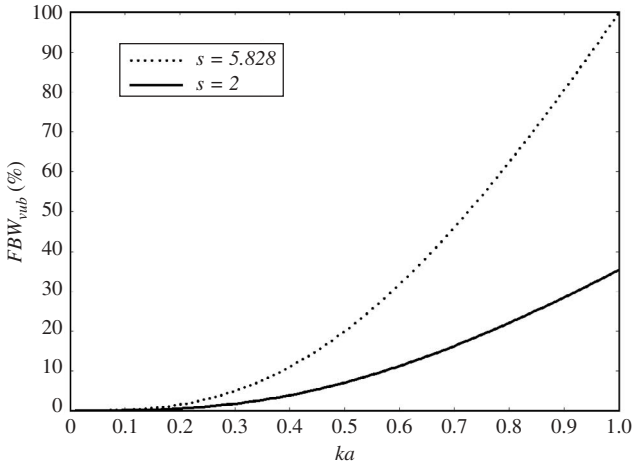


Figure 10.2 The upper bound on fractional matched VSWR bandwidth for the lossless antenna as a function of ka.

the upper bound on fractional matched VSWR bandwidth is approximately 20% and 7.1%, for the half-power and 2:1 VSWR bandwidths, respectively. At $ka = 0.1$, these upper bounds decrease to approximately 0.2% and 0.071%, respectively, illustrating the significant bandwidth limitations that arise as a function of decreasing the antenna size.

In general, the electrically small antenna can be impedance matched and it can be made reasonably efficient. In this case, optimizing the operating bandwidth is perhaps the most significant performance issue. The upper bound on bandwidth presented in Eq. (10.15) is defined for the small antenna exhibiting a single resonance within its operating bandwidth and a perfect impedance match as illustrated in Figure 10.3a. In many cases, a marginal increase in bandwidth can be achieved by purposely mismatching the antenna as illustrated in Figure 10.3b. The half-power and 2:1 VSWR bandwidths for the matched and mismatched antennas are 34.9% and 11.5% and 42.7% and 12.4%, respectively. With the purposely mismatched antenna, the upper bound on bandwidth presented in Eq. (10.15) is only as valid as the relationship between Q and the "mismatched" VSWR bandwidth.

A further increase in bandwidth may be achieved if the small antenna is designed to exhibit multiple resonances within its operating bandwidth as illustrated in Figure 10.4. The impedance curve shown in Figure 10.4 also raises an issue with the approximate relationships between impedance, bandwidth, and Q stated in Eqs. (10.8) and (10.10). In Eqs. (10.8) and (10.10), the frequency derivatives of both resistance and reactance can be used to approximate both Q and bandwidth. The accuracy of these approximations requires the local impedance derivatives at the resonant or tuned frequency to reliably predict the impedance behavior out to the extents of the operating bandwidth. The approximations of Eqs. (10.8) and (10.10) are very accurate for antennas exhibiting a single impedance resonance as illustrated in Figure 10.3. The approximations of Eqs. (10.8) and (10.10) may no longer hold [3, 4] for antennas exhibiting multiple impedance resonances as illustrated in Figure 10.4. If the inverse relationship between bandwidth and Q does not hold for an antenna exhibiting multiple impedance resonances, the question to consider is whether or not the bandwidth of these antennas is limited by the inverse of the lower bound on Q. This is a significant area of current research.

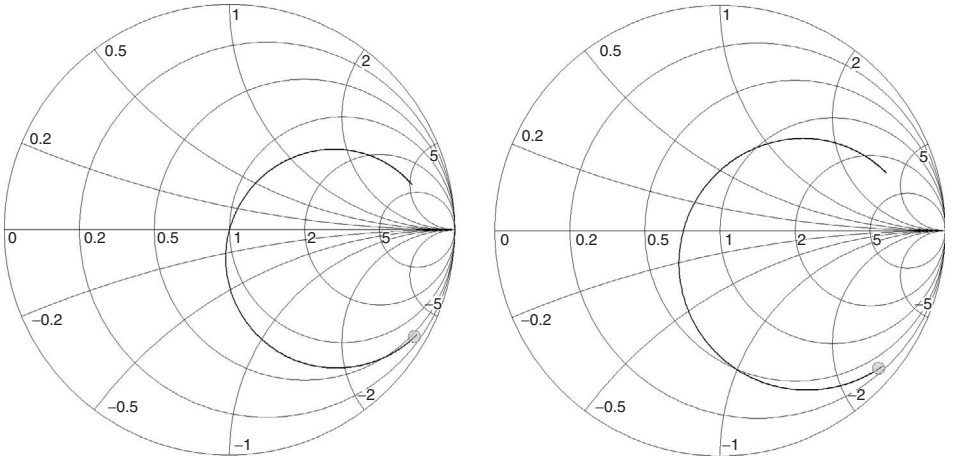


Figure 10.3 Impedance of the (a) matched antenna and (b) mismatched antenna.

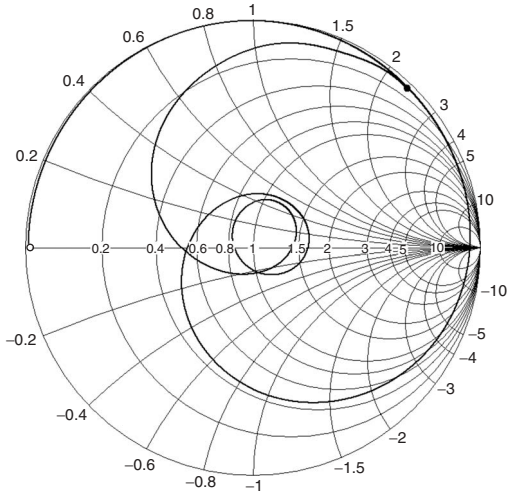


Figure 10.4 Impedance of an antenna exhibiting multiple resonances within its defined VSWR bandwidth. The inverse relationship between bandwidth and Q may not hold for antennas that exhibit this behavior. Additionally, the approximations in Eqs. (10.8) and (10.10), relating Q, bandwidth, and impedance, may not hold.

10.5 SMALL DIPOLE AND LOOP ANTENNAS

Prior to discussing the performance properties of different small antenna designs and associated optimization techniques, it is important to briefly review the performance properties of the small dipole and loop antennas, which are depicted in Figure 10.5. The dipole antenna has an overall length l and a conductor diameter d. The loop antenna has a circumference C and a conductor diameter d.

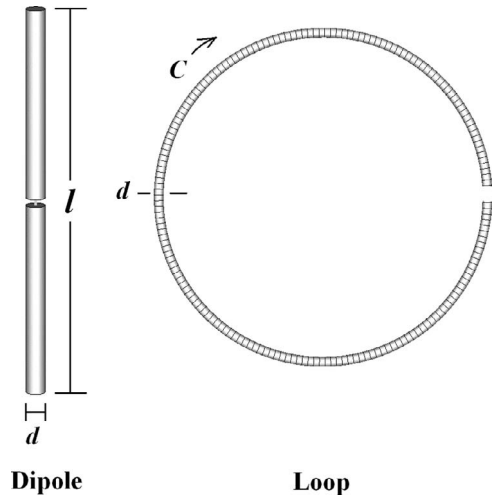


Figure 10.5 Depictions of the straight-wire dipole and circular loop antennas. The straight-wire dipole has an overall length l and a conductor diameter d. The circular loop has a circumference C and a conductor diameter d.

The impedance of a typical dipole antenna is depicted in Smith chart format in Figure 10.6a. The dipole exhibits its first natural resonance, $X_A(\omega_r) = 0$, at a radian frequency $\omega_r \approx c\pi/l$, where c is the speed of light in a vacuum. This first natural resonance is a series resonance, where the frequency derivative of the feed point reactance at the resonant frequency, $X'(\omega_r)$ is greater than zero. The next natural resonant frequency is a parallel or antiresonance, where the frequency derivative of the feed point reactance is less than zero. The antiresonant frequency occurs at a value of $\omega_r \approx 2c\pi/l$. Note that the antiresonant frequency of the dipole is very sensitive to changes in the antenna's conductor diameter, d, and the exact value of ω_r may change considerably as the conductor diameter varies.

Below its first natural resonant frequency, the dipole's feed point reactance is negative (capacitive) and approaches $-\infty$ as ω approaches 0. Behaving as a lossy capacitor, or equivalently as a series RLC circuit, the dipole's reactance takes the form of

$$X(\omega) = \omega L - \frac{1}{\omega C} \tag{10.16}$$

where the total inductance at the dipole's feed point, L, is primarily a function of the self-inductance of the conductor, and the total capacitance at the dipole's feed point, C, is primarily a function of the capacitance between the dipole's upper and lower (relative to the feed point) conductor sections. In the case of a monopole antenna, this total capacitance is primarily a function of the capacitance between the monopole's conductor and the ground plane. At very low frequencies, the $1/\omega$C term dominates the total reactance of the dipole.

The radiation resistance of the dipole approaches 0 as ω approaches 0. At frequencies near and below the small antenna limit ($ka \leq 0.5$), the radiation resistance of the dipole

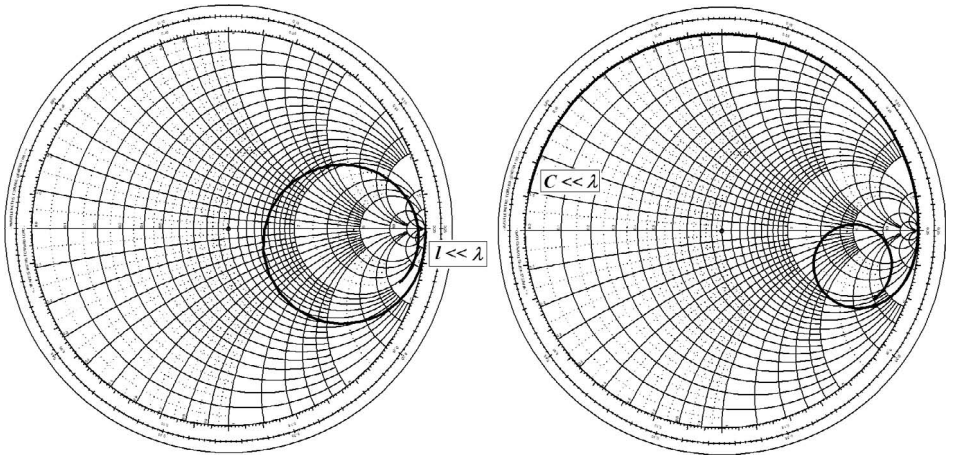


Figure 10.6 (a) Impedance of a typical dipole antenna over a frequency range from where the dipole is very small ($l \ll \lambda$) through the dipole's first resonance and antiresonance. (b) Impedance of a typical loop antenna over a frequency range where the loop is very small ($C \ll \lambda$) through the loop's first antiresonance, first resonance, and second antiresonance.

is approximately given by [10]

$$R_r^d = 20\pi^2 \left(\frac{l}{\lambda}\right)^2 \tag{10.17}$$

The decreasing value of radiation resistance and the increasing value of capacitive reactance result in very high values of VSWR at low frequencies, making it impractical to deliver power to the dipole using a practical transmission line.

To determine the radiation efficiency of the small dipole, its loss resistance must be defined. If the skin depth δ in the conductor is somewhat less than $d/2$, the loss resistance of the dipole can be approximated as [11]

$$R_l^d = \frac{l\rho}{2\pi d\delta} \approx \frac{l}{2\pi d}\sqrt{\frac{kc\mu_0\rho}{2}} \tag{10.18}$$

where ρ is the resistivity of the conductor and μ_0 is the permeability of free space. For δ somewhat greater than $d/2$, the loss resistance of the small dipole can be approximated by $R_l^d = 2l\rho/\pi d^2$. To obtain the approximation of loss resistance in Eq. (10.18), it is assumed that the current on the dipole has a cosine dependence over the length of the conductor and the current density decays exponentially from its value at the surface of the conductor. For reasonable conductor diameters, the radiation efficiency of the small dipole remains in excess of 80% or 90% for values of ka well below 0.5.

A loop antenna, having a circumference C, exhibits its first natural resonance at an approximate radian frequency $\omega_r \approx c\pi/C$. Unlike the dipole, the loop's first natural resonance is a parallel resonance or antiresonance, where its resistance can be several thousand ohms and it essentially behaves as an open circuit. The impedance of a typical loop antenna is depicted in Figure 10.6b.

The first natural resonant frequency of the dipole antenna occurs at a value of $ka \approx 1.57$. The first natural antiresonant frequency of the loop antenna occurs at a value of $ka \approx 0.5$, where the loop may be considered electrically small. Generally, the electrically small loop is considered to behave as a magnetic dipole exhibiting a constant current over its circumference and a typical dipole-like radiation pattern. However, at $ka = 0.5$, the loop has a very high resistance, and a nonconstant current and it does not exhibit a dipole-like radiation pattern. The loop antenna does not behave as a magnetic dipole until the value of ka approaches 0.1.

Below its first natural antiresonant frequency, the loop's feed point reactance is always positive (inductive) and approaches 0 as ω approaches 0. Near and below antiresonance, the loop behaves as a parallel RLC circuit and its reactance takes the form of

$$X(\omega) = \text{Im}\left(\frac{1}{\frac{1}{R} + j\left(\frac{1}{\omega L} + \omega C\right)}\right) \tag{10.19}$$

Well below antiresonance ($ka \ll 0.5$), where the current distribution is near constant and the self-capacitance of the loop diminishes to zero, the loop essentially behaves as a lossy inductor and the reactance takes the form of

$$X(\omega) = j\omega L \tag{10.20}$$

The radiation resistance of the loop also approaches 0 as ω approaches 0. At frequencies well below the small antenna limit, the radiation resistance of the loop is approximately given by [10]

$$R_r^l = 20\pi^2 N^2 \left(\frac{C}{\lambda}\right)^4 \tag{10.21}$$

where N is the number of total turns in the loop.

To calculate the radiation efficiency of the small loop, its loss resistance must be determined. If the skin depth δ in the conductor is somewhat less than $d/2$, the loss resistance of the circular loop can be approximated as [11]

$$R_l^l = \frac{NC\rho}{\pi d\delta} \approx \frac{NC}{\pi d}\sqrt{\frac{kc\mu_0\rho}{2}} \tag{10.22}$$

For δ somewhat greater than $d/2$, the loss resistance of the small loop can be approximated by $R_l^l = 4NC\rho/\pi d^2$. It is assumed that the current is constant around the loop and its density decays exponentially from its value at the surface of the conductor.

The radiation efficiency of the loop exhibits substantially different behavior than that of the dipole as a function of frequency for a given conductor length. The difference is due to the fact that the radiation resistance of the dipole exhibits an ω^2 dependence, whereas the radiation resistance of the loop exhibits an ω^4 dependence. At very small values of ka, the loop is substantially less efficient than a dipole having the same conductor diameter and occupying the same overall volume. The radiation efficiency of the loop can be increased by increasing the number of turns in the loop. As the number of turns increases, the loop's radiation resistance increases as N^2 while its loss resistance increases as N.

10.6 ACHIEVING SELF-RESONANCE

As previously mentioned, one of the most significant performance characteristics of the small antenna to optimize is its impedance match to the transmitter or receiver. With small dipole and loop antennas, it is evident that impedance matching is an important concern since these antennas exhibit very small resistances and, in the case of the small dipole, very high values of capacitive reactance. To impedance match the small dipole and loop, their reactances must be tuned to zero and their low resistances must be transformed to 50 Ω or another desired matching impedance. Any electrically small antenna can be impedance matched at any single frequency using external reactive components. The performance limiting issues to consider with impedance matching are the following: (1) How large is the operating bandwidth about the matched frequency? and (2) How much additional loss is introduced by the impedance matching components? In some instances, the improvement in mismatch loss may be more substantial than the added component loss, improving the overall or realized efficiency, but this is not always the case.

Tuning the high capacitive reactance of the small dipole is generally more of a practical challenge than tuning the low inductive reactance of the small loop. The inductive reactance of a small loop is very often tuned to zero using low loss, variable capacitors. Tuning the capacitive reactance of the small dipole to zero typically requires a large inductance, which is very often an inefficient process. For these reasons, this section focuses on achieving self-resonance with small dipole or monopole antennas. Here, we specifically consider achieving self-resonance and the impedance match within the antenna structure rather than through the use of external reactive components. To facilitate this discussion, we consider monopole elements operating over an infinite ground plane rather than free-space dipole elements. The general discussion presented here applies to virtually all small antenna elements.

The electrically small monopole antenna is assumed to have an overall height $h \ll \lambda$. It exhibits a feed point resistance [10] equal to

$$R_r^m = 40\pi^2 \left(\frac{h}{\lambda}\right)^2 \tag{10.23}$$

and a feed point reactance of the form $X(\omega) = \omega L - 1/\omega C$, where the $1/\omega C$ term dominates at low frequencies, resulting in the monopole's high capacitive reactance. To achieve self-resonance at low frequencies, the equivalent feed point inductance, L, and/or the equivalent feed point capacitance, C, can be increased to a point where $\omega L = 1/\omega C$, resulting in a total feed point reactance equal to zero.

In the case of fixed height wire monopole antennas, the total self-inductance can be increased simply by increasing the total wire length. The total capacitance can be increased using a number of techniques. One of the most effective techniques is to add a capacitive top hat to the structure, increasing the capacitance to ground. This technique is also sometimes referred to as "capacity loading" or "disk loading" [12]. Other techniques to increase the capacitance to ground include increasing the cylindrical diameter of the antenna structure and loading the antenna with a dielectric material. In many instances, this tuning method affects both the self-inductance and the total capacitance of the structure.

10.6.1 Increasing Self-Inductance

The resonant frequency of a small wire antenna is determined by the antenna's total wire length, the wire diameter, the spacing between parallel and orthogonal sections of the wire (self-coupling), and the current vector alignment along the length of the wire [13, 14]. Here, we consider tuning the small antenna (achieving self-resonance) by increasing the total wire length within the antenna structure. We are particularly interested in the significance of antenna geometry in establishing the resonant performance properties of the small antenna. We consider small wire antennas that have the same overall height and wire diameter. The radiation properties of the antennas considered here are modeled using the NEC4 engine of EZNEC Pro [15] or CST's Microwave Studio [16]. Copper conductor loss is included in all of the numerical simulations.

We begin with an arbitrary straight-wire monopole antenna having an overall height of 10 cm and a wire diameter of 1 mm. This monopole exhibits a resonant frequency of 708.3 MHz with a resonant resistance of approximately 36 Ω. At frequencies below 708.3 MHz, the straight-wire monopole exhibits a capacitive (negative) reactance and a resistance less than 36 Ω. At any frequency below 708.3 MHz, we ultimately seek to impedance match the monopole to some desired characteristic impedance.

Consider an arbitrary frequency of 150 MHz, where $ka \approx 0.314$. At this frequency, the monopole's impedance is approximately $1 - j794$ Ω and its radiation efficiency is 96.7%. The monopole's capacitive reactance can be tuned to resonance ($X = 0$) by simply adding wire length while maintaining the same overall height. The increase in total wire length can be accomplished using any geometric configuration. However, since the objective of adding wire length is to increase the self-inductance of the conductor forming the antenna, we seek to use a geometry that adds the most self-inductance for any given wire length.

If tasked with winding an inductor using a fixed length of wire, one would likely choose to use a helical coil structure, the basis of the normal mode helix antenna [17]. Using a normal mode helix to achieve resonance at 150 MHz within a 10-cm overall height requires a total wire length of approximately 86.11 cm (for a 1-mm conductor diameter). The normal mode helix antenna is depicted in Figure 10.7a. Its impedance at 150 MHz is approximately $3.1 + j0.4$ Ω and its radiation efficiency is approximately 81.5%. A larger conductor diameter can be used to increase the radiation efficiency. However, increasing the conductor diameter of the small antenna generally increases the resonant frequency and more wire length must be added to the antenna to maintain resonance at the same frequency.

Other geometries can also be used to achieve resonance at the same frequency within the same overall height. In general, geometries other than the normal mode helix require more total wire length to achieve resonance at the same frequency. For example, a meander line monopole, depicted in Figure 10.7b, can also be made resonant at 150 MHz. In this case, the total wire length is approximately 1.56 m, nearly twice that of the normal mode helix. The meander line monopole's impedance and radiation efficiency are $3.0 + j0.9$ Ω and 64%, respectively. While the impedances of the meander line and normal mode helix antennas are nearly the same, their radiation efficiencies differ due to their respective difference in total wire lengths.

The more significant performance characteristic of each antenna to consider is their operating bandwidth, which is quantified here using the value of Q. The Q values of the normal mode helix and the meander line antennas are 162.2 and 115.5, respectively. The difference in Q can be attributed to the fact that the meander line antenna has a much

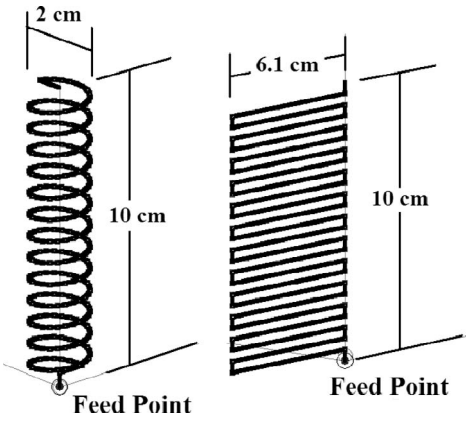


Figure 10.7 (a) Depiction of the normal mode helix monopole antenna having an overall height of 10 cm and an overall diameter of 2 cm. (b) Depiction of the meander line monopole antenna having an overall height of 10 cm and an overall diameter (meander line width) of 6.1 cm. Both antennas have conductor diameters of 1 mm.

wider cylindrical diameter than the normal mode helix. The normal mode helix has a diameter of 2 cm whereas the meander line antenna has a diameter of 6.1 cm. Generally, small antennas of this type with wider diameters exhibit lower Q values.

The relative behavior of these antennas can also be compared over a wider range of resonant frequencies, where the total wire length is adjusted to achieve resonance at any desired frequency. The practical limitation of achieving resonance at much lower frequencies is a function of how much wire (of a specific conductor diameter) can fit within the defined overall height. Here, we remove wire length in each antenna to achieve resonance at several frequencies between 150 and 708 MHz. With the normal mode helix, the total wire length is reduced by decreasing the number of turns. The overall diameter of the helical structure is maintained at 2 cm. With the meander line antenna, the total wire length is reduced by adjusting either the overall diameter (meander section length) or the number of meander sections.

A comparison of the two antenna's resonant performance properties is presented in Table 10.1. In all cases, larger values of resistance and lower values of Q are achieved with increasing values of resonant frequency, since the value of radiation resistance varies in proportion to the value of $(h/\lambda)^2$ and the value of Q varies in proportion to the value of $(\lambda/h)^2$ [13, 18, 19]. Generally, as the height of the small resonant antenna decreases relative to the resonant wavelength, the value of the resonant radiation resistance decreases and the value of resonant Q increases.

Another point to note is the relative Q of each antenna as a function of frequency. At the lower frequencies, where the cylindrical diameter of the meander line antenna is larger than that of the normal mode helix, the meander line antenna exhibits a lower Q. At the highest frequencies, where the overall diameter of the meander line antenna is adjusted to be closer to or the same as that of the normal mode helix, the normal mode helix exhibits a lower Q.

At the lower frequencies, the Q of the normal mode helix can be reduced by increasing the cylindrical diameter of the antenna structure. As the overall diameter of the

antenna is increased, while maintaining the same overall height, the total wire length must be reduced to maintain the same resonant frequency. If the overall diameter of the normal mode helix is increased to 6.1 cm, matching that of the meander line antenna, the total wire length necessary to achieve resonance at 150 MHz decreases from 86.1 to 65.1 cm. This configuration is depicted in Figure 10.8. At 150 MHz, it exhibits a resistance of 2.1 Ω and a Q of 126.9, more closely approaching that of the meander line antenna.

10.6.2 Significance of Wire Geometry in Establishing Resonant Performance

From the results presented in Table 10.1, it is evident that the normal mode helix antenna requires less total wire length than the meander line antenna to achieve the same resonant frequency. This is a result of the fact that the helical coil shape has a higher inductance than the meander line shape. Given its lesser total wire length, the normal mode helix exhibits higher radiation efficiency and can be considered a better choice of geometry.

TABLE 10.1 Performance Comparison of the Normal Mode Helix and the Meander Line Antennas as a Function of Different Resonant Frequencies[a]

	Normal Mode Helix				Meander Line Antenna			
Resonant Frequency (MHz)	Total Wire Length (cm)	Resistance (Ω)	Q	η_r (%)	Total Wire Length (cm)	Resistance (Ω)	Q	η_r (%)
150	86.1	3.1	162.2	81.5	156	3.0	115.5	64
178.25	70	3.6	121	86.4	125.2	3.5	92	74.2
255.8	43.9	6.5	50	94.4	77.2	5.8	50	88.9
313	33.3	9.1	31.1	96.7	58	8	30.7	93.5
419	21.1	15.3	16.3	98.6	26	12.9	19	98
578	13.7	28.2	8.5	99.4	14	23.2	11.2	98.7

[a] All of the antennas have an overall height of 10 cm and a conductor diameter of 1 mm.

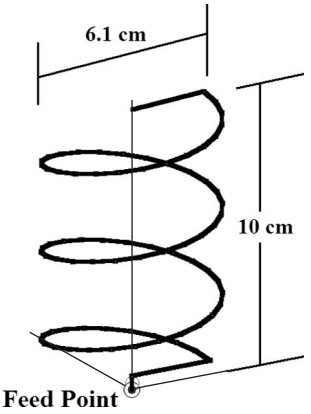


Figure 10.8 Depiction of the normal mode helix monopole antenna having the same overall diameter and resonant frequency as the meander line monopole antenna depicted in Figure 10.6b.

At the lower frequencies, the narrower diameter normal mode helix exhibited higher Q than the meander line antenna and would therefore exhibit lower bandwidth. However, if the diameter of the normal mode helix is increased, its Q can be lowered significantly. At the higher frequencies, where the diameter of the meander line antenna is reduced, the normal mode helix exhibits a slightly lower Q. At all frequencies, the resonant resistances of the two antennas are not significantly different.

In this section, we further examine the significance that antenna geometry plays in establishing the antenna's resonant properties. For a given total wire length, we see that the normal mode helix exhibited a lower resonant frequency. To achieve resonance at a given frequency, the normal mode helix shape required the least amount of wire. Once resonance is established with either geometry, where both have nearly the same overall diameter, the radiation properties were not significantly different, except perhaps for radiation efficiency. If this is universally true, then the overall height and cylindrical diameter of the antenna are the critical physical characteristics that establish the antenna's performance properties.

10.6.2.1 Aspects of Antenna Geometry That Affect Resonant Frequency

The first point discussed here is: Why does the normal mode helix exhibit a lower resonant frequency for the same total wire length? Or, perhaps more important, how can one look at an antenna's geometry and determine if it is well suited for achieving resonance given a fixed wire length and overall volume? To understand the answers to these questions, we begin by looking at the current distribution on the meander line antenna depicted in Figure 10.9.

The meander line antenna shown has an overall height of 10 cm, a total wire length of 30 cm, a meander line section length (antenna diameter) of 3.33 cm, and a conductor diameter of 1 mm. The antenna is resonant at approximately 361 MHz. The arrows shown in Figure 10.9 depict the direction of current along the antenna from the feed point to the end of the wire: the current vector [14]. With small antennas of this type, the current maximum is at the feed point and it diminishes to zero at the end of the wire. Because

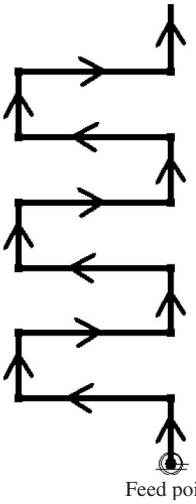


Figure 10.9 Depiction of the current vector alignment along the length of a meander line antenna having a total height of 10 cm and a wire length of 30 cm.

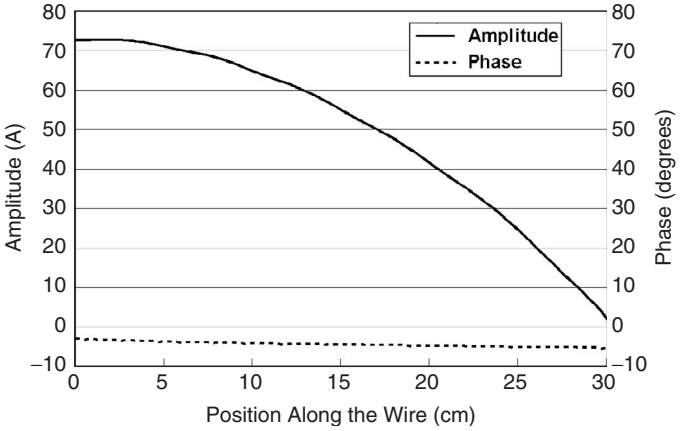


Figure 10.10 Current distribution along the wire of the meander line antenna depicted in Figure 10.9.

the antenna is small relative to the wavelength, the current is nearly in phase along the entire length of wire. The current distribution along the meander line antenna length is presented in Figure 10.10.

Since the current along the total wire length is nearly in phase, the geometry of the wire is significant in establishing the antenna's resonant frequency. The aspects of the antenna geometry that are significant in establishing the resonant frequency are the spacing between parallel sections of wire and the current vector alignment. The spacing and length of horizontal wire above the ground also has a significant impact on establishing the resonant frequency.

In regions where the current vectors oppose in parallel sections of wire, as illustrated in Figure 10.11a, the effective or net self-inductance of the wire is diminished. In regions where the current vectors reinforce in parallel sections of wire, as illustrated with the normal mode helix in Figure 10.11b, the effective or net self-inductance of the wire is increased. In antennas where the geometric layout of wire is such that the current vectors reinforce along the total wire length, rather than oppose, there is more self-inductance for a given wire length and these antennas will exhibit a lower resonant frequency. The meander line antenna depicted in Figure 10.11a exhibits a resonant frequency of 361 MHz, whereas the normal mode helix antenna depicted in Figure 10.11b exhibits a resonant frequency of 312.6 MHz. Both antennas have the same overall height (10 cm), the same total wire length (30 cm), the same conductor diameter (1 mm), and the same overall cylindrical diameter (3.3 cm).

The reinforcing aspects of the current vector alignment along the normal mode helix antenna may not be clearly evident since there is large spacing between the turns of the helical coil structure. This aids in illustrating the significance of the spacing (coupling) between both opposing and reinforcing current vectors. Decoupling sections of opposing current vectors (increasing the spacing) or increasing the coupling between reinforcing current vectors (decreasing the spacing) aids in lowering the antenna's resonant frequency [14].

The meander line antenna of Figure 10.12a illustrates one example of decoupling opposing current vectors, resulting in a decrease in resonant frequency from 361 to

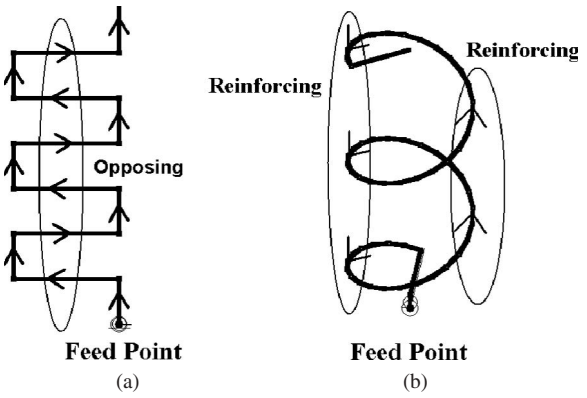


Figure 10.11 Depictions of opposing and reinforcing current vectors for (a) the meander line antenna and (b) the normal mode helix antenna.

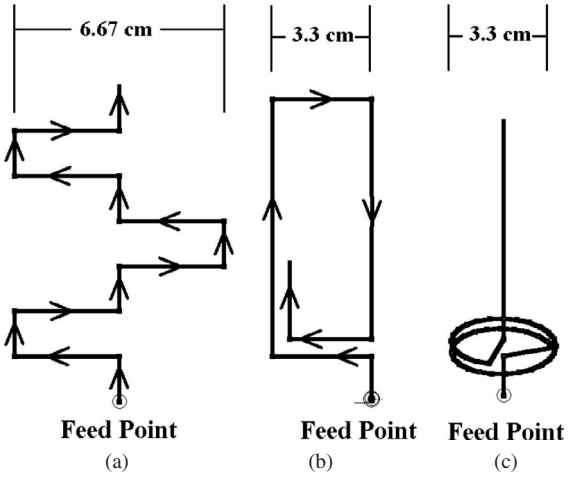


Figure 10.12 (a) Modified meander line antenna with decreased coupling between meander sections exhibiting a resonant frequency of 343.5 MHz. (b) Modified meander line antenna with decreased coupling between parallel sections of wire having opposing current vectors and increased coupling between parallel sections of wire having reinforcing current vectors. This antenna exhibits a resonant frequency of 287 MHz. (c) Modified normal mode helix antenna with the helical coil section tightly coupled near the base of the antenna. This antenna has a resonant frequency of 279.5 MHz.

343.5 MHz. In this configuration of the meander line antenna, the overall diameter has increased to 6.67 cm. An alternate meander line antenna, depicted in Figure 10.12b, illustrates decoupling of opposing current vectors and increased coupling between reinforcing current vectors. This antenna exhibits a resonant frequency of 287 MHz. In the normal mode helix configuration of Figure 10.12c, the horizontal section of wire near the ground plane is elevated to match that of the meander line configurations and the helical coil is

tightly compressed near the base of the antenna to increase self-inductance. This antenna has a resonant frequency of 279.5 MHz, the lowest of the three configurations.

It was briefly noted that the length and spacing of horizontal wire sections above the ground plane can also significantly impact the antenna's resonant frequency. This is particularly significant for sections of wire near the feed point where the current has the largest magnitude. The resonant frequencies of the meander line and normal mode helix antennas can be further reduced by removing the horizontal sections of wire near the feed point or increasing their height above the ground plane. The meander line antenna of Figure 10.12b is modified by removing the horizontal section of wire near the feed point as illustrated in Figure 10.13a. The antenna depicted in Figure 10.13a has the same overall height (10 cm), the same total wire length (30 cm), and the same overall diameter (3.3 cm). Rearranging the wire in this manner reduces the resonant frequency from 287 to 274 MHz. The normal mode helix of Figure 10.11b is modified by completely removing the horizontal sections of wire, as illustrated in Figure 10.13b, while the other physical parameters remain unchanged. This configuration exhibits a resonant frequency of 305 MHz, only a slight reduction from 312.6 MHz, the resonant frequency of the normal mode helix of Figure 10.11b. The final configuration considered in this section is a modified version of the helical coil antenna of Figure 10.12c, with the horizontal sections of wire removed from the structure. The number of turns in the helical coil section is adjusted so that the total wire length is maintained at 30 cm. This configuration is depicted in Figure 10.13c and exhibits a resonant frequency of 251 MHz. A comparison of the resonant frequency of the antennas depicted in Figures 10.11–10.13 is presented in Table 10.2.

In the preceding paragraphs, several aspects of the small antenna's geometry and their impact on resonant frequency were described. In general, if there is a large amount of

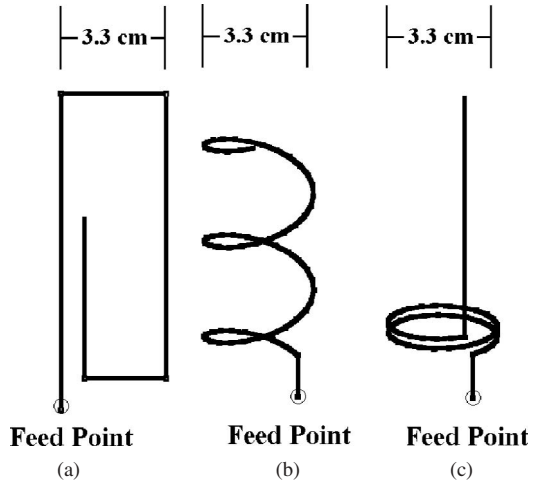


Figure 10.13 (a) Modified meander line antenna with the horizontal section of wire near the feed point removed, exhibiting a resonant frequency of 274 MHz. (b) Modified normal mode helix antenna with horizontal sections of wire removed, exhibiting a resonant frequency of 305 MHz. (c) Modified normal mode helix antenna with the tightly coupled helical coil section near the base of the antenna. Horizontal sections of wire in the helical coil have been removed. This antenna has a resonant frequency of 251 MHz.

TABLE 10.2 Comparison of the Resonant Frequencies of the Antenna Configurations Depicted in Figures 10.11–10.13

Antenna	Resonant Frequency (MHz)
Meander line—Figure 10.11a	361
Normal mode helix—Figure 10.11b	312.6
Meander line—Figure 10.12a	343.5
Meander line—Figure 10.12b	287
Normal mode helix (tight coil at base)—Figure 10.12c	279.5
Meander line—Figure 10.13a	274
Normal mode helix—Figure 10.13b	305
Normal mode helix (tight coil at base)—Figure 10.13c	251

wire to be contained within a small overall height and narrow cylindrical diameter, the helical coil shape is optimum in achieving the lowest resonant frequency. If there is not a substantial amount of wire to be contained within the antenna volume, or the antenna is allowed to have a wide cylindrical diameter, other approaches may be easier to implement and may offer a lower resonant frequency. The helical coil shape is usually the optimum approach when there is very close spacing between parallel sections of wire. In many instances, the helical coil shape may not be an option or even viable—for example, in applications where a planar antenna is required. In general, lower resonant frequency is achieved with closer spacing of parallel sections of wire where the current vectors align. Lower resonant frequency is also achieved when the geometry utilizes the fullest extent of the available height and overall volume.

10.6.2.2 Antennas of Differing Geometry That Are Resonant at the Same Frequency In the preceding section, the resonant frequencies of several antennas were compared as a function of altering details within the antenna geometry. In all cases, each of the antennas had the same overall height and the same conductor diameter and, in most instances, the same overall diameter. In this section, the total wire length in each of these antennas is adjusted so that each antenna operates at the same resonant frequency. The question considered in this section is the following: Once resonance is achieved, is there a unique geometry that provides optimum performance or at least substantially better performance than other geometries? One could argue that without examining all conceivable geometries, this question can never really be answered. However, we seek to examine a sufficient number of geometries that will allow us to draw some reasonable engineering conclusions or guidelines, recognizing that there may or will always be exceptions that can be found.

The total wire length in each of the antennas depicted in Figures 10.11–10.13 is adjusted so that they are all resonant at or very near 361 MHz, the resonant frequency of the meander line antenna depicted in Figure 10.11a. One additional antenna, depicted in Figure 10.14, is also considered. It has a somewhat arbitrary geometry, an overall height of 10 cm, a conductor diameter of 1 mm, and an overall cylindrical diameter of 6.6 cm. A comparison of the antennas' resonant performance properties is presented in Table 10.3.

From the results presented in Table 10.3 it is evident that most of the antennas exhibit very similar electrical performance properties even though their geometries are remarkably different. This is consistent with previous work presented in the literature describing a variety of meander line, normal mode helix, and arbitrary geometries [13]. While one

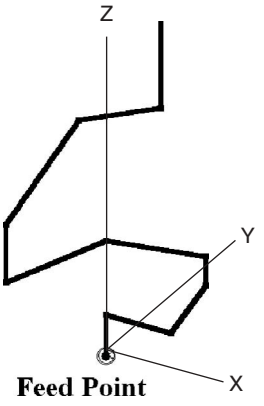


Figure 10.14 Depiction of an antenna having an arbitrary geometry, an overall height of 10 cm, an overall diameter of 6.6 cm, and a conductor diameter of 1 mm.

TABLE 10.3 Comparison of the Resonant Performance Properties of the Antenna Configurations[a] Depicted in Figures 10.11–10.14

Antenna	Resonant Frequency (MHz)	Total Wire Length (cm)	Resistance (Ω)	Q	η_r (%)
Meander line—Figure 10.11a	360.6	30	9.7	24.5	97.3
Normal mode helix—Figure 10.11b	360.1	25	10.9	21.8	98
Meander line—Figure 10.12a	360.5	28.4	9.5	24.3	97.4
Meander line—Figure 10.12b	360.4	23.9	10.7	22.9	97.9
Normal mode helix (tight coil at base)—Figure 10.12c	360.9	24.5	5.2	51.7	95
Meander line—Figure 10.13a	360.4	22.4	9.3	26.3	98
Normal mode helix—Figure 10.13b	360.7	24	11.1	22.1	98.1
Normal mode helix (tight coil at base)—Figure 10.13c	360.3	22.6	5.1	51.5	95.5
Arbitrary antenna—Figure 10.14	360.5	25.6	7.4	30.3	97

[a] All of the antennas have an overall height of 10 cm and a conductor diameter of 1 mm.

will always be able to find exceptions, resonant wire antennas of the same height generally exhibit nearly the same properties independent of differences in their geometries and total wire lengths. This is particularly true for wire antennas that have significant wire length confined to a narrow cylindrical diameter and where the wire is wound away from the feed point with constant or increasing height above the ground plane. With the inductively loaded designs considered here, increased resonant resistance is generally achieved with narrower cylindrical diameter and lower Q (larger bandwidth) is achieved with increasing cylindrical diameter.

Antennas where the wire length is more concentrated in specific regions of the geometry, like a top-loaded antenna or the helical coil antenna of Figure 10.12c, may exhibit significantly different properties. In a subsequent section, a different class of small antennas is presented, the capacitively loaded designs, which will be shown to exhibit

substantially different properties when they are made to be resonant at the same frequency as the inductively loaded designs presented here. However, it will also be shown that, with these antennas, their radiation properties are often independent of the top-loading geometry or technique.

In all cases, the resonant properties of the small antenna are significantly a function of the antenna's overall height and occupied volume.

10.6.2.3 Radiation Pattern Considerations In the previous section, the impedance Q and the radiation efficiency of a number of inductively loaded (increased wire length) antennas were considered. In many instances, the resonant resistances, Q values and radiation efficiencies are very similar even for remarkably different geometries. The same is often but not always true for the radiation pattern shape and polarization. In most cases, the small antenna with more vertical height than horizontal width has predominantly vertical current and the antenna exhibits a vertically polarized dipole or monopole pattern. In some instances, where the horizontal component of current is significant, the antenna will also exhibit substantial horizontal polarization in the overhead (zenith) direction.

The radiation patterns, including both vertical and horizontal polarizations, for the antennas depicted in Figures 10.11, 10.12, and 10.14 are presented in Figure 10.15. In all cases, the dominant pattern shape and polarization are consistent with those of a vertically polarized antenna. With the antennas that have substantial horizontal current, like those of Figures 10.12b and 10.13a, there is significant horizontal polarization in the zenith direction.

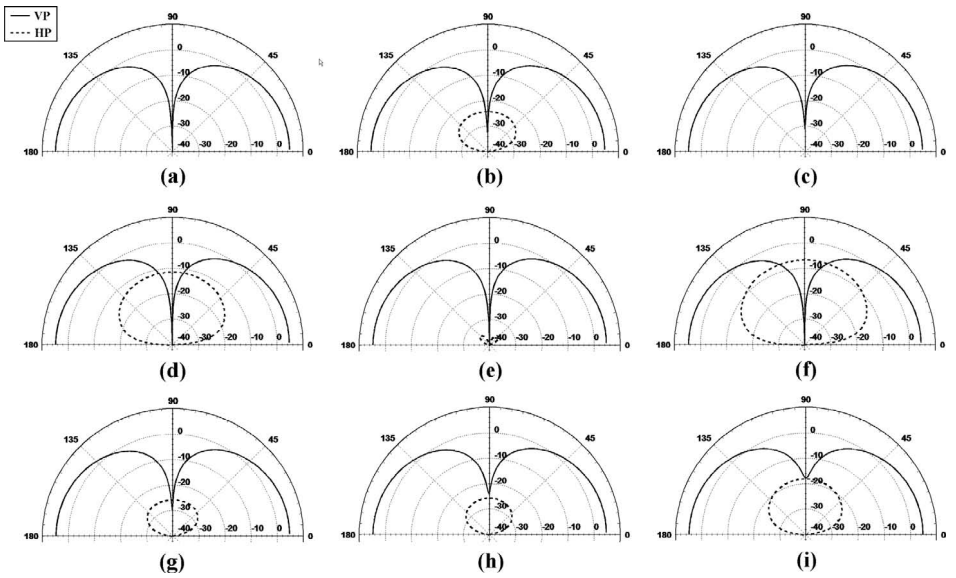


Figure 10.15 Radiation patterns for the antennas depicted in (a)Figure 10.11a; (b) Figure 10.11b; (c) Figure 10.12a; (d) Figure 10.12b; (e) Figure 10.12c; (f) Figure 10.13a; (g) Figure 10.13b; (h) Figure 10.13c; and (i) Figure 10.14. All of the antennas have the same height (10 cm) and they are resonant at or very near the same frequency of 360.5 MHz.

10.6.2.4 Comments on Wire Diameter In the previous sections, the wire diameter was held at a constant value for all of the antennas considered. For antennas having substantial wire length compressed into a small height, the wire diameter also has a significant impact on resonant frequency. For example, if the wire diameters of the normal mode helix and meander line antennas described in Table 10.1 are adjusted, there will be a notable change in resonant frequency. Generally, as the wire diameter is decreased, there is a corresponding decrease in resonant frequency. If the wire diameter is increased (to the physical extent allowed by the interwire spacing), there is a corresponding increase in resonant frequency. A summary of the resonant frequencies of the normal mode helix and meander line antennas as a function of wire diameter is presented in Table 10.4. The wire diameters were adjusted between 0.25 and 2 mm. The normal mode helix and meander line antennas have total wire lengths of approximately 33.3 and 58 cm, respectively. When these configurations have a wire diameter of 1 mm they are resonant at 313 MHz.

With the differences in the wire arrangements and resulting current vector alignments between the two antennas, the change in wire diameter has differing impact on the resonant frequencies. Typically, with geometries having more tightly coupled wires (closer spacing), there is a greater variation in resonant frequency.

The final point to note regarding the wire diameter is the significant impact it has on the antenna's radiation efficiency. As one would expect, with decreasing wire diameter there is a corresponding decrease in the antenna's radiation efficiency.

10.6.3 Increasing Capacitance to Ground

It was illustrated in Sections 10.6.1 and 10.6.2 that the resonant frequency of a fixed height antenna could easily be decreased by simply adding wire length to the antenna structure. This increase in total wire length increases the antenna's self-inductance, thereby lowering the resonant frequency. In addition to implementing an increase in self-inductance, an increase in capacitance to ground can be used to lower the resonant frequency of a monopole antenna. An increase in capacitance is easily accomplished by top-loading the straight-wire monopole antenna with a top hat or disk [12, 20], as illustrated in Figure 10.16a, which depicts top-loading of a 10-cm monopole using a wire-grid disk structure. The wire-grid structure is chosen so that EZNEC could be used to model the antenna, allowing direct comparison with the inductively loaded antennas described in the previous sections. The resonant frequency of the top-loaded monopole is decreased by

TABLE 10.4 Resonant Frequency Comparison of the Normal Mode Helix and the Meander Line Antennas as a Function of Different Wire Diameters[a]

Wire Diameter (mm)	Normal Mode Helix Resonant Frequency (MHz)	Meander Line Antenna Resonant Frequency (MHz)
0.25	293	264
0.5	302.5	284.5
0.75	308	300
1	313	313
1.5	322	336.5
2	329.5	357

[a] All of the antennas have an overall height of 10 cm.

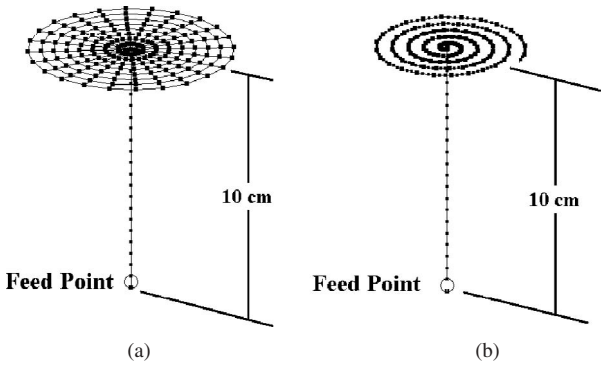


Figure 10.16 (a) Depiction of the 10-cm monopole antenna loaded with (a) a wire-grid top hat and (b) a spiral wire top hat.

increasing the diameter of the top hat. Note that top-loading can also be implemented in conjunction with inductive loading to further lower resonant frequency [20]. For example, the resonant frequency of the inductively loaded configurations discussed in the previous sections can be lowered by adding a top plate to their structure.

Top-loading can also be accomplished using other top-loaded configurations such as the spiral configuration [20, 21] depicted in Figure 10.16b. With the spiral-loaded configuration, the resonant frequency is decreased by both the increase in capacitance and the increase in inductance achieved by increasing the spiral wire length.

For comparative purposes, the resonant frequencies of these two top-loaded configurations are adjusted to match those of the normal mode helix and meander line antennas presented in Table 10.1. Table 10.5 presents the physical and resonant properties of the top-loaded antennas. The maximum radius of the spiral top-loaded antenna is set to match that of the wire-grid top-loaded antenna. Examining the results in Table 10.5 as compared to the results of Table 10.1 illustrates that, for a given resonant frequency, the top-loaded antennas exhibit higher resonant resistance, lower Q (wider bandwidth), and higher radiation efficiency. The differences are more pronounced at the lower resonant

TABLE 10.5 Performance Comparison of the Wire-Grid and Spiral Top-Loaded Antennas[a] at the Same Resonant Frequencies as the Normal Mode Helix and the Meander Line Antennas Described in Table 10.1

	Wire-Grid Top-Loaded Antenna				Spiral Top-Loaded Antenna			
Resonant Frequency (MHz)	Disk Diameter (cm)	Resistance (Ω)	Q	η_r (%)	Total Wire Length (cm)	Resistance (Ω)	Q	η_r (%)
150	23.7	4.7	23.2	82.9	54.8	4.3	65	92
178.25	18.3	6.1	20.6	89.6	46.6	5.8	47.5	95
255.8	10	11.2	14.8	96.5	33.6	11.1	24.7	97.7
313	7	15.7	12.2	98.1	27.8	15.6	17.5	98.4
419	3.6	23.9	9.1	99.4	21.4	24.0	10.9	99.2

[a]Both of the antennas have an overall height of 10 cm and a conductor diameter of 1 mm.

frequencies. In general, capacitive top-loading is a more effective technique for lowering the resonant frequency. The obvious physical trade-off for the improved performance is the substantial increase in cylindrical diameter necessary to accommodate the top hat. At 150 MHz, the wire-grid capacitive top hat has an overall diameter of 23.7 cm. This substantially increases the overall electrical size or volume of the antenna: its *ka* at 150 MHz is 0.488 whereas the *ka* of the normal mode helix described in Table 10.1 is 0.316. This increase in electrical volume contributes to the antenna's lower Q.

Other points to note regarding the results presented in Table 10.5 relate to the relative performance of the wire-grid and spiral top-loaded antennas themselves. In each case, the radiation resistances of the antennas are nearly identical, indicating that this property is primarily established by the physical height of the antenna rather than the implementation methodology of the top hat. At 150 MHz, the two antennas are remarkably different, as illustrated in Figure 10.17. The difference in the top hat configuration plays a significant role in determining the antenna Q. At the lower frequencies, the wire-grid configuration's capacitance (near field) is distributed more uniformly over a larger volume, decreasing the stored or reactive energy surrounding the antenna, resulting in a larger effective volume and lower Q than exhibited by the spiral top hat configuration.

In determining the results presented in Table 10.5, the overall radius of the spiral configuration was set to match that of the wire-grid configuration. Since the spiral begins at the center of the monopole geometry and the wire layout may not exceed more than one turn, the spiral geometry may occupy much less of the overall circular area defined by the geometric radius (this is particularly true for the configuration depicted in Figure 10.17b). The spiral wire layout can be confined to a smaller radius, causing it to occupy more of the total circular area. An alternate version of a 150-MHz spiral-loaded antenna, confined to a smaller circular diameter, is depicted in Figure 10.18. The spiral geometry is confined to an overall diameter of 5 cm rather than the 23.7 cm used with the 150-MHz spiral configuration shown in Figure 10.17b. The spiral top-loaded antenna shown in Figure 10.18 is resonant at approximately 150.8 MHz with a resonant resistance of 5 Ω, just slightly higher than the resistance of the larger diameter configurations. Its Q is approximately 252, a substantial increase relative to the other configurations. The increase in Q occurs as a result of its much narrower diameter and the large amount of wire length confined to the top of the structure. This substantially increases the near fields within the smaller volume defined by the antenna. The total wire length in the antenna is 60.4 cm.

The resonant properties, particularly the resonant resistance, of the wire antennas considered in this and the previous sections are nearly the same independent of the total

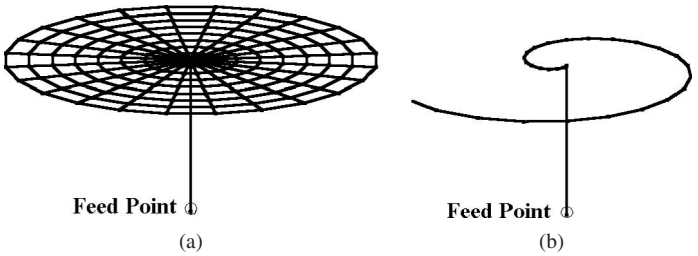


Figure 10.17 (a) Depiction of the 150 MHz monopole antenna loaded with (a) a wire-grid top hat 23.7 cm in diameter and (b) a spiral wire top hat 11.87 cm in maximum radius.

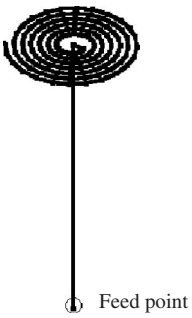


Figure 10.18 Depiction of the 150- MHz monopole antenna loaded with a spiral wire top hat 5 cm in maximum radius.

wire length or the geometric details of the antenna. While different design approaches (e.g., inductive loading versus top-loading) may have substantially different resistances at the same resonant frequency, the resonant resistance achieved with each design approach remains nearly the same. Another approach for achieving a lower resonant frequency is to enclose an antenna within a high permittivity dielectric material. Surrounding the antenna with a high dielectric material effectively increases the capacitance between the antenna and ground (for the monopole configuration) or between the upper and lower arms of the antenna (for the dipole configuration).

Here we use Microwave Studio to examine the resonant resistance properties of a straight-wire dipole antenna surrounded by a dielectric cylinder [22] as illustrated in Figure 10.19. The dipole has an overall length of 12.6 cm and a conductor diameter of approximately 1.2 mm. The dielectric cylinder is assumed to be lossless (more on this shortly) with a diameter that varies from 0.5 to 5 cm with dielectric constants of $\varepsilon_r = 4, 9,$

Figure 10.19 Depiction of the straight-wire dipole enclosed with a dielectric cylinder. The dipole has an overall length of 12.6 cm. The diameter of the dielectric cylinder ranges from 0.5 to 5 cm.

and 12. The resonant frequency as a function of dielectric constant and cylinder diameter is shown in Figure 10.20. As expected, the resonant frequency decreases with increasing dielectric constant and increasing diameter.

More interesting is the comparison of resonant resistance as a function of resonant frequency shown in Figure 10.21. We see that much like the antennas considered in the previous sections, once resonance is established, the resonant resistance is primarily a function of the antenna's overall height rather than the details of how resonance is achieved.

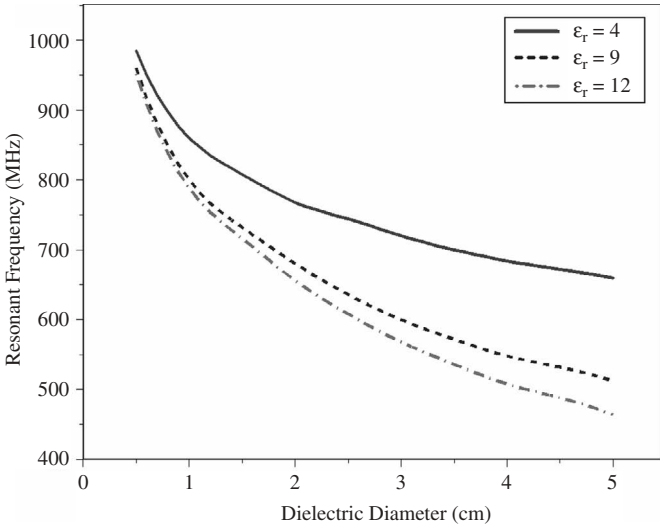


Figure 10.20 Resonant frequency versus dielectric diameter for the dipole antenna enclosed with a dielectric cylinder having values of $\varepsilon_r = 4$, 9, and 12.

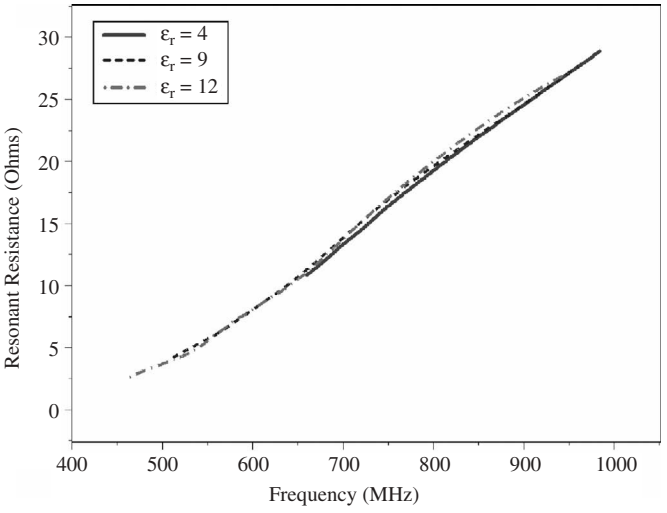


Figure 10.21 Radiation resistance versus resonant frequency for the dipole antenna enclosed within a dielectric cylinder.

Two final points to consider are the antenna's Q and the effect of dielectric losses on the performance of the antenna. With decreasing frequency, the Q of the free-space antenna increases considerably. This increase in antenna Q, coupled with losses in the dielectric, has a significant impact on the antenna's radiation efficiency at lower frequencies [23]. The small antenna's high Q implies the presence of very high values of near fields surrounding the antenna. With the introduction of a lossy dielectric (even those with a low loss tangent), a substantial portion of the high near-field energy can be converted to loss, significantly reducing the antenna's radiation efficiency. This is an important factor to consider when designing small antennas with dielectrics. With the dipole antenna considered above, the radiation resistance of the free-space dipole is quite low at the lower frequencies. This low radiation resistance contributes to the reduction in efficiency, particularly when lossy materials are introduced into the antenna design. To mitigate this issue to some extent, one should begin with small antennas having as high a radiation resistance as possible.

10.7 IMPEDANCE MATCHING AND INCREASING RADIATION RESISTANCE

In the previous sections, a number of simple techniques were presented that can be used to achieve self-resonance with an electrically small antenna. To further implement an impedance match, it is necessary to transform the small antenna's low resonant resistance to the desired value of characteristic impedance. If the small resonant antenna is to be connected to a device having a similarly low input resistance, further matching would not be required. In many instances, we assume that the small antenna is to be used in a 50-Ω system and we seek a method for transforming the antenna's low resistance to 50 Ω. This can be accomplished with a number of methods, such as the use of external matching components or a λ/4 impedance transformer, or the matching can be incorporated directly within the geometry of the antenna.

Any antenna can be impedance matched at a single frequency using external reactive components. This matching technique is straightforward and covered extensively in the open literature and will not be considered here. One of the most significant issues to consider with this matching approach is level of loss introduced by the matching components. This loss contributes directly to the reduction in overall efficiency. In some instances, the increase in dissipation loss may be less than the improvement in mismatch loss and the overall efficiency improves. However, it is generally desired to impedance match the antenna as efficiently as possible. When using external reactive components, the most important first step is to begin with an antenna having as high a radiation resistance as possible [8].

The antennas considered in the previous sections were made to be self-resonant when their overall height or length was much less than the resonant wavelength (λ_r). Since the antenna's radiation resistance behaves as $K(h/\lambda_r)^2$, where K is a constant that varies as a function of design approach, it is always less than the radiation resistance of the full-size monopole or dipole, where $h \approx \lambda/4$ or $l \approx \lambda/2$, respectively. In this case, the impedance will behave as shown in the Smith chart presented in Figure 10.22. This is the impedance of the normal mode helix depicted in Figure 10.11b that was made to be resonant at 360.1 MHz. Any antenna exhibiting this behavior can be impedance matched to 50 Ω by placing a parallel inductor at the feed point. The impedance will move toward 50 Ω as

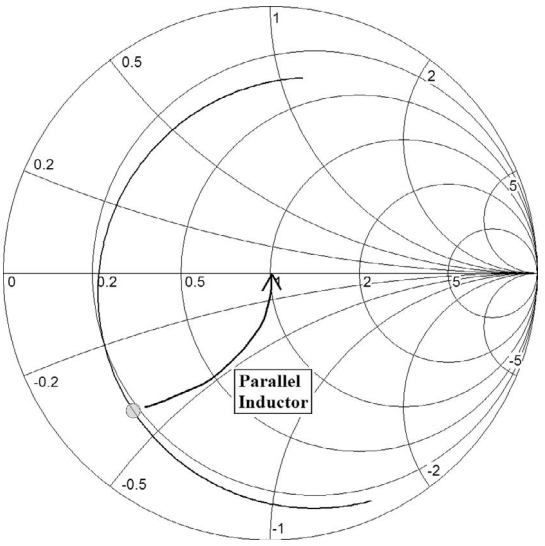


Figure 10.22 Typical impedance of an electrically small monopole (or dipole) antenna made to be self-resonant when the antenna's height or overall length is much less than the resonant wavelength. An antenna of this type can be impedance matched to 50 Ω by using a parallel inductor at the antenna's feed point.

shown in Figure 10.22. With a very large inductance, there will not be an appreciable movement in the impedance toward 50 Ω. With a very small inductance, the impedance will move closely toward the short-circuit end of the Smith chart.

When the parallel inductor is placed exactly between the antenna's feed point and the ground plane (or directly at the feed point between the upper and lower arms in the case of a dipole), the match frequency will be lower than the antenna's resonant frequency. When the parallel inductor is placed away from the feed point, the equivalent impedance being matched is not exactly the same as the feed point impedance and the match frequency and the required matching inductance will vary. If a slight adjustment of match frequency is required, the antenna's wire length and matching inductance can be adjusted accordingly.

A simple method for implementing a parallel matching inductor at or very near the feed point of the small antenna is to use a shunt (or parallel) matching stub as illustrated in Figure 10.23 [24]. Note that this technique is exactly that used to make an inverted-L antenna an inverted-F antenna [25]. As the stub length is increased, the equivalent parallel inductance is increased. A very long stub will not move the impedance appreciably, whereas a very short stub will move the impedance very close to the short-circuit end of the Smith chart as illustrated in Figure 10.24a. The matched impedance for the normal mode helix is shown in Figure 10.24b. Since the shunt stub is located very close to the antenna's feed point, the match frequency of 349 MHz is less than the resonant frequency of 360.1 MHz.

One of the practical challenges associated with matching any small antenna using any matching technique is achieving the impedance match at exactly the desired frequency. With any high Q small antenna, the precise match frequency is very sensitive to changes in the reactance of the matching components. This is true when using external lumped

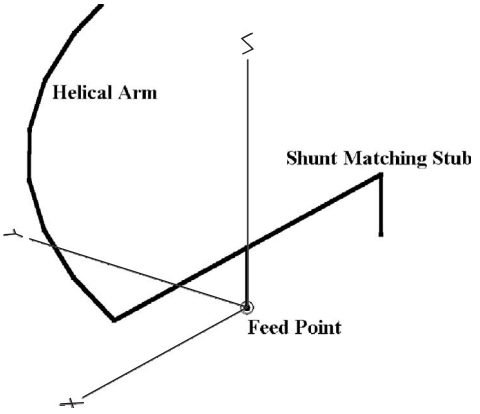


Figure 10.23 Depiction of a shunt or parallel stub that is used to impedance match a small resonant antenna having a resonant resistance less than the desired match impedance. This technique is analogous to the conversion of an inverted-L antenna into an inverted-F antenna.

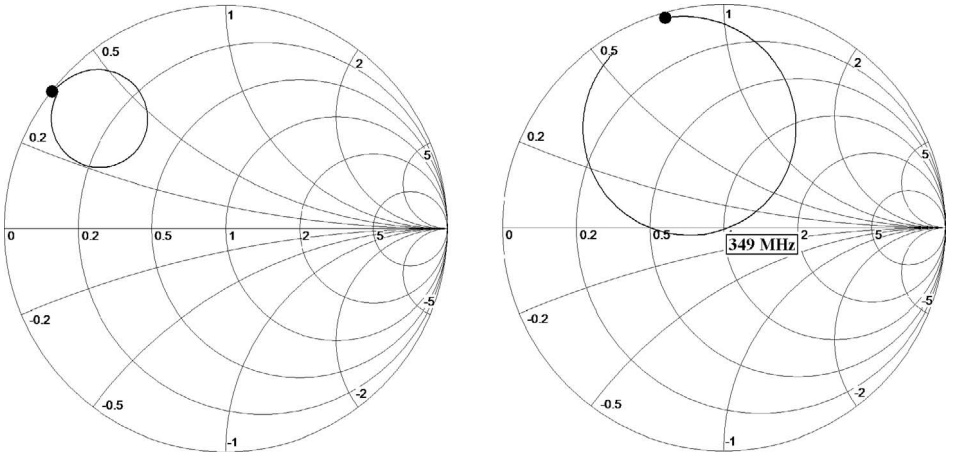


Figure 10.24 (a) Impedance of the normal mode helix with a short shunt stub illustrating how the impedance moves toward the short-circuit side of the Smith chart. (b) Impedance of the normal mode helix matched using an appropriate length stub. The match frequency is 349 MHz, which is lower than the resonant frequency because the shunt stub is located near the antenna's feed point.

matching components or the shunt stub approach illustrated here. For this reason, antenna design approaches that optimize the Q and bandwidth properties of the small antenna are of critical importance. This will be addressed in the next section.

When using a shunt stub to impedance match a small antenna, the Q and therefore operating bandwidth of the antenna does not change significantly since the Q is primarily established by the antenna's height and occupied volume (the antenna's utilization of the available volume defined by the value of ka) [13]. Another issue associated with using this matching technique is that both the antenna's radiation and loss resistance

are simultaneously transformed to 50 Ω and there is not an appreciable change in the antenna's radiation efficiency. If the small antenna exhibits a low radiation efficiency before it is matched, it will exhibit a low radiation efficiency after it is matched. To achieve optimal radiation efficiency with a small antenna it is necessary to achieve as high a radiation resistance as possible.

One technique that can be used to achieve a high radiation resistance is the use of multiple folded arms within the antenna structure [26–29]. For example, consider the approximate λ/4-wavelength single-arm and multiple-arm folded monopoles depicted in Figure 10.25. The single-arm monopole exhibits a radiation resistance of approximately 36 Ω, whereas the two-, three-, and four-arm folded monopoles exhibit radiation resistances of approximately 142, 418, and 543 Ω, respectively. In a 50-Ω system, there would be no advantage to using these multiple-arm folded monopoles since their resonant resistances are too high. However, if this technique is applied to a small antenna having a 3.6-Ω resistance, it may be possible to transform the radiation resistance to 54.3 Ω, resulting in an impedance match and high radiation efficiency, since this impedance transformation technique does not simultaneously transform the conductor's loss resistance. Note that any change in resonant frequency occurring with this process can be mitigated by simply adjusting the wire or conductor length within the antenna structure.

An example of using this technique to efficiently impedance match a small antenna can be illustrated with the inverted-L antenna depicted in Figure 10.26a. The inverted-L

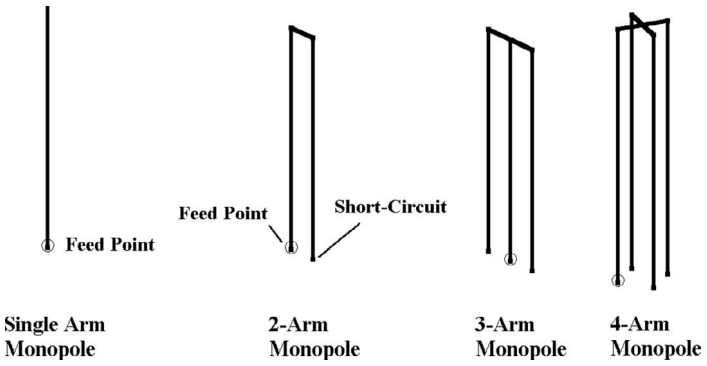


Figure 10.25 Depiction of the approximate λ/4-wavelength single-arm and multiple-arm folded monopole antennas. The single-arm monopole has a radiation resistance of 36 Ω, whereas the two-, three-, and four-arm monopoles have radiation resistances of 142, 418, and 543 Ω, respectively.

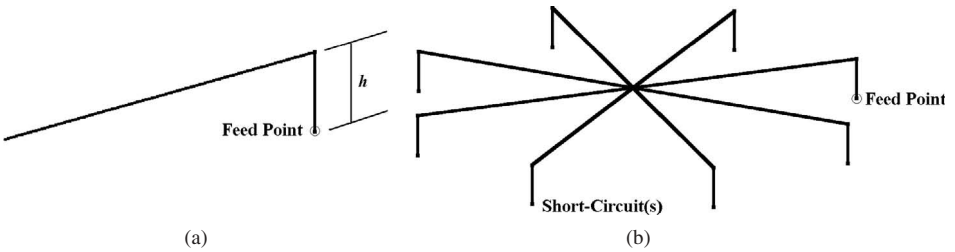


Figure 10.26 (a) Depiction of the inverted-L antenna resonant at 1370.3 MHz with a resonant resistance of (a) 5 Ω, and (b) 53 Ω.

antenna shown in Figure 10.26a has a height, h, equal to 1 cm, and a total conductor length equal to approximately 5.5 cm. The inverted-L's resonant frequency and resonant resistance are 1370.3 MHz and 5 Ω, respectively. With its low resonant resistance, the inverted-L can easily be matched using a shunt stub creating an inverted-F antenna. Here we will match the inverted-L using multiple folded arms.

To increase the radiation resistance of the inverted-L, additional inverted-L arms are added to the structure until the resistance is transformed to a value as close to 50 Ω as possible. The final configuration, an eight-arm inverted-L antenna is depicted in Figure 10.26b [22]. As inverted-L arms are added to the structure, both the resonant frequency and resonant resistance change. Two-arm and four-arm versions of the inverted-L antenna have resonant frequencies of 1421 and 1524 MHz respectively. Their corresponding resonant resistances are 8 and 22.6 Ω. In the final eight-arm inverted-L configuration, the conductor length in each inverted-L section was increased to 6.6 cm (from 5.5 cm) to lower the resonant frequency to 1370.3 MHz. The resonant resistance of the eight-arm inverted-L configuration is 53 Ω. The VSWR of the antenna is presented in Figure 10.27. The antenna exhibits a relatively narrow 2:1 VSWR bandwidth of approximately 1.6% because of its small height ($\approx 0.046\lambda$) and the fact that it utilizes very little of the overall spherical volume defined by the value of ka.

The radiation pattern of the eight-arm inverted-L antenna is presented in Figure 10.28a. It exhibits a pattern similar to that of a vertically polarized monopole antenna because the horizontal components of current cancel within the antenna structure. This cancellation of horizontally polarized current causes the radiation resistance of the horizontal polarization mode to be substantially less than the radiation resistance of the vertical polarization mode. One might assume that this current cancellation is entirely a result of the placement of the horizontal wires close to the conducting ground plane and that the out-of-phase image currents in the ground plane cause the cancellation of horizontally polarized radiation. It is true that it is often difficult to deliver power to this mode since the corresponding radiation resistance is typically very low due to effects of current cancellation. However, this is not always the case as a horizontally polarized antenna closely spaced to a ground plane can exhibit high directivity and gain [10, 30]. As an example,

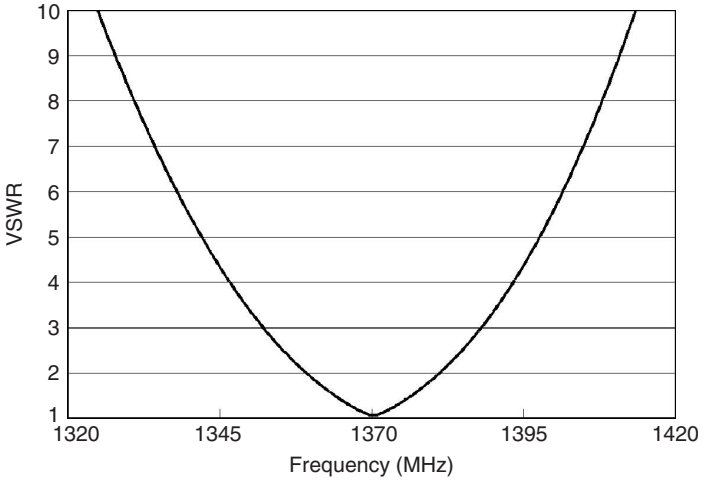


Figure 10.27 VSWR of the eight-arm inverted-L antenna.

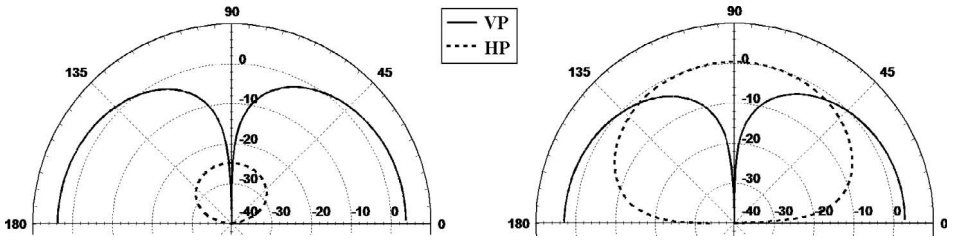


Figure 10.28 (a) Radiation pattern of the eight-arm inverted-L antenna at 1370.3 MHz. (b) Radiation pattern of the single-arm inverted-L antenna at 1370.3 MHz.

the radiation pattern of the single-arm inverted-L antenna shown in Figure 10.26a is presented in Figure 10.28b. The single-arm inverted-L antenna has nearly the same power in the horizontal polarization mode as it does in the vertical polarization mode.

The significant design variables for multiarm antennas of this type include the number of arms, the individual arm wire lengths, and the overall height of the antenna. Achieving a 50-Ω input impedance at any given frequency is a combination of all factors. Adjustment of the individual arm wire length sets the resonant frequency. The input impedance is a function of both the number of arms and the height of the antenna. Increasing the number of arms and/or the antenna height increases the resonant resistance. For example, consider the four-arm inverted-L antenna discussed earlier. It was resonant at 1524 MHz with a resistance of 11.8 Ω. To achieve a 50-Ω input impedance at 1370.3 MHz, we chose to increase the number of arms (to eight) and adjust the wire length to lower the resonant frequency. We could have increased the height of the four-arm antenna and adjusted the wire length to achieve the same result. If the height of the four-arm antenna is increased from 1 to 1.5 cm and the inverted-L wire length is changed to approximately 7.1 cm, the antenna exhibits a 45-Ω resistance at 1370.3 MHz. The bandwidth of the antenna also increases from 1.6% to approximately 2.9% because the antenna is taller and occupies more overall volume. Four-arm antenna designs based on these principles are also presented in Ref. 31.

10.8 OPTIMIZING BANDWIDTH

In the previous section, several techniques were described for efficiently impedance matching the small antenna to a 50-Ω characteristic impedance. Having accomplished an efficient impedance match, the remaining challenge in the design of a small antenna is to optimize its operating bandwidth. For the small antenna that exhibits a single impedance resonance, the maximum achievable bandwidth is limited by the fundamental bounds on Q defined in Eqs. (10.14) and (10.15). To maximize the operating bandwidth within the limits of Eqs. (10.14) and (10.15), the antenna must occupy as much of the spherical volume defined by ka as possible. Additionally, it is advantageous to keep as much of the electric current on the outside of the ka sphere as possible. Ideally, we seek to minimize the reactive fields within the volume of the ka sphere. We focus here on efficient antennas; however, it is well known that increased operating bandwidth can be achieved if the antenna is purposely (or inadvertently) made lossy.

A single impedance resonance antenna that accomplishes these objectives is the four-arm folded spherical helix described in Refs. 28 and 29. The antenna and its VSWR are depicted in Figure 10.29. It was designed to operate at 300 MHz and has an overall height of 8.25 cm ($ka \approx 0.26$). It exhibits a Q that is within approximately 1.5 times the lower bound defined by Eq. (10.14). Even though the antenna has a Q that closely approaches the lower bound, it exhibits a very narrow operating bandwidth. As the value of ka decreases, the operating bandwidth is fundamentally limited by the significantly increasing value of Q_{lb}. For an antenna exhibiting a single resonance, a common option for increasing bandwidth is to increase the loss within the antenna structure.

Recent work in optimizing the bandwidth of small antennas has focused on designing antennas that exhibit multiple impedance resonances within their defined operating bandwidth [32–35]. In Refs. 34 and 35, Stuart and Tran describe multiarm coupled resonator antennas that consist of a single driven arm that is electromagnetically coupled to one or several closely spaced duplicate arms. This design approach is of interest because antennas of this type are capable of exhibiting closely spaced multiple impedance resonances within their defined operating bandwidth. These designs are also of interest because they easily lend themselves to planar implementations, as shown in Figure 10.30, which depicts a two-arm planar, matched version of the multiarm spherical resonator design.

The two-arm planar resonator antenna shown in Figure 10.30 has a radius of approximately 10 cm and was designed to operate near 300 MHz. The asymmetry between the driven and coupled arms was implemented so that the antenna would be resonant and matched without the incorporation of lumped circuit elements within the antenna structure. Smaller antenna sizes can be achieved by increasing the number of coupled arms and by fabricating the antenna on dielectric boards [34, 35]. The impedance and a typical radiation pattern of the two-arm planar resonator are presented in Figure 10.31. The multiple impedance resonances exhibited by the antenna are evident in Figure 10.31a. The antenna first exhibits a 2:1 VSWR at a frequency of 292 MHz, where the value of $ka \approx 0.61$. The radiation pattern of the antenna exhibits both vertical and horizontal polarization since both the vertical and horizontal currents are strong within the structure. There is currently a substantial level of effort by Stuart and others seeking to optimize the impedance and radiation patterns of coupled resonator designs at smaller values of ka.

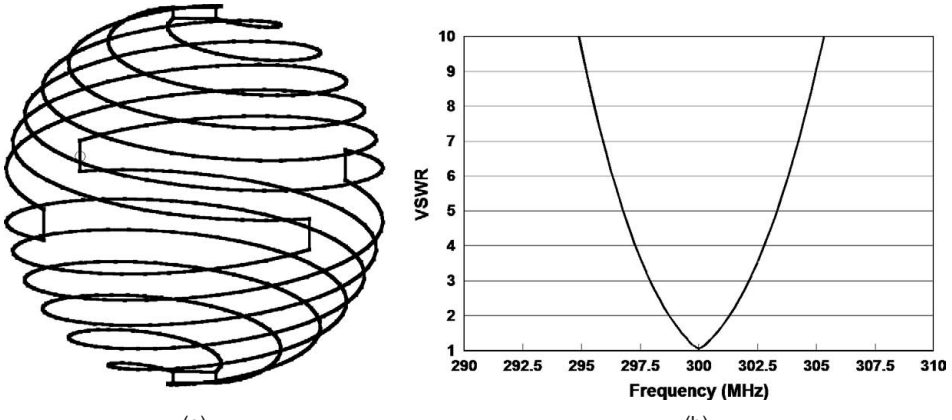


Figure 10.29 (a) Depiction of the four-arm folded spherical helix designed for 300 MHz with a value of $ka = 0.263$. (b) VSWR of the four-arm folded spherical helix antenna.

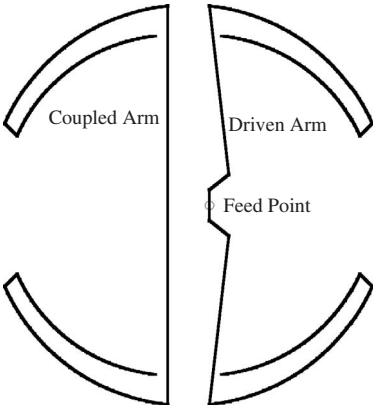
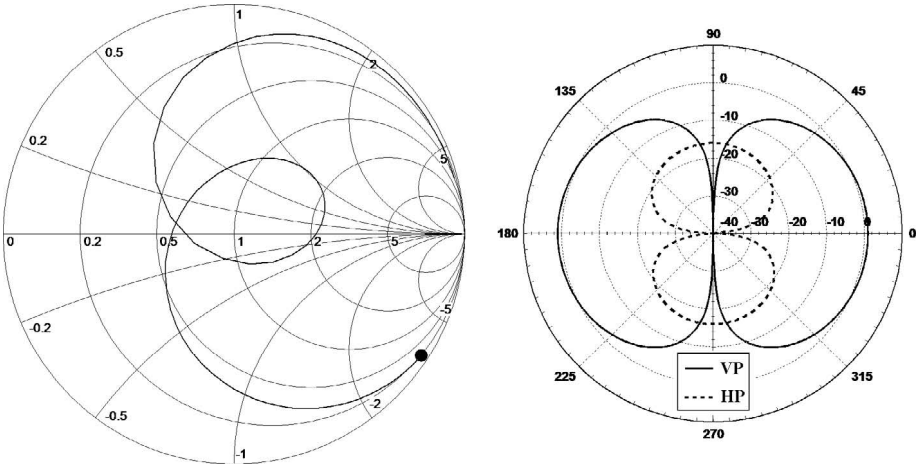


Figure 10.30 Depiction of the matched two-arm, planar spherical resonator antenna.



(a) (b)

Figure 10.31 (a) Impedance of the planar two-arm resonator of Figure 10.30. (b) Radiation pattern of the two-arm resonator antenna of Figure 10.30.

The concept of increasing the operating bandwidth of a small antenna by introducing multiple impedance resonances is not a new concept. One of the most famous small antenna designs that exhibits multiple resonance behavior is the Goubau antenna [36–38], shown by a Microwave Studio model of the structure in Figure 10.32. The antenna has an overall height of 4.3 cm and a top-disk diameter of 12.3 cm. The Goubau antenna was fabricated by the author and was measured to have a VSWR, shown in Figure 10.33a, that is nearly less than 2:1 over a frequency range of 466–916 MHz, a 1.97:1 bandwidth ratio. The antenna's impedance, illustrating the multiple impedance resonances, is presented in Figure 10.33b. The Goubau antenna first exhibits a VSWR equal to 2:1 at a frequency of 466 MHz, where the value of $ka \approx 0.733$. No further optimization of the Goubau antenna's performance was performed at the time of this writing.

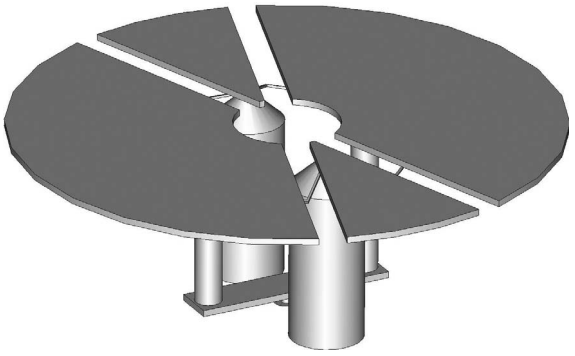


Figure 10.32 Depiction of the Goubau antenna having an overall height of 4.3 cm and a top-disk diameter of 12.3 cm.

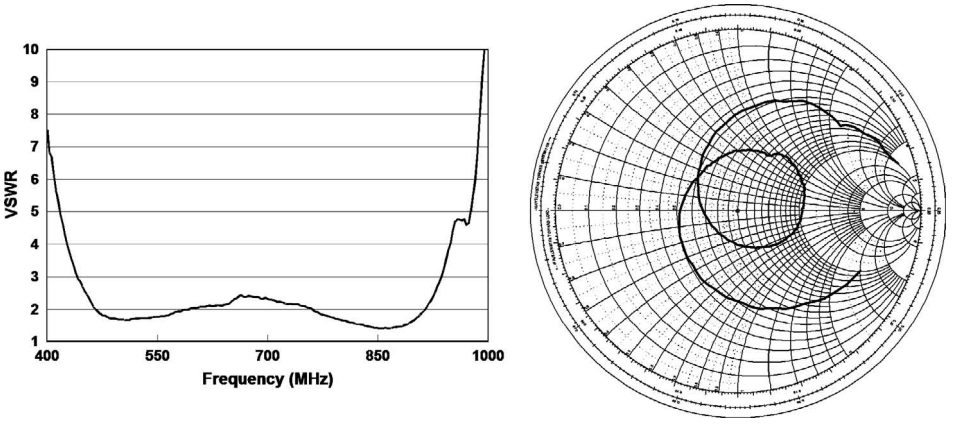


Figure 10.33 (a) Measured VSWR of the Goubau antenna. (b) Measured impedance of the Goubau antenna.

There are two critical issues associated with designing small wideband antennas. First is achieving a good VSWR and high radiation efficiency at the lowest frequency possible. The second is maintaining omnidirectional monopole-like patterns at the highest frequency possible. The Goubau antenna is efficient at the lower operating frequencies, where it exhibits a monopole-like omnidirectional radiation pattern as shown in Figure 10.34a, which presents the radiation pattern calculated at 470 MHz. This pattern is computed assuming the Goubau antenna operates over an infinite ground plane. The Goubau antenna exhibits monopole-like radiation patterns over the entire bandwidth. The radiation pattern at 920 MHz is presented in Figure 10.34b. It exhibits some asymmetry in the azimuth plane.

One disadvantage of the Goubau antenna is that its VSWR properties are very sensitive to the lengths of the tuning wires that are visible in Figure 10.32 just below the top plate. Small variations in wire length dramatically change the VSWR.

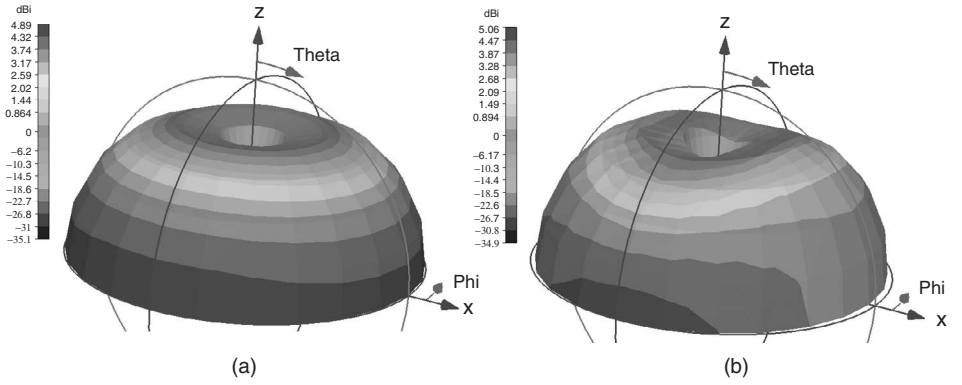


Figure 10.34 (a) Computed radiation pattern of the Goubau antenna at (a) 470 MHz and (b) 920 MHz. These patterns were computed with the Goubau antenna mounted over an infinite ground plane.

It is possible to design reasonably small antennas with wider operating bandwidths, and a substantial amount of work is currently underway in this area. In designing a small wideband antenna, it is important that the antenna structure be circularly symmetric about the feed point. Ideally, the feed point should be located at the center of the antenna and the structure should have the features of a body of revolution. This aids in ensuring that the radiation patterns at the upper end of the operating band are symmetric and ideally monopole-like.

For example, consider the antenna depicted in Figure 10.35. This antenna was designed with the intent of having much simpler construction than the Goubau antenna and a wider operating bandwidth [32, 33]. The antenna is essentially a disk-loaded folded monopole with impedance matching components of capacitance and inductance built into the feed point area of the antenna. The presence of the top disk aids in lowering the operating frequency and the folded arm aids in increasing the radiation resistance to a higher value. Impedance matching is achieved by adding parallel capacitance and series inductance

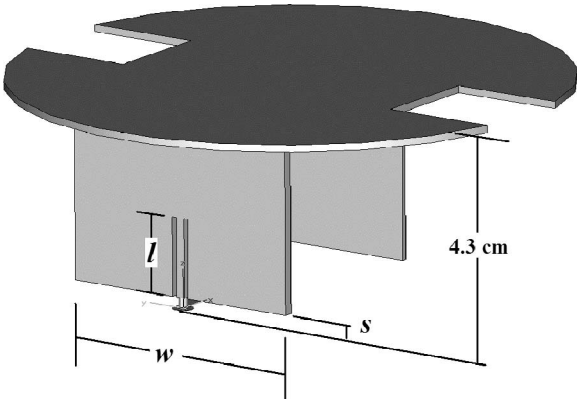


Figure 10.35 Depiction of a wideband antenna designed using a top disk and folded monopole section. Impedance matching is incorporated directly within the structure of the antenna.

directly within the feed point of the antenna. The parallel capacitance is implemented by the width, w, of the feed arm and its spacing, s, above the ground plane. The series inductance is added by introducing slots in the feed arm as shown. The series inductance is controlled by the length, l, of the slots. The antenna has the exact overall dimensions of the Goubau antenna. Its impedance and VSWR are presented in Figure 10.36. The antenna exhibits numerous resonances within its operating band and exhibits a VSWR less than 2:1 over a 3:1 frequency ratio, covering approximately 550–1635 MHz. At its lowest operating frequency, the value of ka is 0.865, a value higher than that of the Goubau antenna. The disadvantage of the antenna is that it is not symmetric about the feed point and, as a result, it should be expected to exhibit asymmetric radiation patterns, particularly at the upper end of the frequency band. One of the issues that becomes significant with these antennas at the upper end of the operating band is the current flow around the outer edge of the top disk. As the wavelength becomes smaller, details of the antenna geometry and the corresponding flow of current become more significant in determining the antenna's performance. The radiation patterns of this antenna at the lower and upper ends of the operating band are depicted in Figure 10.37. The significant asymmetries in the radiation pattern that occur at the higher frequencies are clearly evident.

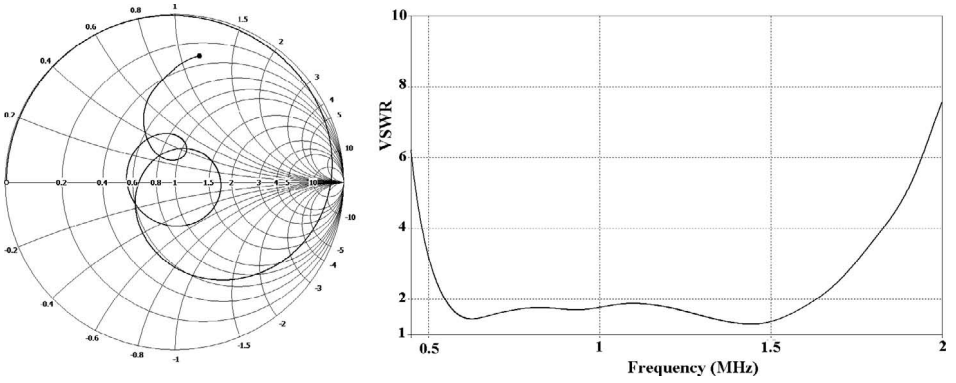


Figure 10.36 (a) Impedance of the wideband antenna depicted in Figure 10.35. (b) VSWR of the wideband antenna depicted in Figure 10.35.

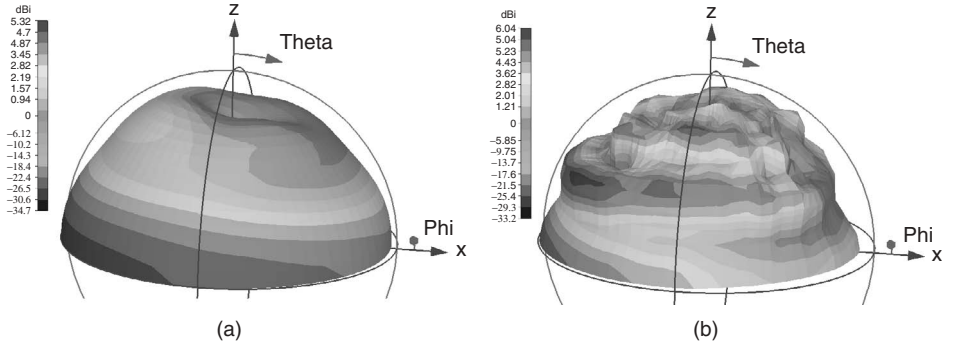


Figure 10.37 Radiation pattern of the wideband antenna at (a) 500 MHz and (b) 1600 MHz.

To illustrate the advantage of using a geometry that is symmetric about a feed point located at the center of the antenna, another disk-loaded folded monopole design is considered and is shown in Figure 10.38. The antenna has the same dimensions as the Goubau antenna, being 4.3 cm tall and having a disk diameter of 12.3 cm. The top disk is supported by a crossed bow-tie like structure with wide planar sections closely approaching ground to introduce a variable capacitance. There are four cylindrical rods connecting the top disk to ground, which act like a multiarm folded monopole. The return loss and radiation patterns of the antenna are presented in Figures 10.39 and 10.40, respectively. While the antenna does not operate at as low a frequency as the Goubau antenna, it does have a slightly wider bandwidth.

The antenna depicted in Figure 10.38 exhibits a 2:1 VSWR over a 2.18:1 frequency ratio and it has symmetric radiation patterns over the entire operating bandwidth. Near the upper end of the operating band, the pattern is symmetric but it does exhibit some variation in the azimuth plane. For many applications the azimuth variation would be acceptable.

The challenges in the design of small wideband antennas of this type are achieving low VSWR at low frequencies and monopole-like radiation patterns at the upper end of the operating bandwidth. Many of the techniques used to accomplish one of these goals

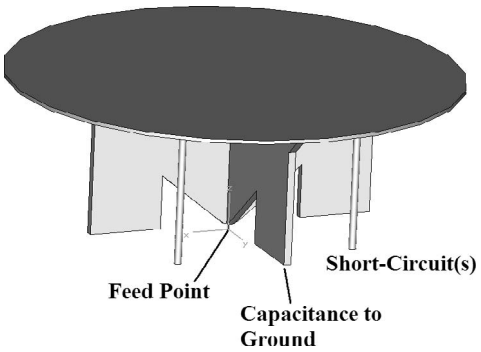


Figure 10.38 Depiction of a wideband antenna designed with a symmetric structure. The antenna has the same dimensions as the Goubau antenna.

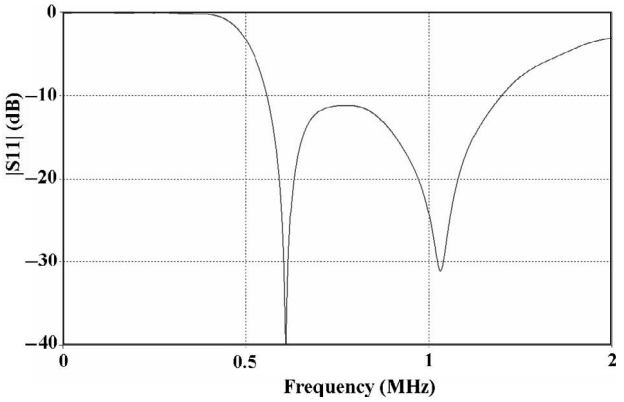


Figure 10.39 Return loss of the wideband antenna depicted in Figure 10.38.

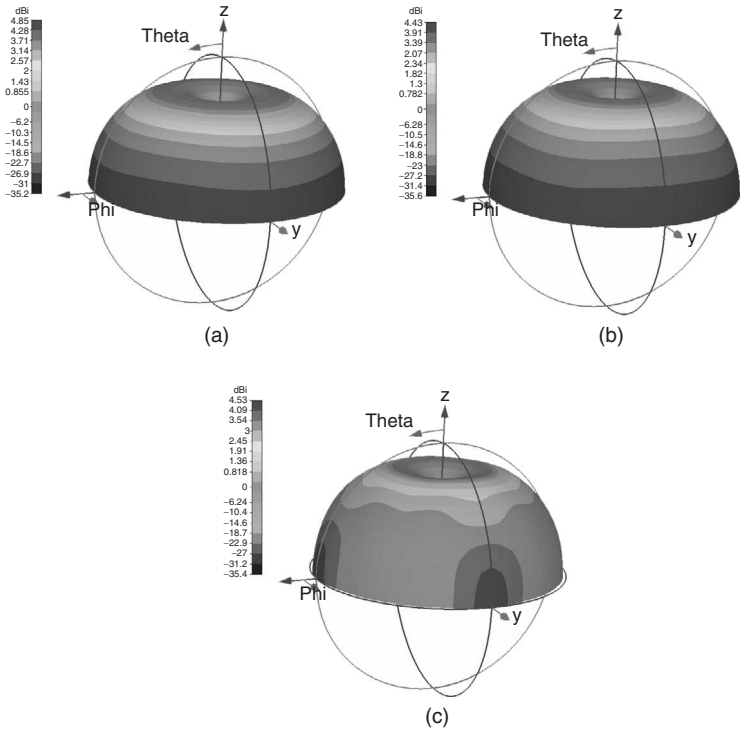


Figure 10.40 Radiation patterns of the wideband antenna depicted in Figure 10.38 at (a) 500 MHz, (b) 1000 MHz, and (c) 1500 MHz.

limits the achievement of the other. The best design approach is to locate the antenna feed point at the center of the structure and make the structure circularly symmetric about the feed point. Generally, as the lowest operating frequency decreases, so does the achievable bandwidth. This is currently a significant area of research and results better than those illustrated here have been achieved. Most recently, designs developed by Ravipati and Reddy [39] have demonstrated substantial increases in bandwidth.

10.9 FRACTAL ANTENNAS

Fractal antennas have been a subject of interest and some controversy since they were first introduced to the antenna community by Cohen in 1996 [40–42]. Cohen introduced the concept of fractal antennas with a family of small Minkowski island fractal loops. The question considered here is not whether fractal antennas work, since they generally work reasonably well. Rather, the questions considered here are: Are their performance properties unique? And more specifically, are their performance properties uniquely a function of the details within their fractal geometry? As is the case with many papers describing new antenna designs, when the performance properties of fractal antennas are described, there is often little or no performance comparisons with other antennas of the same size, volume, wire length, and so on. Early work speculated that fractal antennas might offer performance previously unattainable with other small antenna designs [43]

and that the multiband behavior of fractal antennas was uniquely a function of their self-similar fractal geometry [44]. Recently, in the antenna literature, the term fractal has been replaced with the term prefractal presumably to qualify the fact that the fractal geometry of any fractal antenna cannot be taken to an infinite fractal shape. For example, very often in the design of fractal wire antennas, physical limitations associated with finite conductor diameter and linear segment length limit the number of fractal iterations that can be implemented. Here, we briefly discuss the performance properties of several small and multiband fractal antennas.

What is a fractal antenna? A fractal antenna is one where the antenna geometry is created through the successive iteration of a generator shape to a simple basis shape. For the Koch fractal monopole antenna depicted in Figure 10.41a, the basis geometry is any straight-wire segment within the geometry and the generator geometry is created by dividing this straight-wire segment into thirds and then forming a triangle in the center that has linear segment lengths equal to one-third the length of the original straight-wire segment. This basis-generator geometric conversion process is shown in Figure 10.41b, illustrating the transition from a straight-wire segment to the first iteration of the Koch fractal monopole [43].

Characteristics defining the fractal antenna include *self-similarity*, where small regions of the geometry duplicate the overall geometry on a reduced scale. Another characteristic is *self-affinity*, where small regions of the geometry are not identical to the overall geometry but are skewed or distorted and may have differing scale factors. Mathematically, the fractal geometry is often defined by its *fractal dimension*. For a self-similar fractal, the fractal dimension is given by the expression [45]

$$D_S = \frac{\ln(N)}{-\ln(\gamma)} \tag{10.24}$$

where N is the number of copies of the whole object and γ is a scale factor of each copy. The fractal dimension is generally interpreted as a measure of the space-filling properties and complexity of the fractal shape.

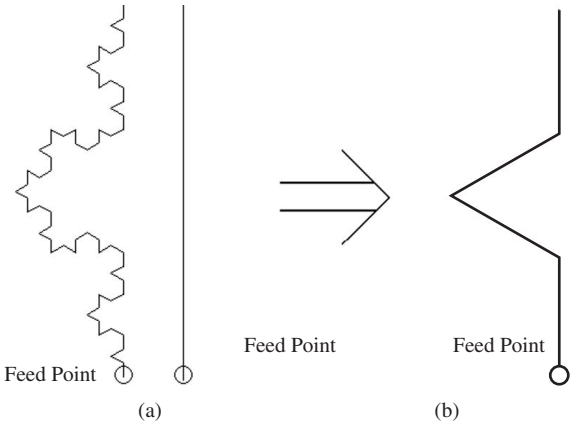


Figure 10.41 (a) Depiction of the Koch fractal monopole. (b) Depiction of the translation between the basis geometry and the generator geometry. This is the creation of the first iteration of the Koch fractal monopole.

10.9.1 Small Fractal Wire Antennas

In previous sections, it has been demonstrated that wire geometry and, specifically, the wire layout have a significant impact on determining the small antenna's resonant frequency. Features like the current vector alignment allow one to easily determine how well one geometry compares relative to another. For example, with the inductively loaded monopole geometries, the normal mode helix generally exhibits the lowest resonant frequency for a given total wire length because it has the most self-inductance. However, once the resonant frequency has been established, antennas of the same basic design methodology exhibit nearly the same resonant performance properties independent of the antenna's total wire length or geometry. In this section, it is demonstrated that small fractals behave in a similar manner and that their resonant performance properties are not uniquely defined by the details of the fractal geometry. Many of the overall results and conclusions in this section were derived from work presented in Refs. 18–19, and 46–50.

10.9.1.1 Hilbert Curve Fractal Dipole The Hilbert curve fractal dipole [51] is a small planar antenna that was presumed to exhibit a lower resonant frequency than any other antenna of the same overall occupied area. The Hilbert curve fractal dipole is shown in Figure 10.42. The arrows in the antenna structure (Figure 10.42b) indicate the current vector alignment. The antenna has a total dipole wire length of 1.2 m compressed into a square area of 7 cm × 7 cm. The antenna has a resonant frequency of 267 MHz. Radiation pattern properties were not studied in Ref. 51.

It was demonstrated in previous sections that achievement of a lower resonant frequency is directly related to increasing the total wire length and the wire layout as viewed from the perspective of the current vector alignment. From a resonant frequency perspective, the objective for a given wire length is to introduce as much self-inductance as possible. Viewing the current vector alignment in the Hilbert curve fractal dipole, it is evident that everywhere the parallel sections of wire closely couple, their current vectors oppose. This indicates that, for a given wire length, the Hilbert geometry is not a suitable candidate for achieving maximum inductance and therefore the lowest resonant frequency.

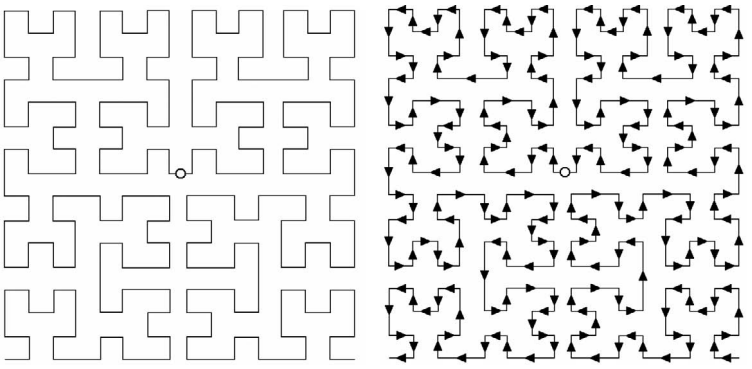


Figure 10.42 (a) Depiction of the Hilbert curve fractal dipole. (b) Depiction of the current vector alignment within the Hilbert curve fractal dipole.

Given the option of any planar geometry to achieve maximum inductance, one would likely wind a spiral-like inductor, as illustrated by the antenna geometry shown in Figure 10.43a [14]. The current vector alignment for this geometry is shown in Figure 10.43b. In this geometry, wherever parallel sections of wire closely couple, their current vectors reinforce, indicating that more self-inductance and a lower resonant frequency can be achieved. The antenna shown in Figure 10.43a exhibits a resonant frequency of 155 MHz. Understanding the relationship between total wire length and resonant frequency, it is obvious that the total wire length in this antenna can be reduced to increase the resonant frequency to 267 MHz. A geometry accomplishing this is depicted in Figure 10.44. The antenna's total wire length is 61.4 cm. This antenna has a resonant resistance of 1.9 Ω, whereas the Hilbert curve fractal antenna has a resonant resistance of approximately 1.2 Ω. In a 50-Ω system, these resistances may be viewed as being essentially identical. From the perspective of achieving the lowest resonant frequency, the Hilbert curve fractal geometry is at a significant disadvantage because of the adverse coupling between closely spaced wire sections.

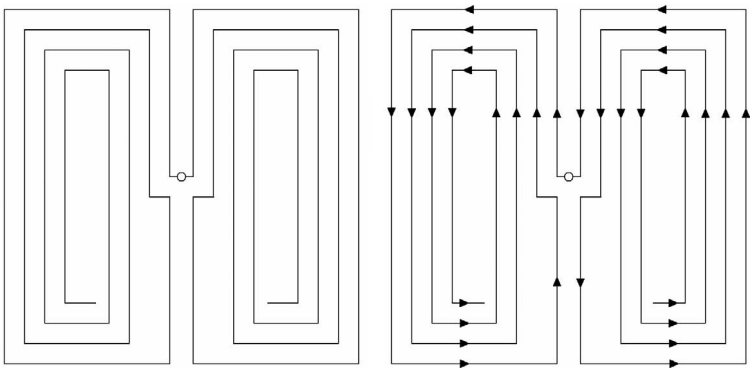


Figure 10.43 (a) Depiction of a spiral wound inductor-like antenna having the same total wire length as the Hilbert curve fractal dipole antenna. (b) Depiction of the current vector alignment within the spiral wound inductor-like antenna.

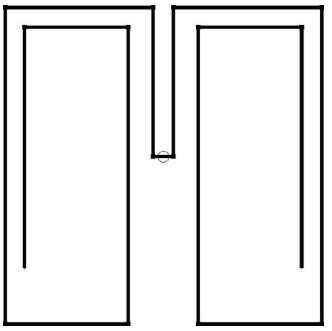


Figure 10.44 Depiction of a spiral wound inductor-like antenna having the same resonant frequency as the Hilbert curve fractal dipole. This antenna has a total wire length of 61.4 cm versus the Hilbert curve's 120 cm.

10.9.1.2 Koch Fractal Monopole The Koch fractal monopole antenna [43] was introduced as a small antenna design that might offer performance properties in a small size that were previously unattainable by other small antenna designs. However, like much of the work on fractal antennas, the Koch fractal monopole was not compared against traditional designs like the meander line monopole and the normal mode helix. With increasing fractal iteration (see Figure 10.41) there is an increase in total wire length within the structure and it would be expected that there would be a corresponding decrease in resonant frequency, much like that which occurs with the normal mode helix and meander line monopoles. To understand the advantages or disadvantages of the Koch fractal geometry in designing monopole antennas, it is necessary to compare its performance to that of other geometries as a function of total wire length and resonant frequency [18, 19, 48].

Here, we compare the Koch fractal monopole to a normal mode helix and meander line monopole—first, having the same total wire length, and then second, having the same resonant frequency. The three antennas are shown in Figure 10.45. The Koch fractal monopole shown in Figure 10.45a is the third fractal iteration. The monopole height of these antennas is 6 cm. All of the antennas have the same conductor diameter.

The first comparison made between the antennas is a comparison of their resonant frequencies as a function of total wire length. While the antenna height is kept fixed at 6 cm, the total wire length is varied from 6 to 25.3 cm. The total wire length used in this comparison is dictated by the fractal iteration. With increasing fractal iteration, the amount of wire length in the structure is constrained by the details of the fractal geometry. This is one disadvantage of using fractal geometry in the design of a wire antenna. The total wire lengths used in this comparison are 6, 8, 10.67, 14.22, 18.96, and 25.3 cm, corresponding to the first through the fifth fractal iteration. The total wire length

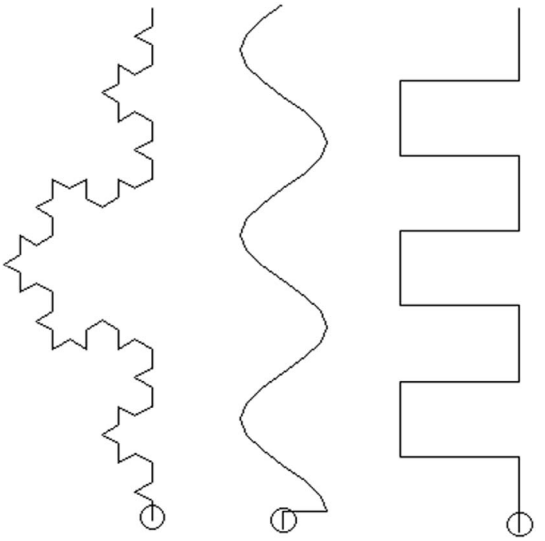


Figure 10.45 (a) Depiction of the third iteration of the Koch fractal monopole. (b) Depiction of the normal mode helix. (c) Depiction of the meander line monopole. All of the antennas have the same overall height and the same conductor diameter.

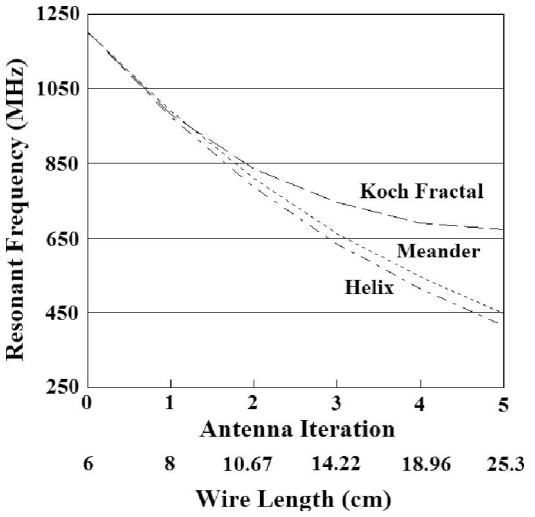


Figure 10.46 A comparison of the resonant frequencies of the Koch fractal monopole, the normal mode helix, and the meander line monopole antenna, where all have the same height, total wire length, and conductor diameter.

in the normal mode helix is adjusted by simply changing the number of turns. The total wire length in the meander line monopole antenna is adjusted by increasing the number of meander sections. A comparison of resonant frequency versus fractal iteration or total wire length is presented in Figure 10.46. It is immediately evident that the normal mode exhibits the lowest resonant frequency for a given wire length. For any given wire length it is the best inductor. If the current vector alignment on the Koch fractal monopole is considered, it becomes evident that this fractal geometry, much like the Hilbert curve, is not an effective inductor.

The next comparison made is that where all three of the antennas are made to have the same resonant frequency. This is accomplished by adjusting the total wire length in the normal mode helix and meander line monopole until they are resonant at the same frequency as the third iteration of the Koch fractal monopole. Both of these antennas have less total wire length than the Koch fractal when they are made to be resonant at the same frequency. In this case the resonant frequency is 745.3 MHz. A comparison of the three antennas' resonant properties is presented in Table 10.6. As evident from the results presented in Table 10.6, when all of these antennas are made to be resonant at the same frequency, they essentially exhibit the same resonant performance characteristics. There is no unique advantage to using the Koch fractal geometry in the design of a small wire antenna.

10.9.1.3 Fractal Tree Dipole The final fractal antenna considered in this section is the fractal tree antenna [52], a dipole version of which is shown in Figure 10.47a (this is the fourth iteration of a four-branch fractal tree). The fractal tree antenna is created much like any fractal antenna, through a successive iteration of applying a geometric generator to a basis geometry, as illustrated in Figure 10.47b, which shows the first iteration of a four-branch fractal tree dipole. Looking at the geometry of the fractal tree shown in Figure 10.47a, it is immediately evident that with increasing fractal iteration, the antenna

TABLE 10.6 Performance Comparison of the Resonant Performance Properties of the Koch Fractal Monopole, the Normal Mode Helix and the Meander Line Monopole When All Are Made to Be Resonant at 745.3 MHz

Antenna	Overall Height (cm)	Total Wire Length (cm)	Resonant Resistance (Ω)	2:1 VSWR Bandwidth (%)
Koch fractal monopole	6	14.22	15.4	3.5
Meander line monopole	6	12.3	15.6	3.5
Normal mode helix	6	11.5	15.8	3.5

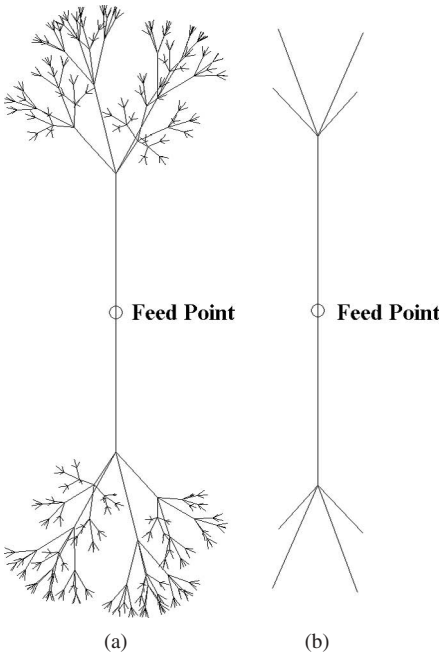


Figure 10.47 (a) The fourth iteration of a four-branch fractal tree dipole antenna. (b) The first iteration of the four-branch fractal tree dipole antenna.

structure becomes increasingly more complex and will eventually reach a point where the physical implementation of the antenna is limited by the constraints imposed by the finite wire diameter.

The fractal tree antenna shown in Figure 10.47 has the appearance of being much like an end-loaded dipole, similar to the capacitive-loaded monopoles described in Section 10.6.3. This being the case, the performance properties of the fractal tree antenna are compared against other top-loaded antennas to see if there are unique advantages to using the fractal tree geometry [53]. The antennas chosen for comparison to the fractal tree antennas are the spiral top-loaded dipole, the helix top-loaded dipole, and the wire-grid top-loaded dipole, all depicted in Figure 10.48. These antennas are designed such that they have the same height and are made to be resonant at the same

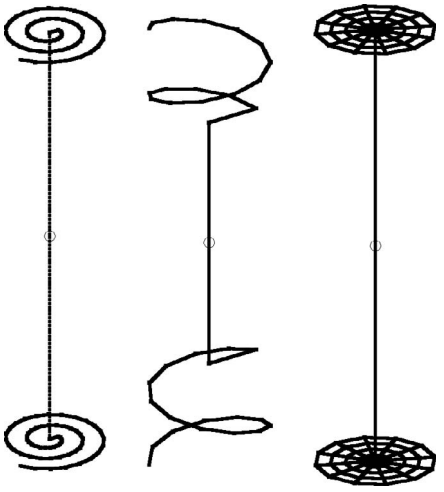


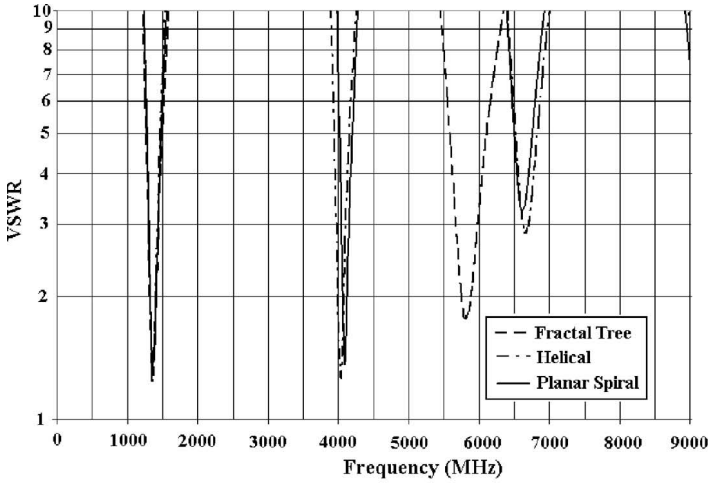
Figure 10.48 (a) The planar spiral top-loaded dipole antenna, (b) the helical top-loaded dipole antenna, and (c) the wire-grid top-loaded dipole antenna.

frequency as the fractal tree antenna they are compared against. The relative comparison is made to determine if the fractal tree dipole exhibits better resonant performance and better multiband performance.

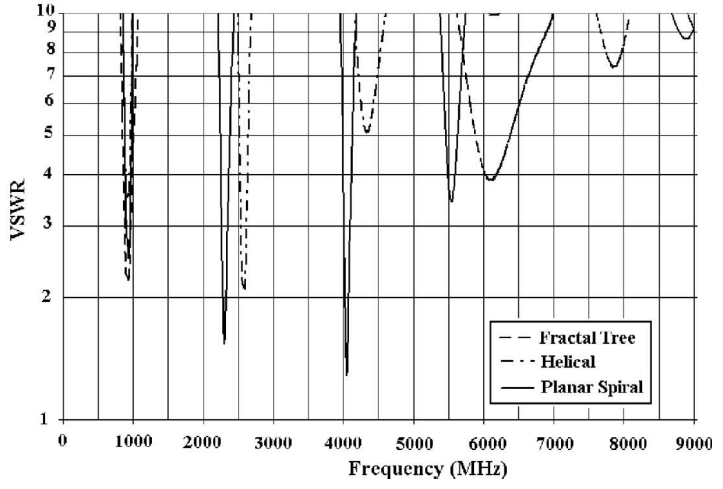
Resonant performance comparisons are demonstrated using the antenna's VSWR properties over a range of frequencies where the antennas are resonant and exhibit low VSWR in a 50-Ω system. The fractal trees used in the comparison are the first iteration of the four-branch fractal tree (Figure 10.47b), the third iteration of the four-branch fractal tree (not shown), and the fourth iteration of the four-branch fractal tree (Figure 10.47a). VSWR comparisons of the planar spiral and helix top-loaded antennas are made against the first iteration of the four-branch fractal tree and the third iteration of the four-branch fractal tree and are shown in Figure 10.49. A VSWR comparison of the wire-grid top-loaded antenna is made against the fourth iteration of the four-branch fractal tree at the lowest resonant frequency and is shown in Figure 10.50. Examining the multiband and resonant VSWR properties of these antennas, it is evident that there is no performance advantage to using the fractal tree geometry, as the antennas behave in a manner consistent with other top-loaded antennas. The complexity of the fractal tree construction represents a significant disadvantage.

10.9.2 Multiband Fractal Antennas

Consideration of multiband fractal antennas is somewhat outside the scope of the previous discussions on small antennas. However, it is a necessary discussion to conclude the section on fractal antennas. There has been a substantial level of work performed in the area of multiband fractal antenna design and a detailed discussion of the topic is beyond the scope of this section. However, it is important to understand and form an initial perspective on whether or not the details of multiband fractal antenna geometry are the significant factors in establishing the multiband behavior of the antenna. This discussion



(a)



(b)

Figure 10.49 (a) VSWR comparison of the planar spiral and helical top-loaded antennas and the first iteration of the four branch fractal tree antenna. (b) VSWR comparison of the planar spiral and helical top-loaded antennas and the third iteration of the four branch fractal tree antenna.

focuses on one of the most recognizable multiband fractal antennas, the Sierpinski gasket [44, 54]. Here, details of the antenn'a fractal geometry are examined to determine whether or not they significantly contribute to the antenna's multiband behavior [55, 56].

The first and fourth iterations of the Sierpinski gasket monopole antenna are depicted in Figure 10.51. These antennas have an overall height of 15.24 cm. The measured multiband return loss of the fourth iteration of this antenna over a frequency range of 50 MHz to 12 GHz is presented in Figure 10.52. The details of the multiband behavior of this antenna (the frequencies of corresponding minimum VSWR) evolve with successive fractal iteration. This implies that specific details of the antenna's geometry directly

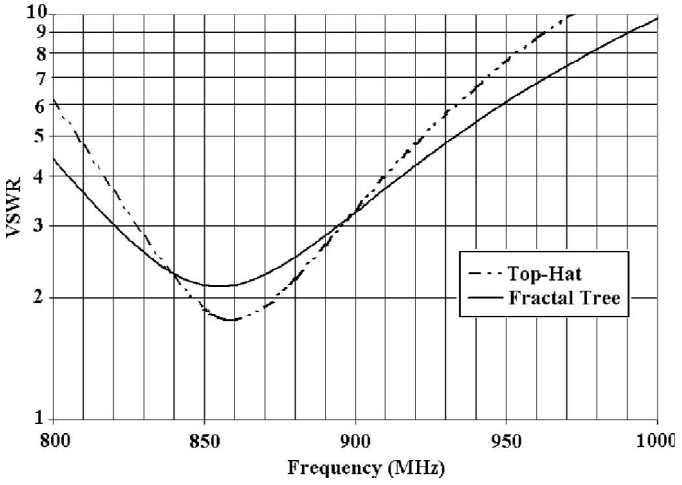


Figure 10.50 VSWR comparison of the wire-grid top-loaded antenna and the fourth iteration of the four branch fractal tree. This comparison is made at the lowest resonant frequency of the fractal tree antenna.

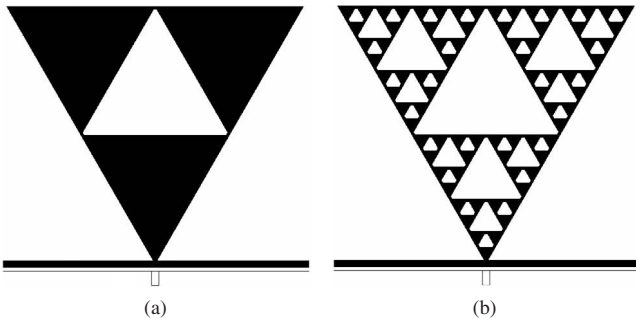


Figure 10.51 (a) The first iteration and (b) the fourth iteration of the Sierpinski gasket monopole antenna.

contribute to its multiband behavior. The question to consider is whether all or some portion of the antenna's fractal geometry is responsible for this multiband behavior. This question is easily answered by removing details in the fractal geometry [56].

If the details of the fractal geometry are responsible for establishing its multiband behavior, then removing specific details or all of the details of the fractal geometry will significantly alter the antenna's multiband behavior. In fact, all of the details of the Sierpinski gasket's fractal geometry can be removed, as shown in Figure 10.53 and the multiband behavior of the antenna is not significantly affected, as illustrated by the measured return loss presented in Figure 10.54. This demonstrates that the details of the fractal geometry are not responsible for its multiband behavior. The details of the Sierpinski gasket's geometry that significantly contribute to its multiband behavior are

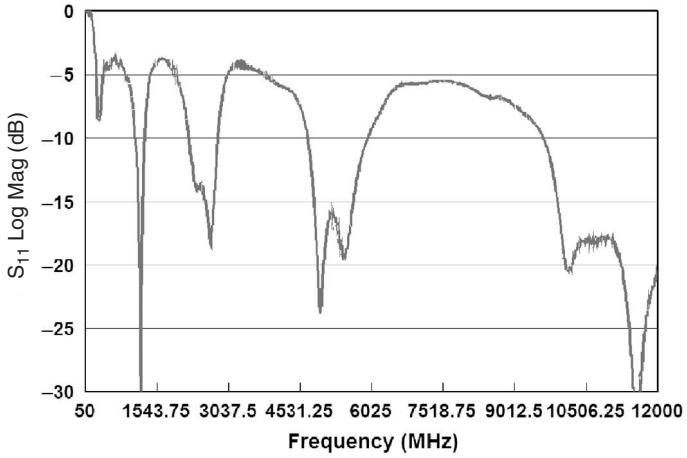


Figure 10.52 Measured return loss of the fourth iteration of the Sierpinski gasket over a frequency range of 50 MHz to 12 GHz.

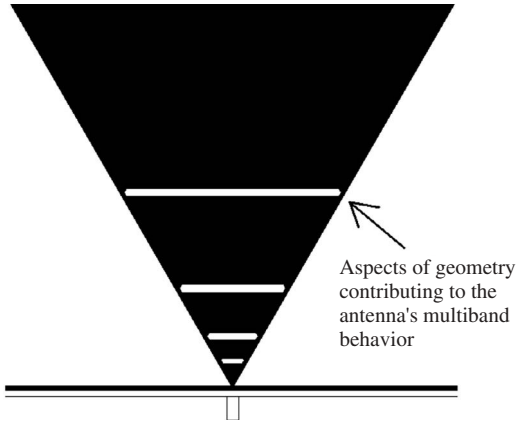


Figure 10.53 The fourth iteration of the Sierpinski gasket monopole antenna with the major aspects of its fractal geometry removed from the antenna structure.

the narrow sections of conductor created by the lower sections of the central triangles, as pointed out in Figure 10.53. At the specific operating bands, there is a high concentration of current at these points within the antenna geometry. This concentration of current is the primary factor in establishing the antenna multiband behavior.

This being the case for the Sierpinski gasket, a very well known multiband fractal antenna, it raises the question for all multiband fractal antennas: What aspects of the antenna's geometry are responsible for its multiband behavior? To determine the significance of the fractal geometry in each case, similar exercises in removing or altering the details of the fractal geometry must be performed.

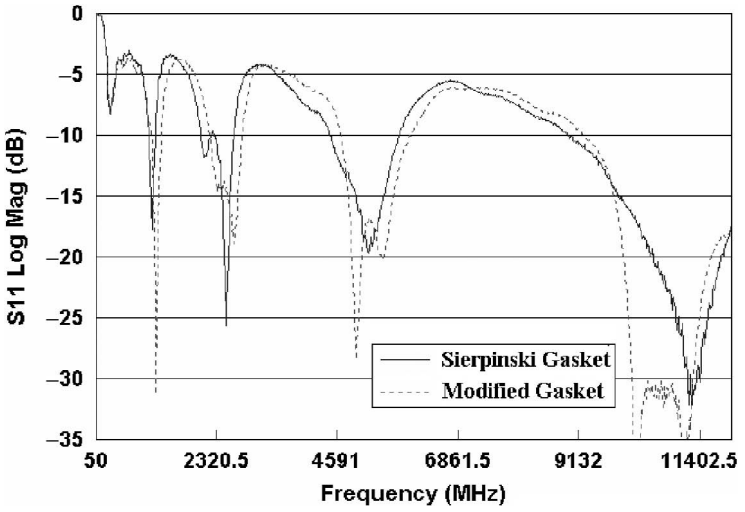


Figure 10.54 Measured return loss of the modified Sierpinski gasket over a frequency range of 50 MHz to 12 GHz.

REFERENCES

1. L. J. Chu, Physical limitations of omni-directional antennas, *J. Appl. Phys.*, Vol. 10, pp. 1163–1175, December 1948.
2. A. D. Yaghjian and S. R. Best, Impedance, bandwidth and Q of antennas, *IEEE Trans. Antennas Propag.*, Vol. 53, No. 4, pp. 1298–1324, April 2005.
3. S. R. Best, The inverse relationship between quality factor and bandwidth in multiple resonant antennas, in *2006 IEEE APS International Symposium*, Albuquerque, NM, July 2006.
4. H. R. Stuart, S. R. Best, and A. D. Yaghjian, Limitations in relating quality factor to bandwidth in a double resonance small antenna, *IEEE AWPL*, Vol. 6, 2007.
5. D. M. Kerns, *Plane-Wave Scattering-Matrix Theory of Antennas and Antenna–Antenna Interactions*, U.S. Government Printing Office, Washington DC; 1981.
6. H. A. Wheeler, Fundamental limitations of small antennas, *Proc. IRE*, Vol. 35, pp. 1479–1484, December 1947.
7. H. A. Wheeler, Small antennas, in *Antenna Engineering Handbook*, 3rd ed., R. C. Johnson (Ed.), McGraw-Hill, New York, 1993.
8. H. A. Wheeler, Small antennas, *IEEE Trans. Antennas Propag.*, Vol. 23, pp. 462–469, July 1975.
9. J. S. McLean, A re-examination of the fundamental limits on the radiation Q of electrically small antennas, *IEEE Trans. Antennas Propag.*, Vol. 44, pp. 672–676, May 1996.
10. C. A. Balanis, *Antenna Theory: Analysis and Design*, 3rd ed., John Wiley & Sons, Hoboken, NJ, 2005.
11. S. R. Best and A. D. Yaghjian, The lower bounds on Q for lossy electric and magnetic dipole antennas, *IEEE AWPL*, Vol. 3, pp. 314–316, December 2004.
12. E. W. Seeley, An experimental study of the disk loaded folded monopole, *IRE Trans. Antennas Propag.*, pp 27–28, January 1956.
13. S. R. Best, A discussion on the performance properties of electrically small self- resonant wire antennas, *IEEE Antennas Propag. Mag.*, Vol. 46, No. 6, pp. 9–22, December 2004.

14. S. R. Best, On the significance of current vector alignment in establishing the resonant frequency of small space-filling wire antennas, *IEEE AWPL*, Vol. 2, pp. 201–204, December 2003.
15. EZNEC, Roy Lewallen, www.eznec.com.
16. Microwave Studio, CST, www.cst.de.
17. J. D. Kraus, The helical antenna, *Proc. IRE*, Vol. 37, pp. 263–272, March 1949.
18. S. R. Best, On the resonant properties of the Koch fractal and other bent wire monopole antennas, *IEEE AWPL*, Vol. 1, No. 3, pp. 74–76, 2002.
19. S. R. Best, On the performance properties of the Koch fractal and other bent wire monopoles, *IEEE Trans. Antennas Propag.*, Vol. 51, No. 6, pp. 1292–1300, June 2003.
20. K. Fujimoto et al., *Small Antennas*, Research Studies Press, England John Wiley & Sons, Hoboken, NJ, 1987.
21. H. R. Bhojwani and L. W. Zelby, Spiral top-loaded antenna: characteristics and design, *IEEE Trans. Antennas Propag.*, Vol. 21, No. 3, pp. 293–298, May 1973.
22. S. R. Best, Advances in the design of electrically small antennas, Short Course, *IEEE AP Symposium*, 2003.
23. E. E. Altshuler, An electrically small genetic antenna immersed in a dielectric, in *IEEE AP Symposium*, Honolulu, HI, June 2007.
24. S. R. Best, A discussion on the quality factor of impedance matched electrically small antennas, *IEEE Trans. Antennas Propagat.*, Vol. 53, No. 12, pp. 502–508, January 2005.
25. R. King, C. W. Harrison, and D. H. Denton, Jr., Transmission-line missile antennas, *IRE Trans. Antennas Propag.*, Vol. 8, pp. 88–90, January 1960.
26. R. Guertler, Impedance transformation in folded dipoles, *Proc. IRE*, Vol. 38, pp. 1042–1047, September 1950.
27. S. R. Best, The performance properties of electrically small resonant multiple arm folded wire antennas, *IEEE Antennas Propag. Mag.*, Vol. 7, No. 4, pp. 13–27, August 2005.
28. S. R. Best, The radiation properties of electrically small folded spherical helix antennas, *IEEE Trans. Antennas Propag.*, Vol. 52, No. 4, pp. 953–960, April 2004.
29. S. R. Best, Low Q electrically small linear and elliptical polarized spherical dipole antennas, *IEEE Trans. Antennas Propag.*, Vol. 53, No. 3, pp. 1047–1053, March 2005.
30. S. R. Best, Improving the performance properties of a dipole element closely spaced to a PEC ground plane, *IEEE AWPL*, Vol. 3, December 2004.
31. P. Hallbjörner, Electrically small unbalanced four-arm wire antenna, *IEEE Trans. Antennas Propag.*, Vol. 52, No. 6, pp. 1424–1428, June 2004.
32. S. R. Best, A small broadband disk loaded folded monopole antenna, in *28th Antenna Applications Symposium*, University of Illinois, Monticello, IL, September 2004.
33. S. R. Best, Progress in the design of a small wideband antenna, *30th Antenna Applications Symposium*, University of Illinois, Monticello, IL, September 2006.
34. H. R. Stuart and C. Tran, Subwavelength microwave resonators exhibiting strong coupling to radiation modes, *Appl. Phys. Lett.*, Vol. 87, p. 151108, 2005.
35. H. R. Stuart and C. Tran, Small spherical antennas using arrays of electromagnetically coupled planar elements, *IEEE Antennas Wireless Propag. Lett.*, Vol. 6, pp. 7–10, 2007.
36. G. Goubau, Multi-element monopole antenna, in *Proceedings of the ECOM-ARO Workshop on Electrically Small Antennas*, Fort Monmouth, NJ, pp. 63–67, October 1976.
37. G. Goubau, N. N. Puri, and F. Schwering, Diakoptic theory for multielement antennas, *IEEE Trans. Antennas Propag.*, Vol. 30, No. 1, pp. 15–26, January 1982.
38. C. H. Friedman, Wide-band matching of a small disk-loaded monopole, *IEEE Trans. Antennas Propag.*, Vol. 33, No. 10, pp. 1142–1148, October 1985.

39. C.B. Ravipati and C. J. Reddy, Low profile disk and sleeve loaded monopole antennas, in *2005 IEEE APS International Symposium*, Washington DC, July 2005.
40. N. Cohen, Fractal antennas Part 1, *Commun. Q.*, pp. 7–22, Summer 1995.
41. N. Cohen, Fractal antennas Part 2, *Commun. Q.*, pp. 53–65, Summer 1996.
42. N. Cohen, Fractal loops and the small loop approximation, *Commun. Q.*, pp. 77–81, Winter 1996.
43. C. Puente-Baliarda, J. Romeu, and A. Cardama, The Koch monopole: a small fractal antenna, *IEEE Trans. Antennas Propag.*, Vol. 48, pp. 1773–1781, November 2000.
44. C Puente-Baliarda, J. Romeu, R. Pous, and A. Cardama, On the behavior of the Sierpinski multiband fractal antenna, *IEEE Trans. Antennas Propag.*, Vol. 46, pp. 517–524, April 1998.
45. Douglas H. Werner and Raj Mittra (Ed.), *Frontiers in Electromagnetics*, IEEE Press, Piscataway, NJ, 2000.
46. S. R. Best, On the resonant behavior of the small Koch fractal monopole antenna, *Microwave Optical Technol. Lett.*, Vol. 35, No. 4, pp. 311–315, November 20, 2002.
47. S. R. Best, A comparison of the properties of small space filling fractal antennas, *IEEE AWPL*, Vol. 2, pp. 197–200, December 2003.
48. S. R. Best, A discussion on the significance of geometry in determining the resonant behavior of fractal and other non-Euclidean wire antennas, *IEEE Antennas Propag. Mag.*, Vol. 45, No. 3, pp. 9–28, June 2003.
49. S. R. Best and J. D. Morrow, The effectiveness of space filling fractal geometry in lowering resonant frequency, *IEEE AWPL*, Vol. 1, pp. 112–115, December 2002.
50. S. R. Best, A comparison of the performance properties of the Hilbert curve fractal and meander line monopole antennas, *Microwave Optical Technol. Lett.*, Vol. 35, No. 4, pp. 258–262, November 20, 2002.
51. K. J. Vinoy, K. A. Jose, and V. K. Varadan, Hilbert curve fractal antenna: a small resonant antenna for VHF/UHF applications, *Microwave Optical Technol. Lett.*, Vol. 29, No. 4, May 20, 2001.
52. J. S. Petko and D. H. Werner, Miniature reconfigurable three-dimensional fractal tree antennas, *IEEE Trans. Antennas Propag.*, Vol. 52, No. 8, pp. 1945–1956, August 2004.
53. J. M. McGinthy, B. C. Kaanta, and S. R. Best, A comparison of fractal and non-fractal end loaded miniature dipole antennas, in *Proceedings of the 2nd IASTED International Conference on Antennas, Radar and Wave Propagation Conference*, Paper 475–801, Banff, Alberta, Canada, July 2005.
54. C. Puente, J. Romeu, R. Pous, X. Garcia, and F. Benitez, Fractal multi-band antenna based on the Sierpinski gasket, *Electron. Lett.*, Vol. 32, No. 1, pp. 1–2, January 1996.
55. S. R. Best, Operating band comparison of the perturbed Sierpinski and modified Parany gasket antennas, *IEEE AWPL*, Vol. 1, pp. 35–38, December 2002.
56. S. R. Best, On the significance of self-similar fractal geometry in determining the multi-band behavior of the Sierpinski gasket antenna, *IEEE AWPL*, Vol. 1, pp. 22–25, December 2002.

PART III

ARRAYS AND SYNTHESIS METHODS

CHAPTER 11

Arrays and Smart Antennas

GEORGE V. TSOULOS and CHRISTOS G. CHRISTODOULOU

11.1 INTRODUCTION

Over the last decade the demand for service provision by wireless communications has risen beyond all expectations. As a result, new improved systems emerged in order to cope with this situation. Global system mobile, (GSM) evolved to general packet radio service (GPRS) and enhanced data rates for GSM evolution (EDGE) and "narrowband" CDMA to wideband code division multiple access (CDMA). Nevertheless, each new system now faces different challenges:

- GPRS consumes GSM user capacity as slots are used to support higher bit rates.
- EDGE faces a similar challenge with GPRS; plus it requires higher SINR to support higher coding schemes (i.e., it also has range problems).
- WCDMA performance depends on interference and hence coverage and capacity are interrelated. High capacity leads to lower range and vice versa.

In order to cope with such system-specific problems but also provide the means for a more general solution to the spectral efficiency problem, spatial filtering has emerged as a promising idea. Spatial filtering can be achieved through adaptive or smart antennas and as a result the area has seriously attracted the interest of both academia and industry over the last decade.

This chapter presents an overview of smart antenna systems for wireless communications. First, it provides the reader with the necessary background on antenna arrays. Then it gives a brief overview for the different options in terms of modeling the propagation characteristics of the radio channel, with particular emphasis on the spatial dimension. These topics are followed by a description of the different smart antenna methods and their characteristics, along with a description of the key algorithms that give "intelligence" to the system. The potential advantages that can be achieved with the different smart antenna methods, along with a discussion of some implementation and system issues that need to be considered when a smart antenna is integrated in a specific wireless system, are then discussed. Finally, the chapter closes with a brief discussion on smart antennas at both ends of the radio link, that is, MIMO (multiple-input multiple-output) systems.

Modern Antenna Handbook. Edited by Constantine A. Balanis

11.2 ANTENNA ARRAYS

Multiple antennas can be arranged in space, in various geometrical configurations, to yield highly directive patterns [1–5]. These antenna configurations are called "arrays." In an array antenna, the fields from the individual elements add constructively in some directions and destructively (cancel) in others. For purposes of analysis, arrays are assumed to consist of identical elements, although it is possible to create an array with elements such that each has a different radiation pattern.

The major advantage of antenna arrays over a single antenna element is their electronic scanning capability; that is, the major lobe can be steered toward any direction by changing the phase of the excitation current at each array element (phased array antennas). Furthermore, by also controlling the magnitude of the excitation current, a large variety of radiation patterns and sidelobe level characteristics can be produced. Adaptive antennas (also called "smart antennas" in mobile communication applications) go a step further than phased arrays and can direct their main lobe (with increased gain) in a desired direction (e.g., a mobile user in a cellular communication system) and nulls in the directions of interference or jammers.

There are five main parameters that affect the overall performance of an antenna array:

1. Geometry (e.g., linear, circular, or planar arrangement of the radiating elements)
2. Distance of separation between adjacent elements
3. Amplitude current excitation of each individual element
4. Phase excitation of each individual element
5. Radiation pattern of each individual element

Figure 11.1 shows an example of a two-dimensional array (planar array), where the individual elements consist of microstrip antennas.

11.2.1 Array Factor

Let us assume first a simple scenario of an antenna array with two infinitesimal horizontal dipoles in free space, positioned as shown in Figure 11.2. The dipole located at (0, 0,

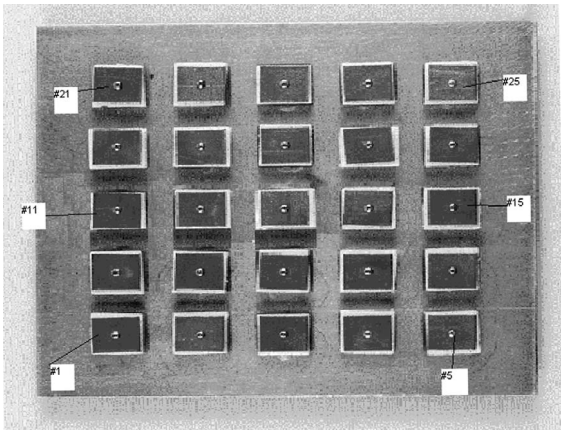


Figure 11.1 Planar array configuration: 5×5 microstrip array.

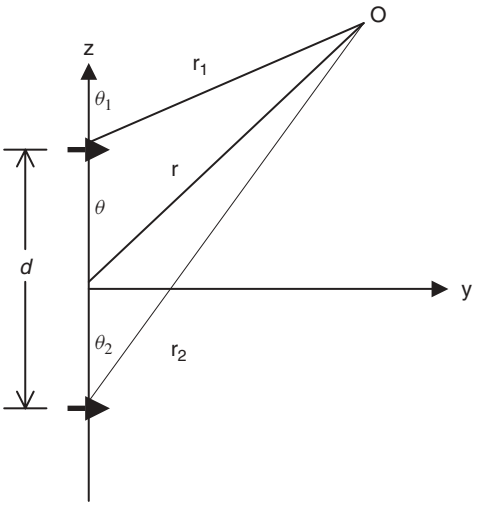


Figure 11.2 Geometry of two-element array separated by a distance d.

$d/2$) carries a current $I_0\angle(\xi/2)$ and the one at (0, 0, $-d/2$) a current $I_0\angle(\xi/2)$ where ξ is the phase difference added to the two dipoles externally through a phase shifter. The total electric field at the observation point O is given as the vectorial summation of the fields due to the two individual antennas [1] (without considering mutual coupling effects):

$$\mathbf{E}_{\text{total}} = \mathbf{E}_1 + \mathbf{E}_2 = \widehat{a}_\theta \frac{j\eta k I_0}{4\pi} l \left(\frac{e^{-j(kr_1 - \xi/2)}}{r_1} |\cos\theta_1| + \frac{e^{-j(kr_2 + \xi/2)}}{r_2} |\cos\theta_2| \right) \tag{11.1}$$

In the far field r_1, r_2, and r are parallel as shown in Figure 11.3, so $\theta_1 \approx \theta_2 \approx \theta$ and $r_1 \approx r_2 \approx r$ for amplitude variations and $r_1 \approx r - (d/2)\cos\theta$ and $r_2 \approx r + (d/2)\cos\theta$

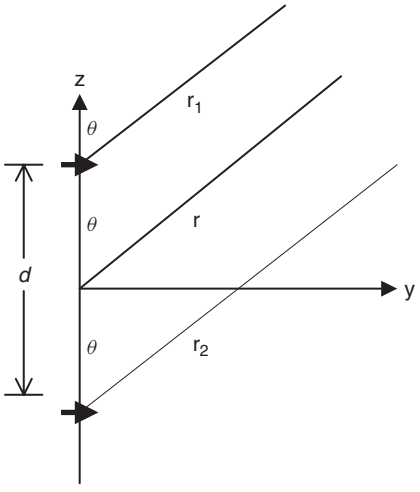


Figure 11.3 Far-field geometry of two dipoles.

for phase variations. Hence Eq. (11.1) can be rewritten

$$\mathbf{E}_{\text{total}} = \mathbf{E}_1 + \mathbf{E}_2 = \hat{a}_\theta \frac{j\eta k I_0}{4\pi r} l e^{-jkr} |\cos\theta| 2\cos\left[\tfrac{1}{2}(kd\cos\theta + \xi)\right] \tag{11.2}$$

Looking at Eqs. (11.1) and (11.2) one can observe that the total field is equal to the field of the single element (*element factor*) located at the origin, multiplied by an *array factor* (AF) given by

$$AF = 2\cos\left[\tfrac{1}{2}(kd\cos\theta + \xi)\right] \tag{11.3}$$

Generally, the far-field pattern of an array is given by the *multiplication pattern* of the single element and the array factor:

$$\textit{Total Pattern} = \textit{Element Factor} \times \textit{Array Factor} \tag{11.4}$$

The array factor is a function of the following parameters:

1. The geometrical arrangement of the radiating elements comprising the array
2. The current excitation of the elements
3. The number of elements
4. The distance of separation d of adjacent elements
5. Frequency (or wavelength) of operation

Example 1 Find the total pattern of two identical horizontal dipoles shown in Figure 11.3, with $d = \lambda/4$ and $\xi = 0$.

Consider the normalized array factor of Eq. (11.3); that is,

$$AF_n = \cos\left[\tfrac{1}{2}(kd\cos\theta + \xi)\right] \tag{11.5}$$

If we substitute d and ξ, the *array factor* becomes $AF_n = \cos[(\pi/4)\cos\theta]$. The final (total) pattern, using the multiplication pattern procedure, becomes $E_{\text{total}} = |\cos\theta| \cos\left[(\pi/4)\cos\theta\right]$. Note that the array factor does not introduce any nulls. The only null that occurs is the one due to the element factor (horizontal dipole) at $\theta = \pi/2$, as shown in Figure 11.4.

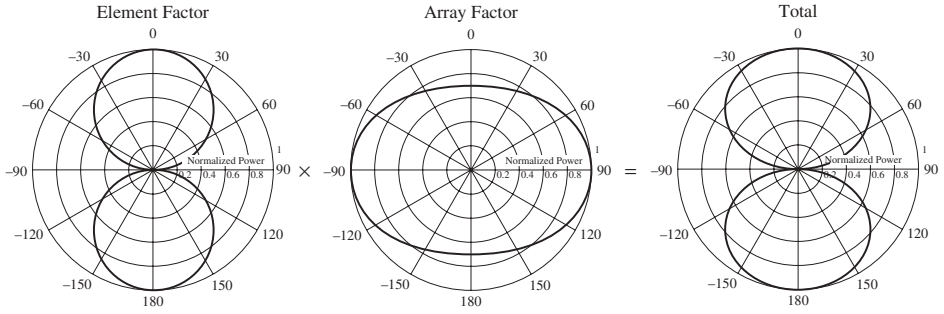


Figure 11.4 Element factor, array factor, and total pattern for a two-element array of infinitesimal horizontal dipoles with $d = \lambda/4$ and $\xi = 0$.

If we change ξ to $\xi = \pi/2$, the array factor changes to $AF_n = \cos\left[(\pi/4)(\cos\theta + 1)\right]$ and the total pattern to $E_{\text{total}} = |\cos\theta|\cos[(\pi/4)(\cos\theta + 1)]$. This change in the externally applied phase introduces an additional null at $\theta = 0^\circ$, as shown in Figure 11.5.

For a phase of $\xi = -\pi/2$, the total pattern is $E_{\text{total}} = |\cos\theta|\cos[(\pi/4)(\cos\theta - 1)]$ and an additional null is introduced at $\theta = 180^\circ$, as shown in Figure 11.6.

11.2.2 Uniform *N*-Element Linear Array

Consider now an N-element array of isotropic radiators shown in Figure 11.7. It is a linear array since all elements of the array are positioned along a single line and it is called uniform because each identical element is fed with a current of the same magnitude but with a progressive phase shift of ξ. The distance of separation between adjacent elements is d.

In this case the array factor can be expressed as the sum of all single-element contributions:

$$AF = 1 + e^{j\psi} + e^{j2\psi} + e^{j3\psi} + \cdots + e^{j(N-1)\psi} \tag{11.6}$$

where $\psi = kd\cos\theta + \xi$. Equation (11.6) is a geometric series that can be expressed in compact form:

$$AF = \frac{\sin(N\psi/2)}{\sin(\psi/2)} \tag{11.7}$$

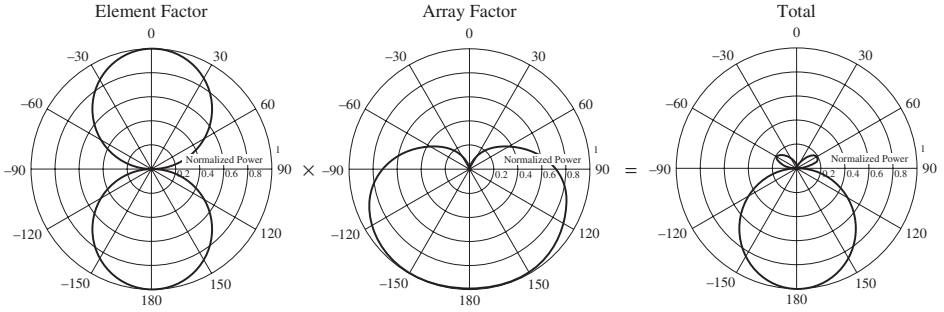


Figure 11.5 Element factor, array factor, and total pattern for a two-element array of infinitesimal horizontal dipoles with $d = \pi/4$ and $\xi = \pi/2$.

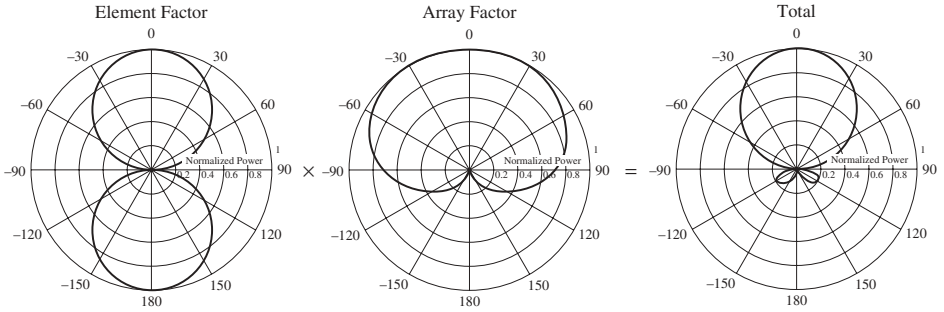


Figure 11.6 Element factor, array factor, and total pattern for a two-element array of infinitesimal horizontal dipoles with $d = \pi/4$ and $\xi = -\pi/2$.

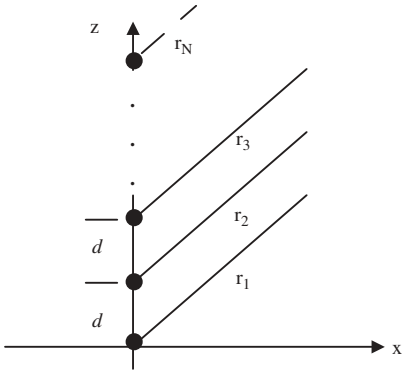


Figure 11.7 Geometrical configuration of N isotropic elements along the z-axis, separated by a distance d and fed with a progressive phase ξ.

Examination of Eq. (11.7) reveals the following points about the array factor of a uniform linear array:

1. The principal maximum (major lobe) occurs when, $\psi = 0$; that is,

$$kd\cos\theta_{\text{major}} + \xi = 0 \quad \text{or} \quad \theta_{\text{major}} = \cos^{-1}\left(-\lambda\xi/2\pi d\right).$$

2. The principal maximum (major lobe) occurs when $\psi/2 = \pm m\pi$; that is,

$$kd\cos\theta_{\text{major}} + \xi = \pm 2m\pi$$

or

$$\theta_{\text{major}} = \cos^{-1}\left[(\lambda/2\pi d)(-\xi \pm 2m\pi)\right], \quad m = 0, 1, 2, \ldots$$

3. The nulls occur when $\sin(N\psi/2) = 0$; that is, $N\psi/2 = \pm n\pi$ for $n = 1, 2, 3, \ldots$ and $n \neq N, 2N, \ldots$.

Figure 11.8 shows an example of a linear array of eight microstrip antennas mounted on a ground plane.

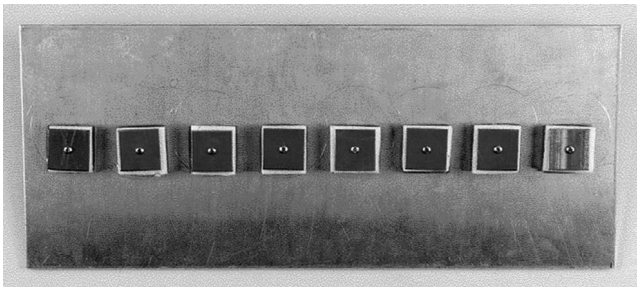


Figure 11.8 A linear array antenna comprised of eight identical microstrip elements.

If the aim is to steer the main beam at $\theta_{\text{major}} = 90^\circ$ (broadside array), the progressive phase shift should be equal to zero, provided that $d \neq n\lambda$ for $n = 1, 2, 3, \ldots$. If we want the major lobe to appear at $\theta_{\text{major}} = 0^\circ$ or $\theta_{\text{major}} = 180^\circ$ (end fire array), then (1) for $\theta_{\text{major}} = 0^\circ$, the progressive phase shift should be $\xi = -kd$; and (2) for $\theta_{\text{major}} = 180^\circ$, $\xi = kd$.

Example 2 To illustrate the procedure for designing uniform array antennas and to introduce the concept of "grating lobes," a 10-element uniform array with $\xi = 0$ and two different distances of separation d ($d = \pi/4$ and $d = \lambda$) are considered. Since $\xi = 0$ we expect the major lobe to appear at $\theta = 90^\circ$. However, for $d \geq \lambda$, two more maxima appear at $\theta = 0^\circ$ and $\theta = 180^\circ$ (Figure 11.9). These two extra unexpected lobes are referred to as "grating lobes." These lobes are undesired lobes that occur due to the constructive interference of the individual element fields. To avoid grating lobes d should always be less than λ.

11.2.3 Planar Arrays

Unlike linear arrays that can only scan the main beam in one polar plane (ϕ or θ), planar arrays scan the main beam along both ϕ and θ. Planar arrays offer more gain and lower sidelobes than linear arrays, at the expense of using more elements. The design principles for planar arrays are similar to those presented earlier for the linear arrays. Since the elements are placed in two dimensions (see Figure 11.10), the array factor of a planar array can be expressed as the multiplication of the array factors of two linear arrays: one along the x-axis and one along the y-axis [1]:

$$AF_{\text{planar}} = (AF_x) \cdot (AF_y) \tag{11.8}$$

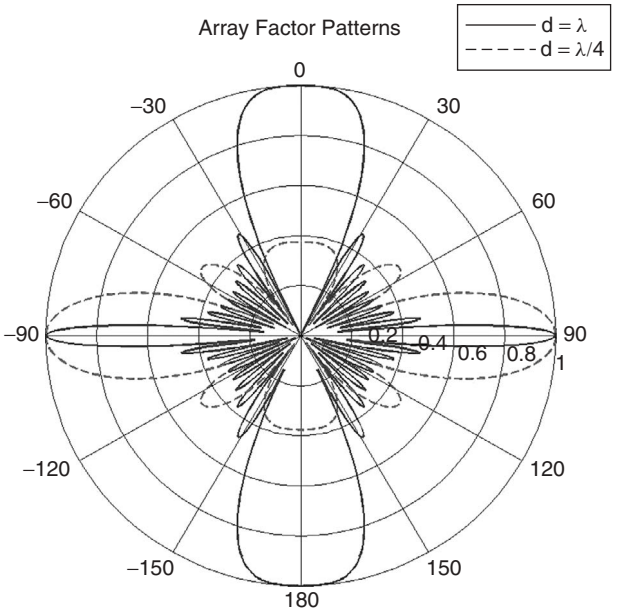


Figure 11.9 Array factor patterns for a 10-element, uniform broadside array with $\xi = 0$.

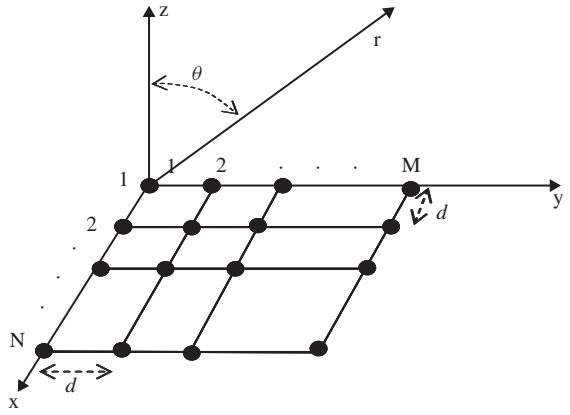


Figure 11.10 Geometry of a rectangular $M \times N$ planar array with interelement distance d.

or

$$AF = \left(\frac{\sin(N\psi_x/2)}{N\sin(\psi_x/2)}\right)\left(\frac{\sin(M\psi_y/2)}{M\sin(\psi_y/2)}\right) \tag{11.9}$$

where

$$\psi_x = kd_x \sin\theta\cos\phi + \xi_x$$
$$\psi_y = kd_y \sin\theta\sin\phi + \xi_y$$

Figure 11.11 shows an example for the radiation pattern produced by an 11×11 planar array with $\lambda/2$ spacing and the major lobe steered towards $(20^\circ, 30^\circ)$.

By using a progressive phase shift applied to the elements of the array we can scan the main beam in certain angles (phased array), but it is also possible to use different

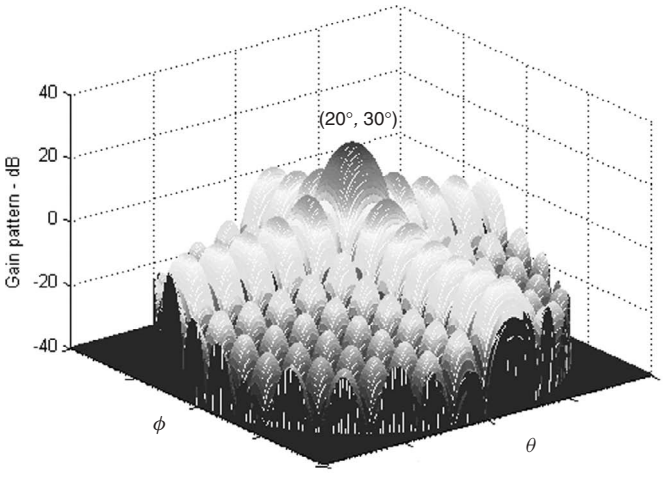


Figure 11.11 Example of radiation pattern from an 11×11 planar array.

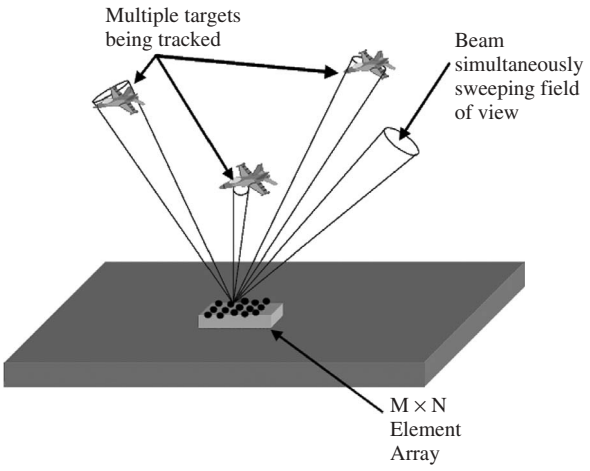


Figure 11.12 Planar $M \times N$ array scanning and tracking multiple targets simultaneously.

excitation schemes for the individual elements to track multiple sources or targets along θ and ϕ, as shown in Figure 11.12.

11.2.4 Circular Arrays

In a circular array configuration, the elements are placed in a circle, as shown in Figure 11.13. The array factor in this case is given by [1]

$$AF(\theta, \phi) = \sum_{n=1}^{N} I_n e^{jka[\sin\theta\cos(\phi-\phi_n)-\sin\theta_0\cos(\phi_0-\phi_n)]} \tag{11.10}$$

where ϕ_0 and θ_0 are the angles of the main beam, a is the radius of the circular array, and I_n are the excitation coefficients of the elements.

Figure 11.14 shows an example radiation pattern produced with a 15-element circular array, with $\lambda/2$ spacing.

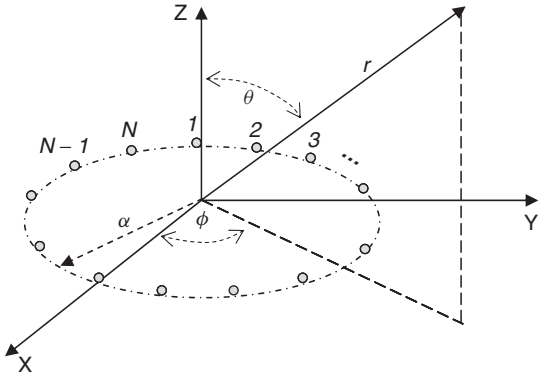


Figure 11.13 Geometry of an N-element circular array.

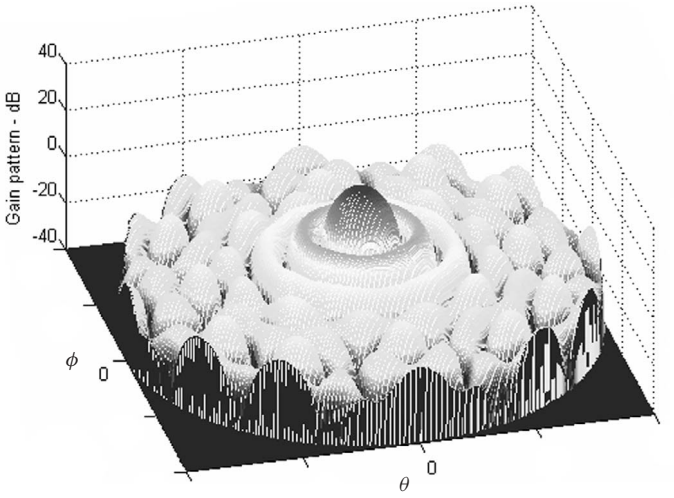


Figure 11.14 Example of radiation pattern from a 15-element circular array with main lobe at $(0^\circ, 0^\circ)$.

11.2.5 Nonuniform Arrays

Nonuniform arrays are arrays whose elements can have variable excitations. There are several ways of exciting the elements of an array to achieve certain goals. For example, one can use the Schelkunoff zero placement method [1, 3, 5] to place nulls in certain directions in the radiation pattern. Another well known technique is the binomial array approach, which one can use to achieve zero sidelobe levels when the interelement spacing is $\lambda/2$. However, this will also yield a larger beamwidth. Finally, the one technique that offers a compromise between the uniform and binomial arrays is the Dolph–Chebyshev array. Here we describe both the binomial and Chebyshev design methods.

11.2.5.1 Binomial Arrays Consider the array factor for uniform arrays given by Eq. (11.6):

$$AF = 1 + e^{j\psi} + e^{j2\psi} + e^{j3\psi} + \cdots + e^{j(N-1)\psi}$$

In the case of the binomial array with N elements, one can show [3, 5] that the array factor with nonuniform excitation coefficients can be expressed as

$$AF = (1 + e^{j\psi})^{N-1} = 1 + (N-1)e^{i\psi} + \frac{(N-1)(N-2)}{2!}e^{j2\psi} + \frac{(N-1)(N-2)(N-3)}{3!}e^{j3\psi} + \cdots \tag{11.11}$$

From Eq. (11.11), it is evident that the current coefficients of the resulting N-element array take the form of a binomial series and that is why this array is known as a *binomial array*. Table 11.1 shows some of the excitation coefficients for various N values.

TABLE 11.1 Pascal's Triangle and the Binomial Expansion Coefficients

1	$N = 1$
1 1	$N = 2$
1 2 1	$N = 3$
1 3 3 1	$N = 4$
1 4 6 4 1	$N = 5$
1 5 10 10 5 1	$N = 6$
1 6 15 20 15 6 1	$N = 7$
1 7 21 35 35 21 7 1	$N = 8$
1 8 28 56 70 56 28 8 1	$N = 9$
1 9 36 84 126 126 84 36 9 1	$N = 10$

Advantages

- The binomial array has very low sidelobes.
- You can actually get the sidelobes equal to zero if $d = \lambda/2$.

Disadvantages

- They exhibit larger beamwidth compared to the uniform array.
- There are wide variations between the amplitudes of the different elements.
- There are low efficiencies.
- It is difficult to implement in practice since the electronics required to create the wide variation in magnitudes becomes complex and costly.

11.2.5.2 Dolph–Chebyshev Method It is a compromise between the *uniform* and *binomial* arrays. The idea is that for a given sidelobe level you can design an array with the narrowest beam, or vice versa.

In this case the array factor of P elements can be expressed in terms of cosine terms [1, 3, 5]:

$$(AF)_P = \sum_{n=1}^{M} a_n \cos[(2n-1)u], \quad \text{for } P = 2M \text{ (even number)} \tag{11.12}$$

where $u = (\pi d/\lambda)\cos\theta$

$$(AF)_P = \sum_{n=1}^{M+1} a_n \cos[2(n-1)u], \quad \text{for } P = 2M+1 \text{ (odd number)} \tag{11.13}$$

Table 11.2 shows the relationship between the various Chebyshev polynomials and the cosine terms that appears in Eqs. (11.12) and (11.13).

Design Procedure For a given sidelobe ratio R_0 (in dB) and the number of elements find the excitation coefficients.

Step 1: Replace each $\cos(mu)$ term in the array factor by its expansion in terms of powers of $\cos(u)$.

TABLE 11.2 Chebyshev Polynomials in Terms of Cosine Terms

$\cos 0u = 1$	$\cos 0u = 1 = T_0(x)$
$\cos 1u = \cos u$	$\cos 1u = x = T_1(x)$
$\cos 2u = 2\cos^2 - 1$	$\cos 2u = 2x^2 - 1 = T_2(x)$
$\cos 3u = 4\cos^3 u - 3\cos u$	$\cos 3u = 4x^3 - 3x = T_3(x)$
$\cos 4u = 8\cos^4 u - 8\cos^2 u + 1$	$\cos 4u = 8x^4 - 8x^2 + 1 = T_4(x)$
$\cos 5u = 16\cos^5 u - 20\cos^3 u + 5\cos u$	$\cos 5u = 16x^5 - 20x^3 + 5x = T_5(x)$
$\cos 6u = 32\cos^6 u - 48\cos^4 u + 18\cos^2 u - 1$	$\cos 6u = 32x^6 - 48x^4 + 18x^2 - 1 = T_6(x)$
$\cos 7u = 64\cos^7 u - 112\cos^5 u + 56\cos^3 u - 7\cos u$	$\cos 7u = 64x^7 - 112x^5 + 56x^3 - 7x = T_7(x)$
$\cos 8u = 128\cos^8 u - 256\cos^6 u + 160\cos^4 u - 32\cos^2 u + 1$	$\cos 8u = 128x^8 - 256x^6 + 160x^4 - 32x^2 + 1 = T_8(x)$
$\cos 9u = 256\cos^9 u - 576\cos^7 u + 432\cos^5 u - 120\cos^3 u + 9\cos u$	$\cos 9u = 256x^9 - 576x^7 + 432x^5 - 120x^3 + 9x = T_9(x)$

Step 2: Express the sidelobe ratio R_0

$$\frac{\text{Height of major lobe}}{\text{Height of first sidelobe}} = \frac{H_0}{H_1} = R_0 \quad \text{(dB)}$$

as a voltage ratio using

$$R_0\big|_{\text{Voltage}} = 10^{R_0/20} \tag{11.14}$$

Step 3: Use R_0 (voltage ratio) to determine the normalization number z_0 from

$$R_0\big|_{\text{Voltage}} = T_{P-1}(z_0) = \cosh[(P-1)\cosh^{-1}(z_0)] \tag{11.15}$$

which yields

$$z_0 = \cosh\left(\frac{\cosh^{-1} R_0}{P-1}\right) \tag{11.16}$$

Step 4: Substitute $\cos(u) = z/z_0$ in the array factor of Step 1. The idea of using z_0 in the array factor is to normalize the sidelobes to the value of 1.

Step 5: Equate the array factor of Step 5 to the $T_{P-1}(z)$ Chebyshev polynomial. By equating the array factor and the Chebyshev polynomial we can determine the excitation coefficients.

Example Design a five-element Dolph–Chebyshev array that will yield sidelobes 20 dB below the main beam. Let $d = \lambda/2$.

Using Eq. (11.13) we have

$$(AF)_P = \sum_{n=1}^{M+1} a_n \cos[2(n-1)u]$$

with $M = 2$ and $P = 5$.

Step 1: Write the array factor in terms of powers of cos u:

$$\begin{aligned}(AF)_P &= \sum_{n=1}^{M+1} a_n \cos[2(n-1)u] = a_1 \cos 0u + a_2 \cos 2u + a_3 \cos 4u \\ &= a_1 + a_2(2\cos^2 u - 1) + a_3(8\cos^4 u - 8\cos^2 u + 1) \\ &= (a_1 - a_2 + a_3) + (2a_2 - 8a_3)\cos^2 u + 8a_3 \cos^4 u \end{aligned} \tag{11.17}$$

Step 2:

$$20\text{dB} = 20\log_{10} R_0 \Rightarrow \quad R_0\big|_V = 10^{R_0/20} = 10^1 = 10$$

Step 3:

$$z_0 = \cosh\left(\frac{\cosh^{-1} R_0}{P-1}\right) = \cosh\left(\frac{\cosh^{-1} 10}{4}\right) = 1.293$$

Step 4:

$$\begin{aligned}(AF)_5 &= a_1 + a_2\left[2\left(\frac{z}{z_0}\right)^2 - 1\right] + a_3\left[8\left(\frac{z}{z_0}\right)^4 - 8\left(\frac{z}{z_0}\right)^2 + 1\right] \\ &= (a_1 - a_2 + a_3) + (2a_2 - 8a_3)\left(\frac{z}{z_0}\right)^2 + 8a_3\left(\frac{z}{z_0}\right)^4\end{aligned}$$

Step 5:

$$\begin{aligned}(AF)_5 &= (a_1 - a_2 + a_3) + (2a_2 - 8a_3)\left(\frac{z}{z_0}\right)^2 + 8a_3\left(\frac{z}{z_0}\right)^4 \\ &= T_4(x) = 8z^4 - 8z^2 + 1\end{aligned} \tag{11.18}$$

Hence

$$a_1 = 2.698, \quad a_2 = 4.493, \quad \text{and} \quad a_3 = 2.795$$

Directivity The directivity for a large Dolph–Chebyshev array with sidelobes between -20 and -60 dB, and for a direction near broadside, is given by [1]

$$D_0 = \frac{2R_0^2}{1 + (R_0^2 - 1) f \frac{\lambda}{L+d}} \tag{11.19}$$

where L is the length of the array and f is the broadening factor given by [1]

$$f = 1 + 0.636\left\{\frac{2}{R_0}\cosh\left[\sqrt{(\cosh^{-1} R_0) - \pi^2}\right]\right\}^2 \tag{11.20}$$

Figures 11.15 and 11.16 depict the broadening factor and directivity for Dolph–Chebyshev arrays, respectively.

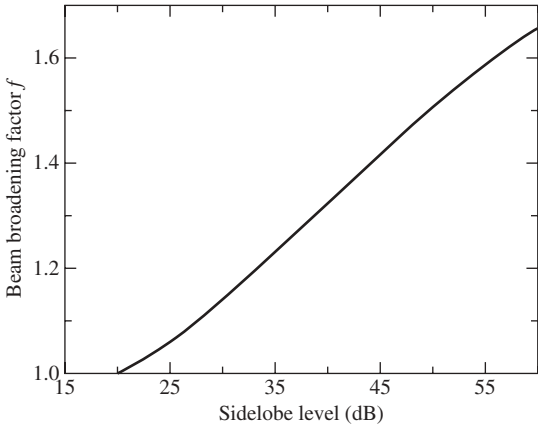


Figure 11.15 Beam broadening factor for Chebyshev arrays. (From R. S. Elliott, Beamwidth and directivity of large scanning arrays, Part 1, *Microwave J*., Vol. 6, pp. 53–60, December 1963.)

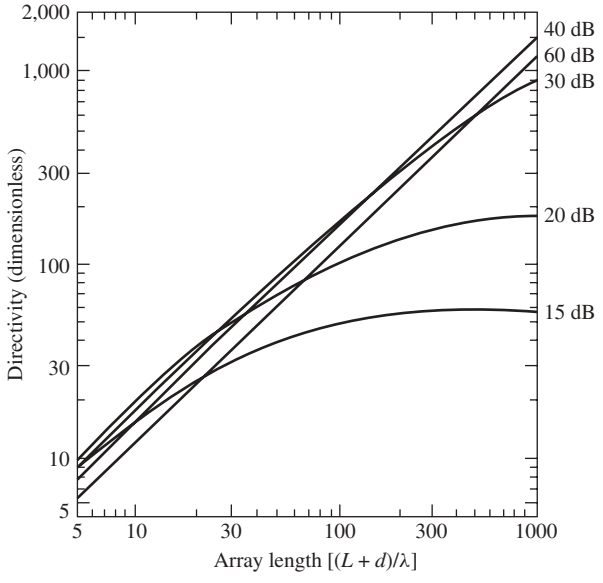


Figure 11.16 Directivity of Chebyshev arrays. (From R. S. Elliott, Beamwidth and directivity of large scanning arrays, Part 1, *Microwave J*., Vol. 6, pp. 53–60, December 1963.)

11.3 PROPAGATION MODELS FOR SMART ANTENNA ANALYSIS

There are a plethora of radio wave propagation models, with varying complexity and accuracy. The right choice depends on system and operational parameters like the specific environment, simulation run time, achieved accuracy, cost, and ease of use. Generally, experience has shown that for nonsite-specific scenarios, sufficient accuracy can be

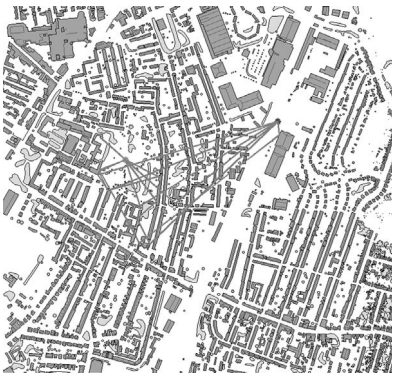
achieved at reasonable simulation times, with stochastic models. For site-specific scenarios, more complex deterministic (ray tracing) models, which employ realistic information for the environment under consideration (geographical databases), are required to provide reasonable accuracy, but at the cost of increased computational complexity and hence simulation times.

Table 11.3 provides a brief overview of the most representative propagation models that include spatial information for smart antenna analysis, along with some key references. For more information, Refs. 6 and 7 and the references therein are cited.

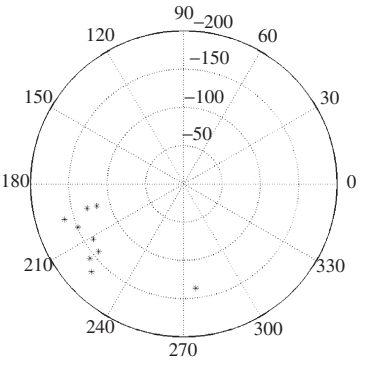
Ray tracing techniques, in particular, although computationally complex, can provide accurate site-specific information for the three-dimensional (3D) impulse response of the radio channel. This modeling approach has evolved from "point-to-point" analysis to "multipoint-to-multipoint" scenarios and hence is a valuable tool for smart antenna (and MIMO) system analysis. An example for the produced results from a ray tracing model is shown in Figure 11.17. Figures 11.17a and 11.17b show the area under study, a typical large cell operational environment, along with the multipath between a base station (BS) and a mobile station (MS) position, in 2D and 3D, respectively. Figures 11.17c and 11.17d show the angle of arrival (in azimuth) versus power for each of the multipath rays, at the BS and MS, respectively. The spatial characteristics of the radio channel differ between the two ends; they are generally more uniform at the MS while they are more directive at the BS. Figures 11.17e and 11.17f show the angle of arrival (in elevation) versus power for each of the multipath rays, at the BS and MS, respectively. Similar behavior between

TABLE 11.3 Spatiotemporal Propagation Models for Smart Antenna Analysis

Model	Brief Description
Ring of scatterers (Lee's model)	Effective scatterers are uniformly spaced on a circular ring around the mobile [8]
Discrete uniform distribution (macrocell)	Scatterers are evenly located within a narrow beamwidth centered around the direction of the mobile [9]
Geometrically based single-bounce (GBSB) circular model (macrocell)	Scatterers are uniformly placed within a disk of radius R, around the mobile [10, 11]
Geometrically based single-bounce elliptical model (microcells)	Same concept with the previous model (GBSB) but now scatterers are uniformly distributed within an ellipse [12]
Gaussian wide sense stationary uncorrelated scattering (GWSSUS) model	Scatterers are grouped in clusters and time delay–angle spread differences within clusters are not resolvable within the signal's bandwidth (narrowband assumptions) [13, 14]
Combined GBSB and GWSSUS	Combination of the geometrically based single bounce and the Gaussian wide sense stationary uncorrelated scattering models, further enhanced to handle nonstationary scenarios [15–17]
Modified Saleh–Valenzuella (indoor)	Extended indoor model based on measurements and the assumption that time and angle are statistically independent [18]
Measurement based models—statistical	Statistical description of propagation characteristics based on measurements [19–21]
Ray tracing	Deterministic models that combine geometrical optics, uniform theory of diffraction, and scattering models with geographical information of the environment [22–24]

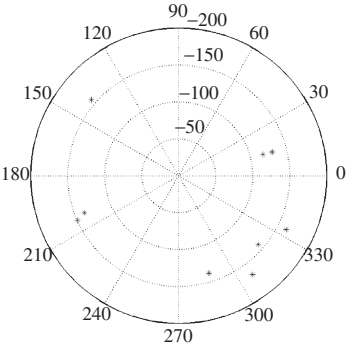


(a)

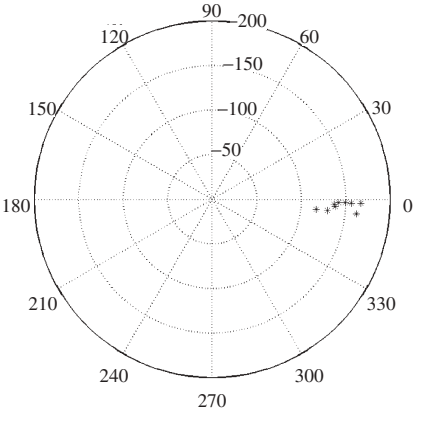


(c)

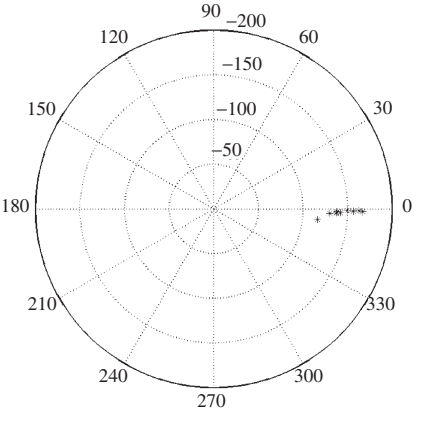
(b)



(d)



(e)



(f)

Figure 11.17 Map of the (a) 2D and (b) 3D environment and multipath example. Example of ray tracing 3D impulse response: angle of arrival—AoA (azimuth)—versus power at the (c) BS and (d) MS. AoA (elevation) versus power at the (e) BS and (f) MS. (g) Time of arrival (ToA) versus power.

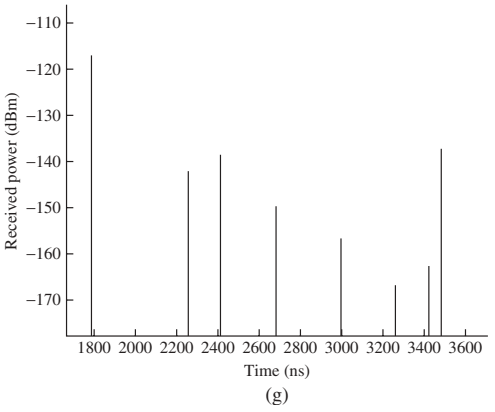


Figure 11.17 (*continued*)

the ends is shown here, with the multipath rays confined within a small angular spread, although there are cases where ground reflections can increase the elevation angular spread. Figure 11.17g shows the time of arrival versus power for the multipath rays.

11.4 SMART ANTENNAS

11.4.1 Beam-Forming Basics

The term *beam forming* relates to the capability of the antenna array to focus energy toward a specific direction in space and nulls in the undesired directions. For this reason, beam forming is often referred to as spatial filtering. Just spatial filtering or beam forming was the first approach to carrying out space–time processing of data sampled at antenna arrays.

The Bartlett (*conventional*) beam former was the first to emerge during World War II [25]. Later, adaptive beam formers and classical time-delay estimation techniques were applied to enhance the ability to resolve closely spaced signal sources [26, 27]. From a statistical point of view, the classical techniques can be seen as spatial extensions of the spectral Wiener (or *matched*) filtering method [28]. However, the conventional beam former has some fundamental limitations connected to the physical size of the array, the available data collection time, and the signal-to-noise ratio (SNR). Some aspects of *analog* and *digital* beam forming are introduced next.

11.4.1.1 Analog Beam Forming Figure 11.18 depicts a radiofrequency (RF) beam-former example for creating only one beam at the output [29]. In practice, RF beam formers can employ microwave waveguides, microstrip structures, transmission lines, or printed microwave circuits.

Multiple-beam beam formers are more complex configurations, based mathematically on the beam-forming matrix (the Butler matrix is the most well known and widely used matrix [30, 31]). Figure 11.19a shows a Butler beam-forming matrix for a four-element antenna array. This matrix utilizes two 45° fixed-phase shifters and four 90° phase-lag hybrid junctions (Figure 11.19b). Typically, the number of beams is equal to the number of antenna elements in the arrays.

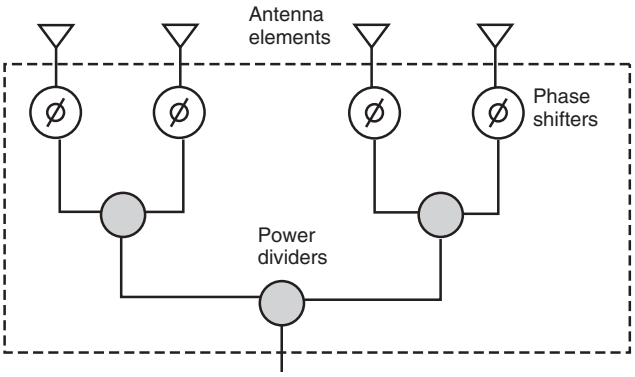
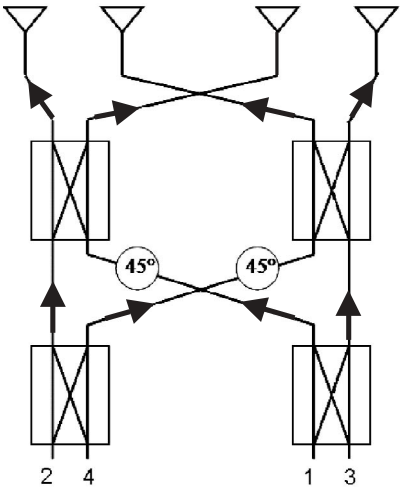
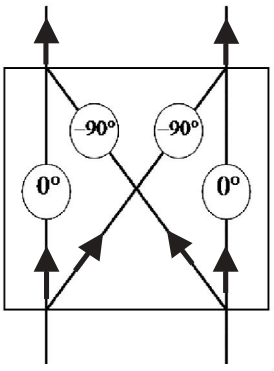


Figure 11.18 Architecture of a simple, on-beam RF beam former.



(a)



(b)

Figure 11.19 (a) A Butler beam-forming matrix for a four-element antenna array, and (b) its phasing scheme.

By tracing the signal from the four ports to the array elements, one can show that the phase distribution at the antenna aperture corresponds to the individual ports of the four-port Butler matrix. Figure 11.20 depicts the radiation pattern from a four-element antenna array with elements spaced at $\lambda/2$ using a Butler matrix feed structure. Although these four beams are overlapping, they are mutually orthogonal [32].

Although the Butler matrix was developed before the fast Fourier transform (FFT), they are completely equivalent. The Butler matrix is used with analog beam forming, whereas the FFT is used in digital beam forming [33].

11.4.1.2 Digital Beam Forming Digital beam forming is achieved by converting the incident RF signal at each antenna element into two streams of binary complex baseband signals that represent the in-phase component (I) and the 90° phase shifted or quadrature component (Q). These weighted signals, from each element, are sampled and stored, and beams are then formed summing the appropriate samples [34]. Depending on the choice of weights, one can use this technique to realize a multibeam antenna array for a switched or an adaptive system. Processing speed and cost have been problems traditionally, but today, inexpensive digital processors, such as field programmable gate arrays (FPGAs), and advanced digital signal processing (DSP) techniques have made the use of smart antennas a reality in wireless communication systems.

A simple structure where a processor can be inserted into an antenna array to achieve beam forming is shown in Figure 11.21 [35]. Let us assume that this array has M elements (sensors). The output $y_n(\theta)$, $(t = nT)$ is given by the sum of the sampled baseband data at M sensors or

$$y_n(\theta) = \sum_{m=0}^{m-1} w_m^* \, x_m(n) \tag{11.21}$$

where x_m is the signal from the mth element of the array, w_m^* is the weight applied to the antenna element m, and * denotes the complex conjugate operation. Equation (11.21)

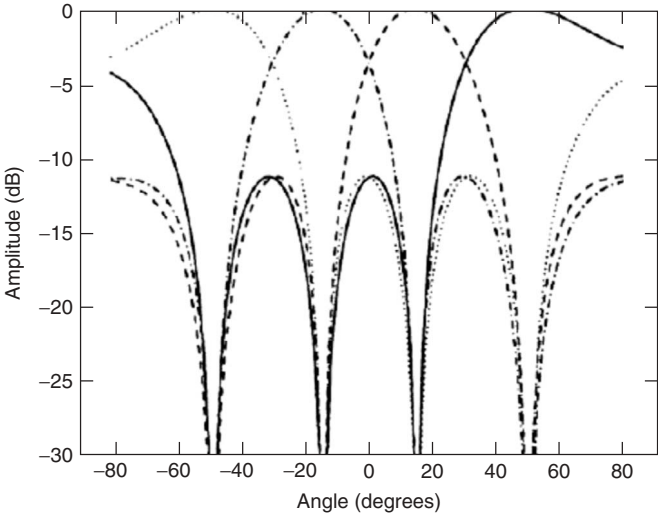


Figure 11.20 Computation of the antenna pattern corresponding to the beam former in Figure 11.19. (From Ref. 32.)

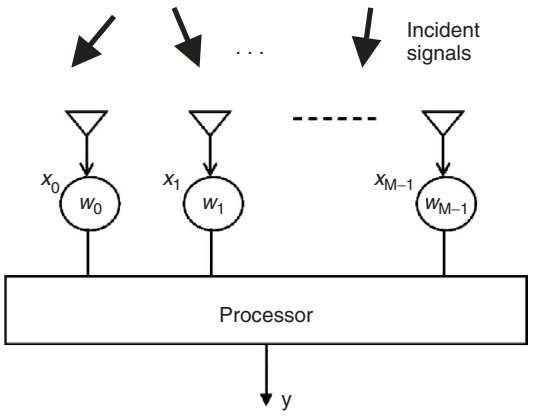


Figure 11.21 Simple digital beam-former architecture.

can be also written

$$y_n(\theta) = \mathbf{w}^H \cdot \mathbf{x}(n) \tag{11.22}$$

where ^{H} represents the Hermitian transpose. The nth sample of the output $y_n(\theta)$ is given by the array snapshot multiplied by a set of weights chosen to form a beam in a desired direction θ.

For generating an arbitrary number of simultaneous beams from an M antenna element array, we can use the arrangement shown in Figure 11.22. Here, each beam former generates a separate beam by applying independent weights to the array signals [36], that is,

$$y(\theta_i) = \sum_{m=0}^{M-1} w_m^{i*} x_m \tag{11.23}$$

where $y(\theta_i)$ is the output of the beam former for a beam in direction θ_i, x_m is a sample from the mth array element, and w_m^i are the weights required to form a beam in direction θ_i.

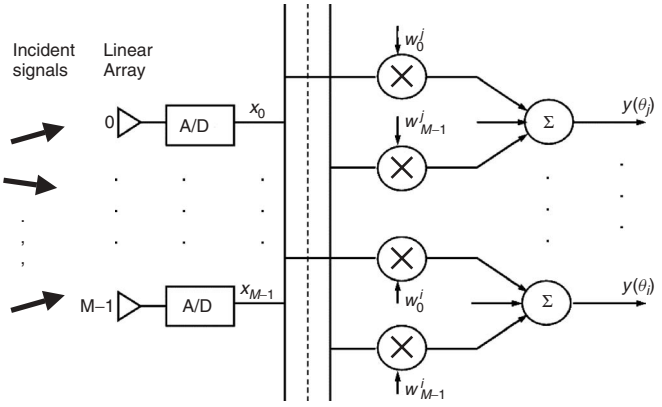


Figure 11.22 A simple beam former capable of generating an arbitrary number of beams.

It should be mentioned here that by choosing the appropriate weight vectors, *beam steering, adaptive nulling*, and *beam shaping* can be achieved, as will be described next.

Example 3 (Beam Steering) Let us use a two-element array antenna to steer the main beam in the direction of a desired signal and a null in the direction of an interference. Let the two antenna elements be separated by a distance of $d = \lambda/2$, as shown in Figure 11.23. Also, let us assume that the desired signal arrives at $\theta_1 = 0^\circ$ and the interfering signal at $\theta_2 = \frac{\pi}{6}$ Both signals use the same carrier frequency f.

The incoming desired signal, given by $s_d = V_d e^{j2\pi ft}$, is received by the two elements with the same phase. The output y_d for the desired signal is a linear combination of the incoming desired signal and the two corresponding weight coefficients, w_1 at the first element and w_2 at the second element, or

$$y_d = V_d e^{j2\pi ft}(w_1 + w_2) \tag{11.24}$$

Simultaneously, the interfering signal $s_{\mathrm{I}}(t) = V_{\mathrm{I}} e^{j2\pi ft}$ is received by the first and second antenna elements with a phase shift of $\phi = kd\,\sin(\pi/6) = 2\pi(\lambda/2)(\frac{1}{2}/\lambda) = \pi/2$ Thus the array output for the interfering signal equals [37]

$$y_{\mathrm{I}} = V_{\mathrm{I}} e^{j2\pi ft} w_1 + V_{\mathrm{I}} e^{j2\pi ft + \pi/2} w_2 \tag{11.25}$$

To guarantee that the desired signal appears at the output, the following conditions must be satisfied:

$$\mathrm{Re}[w_1] + \mathrm{Re}[w_2] = 1 \quad \text{and} \quad \mathrm{Im}[w_1] + \mathrm{Im}[w_2] = 0 \tag{11.26}$$

At the same time, to minimize (or eliminate) the effect of the interfering signal, without affecting the desired signal, one must enforce the condition that the array output for the

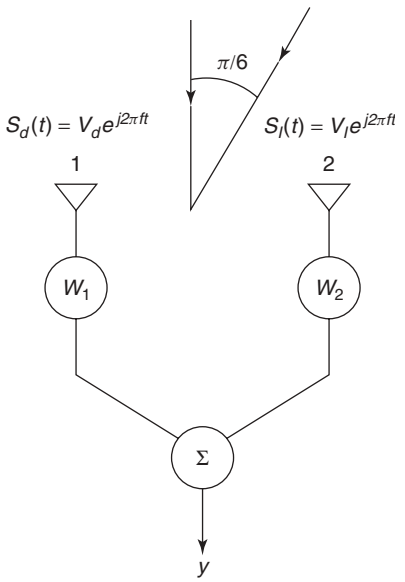


Figure 11.23 An adaptive array of two elements.

interference response must be zero or

$$\mathrm{Re}[w_1] + \mathrm{Re}[jw_2] = 0 \quad \text{and} \quad \mathrm{Im}[w_1] + \mathrm{Im}[jw_2] = 0 \tag{11.27}$$

Expressions (11.26) and (11.27) are then solved simultaneously to yield

$$w_1 = \tfrac{1}{2} - j(\tfrac{1}{2}) \quad \text{and} \quad w_2 = \tfrac{1}{2} + j(\tfrac{1}{2}) \tag{11.28}$$

Using these two weights, the two-element array will accept the desired signal while simultaneously it will eliminate the interfering signal. In this example we assumed prior knowledge of the direction of arrival for the interfering and desired signals. Since this is not the case in practice, adaptive arrays can actually be used to determine the angle of arrival (AoA) of a desired signal before it is used for beam steering [38].

11.4.2 Smart Antenna Methods

11.4.2.1 Switched Beams: Adaptive Antennas Smart antenna systems are usually categorized as either switched beam or adaptive antenna array systems. Although both systems attempt to increase gain in the direction of the user, only the adaptive array system can offer optimal gain since it can actively identify and track both the desired and the interfering signals. It is the adaptive system's active interference capability that offers substantial performance advantages and flexibility over the more passive switched beam approach [39]. The switched beam method can be considered as an extension of traditional cellular sectorization, where a typical cell is composed of three 120° sectors. The switched beam approach can further subdivide each sector into several sub sectors (see Figure 11.24). As a result, the output from a switched beam is a grid of orthogonal

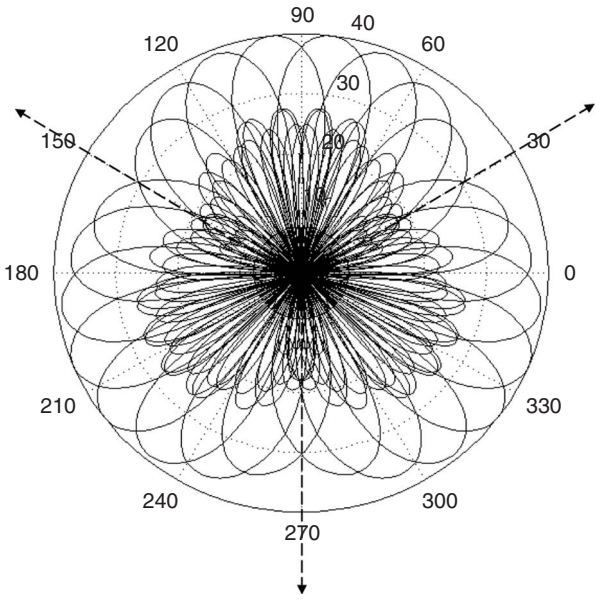


Figure 11.24 Example of switched beams: grid of eight beams produced by a six-element linear array per sector.

beams, where each beam has a maximum toward some direction and minima toward the directions of all other beams. The best beam from the grid is selected based on an algorithm that employs maximum power or SINR as criteria. Due to the fixed pointing direction nature of the switched beam approach, it introduces cusping loss between beams, offers limited interference suppression, and is sensitive to false beam locking due to interference, shadowing, and wide angular spread.

The switched beam technique can be extended to achieve dynamic sectorization by optimizing sector coverage as a function of traffic requirements. Figure 11.25b shows an example of such a scenario where a narrower sector has been created from 60° to 120° by choosing the appropriate beams in order to support the increased traffic in this area (see Figure 11.25a), while the other two sectors have been widened accordingly. The result is to balance the load throughout the whole cell, hence improving the overall capacity. This technique gives reasonable benefits for nonuniform traffic scenarios (in practice, traffic during a day is highly time variant, e.g., high traffic demands in highways during

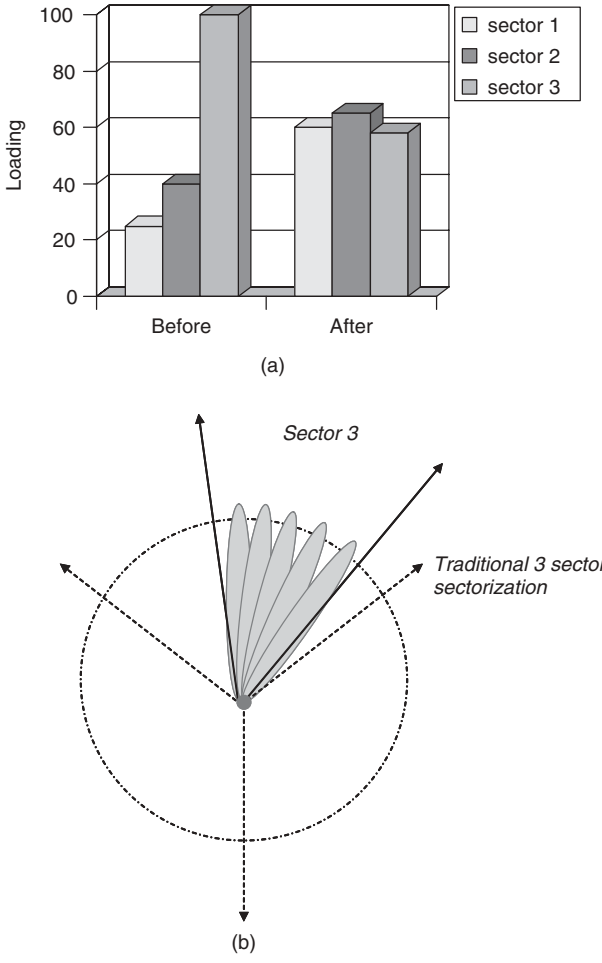


Figure 11.25 Dynamic sectorization example: (a) loading before and after dynamic sectorization and (b) dynamic sectorization layout.

morning–afternoon rush hours but lower during office hours and high traffic demands in office areas during office hours but low in the afternoon–night).

Adaptive antennas take a different approach. By adjusting to an RF environment as it changes, they can dynamically alter the signal patterns to optimize the performance of the wireless system. The adaptive approach continuously updates its beam response based on changes in both the desired and interfering signals. The ability to track users with main lobes and interferers with nulls ensures that the link budget is constantly maximized. Figure 11.26 shows the architecture of a basic adaptive array antenna. By continuously adjusting the complex weights of the array with sophisticated signal processing (the weight controller is the "brain" of the technique or where the intelligence comes from), beam steering as well as desired or interfering user location and/or SINR optimization can be achieved.

Figure 11.27 illustrates an example for the beam pattern response that each smart antenna method might choose in a scenario involving one desired signal and two

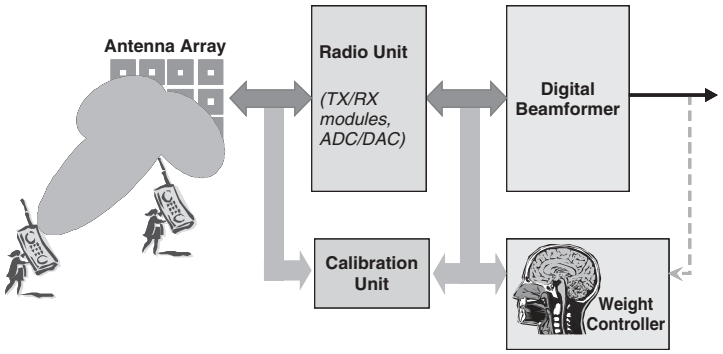


Figure 11.26 Basic architecture of an adaptive array antenna.

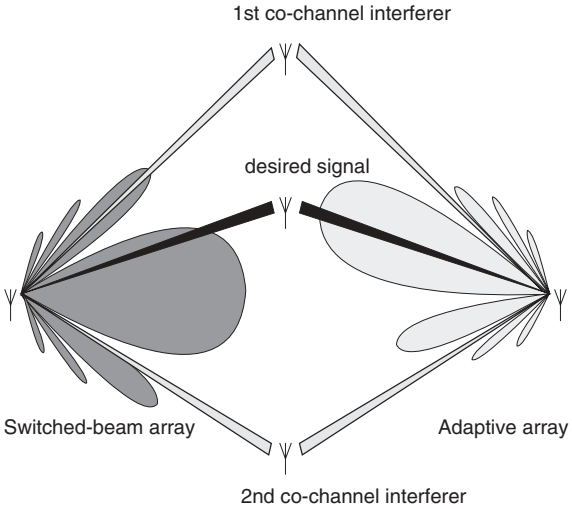


Figure 11.27 Smart antenna response: switched beam system is on the left and the adaptive antenna system on the right.

co-channel interferers. The switched beam system is depicted on the left, while the adaptive antenna system on the right. Both systems direct their major lobes in the general direction of the signal of interest. The switched beam has to choose the "best" beam from a grid of beams with fixed directions while the adaptive system chooses a more accurate placement, hence providing greater signal enhancement. Similarly, the interfering signals are reduced due to the lower gain outside the main lobe of the chosen radiation pattern for the switched beam method, while again, in the adaptive antenna system, the interfering signals experience maximum suppression.

11.4.2.2 Smart Antenna Network Implementations Different network implementations of smart antennas include spatial filtering for interference reduction (SFIR) and space division multiple access (SDMA) (see Figure 11.28). When a smart antenna is deployed in a network for SFIR, the goal is to support one user in each of the co-channel cells and through interference reduction in the spatial domain, to achieve a lower reuse pattern (and hence increase capacity). SDMA, on the other hand, is deployed by exploiting the spatial separation of the users, so that multiple users can operate simultaneously within the same cell (same time and frequency channel). This concept can be seen as a dynamic (as opposed to fixed) sectorization approach, where each mobile defines its own sector as it moves.

The major advantages and disadvantages of these techniques are summarized in Table 11.4 [39]

TABLE 11.4 Advantages and Disadvantages of SDMA and SFIR

Implementation	Advantages	Disadvantages
SDMA	• No need for revised frequency planning to exploit capacity gain • Limited cell deployment for local capacity improvement	• Requires discrimination between intracell SDMA users • More complex radio resource management (angle and power)
SFIR	• No need for major air interface changes • Minor changes to the radio resource management	• Relies on intelligent intracell handover • Large deployments necessary to exploit the full capacity potentials

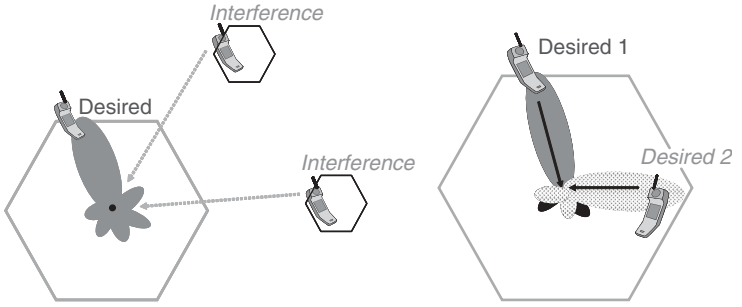


Figure 11.28 SFIR and SDMA concepts.

11.5 SMART ANTENNA ALGORITHMS

11.5.1 Adaptive Beam Forming

Howells was the first to implement an adaptive *sidelobe canceler* (SLC) at intermediate frequency (IF), [40] (see Figure 11.29). This was the first adaptive antenna system that was capable of placing nulls in the direction of interfering signals. His system was a combination of an antenna array with a high gain main-beam "dish" antenna. The dish antenna had a weight of w_0, whereas the surrounding linear array elements were assigned weights $w_m, m = 1, 2, \ldots, M-1$.

Applebaum developed the well-known Howells–Applebaum algorithm [27], which was the starting point on how to control the weights of the adaptive beam former to maximize the SNR at the array output. For the analog SLC multibeam antenna loop, Applebaum expressed a differential adaptive processing equation, given by

$$T\frac{dw_i}{dt} + w_i = G\left(p_i - x_i^*(t)\sum_{m=0}^{M-1} w_m\, x_m(t)\right) \tag{11.29}$$

where T is the smoothing filter time constant, G is the amplifier gain, w_i is the i th weight at the i th element output, $x_i(t)$ is a signal from the i th antenna element, and p_i is the cross correlation between $x_i(t)$ and the output of the main dish antenna channel (reference signal).

The above equation can be expressed in a matrix form:

$$\mathbf{w} = \mu\mathbf{R}^{-1}\mathbf{p} \tag{11.30}$$

where μ is a scalar used for convergence purposes of the algorithm and $\mathbf{R}$ is the $M \times M$ covariance matrix $\mathbf{R} = E[\mathbf{x}(n)\mathbf{x}^H(n)]$. Here it is assumed that the correlation matrix is nonsingular so its inverse exists (it is true when noise is present). This matrix is formed

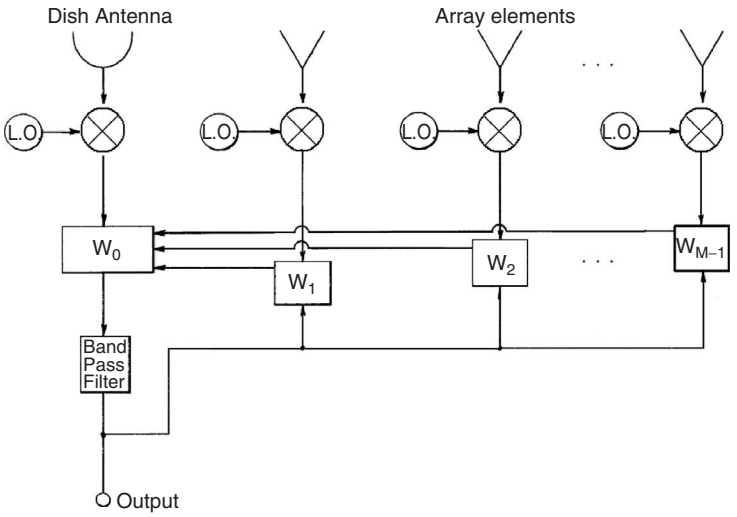


Figure 11.29 Adaptive sidelobe canceler.

from the expected values of the array signal correlation. The above solution is the optimal antenna array weight vector in the MMSE (minimum mean square error) sense [28].

The mean output power from this optimum processor is

$$P = \mathbf{w}^H \mathbf{R}\mathbf{w} \tag{11.31}$$

11.5.1.1 Sample Matrix Inversion (SMI) Algorithm Reed developed the *sample matrix inversion* (SMI) technique for a fast adaptive beam-forming scheme [41, 42]. This algorithm provides estimates for the array weights by replacing the correlation matrix $\hat{\mathbf{R}}$ an estimate from multiple snapshots (N) of the array signal:

$$\widehat{\mathbf{R}} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{x}(n)\,\mathbf{x}^H(n) \tag{11.32}$$

The estimate of $\mathbf{R}$ can be updated when new samples arrive:

$$\widehat{\mathbf{R}}(n+1) = \frac{n\widehat{R}(n) + x(n+1)x^H(n+1)}{n+1} \tag{11.33}$$

11.5.1.2 Least Mean Squares (LMS) Algorithm Widrow developed the *least squares or LMS algorithm*, which is the most commonly used adaptive algorithm [26]. The LMS algorithm updates the weights at each iteration by estimating the gradient of the quadratic surface and then moving the weights in the negative direction of the gradient by a constant, which is referred to as "step" [43]:

$$\nabla\left(E\left(e(n)^2\right)\right) = -e^*(n)\cdot\mathbf{x}(n) \tag{11.34}$$

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \text{step}\cdot\nabla\left(E\left(e(n)^2\right)\right) \tag{11.35}$$

From these two equations it can be seen that

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \text{step}\cdot[e^*(n)\cdot\mathbf{x}(n)] \tag{11.36}$$

When the step size is smaller than the maximum eigenvalue of the correlation matrix $\mathbf{R}$, the algorithm is stable and the mean values of the estimated weights converge to the optimal weights. Since the sum of the eigenvalues of $\mathbf{R}$ is its trace, the step size is usually chosen as

$$0 < \text{step} < \frac{1}{\text{Trace}(\mathbf{R})} \tag{11.37}$$

Clearly, the LMS algorithm does not require prior knowledge of the signal statistics (i.e., $\mathbf{R}$ and $\mathbf{p}$). The LMS algorithm amounts to approximating the expected value by the product of the observation vector and the estimation error, a technique known as *stochastic approximation*. Because the filter coefficients depend on new observations and are continuously updated via the estimate of the gradient, the LMS filter coefficients do not converge to a stable set of values even when the observation is stationary. Rather, after an initial transient response to the set of observations, the coefficients settle toward

a set of values and vary randomly to some degree. The variance of the filter's coefficients is related to the adaptation parameter "step," but in the opposite way to the settling time. A small value for the "step" results in slow convergence of the filter coefficients and their long-term values do not vary greatly about the optimal values. Hence the value for the parameter "step" depends on the trade-off between convergence rate and misadjustment (misadjustment is the ratio of the excess squared error to the minimum squared error and is a measure of how closely the adaptive process tracks the optimum Wiener solution).

The LMS algorithm is the least computationally complex weight adaptation algorithm. However, the rate of convergence to the optimum weights depends on the eigenvalues of the correlation matrix $\mathbf{R}$, that is, the power of the desired and interfering signals.

Although both Applebaum's maximum SNR algorithm and Widrow's LMS error algorithm were discovered independently using different approaches, they are basically very similar and in the case of stationary signals they both converge to the optimum Wiener solution.

The LMS algorithm is a simple adaptive algorithm, but with data-dependent behavior. The influence of the data mean power on the convergence time and the final misadjustment can be removed by normalization of the step size by the data mean power estimate, which effectively leads to the normalized LMS (NLMS) algorithm.

11.5.1.3 Recursive Least Squares (RLS) Algorithm In Ref. 43 it is shown that the weights can be calculated in a recursive manner:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mathbf{R}^{-1}(n)e^*(n+1)\mathbf{x}(n+1)}{\lambda + \mathbf{x}^H(n+1)\mathbf{R}^{-1}(n)\mathbf{x}(n+1)} \tag{11.38}$$

$$\mathbf{R}^{-1}(n+1) = \frac{1}{\lambda}\left[\mathbf{R}^{-1}(n) - \frac{\mathbf{R}^{-1}(n)\mathbf{x}(n+1)\mathbf{x}^H(n+1)\mathbf{R}^{-1}(n)}{\lambda + \mathbf{x}^H(n+1)\mathbf{R}^{-1}(n)\mathbf{x}(n+1)}\right] \tag{11.39}$$

$$\mathbf{R}^{-1}(0) = \delta^{-1}\mathbf{I} \tag{11.40}$$

where δ is a small positive constant. The *forgetting factor* λ depends on the fading rate of the channel and the best value is close to 1. The RLS algorithm is extremely fast compared to the LMS algorithm (it generally needs about $2M$ iterations, with M the number of antenna elements) and does not suffer from poor convergence in variance (inconsistency) as with the LMS, but requires several matrix multiplications at each iteration (i.e., has greater computational complexity).

The RLS algorithm is susceptible to round-off noise in an implementation of the algorithm with finite precision arithmetic. The major problem with round-off errors occurs in the updating of $\mathbf{R}^{-1}(n)$. To avoid this problem a decomposition of the correlation matrix or its inverse may be performed. The resulting algorithm is then dependent on the data vector and not the *squared* data vector (as in the correlation matrix), consequently reducing the effect of the round-off errors [28, 43, 44]:

$$\mathbf{R}^{-1}(n) = \mathbf{A}(n)\mathbf{A}^H(n) \tag{11.41}$$

Thus $\mathbf{A}(n)$ instead of $\mathbf{R}(n)$ can be updated:

$$\mathbf{A}(n+1) = \frac{1}{\sqrt{\lambda}}\left[\mathbf{A}(n) - \gamma\mathbf{A}(n)\mathbf{A}^H(n)\mathbf{x}(n)\mathbf{x}^H(n)\mathbf{A}(n)\right] \tag{11.42}$$

The initial value for the $\mathbf{A}(k)$ matrix is chosen again as for the RLS algorithm:

$$\mathbf{A}^{-1}(0) = \delta^{-1}\mathbf{I} \tag{11.43}$$

and γ is a parameter that is calculated from the element signal vector and the correlation matrix.

11.5.2 Direction Finding Methods

Direction finding methods have been studied extensively in the literature. The problem of direction estimation is significantly simplified if it is assumed that each source transmits in such a way that one plane wave arrives at a receiving antenna array from a specific direction (estimation of point sources). Here, only some of these algorithms are presented, the most widely used for smart antenna applications in wireless communication systems.

11.5.2.1 Conventional Beam Forming (CBF) This method (also referred to as delay and sum or Fourier method) is possibly the simplest technique for direction finding [45]. The output power from the conventional beam former as a function of the angle of arrival (spatial power spectrum) is given by

$$P_{\text{CBF}} = \frac{\mathbf{a}^H(\phi)\mathbf{R}\mathbf{a}(\phi)}{\mathbf{a}^H(\phi)\mathbf{a}(\phi)} \tag{11.44}$$

where $\mathbf{a}(\phi)$ is the steering vector.

If a signal is received from the array from a direction ϕ_0, then the output power of the conventional beam former is maximized when $\mathbf{w} = \mathbf{a}(\phi_0)$. Therefore, in order to produce the spatial power spectrum with a conventional beam former, we need to estimate the correlation matrix of the received signal and calculate the steering vector for all the angles of interest. Generally, it is a simple and hence robust technique but suffers from poor resolution.

11.5.2.2 Minimum Variance Distortionless Response (MVDR) The MVDR method (also referred to as maximum likelihood of spectrum estimation or Capon's method [46]) employs array weights that minimize the mean output power subject to maintaining constant gain in the desired direction. Using a Lagrange multiplier, the weights that satisfy this constrained optimization problem are

$$\mathbf{w} = \frac{\mathbf{R}^{-1}\mathbf{a}(\phi)}{\mathbf{a}^H(\phi)\mathbf{R}^{-1}\mathbf{a}(\phi)} \tag{11.45}$$

This can also be related to classical adaptive antenna array terminology by recognizing that such weights produce the optimum adaptive antenna response.

The expression for the spatial power spectrum is given by

$$P_{\text{MVDR}} = \frac{1}{\mathbf{a}^H(\phi)\mathbf{R}^{-1}\mathbf{a}(\phi)} \tag{11.46}$$

The MVDR method is more complex than the conventional beam former but offers improved resolution at the cost of lower noise reduction.

11.5.2.3 Maximum Entropy (ME) The ME method (or Burg method [47]) calculates the spatial power spectrum so that its Fourier transform equals the measured autocorrelation subject to the constraint that its entropy is maximized. It has been shown by Gabriel [48] to be equivalent to an LMS error linear prediction. Also, Gabriel [48] showed that the linear prediction filter is identical in configuration to the Applebaum sidelobe canceler, and hence the expression for the spatial power spectrum is given by

$$P_{\mathrm{ME}}(\phi) = \frac{1}{\mathbf{a}^H(\phi)\mathbf{w}\mathbf{w}^H\mathbf{a}(\phi)} \tag{11.47}$$

where $\mathbf{w} = \mu\mathbf{R}^{-1}c$, μ is a scalar, and c is the quiescent weight steering vector, which for the Applebaum sidelobe canceler is $[1 \quad 0 \quad \cdots \quad 0]^T$. ME offers better resolution than the MVDR algorithm.

11.5.2.4 Eigenstructure Methods These methods exploit the eigenstructure of the array correlation matrix in order to achieve better resolution for the angle of arrival estimation. The space spanned by the eigenvectors of the array correlation matrix can be split into two orthogonal subspaces: the signal (associated with the larger eigenvalues) and noise (associated with the smaller eigenvalues) subspaces:

$$\mathbf{R} = \sum_{i=1}^{M} \lambda_i e_i e_i^H = \mathbf{E}_{\mathrm{S}}\Lambda_{\mathrm{S}}\mathbf{E}_{\mathrm{S}}^H + \sigma^2\mathbf{E}_{\mathrm{N}}\mathbf{E}_{\mathrm{N}}^H \tag{11.48}$$

where $\mathbf{E}_s$ and $\mathbf{E}_N$ represent the signal and noise eigenvectors, respectively, and Λ_s are the signal eigenvalues.

Since the steering vectors associated with the desired directions are orthogonal to the noise subspace, they are contained in the signal subspace. As a result, such methods search in principle for directions where the corresponding steering vectors are orthogonal to the noise subspace.

Multiple Signal Classification (MUSIC) MUSIC estimates the noise subspace $\mathbf{E}_N$, either by eigenvalue decomposition of the array correlation matrix or singular value decomposition of the data matrix (several snapshots) [49]. The MUSIC spatial power spectrum is given by

$$P_{\mathrm{MUSIC}}(\phi) = \frac{1}{\mathbf{a}^H(\phi)\mathbf{E}_{\mathrm{N}}\mathbf{E}_{\mathrm{N}}^H\mathbf{a}(\phi)} \tag{11.49}$$

In this spatial power spectrum, peaks occur in directions where the steering vector is orthogonal to the noise subspace. This is demonstrated in Figure 11.30, where it can be seen that the MUSIC algorithm can resolve two closely spaced signals.

The MUSIC algorithm offers high resolution (see Table 11.5) but has certain problems in some situations:

1. It is difficult in practical situations to estimate the real number of sources and hence separate noise and signal eigenvectors. This is due to the fact that in theory all noise eigenvectors should correspond to the same eigenvalues, which is not true in real environments (see Figure 11.31). Practical measures to solve such problems could include adaptive "windowing" of the eigenvalue profile.

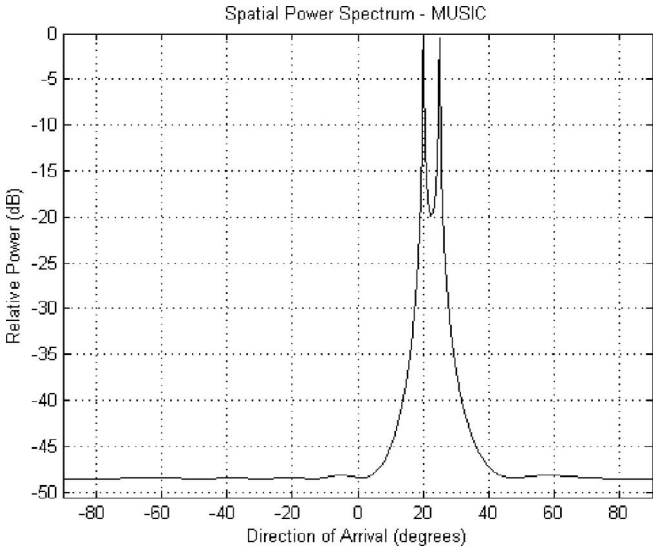


Figure 11.30 Spatial power spectrum for the MUSIC algorithm. Two signals with 20-dB SNR arriving at an eight-element array from 20° and 25°.

TABLE 11.5 Electrical Angle Separation (Resolution Δ_r) for Conventional Beam Forming, MVDR, and MUSIC Methods[a]

CBF	MVDR [50]	MUSIC [51]
$\frac{2\pi}{M}$	$8.71\sqrt{\frac{1}{M^5 SNR}}$	$\Delta_r : \frac{2880(M-2)}{NM^4\Delta_r}\left[1+\sqrt{1+\frac{NM^2\Delta_r^2}{60(M-1)}}\right] = SNR$

[a] N = number of snapshots, M = number of array elements

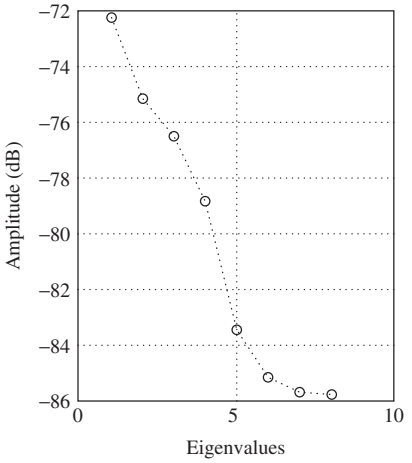


Figure 11.31 Eigenvalues measured at an eight-element antenna array, from field trials in an urban operational scenario.

2. In the presence of highly correlated signals (e.g., due to multipath), MUSIC fails (singularity problems). Nevertheless, in most wireless scenarios, clustering of the multipath [52] results in lower correlation. Furthermore, spatial smoothing or multidimensional searches are employed to decorrelate the signals (presented in the following).

3. MUSIC requires accurate calibration. Generally, calibration is required for digital beam-forming systems where each channel contains a separate transceiver, each of which must be amplitude and phase matched over the entire signal bandwidth. Any distortions introduced will effectively preclude precise pattern control and hence significantly reduce the overall benefit offered.

The accuracy required from the calibration process is mainly driven by the effects of the element mismatches on the adaptive signal processing. For algorithms such as selection diversity, which rely on power measurements, it is only necessary to ensure that any gain mismatches introduced in the receive chains do not bias the power measurements made across the array—and hence the selection decision. For switched beams the distortion of the beam pattern resulting from the element mismatch must be considered. For super-resolution algorithms, such as the MUSIC algorithm, which are susceptible to array mismatch errors, the calibration requirements are significantly more demanding. In this case the quality of the calibration will be evident from the noise floor of the MUSIC spectrum, which is an indicator of the null depth that is achieved in the field. Field trials [53] demonstrated the need for accurate calibration. The tests were performed for a range of environmental conditions and showed that the gross fixed offset distortion across an eight-element antenna array varied over almost 2 dB and 180° across the different elements, and that the worst dynamic offset about the mean was 4.1 dB in terms of amplitude and 35° in terms of phase variations.

4. With low SNR and number of samples, the MUSIC resolution degrades. To improve the situation, schemes like the Min-Norm (MN) are employed. The MN method [54] employs a weight vector that is of the minimum norm, has the first element equal to unity, and is contained in the noise subspace. This gives the following expression for the spatial power spectrum:

$$P_{\mathrm{MN}}(\phi) = \frac{1}{\left|\mathbf{a}^H(\phi)\mathbf{E}_{\mathrm{N}}\mathbf{E}_{\mathrm{N}}^H\mathbf{e}_1\right|^2} \tag{11.50}$$

Several modifications of the standard MUSIC method have been proposed in an attempt to improve its performance and reduce computational complexity:

- For a uniform linear array, Barabell [55] proposed Root MUSIC, which calculates the roots of an appropriate polynomial for the AoA search, instead of MUSIC's spectral peak identification. Root MUSIC offers better resolution but is limited to uniform linear arrays only.
- Cyclic MUSIC [56] uses the spectral coherence properties of the signal to improve the standard MUSIC.
- Beam space MUSIC [57, 58] passes the input array data through a beam former before MUSIC processing, hence achieving reduced computation, improved resolution, and reduced sensitivity to errors (beam space processing is described next).

Estimation of Signal Parameters Via Rotational Invariance Techniques (ESPRIT) The ESPRIT method exploits the shift structure that is generated if the array can be divided into two identical arrays displaced by a known translation distance $\triangle$ [59]. An estimate of the input covariance matrix from the measurements of the total array output (received signal vector at the two subarrays) is then eigendecomposed and, using the multiplicity of the smallest eigenvalue, the number of signals is estimated (d_{est}). Based on that and the previous eigendecomposition, the signal subspace is estimated and decomposed into two subarray matrices, which are eigendecomposed as $\mathbf{E}\Lambda\mathbf{E}^H \cdot \mathbf{E}$ is then partitioned into $d_{\text{est}} \times d_{\text{est}}$ submatrices:

$$\mathbf{E} = \begin{bmatrix} \mathbf{E}_{11} & \mathbf{E}_{12} \\ \mathbf{E}_{21} & \mathbf{E}_{22} \end{bmatrix} \tag{11.51}$$

The ith direction of arrival (DoA_i) is estimated as

$$DoA_i = \sin^{-1}\left(\frac{\arg(\lambda_i)}{2\pi\,\Delta}\right) \tag{11.52}$$

where λ_i is the ith eigenvalue of the matrix:

$$\Psi_{\text{TLS}} = -\mathbf{E}_{12}\,\mathbf{E}_{22}^{-1} \tag{11.53}$$

The total least squares (TLS) criterion is described above since, due to the problem's nature (two equally noisy matrices), TLS gives better results.

ESPRIT reduces the computational and storage requirements compared with MUSIC, does not involve exhaustive search through all possible steering vectors to estimate the direction of arrival, and does not require precise knowledge of the array manifold vectors (low calibration requirements). One disadvantage is that, since it requires two identical subarrays, it experiences higher estimation errors compared with other methods that exploit the full array and not only the translation (e.g., stochastic maximum likelihood and subspace fitting).

11.5.2.5 Beam Space Processing The idea behind this method is to preprocess the received data so that only a spatial sector is chosen to be analyzed. This spatial prefiltering of the received data results in computational gain, since the analysis is now performed only for the focused sector. A simple choice to achieve this (apart from knowing broadly where signals come from) is to employ conventional beam forming with steering vectors for a set of chosen directions or spanning a sector with some predefined angular resolution. Several algorithms with high computational complexity (like MUSIC and ESPRIT) have been used with beam space preprocessing. As an example, beam space MUSIC has better resolution and less sensitivity but worse variance than element space MUSIC [60].

11.5.2.6 Algorithms for Estimating the Wireless Channel Although knowledge of the direction of arrival of the desired signal, as outlined above, is useful, in wireless communication applications there are other channel characteristics, like time and frequency spread, that play an important role in the performance of a wireless communication system. For this reason there are many algorithms presented in the literature

that perform joint angle and time-delay estimation. The JADE (joint angle and delay estimation) [61] algorithm is based on fitting a parameterized model of the channel matrix to an unstructured estimate of the channel.

In the next section the SAGE algorithm is briefly discussed.

SAGE (Space-Alternating Generalized Expectation Maximization) This algorithm employs the technique of expectation and maximization in order to optimize the maximum likelihood function, with the parameters of time, direction of arrival, and the Doppler frequency [62]. Essentially, the algorithm splits the problem and solves it for each wave and then performs multiple iterations between the local solutions until a global optimum is found. Assuming that $b(t)$ is the transmitted signal, and the unknowns associated with each path are $\boldsymbol{\delta}_i = [\tau_i, \phi_i, \nu_i, \rho_i]$ (time, angle, speed, wave complex amplitude), E is the energy of the transmitted signal, and p is the number of waves arriving at the array, SAGE does the following calculations:

$$
\begin{aligned}
R_i(\tau, \phi, \nu_i) &= \sum_t b^*(t-\tau_i) e^{j2\pi\nu_i t} \mathbf{a}^H(\phi_i) \widehat{\mathbf{x}}_i(t; \widehat{\delta}) \\
\widehat{\mathbf{x}}_i(t; \widehat{\delta}) &= \widehat{\mathbf{x}} \; - \sum_{i'=1, i' \neq i}^{p} \mathbf{s}(t; \widehat{\delta}_{i'}) \\
\widehat{\tau}_i' &= \arg\max_{\tau_i} \left| R_i(\tau_i, \widehat{\phi}_i, \widehat{\nu}_i) \right| \\
\widehat{\phi}_i' &= \arg\max_{\phi_i} \left| R_i(\widehat{\tau}_i', \phi_i, \widehat{\nu}_i) \right| \\
\widehat{\nu}_i' &= \arg\max_{\nu_i} \left| R_i(\widehat{\tau}_i', \widehat{\phi}_i', \nu_i) \right| \\
\widehat{\rho}_i' &= \frac{R_i(\widehat{\tau}_i', \widehat{\phi}_i', \widehat{\nu}_i')}{\left\| a(\widehat{\phi}_i') \right\|^2 E}
\end{aligned}
\tag{11.54}
$$

Although the general optimization problem is quite complicated, it can be shown that all that SAGE does is to estimate the delay of a path based on the correlation peak between the transmitted signal and the path signal $\mathbf{x}_i(t)$ [63]. For the azimuth estimation, SAGE employs simple beam forming and for the Doppler frequency it employs a Doppler correction.

11.5.2.7 Detection of Number of Sources Since detection of the number of sources/signals arriving at the antenna array is important for most super-resolution direction finding methods, several techniques have been proposed to achieve this goal. Two of the most widely used are based on the Akaike information theoretic criteria (AIC) and the minimum descriptive length (MDL) criteria [64].

In the AIC method, the number of sources is the value $k \in \{0, 1, \ldots, M-1\}$, which minimizes the following criterion:

$$
AIC(k) = -\log \left[\frac{\prod_{i=k+1}^{M} \lambda_i^{1/(M-k)}}{\frac{1}{M-k} \sum_{i=k+1}^{M} \lambda_i} \right]^{(M-k)N} + k(2M-k) \tag{11.55}
$$

In the MDL method the number of sources is the value $k \in \{0, 1, \ldots, M-1\}$, which minimizes the following criterion:

$$MDL(k) = -\log \left[\frac{\prod_{i=k+1}^{M} \lambda_i^{1/(M-k)}}{\frac{1}{M-k} \sum_{i=k+1}^{M} \lambda_i} \right]^{(M-k)N} + \tfrac{1}{2} k(2M-k) \log N \tag{11.56}$$

In Ref. 64 it was shown that the MDL criterion is more consistent while the AIC tends to overestimate the number of sources.

11.5.2.8 Spatial Smoothing Techniques The principle behind this technique is to preprocess the received data by splitting the array into subarrays (in the forward and/or backward directions, see Figure 11.32) and then forming the array correlation matrix by averaging all the subarray correlation matrices [65]. If the antenna array is split into L subarrays both in the forward and backward directions, then the forward only ($\mathbf{R}^{\mathrm{f}}$) and backward only ($\mathbf{R}^{\mathrm{b}}$) spatially smoothed correlation matrices are the mean of the forward/backward subarray correlation matrices, respectively:

$$\mathbf{R}^{\mathrm{f}} = \frac{1}{L} \sum_{l=1}^{L} \mathbf{R}_l^{\mathrm{f}} \tag{11.57}$$

$$\mathbf{R}^{\mathrm{b}} = \frac{1}{L} \sum_{l=1}^{L} \mathbf{R}_l^{\mathrm{b}} \tag{11.58}$$

If $L \geq M$ (the number of subarrays is greater than or equal to the number of signals), then either of the modified correlation matrices will be nonsingular regardless of the coherence of the signals. The maximum number of sources that can be estimated now is half the number of array elements (normally an M-element array can resolve up to $M-1$ uncorrelated sources). In order to increase the available degrees of freedom, both the forward and the backward subarray smoothing techniques can be used:

$$\mathbf{R}^{\mathrm{fb}} = \tfrac{1}{2}\left(\mathbf{R}^{\mathrm{f}} + \mathbf{R}^{\mathrm{b}}\right) \tag{11.59}$$

In this case the array can resolve up to $2M/3$ sources.

Table 11.6 [39] summarizes the most important advantages and disadvantages of the three major methods discussed.

11.6 SMART ANTENNA ADVANTAGES

11.6.1 Coverage Extension

Smart antennas [39] can increase network coverage by exploiting the antenna array directivity (assuming low angular spread). Since the gain G (assuming 100% antenna efficiency and no mutual coupling) with an M-element antenna array is

$$G \approx 10 \log_{10} M \tag{11.60}$$

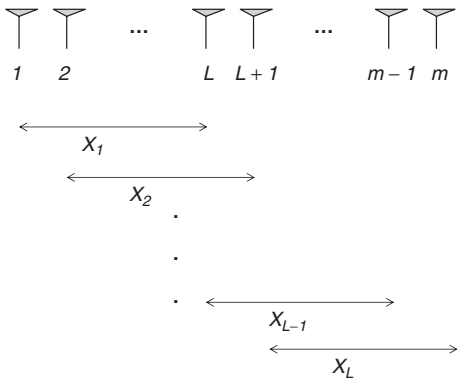


Figure 11.32 Subarray spatial smoothing.

TABLE 11.6 Major Advantages and Disadvantages of Different Smart Antenna Approaches

Method	Advantages	Disadvantages
Switched beams	• Easily deployed • Tracking at beam switching rate	• Cusping loss • Lower overall SINR gain • False locking (due to shadowing, interference, and wide angular spread)
Direction finding	• Tracking at angular change rate • No reference signal required • Easier downlink beam forming	• Lower overall SINR gain • Usually needs good calibration • Concept not applicable to small cell NLOS environments
Optimum combining	• Optimum SINR gain • No need for accurate calibration • Acceptable performance with MMSE approach	• Difficult downlink beam forming with FDD and fast TDD • Needs good reference signal for optimum performance • Requires high update rates

it can be exploited to extend range (from r_1 to r_2):

$$\Delta R = \frac{r_2}{r_1} = M^{1/n} \tag{11.61}$$

assuming a single slope path-loss model with exponent n, or

$$\Delta R = M^{0.3} \tag{11.62}$$

with the typical Hata path-loss model.

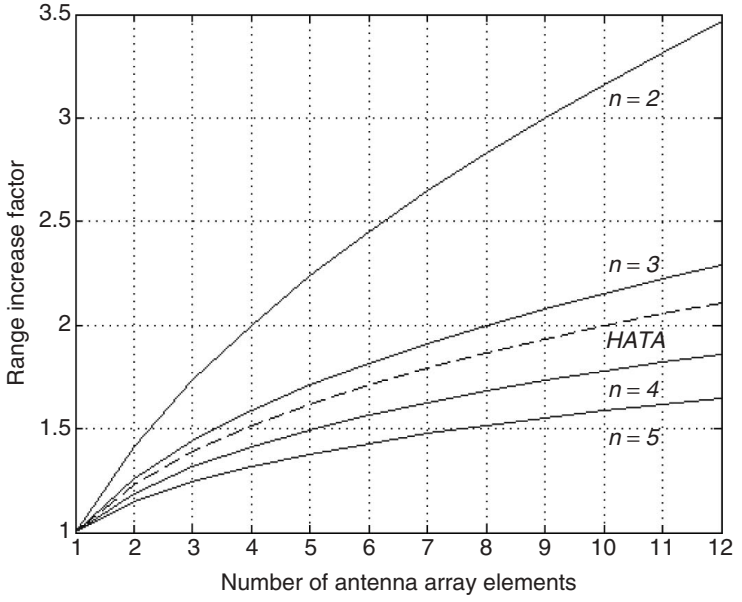


Figure 11.33 Range increase as a function of the number of antenna array elements.

It follows that range extension gives an area improvement factor and hence the number of base stations required to cover a given area is improved by a factor of $M^{2/n}$ (single slope path-loss model).

Figure 11.33 shows that with eight antenna elements the range increases by a factor of 1.5 to almost 3 ($n = 5$ to $n = 2$), or the number of base stations is decreased by a factor of 2.3 to 8. It follows from the analysis that smaller path loss exponents offer higher improvement factors; that is, the improvements are better in line of sight (LOS) rather than nonline of sight (NLOS) scenarios.

The full range extension offered by the smart antenna can be exploited when the angular spread of the signal is less than the antenna array beamwidth (otherwise the desired signal is reduced). This is obviously a greater problem for switched beams since adaptive antennas can also provide diversity gain, which is higher for high angular spread (spatial diversity) [66].

11.6.2 Capacity

Controlling the transmitted and received power with a smart antenna effectively allows for interference reduction (improving the SINR) and hence more users can now share (SFIR) or reuse (SDMA) the available resources. The following section presents some capacity case studies that demonstrate this behavior.

11.6.2.1 Capacity Case Studies

FDMA–TDMA System with Switched Beams Following the analysis in, Ref. 67, the overall outage probability for the scenario with one tier of interfering cells can be

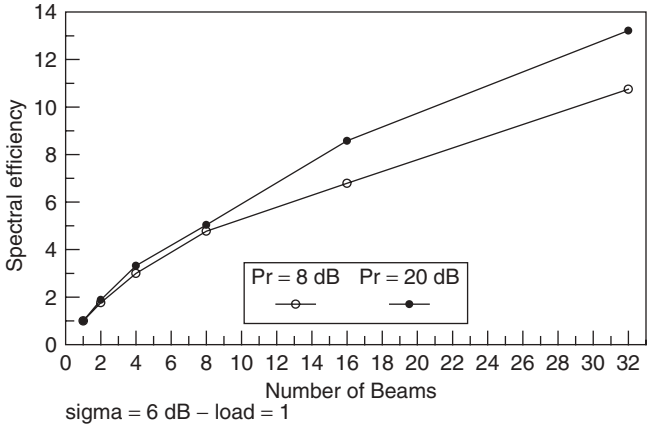


Figure 11.34 Spectral efficiency with an adaptive multibeam antenna.

calculated as

$$P(s_w \le p_R s_I, N_B) = \sum_{CCI} P(s_w \le p_R s_I/cci) \cdot (6 - cci) \left(\frac{LF}{N_B}\right)^{cci} \left(1 - \frac{LF}{N_B}\right)^{6\text{-}cci} \tag{11.63}$$

where N_B is the number of beams, LF is the loading factor, s_w is the wanted signal, s_i is the interfering signal, p_r is some protection ratio, and $P(s_w \le p_R s_I/cci)$ is the conditional outage probability, that is, the probability of co-channel interference (CCI) given that there are cci active interfering cells,

$$\binom{k}{l} = \frac{k!}{l! \cdot (k-l)!}$$

and the omnidirectional case is given for $N_B = 1$.

In Ref. 67 it was also shown that the ratio of the achieved spectral efficiencies with/without smart antennas is inversely proportional to the achieved cluster sizes. Based on the above, the cluster size can be calculated for given shadow fading and loading conditions for an outage criterion. The example plot shown in Figure 11.34 considers 6-dB shadow fading, 100% loading, and 1% outage. The values of 8 and 20 dB were used for the protection ratio, in order to cover a variety of modulation schemes.

It can be seen from Figure 11.34 that the introduction of an adaptive multibeam antenna capable of forming 24 beams (8 beams per sector, as in Figure 11.24) will result in an eight- to tenfold increase in the spectral efficiency (depending on the chosen modulation). It has to be mentioned here that this result represents rather a lower bound for the spectral efficiency, due to the fact that the worst case of cell loading was considered (100%).

Macrocells with CDMA The performance achieved with adaptive antennas in large cell CDMA networks has been studied extensively in the literature [68–71]. A slightly different approach for the calculation of the performance of CDMA with smart antennas (could be characterized as network-based analysis) was proposed in Refs. 72–74. In Refs. 73 and 74, a method based on the radiation pattern characteristics was analyzed. There,

the underlying idea was the fact that the characteristics of the produced adaptive antenna pattern response (and, as a consequence, the performance of the system) are affected from parameters such as scattering of the environment, mutual coupling of the array elements, and other mismatches. If these effects are ignored, the estimated performance will be overly optimistic.

In Refs. 73 and 74, the BER probability was calculated from the equation

$$P_b = Q\left(\sqrt{3 \cdot SF \cdot SIR_{\text{omni}} \cdot G_{\text{AA}}}\right) \tag{11.64}$$

where SF is the spreading factor, SIR_{omni} is the signal-to-interference ratio with an omnidirectional antenna, and G_{AA} is the gain with adaptive antennas, calculated using the approximate interference model for uniform distribution of users from Ref. 75 and a radiation pattern that is approximated as a stepped function with the ideal or effective† beamwidth and average sidelobe level:

$$G_{\text{AA}} = \begin{cases} \dfrac{1}{k}, & \text{central cell} \\ \dfrac{1}{k}\dfrac{5D}{3D+2}, & \text{all cells} \end{cases} \tag{11.65}$$

where "central cell" and "all cells" refer to how many base stations employ smart antennas (four tiers are assumed for the latter scenario), D is the directivity of the radiation pattern, and

$$k = \frac{BW}{2\pi} + SLL\left(1 - \frac{BW}{2\pi}\right) \tag{11.66}$$

with BW and SLL the ideal or effective beamwidth and sidelobe level, respectively.

A Monte Carlo simulation was also developed in Refs. 73 and 74 to verify the results of the above model. The difference between the simulation results and the above model was shown to be less than 0.035 dB for a wide range of beamwidth and sidelobe level values.

Figure 11.35 shows a plot of the SIR gain as a function of the achieved beamwidth and average sidelobe level, when the smart antenna is deployed only at the central cell. When the effective beamwidth is 10° to 40° and the effective average sidelobe level is -10 to -20 dB, the results show that spatial filtering can offer gains between 7 and 14 dB.

Macrocells with WCDMA Table 11.7 shows some basic operational characteristics of the UMTS WCDMA system (see Ref. 76 for more details). In order to perform realistic simulations with adaptive antennas, the problem of accurate radio channel characterization (space and time) for the operational environments under study needs to be resolved. One possible way to do this, as mentioned in Section 11.3, is with ray tracing propagation modeling. The advantage of this technique is that it offers the necessary 3D (site-specific) impulse response of the radio channel with a high degree of accuracy (see Figure 11.18).

By employing such a technique for propagation modeling for a 4-km^2 3D geographical database in conjunction with WCDMA and smart antennas for a typical large cell

†The word *effective* is used in order to imply that the value for a parameter takes into account scattering and signal mismatching effects.

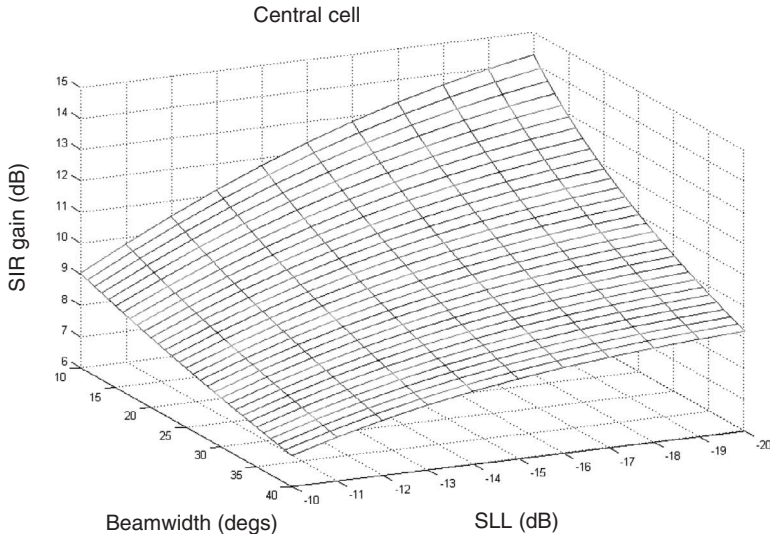


Figure 11.35 SIR gain as a function of beamwidth and sidelobe level.

TABLE 11.7 Basic Characteristics of the UMTS WCDMA System

Multiple access method	DS-CDMA
Duplexing method	Frequency division duplex (FDD)
Chip rate	3.84 Mcps
Frame length	10 ms
Number of slots/frame	15
Multirate concept	Variable spreading factor and multicode
User data rate	12.2, 64, 144, and 384 kbps
Interleaving	Interframe/intraframe
Spreading factor	4 to 256
Modulation	Dual channel QPSK with complex scrambling
Handover	Intra–interfrequency, soft–softer
Power control	Open loop and fast closed loop
Diversity	Multipath diversity with RAKE, antenna diversity
Adaptive antenna support	Connection dedicated pilot bits on UL and DL

scenario, the performances of several adaptive algorithms (conventional and switched beam forming with/without diversity, beam space optimum combining with the largest eigenvalue) were studied in terms of the achieved SIR gain, as shown in Figure 11.36 [77].

For the scenario of a voice loaded system and 99% probability, Figure 11.36 shows that the switched beam version with 13 beams (SB 13 beams) is better than the 9-beam version ($\sim$1 dB), the conventional beam forming (CBF) method is slightly better than the two switched beam versions, and the best performance is achieved with the beam space optimum combining (BSOPC) algorithm.

The analysis also showed that although the relatively simple switched beams method can offer acceptable gains, higher gains can be achieved with adaptive beam forming (i.e.,

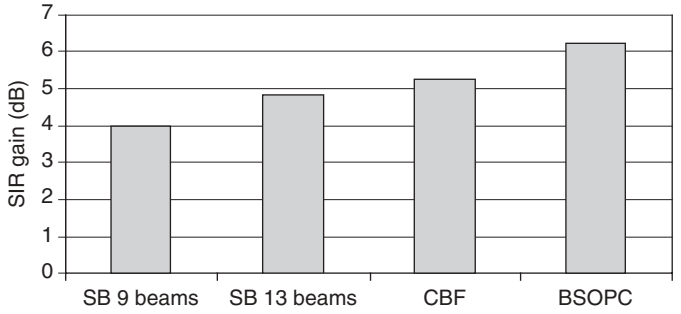


Figure 11.36 Adaptive antenna gain for different algorithms for voice loaded system.

steering the main lobe to the direction of maximum energy). Furthermore, selection and subsequent beam-diversity combining produces better gains; but one should bear in mind in this case the conflicting requirements of desired signal maximization and interference reduction.

For different mixed service requirements it was seen that, for the algorithms examined, the higher the desired user bit rate, the higher the output gain, while the higher the interferer's bit rate, the more the output gain reduction.

11.6.3 More Efficient Power Control: Smart Handover

A smart antenna can offer diversity gain and hence fading of the radio signal can be reduced, leading to better performance of the power control technique. Furthermore, if user location information is combined with the handover mechanism, "soft/softer" and "hard" handovers can now become "smart handovers."

Even more, if handover between cells of different layers in a hierarchical cell structure needs to be supported on the same carrier in order to increase spectral efficiency, spatial filtering with smart antennas can be employed to provide the necessary RF power balancing between the different cells [78]. Nevertheless, more complex radio resource management is necessary in this case, to handle near–far situations between the different cell layers.

11.6.4 Better QoS

In noise or interference limited environments, the gain that can be achieved with an antenna array can be exchanged for signal quality enhancement (i.e., lower BER). This is demonstrated in Figures 11.37 and 11.38. Figure 11.37 considers the probability of detection (P_D) for the case when a matched filter is employed for each antenna array element. From Ref. 45,

$$P_D = Q\left(Q^{-1}(P_F) - \sqrt{M \cdot SNR}\right) \tag{11.67}$$

where $Q(\cdot)$ is the Q function, P_F is the threshold probability of false alarm, M is the number of elements, and SNR is the signal-to-noise ratio.

From Figure 11.37 it can be seen that the detection performance (and hence QoS) can be significantly enhanced with an antenna array. For example, with 0-dB SNR there is

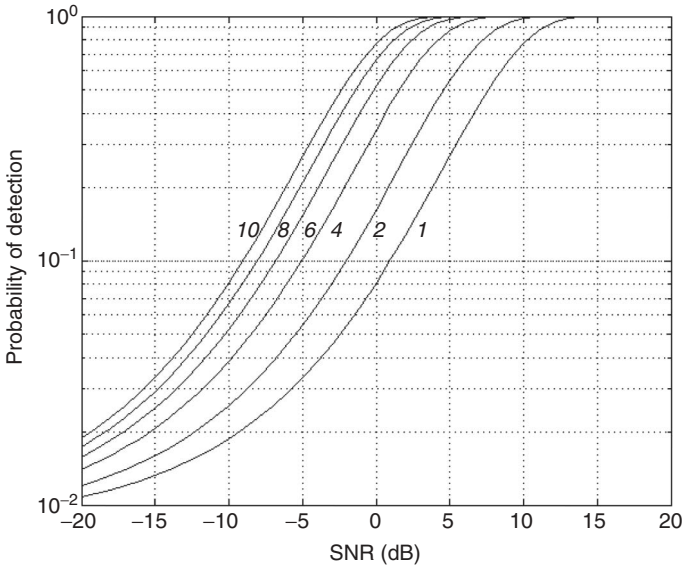


Figure 11.37 Probability of detection for matched filtering with different antenna array elements.

8% probability of detection with a single element, 16% with two elements, 34% with four, 52% with six, 66% with eight, and 77% with ten.

Considering the large cell scenario for DS-CDMA discussed in Section 11.6.2.1, Figure 11.38 shows results for the achieved BER as a function of beamwidth and sidelobe level produced by the smart antenna radiation pattern. It can be seen that for beamwidth and sidelobe levels between 10° and 40°, and -10 and -20 dB, respectively, considerable BER gain can be achieved. If the smart antenna can achieve 30° beamwidth, then an improvement of 1–3 orders of magnitude for the BER can be accomplished with average sidelobe levels between -10 and -20 dB.

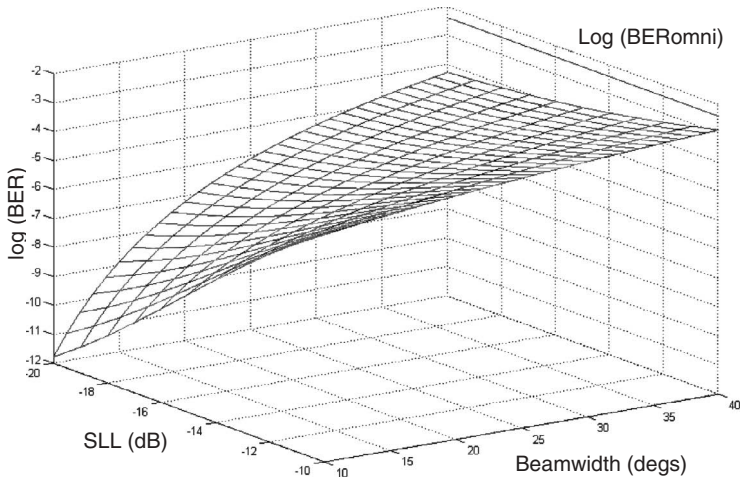


Figure 11.38 BER as a function of the "effective" beamwidth and sidelobe level.

11.6.5 Lower Transmitted Power

Wireless standards usually introduce limitations on the maximum EIRP (effective isotropic radiated power). This effectively means that in some situations the smart antenna array gain cannot be exploited for coverage extension. As a result, it is possible to exchange the antenna array gain in order to reduce the maximum output power of the base station and subsequently of the power amplifiers (PAs). With an M-element antenna array, the reduction is a factor of M^{-2} for the PAs and M^{-1} for the total transmitted power. Eight antenna elements will reduce the total transmitted power by 9 dB, while the output of each PA will be reduced by 18 dB. Obviously, if one (rather than M) PA is employed, the power reduction for the PA and the total base station output power will be the same.

In a similar context, recent public worry for health issues arising from exposure to electromagnetic radiation (we are not concerned here if it is reasonable or not) means that if reduction of the emitted power by the handset could be achieved somehow, it would be highly desirable. This handset reduction can be achieved if the base station (smart antenna) array gain is exploited to accordingly reduce the power transmitted by the mobile. This power reduction is also useful because it relaxes the battery requirements, and hence talk times or handset size and weight can be increased or decreased, respectively.

11.6.6 Ability to Support User Location Services

There are a number of added value services that can be provided more efficiently if the user location capabilities of smart antennas are exploited:

- Emergency call location: The FCC has recently passed to Phase II E911 requirements. Based on that, wireless carriers are required to provide automatic location identification (ALI) for at least 95% of their subscribers, with location accuracy of 100 m for 67% of calls and 150/300 m for 95% of calls (the latter depending on handset/network implementation).
- Location based billing: This will provide the operator with the ability to control its network usage. Furthermore, this method enables tariff plans to be tailored to individual user needs, for example, by providing low cost zones when the user is in a preferred zone.
- On-demand location-specific services: These could include roadside assistance, real-time traffic updates, tourist information, and electronic yellow pages (e.g., with local entertainment and dining information).
- Location and subsequent navigation of a vehicle with possible extensions to package monitoring and stolen vehicle recovery.

11.7 SMART ANTENNA IMPLEMENTATION AND SYSTEM ISSUES

Up to now the discussion on smart antennas has focused mostly on algorithmic and performance issues. Nevertheless, there are a number of other factors that affect both the implementation and the deployment of smart antennas. The following sections discuss some of these issues.

11.7.1 Implementation Issues

- *Antennas.* These will be different for different environments, since performance of smart antennas depends on the spatial characteristics of the operational environment. Zoning issues will probably limit the number of elements to four to eight, depending on interelement distance. Polarization is also important, since it can be employed for diversity and/or extending the number of elements. Aesthetic design for real-world deployments must become an integral part of network development, especially since appearance is strongly connected to health hazards perceptions. Also, mutual coupling and element failure problems should be taken into account.
- *RF Front-Ends.* These can be wideband and possibly combined with software radio for reconfigurability and maximum flexibility.
- *Calibration.* This needs to be regular and must be transparent. Different smart antenna implementations require different levels of accuracy.
- *Digital Beam Former.* It forms multiple beams by finding the inner product of the set of received samples from the array elements and the set of weights produced by the weight controller. Speed requirements depend on the number of elements and the bandwidth. A system with 10-MHz bandwidth and 8 or 12 antenna elements requires 300 or 500 MOPS (millions of operations per second), while the same antenna with a 25-MHz system would require around 1 GOPS.
- *Weight Controller.* This generates a weight vector corresponding to the operational requirements. The computational load varies greatly with the employed algorithm and the required update rate. This effectively means that optimum combining techniques introduce high update rates (fading rate), while direction finding algorithms and switched beams require updates of the order of angular/beam change.
- *A/D and D/A Converters.* They determine the dynamic range of the system. Also, the sensitivity loss due to quantization distortion from limited word lengths affects the null depth and the sidelobe levels of the produced radiation pattern.
- *Intermodulation Distortion.* An important issue is the spurious free dynamic range over which all generated signals or intermodulation products are maintained below some threshold level. The use of DBF techniques poses linearity demands since any distortion in the up/downconversion chains will distort the produced radiation pattern and hence the overall performance of the smart antenna. Tsoulos et al. [53] demonstrated with field trials that as the level of distortion from nonlinearities increases, the sidelobe level increases (5–12-dB null depth reduction) and the null directions shift.

11.7.2 System Issues

- *Control Channels.* They are omnidirectional, and hence there is a problem in terms of range extension. Techniques that could offer additional gain to balance this asymmetry include revolving beams or increased sectorization (e.g., with different scrambling codes for each sector).
- *Initial Access and Handover.* Initial access is based on the omnidirectional random access channel. With smart antennas the capacity increase will lead to a CIR reduction and hence DoA estimation will be problematic. The handover process has

similar problems since the new cell has no information for the user location. Furthermore, soft handover will be problematic if no constraint is introduced, because up to 6 BSs may need to keep tracking all possible MSs. New radio resource management methods that take into account the user location information and limit the available links to two or three, are now necessary. More generally, the radio resource management needs to take into account the interference environment and the smart antenna capabilities (i.e., include the spatial dimension in the optimization process).

- *Dynamic Channel Allocation.* In order to avoid collisions the employed algorithm needs to also consider location information.
- *Discontinuous Transmission.* In some cases (e.g., downlink beam forming based on uplink information) discontinuous transmission should be reduced (e.g., dummy bursts) in order to avoid erroneous estimates due to long periods of silence.
- *Reference Signals.* We need more training sequences for GSM systems with SDMA, and with better cross-correlation properties. For CDMA/WCDMA systems, there is a potential hard blocking with one scrambling code and hence deployment of an additional is required (reduced orthogonality due to the second scrambling code will introduce some loss but only marginal compared with the additional capacity gain [79]).
- *Downlink Beam Forming for FDD.* In frequency division duplex (FDD) systems, the downlink fading characteristics are independent of the uplink characteristics due to the frequency difference between the two bands. In time division duplex (TDD) systems the two links can be considered reciprocal provided that the channel characteristics have not changed considerably between the receive–transmit slots. Under this condition, the weights calculated by the adaptive antenna for the uplink can be used for the downlink. For current FDD systems (e.g., GSM, DCS1800, IS-95, WCDMA), the processing performed in the uplink cannot be exploited directly in the downlink. There are several approaches proposed in the literature which attempt to solve the downlink problem:
  - Dedicated pilot (WCDMA)
  - Feedback from the mobile
  - Beam forming with respect to angle
- *Network Dimensioning and Planning.* The introduction of smart antennas makes things worse for the planning and optimization of a wireless communication network, due to the need to consider an additional parameter—space. Smart antenna implementation, antenna array configuration, spatial characteristics of the radio channel, desired and interfering traffic and control signals (intra/intercell), uplink and downlink for the different services, and radio resource management algorithms with smart antennas are some of the additional issues that need to be considered. Nevertheless, it is rather obvious that this kind of overall optimization, although desirable, soon becomes an extremely complex task. If accuracy is traded off for reduced complexity, then the following two approaches could be used: (1) look-up table for the SINR gain that a certain smart antenna can achieve in different operational environments and conditions for each service; and (2) the method proposed in Refs. 72–74, where the performance of a smart antenna is analyzed based on two parameters—the effective beamwidth and average sidelobe level of the produced radiation pattern (both depend on the chosen antenna array characteristics and the environment).

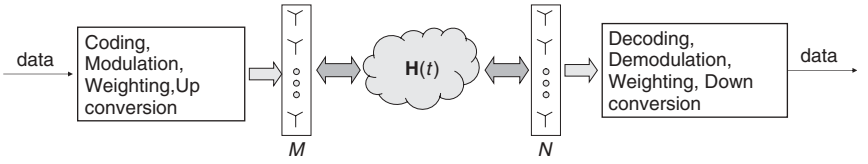


Figure 11.39 MIMO system architecture.

11.8 DISCUSSION: MIMO SYSTEMS

As discussed earlier in this chapter, smart antennas refer to processing of data at one end of the radio link—either at the base station or the mobile (typically at the base station where space, processing power, and power consumption are more readily available). When multiple antennas are used at both ends of the radio link, MIMO (multiple-input multiple-output) systems emerge [80–82]. The underlying idea is that on transmission and reception from multiple antennas, the signals are combined appropriately so that the quality (BER) or data rate (bps) is improved. The way to achieve this is through processing an $M \times N$ channel matrix that is created from the M transmit and N receive antenna elements (Figure 11.39).

Example 4 A simple intuitive example in order to understand the MIMO concept is *spatial multiplexing* [82]. Assume a 2×2 system. A high bit rate data stream is split into two streams and transmitted through the two antennas. Upon reception from the two antennas at the other end, the signals are separated and estimated (exploiting knowledge of the radio channel provided by pilot signals). Effectively, this is achieved by solving a linear system with two unknowns. The separation is possible only if the equations are independent, a condition that is satisfied in rich multipath scenarios.

It was shown in Refs. 80–82 that the capacity (bps/hertz) achieved with an $M \times N$ MIMO system (channel matrix with random independent elements) is given by

$$C = \log_2 \left[\det \left(\mathbf{I}_M + \frac{SNR}{N} \mathbf{H}\mathbf{H}^* \right) \right] \tag{11.68}$$

where the SNR refers to the average SNR at any receive antenna. For a large number of antennas, the average capacity increases linearly with $M (M = N)$. This is a very significant result since it implies limitless capacity, if the additional cost of multiple RF chains can be afforded. In practice, capacity is shown to increase linearly with the smaller number of antennas $(\min(M, N))$ and will be limited by radio channel characteristics and the chosen transmission algorithm.

REFERENCES

1. C. A. Balanis, *Antenna Theory: Analysis and Design*, John Wiley & Sons, Hoboken, NJ, 1997.
2. R. E. Collin, *Antennas and Radiowave Propagation*, McGraw-Hill, New York, 1985.
3. J. D. Kraus, *Antennas*, 2nd ed., McGraw-Hill, New York, 1988.
4. K. F. Sander, and G. A. L. Reed, *Transmission and Propagation of Electromagnetic Waves*, 2nd ed., Cambridge University Press, Cambridge, UK, 1986.

5. W. L. Stutzman, and G. A. Thiele, *Antenna Theory and Design*, John Wiley & Sons, Hoboken, NJ, 1981.
6. R. B. Ertel, P. Cardieri, K. W. Sowerby, T. S. Rappaport, and J. H. Reed, Overview of spatial channel models for antenna array communication systems, *IEEE Personal Commun. Mag.*, Vol. 5, No. 1, pp. 10–22, February 1998.
7. G. E. Athanasiadou, Spatio-temporal propagation modelling, in *MIMO System Technology for Wireless Communications*, G. Tsoulos (Ed.), CRC Press, Boca Raton, FL, 2006, Chap. 1.
8. W. C. Y. Lee, Effects on correlation between two mobile radio basestation antennas, *IEEE Trans. Commun.*, Vol. 21, No 11, pp. 1214–1224, 1973.
9. D. Aszetly, On antenna arrays in mobile communication systems: fast fading and GSM base station receiver algorithm, Ph.D. dissertation, Royal Institute of Technology, March 1996.
10. P. Petrus, J. H. Reed, and T. S. Rappaport, Geometrically based statistical channel model for macrocellular mobile environment, in *IEEE GLOBECOM*, 1996, pp. 1197–1201.
11. J. Fuhl, A. Molisch, and E. Bonek, Unified channel model for mobile radio systems with smart antennas, *Proc. IEE Radar, Sonar. Navig.*, Vol. 145, No. 1, pp. 32–41, February 1998.
12. J. Liberty, and T. Rappaport, A geometrically based model for line-of-sight multipath radio channels, in *IEEE Vehicular Technology Conference (VTC)*, 1996, pp. 844–848.
13. P. Zetterberg, and B. Ottersten, The spectrum efficiency of a basestation antenna array system for spatially selective transmission, in *IEEE VTC*, 1994.
14. P. Zetterberg, and P. L. Espensen, A downlink beam steering technique for GSM/DCS1800/PCS1900, in *IEEE PIMRC*, Taipei, Taiwan, October 1996.
15. R. J. Piechocki, and G. V. Tsoulos, Combined GWSSUS and GBSR channel model with temporal variations, in *Joint COST259/260 Workshop*, Vienna, Austria, April 1999.
16. R. J. Piechocki, G. V. Tsoulos, and J. P. McGeehan, Simple general formula for PDF of angle of arrival in large cell operational environment, *IEE Electron. Lett.*, Vol. 34, pp. 1784–1785, September 3, 1998.
17. R. J. Piechocki, J. P. McGeehan, and G. V. Tsoulos, A new stochastic spatio-temporal propagation model (SSTPM) for mobile communications with antenna arrays, *IEEE Trans. Commun.*, Vol. 49, No. 5, pp. 855–862, May 2001.
18. Q. H. Spencer, B. D. Jeffs, M. A. Jensen, and A. L. Swindlehurst, Modeling the statistical time and angle of arrival characteristics of an indoor multipath channel, *IEEE J. Selected Areas Commun.*, Vol. 18, No. 3, pp. 347–360, March 2000.
19. K. I. Pedersen, P. E. Mogensen, and B. H. Fleury, A stochastic model of the temporal and azimuthal dispersion seen at the base station in outdoor propagation environments, *IEEE Trans. Vehicular Technol.*, Vol. 49, No. 2, pp. 437–447, March 2000.
20. A. Algans, K. I. Pedersen, and P. E. Mogensen, Experimental analysis of the joint statistical properties of azimuth spread, delay spread, and shadow fading, *IEEE J. Selected Areas Commun.*, Vol. 20, No. 3, pp. 523–531, April 2002.
21. V. Perez et al., Final propagation model, CODIT Deliverable No. R2020/TDE/PS/DS/P/040/b1, 1994/6.
22. S. Y. Seidel, and T. S. Rappaport, Site-specific propagation prediction for wireless in-building personal communications system design, *IEEE Trans. Vehicular Technol.*, Vol. 43, No. 4, pp. 1058–1066, November 1994.
23. G. E. Athanasiadou, and A. R. Nix, A novel 3D indoor ray-tracing propagation model: the path generator and evaluation of narrowband and wideband predictions, *IEEE Trans. Vehicular Technol.*, Vol. 49, No. 4, pp. 1152–1168, July 2000.
24. G. E. Athanasiadou, A. R. Nix, and J. P. McGeehan, A microcellular ray-tracing propagation model and evaluation of its narrowband and wideband predictions, *IEEE J. Selected Areas Commun., Wireless Commun., Series*, Vol. 18, No. 3, pp. 322–335, March 2000.

25. M. S. Bartlet, *An Introduction to Stochastic Process*, Cambridge University Press, Cambridge, UK, 1956.
26. B. Widrow, P. E. Mantey, L. J. Griffiths, and B. B. Goode, Adaptive antenna systems, *IEEE Proc.*, Vol. 55, No. 12, pp. 2143–2159, December 1967.
27. S. P. Applebaum, Adaptive arrays, *IEEE Trans. Antennas Propag.*, Vol. 24, No. 5, pp. 585–598, September 1976.
28. R. Monzingo and T. Miller, *Introduction to Adaptive Arrays*, John Wiley & Sons, Hoboken, NJ, 1980.
29. L. C. Godara, Application of antenna arrays to mobile communications, Part II: beam-forming and direction-of-arrival considerations, *Proc. IEEE*, Vol. 85, No. 8, pp. 1195–1245, August 1997.
30. J. Butler and R. Lowe, Beamforming matrix simplifies design of electronically scanned antennas, *Electron. Design*, April 1961.
31. J. P. Shelton and K. S. Kelleher, Multiple beams from linear arrays, *IRE Trans. Antennas Propag.*, March 1961.
32. J. Litva, and T. Lo, *Digital Beamforming in Wireless Communications*, Artech House, London, 1996.
33. D. E. Dudgeon, Fundamentals of digital array processing, *Proc. IEEE*, Vol. 65, No. 5, pp. 898–904, June 1977.
34. R. G. Pridham and R. A. Mucci, Digital interpolation beamforming for low-pass and bandpass signals, *Proc. IEEE*, Vol. 67, No 5, pp. 904–919, 1979.
35. S. Anderson, M. Millnert, M. Viberg, et al., An adaptive array for mobile communication systems, *IEEE Trans. Vehicular Technol.*, Vol. 40, No 2, pp. 230–236, 1991.
36. A. H. El Zooghby, C. G. Christodoulou, and M. Georgiopoulos, Adaptive interference cancellation with neural networks, in *Wireless Personal Communications: Emerging Technologies for Enhanced Communications*, The Kluwer International Series in Engineering and Computer Science, Kluwer, Norwell, MA, 1998.
37. N. Blaunstein and C. G. Christodoulou, *Radio Wave Propagation and Adaptive Antennas for Wireless Communication Links: Terrestial, Atmospheric and Ionospheric*, John Wiley & Sons, Hoboken, NJ, 2006.
38. C. G. Christodoulou and M. Georgiopoulos, Smart adaptive array antennas for wireless communications, *Proc. SPIE*, Vol. 4395, pp. 75–83, 2001.
39. G. V. Tsoulos, Smart antennas for mobile communication systems: benefits and challenges, *IEE Electron. Commun. Eng. J.*, Vol. 11, No. 2, pp. 84–94, April 1999.
40. P. W. Howells, Explorations in fixed and adaptive resolution at GE and SURC, *IEEE Trans. Antennas Propag.*, Vol. 24, No. 5, pp. 575–584, September 1976.
41. I. S. Reed, J. D. Mallett, and L. E. Brennan, Rapid convergence rate in adaptive arrays, *IEEE Trans. AES*, Vol. 10, pp. 853–863, 1974.
42. T. K. Sarkar, M. C. Wicks, and M. Salazar-Palma, *Smart Antennas*, John Wiley & Sons, Hoboken, NJ, 2003.
43. S. Haykin, *Adaptive Filter Theory*, Prentice Hall, Englewood Cliffs NJ, 2002.
44. S. Mitra and J. Kaiser, *Handbook for Digital Signal Processing*, John Wiley & Sons, Hoboken, NJ, 1993.
45. D. Johnson and D. Dudgeon, *Array Signal Processing*, Prentice Hall, Englewood Cliffs, NJ, 1993.
46. J. Capon, High resolution frequency wave number spectrum analysis, *Proc. IEEE*, Vol. 57, pp. 1408–1418, 1969.
47. J. P. Burg, Maximum entropy spectral analysis, in 37th *Annual Meeting of the Society of Exploration in Geophysics*, Oklahoma, USA, 1967.

48. W. F. Gabriel, Spectral analysis and adaptive array superresolution techniques, *Proc. IEEE*, Vol. 68, No. 6, pp. 654–666, June 1980.

49. R. O. Schmidth, Multiple emitter location and signal parameter estimation, *IEEE Trans. Antennas Propag.*, Vol. 34, No. 3, pp. 276–280, March 1986.

50. G. V. Serrebryakov, Direction of arrival estimation of correlated sources by adaptive beamforming, *IEEE Trans. SP*, Vol. 43, No. 11, pp. 2782–2787, November 1995.

51. M. Kaveh and A. J. Barbell, The statistical performance of the MUSIC and the minimum norm algorithms in resolving plane waves in noise, *IEEE Trans. ASSP*, Vol. 34, pp. 331–341, April 1986.

52. G. V. Tsoulos and G. E. Athanasiadou, On the application of adaptive antennas to microcellular environments: radio channel characteristics and system performance, *IEEE Trans. Vehicular Technol.*, Vol. 51, No. 1, pp. 1–16, January 2002.

53. G. V. Tsoulos, J. P. McGeehan, and M. A. Beach, Space division multiple access (SDMA) field trials—Part II: calibration and linearity issues, *IEE Proc. Radar Sonar Navig.*, Special Issue on Antenna Array Processing Techniques, pp. 79–84, February 1998.

54. R. Kumaresan and D. W. Tufts, Estimating the angles of arrival of multiple plane waves, *IEEE Trans. Aerospace Electron. Systems*, Vol. 19, pp. 134–139, January 1983.

55. A. Barabell, Improving the resolution of eigenstructured based direction finding algorithms, *Proc. ICASSP*, pp. 336–339, 1983.

56. S. V. Schell, R. A. Calabretta, W. A. Gardner, and B. G. Agee, Cyclic MUSIC algorithms for signal selective DOA estimation, *Proc. ICASSP*, pp. 2278–2281, 1989.

57. J. T. Mayhan and L. Niro, Spatial spectral estimation using multiple beam antennas, *IEEE Trans. Antennas Propag.*, Vol. 35, pp. 897–906, 1987.

58. I. Karasalo, A high-resolution postbeamforming method based on semidefinite linear optimisation, *IEEE Trans. ASSP*, Vol. 38, pp. 16–22, 1990.

59. R. Roy and T. Kailath, ESPRIT—estimation of signal parameters via rotational invariance—techniques, *IEEE Trans. ASSP*, Vol. 37, pp. 984–995, July 1986.

60. K. Buckley and X. L. Xu, Spatial spectrum estimation in a location section, *IEEE Trans. ASSP*, Vol. 38, pp. 1842–1852, November 1990.

61. M. C. Vannderveen, C. B. Papadias, and A. Paulraj, Joint angle and delay estimation (JADE) for multipath signals arriving at an antenna array, *IEEE Commun. Lett.*, Vol. 1, No. 1, pp. 12–14, January 1997.

62. J. A. Fessler and A. Hero, Space-alternation generalised expectation-maximisation algorithm, *IEEE Trans. SP*, Vol. 42, No. 10, pp. 2664–2675, October 1994.

63. M. Viberg and T. Svantesson, Direction of arrival estimation in mobile communication environments, in *Handbook of Antennas in Wireless Communications*, L. Godara, (Ed.), CRC Press, Boca Raton, FL, 2002, Chap. 20.

64. M. Wax and T. Kailath, Detection of signals by information theoretic criteria, *IEEE Trans. ASSP*, Vol. 33, No. 2, pp. 387–392, April 1985.

65. T. J. Shan, M. Wax, and T. Kailath, On spatial smoothing for direction of arrival estimation of coherent signals, *IEEE Trans. ASSP*, Vol. 33, No. 4, pp. 806–811, August 1985.

66. J. H. Winters, Smart antennas for wireless systems, *IEEE Personal Commun. Mag.*, pp. 23–27, February 1998.

67. S. C. Swales, M. A. Beach, D. J. Edwards, and J. P. McGeehan, The performance enhancement of multibeam adaptive base-station antennas for cellular land mobile radio systems, *IEEE Trans. Vehicular Technol.*, Vol. 39, No. 1, pp. 56–67, February 1990.

68. R. Kohno, H. Imai, M. Hatori, and S. Pasupathy, Combination of an adaptive array antenna and a canceller of interference for direct sequence spread spectrum multiple access system, *IEEE J. Selected Areas Commun.*, Vol. 8, No. 4, pp. 675–682, May 1990.

69. A. Naguib, A. Paulraj, and T. Kailath, Capacity improvement with base station antenna arrays in cellular CDMA, *IEEE Trans. Vehicular Technol.*, Vol. 43, No. 3, pp. 691–698, August 1994.
70. J. Thompson, P. Grant, and B. Mulgrew, Smart antenna arrays for CDMA systems, *IEEE Personal Commun.*, Vol. 3, No. 5, pp. 16–25, October 1996.
71. S. Choi and D. Yun, Design of an adaptive antenna array for tracking the source of maximum power and its application to CDMA mobile communications, *IEEE Trans. Antennas Propag.*, Vol. 45, No. 9, pp. 1393–1404, September 1997.
72. J. Liberti and T. Rappaport, Analytical results for capacity improvements in CDMA, *IEEE Trans. Vehicular Technol.*, Vol. 43, No. 3, pp. 680–690, August 1994.
73. G. V. Tsoulos, On the single and multiple cell deployment of adaptive antennas with CDMA, *IEE Electron. Lett.*, Vol. 34, No. 23, pp. 2196–2197, November 12, 1998.
74. G. V. Tsoulos, Approximate SIR and BER formulas for DS-CDMA based on the produced radiation pattern characteristics with adaptive antennas, *IEE Electron. Lett.*, Vol. 34, No. 19, pp. 1802–1804, September 17, 1998.
75. Qualcomm Inc., An overview of the application of code division multiple access (CDMA) to digital cellular systems and personal cellular networks, EX60–10010, May 1992.
76. H. Holma and A. Toskala, *WCDMA for UMTS*, John Wiley & Sons, Hoboken, NJ, 2002.
77. G. V. Tsoulos, G. E. Athanasiadou, and R. J. Piechocki, Low complexity smart antenna methods for third generation WCDMA systems, *IEEE Trans. Vehicular Technol.*, Vol. 49, No. 6, pp. 2382–2396, November 2000.
78. G. V. Tsoulos, G. E. Athanasiadou, M. A. Beach, and S. C. Swales, Adaptive antennas for microcellular and mixed cell environments with DS-CDMA, *Wireless Personal Commun. J.*, Special Issue on CDMA for Universal Personal Communications Systems, Vol. 7, No. 2/3, pp. 147–169, August 1998.
79. K. Ingemann, P. Mogensen, and J. Ramiro-Moreno, Application and performance of downlink beamforming techniques in UMTS, *IEEE Commun. Mag.*, pp. 134–143, October 2003.
80. G. J. Foschini and M. J. Gans, On limits of wireless communications in fading environment when using multiple antennas, *Wireless Personal Commun. J.*, Vol. 6, No. 3, pp. 311–335, March 1998.
81. I. E. Telatar, Capacity of multi-antenna Gaussian channels, Bell Labs Technical Memorandum, 1995. Also in *Eur. Trans. Telecom.*, Vol. 10, No. 6, December, 1999.
82. G. J. Foschini, Layered space–time architecture for wireless communications in a fading environment when using multi-element antennas, *Bell Labs Tech. J.*, pp. 41–59, Autumn 1996.

CHAPTER 12

Wideband Arrays

WILLIAM F. CROSWELL, TIM DURHAM, MARK JONES, DANIEL SCHAUBERT, PAUL FRIEDERICH, and JAMES G. MALONEY

12.1 INTRODUCTION

The purpose of this chapter on wideband arrays is to introduce the antenna design community to a new method of array design. The traditional method of wideband array design is to first design an antenna element that has, for example, an impedance and gain bandwidth of 5:1 or 10:1. Then one tries to choose an array lattice that is devoid of major grating and surface wave effects over the operating bandwidth and scan volume. Another traditional design goal is to minimize interelement coupling. One soon determines that the radiation pattern of an array composed of wideband elements often exhibits grating and surface wave modes which can result in scan blindness at certain angles.

Three independent groups of people have been investigating wideband array technology over the last decade. Breaking with the conventional array methodology, each group independently discovered that tightly coupled elements were necessary to achieve wide bandwidth performance. This assessment by the three groups was based upon extensive analysis using well-validated numerical methods such as the finite difference time domain, (FDTD), method of moments (MOM), and the finite element method (FEM).

Several fundamental rules were established by each group independently, such as:

- The unit cell which contains the basic antenna element must be no larger than 0.5 wavelengths at the highest frequency. For a 10:1 bandwidth, this means the unit cell is .05 wavelengths at the lowest frequency. If the unit cell is larger than 0.5 wavelengths, grating lobes and surface wave lobes can occur in the frequency band and scan volume of interest.
- The maximum theoretical gain of the unit cell is $4\pi A/\lambda^2$, where A is the area of the unit cell.
- For planar arrays, the distance from the aperture layer to the groundplane is determined by the lowest frequency of operation.

Modern Antenna Handbook. Edited by Constantine A. Balanis

The approach developed at Harris Corporation, known as the current sheet array (CSA), consists of dipole elements capacitively coupled at their end points placed over a groundplane. The element layer can be embedded in multiple layers of dielectric which serve to improve the impedance match over scan. This resulted from a fundamentally different perspective in which mutual coupling is intentionally introduced in a controlled manner between the elements. The initial concept was developed by Dr. Ben Munk based on his extensive study of periodic structures over several decades. Numerous current sheet array breadboards have been built and tested that demonstrate bandwidths approaching a decade, validating the theoretical predictions.

Extensive studies of wideband arrays of Vivaldi elements have been conducted by Dr. Daniel Schaubert at the University of Massachusetts. This research has produced Vivaldi array designs having VSWR <2:1 over bandwidths of at least 10:1 and nearly ideal active element patterns for angles out to 60°. A balun is incorporated into the antenna structure, which allows printed transmission lines such as microstrip and stripline to be used to feed the elements. It is now understood that a successful array design for wide bandwidth and electronic scanning using Vivaldi elements depends strongly on management of coupling between array elements. A result of using coupling between elements is that finite array size affects elements near the array edges more than in conventional array designs.

Georgia Tech Research Institute (GTRI) began investigating wideband planar arrays at approximately the same time as the previously discussed research. Their approach treats the aperture as a blank canvas and uses optimization routines to synthesize elements, which can be located on multiple layers. The results of the numerical optimization is a class of antennas known as fragmented apertures due to the complex distribution of conducting regions over the unit cell. An important aspect of the design is that the elements are connected across the unit cell boundaries. The fragmented aperture is located above multiple resistive cards and dielectric layers which are used to create a wideband backplane. Measured results of embedded-element breadboards have shown bandwidths of 33:1.

12.2 CURRENT SHEET ARRAY

12.2.1 Introduction

The design of wideband phased arrays with operational bandwidths approaching a decade or more is a challenging area of current research. The use of these arrays would result in benefits such as the ability to use a single array for wideband or widely separated signals and the ability to share a common aperture for multiple functions. Since fewer openings would be required in a host platform needing to communicate on widely spaced frequency bands, the use of wideband arrays could reduce integration cost and also ease other system-level requirements.

Significant analytical and empirical effort is usually required in order to design wideband arrays. This is in part due to mutual coupling between array elements, which complicates the array design. Mutual coupling results when array elements located in close proximity interact in a manner that alters the impedance and radiation characteristics of the elements. Because of the difficulty associated with predicting these effects, mutual coupling is traditionally considered an obstacle to array design. Mutual coupling typically tends to make the active element pattern more directive than the ideal element pattern and therefore increases the scan rolloff of the array.

Fortunately, an analysis approach based on Floquet's theorem (i.e., periodic boundary or "unit cell" approach) can be applied to large array design in order to simplify the prediction of array parameters such as active impedance and active element patterns. This method assumes that the array size is infinite and therefore does not include edge effects, but it inherently includes all mutual coupling between array elements. This approach can generally be used for arrays larger than five wavelengths at the lowest frequency of operation [1].

A relatively recent development in the field of wideband arrays is described in the following sections, which provides new degrees of freedom to the array designer by applying a fundamentally different approach. This new methodology leverages from extensive knowledge of the behavior of the passive wideband periodic structures known as frequency selective surfaces. In this approach, some of the limitations present in traditional array design are no longer obstacles to achieving wideband operation.

12.2.2 Limitations of Conventional Wideband Array Design

The conventional approach to wideband array design is to start from an element designed in isolation and attempt to make it function over the desired bandwidth in the array environment. This often yields unsatisfactory results since mutual coupling changes the element performance from the isolated response. Furthermore, uncontrolled mutual coupling between array elements can produce deleterious effects such as an array blindness when the array beam is scanned away from broadside. An overview of mutual coupling effects in arrays can be found in phased array texts [2].

In the conventional approach, the array designer must balance avoiding undesired mutual coupling effects with eliminating grating lobes due to the element spacing at the high end of the frequency band. This limits the array bandwidth when designing with traditional elements. Elements that are typically used in wideband arrays also tend to be deep and not amenable to conformal applications. These limitations have prevented array designers from providing array systems that are wideband, planar, and free from grating lobes over a large scan volume.

12.2.3 Current Sheet Array Concept

A fundamentally different approach to the design of wideband arrays has been developed in which mutual coupling is intentionally introduced in a controlled manner between the elements. This methodology treats the array aperture as a periodic surface and largely ignores isolated element performance, so that the performance of elements within the array environment is assured. The same approach is used to design periodic, highly coupled structures such as a frequency selective surface (FSS) [3].

In 1999, Harris Corporation sought to develop a decade-bandwidth scanning array for a variety of applications. Harris approached Dr. Ben Munk of The Ohio State University Electroscience Laboratory for concepts for a nominal 9:1 bandwidth phased array targeted for 2–18 GHz operation. Dr. Munk drew upon his experience in wideband periodic structures to develop an initial design concept in which electrically small dipole elements above a ground plane are capacitively coupled at their end points. This approach was termed the "current sheet array" (CSA) since it achieves something very similar to Wheeler's theoretical current sheet concept [4].

Dr. Munk and his associates had discovered that a closely spaced overlapping dipole array exhibits a wide bandwidth when employed as an FSS. This was recognized to be

due to the inherent capacitance between the overlapping portions of the elements [3, 5]. This led to the realization that the same effect can be achieved in a controlled manner by introducing capacitance between the ends of the dipoles, as shown in Figure 12.1. In fact, an array of dipoles with small element spacings above a ground plane can be designed to have a 2:1 VSWR bandwidth of around 4:1. Investigations of similar dipole geometries performed at Harris in the early 1990s had found that approximately the same bandwidth can be obtained from finite dipole arrays. Previous work by Dr. Munk had proved that the impedance variation with scan angle could be minimized by using closely spaced elements [6].

The question then became how to achieve an even greater bandwidth from this capacitively coupled dipole configuration. By studying the behavior of the equivalent circuits of the array and ground plane combination, it was observed that placing a single dielectric layer over the array yielded even greater bandwidth. The investigation was continued using numerical analysis techniques to augment the equivalent circuit analysis used up to this point. The periodic moment method (PMM) code used for this numerical analysis was originally developed to predict scattering from infinite periodic arrays of wire and slot elements embedded in multiple dielectric layers. It was later extended to calculate input impedance for active arrays of wire elements [7].

The result of performing extensive calculations using PMM was that a capacitively coupled array of dipole elements embedded in a carefully chosen dielectric profile can provide wideband performance. Figure 12.2 shows the dielectric structure for a typical example. Note that thin dielectric layers are placed around the array, and there are two dielectric slabs above the array to further improve the wideband match. Figure 12.3 shows the input impedance computed by PMM for this array in series with a short transmission-line matching section. This design exhibits a VSWR below 2:1 over an approximately 7:1 bandwidth. Subsequent development of CSA test articles at Harris has demonstrated that the usable bandwidth approaches a decade.

There are numerous benefits of this new approach to wideband array design. The controlled interelement coupling, dielectric profile, and avoidance of resonant behavior eliminate the rapid variations in impedance and yield a bandwidth approaching a decade.

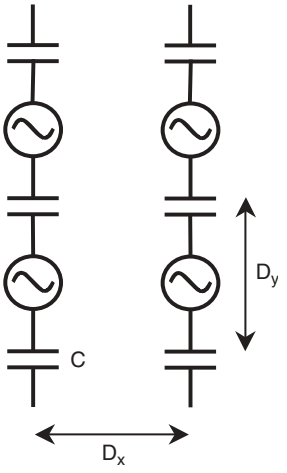


Figure 12.1 Schematic circuit model of current sheet array.

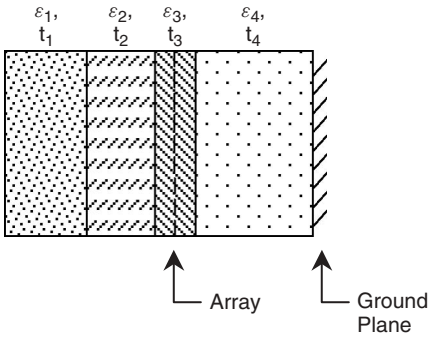


Figure 12.2 Dielectric profile of example current sheet array.

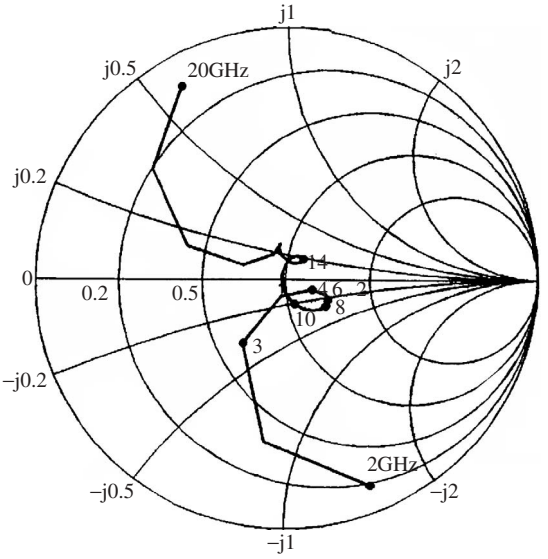


Figure 12.3 Predicted impedance of the example current sheet array in series with a short matching transmission-line section.

The CSA elements are nominally sized for operation at the highest frequency of interest, with a square array lattice near one-half wavelength at the high end of the band. This small element spacing eliminates in-band grating lobes and provides a large potential scan volume. The CSA is a very thin planar aperture with a thickness typically on the order of $\lambda/10$ at the lowest frequency. The exact thickness at the low end of the operating band is determined by the maximum allowable impedance mismatch and is dependent on the specific requirements. This makes it suitable for applications requiring conformal or low profile installations. Two orthogonal arrays can be used to provide dual polarization with very good cross-polarization isolation. The phase center of the array does not vary with frequency, which is an important feature in applications such as an array-fed reflector.

Industry-standard full-wave electromagnetic analysis tools such as Ansoft HFSS and CST Microwave Studio can be utilized for design purposes and performance predictions.

12.2.4 CSA Design Principles

As stated earlier, the performance of an isolated element cannot be used to predict performance of an array since mutual coupling can rarely be neglected. Therefore mutual coupling is utilized as a fundamental design parameter in the CSA approach and the methodology is focused on the aperture rather than on the isolated elements. Features and dimensions that create undamped resonances are avoided. Away from resonance, the impedance properties of elements of a wideband array design vary relatively slowly with frequency. These gradual variations can be played off each other to provide an overall uniform impedance match across a very wide bandwidth. It is important to note that the bandwidth of the combination can be greater than the bandwidth of any of the individual constituent elements. The CSA addresses impedance matching by deliberately introducing design elements to counteract the natural resonant behavior of dipole arrays. Implementation of the CSA has required a considerable amount of both empirical and analytic effort to determine the practical constraints on the design.

The principal features of the CSA design are a periodic lattice selected to avoid grating lobes, a ground-plane spacing selected to avoid the half wavelength radiation null at the upper end of the band, and a dielectric profile and interelement coupling selected using wideband matching techniques. The design is not a linear process with each parameter fixed in turn. Instead, all of the features must be chosen to be compatible. Understanding the interactions between these aspects of the design is a crucial part of obtaining wideband performance. Reducing the CSA to practice introduces several additional features that are also critical to obtaining good performance, such as finite array size, the feed structure, and physical implementation of the element coupling.

The lattice spacing of the array can be varied somewhat, but it is bounded by an extremely important constraint. If the lattice spacing exceeds a half-wavelength, grating lobes can enter real space for wide scan angles. This constraint turns out in practice to be a determining factor in the overall design. Many factors in implementation become more difficult as the elements become physically smaller. This places a practical limit on the upper frequency of operation.

12.2.5 Wideband Impedance Matching

The theoretical basis for the wideband impedance matching techniques used in the CSA design can be found in the text by Munk [8]. A brief summary of the basic principles is given below to provide an understanding of the equivalent circuit and the relevant impedance terms that contribute to the active impedance of the array. Certain assumptions are made in the basic theoretical development of this equivalent circuit model in order to reduce the complexity of the problem. Among these are that the elements are single polarized, the planar array is infinite in two directions, and the element spacing is such that no grating lobes exist in free space. The original analysis is conducted for a single scan angle, which is assumed to be broadside. The scan dependence of the ideal current sheet as theorized by Wheeler is known for the two principal planes. The proper use of dielectric layers has been shown for other periodic structures to compensate for these basic changes in array impedance with scan angle.

The fundamental behavior of a periodic structure can be represented as a plane wave spectral decomposition. This yields an equation with the same functional form as the transmission-line equations, and the equivalent circuit model for a periodic structure is derived from the congruence of the equations. Figure 12.4 shows an example array geometry and the equivalent circuit. The actual dielectric profile for a CSA design can be different from this example, but this profile shows the basic impedance matching techniques used in the theoretical design. The impedance at the antenna terminals is $Z_A = jX_{A1} + Z_1^- \| Z_1^+$, where the phased array in free space would have a terminal impedance $Z_{A0} = R_{A0} + jX_{A0}$, and in an infinite dielectric medium would have a terminal impedance $Z_{A1} = R_{A1} + jX_{A1}$.

As shown in Figure 12.5, the impedance of the ground plane viewed from the array position varies rapidly with frequency. At the reference frequency f_0, the array is $\lambda/4$ away from the ground plane, and Z_1^+ is infinite. As shown in Figure 12.6, the impedance of free space viewed through the dielectric layer also varies with frequency but remains on the circle passing through the center of the Smith chart and through the real impedance $Z_1^- = (2R_{A1})^2/2R_{A0}$ at f_0, the frequency where the dielectric thickness is $\lambda/4$. Note that the impedances are referenced to $2R_{A0}$, the center of the Smith chart. The parallel combination of Z_1^- and Z_1^+ is found by adding the pure susceptance of Y_1^+ to the complex admittance Y_1^-. As shown in Figure 12.7, this is easily performed on an admittance Smith chart by taking each point on the Z_1^- curve and moving along the circle passing through that point and the Smith chart origin a distance equal to the value of Z_1^+ for that frequency. Finally, the antenna terminal impedance is found as shown in Figure 12.8 by adding the

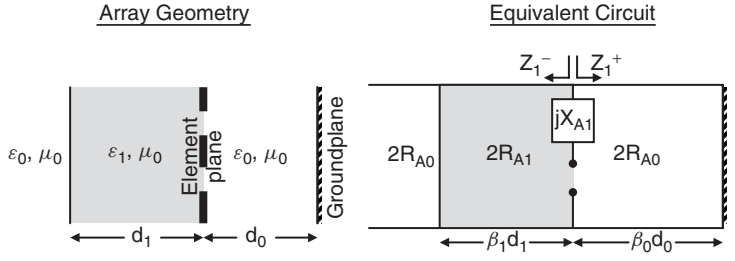


Figure 12.4 Example array geometry and equivalent circuit.

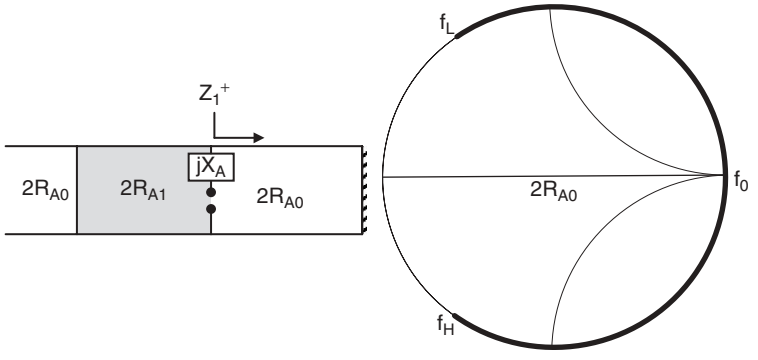


Figure 12.5 Impedance of the array ground plane versus frequency as viewed from the array terminals.

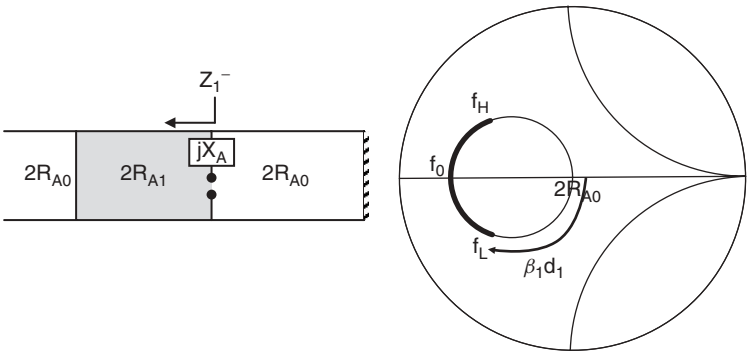


Figure 12.6 Impedance of free space versus frequency as viewed from the array terminals.

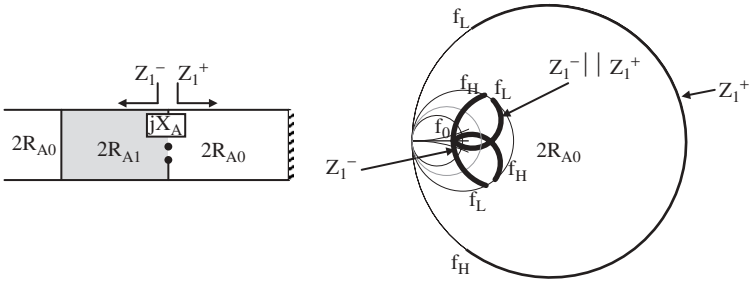


Figure 12.7 Impedance of the parallel combination of Z_1^- and Z_1^+ versus frequency as viewed from the array terminals.

reactive component jX_A to $Z_1^- \| Z_1^+$ to yield Z_A. Since jX_A is negative for $f < f_0$ and positive for $f > f_0$, the outer points of the $Z_1^- \| Z_1^+$ locus are drawn in toward the real axis, yielding a lower VSWR. This final Smith chart shows Z_A and how the combination has the potential to achieve a wideband match.

Dr. Munk's extensive experience with wideband FSS design indicated that an array of short, coupled dipoles is amenable to wideband matching. Recall that a half-wave dipole

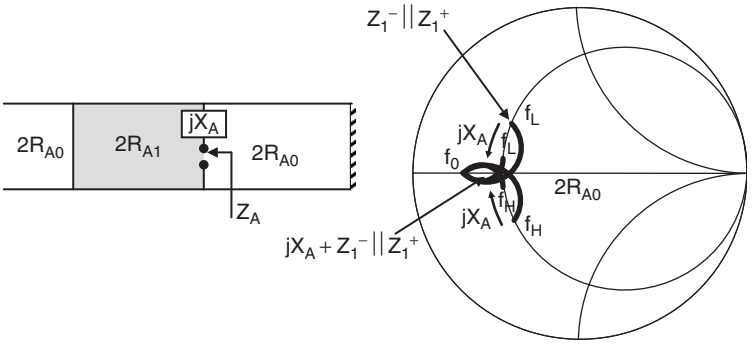


Figure 12.8 Final array terminal impedance versus frequency.

has a feed-point impedance that is purely resistive at the reference frequency. This is because the inductance due to the wire is balanced at that frequency by the terminal capacitance. A shorter dipole has less inductance; thus in order to maintain the condition of zero reactance at the reference frequency, the terminal capacitance must increase. This explains why capacitive rather than inductive coupling is needed between the ends of the dipole elements.

12.2.6 CSA Implementation

The promise of this concept spurred an intensive effort at Harris to implement the idealized current sheet array in hardware in order to validate the predicted performance. An early experimental article was a single-polarized 12-in. × 18-in. array, which achieved wideband gain performance similar to predictions and validated the theoretical basis of the CSA. The capacitive coupling between the elements was obtained using an interdigital capacitor as shown in Figure 12.9. Since the balanced feed of the dipole elements requires a balun, commercially available $0^\circ/180^\circ$ hybrids were used to provide the 180° phase shift between the element arms. Only one element was excited with the remaining elements resistively terminated with 100-Ω chip resistors. The measured embedded-element VSWR was below 3:1 from 3 to 18 GHz. This first experiment proved the fundamental promise of the wideband matching approach.

Subsequent development at Harris has substantially refined the CSA concept. This optimization has resulted in dual-polarized designs having an operational bandwidth of 9:1. Figure 12.10 shows a dual-polarized breadboard that operates over 2–18 GHz, and Figure 12.11 shows the feed network used for this breadboard. The aperture size is 22 in. × 22 in. with a total of 2664 elements, of which only 64 dual-polarized elements located in the center of the array are connectorized. The remaining elements are terminated using 100-Ω chip resistors. The large array size demonstrated that an array of this size could be manufactured as a single unit. One or two rows of resistively terminated elements must be placed around the perimeter of the active region in order to minimize edge effects. The total resulting edge termination region for the CSA to operate efficiently is similar to that of an array of conventional elements.

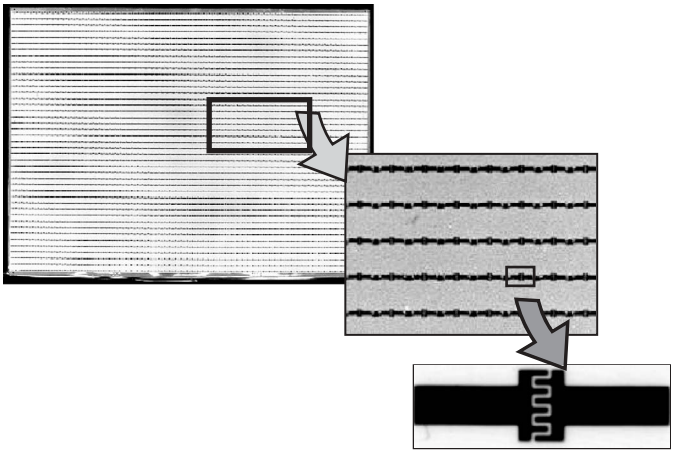


Figure 12.9 An early single-polarization CSA breadboard.

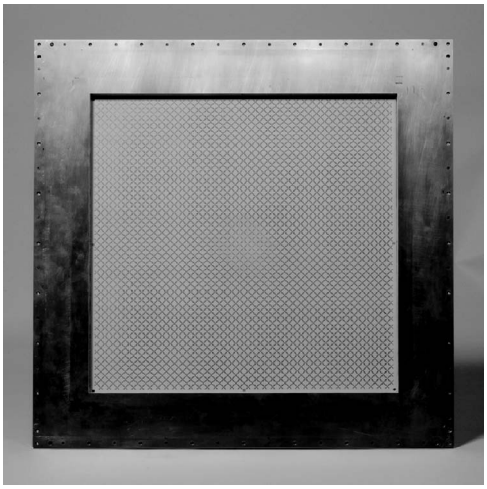


Figure 12.10 Dual-polarization 2–18-GHz CSA breadboard.

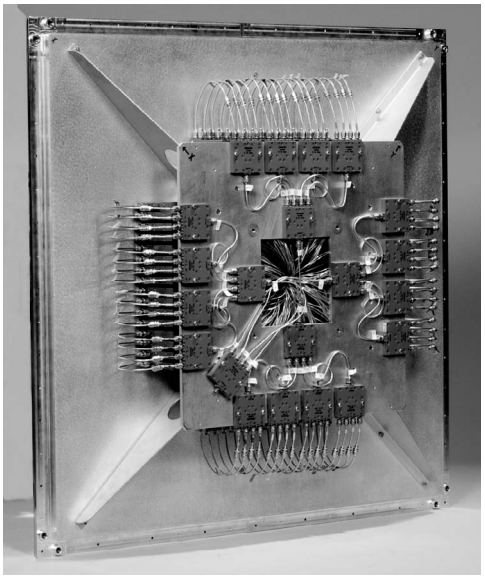


Figure 12.11 Feed network for dual-polarization 2–18-GHz CSA breadboard.

The measured gain of the 8×8 array is shown in Figure 12.12. Since the CSA is an aperture antenna, the maximum theoretical directivity is calculated using the well known equation

$$D_{\max} = \frac{4\pi A}{\lambda^2} \cos(\theta)$$

where A is the physical area of the active region of the array and θ is the scan angle away from broadside. The low end of the frequency band is determined by the ability to

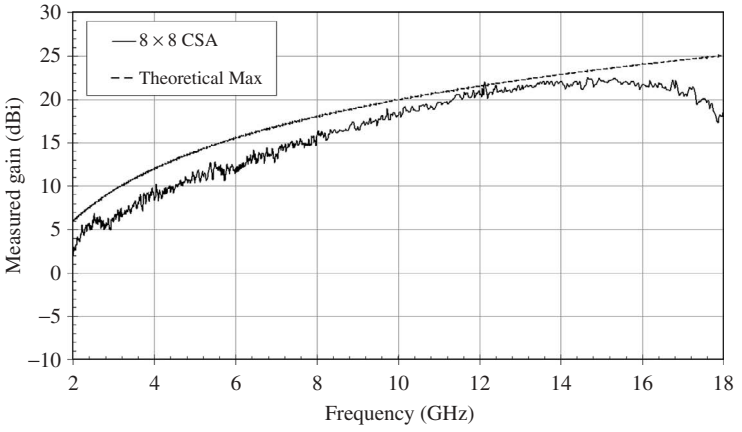


Figure 12.12 Measured gain of 8 × 8 CSA active region.

match the element as the distance to the ground plane becomes smaller, and the high end of the frequency band is limited by the frequency at which the element spacing becomes greater than one-half wavelength. There will also be a radiation null at broadside when the element spacing above the ground plane is one-half wavelength. Multiple design iterations have shown that bandwidth can be traded for efficiency depending on the specific application and requirements.

Measured embedded-element impedance and embedded-element VSWR of a typical element in this array are shown in Figure 12.13. Measurements were referenced to the difference port of the commercially available $0^\circ/180^\circ$ hybrid used to feed the array elements. The resistive component of the complex impedance is fairly constant and near 50 Ω over the operating band, while the reactive component is constant and near 0 Ω over the operating band. Note that an impedance of 50 Ω referenced to the hybrid port corresponds to an impedance of 100 Ω across the antenna terminals, since there is a

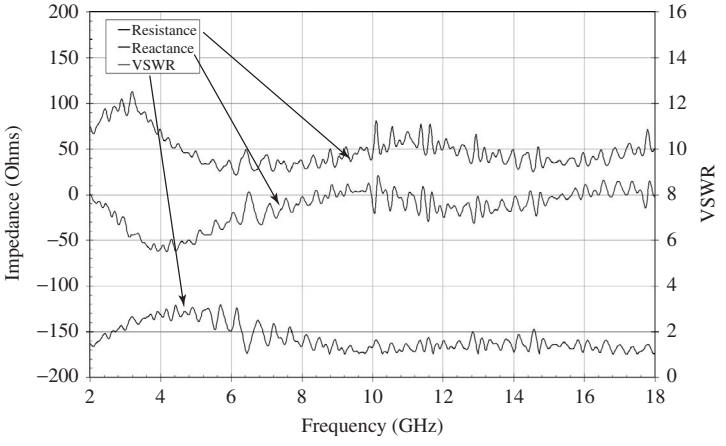


Figure 12.13 Measured embedded-element impedance and VSWR of a centrally located element of 8 × 8 active region.

virtual ground at the center of the feed point. This relationship has been verified using a measurement technique described in the literature [9].

This breadboard exhibited well behaved radiation patterns at normal incidence and at scan. The element spacing of this particular design was limited by manufacturing considerations. This limited the scan volume at the high end of the band. Several patterns of the 64-element array at broadside and also at 35° scan are shown in Figures 12.14 and 12.15. The beam scanning was accomplished using true time delays. Other CSA designs have been shown to scan over a 50° scan volume without grating lobes and pattern distortions up to the highest operating frequency.

Measurements of the CSA have consistently demonstrated very low cross-polarization levels. This is due to the inherent isolation of the orthogonal elements and to the use of a symmetric and controlled feed structure that eliminates common-mode currents. Figure 12.16 shows mutual coupling measurements for various elements of a 4 × 4 array from 2 to 18 GHz. Note that the cross-polarization isolation is greater than approximately 40 dB over the entire operating band. The most significant coupling across the aperture appears between adjacent elements, which are intentionally coupled using capacitors in the line of the element. Coupling decreases rapidly between adjacent rows of elements.

12.2.7 Robust Element Feed

As mentioned earlier, the CSA elements require a balanced line feed just as any other dipole element. For each dipole element, two coaxial cables are fed through the ground plane and lower foam layer up to the thin substrate layer containing the printed elements. The center conductor of the cables passes through the thin substrate layer and is connected to the dipole as shown in Figure 12.17. The cables are connected to a commercially available 0°/180° hybrid power combiner that serves as a balun.

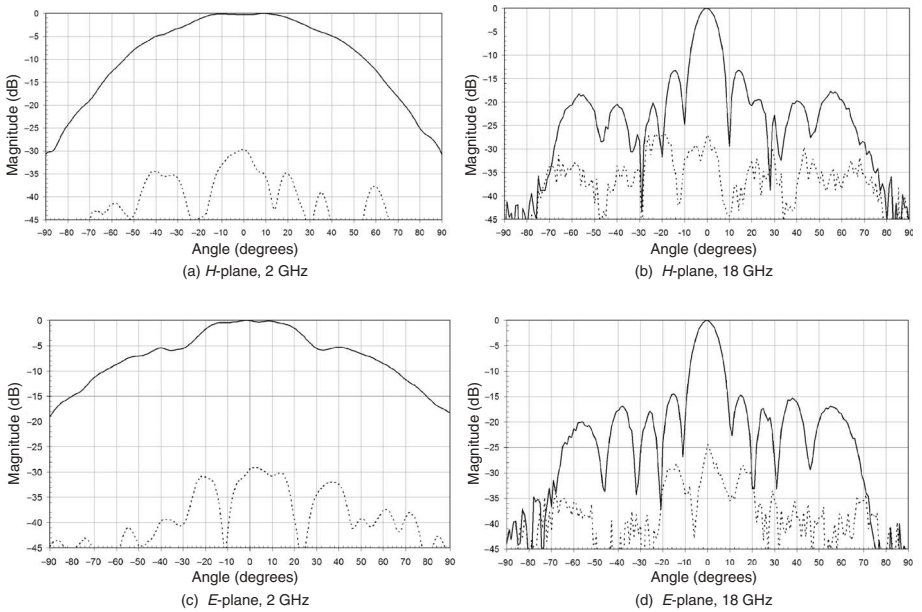


Figure 12.14 Broadside pattern of 8 × 8 CSA at 2 and 18 GHz.

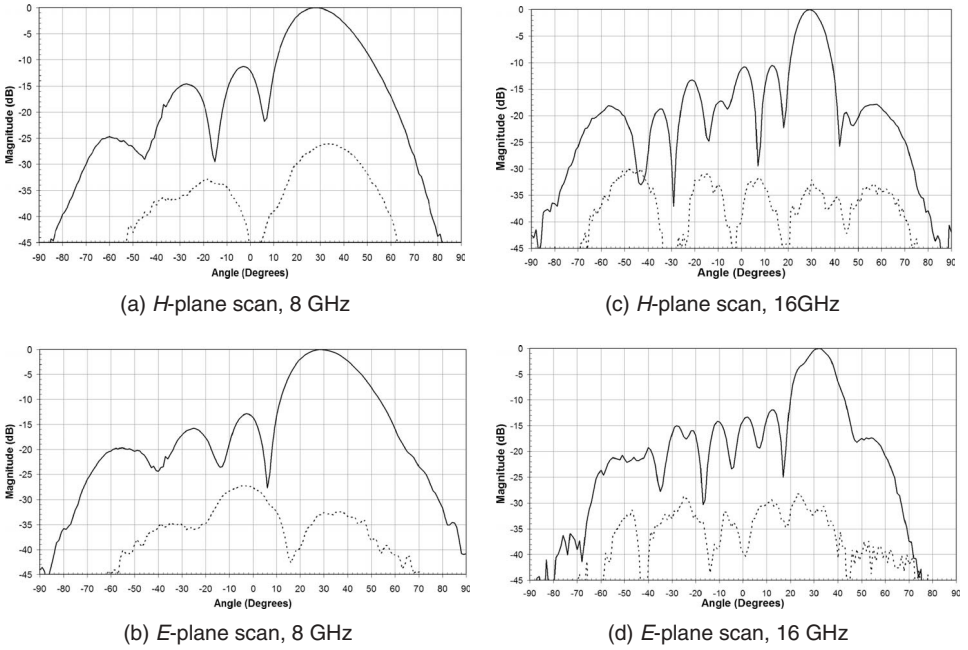


Figure 12.15 Scanned beam pattern of 8×8 CSA at 8 and 16 GHz.

Breadboard tests revealed that a critical aspect of this feed is the proper grounding of the two coaxial cables feeding up through the foam to the dipole elements. If these cables are not properly grounded, common-mode currents cause anomalies in the mutual coupling and surface waves, which appear in the radiation patterns, especially the cross-polarization patterns.

The solution to the common-mode currents traveling on the feed lines was the development of the feed organizer [9] shown in Figure 12.18. This simple device provides an electrical connection between the outer ground conductors of the coaxial cables as well as to the array ground plane. Feed organizers and cables are assembled separately from the array and inserted later in the manufacturing process. The use of machined parts to route the feed cables to etched components enables very repeatable antenna performance.

12.2.8 Beam-Former Development

All phased arrays require a beam former to combine the active element patterns into a useful beam. Commercially available components can be used to feed the array, but this results in a total height much greater than the antenna aperture itself. The total height from the ground plane to the CSA element layer is nominally $\lambda/10$ at the low end of the band. Custom components are required in order to exploit the inherently thin nature of the CSA aperture.

The small element spacing of the CSA provides a large scan volume but also limits the available area behind each element, where feed electronics can be located. Considerable effort has been placed on providing a balanced feed to each dipole element. A monolithic microwave integrated circuit (MMIC) has been developed that combines a balun, low

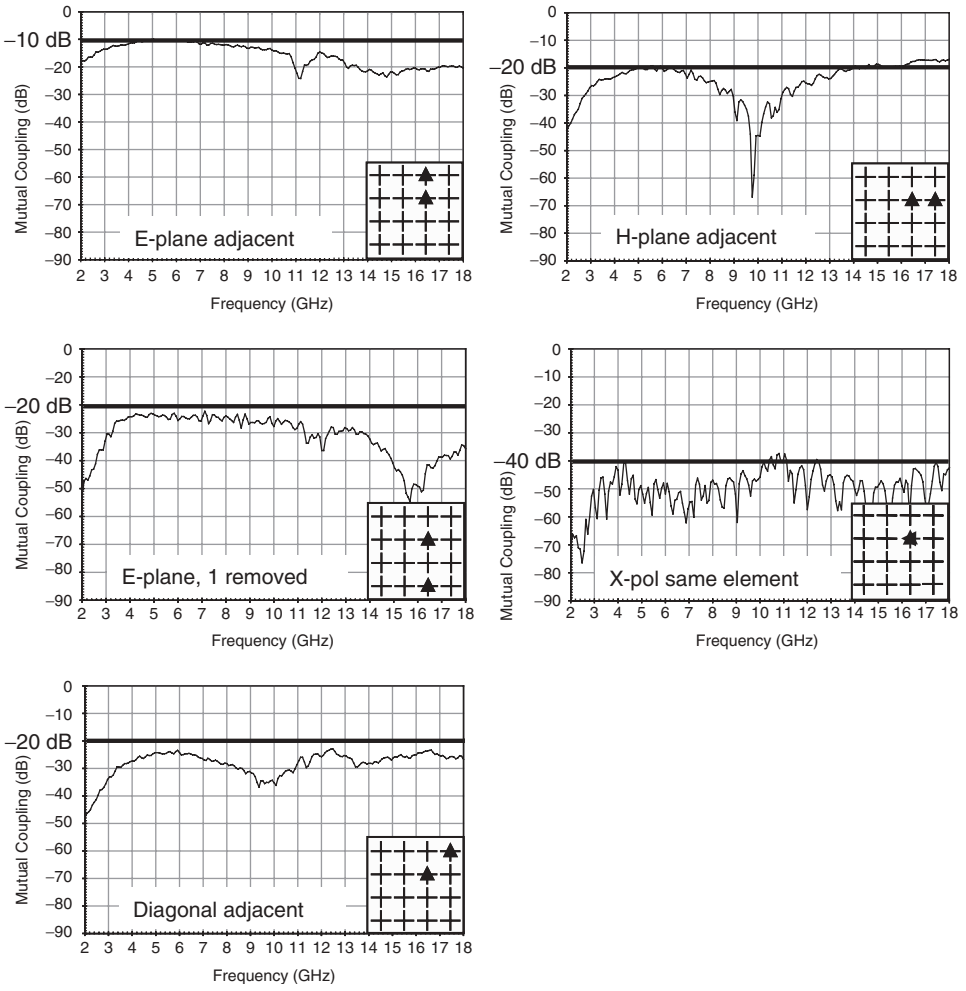


Figure 12.16 Measured coupling between adjacent elements of 4 × 4 CSA.

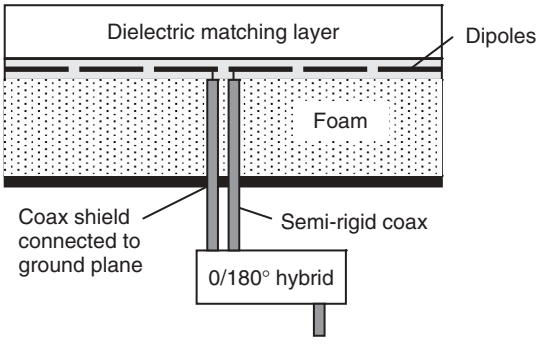


Figure 12.17 CSA element feed using coaxial lines and hybrid.

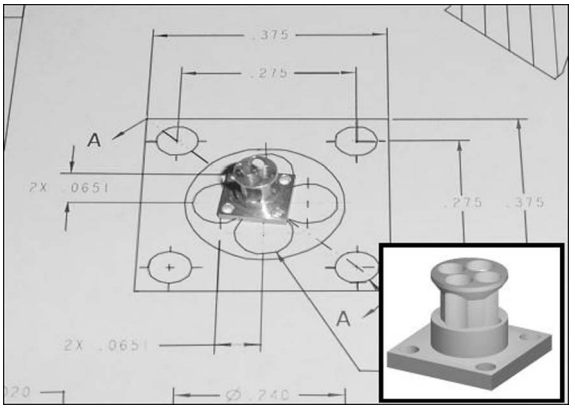


Figure 12.18 CSA feed organizer developed to eliminate common-mode feed currents.

noise amplifier, and time delay unit. Measured results show that this device has the required bandwidth to operate over the entire CSA frequency range. This active balun can provide the building block for a very thin 2–18-GHz phased array.

12.2.9 Electromagnetic Analysis

Numerous electromagnetic analysis software tools have been used to predict the wideband performance of the CSA and to provide insight into its operation. The fundamental wideband nature of capacitively coupled dipoles on dielectric layers was confirmed using PMM, which is a moment-method wire code for infinite arrays. Early breadboards were based on designs optimized using this tool and good agreement was obtained between PMM analysis and gain measurements. The primary limitations of PMM are that it uses a voltage gap feed and a bulk coupling capacitor to connect the elements.

Other commercially available electromagnetic design tools such as Ansoft HFSS, Zeland IE3D, MicroStripes, and CST Microwave Studio have been used to complement PMM analysis with detailed three-dimensional (3D) models of the CSA for finite and infinite arrays. The use of these tools enable nonidealized factors such as finite element width, capacitor implementation, and feed details to be included in the performance predictions. It has been found that single-polarization models are sufficient to predict broadside performance of the antenna due to extremely high cross-polarization isolation. The level of agreement between these EM solvers and measurement of an early CSA breadboard is shown in Figure 12.19. Computer analysis has consistently shown that the embedded-and active-element impedances are similar for the CSA.

12.2.10 Small, Finite Low Frequency Arrays

After the design was proven to operate over the 2–18-GHz band, the CSA was frequency-scaled to operate in the VHF/UHF band. An array was designed for 110–1000 MHz to fit inside an 18-in. × 36-in. × 16-in. metal cavity. This electrically small array, shown in Figure 12.20, consisted of 18 dual-polarized radiating elements arranged in a 3 × 6 array and required a new edge termination technique in which the resistively terminated elements were located on the vertical side walls. These terminated

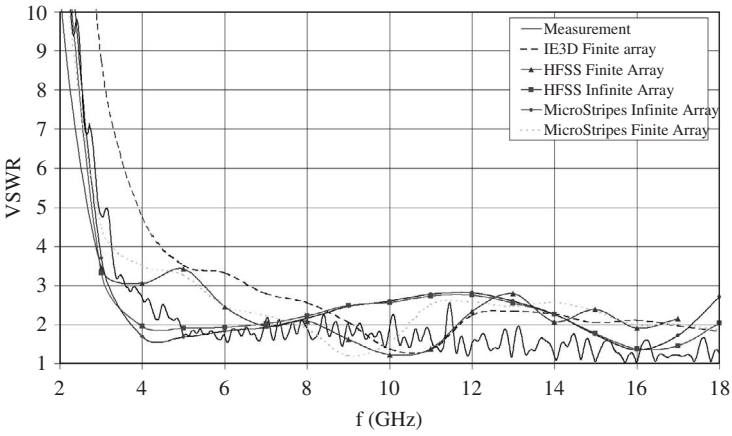


Figure 12.19 Predicted VSWR using infinite and finite arrays compared with measurements for an embedded element in a 28 × 28 CSA.

Figure 12.20 CSA breadboard designed for 110-MHz to 1-GHz operation.

elements did not contribute to the radiating aperture since ferrite absorbing tiles were placed on the cavity walls down to the antenna ground plane.

This array also utilized commercially available components for the beam-former assembly. For small, low frequency arrays such as this, the beam-former design is not a difficult task due to the increased volume beneath the unit cell and the small number of antenna elements. The elements parallel to the 36-in. cavity dimension are designated as H-pol and the elements parallel to the 18-in. cavity dimension are designated as V-pol. The overall thickness of this array was approximately 10 inches including the beam-former network.

The measured H-pol array gain at broadside compared to the theoretical maximum directivity of the planar aperture is shown in Figure 12.21. Note that the antenna efficiency varies from 30% to 50% over the 100-MHz to 1-GHz band. The low frequency

performance is limited by the size of the cavity, which acts as a waveguide below cut-off for frequencies below 150 MHz. Measured patterns for the H-pol array, shown in Figures 12.22 and 12.23, show well behaved performance across the band. The VSWR for the H-pol and V-pol arrays, shown in Figure 12.24, is below 3:1 across the band.

To determine the limits of the CSA concept, an even smaller breadboard was constructed consisting of four radiating elements with a total size of 12 in. × 12 in. × 6 in. This breadboard, shown in Figure 12.25, utilizes the same element design as the above array but is reduced to only four radiating elements in a 2 × 2 configuration. The five visible sides are conventional thin substrates with the interior volume occupied by a low density, lightweight foam. The weight of this breadboard is approximately 10 pounds, which is primarily determined by the beam-former components and mounting hardware.

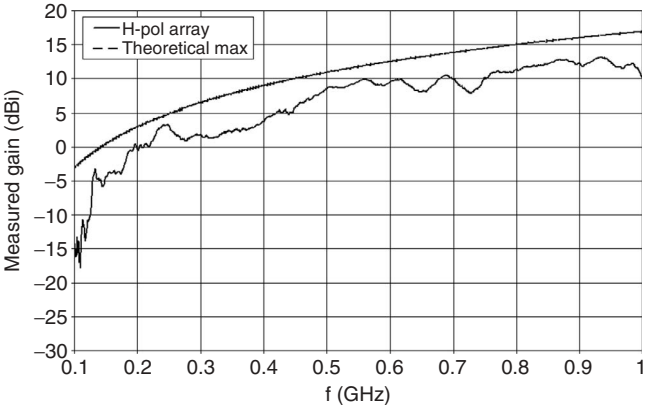


Figure 12.21 Measured broadside gain of the H-pol array compared to the theoretical maximum gain.

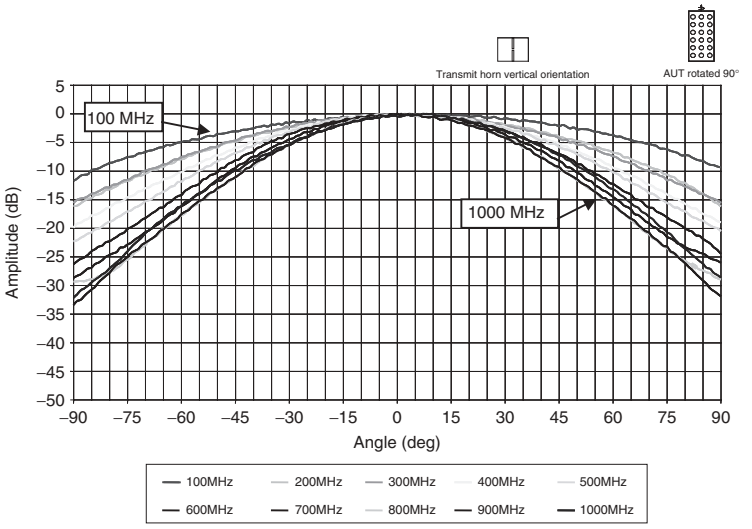


Figure 12.22 Measured H-plane antenna patterns for H-pol array.

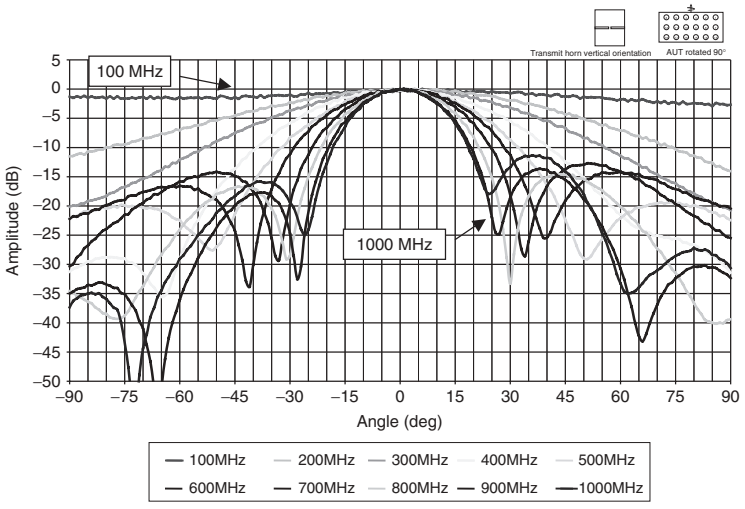


Figure 12.23 Measured E-plane antenna patterns for H-pol array.

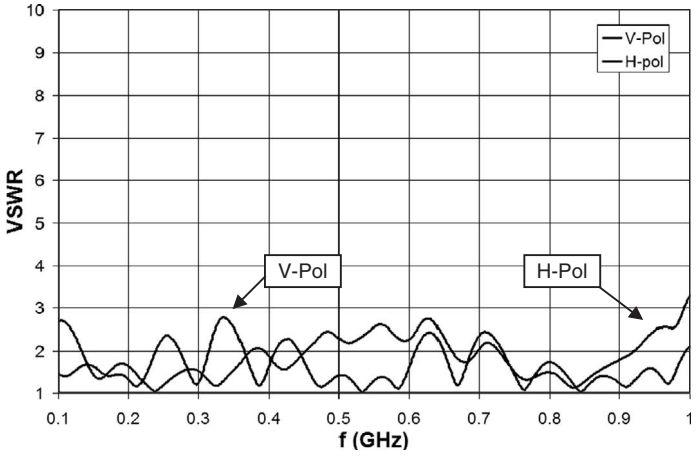


Figure 12.24 Measured VSWR of the V-pol and H-pol arrays.

The gain of this breadboard is shown in Figure 12.26 and shows an operating bandwidth of 150 MHz to 1 GHz, which is remarkable for an array of this size. The measured VSWR is shown in Figure 12.27. The small finite CSA arrays radiate largely from the top surface. The nonradiating vertical elements provide a direct connection for the currents to terminate into the ground plane. This method of edge termination is a result of the intentional mutual coupling used in the CSA approach and is not possible in conventional arrays.

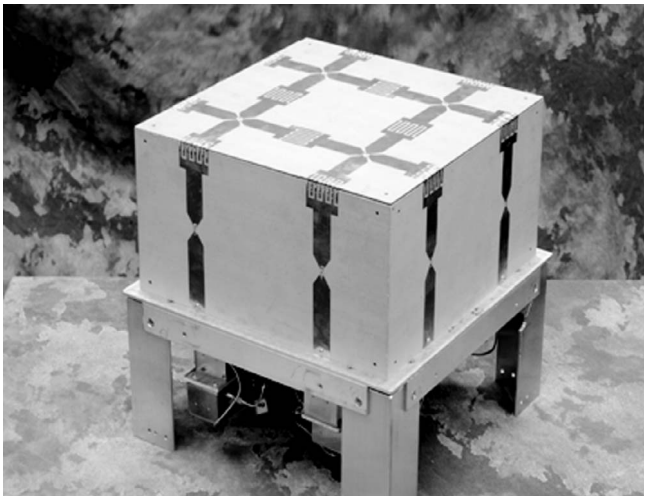


Figure 12.25 A 2 × 2 CSA breadboard designed for 150 MHz to 1 GHz.

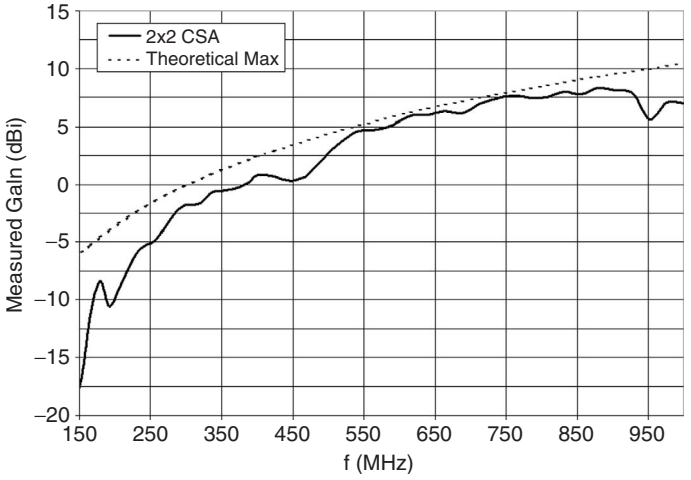


Figure 12.26 Measured broadside gain of the 2 × 2 CSA compared to the theoretical maximum gain.

12.2.11 Summary

A fundamentally different approach to wideband array design has been introduced and extensively validated with measured hardware. Conventional array design attempts to minimize mutual coupling between elements designed in isolation. This approach intentionally couples the array elements to provide wideband operation and also results in a low profile array due to the use of printed dipole elements. The dielectric layers placed above the array to enhance bandwidth can also serve as a radome to protect the elements.

The CSA was initially conceived to solve a 9:1 bandwidth problem. The design has proved to be scalable across frequency. Lower frequency versions have demonstrated a

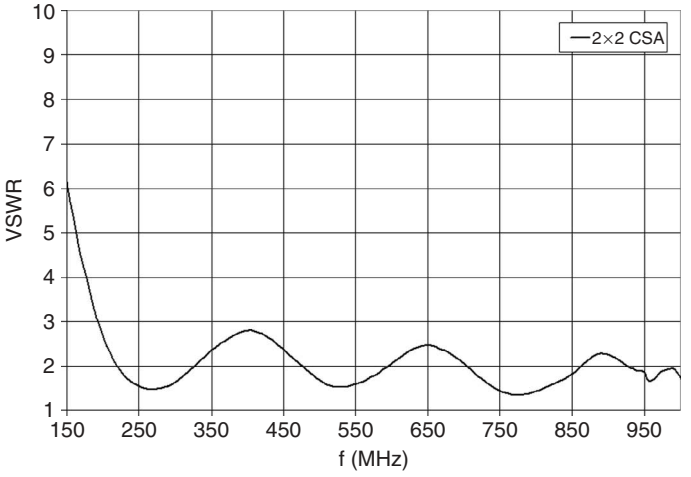


Figure 12.27 Measured VSWR of the 2 × 2 CSA.

usable bandwidth of a decade. One of the remarkable features of the CSA is its ability to maintain performance for finite arrays as small as 2 × 2 active elements. Since the initial conceptual description of the CSA, more than 20 breadboards have been constructed and measured covering frequencies ranging from 100 MHz to 18 GHz. The development of this technology has fundamentally altered the design approach for achieving wideband performance using a planar phased array.

12.3 VIVALDI ANTENNA ARRAYS

Current designs of Vivaldi antenna arrays can yield VSWR < 2 for bandwidths of 10:1 or more. Nearly ideal element patterns ($\sim\cos\theta$) are obtained in all scan planes for θ as great as 50°–60°. Properly designed and fabricated Vivaldi antenna arrays do not require lossy materials to achieve this performance or to suppress anomalous behavior. Furthermore, the Vivaldi antenna incorporates a balun into its basic structure, making it compatible with commonly used printed transmission lines, such as microstrip and stripline.

Throughout their development, Vivaldi-like antennas have been referred to by many names: notch or flared notch, tapered slot antenna (TSA), endfire slotline antennas, and variations of these names. The name Vivaldi was associated with an exponentially flared slotline antenna by Gibson in his 1979 paper [11]. Since that time, the name has come to be associated with any endfire TSA that resembles Gibson's original structure. Lewis et al. [12] suggested wide bandwidth operation of Vivaldi-like antennas in an array environment.

Development of operational Vivaldi antenna arrays progressed very slowly until the 1990s, when computer capabilities and numerical analysis techniques were able to accurately characterize realistic antenna structures in large array environments. Use of periodic boundary conditions provided full-wave treatment of the critical coupling effects between densely packed elements of the array [13–16]; more than 100 elements per square wavelength at the lower operating frequencies. It is now understood that a successful array

design for wide bandwidth and electronic scanning depends strongly on modeling and management of coupling between elements of the array.

12.3.1 Background

When used as single antennas, TSAs usually are a few wavelengths long and the slot opens to a width of one wavelength or more. The antenna behaves like an endfire traveling wave radiator [17]. Small arrays operating over a few percent bandwidth were found to be useful for millimeter wave spatial power combining [18]. Vivaldi antennas do not include any resonant structures, so manufacturing tolerances do not have a strong influence on the operating frequency or radiation of the antennas. Recent interest in imaging radio telescopes has spurred studies of wideband Vivaldi arrays in the focal plane of large reflectors [19].

12.3.1.1 Some Examples Large aperture arrays that must operate simultaneously over wide bandwidths or at widely separated frequency bands with electronic scanning demand performance that can be provided by Vivaldi antenna arrays. Some examples of Vivaldi antenna arrays that have been fabricated are shown in Figure 12.28. Most Vivaldi antenna arrays are planar structures, but conformal arrays have been designed. Single- and dual-polarized arrays with singly and doubly curved aperture surfaces can be designed to work satisfactorily for sufficiently large radii of curvature.

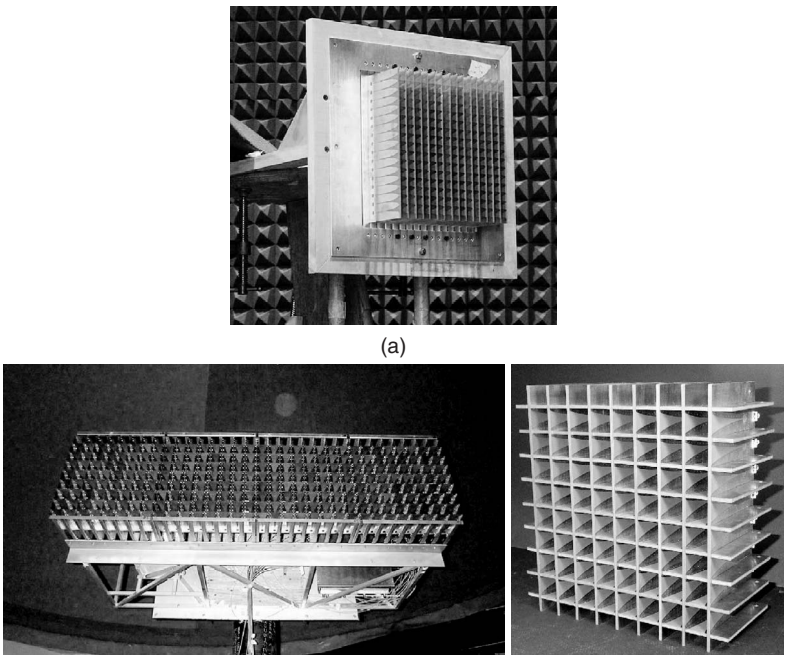


Figure 12.28 Vivaldi antenna arrays: (a) 16×15 single-polarized array, (b) $8 \times 32 \times 2$ dual-polarized array, and (c) $9 \times 8 \times 2$ dual-polarized array. (Photographs (a) and (c) courtesy of University of Massachusetts Center for Advanced Sensor and Communication Antennas; photograph (b) courtesy of L-3 Randtron Antenna Systems.)

Antenna arrays operating as apertures (as opposed to focal plane feeds for reflectors or lenses) must be at least a few wavelengths in size to produce useful beamwidths. Since array element spacing is slightly less than 0.5 wavelength at the highest frequency of operation, useful arrays covering two or more octaves will contain 30 or more elements along any direction for which beam forming and beam scanning are desired. If the number of elements in any linear direction is less than 15–20, the performance of individual elements and of the entire array can be affected by the truncation, especially at the lower frequencies of the operating band.

12.3.1.2 Polarization Considerations Vivaldi antenna elements radiate fields that are predominantly linearly polarized with the electric field vector at bore sight parallel to the plane of the Vivaldi antenna. In the principal planes, the cross polarization is quite small and is created by fabrication tolerances for stripline-fed antennas. Microstrip-fed antennas are not symmetrical, so they typically have higher cross-polarized radiation in the principal planes. Vivaldi antennas are known to produce elliptical polarization in the intercardinal planes [20], with the lowest axial ratio typically observed in the diagonal plane. Figure 12.29 shows the ratio $|E_y/E_x|$ in the diagonal plane of a large Vivaldi antenna array (an infinite array is used for this simulation). The VSWR of the array is mostly less than 2:1 over the frequency range and scan angles shown. The cross polarization is mostly below -20 dB up to 30° from broadside, but degrades to -15 dB or worse for scan angles greater than 45° in the diagonal plane. McGrath et al. [21] found that the cross-polarized radiation is higher in the diagonal plane of single-polarized Vivaldi arrays than dual-polarized arrays.

The most common array configuration to obtain dual polarization is the "egg crate" arrangement with Vivaldi antennas on the four sides of a square unit cell, Figure 12.30a. In this arrangement, the phase centers of the horizontally and vertically polarized elements

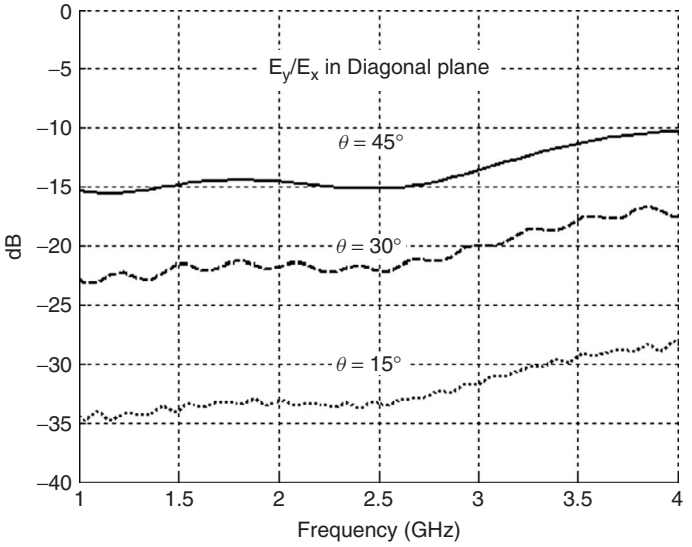


Figure 12.29 Ratio of E_y to E_x for scan in the diagonal plane of an infinite, dual-polarized Vivaldi array when only the x-polarized elements are excited (y-polarized elements are terminated). Scan angles of 15°, 30°, and 45° from broadside are shown.

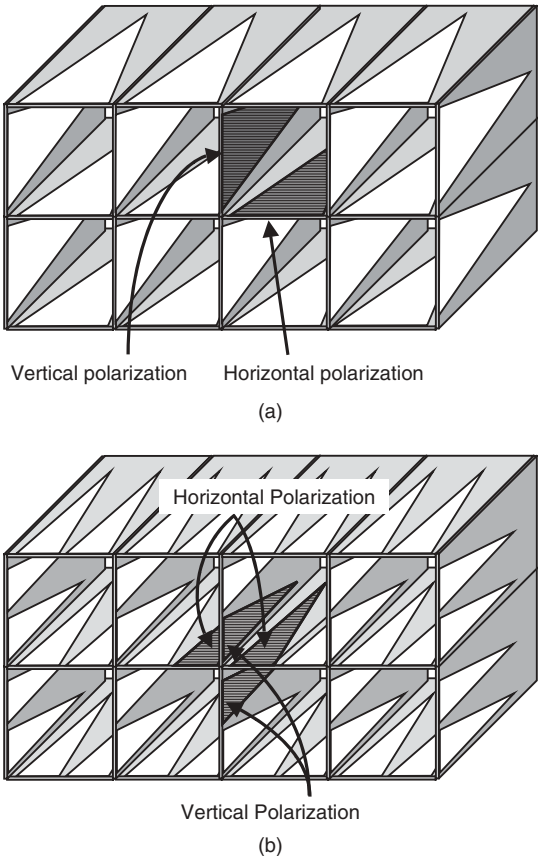


Figure 12.30 Two ways to obtain dual polarization. (a) Each unit cell of the array is comprised of one horizontally polarized radiator and one vertically polarized radiator. The phase centers for H and V polarization are not coincident. (b) Each unit cell is comprised of a pair of horizontally polarized radiators and a pair of vertically polarized radiators. The two horizontally (vertically) polarized elements are fed in phase from a single input connection. The phase centers are more nearly coincident, but the smaller radiator size limits bandwidth.

are not coincident, which may necessitate a more complicated feed network to achieve dual-polarized scanning. An alternative configuration to achieve dual-polarized radiation is shown in Figure 12.30b.

12.3.2 Design and Fabrication Considerations

Vivaldi antenna arrays are usually constructed with printed circuit techniques, employing stripline or microstrip input to the elements; see Figure 12.31. The Vivaldi antenna includes as part of each element a balun transition to feed the slotline in the required balanced mode. No external balun is required to operate with unbalanced transmission lines that are typically used for microwave feed networks. If desired, a balanced connection directly to the Vivaldi antenna element at the narrow portion of the slotline yields performance similar to that illustrated in these sections.

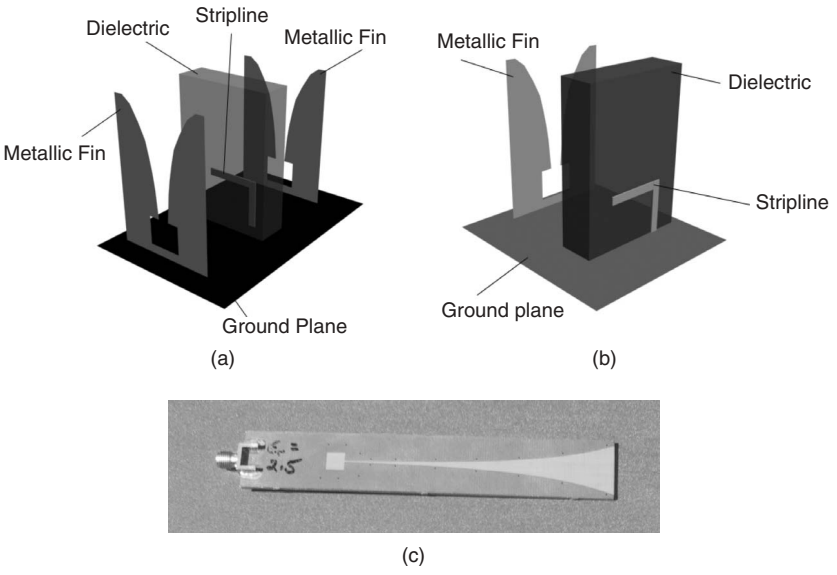


Figure 12.31 Exploded views of Vivaldi antenna elements: (a) stripline-fed antenna, (b) microstrip-fed antenna, and (c) photo of a stripline-fed antenna with coaxial connector. This element operates at 3–12 GHz in a large array.

12.3.2.1 Baluns for Vivaldi Antennas The most common balun structure used with Vivaldi antennas is adapted from Knorr's microstrip-to-slot transition [22]. A six-port balun also has been used by Kragalott et al. [23]. Figure 12.32 shows these two types of baluns. The stripline radial stub in Figure 12.32a can be replaced by a via connecting the stripline directly to the ground planes, but the reactance variation of the radial stub has been found to be beneficial for impedance matching over wide bandwidths.

12.3.2.2 Continuous Electrical Contact Gaps between the elements of Vivaldi antenna arrays cause undesirable resonances that disrupt the wideband performance of the arrays; see Figure 12.33. Impedance anomalies can occur also when a large array is comprised of subarrays, each subarray having contiguous elements but gaps between the subarrays [24]. Elements near the edges of the subarrays are affected more by the gaps than elements in the interior of the subarrays.

Single-polarized arrays should be fabricated with continuous metallization on each row of the array, and dual-polarized arrays should have electrical contact at all junctions of orthogonal antennas. This requirement makes it difficult to build wideband Vivaldi antenna arrays in a modular fashion, that is, with field replaceable modules comprised of one element and a T/R module. Rather, the Vivaldi antenna aperture must be a single unit or must have reliable means to maintain electrical continuity element-to-element. Partitioning the aperture into subarrays of connected elements, where the subarrays are not in direct contact, is an alternative to element-by-element modularity but, as noted previously, this also causes anomalies in the array performance.

The balanced antipodal Vivaldi antenna (BAVA) [25] is a variation of the traditional Vivaldi antenna that is inherently separated from its neighbors. When used in large

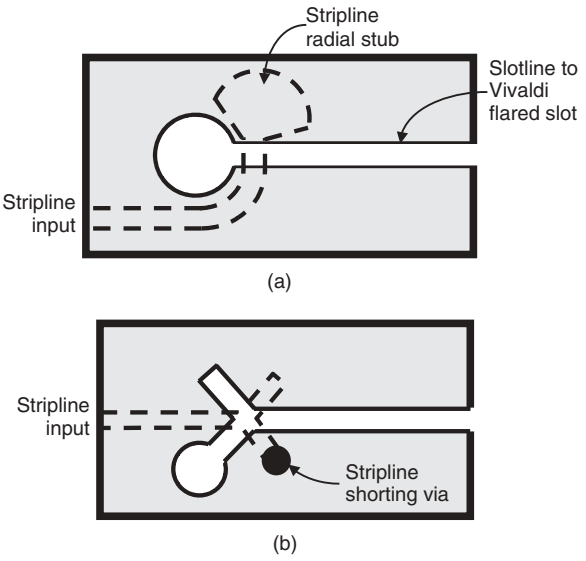


Figure 12.32 Two types of baluns that have been used successfully for stripline-fed and microstrip-fed Vivaldi antennas.

arrays, it also suffers from bandwidth-limiting resonances [26]. However, recent work has extended the useful bandwidth of BAVA arrays to nearly two octaves [27].

12.3.2.3 Element Spacing To achieve wide scan without grating lobes, array element spacing must not exceed $0.5\lambda_{hi}$, where λ_{hi} is the wavelength at the highest frequency of operation. Also, grating lobe theory suggests that scan performance is improved by reducing element spacing as much as possible. However, feed network complexity and T/R module count increase when element spacing decreases. Experience has shown that element spacing should not be greater than approximately $0.45\lambda_{hi}$. For large element spacings, the air-filled region between the elements supports waveguide modes that cause severe impedance anomalies. In addition, a parallel plate mode is responsible for one type of anomaly in single-polarized arrays with spacing greater than 0.5λ [28]. Modes of a square waveguide account for some of the anomalies in dual-polarized arrays with spacing greater than $0.5\lambda_{hi}$.

12.3.2.4 Vias to Suppress Resonance in Dielectric The dielectric-filled region between the conducting ground planes of stripline-fed Vivaldi antenna elements can support the dominant mode of a parallel plate or rectangular waveguide and this mode can be excited in dual-polarized arrays or by H-plane scan of single-polarized arrays. However, it is easy to suppress this mode by vias [29].

12.3.2.5 Size and Weight Vivaldi antennas radiate preferentially in the endfire direction, necessitating a finite depth for the antenna array. This depth has been a disadvantage of Vivaldi antenna arrays, but alternative array configurations that can provide wide bandwidth and wide scanning while occupying less volume have not yet reached similar levels of development and implementation. A dual-polarized Vivaldi array has

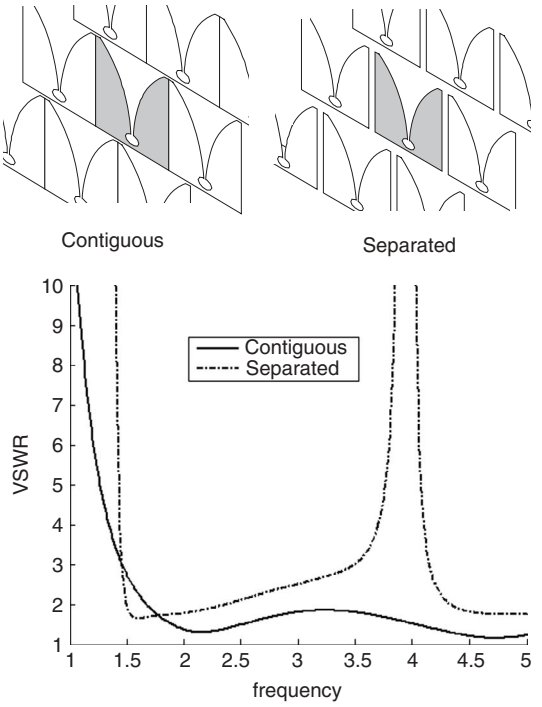


Figure 12.33 Vivaldi arrays with gaps separating adjacent elements have impedance anomalies that disrupt wideband performance. The example is computed for a single-polarized, infinite array of stripline-fed antennas. A VSWR spike near 4 GHz occurs when the gap resonates and radiates field strength sufficient to cancel the desired radiation from the Vivaldi slot.

achieved the performance shown in Figure 12.34 with a total depth of 0.23 wavelength at the lowest operating frequency.

When comparing alternatives for wide bandwidth antenna arrays, it is important to consider the total volume, weight, and complexity of the array. Vivaldi antenna arrays have the balun integrated into the radiating element; there is no need for additional circuitry that may add weight, volume, and loss to the array.

The weight of a Vivaldi antenna array is dominated by the dielectric substrate that comprises the elements. (Antenna mounting structures are not included in this estimate. The egg-crate structure of the dual-polarized array is very strong and requires little or no element-by-element support over moderately large distances.) The total length of a typical Vivaldi antenna array element operating over approximately three octaves is 4 times its width, which is approximately 0.45 wavelength at the highest operating frequency. Furthermore, the substrate thickness needed for good antenna performance is approximately 0.1 times the antenna width. Therefore the total volume of dielectric substrate in a Vivaldi antenna element is $0.03\lambda_{\text{hi}}^3 - 0.04\lambda_{\text{hi}}^3$. This can be used to estimate the weight of a Vivaldi antenna array before completing a detailed design.

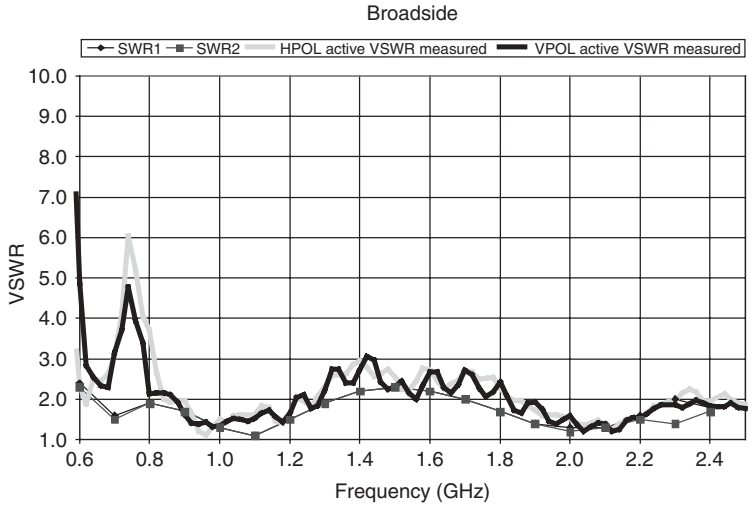


Figure 12.34 VSWR predicted (for infinite array) and measured (for central element in $8 \times 8 \times 2$ array) for dual-polarized array; array depth $= 0.23\lambda$ at lowest frequency; broadside beam. Peaks of measured VSWR near 0.8 GHz are likely due to truncation effects as discussed in Section 12.3.4. (From Ref. 30 with permission.)

12.3.3 Electrical Design in Infinite Arrays

Figure 12.35 shows the parameters that define the Vivaldi antenna element and array. All of these parameters influence antenna performance, but some general guidelines can be used to minimize the parameter space that must be searched to obtain a good design.

1. Element spacings a and b should be approximately $0.45\lambda_{\text{hi}}$.
2. Longer elements usually work better at low frequencies, yielding greater bandwidth.
3. Substrate thickness approximately 0.1 times element width usually yields good results.
4. The exponential opening rate, R_a, and cavity size, D_{sl}, strongly influence antenna performance.

Vivaldi array design requires full-wave numerical simulation of many geometries to obtain satisfactory performance over wide bandwidth and scan angles. Vivaldi antenna arrays that utilize the Knorr balun [22] as indicated in Figure 12.35 generally have input impedances that resemble the plots in Figure 12.36. The lower operating frequencies are limited by low input resistance and large variation of the reactance from capacitive at low frequencies to an inductive peak that must be controlled for good impedance match at low frequencies. Larger cavity sizes, D_{sl}, increase the antenna resistance at the lowest frequencies, but overly large cavities produce an undesirable overshoot as the resistance rises at lower frequencies and also causes large reactance. Higher permittivity substrates usually result in lower operating frequencies but can adversely affect high frequency performance [31].

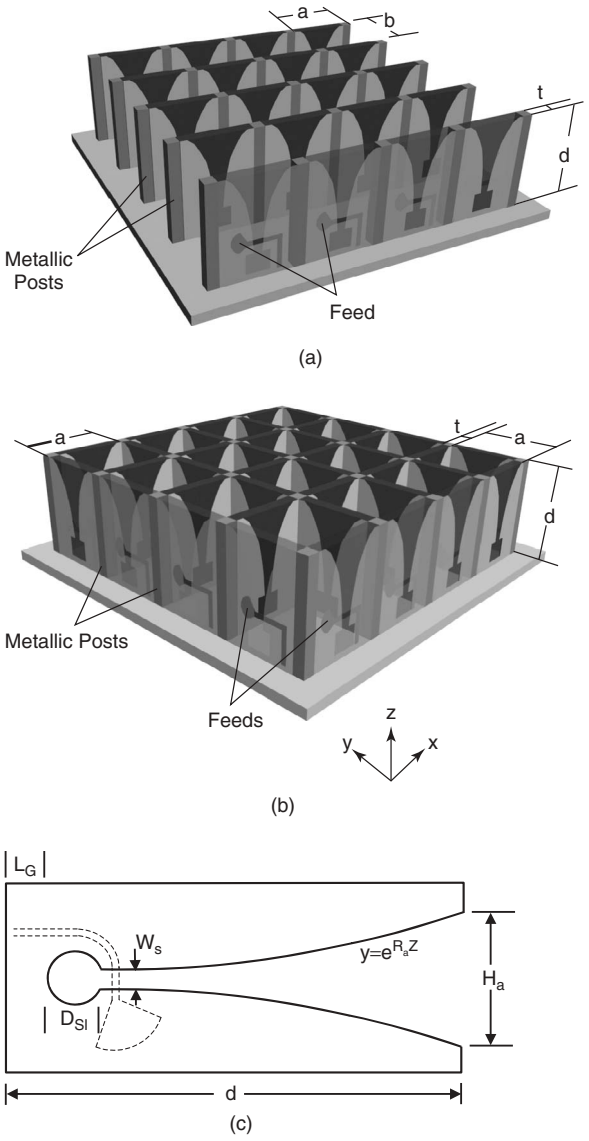


Figure 12.35 Parameters that describe Vivaldi antenna arrays: (a) single-polarized array parameters, (b) dual-polarized array parameters, and (c) design parameters of Vivaldi element.

The exponential flare of the Vivaldi slot is defined by $y = C_1 e^{R_a Z} + C_2$. Larger opening rates, R_a, improve low-frequency resistance but increase the variations of resistance and reactance throughout the operating band. Thus a compromise is usually required when choosing D_{sl} and R_a for wide bandwidth operation.

Like all phased arrays, the impedance of Vivaldi antenna arrays changes with scan angle. Simulated VSWR for three scan angles of a particular array design are shown in Figure 12.37. This array operates well over a bandwidth of two octaves for scan angles

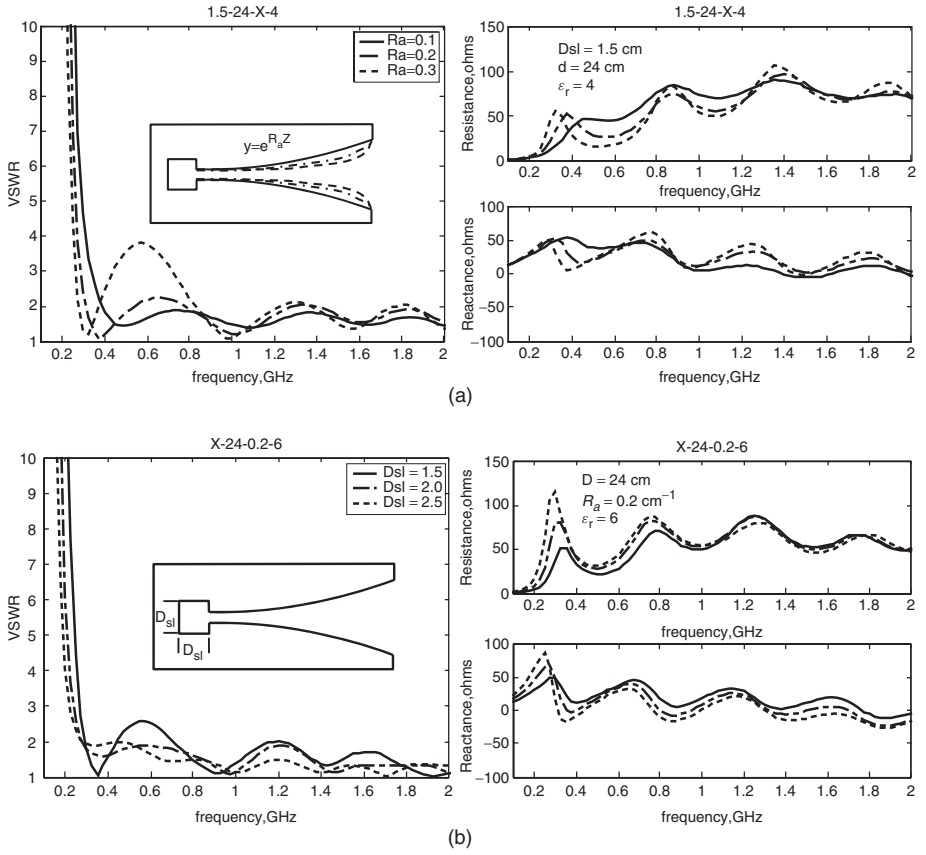


Figure 12.36 Variation of active input impedance. Vivaldi element in single-polarized infinite array. (a) Effect of exponential opening rate. (b) Effect of slotline cavity size. (From Ref. 33 with permission.)

to 45°. Using the standard formula $F(\theta, \phi) = [1 - |\Gamma(\theta, \phi)|^2] \cos\theta$, the element pattern can be evaluated from the infinite array reflection coefficient. Measured element gain for a central element of a single-polarized array is shown in Figure 12.38.

Design curves for a dual-polarized array resemble those of a single-polarized array; see Figure 12.39. The antenna designer can analyze a few element designs in an infinite array by using a full-wave electromagnetic simulator and then use these plots to guide element optimization for desired performance. The performance of a particular array designed in this fashion is shown in Figure 12.40.

12.3.4 Truncation Effects in Vivaldi Antenna Arrays

Since Vivaldi antenna arrays utilize coupling between neighboring elements to achieve wide bandwidth performance, truncation of the array to finite size affects elements near the array edges. Holter and Steyskal [32] explored the extent of truncation effects by analyzing finite arrays of various sizes and observing the input impedance as a function

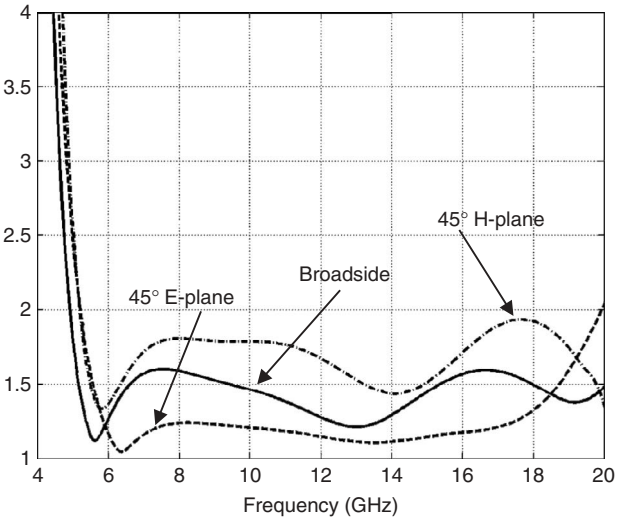


Figure 12.37 VSWR of an infinite dual-polarized Vivaldi array similar to Figure 12.28c at three scan angles. Only one polarization is excited for these results; the other is terminated. Element is stripline-fed; ε_r of substrate is 2.5. Element spacing is 0.45λ at 20 GHz.

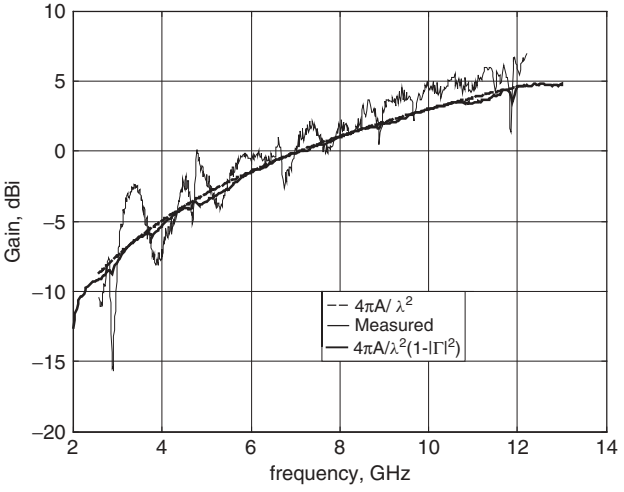


Figure 12.38 Element gain measured at broadside for a central element of a 16 × 15 array in Figure 12.28a. Truncation effects are evident in these measurements. Element spacing is λ/2 at 12 GHz; 16 elements = 20 cm. (From Ref. 33 with permission.)

of element distance from the array edge. They show that the element impedance approximately converges to the value in an infinite array for elements that are at least two wavelengths from the array edge. This conclusion is consistent with the rule of thumb often applied to narrow bandwidth arrays, where truncation effects are significant for the outer 3–4 rows of the array. At the upper edge of their operating bandwidth, Vivaldi

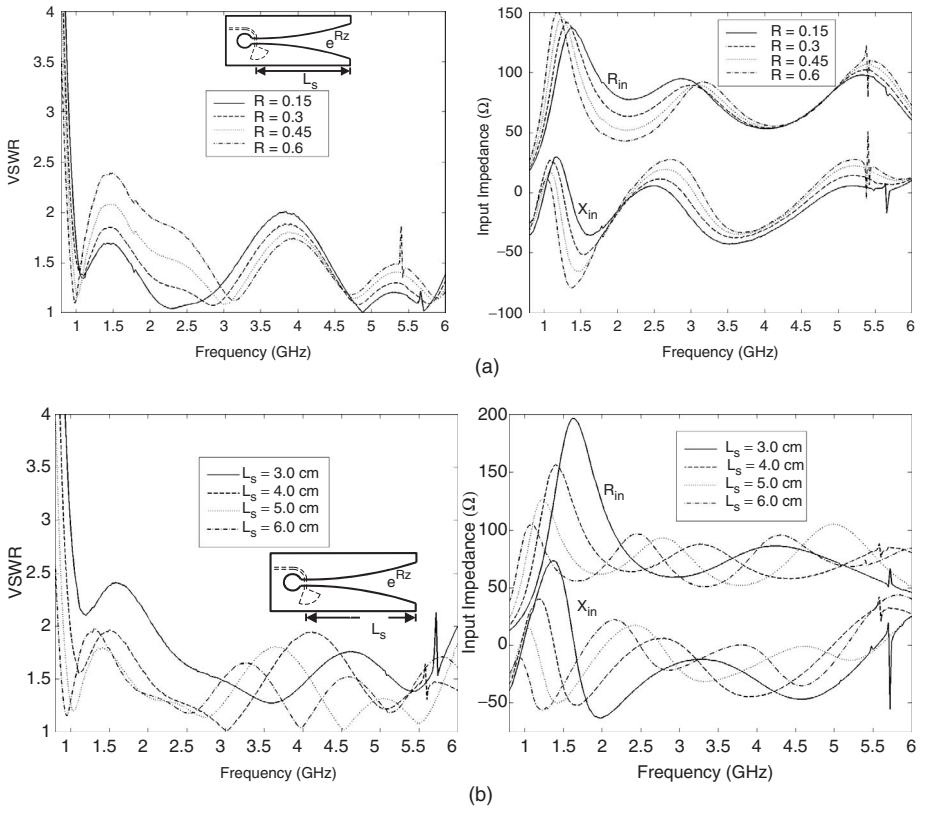


Figure 12.39 Effects of (a) opening rate R and (b) exponential flare length on active input impedance of infinite array at broadside scan. Dual-polarized, stripline-fed elements. Anomaly near 5.3 GHz limits upper operating frequency. (From Ref. 38, copyright © 2000 IEEE.)

arrays have element spacing approximately $\lambda/2$ and the truncation effects are limited to (approximately) the outer 3–4 elements. However, truncation effects at the lower end of the operating band, where element spacing may be 0.1λ or less, are significant 10–15 elements from the array edge.

Truncation effects in small arrays can cause severe mismatch of some array elements over portions of the desired operating band. Figure 12.41 shows the VSWR of two columns in an 8×8 array of Vivaldi antennas. The strong anomaly in the middle of the operating band occurs for a few of the interior elements, but not for edge elements, nor for most other interior elements. Redesign of the element geometry can reduce or eliminate this midband anomaly.

12.4 FRAGMENTED ARRAY ANTENNAS

12.4.1 Introduction

Students of antenna design are taught that one can estimate the gain of an array antenna by multiplying the pattern of a single element by the array factor. This simplistic formula

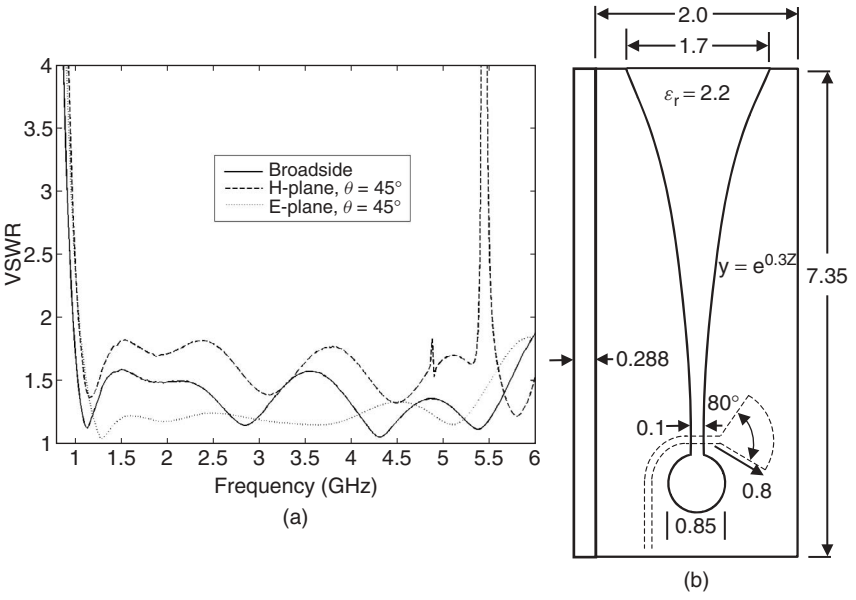


Figure 12.40 Simulated performance of dual-polarized array designed by using full-wave analysis and guided by design curves like those in Figure 12.39. (a) VSWR for three scan angles. (From Ref. 38, copyright © 2000 IEEE.) (b) Element dimensions in centimeters.

ignores mutual coupling between elements, which has traditionally bedeviled designers of phased arrays by introducing areas of scan blindness (combinations of frequency and scan angle for which the array is poorly matched). As the sophistication of numerical modeling codes has increased in concert with the availability of inexpensive parallel computing power, antenna designers have developed the ability to include the effects of mutual coupling in performance predictions. This in turn suggests the possibility of exploiting mutual coupling rather than avoiding it.

As GTRI was addressing this question in the mid-1990s, work was also being done on single feed planar antennas to maximize both gain and bandwidth. Consideration of canonical broadband antenna types such as spirals or bow ties (Figure 12.42) suggested that these radiating structures fail to achieve the theoretical limits of gain over a broad bandwidth (diffraction limited gain, approximated by $2\pi A/\lambda^2$ for a uniform current sheet without a ground plane radiating into both hemispheres). The question naturally arises: Are there other radiating structures that can improve the performance of traditional broadband antennas? Based on these considerations, GTRI decided to treat the allocated aperture as a blank canvas, using optimization routines to determine placement of conducting structures to most efficiently radiate energy over the broadest bandwidth. The result is a class of antennas we have dubbed "fragmented apertures," due to the typically complex distribution of conducting regions over the surface. Although the names sound similar, fragmented apertures are not fractal antennas and are not based on replication of geometry at multiple scales [35, 36].

Figure 12.43 shows several examples of early fragmented aperture antenna designs. These examples indicate the trade space of bandwidth versus realized gain for single

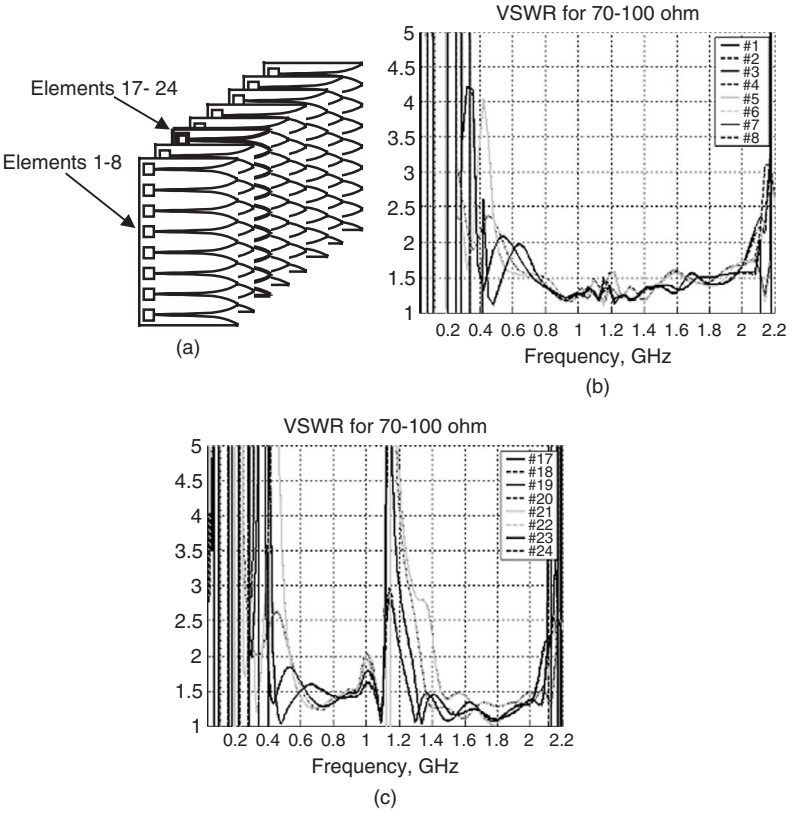


Figure 12.41 VSWR of elements in two columns of 8 × 8 array, broadside beam. (a) Locations of edge column (elements 1–8) and interior column (elements 17–24). (b) Edge column operates well at 0.5–2.0 GHz. (c) Most elements of third column are mismatched near 1.2 GHz. (Parts (a) and (c) from Ref. 34, copyright © 2003 IEEE.)

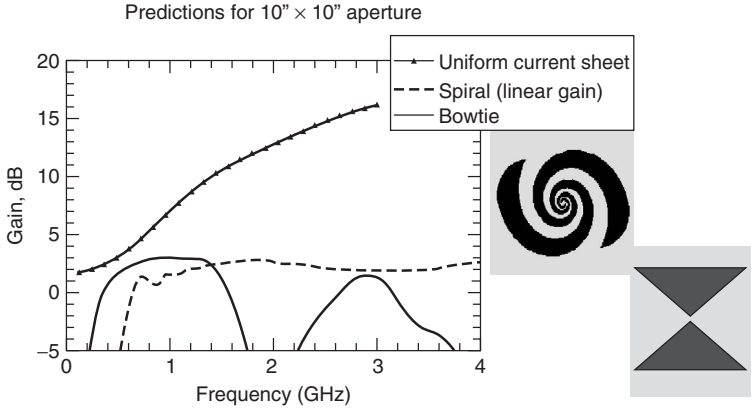


Figure 12.42 Gain available from two standard broadband antenna types compared to area gain theoretically available from a 10-inch square aperture (with no ground plane).

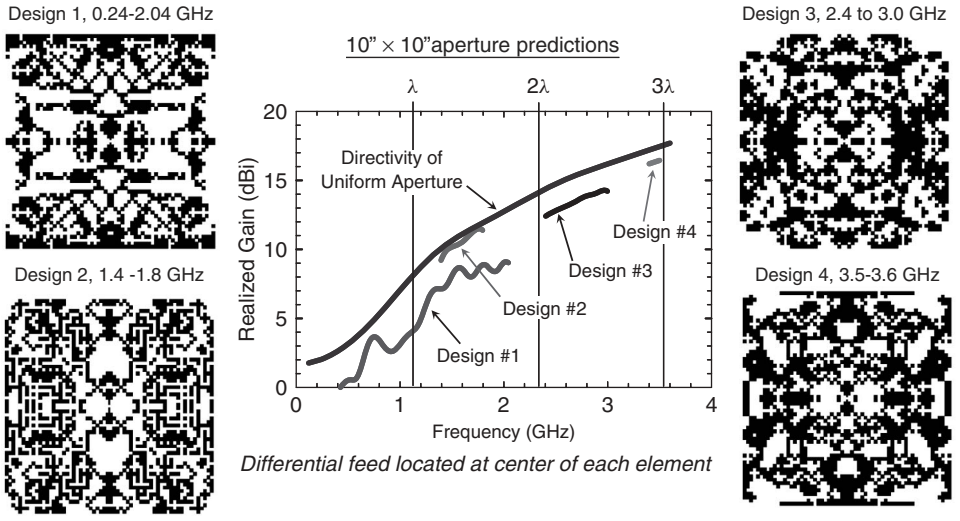


Figure 12.43 Early development of fragmented apertures focused on the relationship between gain and bandwidth for single feed antennas. The plot shows four early designs for antennas spanning up to 3 wavelengths.

element planar antennas. Note, in particular, that the 3λ square design has achieved realized gain (gain with mismatch) of better than 16 dBi.

The fragmented aperture design approach extended naturally into efforts to extend the instantaneous bandwidth of practical array antenna elements. The insight that led to a breakthrough in achievable bandwidths was the recognition that dc electrical connection between elements was actually useful and should be exploited. Subsequent multiple-octave array designs consistently featured these connections, which support continuous current paths that span multiple elements. For example, in an array with an 8:1 bandwidth, the radiated wavelength changes from approximately the width of two elements at the highest frequency to 16 elements at the lowest. With the connected array, continuous conducting paths of sufficient length to support the necessary currents are present on the aperture.

As an experiment to demonstrate the importance of connected arrays, the 6-cm elements of Figure 12.44 were designed to operate from 0.25 to 2.5 GHz in an array with no ground plane. The aperture limited gain was thus approximately $2\pi A/\lambda^2$, since the apertures radiate equally in both hemispheres. The first element design was optimized with electrical connections between elements permitted, that is, a connected array. The second element was optimized with a boundary enforced around each element to prevent conducting pathways between elements. The realized gain achieved by an 8×8 finite array of each element design is shown in the figure. Because of the continuous current paths across element boundaries, the connected design is able to maintain a good impedance match over the full 8:1 bandwidth and thus achieves superior performance.

Another key feature of the connected geometry is that the overall size of the array becomes a limiting factor on the lowest operating frequency. When the previous connected design was modeled in arrays of various sizes (again without a ground plane), the resulting performance was proportional to array size, as shown in Figure 12.45. Arrays

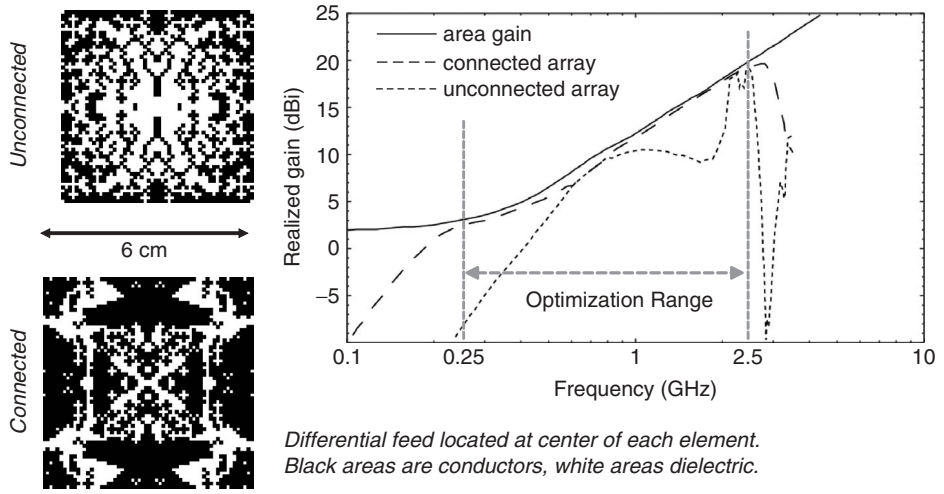


Figure 12.44 This design experiment compared performance of two 8×8 arrays using the elements pictured on the left. The first was optimized with the constraint that the elements not be electrically connected; the second was optimized with no such constraint. The connected array achieved much broader bandwidth.

of $2 \times 2, 4 \times 4, 8 \times 8,$ and 16×16 elements were simulated. For all cases, the upper frequency limit remained relatively constant, being limited by the element lattice spacing and resulting grating lobe formation. The low frequency limit, on the other hand, was approximately proportional to the overall array dimension.

To confirm the validity of these simulations, a fragmented array with 3-cm elements was designed and measured in 1999. The quantity modeled and measured was the embedded element realized gain (EERG), where one central element was active and others were terminated in matched resistive loads. This quantity measures the performance of an element in the array environment. An angle pattern cut of the EERG can be used to predict the scan performance of an array of these elements. This highlights the performance of the radiator without the need for expensive beam-forming networks. As can be seen from Figure 12.46, the array achieves near aperture-limited gain at broadside over a 10:1 bandwidth, with excellent model-measurement agreement.

12.4.2 Wideband Backplanes: Planar 10:1 Arrays

Early explorations of the fragmented arrays (2000 and earlier) focused on basic questions of element connections, bandwidth limits, and natural impedance values [37]. These investigations typically either used no ground plane behind the radiating surface or accepted the limitations of simple ground planes. Ideally, a ground plane should be located $\lambda/4$ behind the broadside radiating surface of a planar antenna. Notionally, the backward-radiated energy travels a path length with a phase progression of 180° that, together with the 180° phase inversion at the perfect electrical conductor (PEC) surface of the ground plane, causes the reflected energy to arrive in phase with the forward going radiation. Wideband antennas pose a difficulty, however, since λ varies widely over the operating bandwidth. In fact, when the ground plane is $\lambda/2$ behind the radiating

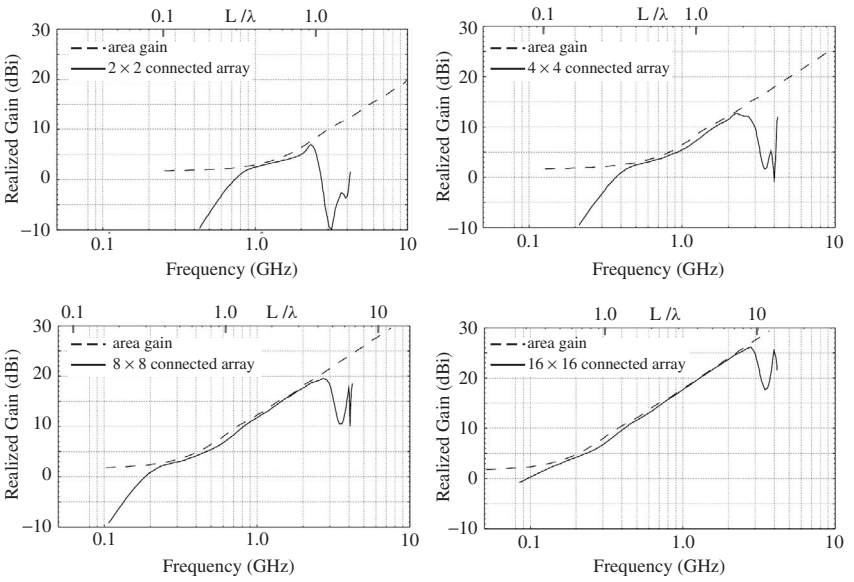


Figure 12.45 The connected element from Figure 12.44 was simulated in arrays of various sizes. The results show that the low-frequency performance limit is essentially proportional to overall array size, L.

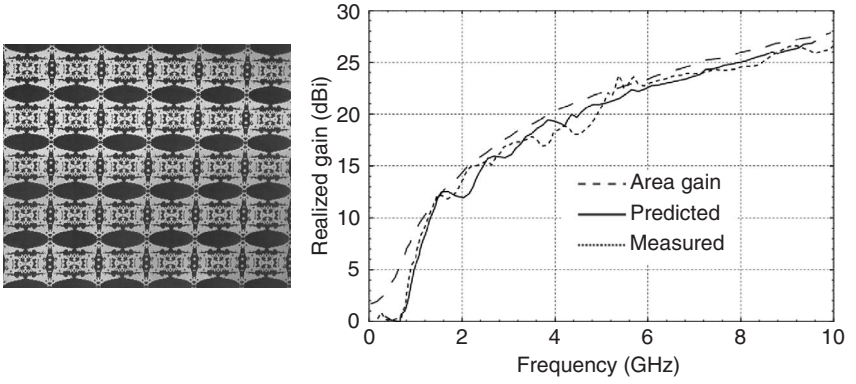


Figure 12.46 Embedded element realized gain for a central element of a 10 × 17 array with 3-cm square unit cells.

surface (or an integer multiple of $\lambda/2$) the backward going radiation is reflected and arrives exactly out of phase with the forward radiation. This situation is illustrated in Figure 12.47, which shows the results of a simulation of a fragmented aperture radiator placed 2.5 cm in front of a PEC ground plane. The broadside gain is normalized to the area gain for this antenna. Without the ground plane, the radiator is well matched across the band, but because it is radiating in both directions the forward radiation only approaches −3 dB, represented by the dashed line in the plot. With the ground plane,

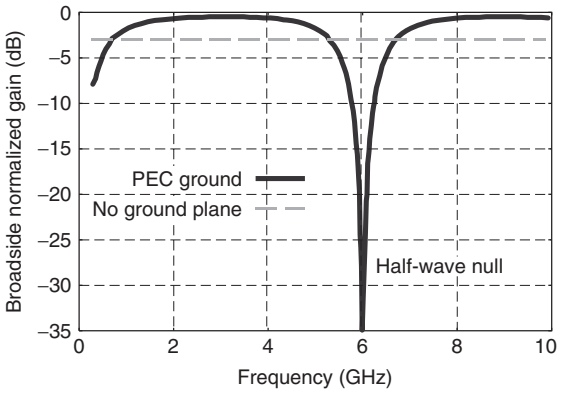


Figure 12.47 When a broadband radiating sheet is placed in front of a simple PEC ground plane, the resulting gain pattern will suffer nulls at frequencies where the separation distance is an integer multiple of a half-wavelength (in this case, 6 GHz for a 2.5-cm separation).

the gain approaches the maximum around 3 GHz, where the 2.5 cm represents a quarter of the free-space wavelength and the ground plane provides almost a 3-dB increase in broadside gain. At 6 GHz, however, the ground plane is a half-wavelength behind the radiating surface and the gain suffers a deep null. The null will be repeated every multiple of $\lambda/2$ (12 GHz, 18 GHz, etc.) Practical experience indicates that fragmented aperture designs can be extended to approximately 8:1 bandwidths before the half-wave null impact must be addressed.

The problem is made more complicated if the array is intended to scan over any significant volume, because the null frequency is dependent on the scan angle, as illustrated in Figure 12.48. As the scan angle moves away from broadside the null frequency increases. The contour plot in the figure shows this trend.

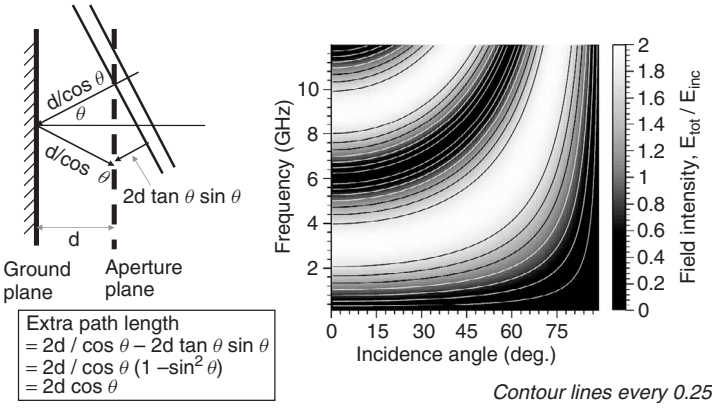


Figure 12.48 For normal incidence (or when a phased array is scanned to broadside) pattern nulls will occur when the ground plane is at a half-wavelength separation. At scan angles off normal, the null will occur at higher frequencies. This geometry is illustrated in the diagram on the left. The contour plot on the right shows the relationship between field intensity at the radiating surface and frequency and angle for a 2.5-cm separation.

Since the problem can be attributed to backward radiated energy, one is tempted to address it with absorbing solutions. Interestingly, the Salisbury screen absorbing structure has the desirable characteristic that its tuned absorption frequency increases with incidence angle, exactly analogous to the scan angle–frequency dependence of the half-wave null. This was the inspiration for the "broadband screen" backplane, which GTRI developed to extend the frequency performance of fragmented apertures over a ground plane. As an example, Figure 12.49 shows the performance of a typical planar fragmented aperture over a classical Salisbury screen, that is, a 377-ohm R-card positioned $\lambda/4$ in front of a ground plane. The backplane is most absorptive at exactly the frequency/angle combinations where the half-wave null occurs (and in fact at every odd multiple of half-wavelengths).

Figure 12.50 shows the normalized realized gain at broadside with the first generation broadband screen backplane. Now the aperture has recovered enough gain at the problem frequency to achieve near 50% efficiency. However, we can do better. For overall antenna performance, the impedance value, position, and even the number of R-card layers may

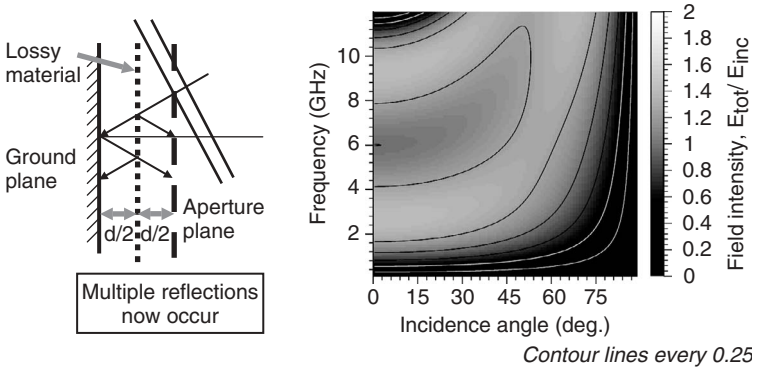


Figure 12.49 Here the radiating surface is located 2.5 cm in front of the PEC ground plane, but a 377-ohm/square R-card layer is placed halfway between the radiator and ground plane, eliminating the deep null at $\lambda/2$.

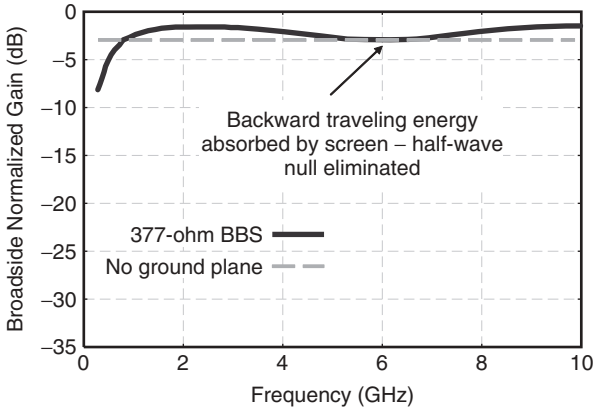


Figure 12.50 This plot of the normalized realized gain at broadside for the configuration of Figure 12.49 shows that the deep null at 6 GHz has been improved to only 3-dB insertion loss.

be considered as free variables in the design. For example, if the 377-ohm R-card is replaced with a 225-ohm card, realized gain is maintained within 2 dB of the aperture limit across the operating band of the antenna.

Figure 12.51 pictures a 10:1 design that was developed as a proof of concept using a single R-card broadband screen. The array demonstrated better than 50% efficiency over the operating band of 1–10 GHz. The plot in the right half of the figure shows normalized predictions of realized gain or, equivalently, insertion loss. The top curve (normalized gain) shows the effects of resistive loss. The bottom curve (normalized realized gain) shows the combined effects of resistive and mismatch loss. Thus the distance between the two curves is a measure of the impedance match for this design.

With a simple conducting ground plane behind a planar radiating surface, a standing wave occurs when the separation distance is one-half wavelength (or an integer number of half-wavelengths), which places a field null at the radiating surface. The resulting impedance mismatch is the cause of the deep dropout in the gain curve. In addition to the energy they dissipate, R-cards inserted in the backplane stack introduce additional reflection boundaries that "break up" or redistribute the standing wave to avoid field cancellation at the radiating surface. As the operating bandwidth of the array spans more octaves, a simple ground plane introduces more half-wave nulls and the problem of defeating the standing wave becomes more complicated. Figure 12.52 shows an example of a radiating surface located 3 inches in front of a simple conducting ground plane. The resulting standing wave produces interference nulls approximately every 2 GHz. When the empty cavity is replaced by an optimized broadband screen with six R-card layers in the backplane stack, the standing wave nulls are eliminated. Figure 12.53 compares the performance of the empty cavity with the broadband screen over frequency and scan angle. The nulls are effectively controlled to scan angles of 60° or more.

12.4.3 Multilayer Radiators: 33:1 Bandwidth Arrays

A broadband screen backplane can control half-wave nulls, but it uses a loss mechanism to do so. While it is not necessary to attenuate all of the backward radiated energy, some

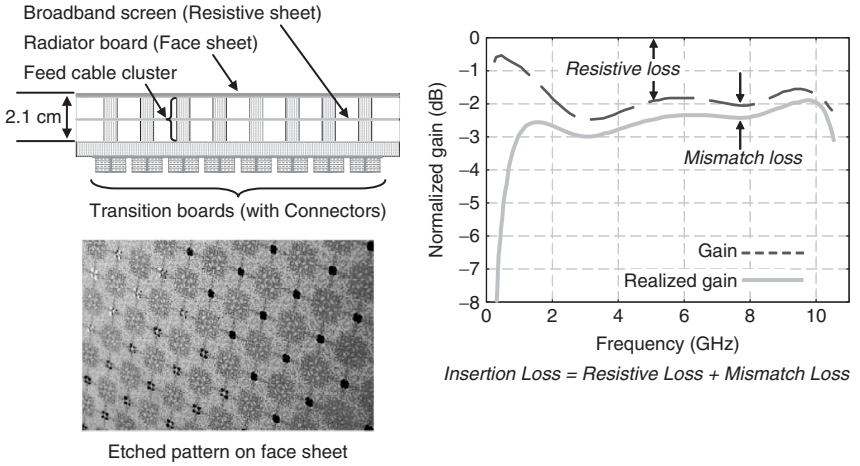


Figure 12.51 The first array built using the broadband screen backplane was this 10:1 design. Efficiency was better than 50% (<3-dB insertion loss) from 1 to 10 GHz.

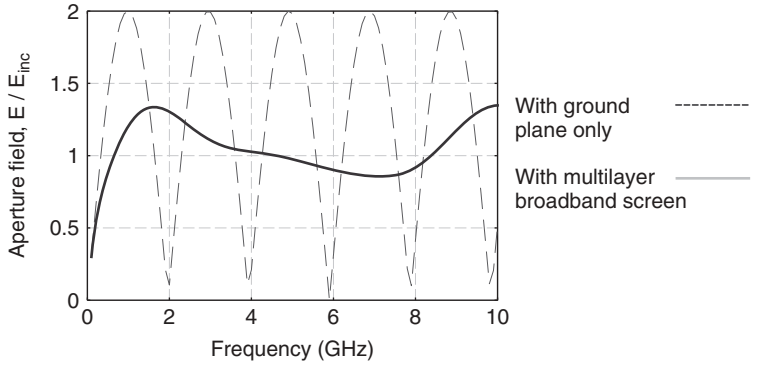


Figure 12.52 Aperture fields 3 inches in front of a PEC surface, with and without a broadband screen backplane in place.

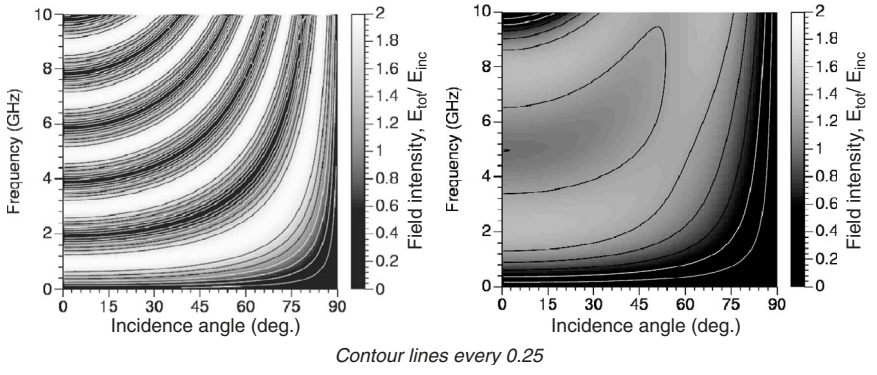


Figure 12.53 Contour plots comparing the configurations of Figure 12.52 over a range of scan angles.

loss is inevitable with this approach. It would be better to radiate energy only into the forward hemisphere and eliminate the possibility of backward-radiated energy reflecting off a ground plane to add out of phase with the forward directed radiation. As a thought experiment, consider the ideal planar radiator with no thickness in Figure 12.54. Notionally, radiation must occur equally into both hemispheres since nothing distinguishes one side from the other. However, if the radiating layer has some thickness, then asymmetries may be introduced that cause the surface to radiate preferentially in one direction, as in the second antenna of Figure 12.54. For example, if 90% of the energy is made to radiate into the forward hemisphere, then even if the backward-radiated energy is reflected and returns 180° out of phase, it will only reduce the transmitted power to 80% of the maximum value.

This principle may be exploited by using multiple radiating layers in front of the ground plane. The radiating layers may be actively driven, or they may be parasitic, analogous to the directors in a Yagi–Uda antenna. In Figure 12.55, two radiating layers approximately 8 mm apart with no ground plane were optimized using the fragmented aperture design process. The design goal was to maximize gain in the forward hemisphere.

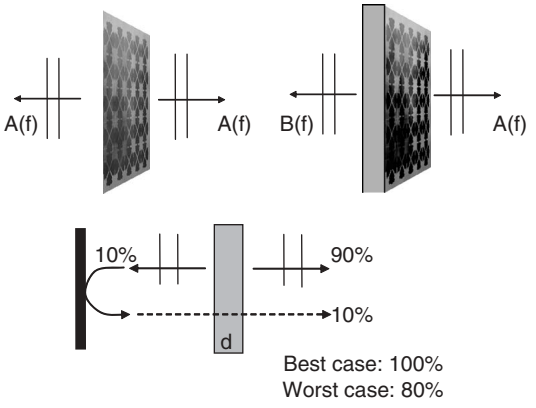


Figure 12.54 Thought experiment demonstrating the benefit of preferentially radiating in one direction to mitigate ground plane nulls. This is possible with asymmetric radiation, which can be achieved with a radiator thickness > 0.

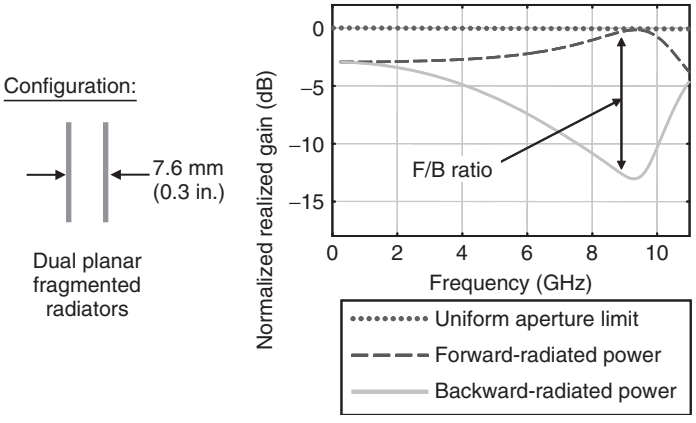


Figure 12.55 Idealized design with simultaneously optimized radiating layers. The design goal was to maximize front-to-back ratio.

The plot shows the normalized forward-going and backward-going radiation, demonstrating good preferential radiation, or front-to-back ratio (F/B), over the upper octave of the design region (1–10 GHz). As the wavelength gets longer, the electrical separation between radiating layers becomes insufficient to direct the radiation. The region of effectiveness for this approach is enough to produce the 10:1 design of Figure 12.56, where the realized gain remains within 3 dB of the maximum across the band. This design is for illustrative purposes only, as it has no realistic feed structures or loss.

The extent to which using multiple radiating face sheet layers may improve the bandwidth of the antenna depends on the number and spacing of the face sheets. In Figure 12.57, the antenna designs use two and three face sheets, respectively. These simulations include realistic feed structures, but the ground planes have been replaced by perfectly matched absorbing layers in the simulation. With two face sheets, the antenna

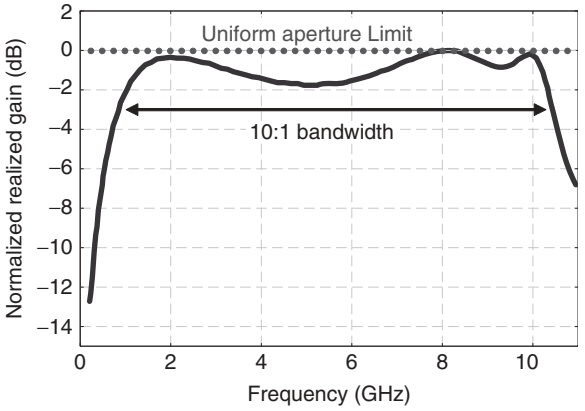


Figure 12.56 The use of parasitic layers to direct radiation forward may also be accomplished in the presence of a PEC backplane, as in this example, where insertion loss is kept below 2 dB over most of a 10:1 bandwidth.

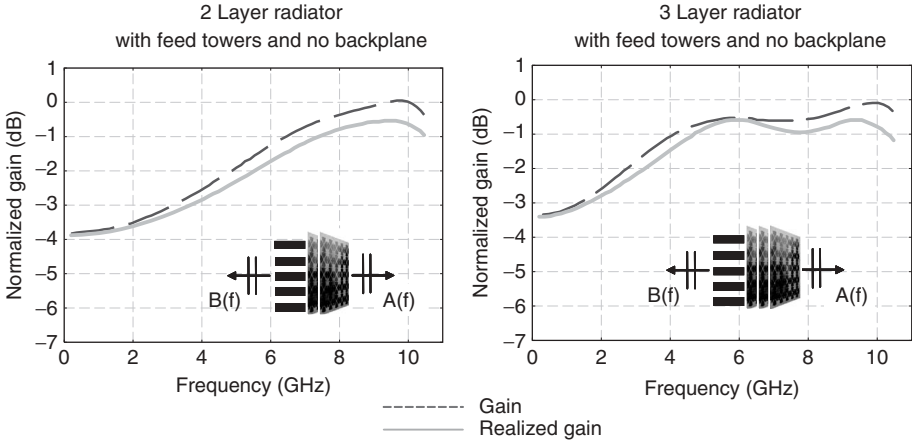


Figure 12.57 Design experiments with two and three radiating face sheets. The ground plane has been replaced in each simulation with a perfectly absorbing layer as the back boundary condition. Thus normalized gain levels above −3 dB may be attributed to the F/B ratio.

exhibits enhanced gain over most of the upper octave. With three, antenna gain is enhanced over the upper two octaves.

Design of fragmented elements for phased arrays with operating bandwidths beyond 10:1 are best executed through a judicious combination of a multilayer radiator with a broadband screen backplane. In partnership with Northrop Grumman Electronics Systems, GTRI has built and measured two laboratory proof-of-concept radiators with 33:1 bandwidths, each incorporating both design strategies. Each design consisted of a three-layer radiator stack over a six R-card backplane stack. In the first design, two face sheets were driven by the feeds and the third was parasitic. In order to simplify the manufacturing process, the second design had only the innermost face sheet driven, with two parasitic

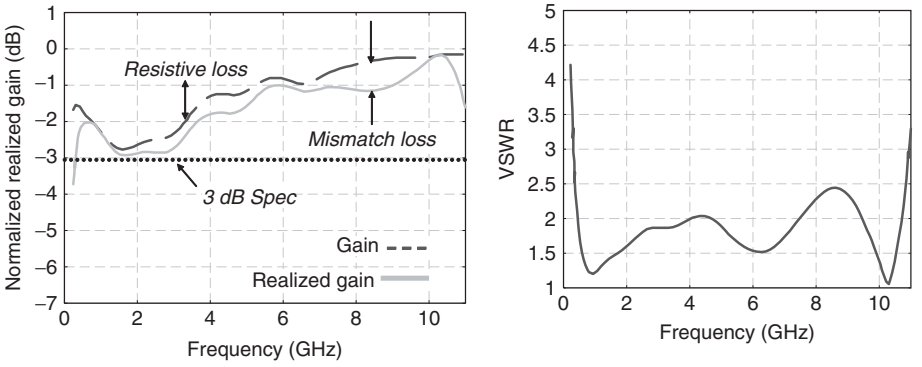


Figure 12.58 Predicted performance of the 33:1 antenna design. In this periodic simulation, antenna efficiency is shown to be better than 50% over the entire bandwidth of 0.3–10 GHz for an actual antenna designed with realistic feed structures.

outer layers. Figure 12.58 shows the performance of the second design in a periodic simulation, which eliminates finite array edge effects. The simulation includes realistic feed structures and material properties. Gain is normalized to the element area gain, so the 0-dB line represents ideal performance. Note that the design achieves nearly 1 dB or better insertion loss over the upper octave, with better than 3-dB insertion loss over the entire 0.3–10-GHz design bandwidth. Representative measurements of the second design are presented below.

The simulations were supported with measurements of a test piece on three different antenna ranges covering the entire operating bandwidth. These measurements were not only consistent across all three ranges, but they validated the simulation results. The test antenna was a 23×23 element, dual linear polarized array with the center element actively driven and all surrounding elements terminated in matched 188-ohm impedances at the feed points. Figure 12.59 illustrates the composition of one element of the array in cross section, with cartoons of the etched unit cell pattern on each face sheet and a photo of the test antenna.

Broadside frequency scans of the embedded element realized gain (EERG) are plotted in Figure 12.60. The EERG is obtained by driving one element and terminating the rest with matched loads. This greatly reduces the cost of the measurement setup and test piece, as beam-former electronics are not required. Achievable scan volume may be inferred from the beamwidth of the EERG angle pattern cuts.

The figure compares the measurement results at broadside to predictions. The element area gain, which represents the physical limit for antenna performance, is denoted by the dashed line. The predicted EERG at broadside is denoted by the solid line. Compared to these are four measured data sets from the three different antenna ranges, including three different calibration horns spanning the 33:1 bandwidth. Data was also measured on both V-pol and H-pol feeds (both sets should be equivalent at broadside for this symmetric design). The measured data show excellent consistency across ranges and at both sets of feed points, and excellent agreement with the predictions. The measurements showed approximately 1-dB more insertion loss at the high end than predicted. The difference is more than can be attributed to resistive loss in the feed cables and on the metal radiating surfaces. It is likely due to slight imperfections in the assembly of the radiator.

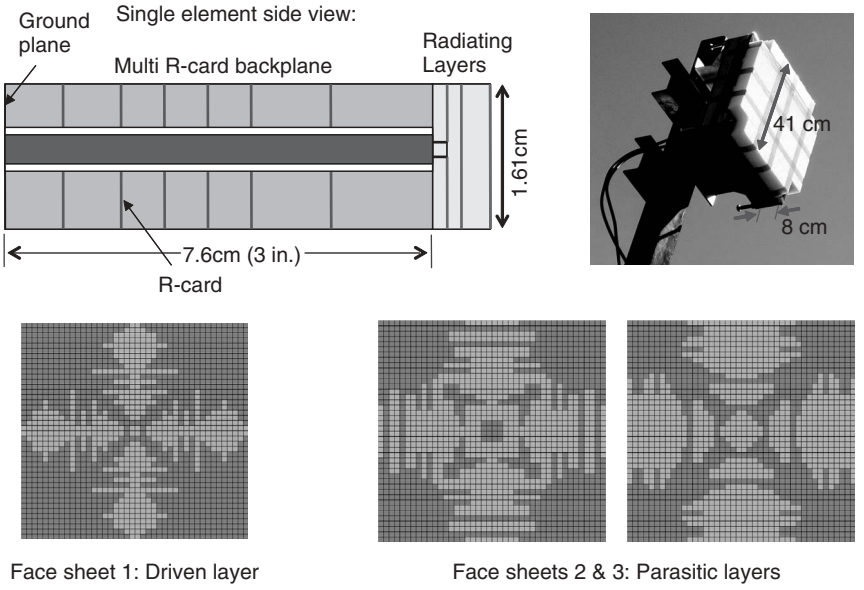


Figure 12.59 Construction details of one 33:1 antenna design, including a photo of the test piece used to measure embedded element realized gain (EERG).

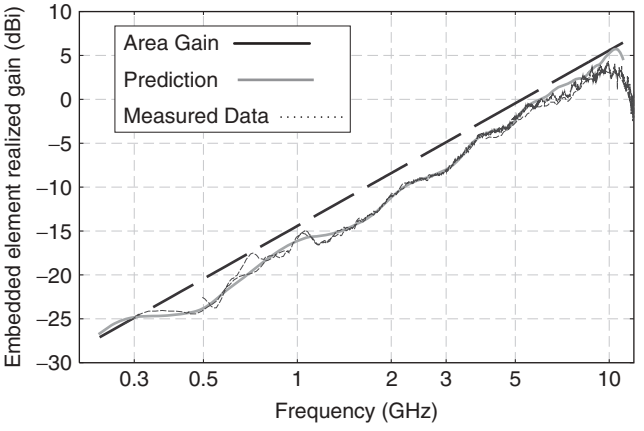


Figure 12.60 The plot shows a compilation of measured data at broadside for the 33:1 test antenna (three antenna ranges, two polarizations). The measured data is plotted against numerical predictions of performance, along with the element area gain representing ideal performance.

Performance at the high end is particularly sensitive to the position of the three layers at the end of the feed cables, and their planarity in the assembled test piece was affected by warping in the etched sheets.

This slight drop-off is removed in Figure 12.61 to facilitate angle pattern comparisons. These patterns allow detailed comparison of measured and modeled EERG over angle cuts at several discrete frequencies. Again, model-measurement agreement is excellent,

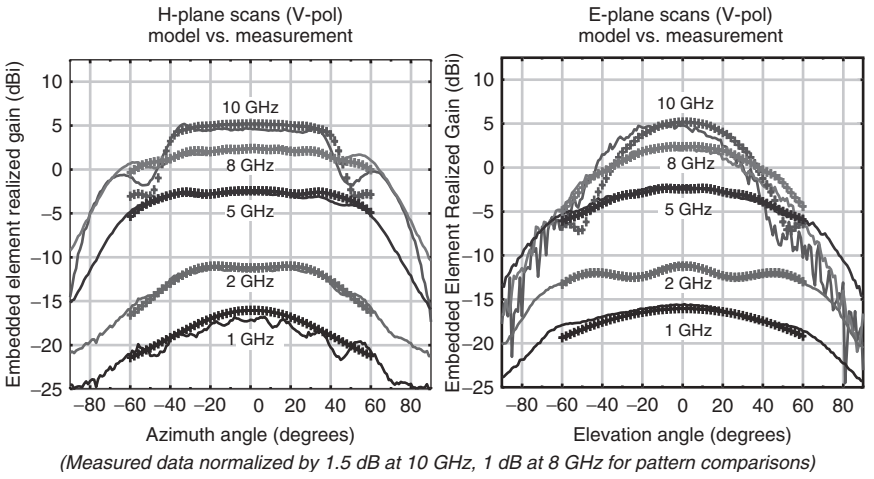


Figure 12.61 Comparison of modeled (solid lines) and measured (data markers) EERG pattern cuts at several discrete frequencies. Again, note the excellent agreement between prediction and measurement.

with the models predicting features such as the ripple at 2 GHz due to finite array edge effects and the narrowing of the scan volume above 8 GHz.

Figure 12.62 presents the measured EERG data from three different overlapping data sets for H-plane scans and two for E-plane scans. The contour plots show angle cuts plotted horizontally at each frequency. Each angle cut (horizontal line across the contour plot) has been normalized so that the maximum value at each frequency is zero (i.e., frequency slope has been removed from the data sets). The resulting image shows the achievable scan volume as a function of frequency for a fully driven array antenna with this design. That scan volume, as defined by the 3-dB points, is approximately $\pm 60^{\circ}$ over most of the band, with some narrowing above 9 GHz in the H-plane and above

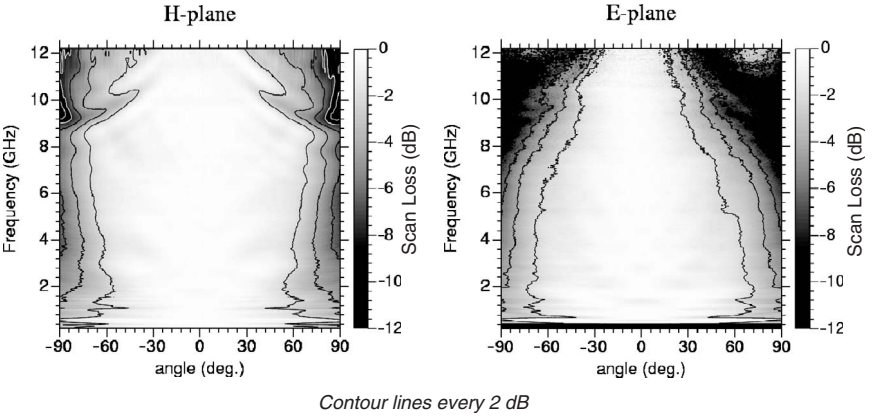


Figure 12.62 Compilation of measured angle cuts normalized to the maximum value at each frequency. Resulting contours indicate achievable scan volume. Note the lack of scan blindness in the operating regions.

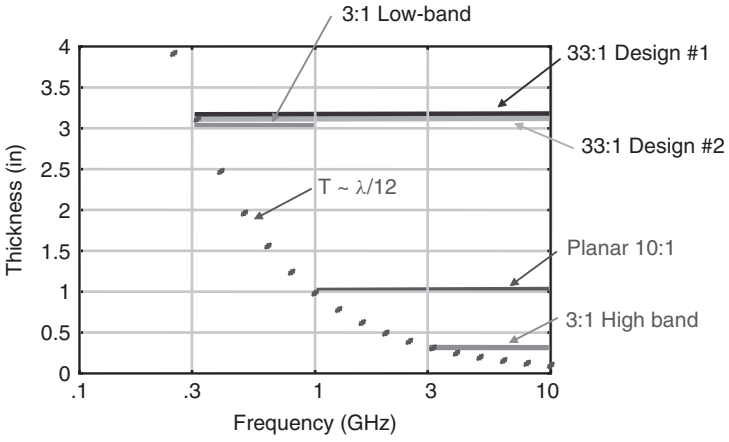


Figure 12.63 Results of several design exercises for fragmented arrays. For air-filled cavities, the antenna thickness is approximately $\lambda/12$ at the lowest operating frequency.

7 GHz in the E-plane. Note also that the featured scan volume shows no evidence of any suck-outs or scan blindness over the operating bandwidth.

Our experience with several wideband phased array designs has produced empirical evidence for a rule of thumb regarding the thickness of these wideband radiators. Figure 12.63 compiles results for five fragmented array designs with bandwidths greater than an octave. Designs with bandwidths less than 10:1 used simple ground planes; antennas with bandwidths of 10:1 or greater incorporated broadband screen backplanes. In each case, the overall thickness is dictated not by the bandwidth, but by the lowest operating frequency. For cavities filled with air or low-dielectric foams, the antenna thickness will be approximately $\lambda/12$ at the lowest frequency.

12.4.4 Conclusions

The successful design of ultrawideband phased arrays has been enabled by several factors. These designs require high fidelity time-domain EM solvers. The necessity to optimize over many frequencies would be time prohibitive if the designs were done with frequency-domain codes. Designs referenced in this chapter were all developed using a finite-difference time-domain code, the results of which can then be Fourier transformed to produce the requisite range of frequency predictions. The code is highly validated, but it was developed in house so we understand how to enhance the speed of simulations by sacrificing some accuracy when necessary. In early design stages, it is important to move through iterations rapidly to converge on a "pretty good" design. Then in the design's final stages, accurate predictions are required to account for realistic details such as feed structures and material characteristics. These highly accurate modeling codes require computing hardware with sufficient processing power and memory to handle fine details, and this is accomplished using relatively inexpensive Beowulf clusters of Linux-based PCs.

With the appropriate modeling tools and computing infrastructure, GTRI was able to develop the essential features for ultrawideband planar phased arrays: (1) connected arrays to span several octaves; (2) broadband backplane stacks to mitigate half-wave

nulls introduced by a ground plane; (3) multilayer radiators to enhance high frequency performance with front-to-back ratio, thus leveraging the backplane improvements at low frequencies; and (4) fragmented aperture radiators to accomplish impedance matching in the presence of feed structures, material substrates, multiple layers, and so on. Measured results confirm the success of these designs for bandwidths up to 33:1. Preliminary work suggests that phased array operation over bandwidths of 100:1 or more is possible.

ACKNOWLEDGMENTS

The three groups of authors of this chapter were joined by numerous colleagues and students, all of whom contributed to the work. This chapter is dedicated to them and their outstanding contributions.

Harris Corporation: Randy Boozer, Steve Brown, Grif Gothard, Dieter Gum, Jay Kralovec, Keith Krebs, Charlie Kulisan, Sean Ortiz, Brett Pigon, Jonothan Pryor, Jim Rawnick, Roger Strange, Bob Taylor, Don Weibelt, and Mark Vanstrum.

University of Massachusetts: Ajay Asija, Tan-Huat Chio, Michael E. Cooley, Wajih Elsallal, Ramakrishna Janaswamy, Sreenivas Kasturi, Jennifer Munro, Joon Shin Gregory J. Wunsch, K. Sigfrid Yngvesson, Yana Zilberberg.

Georgia Tech Research Institute/STL: Jim Acree, Brad Baker, Stephen Blalock, Doug Denison, Lynn Fountain, James Fraley, Morris Kesler, Eric Kuster, and Glenn Smith.

REFERENCES

1. H. Holter and H. Steyskal, On the size requirement for finite phased-array models, *IEEE Trans. Antennas Propag*., Vol. 50, No. 6, pp. 836–840, June 2002.
2. R. J. Mailloux, *Phased Array Antenna Handbook*, Artech House, Norwood, MA, 2005, pp. 291–352.
3. B. A. Munk, *Frequency Selective Surfaces: Theory and Design*, John Wiley & Sons, Hoboken, NJ, 2000, pp. 28–32.
4. H. A. Wheeler, Simple relations derived from a phased-array antenna made of an infinite current sheet, *IEEE Trans. Antennas Propag*., Vol AP-13, pp. 506–514, July 1965.
5. T. W. Kornbau, Analysis of periodic arrays of rotated linear dipoles, rotated crossed dipoles, and of biplanar dipole arrays in dielectric, Ph.D. dissertation, Ohio State University, Department of Electrical Engineering, Columbus, 1984.
6. B. A. Munk, T. W. Kornbau, and R. D. Fulton, Scan independent phased arrays, *Radio Sci.*, Vol. 14, No. 6, pp. 978–990, November–December 1979.
7. L. W. Henderson, Introduction to PMM, Tech. Rept. 715582–5, OSU ESL, Department of Electrical Engineering, prepared for Wright-Patterson Air Force Base, OH, February 1986.
8. B. A. Munk, *Finite Antenna Arrays and FSS*, John Wiley & Sons, Hoboken, NJ, 2003, pp. 181–213.
9. W. A. Davis, J. R. Nealy, G. F. Ricciardi, and W. L. Stutzman, Techniques for the measurement of the impedance of wideband balanced antennas, in *IEEE AMTA Symposium*, November 1995.
10. J. Rawnick, et al. Patch dipole array antenna including a feed line organizer body and related methods, U. S. Patent No. 6,483,464.

11. P. J. Gibson, The Vivaldi Aerial, *9th European Microwave Conference*, Brighton, UK, 1979, pp. 101–105.
12. L. R. Lewis, M. Fasset and J. Hunt, A broadband stripline array element, in *IEEE Antennas and Propagation Symposium Digest*, Atlanta, 1974, pp. 335–337.
13. P. S. Simon, K. McInturff, R. W. Jobsky, and D. L. Johnson, Full-wave analysis of an infinite, planar array of linearly polarized, stripline-fed, tapered notch elements, in *IEEE Antennas and Propagation Symposium Digest*, London, Ontario, 1991, pp. 334–337.
14. D. H. Schaubert, J. A. Aas, M. E. Cooley, and N. E. Buris, Moment method analysis of infinite stripline-fed tapered slot antenna arrays with a ground plane, *IEEE Trans. Antennas Propag.*, Vol. 42, No. 8, pp. 1161–1166, August 1994.
15. E. W. Lucas and T. P. Fontana, A 3-D hybrid finite element boundary element method for the unified radiation and scattering analysis of general infinite periodic arrays, *IEEE Trans. Antennas Propag.*, Vol. 43, No. 2, pp. 145–153, February 1995.
16. D. T. McGrath and V. P. Pyati, Phased array antenna analysis with the hybrid finite element method, *IEEE Trans. Antennas Propag.*, Vol. 42, No. 12, pp. 1625–1630, December 1994.
17. K. S. Yngvesson, D. H. Schaubert, T. L. Korzeniowski, E. L. Kollberg, T. Thungren, and J. F. Johansson, Endfire tapered slot antennas on dielectric substrates, *IEEE Trans. Antennas Propag.*, Vol. 33, No. 12, pp. 1392–1400, December 1985.
18. J. Chang, D. H. Schaubert, K. S. Yngvesson, J. Huang, V. Jamnejad, D. Rascoe, and L. Riley, 32 GHz Power-combining TSA array with limited sector scanning, in *IEEE Antennas and Propagation Symposium Digest*, Dallas, 1990, pp. 1150–1153.
19. M. V. Ivashina, J. G. Bij de Vaate, R. Braun, and J. D. Bragman, Focal plane arrays for large reflector antennas: first results of the demonstrator project, Ground-Based Telescopes and Instrumentation, in *Proceedings of the SPIE Conference*, Glasgow Scotland, 21–25 June 2004, Vol. 5489, pp. 1127–1138.
20. D. H. Schaubert, Endfire tapered slot antenna characteristics, in *Sixth International Conference on Antennas and Propagation, ICAP 89*, Coventry, UK, 1989, pp. 432–436.
21. D. T. McGrath, N. Schuneman, T. H. Shively, and J. Irion II, Polarization properties of scanning arrays, in *IEEE International Symposium on Phased Array Systems and Technology*, Boston, October 2003, pp. 295–299.
22. J. B. Knorr, Slot-line transitions, *IEEE Trans. Microwave Theory Tech.*, Vol. 22, No. 5, pp. 548–554, May 1974.
23. M. Kragalott, W. R. Pickles, and M. S. Kluskens, Design of a 5:1 bandwidth stripline notch array from FDTD analysis, *IEEE Trans. Antennas Propag.*, Vol. 48, No. 11, pp. 1392–1400, November 2000.
24. D. H. Schaubert, S. Kasturi, M. W. Elsallal, and W. van Cappellen, Wide bandwidth Vivaldi antenna arrays—some recent developments, in *Proceedings of the First European Conference on Antennas and Propagation*, Nice, France, November 2006.
25. J. D. S. Langley, P. S. Hall, and P. Newman, Balanced antipodal Vivaldi antenna for wide bandwidth phased arrays, *IEE Proc. Microwaves Antennas*, Vol. 143, No. 2, pp. 97–102, April 1996.
26. M. W. Elsallal and D. H. Schaubert, Parameter study of a single isolated element and infinite arrays of balanced antipodal Vivaldi antennas, in *Proceedings of the 2004 Antenna Applications Symposium*, Monticello, IL, September 2004, pp. 45–69.
27. M. W. Elsallal and D. H. Schaubert, On the performance trade-offs associated with modular element of single- and dual-polarized DmBAVA, in *Proceedings of the 2006 Antenna Applications Symposium*, Monticello, IL, September 2006, pp. 166–187.
28. D. H. Schaubert, A class of E-plane scan blindnesses in single-polarized arrays of tapered-slot antennas with a ground plane, *IEEE Trans. Antennas Propag.*, Vol. 44, No. 7, pp. 954–959, July 1996.

29. H. Holter, T.-H. Chio, and D. H. Schaubert, Elimination of impedance anomalies in single- and dual-polarized endfire tapered slot phased arrays, *IEEE Trans. Antennas Propag.*, Vol. 48, No. 1, pp. 122–124, January 2000.

30. D. Norman, D. H. Schaubert, B. DeWitt, and J. Putnam, Design and test results for dual polarized Vivaldi antenna array, in *Proceedings of the 2005 Antenna Applications Symposium*, Monticello, IL, September 2005, pp. 243–267.

31. S. Kasturi and D. H. Schaubert, Effect of dielectric permittivity on infinite arrays of single-polarized Vivaldi antennas, *IEEE Trans. Antennas Propag.*, Vol. 54, No. 2, pp. 351–358, February 2006.

32. H. Holter and H. Steyskal, On the size requirement for finite phased-array models, *IEEE Trans. Antennas Propag.*, Vol. 50, No. 6, pp. 836–840, June 2002.

33. S. Kasturi, Design parameters in single-polarized, infinite arrays of Vivaldi antennas, M.S. Thesis, University of Massachusetts, Electrical and Computer Engineering, September 2004.

34. D. H. Schaubert, A. O. Boryssenko, A. van Ardenne, J. G. bij de Vaate, and C. Craeye, The square kilometer array (SKA) antenna, in *IEEE International Symposium on Phased Array Systems and Technology 2003*, Boston, October 2003, pp. 351–358.

35. D. H. Werner, R. L. Haupt, and P. L. Werner, Fractal antenna engineering: the theory and design of fractal antenna arrays, *IEEE Antennas Propag. Mag.*, Vol. 41, No. 5, pp. 37–59, October 1999.

36. D. H. Werner and S. Ganguly, An overview of fractal antenna engineering research, *IEEE Antennas Propag. Mag.*, Vol. 45, No. 1, pp. 38–57, February 2003.

37. P. Friederich et al., A new class of broadband planar apertures, in *Proceedings of the 2001 Antenna Applications Symposium*, Allerton Park, IL, September 2001, pp. 561–587.

38. T.-H. Chio and D. H. Schaubert, Parameter study and design of wide-band widescan dual-polarized tapered slot antenna arrays, *IEEE Trans. Antennas Propag.*, Vol. 48, No. 6, pp. 879–886, June 2000.

CHAPTER 13

Synthesis Methods for Antennas

WARREN STUTZMAN and STANISLAV LICUL

13.1 INTRODUCTION

Most engineering problem types can be classified as either analysis or synthesis problems. The field of antennas is no different. *Antenna analysis* addresses a given antenna by applying techniques (analytical formulation, simulation, and/or measurement) to understand its operation and to determine its performance characteristics. *Antenna synthesis* is the reverse of analysis; an antenna structure is derived from a given set of performance characteristic specifications, usually including the radiation pattern. As an analogy, in circuit theory it is possible to determine the needed circuit components and their values for achieving a specified frequency response characteristic. In ideal antenna synthesis, one would start with a set of electrical, mechanical, and system specifications that would lead to a particular antenna together with its specific geometry and material composition. This ideal general synthesis method does not exist in the antenna field. Antenna synthesis is usually only possible after the problem is limited to a specific antenna type or a narrow range of antenna types, but this is entirely adequate for many applications.

Traditionally, antenna synthesis has been applied to narrowband antennas in the frequency domain. This chapter includes in Section 13.5 a treatment of methods that can be used in both the time and frequency domains. It is shown how pole-residue methods are applied to synthesis of antennas in the time domain, which is useful in emerging areas such as ultrawideband applications.

Antenna design is the determination of a specific geometry for an antenna or antenna type satisfying performance specifications that are known from experience to be achievable in practice. Very often antenna design is accomplished through repeated analysis. This is a trial-and-error approach based on organized variation of parameter values such as lengths and material parameters for the specific antenna or antenna type. This is usually started with numerical simulations that are performed until acceptable performance characteristics are achieved followed by experimental validation at appropriate points in the design process. Rarely is it known how close to optimum the final design is. Repeated analysis will not necessarily yield, for example, the smallest size antenna for the achieved performance. However, repeated analysis produces a realizable antenna

Modern Antenna Handbook. Edited by Constantine A. Balanis

because the starting point is realizable. There are many antenna simulation code packages available that facilitate repeated analysis. Some packages also include optimization routines, often based on the genetic algorithm, to reach an "optimum" design.

Antenna synthesis, in contrast to antenna analysis that does not yield the antenna excitations in a straightforward manner, yields a current distribution that will produce the specified radiation pattern. Realization of required current distribution (either real or equivalent) in antenna hardware may at times be difficult. Nevertheless, antenna synthesis methods are very valuable in guiding the antenna designer in the pursuit of near optimal solutions for the problem class. For example, the synthesis of a high directivity, single beam, low sidelobe pattern is achieved by simply using a uniform phase, equal amplitude current distribution that is long enough to achieve the required directivity. In an application such as a satellite-platform based antenna that must illuminate the entire globe, a single beam centered on the Earth with about 34-dB gain is required. An efficient design is to use an antenna with a uniform phase, equal amplitude current distribution giving 34-dB directivity.

Synthesis becomes more difficult with increasingly complex antenna requirements. The general pattern synthesis problem is to specify the pattern variables and then to determine the required antenna variable values for a given antenna geometry type. The possible variables for the antenna and pattern are listed in Table 13.1 [1, Chap. 8]. Synthesis methods that apply to various combinations of antenna variables and pattern variables are the focus of this chapter. As a final note, this chapter emphasizes practical synthesis methods and provides references to methods with primarily theoretical interest.

13.2 SYNTHESIS PRINCIPLES

Antenna analysis in the frequency domain relates an antenna current distribution to its radiation pattern. The radiation pattern is the Fourier transform of the current distribution altered by obliquity factors that arise due to antenna geometry class and the

TABLE 13.1 Antenna Synthesis Variables

Antenna Variables	Radiation Pattern Variables
Antenna continuity	Main beam
Continuous	Narrow main beam
Discrete—array	Single beam
	Multiple beams
Antenna shape	Shaped main beam
Linear	
Planar	Pattern nulls at Specified Angles
Conformal	Sidelobes
Three-dimensional	Nominal sidelobes
	Low sidelobes
Antenna size	Shaped envelope sidelobes
Domain	
Frequency	
Time	

coordinate system choice. That is, each far-field component, with spherical spreading loss and constants removed, can be factored into an element factor and a pattern factor:

$$F(\theta, \phi) = g(\theta, \phi) f(\theta, \phi) \tag{13.1}$$

where $g(\theta, \phi)$ is the *element factor* (or obliquity factor), which is the pattern of an infinitesimal element of current of the actual antenna, and $f(\theta, \phi)$ is the *pattern factor*. For example, the E-plane pattern of a half-wave dipole oriented along the z-axis of a spherical coordinate system can be expressed as

$$F(\theta) = g(\theta) f(\theta) = \sin\theta \frac{\cos[(\pi/2)\cos\theta]}{\sin^2\theta} = \frac{\cos[(\pi/2)\cos\theta]}{\sin\theta} \tag{13.2}$$

The pattern factor of any linear current distribution along the z-axis is found as

$$f(\theta) = \frac{1}{\lambda} \int_{-L/2}^{L/2} i(z) e^{j\beta z \cos\theta} dz \tag{13.3}$$

where $\beta = 2\pi/\lambda$ is the phase constant (or wavenumber) and λ is the wavelength. The current distribution $i(z)$ is unitless and such that the far-field pattern $f(\theta)$ is normalized to unity maximum. The current is complex-valued to include beam scanning through phase control. The mathematics simplify by using the following normalized independent variables: s, the spatial variable on the antenna, and w, the associated far-field angular variable. Then Eq. (13.3) becomes

$$f(w) = \int_{-L/2\lambda}^{L/2\lambda} i(s) e^{j2\pi ws} ds \tag{13.4}$$

where

$$w = \cos\theta \quad \text{and} \quad s = z/\lambda$$

The current distribution is of finite extent, spanning $-L/2\lambda$ to $+L/2\lambda$. In communication theory terminology, this would be referred to as a "band-limited" function. With the current being zero valued for $|s| > L/2\lambda$, the limits on the integral can be extended to infinity, giving

$$f(w) = \int_{-\infty}^{\infty} i(s) e^{j2\pi ws} ds \tag{13.5}$$

This pattern factor is the Fourier transform of the current distribution.

From Fourier transform theory, we can find the current distribution through the inverse transform:

$$i(s) = \int_{-\infty}^{\infty} f(w) e^{-j2\pi sw} dw \tag{13.6}$$

This result forms the synthesis solution. The current distribution $i(s)$ required to produce the far-field pattern $f(w)$ is found by the inverse transform, Eq. (13.6). This result is very comforting because the large body of Fourier transform theory can be borrowed. However, there are further considerations for antenna synthesis. A *desired pattern* $f_d(w)$ used in the inverse transform in Eq. (13.6) will, in general, not produce a current distribution of finite extent fitting into $-L/2\lambda$ to $+L/2\lambda$.

Example 1: The Sector Pattern A radiation pattern that provides uniform response over a limited angular sector and zero response over the remaining angular space has several applications, including direction finding. A perfect sector pattern extending from $-c$ to $+c$ is

$$f_d(w) = \begin{cases} 1, & |w| \le c \\ 0, & |w| > c \end{cases} \tag{13.7}$$

Then Eq. (13.6) yields

$$i(s) = \int_{-\infty}^{\infty} f(w) e^{-j2\pi sw} dw = \int_{-c}^{c} 1 e^{-j2\pi sw} dw = \frac{e^{-j2\pi sc} - e^{-j2\pi s(-c)}}{-j2\pi s}$$
$$= 2c \frac{\sin(2\pi cs)}{2\pi cs} = Sa(2\pi cs) \tag{13.8}$$

where $Sa(x) = (\sin x)/x$ is the *sampling function*. This should be a familiar result from communication theory; the Fourier transform of a pulse is a sampling function. The current distribution, Eq. (13.8), is infinite in both directions, which is clearly impractical. We will return to practical forms of Fourier transform synthesis in Section 13.3.1.1.

Fourier transform theory can be used to derive properties that provide guidance in antenna synthesis, especially relating to the symmetry characteristics. Current distribution symmetry is relative to the center of the aperture and pattern symmetry is relative to the $w = 0$ point. The following properties that can be derived from Fourier transform theory are useful in determining the current distribution types required for achieving certain pattern symmetries [1, Sec. 8.1].

Property 1: A real-valued pattern is achieved if and only if the current distribution amplitude is symmetric and the phase is odd. Real-valued patterns are often assumed for mathematical simplicity, but patterns can be complex valued. A real-valued pattern is not required in most applications, but it turns out that introducing pattern phase as an additional variable adds little to the synthesis capability.

Property 2: A real-valued current distribution produces a symmetric pattern. Thus a symmetric pattern is obtained from a zero-phase current distribution.

Property 3: An asymmetric pattern can only be achieved through the use of current phase control. For example, a linear phase taper in the current will steer the beam off axis, which is a pattern that is asymmetric about $w = 0$, although it can be symmetric about the beam peak point.

Many applications require the pattern to be of a specified shape and, in some cases, pattern control must be dynamically controlled to respond to changing operational conditions. Pattern shape requirements often include the main beam shape, multiple main beams, all sidelobes of low level, or a maximum sidelobe level envelope that varies with angle. Associated system level constraints are frequency of operation, bandwidth, allowed array geometries, antenna size, and profile limits. A completely general synthesis problem is one of determining the antenna geometry and excitation that satisfies the pattern and system requirements within acceptable limits. No such general method exists, but the framework for a general approach has been reported [2]. In practice, antenna design involves the following process. First, the search is narrowed to a few candidate antenna

types using judgment and experience. Then, for each candidate antenna type, a synthesis procedure is applied to determine the excitations that produce a radiation pattern that fits the pattern specifications within acceptable tolerances. Finally, a down-selection process of the candidate solutions is used to select the final design.

This chapter presents classical and modern antenna synthesis methods. The primary focus is on one-dimensional linear current distributions. However, most such one-dimensional problems are extendable to two dimensions in a straightforward fashion. Also, discussions focus on real-valued patterns in order to limit unnecessary mathematical difficulties. Emphasis is twofold: (1) presentation of the methods that are useful to the practicing engineer, and (2) a survey of important literature for locating more details and advanced applications. Numerical examples and complete developments of some methods are found in Refs. 1 and 3–6. Also see Hansen [7] for a succinct summary of synthesis methods.

13.3 SHAPED BEAM SYNTHESIS METHODS

The main beam (or beams in the case of multiple beam patterns) can be made to conform closely to a prescribed shape. This section concentrates on one-dimensional current distributions, but the methods also apply to the principal plane patterns of two-dimensional currents that are separable (i.e., factorable into a product of two functions, each depending only on the associated source coordinate in the corresponding plane). As pointed out in Section 13.2, current phase control is required to synthesize an asymmetric beam. Usually arrays are more suitable for realizing the currents required to produce shaped beam antennas.

13.3.1 Shaped Beam Synthesis Methods for Line Sources

13.3.1.1 The Fourier Transform Method The most direct synthesis method makes use of the inverse Fourier transform relationship of Eq. (13.6) by merely inserting the desired pattern $f_d(w)$ in the transform:

$$i_d(s) = \int_{-\infty}^{\infty} f_d(w) e^{-j2\pi sw} dw \tag{13.9}$$

The resulting current distribution $i_d(s)$ does produce the exact desired pattern. However, the desired current, in general, will be of infinite extent, producing an impractical result. The Fourier transform antenna synthesis method proceeds by using a current distribution that is a truncated version of the desired current limited to an antenna of length L:

$$i_d(s) = \begin{cases} i_d(s), & |s| \le L/2\lambda \\ 0, & |s| > L/2\lambda \end{cases} \tag{13.10}$$

Substituting this into Eq. (13.9) yields

$$f(w) = \int_{-L/2\lambda}^{L/2\lambda} i_d(s) e^{j2\pi ws} ds \tag{13.11}$$

The Fourier transform synthesized pattern produces a very good approximation to the desired pattern and yields the minimum mean squared error (MSE) deviation from the desired pattern defined as

$$MSE = \int_{-\infty}^{\infty} |f(w) - f_d(w)|^2 dw \tag{13.12}$$

The visible region extends from 0° to 180° in θ, which corresponding to -1 to $+1$ in the variable w. So the MSE result in Eq. (13.12) extends over all space, both visible ($|w| \leq 1$) and invisible ($|w| > 1$); it does not give the desired measure of mean squared error in the visible region but is a useful guideline.

Example 2: Fourier Transform Synthesis of a Sector Pattern with a 10-λ *Line Source* An infinitely long linear current distribution is required to produce a perfect sector pattern, as we saw in Example 1. The Fourier transform antenna synthesis method consists of evaluating Eq. (13.10) by truncating the infinite current in Eq. (13.8) and applying the forward transform of Eq. (13.11) to obtain the approximate pattern:

$$f(w) = \frac{1}{\pi}\left\{Si\left(\frac{L}{\lambda}\pi(w+c)\right) - Si\left(\frac{L}{\lambda}\pi(w-c)\right)\right\} \tag{13.13}$$

where $Si(x) = \int_0^x [(\sin\tau)/\tau] d\tau$ is the sine integral function. Direct numerical integration of Eq. (13.11) is usually easier than working with sine integral functions. As a specific example, we choose $c = 0.5$ and $L = 10\lambda$. Figure 13.1 plots the synthesized pattern of Eq. (13.13) and required current distribution:

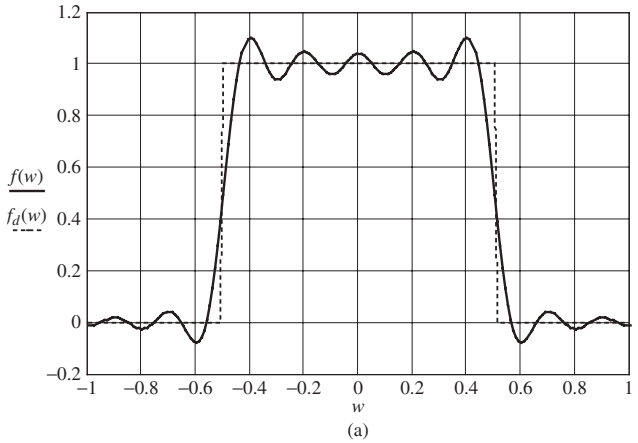
$$i(s) = 2c\frac{\sin(2\pi cs)}{2\pi cs} = Sa(2\pi cs) = Sa(\pi s), \quad |s| < 2.5 \tag{13.14}$$

This example produces an expected result. The approximate sector pattern with a finite pulse type shape has an approximately sampling function shaped current distribution. The overshoots and ripple in the synthesized pattern are representative of deviations from the desired pattern if it has discontinuities like this sector pattern. This *Gibbs phenomenon* is a familiar result from Fourier theory and signal theory. An infinitely long aperture is required to create the exact desired pattern with no ripple.

13.3.1.2 The Woodward–Lawson Sampling Method

The Fourier transform synthesis method produces a synthesized pattern with good overall agreement to the desired pattern. The Woodward–Lawson sampling method, first presented by Woodward [8] and expanded by Woodward and Lawson [9], provides a direct way to ensure that the synthesized pattern agrees with the desired pattern exactly at discrete *sample points* where the pattern equals specified *sample values*. The Woodward–Lawson method is based on decomposing the current distribution into a sum of uniform amplitude, linear phase slope sources:

$$i(s) = \frac{1}{L/\lambda}\sum_{n=-M}^{M} a_n e^{-j2\pi w_n s} \tag{13.15}$$



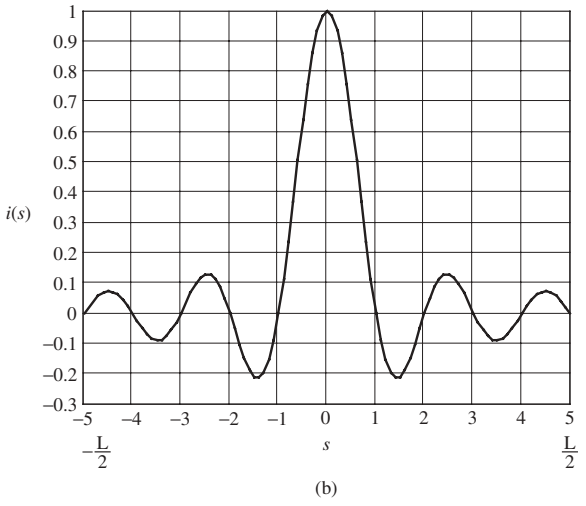


Figure 13.1 Example of Fourier transform synthesis of a sector pattern using a 10λ line source. (a) The synthesized pattern (solid curve) and the desired sector pattern (dashed curve). (b) The current distribution for the synthesized pattern.

There are $2M + 1$ component currents, sample points $\{w_n\}$, and sample values $\{a_n\}$. The corresponding component beams are orthogonal if the samples are spaced properly; the original method prescribes the following:

$$w_n = \frac{n}{L/\lambda}, \quad |n| \le M, \quad |w_n| \le 1.0 \tag{13.16}$$

The pattern associated with Eq. (13.15) is found by taking its Fourier transform, giving a summation of component beams of the sampling function type:

$$f(w) = \sum_{n=-M}^{M} a_n Sa\left(\pi \frac{L}{\lambda}(w - w_n)\right) \tag{13.17}$$

At sample locations $w = w_n$, the pattern equals the sample value $f(w_n) = a_n$ because

$$Sa(0) = 1 \quad \text{and} \quad Sa\left(\pi \frac{L}{\lambda}(w_m - w_n)\right) = Sa\left(\pi \frac{L}{\lambda}(m-n)\frac{\lambda}{L}\right) = 0 \quad \text{for } m \neq n$$

Then the sample value at the sample point is

$$a_n = f_d(w = w_n) = f(w = w_n) \tag{13.18}$$

Thus the synthesized pattern equals the desired pattern at the sample points and the sample values at the sample points are unaffected by the other component beams; that is, the component beams are orthogonal. The MSE will not be greatly different from that of the Fourier transform method.

Example 3: Woodward–Lawson Synthesis of a Sector Pattern with a 10λ *Line Source* The same sector pattern as in Example 2 is synthesized by the Woodward–Lawson method again using a 10λ-long line source so that the methods can be compared directly. The sector pattern is sampled at locations $w_n = n\lambda/L = 0.1n$ as prescribed in Eq. (13.16). The sample values are equal to the desired pattern at the sample points as in Eq. (13.18); that is, 1 for $|n| \leq 4$ and 0 for the rest except for $|n| = 5$ at the discontinuity, which can have a sample value of 0, 0.5, or 1. The best approximation results from using $a_{\pm 5} = 0.5$. The pattern formed by summing component beams using Eq. (13.17) with the sample locations and values yields the pattern in Figure 13.2a. Note that, unique to this method, the synthesized pattern exactly equals the desired pattern at the sample points due to the orthogonality of the component beams. The current distribution required to realize the synthesized pattern found by evaluating Eq. (13.15) is shown in Figure 13.2b.

Examples 2 and 3 for a 10λ line source with a sector shaped desired pattern show that the Woodward–Lawson synthesized pattern provides a better approximation to the desired pattern than the Fourier transform method. Rhodes [10] extended the Woodward–Lawson method for component current functions with arbitrary decay at the edge of the aperture, including a zero edge illumination.

13.3.1.3 The Iterative Sampling Method Classical synthesis methods usually result in a pattern that does not approach the desired pattern sufficiently close over the whole pattern region. In addition, classical methods have limited ability to refine the synthesized pattern. In essence, classical methods are "one-shot" methods in the sense that if the synthesized pattern with one method is not good, a different method must be tried. Iterative methods, on the other hand, begin with a pattern that roughly approximates the desired pattern and adjustments are made to the pattern to improve the pattern fit to the desired pattern. Convergence is not guaranteed because the desired pattern can be overspecified such that the specifications cannot be realized with the aperture size; this is true for any method. If this occurs, the choices are to relax the pattern fit constraints or to increase the aperture size. A better approach is to use an iterative method with flexibility to control the pattern fit. One such method, the *iterative sampling method*, is appealing because of its simplicity and because it is applicable to nearly any antenna type, including linear arrays, linear apertures, nonuniformly spaced arrays, multidimensional antennas, and even reflector antennas with array feeds. In this section we present the basic form of the iterative sampling method for linear apertures.

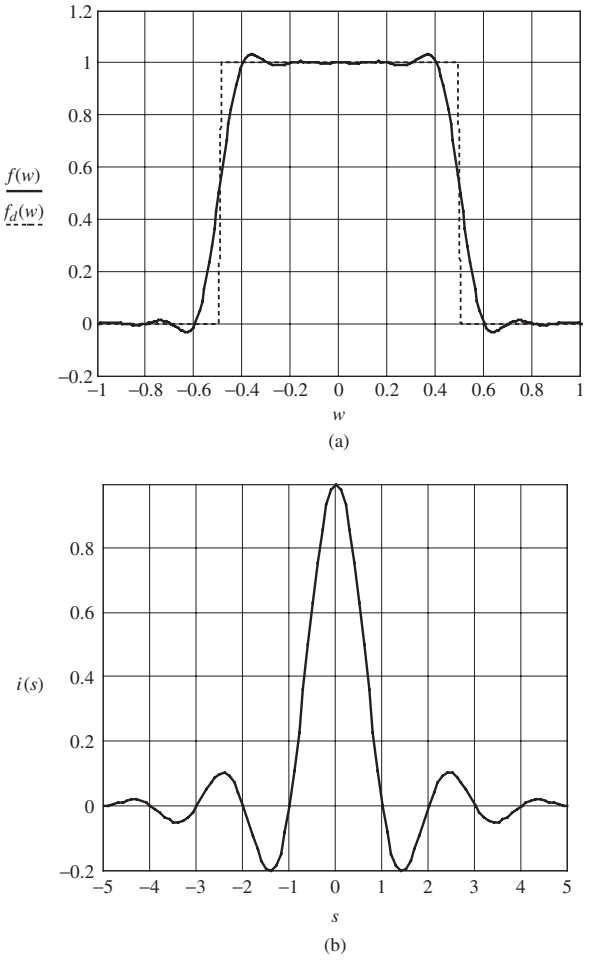


Figure 13.2 Woodward–Lawson synthesis of the sector pattern using a 10λ line source of Example 3. (a) The synthesized pattern (solid curve) and the desired sector pattern (dashed curve). (b) The current distribution for the synthesized pattern.

By extending the Woodward–Lawson method from one based on uncorrelated pattern samples to one of weakly correlated samples applied iteratively, a customized pattern fit to the desired pattern can be achieved. The goal of low MSE (which is an integrated error) between the desired and synthesized pattern is replaced by a minimax error, where the maximum deviation from the desired pattern is minimized. The maximum allowed deviation is specified over the far-field region in accordance with the criticality of the pattern fit as a function of angle. The *iterative sampling method* has its roots in the Woodward–Lawson formulation of Eqs. (13.15)–(13.18) [11]. But instead of requiring the individual pattern samples to be uncorrelated, they need only be weakly correlated. This gives the method great flexibility and permits iterative application.

The iterative sampling method begins with any initial pattern, $F^{(0)}(s)$, which is usually obtained by a classical synthesis method such as the Woodward–Lawson method, or it could even be a measured pattern. Here we are using the notation associated with the

total pattern, in order to include the element factor. The initial pattern is compared to the desired pattern at as many pattern points as appropriate. Next, correction patterns $\Delta F^{(i)}(w)$ are added to the initial pattern $F^{(0)}(w)$ to adjust the total pattern toward the desired pattern. Note that many classical synthesis methods provide an integrated error solution, whereas the iterative sampling method gives a minimax error solution by minimizing the maximum deviations from the desired pattern according to specifications by the designer, which is the most appropriate approach in most applications [12, p. 293].

The general expression for the synthesized pattern after K iterations is

$$f^{(K)}(w) = f^{(0)}(w) + \sum_{i=1}^{K} \Delta f^{(i)}(w) = f^{(0)}(w) + \sum_{i=1}^{K} \sum_{n=-M}^{M} a_n^{(i)} G_n^{(i)}(w) \tag{13.19}$$

A maximum of $2M + 1$ component correction patterns $G_n^{(i)}(w)$ are possible for each iteration, but usually fewer are used. The component corrections can be about any function with the simplest being uniform line source (sampling) patterns as used in the Woodward–Lawson method. The component patterns weighted by coefficients $a_n^{(i)}$ can be determined in a variety of ways, with the easiest being the difference between the previous iteration value and the desired pattern value at the sample point:

$$a_n^{(i)} = f_d(w_n^{(i)}) - f^{(i-1)}(w_n^{(i)}) \tag{13.20}$$

The locations of the correction patterns are often applied to points where the pattern deviates the most from the desired pattern. For blind application in computer codes, the algorithm that works the best is to use the mean value between previous sample points:

$$w_n^{(i)} = \frac{w_n^{(i-1)} + w_{n-1}^{(i-1)}}{2} \tag{13.21}$$

To maintain any pattern symmetry and to have real-valued pattern and current, corrections should be applied symmetrically about the coordinate origin. The source distribution corresponding to the synthesized pattern is simply given by the Fourier transform of Eq. (13.19):

$$i^{(K)}(s) = i^{(0)}(s) + \sum_{i=1}^{K} \sum_{n=-M}^{M} a_n^{(i)} g_n^{(i)}(s) \tag{13.22}$$

Many examples of patterns synthesized with the iterative sampling method are presented in Refs. 11, 13, and 14, including line sources, two-dimensional continuous sources, linear arrays, and planar arrays for shaped beam and/or low sidelobe applications. Thus the iterative sampling method can be used for all antenna types and desired patterns within one simple framework.

As mentioned earlier, the component illumination function need not be that of a uniform line source, as proposed by Raabe [15]. In the Woodward–Lawson method, such component illumination functions cause the method to lose its simplicity, but it can offer improvement to an iterative sampling approach with no added complexity.

13.3.1.4 The Optimum Line Source Method The Fourier transform method yields the minimum MSE between the desired and synthesized patterns but over all space, visible and invisible. The optimum line source method gives the minimum mean squared error (MSE) pattern over the visible region (θ from 0° to 180°). The method

places a constraint on the superdirective ratio, γ, which is the ratio of total power in all space to the radiated power in visible space [12, Chap. 7]. The higher the superdirective ratio value, the higher are the currents on the antenna, leading to higher ohmic losses and lower efficiency. In addition, a high superdirective ratio ($\gamma \gg 1$) is accompanied by highly oscillatory currents that are sensitive to errors, giving realization difficulties. This leads to a narrow bandwidth. Yaru [16] used numerical examples of arrays to quantify the deleterious effects of superdirectivity to show that if the required current distribution is not realized exactly, the synthesized pattern will be nowhere close to the desired pattern and the array is sensitive to excitation errors and has very low bandwidth and efficiency. Thus high superdirective ratio synthesis is mostly of theoretical interest.

Woodward and Lawson [9] in 1948 showed that for finite sized two-dimensional apertures, continuous or discrete (arrays), it is theoretically possible to produce a radiation pattern that approaches the desired pattern as closely as wanted. Rhodes [17] and Fong [18] presented mathematical approaches to synthesizing optimum line source distributions subject to not exceeding a specified superdirective ratio. The mean squared error of the optimum pattern can be evaluated using a simple technique [19].

13.3.1.5 Search Methods There are synthesis methods based on a computerized search for a solution or solutions subject to a cost function that is usually the deviation between the desired and synthesized pattern with an error limit that is either integrated over the pattern or continuously specified over the radiation space. Fong and Birgenheier [20] introduced a conjugate gradient iterative search method that miminizes an error functional such as a weighted squared error. Their method applies to line sources or linear arrays and is extendable to two-dimensional antennas. Genetic algorithms provide an organized guessing to yield a good solution, albeit not necessarily optimum [21, Chap. 9; 22].

13.3.1.6 Other Methods Hyneman and Johnson [23] developed a line source technique for achieving low ripple and sidelobes through a procedure to deterministically perturb the pattern zeros. For antennas that are many wavelengths in size, geometrical optics methods can be used for synthesis by relating a section of the source to an angular wedge in the far field [24, Sec. 3-2E]. Optical synthesis of shaped beams can be applied to reflector antennas to shape the reflector surface [21, Chap. 15; 25, Chap. 13; 26, 27]. The radiation integral of Eq. (13.4) can be recast as a Laplace transform [24, Sec. 3-2 C] and then circuit synthesis methods such as Butterworth (maximally flat) and Chebyshev (equal ripple on the main beam) can be used to synthesize patterns.

13.3.2 Shaped Beam Synthesis Methods for Linear Arrays

Many of the shaped beam pattern synthesis methods for linear arrays follow directly from corresponding line source methods; good literature summaries are found in Refs. 12 (Chap. 7), 21 (Chap. 16), 29 and 30. For terminology, we use equal and unequal to refer to array interelement spacings and uniform and nonuniform to refer to excitations.

Analogous to Eq. (13.1), the total pattern for an array antenna can be expressed as a product given by Ref. 1, Eq. (3-66)

$$F(\theta, \phi) = g_a(\theta, \phi) f(\theta, \phi) \tag{13.23}$$

where $g_a(\theta, \phi)$ is the normalized element pattern and can include average mutual coupling effects [31]. The array factor is determined by the element spacing d and excitations $\{i_n\}$ for an odd number of elements $P = 2N+1$:

$$f(w) = K \sum_{m=-N}^{N} i_m e^{j2\pi m(d/\lambda)w} \tag{13.24}$$

where K is a constant required to normalize the pattern to unity maximum. Results for an even number of elements is found in Ref. 1 (Sec. 8.3). The desired array factor f_d can be found for a known element pattern by back solving Eq. (13.23):

$$f_d(w) = \frac{F_d(w)}{g_a(w)} \tag{13.25}$$

13.3.2.1 The Fourier Series Method The observation that an array factor for a linear array has the same mathematical form as a Fourier series was first noted by Wolff [32]. The Fourier series method gives a minimum MSE pattern over one period of the array factor, which corresponds to the visible region if the element spacing is one-half wavelength. The desired pattern can be expanded into a Fourier series:

$$f_d(w) = K \sum_{-\infty}^{\infty} b_m e^{j2\pi m(d/\lambda)w}, \quad -\lambda/2d < w < \lambda/2d \tag{13.26}$$

where the Fourier coefficents are

$$b_m = \frac{d}{\lambda} \int_{-\lambda/2d}^{\lambda/2d} f_d(w) e^{-j2\pi m(d/\lambda)w} dw \tag{13.27}$$

Identifying d as the spacing between elements and $w = \cos\theta$, where θ is the angle from the line of the array, the sum in Eq. (13.26) is recognized as the array factor of an equally spaced linear array with array factor Eq. (13.24) and an infinite number of elements. An infinite array, of course, is not practical as was the infinite line source for the inverse Fourier transform of Eq. (13.9). The practical Fourier series synthesis method results by truncating the sum in Eq. (13.26), giving the array factor in Eq. (13.24) and choosing the currents equal to the Fourier series coefficients:

$$i_m = b_m, \quad 1 \le |m| \le N \tag{13.28}$$

Example 4: Fourier Series Synthesis of a Sector Pattern with a Linear Array The same pattern of Examples 2 and 3 is synthesized for a linear array of 21 elements that are half-wavelength spaced apart, giving a linear array of length $L = 10.5\lambda$, which is close to the 10λ antenna length of the previous two examples. The current values found from Eqs. (13.27) and (13.28) are 0.5 at $z = 0$, 0.318 at $z = \pm 0.5\lambda$, 0 at $z = \pm 1\lambda$, -0.106 at $z = \pm 2\lambda$, 0.064 at $z = \pm 2.5\lambda$, 0 at $z = \pm 3\lambda$, -0.046 at $z = \pm 3.5\lambda$, 0 at $z = \pm 4\lambda$, 0.035 at $z = \pm 4.5\lambda$, 0 at $z = \pm 5\lambda$. So there are only 11 elements with nonzero excitations in this example. The Fourier series synthesized pattern calculated using these currents in Eq. (13.24) is shown in Figure 13.3. Note that it is very similar to the Fourier transform synthesized pattern of Figure 13.1.

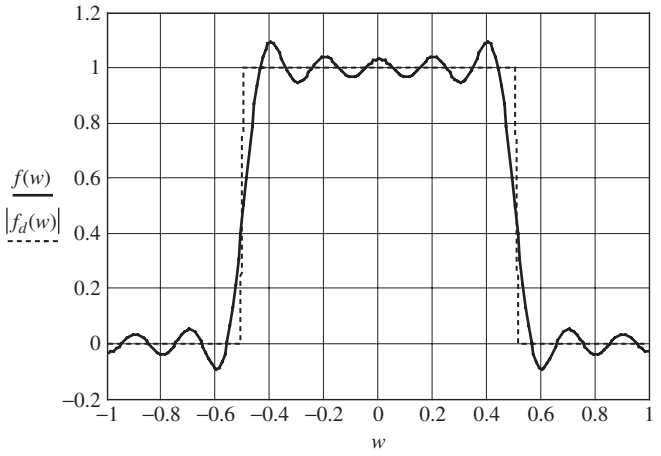


Figure 13.3 Pattern (solid curve) of the 10.5λ-long linear array of 21 half-wavelength spaced elements of Example 4 synthesized for a sector pattern shape (dashed curve).

The Fourier series method provides the best MSE approximation to the desired pattern over one array factor period. However, if the desired pattern has a discontinuity, such as with a sector pattern, there are large overshoots in the main beam and large near-in sidelobes—the Gibbs phenomenon. These oscillating deviations can be reduced by weighting the Fourier determined coefficients. One popular weighting is the Fejer sum. Other weightings have been used in array synthesis too [33–36]. Chiba [36] presents measured shaped beam patterns implemented with waveguide slot arrays.

13.3.2.2 The Woodward–Lawson Method The original Woodward–Lawson method for synthesizing patterns based on sampling methods was applied to continuous currents as in Section 13.3.1.1, but due to its simplicity and flexibility, the approach is easy to apply to arrays. Somers [37] presents formal proofs of the extension of the Woodward–Lawson method to arrays and that the correct array length to use is Nd, not $(N-1)d$. Shanks [38] showed that in the limit of infinite array length, the Woodward–Lawson method is exact. The Woodward–Lawson array synthesis method is actually a perfect match to array beam formers such as a Butler matrix or Rotman lens-fed array [39]. When the input ports to these feed networks are in proportion to the synthesized currents, the Woodward–Lawson pattern will be produced. An example similar to Example 3 in Stutzman and Thiele [1, Sec. 8.3] produces a nearly identical pattern to the Woodward–Lawson line source case.

13.3.2.3 The Iterative Sampling Method The iterative sampling method described in Section 13.3.1.3 for line sources applies in exactly the same fashion to linear arrays. Werner and Ferraro [40] show other examples of the iterative sampling method.

13.3.2.4 Array Polynomial Zero Manipulation Methods Schelkunoff [41] in 1943 originated shaped pattern synthesis for equally spaced linear arrays by manipulating the zeros of the array polynomial. To understand the technique, we first examine the roots

of the array factor. By the fundamental theorem of algebra the array factor of Eq. (13.26) can be expressed as [12, Sec. 7.10]

$$f(w) = K \sum_{m=-N}^{N} b_m z^m = K \frac{b_P}{b_1}(z - z_1)(z - z_2) \cdots (z - z_{2N}) \tag{13.29}$$

where $\{z_m = e^{j2\pi w_m d/\lambda}\}$ are the locations of the zeros of the pattern. The variable $z = e^{j(2\pi/\lambda)dw}$ has unity magnitude as $w = \cos\theta$ varies with angle θ. Thus if a zero, say, z_k, lies on the unit circle, it will produce a null in the pattern at the angle θ_k. If all zeros lie on the unit circle, there are $P-1$ nulls in the visible region and the element currents are unique; see Taylor and Whinnery [42] for this and other theorems.

Orchard and Elliott developed a synthesis method generally referred to as Orchard's method that applies to equally spaced linear arrays [43, 44]. It builds on Schelkunoff's method by manipulating the zeros of the pattern function in the complex plane using an iterative numerical method. The method works with the power pattern, which makes the problem nonlinear, leading to multiple solutions and associated multiple current distributions. Orchard's method is capable of good pattern fits, but its drawbacks are that it is computer intensive, only applies to equally spaced linear arrays, and it may converge to a local minimum rather than the global minimum [45]. If the array polynomial (i.e., the array factor) is specified in amplitude and phase in far-field directions equal in number to the number of elements, the solution is unique and can be found from matrix inversion [46]. If the selected far-field directions are equally spaced in terms of w_n, the approach is equivalent to the Fourier series method of Section 13.3.2.1. These methods do not extend to continuous current antenna methods as does the Woodward–Lawson method and its variants.

13.3.2.5 Search Methods Most of the search methods mentioned in Section 13.3.1.5 can be used for arrays also. Quadratic programming uses the mean square error between the synthesized and desired pattern as a cost function to determine an optimum excitation of a specified array geometry [21, Chap. 16; 47]. The genetic algorithm is especially well suited to array synthesis [21, Chap. 16; 48]. Mitchell et al. [49] applied the genetic algorithm to roots in the complex plane. Array element excitations and/or locations can be optimized in the presence of its environment, such as a vehicle, by working with *in situ* element patterns that include coupling and scattering effects [50]. Michelson and Schomer [51] used a nonlinear programming approach to shaped beam syntheses that treats the element amplitudes, phases, and positions as variables and permits the array geometry to be conformal to any nonlinear curve. Buckley [52] used a search method with constraints on the element currents to avoid large current variations, which he noted arise with the Orchard method. Neural network methods have been applied to arrays for shaped beam and adaptive null forming applications [53, Chap. 6]. The particle swarm method has recently been introduced to array synthesis. It provides better results than the quadratic programming method and is easier to use than the genetic algorithm [54]. The particle swarm method is easier to implement than the genetic algorithm and usually produces better results, but not always [55].

13.3.2.6 Aperture Sampling Method Complex shaped patterns synthesized for continuous antennas require current amplitude and phase distributions with many shape details that are difficult to realize in practice. But one can design an array antenna, which often is much easier to realize than a continuous source, by first synthesizing a

continuous source and then sampling the source. That is, the array element excitations are determined by equating them to samples of the aperture distribution at the element locations [12, Sec. 7.4]. The resulting pattern will in most cases closely approximate that of the continuous distribution for both shaped and narrow beam patterns [56]. Walter [24, Sec. 2-18] gives examples showing that the approximation improves with smaller element spacings.

13.3.2.7 Unequally Spaced Array Methods Unequally spaced arrays offer several advantages with narrow beam patterns (see Section 13.4.4), but are also sometimes used for shaped beam patterns. Ishimaru [57] introduced the "source position function" concept as a model function to determine positions of unequally spaced uniformly excited elements. The method was implemented experimentally for a cosecant shaped beam with a series-fed waveguide with varying width and equally spaced slots [58].

Density tapering methods produce a desired aperture amplitude distribution by spacing uniformly excited elements unequally. Density tapering is often used with narrow beam patterns (see Section 13.4.4), but there is not a direct trade-off between the element amplitude distribution and element positions. Density tapering has limited application to desired pattern shapes requiring control of the excitations.

Another way to achieve element count reduction is to aperture sample, as described in Section 13.3.2.6, in a way that elements are located at low or very low model current amplitude locations, and then remove that element [59]. Unequally spaced, shaped beam arrays can be synthesized based on a numerical approximation (i.e., quadrature) method to the radiation integral in Eq. (13.11). The approximating summation represents an array factor and the quadrature weights are the element excitations [13].

13.3.2.8 Array Synthesis Including Mutual Coupling Effects Arrays are designed for the widest possible interelement spacings to reduce the number of elements in order to limit the cost and complexity of the array system. Wide spacings reduce mutual coupling and, in most cases, permit factoring the array pattern expression into a product of an active element pattern that includes averaged mutual coupling effects and the array factor as in Eq. (13.23). Then synthesis is applied to the array factor. In some cases (especially for low sidelobe desired patterns), mutual coupling effects cannot be ignored or included in an average active element pattern, and special techniques are required.

Sahalos [6] presents a general synthesis method that applies to arrays that can be unequally spaced and multidimensional by manipulating the traditional array factor from the nonorthogonal basis functions in Eq. (13.2.4) into a form that has orthogonal basis functions and then solving for the element currents. The orthogonal method can also include mutual coupling effects and the obtained element current values compensate for coupling effects.

Arrays that can be analyzed using moment method numerical methods [1, Chap. 10] can be synthesized including all element mutual coupling and the environment around the antenna through a matrix inversion process [31, 60–62].

13.3.2.9 Pattern Synthesis for Reflector Antenna Array Feeds Array antennas placed in the focal region of a reflector antenna are used to shape the main beam of the array–reflector antenna system. Spacecraft reflector antenna requirements are the most

demanding and require sculpting the beam shape to match the contour, for example, of a land region as viewed from geostationary orbit. Dual offset reflectors with array feeds represent the state of the art [63]. Advanced applications are possible with array-fed reflector antennas, such as correcting for reflector surface distortions using an array feed designed using the iterative sampling method [64]. An example situation where this would be useful is a mesh reflector that deviates from its canonical shape due to structural details such as support ribs or due to, for example, imperfect deployment from a space platform.

13.3.2.10 Wide Bandwidth Methods Arrays with elements in a periodic grid (i.e., equal interelement spacings) have limited bandwidth, especially for phased array applications. For example, an array with half-wavelength spaced elements has full-wavelength spacings when the frequency doubles and grating lobes become significant. In practice, arrays cannot be operated with one-wavelength spacings, and spacings must be much less than a wavelength for wide angle electronic scanning [65].

There are some techniques for broadbanding arrays. Uthansakul and Bialkowski [66] discuss a wideband feeding technique for synthesizing a scanned narrow beam pattern and report an experiment demonstrating 27% bandwidth. Aperiodic grids reduce grating lobe effects and permit wider bandwidth [67]. King et al. [68] found several arrangements of unequally spaced arrays that give stable, usable patterns over a 2:1 bandwidth. Finally, elements of different sizes can be used in a shared aperture geometry to widen bandwidth to at least two octaves (4:1) [69, 70].

13.3.2.11 Other Methods Guy [29] presented a method for synthesizing arrays of elements that lie on a curve and are composed of dissimilar elements. This iterative method has been used to design many practical antennas. Baker [71] extended the work of Hyneman and Johnson, which was mentioned in Section 13.3.1.6, to arrays for synthesizing shaped patterns with low ripple. The array synthesis problem can be expanded to include optimization of other system parameters such as signal-to-noise ratio [72] and beam efficiency [73].

As noted earlier, power pattern array synthesis methods exist [12, Chap. 7; 74], but at the expense of making the synthesis problem more difficult due to the nonlinearity introduced by squaring the field pattern. There are two reasons to use field pattern synthesis, as we do in this chapter: The source distribution solution is unique and the benefits of power pattern synthesis over field pattern synthesis are marginal [46].

13.3.3 Shaped Beam Synthesis Methods for Multidimensional Antennas

Synthesis of multidimensional antennas can be treated in parts as one-dimensional synthesis problems for the principal planes if the current distributions are *separable*—that is, the current variations in each principal plane of the antenna are of the same relative shape. However, for large antennas better results are obtained by using a nonseparable distribution [4, 75]. Elliott and Stern [76] present an iterative method to synthesize sector patterns using a circular planar aperture. Richie and Kritikos [77] extended the iterative sampling method to avoid iterations using a linear programming solution of simultaneous equations for application to planar arrays. Hodges and Rahmat-Samii [78] presented a general planar array synthesis method. Elliott [4, Chap. 6] has a good treatment of planar arrays, including the *projection method*, where planar array element currents are projected onto the plane of interest and the resulting linear array is used to find the pattern

in that plane. Collin and Zucker [12, Chap. 7] also treat two-dimensional apertures and arrays. Josefsson and Persson [79] present an aperture projection method for synthesizing conformal arrays, as well as several other methods for conformal arrays, some including mutual coupling effects.

13.4 LOW SIDELOBE SYNTHESIS METHODS

Sidelobes should be of a low level when it is important to reject unwanted signals arriving outside the main beam. Moderate sidelobe reduction (below those obtained with a uniformly illuminated aperture) is required for interference control, and extreme sidelobe reduction may be necessary for a receiver to cope with high power jamming transmitters. In some cases, dynamic (adaptive) pattern control is needed to respond to changing interference/jamming conditions. Low sidelobe designs can be combined with a shaped main beam, with multiple main beams, or, more commonly, with a single pencil beam. A consequence of low sidelobes is reduced directivity due to main beam widening.

13.4.1 Taylor Line Source Methods

The optimum (narrowest) beamwidth for a given sidelobe level occurs when all sidelobes are of equal level [80]. The ideal Taylor line source, although unrealizable, provides such a pattern. However, a realizable current distribution is obtained when the perfectly equal sidelobe level requirement is relaxed somewhat to produce the Taylor line source with approximately equal sidelobes. While it is hard to achieve the required continuous current distribution in practice for Taylor line source forms, they are often used as a model function for sampling to produce an array excitation.

The *ideal Taylor line source* pattern has all sidelobes of equal peak level $1/R$ relative to the main beam peak, giving a sidelobe level (SLL) in decibels as

$$SLL = -20 \log R \quad \text{(dB)} \tag{13.30}$$

where R is the ratio of the main beam peak to peak sidelobe level values.

The pattern is given by

$$f(w) = \begin{cases} \dfrac{\cos\left[\pi\sqrt{(L/\lambda)^2 w^2 - A^2}\right]}{\cosh \pi A}, & |w| \geq A\lambda/L \\ \dfrac{\cosh\left[\pi\sqrt{A^2 - (L/\lambda)^2 w^2}\right]}{\cosh \pi A}, & |w| \leq A\lambda/L \end{cases} \tag{13.31}$$

This pattern is unity for $w = 0$ and is a function of the sidelobe level through the parameter of convenience. A defined with the following relation:

$$R = \cosh \pi A \tag{13.32}$$

In the sidelobe region, this pattern oscillates between $-1/R$ and $+1/R$ to infinity in w, leading to infinite power in the pattern and consequently infinite power in the source due to current singularities at the aperture edges. Clearly, this ideal pattern cannot be realized in exact form.

For large w, the argument of the cosine function in Eq. (13.31) for the ideal Talyor line source pattern is approximately $\pi wL/\lambda$. So the pattern zero locations, which occur for cosine argument multiples of π, are $w_n \cong \pm\lambda(n - \frac{1}{2})/L$, which are regularly spaced. Taylor [80] devised an approximate pattern, usually referred to as the *Taylor line source*, based on modifying the interior zero locations by a scale factor σ, giving

$$w_n = \begin{cases} \pm(\lambda/L)\sigma\sqrt{A^2 + \left(n - \frac{1}{2}\right)^2}, & 1 \le n < \overline{n} \\ \pm(\lambda/L)n, & \overline{n} \le n < \infty \end{cases} \tag{13.33}$$

where

$$\sigma = \frac{\overline{n}}{\sqrt{A^2 + \left(\overline{n} - \frac{1}{2}\right)^2}} \tag{13.34}$$

The zero expressions in Eq. (13.33) are equal for $n = \overline{n}$. The scale factor σ is only slightly above unity, so the pattern is close to the ideal Taylor line source inside the transition point of $n = \overline{n}$. Values of σ for various SLL and $\overline{n}$ values are found in Table 13.2. Beyond nth zero the pattern zeros align with those of the *Sa* function, giving a sidelobe envelope of $1/w$. This sidelobe decay leads to a reduction of the current distribution edge singularities to secondary peaks with height depending on the SLL.

Taylor line sources are actually a member of the Woodward–Lawson family, as shown in [1, Sec. 8.4], and the pattern and corresponding current distribution are given by

$$f(w) = \sum_{n=-\overline{n}+1}^{\overline{n}-1} f(n, A, \overline{n}) Sa\left[(w - \frac{\lambda}{L}n)\frac{L}{\lambda}\pi\right] \tag{13.35}$$

$$i(s) = \frac{\lambda}{L}\left[1 + 2\sum_{n=1}^{\overline{n}-1} f(n, A, \overline{n}) \cos\left(2\pi\frac{\lambda}{L}ns\right)\right] \tag{13.36}$$

Formulas [1, Sec. 8.4] and tables [81, Appendix I] are available to determine the coefficients $f(n, A, \overline{n})$, which are samples of the Taylor line source pattern. For the ideal Taylor line source, $\overline{n} = \infty$ and the summations extend to infinity—but the pattern is more simply expressed by Eq. (13.31).

The half-power beamwidth expression for the Taylor line source in angle space (i.e., not the space variable w) is

$$HP \approx 2\sin^{-1}\left\{\frac{\lambda\sigma}{L\pi}\left[(\cosh^{-1} R)^2 - (\cosh^{-1}(R/\sqrt{2})^2\right]^{1/2}\right\} \tag{13.37}$$

This formula applies to the ideal Taylor line source by setting σ to unity.

TABLE 13.2 Design Parameters for Taylor Line Sources

SLL		σ							
(dB)	A	$\overline{n} = 3$	$\overline{n} = 4$	$\overline{n} = 5$	$\overline{n} = 6$	$\overline{n} = 7$	$\overline{n} = 8$	$\overline{n} = 9$	$\overline{n} = 10$
−20	0.9528	1.1213	1.1027	1.0870	1.0749	1.0655	1.0582	1.0522	1.0474
−25	1.1366	1.0924	1.0870	1.0773	1.0683	1.0608	1.0546	1.0495	1.0452
−30	1.3200		1.0693	1.0662	1.0608	1.0554	1.0505	1.0463	1.0426
−35	1.5033			1.0539	1.0523	1.0492	1.0459	1.0426	1.0397
−40	1.6865				1.0430	1.0424	1.0407	1.0386	1.0364

Example 5: A 5λ Taylor Line Source with −30-dB Sidelobes In this design example, the near-in sidelobes of a 5λ-long line source are to be approximately −30 dB. A Taylor line source with $\overline{n} = 4$ from Eq. (13.32) has $A = 1.32$ and from Eq. (13.34) $\sigma = 1.0693$. From formulas in Stutzman and Thiele [1, Sec. 8.4] or using the tables in Hansen [81, Appendix I]

$$f(n, A, \overline{n}) = \{0.29266, -0.01578, 0.00218\}$$

The pattern and current distribution using Eqs. (13.35) and (13.36) are shown in Figure 13.4. The half-power beamwidth from Eq. (13.37) is HP = 12.98°; the corresponding ideal Taylor line source has a beamwidth of 12.13°.

There is another Taylor line source method referred to as the one-parameter Taylor line source method, which has a sampling function type pattern but with the first sidelobe level below that for the uniform line source of −13.3 dB [81, Chap. 1].

13.4.2 Linear Array Methods

13.4.2.1 Dolph–Chebyshev Method Dolph [82] exploited the equal ripple properties of Chebyshev polynomials for the design of linear arrays with a narrow main beam

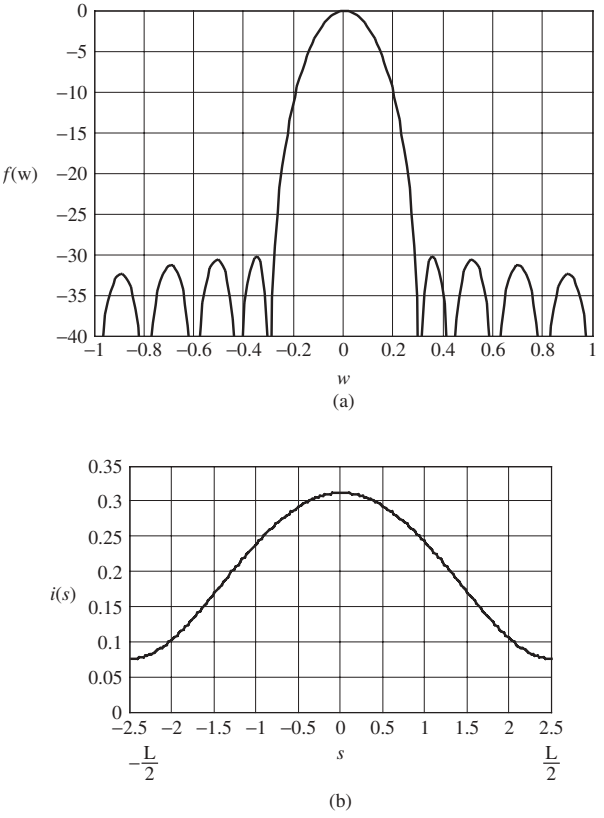


Figure 13.4 The (a) radiation pattern and (b) current distribution for the 5λ Taylor line source of Example 5 for SLL = −30 dB, $\overline{n} = 4$.

and equal sidelobes. Because of the trade-off between beamwidth and sidelobe level, an equal sidelobe pattern yields the narrowest beamwidth possible for a given sidelobe level. So the *Dolph–Chebyshev array* provides the optimum compromise between beamwidth and sidelobe level. Of course, the following fundamental antenna principle still holds: lower sidelobes are associated with a wider main beam and require a more tapered current distribution. This section presents the essential formulas required for designing Dolph–Chebyshev arrays. Derivation details are found in the literature. [1, Sec. 8.4; 12, Sec. 5.10; 82]. Villeneuve [83] developed a method that is a discrete array equivalent to the Taylor method for continuous line sources.

The Dolph–Chebyshev synthesis process begins by specifying a main beam peak to sidelobe ratio, R, and then finding the array currents that yield the narrowest beamwidth for the number of elements P. Alternatively, one could specify the beamwidth and find the lowest sidelobe design. The array factor is expressed as a Chebyshev polynomial:

$$f(\psi) = T_{P-1}\left(x_0 \cos\frac{\psi}{2}\right) \tag{13.38}$$

where the pattern variable is $\psi = 2\pi(d/\lambda)w$. The main beam peak value is not normalized to unity as usual, but the peak value is R and occurs for $\psi = 0$:

$$f(\psi = 0) = R = T_{P-1}(x_0) = \cosh[(P-1)\cosh^{-1} x_0] \tag{13.39}$$

where the definition of Chebyshev polynomial was used to obtain the last form [1, Sec. 8.4]. Solving this equation for x_0 gives

$$x_0 = \cosh\left(\frac{\cosh^{-1} R}{P-1}\right) \tag{13.40}$$

Summarizing the synthesis procedure, first the sidelobe level SLL is specified and R is calculated from Eq. (13.30). Then x_0 is calculated using Eq. (13.40). Finally, the array factor is evaluated from Eq. (13.38).

There are several ways to find the element currents. Dolph [82] presented a method for calculating the currents, but it is difficult to evaluate and is inaccurate for large arrays. A second method involves manipulating the array factor to be in the same form as the Chebyshev polynomial and equating the coefficients of each term to find the element currents [1, Sec. 8.4; 84] Third, direct evaluation methods are appropriate for most cases; see Safaai-Jazi [85], Bresler [86], and Elliott [89, Sec. 5.3]. Finally, a formula by Bresler [86] for the currents based on the work of Van der Mass [87] is the most accurate [88]. Table 13.3 gives the currents for popular SLL and P values computed using Bresler's formula.

The Dolph–Chebyshev pattern expression depends only on the SLL and the number of elements, not the element spacing and, thus, array length. The narrowest beamwidth in the visible region while maintaining the desired SLL is the practical optimum solution. This is obtained by increasing the element spacing to a point where a piece of the grating lobe equaling the SLL is just visible. This optimum spacing is [85]

$$d_{\text{opt}} = \lambda\left[1 - \frac{\cos^{-1}(1/\gamma)}{\pi}\right], \quad \text{broadside} \tag{13.41a}$$

$$d_{\text{opt}} = \frac{\lambda}{2}\left[1 - \frac{\cos^{-1}(1/\gamma)}{\pi}\right], \quad \text{endfire} \tag{13.41b}$$

TABLE 13.3 Current Magnitudes for Equally Spaced Dolph-Chebyshev Linear Arrays

	SLL (dB)				
P	−20	−25	−30	−35	−40
3	1.6364	1.7870	1.8774	1.9301	1.9604
4	1.7357	2.0699	2.3309	2.5265	2.6688
5	1.9319	2.5478	3.1397	3.6785	4.1480
	1.6085	2.0318	2.4123	2.7401	3.0131
6	1.8499	2.5876	3.3828	4.1955	4.9891
	1.4369	1.8804	2.3129	2.7180	3.0853
7	1.8387	2.7267	3.7846	4.9811	6.2731
	1.6837	2.4374	3.3071	4.2625	5.2678
	1.2764	1.7081	2.1507	2.5880	3.0071
8	1.7244	2.6467	3.8136	5.2208	6.8448
	1.5091	2.2296	3.0965	4.0944	5.1982
	1.1386	1.5464	1.9783	2.4205	2.8605
9	1.6627	2.6434	3.9565	5.6368	7.6989
	1.5800	2.4751	3.6516	5.1308	6.9168
	1.3503	2.0193	2.8462	3.8279	4.9516
	1.0231	1.4036	1.8158	2.2483	2.6901
10	1.5585	2.5318	3.8830	5.6816	7.9837
	1.4360	2.2770	3.4095	4.8740	6.6982
	1.2125	1.8265	2.5986	3.5346	4.6319
	0.9264	1.2802	1.6695	2.0852	2.2182

(The central element current values are listed first and the edge elements have unity magnitude.)

where

$$\gamma = \cosh\left[\frac{1}{P-1}\ln(R+\sqrt{R^2-1}\right] \tag{13.41c}$$

Optimum patterns are possible for broadside arrays with spacings not less than a half-wavelength and for endfire arrays with no lower limit on spacing [12, Sec. 5.10].

Accurate formulas for the beamwidth are available, but the following is a good approximation [89]

$$HP \approx b_{\text{HP}}\ 0.886\frac{\lambda}{L}$$

$$b_{\text{HP}} = 1 + 0.636\left\{\frac{2}{R}\cosh\left[\sqrt{(\cosh^{-1}(R))^2 - \pi^2}\right]\right\}^2, \qquad \text{broadside} \tag{13.42}$$

where b_{HP} is the beam broadening factor relative to a uniform line source of length $L = Pd$. Exact methods are available to calculate the directivity of Dolph–Chebyshev arrays [90, 91], but the following formula provides an approximation [1, p. 382]:

$$D \approx \frac{2R^2}{1 + R^2 HP}, \qquad \text{broadside} \tag{13.43}$$

Another useful approximate relation is [4, p. 157]

$$D \approx \frac{100}{HP_\circ} = \frac{1.75}{HP} \tag{13.44}$$

where $HP_\circ$ and HP are the half-power beamwidths in degrees and in radians, respectively. Interestingly, the directivity value is identical for broadside and endfire cases for both half-wavelength and optimum spacings [90]. At broadside, the beamwidth is narrow, but the three-dimensional pattern is a fan beam. The main beam is broader at endfire but is a pencil beam. In general, the directivity of an endfire array equals that of the broadside array with the same number of elements and sidelobe level, but with twice the spacing [90]:

$$D_{\text{endfire}}\,(d) = D_{\text{broadside}}\,(2d) \tag{13.45}$$

Example 6: A 5λ Dolph–Chebyshev Array with −30-dB Sidelobes In order to have a direct comparison with Example 5, the same length antenna of 5λ and −30-dB ($R = 31.6$) sidelobes are used in this Dolph–Chebyshev array example. The 10-element, half-wavelength spaced array has an excitation taper from center to edge of the array from Table 13.3

$$3.8830 : 3.4095 : 2.5986 : 1.6695 : 1.0000 \tag{13.46}$$

Figure 13.5 shows the pattern calculated with Eq. (13.24) using these currents. Note the similarities between the patterns of Figures 13.4a and 13.5. The half-power beamwidth from Eq. (13.42) is 11.6° and is 13.0° for the comparable Taylor line source of Example 5. The directivity from Eqs. (13.43) and (13.44) are 9.9 dB and 9.4 dB, respectively. The ratio of maximum to minimum current for the Taylor line source of Figure 13.4b is about 4.2 and is 3.9 for this example.

13.4.2.2 Sampled Aperture Method Just as with shaped beam synthesis discussed in Section 13.3.2.6, spatial samples of the continuous aperture distribution can used as element excitations in array to produce a good approximation to the desired pattern. Elliott [4, Sec. 5.10] gives examples of array designs based on aperture sampling. As a note of caution, if the sampling interval (i.e., element spacing) becomes too large, grating lobe effects (i.e., aliasing effects) become significant.

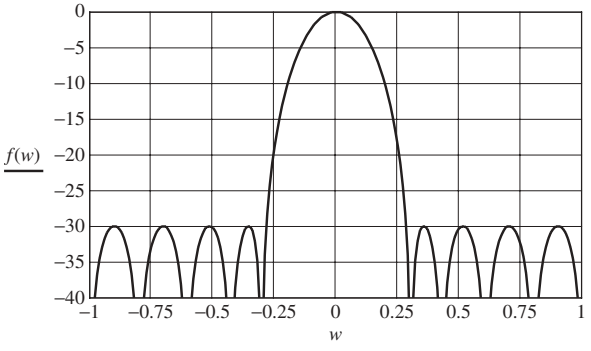


Figure 13.5 Radiation pattern for the Dolph–Chebyshev array of 10 half-wave spaced elements in Example 6 for SLL = −30 dB.

13.4.3 Sidelobe Envelope Control Methods

Equal sidelobe levels are desired when interfering signals can arrive from any direction outside the main beam. In cases where specific angular regions are of more concern than others, it is best to lower sidelobes more in those regions and allow remaining sidelobes to be higher. This approach will yield a main beam that is narrower than if the sidelobes were all reduced. In addition, the excitation currents will have less taper and be easier to implement. The iterative sampling method developed in Sections 13.3.1.3 and 13.3.2.3 for line source and linear array shaped beam patterns is also directly applicable to low sidelobe patterns, again for both line sources and arrays. Several example results for line sources and arrays with a variety of sidelobe envelope functions are found in Stutzman [13]. As with any method, if constraints on the pattern are severe in one region, the achievable quality of the pattern will degrade in other regions [92].

Hyneman [93] and Elliott [94] presented methods to customize the envelope of the sidelobes using deterministic perturbation of the Taylor line source pattern zero locations. Ma [95] treats several low sidelobe array synthesis methods, including endfire arrays.

13.4.4 Low Sidelobe Unequally Spaced Array Methods

The number of elements in an unequally spaced array, especially for large arrays, can be reduced significantly from that for a comparable equally spaced array. Wider average element spacing can be used because the grating lobes that appear with equally spaced elements tend to spread out and lower, forming a grating plateau [96]. The disruption of grating lobes found in equally spaced arrays is especially important when scanning the pattern. Also, the array bandwidth is increased when using unequally spaced elements instead of a periodic grid; see Section 13.3.2.10. Element locations for unequally spaced arrays can be determined by either deterministic or probabilistic methods. An excellent treatment is found in Collin and Zucker [12, Chap. 6].

Deterministic methods provide a procedure of uniquely determining element locations rather than using a trial-and-error approach. Unz [97] first recognized the value of using unequal element spacings. Unequally spaced arrays are sometimes used in order to preserve uniform excitation amplitudes for feed simplicity. For example, to realize an effective tapered amplitude illumination, uniformly excited elements can be spaced increasingly farther apart with greater distance from the center of the array, producing *density taper thinning* or *space tapering* [12, Chap. 6; 98]. Uniform excitations ease the feed implementation and yield the largest directivity for the number of elements. Density thinning methods are most appropriate for large arrays, especially large planar arrays. Adaptive array methods have been applied to low sidelobe patterns realized with unequally spaced arrays; the element patterns can be included and can be different for different elements [99].

Ishimaru [57] introduced the "source position function" concept as a model function to determine positions of unequally spaced, uniformly excited elements and showed that any sidelobe level can be achieved provided the number of elements is large enough and the average element spacing is less than a half-wavelength. Ishimaru gave a method for an unequally spaced, uniform amplitude array to approximate a Taylor type narrow beam, low sidelobe pattern.

In probabilistic array design, a normalized excitation amplitude illumination model function is interpreted as a probability density function for determining whether or not an element should be placed at discrete allowed element locations. Such arrays are often

called "random arrays," but there is really nothing random about the array, just the method used to obtain the element locations. One could also consider trial-and-error and perturbational methods in this category. Lo and Lee [100] quantized the allowed positions of elements in a linear aperture to examine all possible array configurations and found few differences in the patterns except for near-in sidelobes. In a classic work, Lo [101] developed statistical quantities to evaluate density tapered linear arrays with many randomly placed elements and found that mainly the number of elements determines sidelobe level and aperture size mainly determines beamwidth. Mailloux and Cohen [102] evaluated statistical average parameters for planar arrays with circular symmetry and density taper thinned with quantized element amplitudes.

Statistical methods for unequally spaced arrays are also useful in analyzing errors in array element positions and excitations. This is of value because it is important to know how accurately elements must be positioned and to estimate array performance as a function of element position and excitation errors, including the total failure of an element [12, Chap. 6; 103].

13.4.5 Low Sidelobe Synthesis Methods for Multidimensional Antennas

Just as for shaped beam patterns, synthesis of multidimensional antennas for low sidelobes can be treated in parts composed of one-dimensional synthesis problems for the principal planes if the current distributions are separable. For example, Dolph–Chebyshev or Taylor distributions can be applied to the x- and y-dimensions of a planar aperture in the xy-plane to produce the canonical low sidelobe structures in the principal pattern planes. In other directions, the sidelobes are a product of the low sidelobe principal plane patterns, producing very low sidelobes. This will make the main beam wider than it needs to be for the sidelobe levels in the principal planes. Autrey [104] presented an approximate method for synthesizing low sidelobe patterns for rectangular arrays with nonseparable distributions by extending the method of Villeneuve [83]. Hodges and Rahmat-Samii [78] presented an integrated aperture sampling method for the synthesis of low sidelobe planar arrays.

Circular apertures can be synthesized for low sidelobes using a method analogous to the Taylor line source method [79; 105, pp. 67–71; 106] Elliott [4, Chap. 6] reviews and extends these methods for planar arrays. Also, there is a one-parameter circular aperture distribution analogous to a Taylor one-parameter line source method [107]. Ma [95, Chap. 4] presents methods for the synthesis of two-dimensional arrays, including ring arrays. Josefsson and Persson [79, Chap. 10] gives a number of methods for synthesizing low-sidelobe arrays constrained to a three-dimensional surface such as a cylinder.

13.5 POLE–RESIDUE METHODS FOR TIME- AND FREQUENCY-DOMAIN MODELING AND SYNTHESIS

The singularity expansion method (SEM) allows modeling of the antenna response in both the time and frequency domains with only one set of parameters—poles and residues. A pole–residue model of the antenna realized effective length is a complete antenna model. In essence, once the model is available, one can obtain antenna pattern, directivity, and gain in the frequency domain, as well as the radiated transient waveform for an arbitrary excitation and an arbitrary antenna orientation. Modeling and synthesis of the antenna

using the SEM point of view is discussed in this section. The SEM is not a numerical method or analytical formulation, but rather is a representation of the antenna in terms of the s-plane singularities and has a physical basis [109, 110].

Two common methods used to obtain a pole–residue antenna representation are discussed here: the matrix-pencil method and Prony's method. Next, pole–residue linear array pattern synthesis using Prony's method is described. Finally, some results are presented on antenna modeling and synthesis using antenna effective length.

13.5.1 Singularity Expansion Method (SEM) Formulation

The singularity expansion method is used to characterize the electromagnetic response of structures (e.g., aircraft, antennas) in both the time and complex frequency domains. SEM was first introduced by C. E. Baum and was inspired by typical transient responses of various complicated scatterers [109, 110]. The measured transient responses were usually a few cycles of damped sinusoids. The two-sided Laplace transform† of a damped sinusoid has one real pole or a pair of conjugate poles in the complex frequency plane. Thus the resonant frequencies of the specific object are expected to correspond to frequencies near these poles.

The development of the general SEM equation begins with the integral equation in terms of the source current density, which requires a dyadic operator. The SEM formulation presented below follows the development of C. E. Baum [109]. The following expression is a convenient starting point [109]:

$$\int_V \overline{\overline{\mathbf{\Gamma}}}(\mathbf{r}, \mathbf{r}', s) \cdot \mathbf{J}(\mathbf{r}, s)\, dv' = \mathbf{I}'(\mathbf{r}, s) \tag{13.47}$$

In this relationship, $\overline{\overline{\mathbf{\Gamma}}}(\mathbf{r}, \mathbf{r}', s)$ is the dyadic Green's function (Baum calls this the dyadic kernel of the integral equation) and $\mathbf{I}'(\mathbf{r}, s)$ is a forcing function (e.g., electric field). Singular forms of the spatial integral may be required for proper definition in the region of the spatial singularity. The current source $\mathbf{J}(\mathbf{r}, s)$could be a volume current density, a surface current density, or a line current.

To obtain the most convenient expression for the development of the general SEM equations, the forcing function $\mathbf{I}'(\mathbf{r}, s)$ needs to be replaced with an appropriately normalized temporal delta function or frequency equivalent. With this new formulation Eq. (13.47) can be written

$$\int_V \overline{\overline{\mathbf{\Gamma}}}(\mathbf{r}, \mathbf{r}', s) \cdot \mathbf{U}(\mathbf{r}, s)\, dv' = \mathbf{I}'(\mathbf{r}, s) \tag{13.48}$$

where $\overline{\overline{\mathbf{\Gamma}}}$ represents positions within the source and $\mathbf{r}$ represents the observation location. $\mathbf{U}(\mathbf{r}, s)$ are surface currents induced by a delta function excitation on finite-size perfectly conducting objects in free space. This equation can be solved numerically using the method of moments (MoM) by first converting it into a matrix equation:

$$\left[\overline{\overline{\Gamma}}_{n,m}(s)\right] \cdot [U_n(s)] = [I_n(s)], \quad n, m = 1, 2, 3, \ldots, N \tag{13.49}$$

†The two-sided Laplace transform of an arbitrary function $f(t)$ is defined as $F(s) = \int_{-\infty}^{\infty} f(t)e^{-st}dt$, where $s = \sigma + j\omega$ is a complex variable.

The $\left[\overline{\overline{\Gamma}}_{n,m}(s)\right]$ are analytic functions of s and are related to the dyadic Green's function. Baum expanded the currents induced on finite-size perfectly conducting objects in free space in a pole series as

$$\mathbf{U}(\mathbf{r}, s) = \sum_{\alpha} \frac{\eta_\alpha(1_0, s)\mathbf{v}_\alpha(\mathbf{r})}{(s - s_\alpha)^{m_\alpha}} + W(1_0, s) \tag{13.50}$$

In this expression η_α is a coupling coefficient, v_α is natural mode, s_α is a natural frequency (pole), 1_0 is unity dyadic, m_α is a positive integer (often unity), and W is an entire function. The coupling coefficients depend on the antenna orientation while the poles are independent of antenna orientation.

In SEM computations, the incident waveform is usually known. Thus SEM is used to represent an antenna response in terms of the natural frequencies s_α, coupling coefficients η_α, and natural modes v_α. The natural modes and natural frequencies can be calculated from

$$\int_V \overline{\overline{\Gamma}}(\mathbf{r}, \mathbf{r}', s_\alpha) \cdot \mathbf{v}_\alpha(\mathbf{r}') = 0 \tag{13.51}$$

where s_α define the null space of the operator. The MoM formulation of this equation is

$$\left[\overline{\overline{\Gamma}}_{n,m}(s_\alpha)\right] \cdot [v_n]_\alpha = [0_n] \tag{13.52}$$

The natural frequencies are found through

$$\det\left[\overline{\overline{\Gamma}}_{n,m}(s_\alpha)\right] = 0 \tag{13.53}$$

To find the coupling coefficients, it is useful to find the coupling vectors from

$$\int_V \mu_\alpha(\mathbf{r}) \cdot \overline{\overline{\Gamma}}(\mathbf{r}, \mathbf{r}', s_\alpha) dv' = 0 \tag{13.54}$$

or

$$[\mu_n]_\alpha \cdot \left[\overline{\overline{\Gamma}}_{m,n}(s_\alpha)\right] = [0_n] \tag{13.55}$$

From these coupling vectors, Baum shows that the coupling coefficients are given by

$$\eta_\alpha(s_\alpha) = \frac{\int_V \mu_\alpha(\mathbf{r})\vec{\mathbf{I}}'(\mathbf{r}, s) dv}{\int_V \mu_\alpha(\mathbf{r}) \int_{V'} \frac{\partial}{\partial s}\overline{\overline{\Gamma}}(\mathbf{r}, \mathbf{r}', s) v_\alpha(\mathbf{r}')\, dv'\, dv} \tag{13.56}$$

Once the poles, coupling coefficients, and natural modes are obtained, one is able to describe the antenna response.

Tesche used Pocklington's equation to model the current on a dipole in terms of poles and residues [111, 112]. Figure 13.6 shows a typical pole plot of a thin wire dipole antenna of length L. The SEM representation of the current in terms of poles and residues provides some interesting observations. First, for the late time response ($ct/L > 3$) only 3–4 poles may be required for an accurate description. The complete time response can be accurately described with 10 poles for practical systems. The advantage of fewer poles is that only a few parameters are needed to represent the antenna current. Second, it is important to not that any pole contributing to radiation cannot lie on the $j\omega$-axis. The reason for this condition is that the radiation process requires the current on the wire

to eventually reach zero value as power is radiated. Next, increasing the wire diameter moves the poles further away from the $j\omega$-axis. This indicates that broadband antenna structures have poles located further from the $j\omega$-axis than with narrowband antenna structures. A more computationally efficient method is the time-domain SEM proposed by Cordaro and Davis [113].

Van Blaricum and Mittra [114] numerically obtained a time-domain current on a dipole antenna and used Prony's method to obtain a pole–residue representation of the current. Hua and Sarkar [115] introduced the matrix-pencil method as a more elegant way of obtaining the poles and residues of the system. This method is a one-step process, compared to Prony's method where the poles are obtained by finding the roots of a characteristic polynomial. In the next section, we present derivations of the matrix-pencil method and Prony's method for obtaining a pole–residue representation of the antenna response (e.g., current, radiated field, pattern).

13.5.2 The Matrix-Pencil Method

The matrix-pencil method was introduced by Sarkar and others [115, 116, 128]. They defined the noise affected data as

$$y(t_k) = \sum_{n=1}^{N} R_n e^{s_n k \Delta t} \tag{13.57}$$

where $k \in (0, K)$ (K is the number of data samples), R_n are the complex residues, s_n are the complex poles, and Δt is the sampling interval. It is important to note that $z_n = e^{s_n \Delta t}$ are the poles on the z-plane.

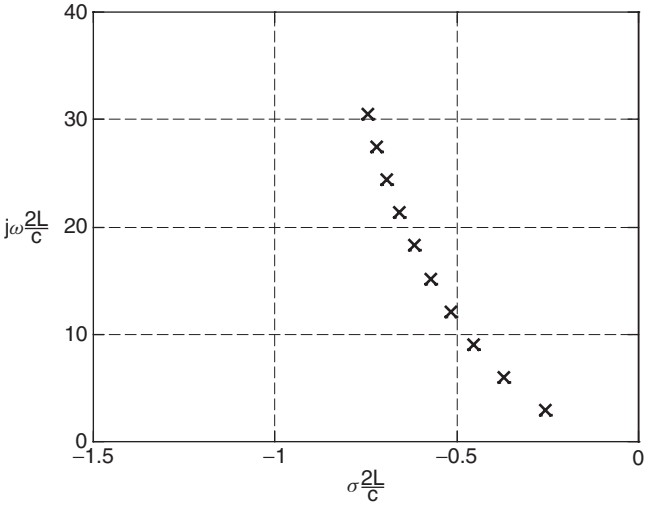


Figure 13.6 Typical pole plot in the s-plane ($s = \sigma + j\omega$) of a dipole antenna of length L, c is wave velocity. (From Ref. 111.)

It is possible to create a data matrix $[Y]$ for the K elements in terms of an $(L+1) \times (L+1)$ matrix where $L = K/2$:

$$[Y] = \begin{bmatrix} y(0) & y(1) & \cdots & y(L) \\ y(1) & y(2) & \cdots & y(L+1) \\ \vdots & \vdots & & \vdots \\ y(L) & y(L+1) & \cdots & y(2L) \end{bmatrix}_{(L+1)\times(L+1)} \tag{13.58}$$

The singular value decomposition (SVD) of the matrix $[Y]$ can be written

$$[Y] = [U][\Sigma][V] \tag{13.59}$$

A significant parameter, N, is chosen such that the singular values in Σ beyond N are small and can be approximated as zero. The value of N is typically selected by examining the ratio between the maximum singular value and all other singular values in the matrix:

$$\frac{\sigma_N}{\sigma_{\max}} \approx 10^{-p} \tag{13.60}$$

where p is the number of significant decimal digits in the data.

Next, a reduced matrix $[V']$ is constructed using only the rows corresponding to N dominant singular values:

$$[V'] = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{bmatrix} \tag{13.61}$$

The singular vectors that correspond to small singular values are discarded. Submatrices V_1 and V_2 are defined from V by deleting the first and last column of V', respectively. The problem then reduces to solving a left-hand eigenvalue problem given by

$$v[V_2] = zv[V_1] \Rightarrow v[V_2][V_1]^H = zv[V_1][V_1]^H \tag{13.62}$$

where the H refers to the Hermitian transpose. The eigenvalues z correspond to the poles of the system in terms of $e^{s\Delta t}$. Once the reduction to N has been performed and the corresponding poles are obtained, the residues R_i are found as the least squares solution to

$$\begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(K-1) \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_N \\ \vdots & \vdots & & \vdots \\ z_1^{K-1} & z_2^{K-1} & \cdots & z_N^{K-1} \end{bmatrix} \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_N \end{bmatrix} \tag{13.63}$$

Improvements are possible by deleting poles with low residue values and recomputing the residues using the retained poles.

13.5.3 Prony's Method

In 1795, Baron de Prony introduced a technique for modeling sampled data as a linear combination of damped exponentials [117]. Prony's method was used in the description of laws governing the expansion of gases. This method has been applied to various areas, notably electromagnetic scattering and antenna problems.

We start with a basic derivation of Prony's method. For particular applications, it is desired to determine an approximation of the form

$$v(t) = \sum_{n=1}^{N} R_n e^{s_n t}, \quad t > 0 \tag{13.64}$$

In practice, we usually deal with discrete and sampled data, so it is useful to express Eq. (13.64) as [118]

$$v(t_k) = \sum_{n=1}^{N} R_n e^{s_n k \Delta t} = \sum_{n=1}^{N} R_n z_n^k, \quad k \in (0, K) \tag{13.65}$$

where $z_n = e^{s_n \Delta t}$ and the time sampling interval is Δt.

Prony's method is separated into two problems. The first problem consists of determining the poles (s_n), and the second problem determines the corresponding residues (R_n). Prony has shown that there exists a vector $\{\alpha\}$ with elements $\alpha_1, \alpha_2, \ldots, \alpha_N$ such that [118]

$$\sum_{n=0}^{N} \alpha_n v_{n+m} = 0, \quad m \in (0, M-1) \tag{13.66}$$

If we set the leading coefficient $\alpha_N = 1$, without loss of generality, we obtain the following expression [118]:

$$\sum_{n=0}^{N-1} \alpha_n v_{n+m} = -v_{N+m} \tag{13.67}$$

In this expression $\{\alpha_n\}$ are unknown coefficients and $\{v_{n+m}\}$ are the sampled transient data values. If $K = 2N$ in this matrix system, then the solution is just the inverse of a real symmetric matrix. If $K > 2N$, then the solution is obtained through a least squares problem. Substituting Eq. (13.65) into Eq. (13.66), we obtain the polynomial

$$\sum_{n=0}^{N} \alpha_n z^n = 0, \quad \{z : z_1, \ldots, z_N\} \tag{13.68}$$

The roots of this polynomial are obtained by a root finding routine. From the roots we obtain the pole values:

$$s_n = \frac{\ln z_n}{\Delta t} \tag{13.69}$$

The corresponding residues are obtained from Eq. (13.64). Figure 13.7 illustrates the parameters in the Prony's method algorithm.

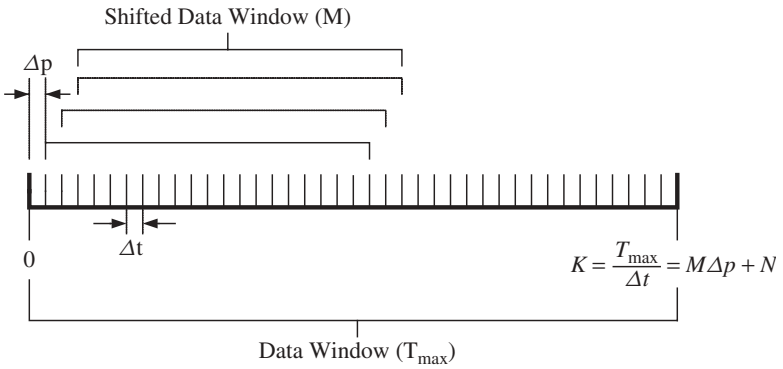


Figure 13.7 Parameters in Prony's method algorithm implementation.

13.5.4 Pole–Residue Linear Array Synthesis Using Prony's Method

Using a pole–zero (pole–residue) procedure for array synthesis, one needs only to specify the array pattern and the number of source elements. The synthesis procedure provides both source locations along with their excitations. The pole–residue modeling approach for linear array synthesis was considered by Miller [119, 120] and Prony's method was used to obtain the array source locations and the strength of the source.

The array factor of a linear array of N sources can be expressed as [12]:

$$f(\theta) = \sum_{n=1}^{N} R_n e^{j\beta d_n \cos\theta} \tag{13.70}$$

where d_n is the position of the nth source on the z-axis, R_n is the strength of the source, β is the wavenumber, and θ is observation angle. Now one is able to provide estimates for locations d_p and strengths R_p of N discrete sources using Prony's method with the following pattern representation [119]:

$$f(\theta) \approx f_{\text{est}}(\theta) = \sum_{p=1}^{N} R_p e^{j\beta d_p \cos\theta}, \quad 0 < \theta < \pi \tag{13.71}$$

One can follow Prony's development in Eqs. (13.64)–(13.69) to extract residues and poles (source locations) by defining $z_p = e^{j\beta d_p \Delta}$ in Eq. (13.65). The pattern sampling interval is Δ. The desired pattern needs to be uniformly sampled in $w = \cos\theta$ rather than θ. In this formulation, the phase of $f(\theta)$ is specified to be zero for all values of θ, but the source locations can be complex valued. One can rewrite Eq. (13.71) with both real and imaginary parts of source locations d_p [106]:

$$f(\theta) \approx f_{\text{est}}(\theta) = \sum_{p=1}^{N} R_p e^{j\beta d_{pr} \cos\theta} e^{-\beta d_{pi} \cos\theta} \tag{13.72}$$

where d_{pr} and d_{pi} are the real and imaginary parts of complex location d_p; however, the imaginary part has no physical basis.

Miller and Goodman [119] studied variations of desired patterns of the form $\sin^P \theta$ and concluded that it is very important to use a good numerical matrix inversion procedure to solve for d_p and R_p in Prony's method. The calculated source spacing is larger than 0.7λ and the spacings are nonuniform, which eliminated grating lobes for studied patterns. The effects of reducing the number of sources on the pattern accuracy were also investigated. The sources with lowest residues were successively eliminated and new sets of residues were computed with new poles (source locations). It is interesting that for a particular range of source locations the sidelobe level is nearly uniform. This property is included explicitly in Dolph–Chebyshev synthesis discussed in Section 13.4.2.1 [119].

Miller and Burke studied a relationship between the aperture width and number of point sources in a linear array [121]. Increasing the number of sources by increasing the aperture size produces more accurate results for the source location values. The results stabilize after the source separations exceed one-half wavelength.

Finally, the pattern generated by a continuous source distribution is used as a desired pattern in the pole–residue synthesis technique. The same case was treated by Elliott who employed a modification of Schelkunoff's synthesis procedure [122]. Table 13.4 shows the results obtained by Elliott [119, 122] and the results obtained using the pole–residue synthesis method. Elliott's technique assumed equally spaced sources with $d = 0.475\lambda$. The pole–residue synthesis produces a linear array with unequal spacings and with two fewer elements (i.e., 13 elements compared to 15 elements used by Elliott). Elliott obtained a very accurate pattern approximation within 0.25 dB of the peak sidelobe specification. A constrained solution using Prony's method satisfied the sidelobe specifications to within 1 dB. Both techniques yield the same aperture length of 6.64λ. Figure 13.8 shows that the patterns for Elliott's technique and pole–residue synthesis are in good agreement. Figure 13.9 shows the poles and residue strengths and the locations in the complex plane. In this case, all the poles are real valued due to the constrained problem of equal source spacing.

TABLE 13.4 Source Locations and Strengths for the Pole–Residue Synthesis Method and Elliott's Synthesis Method Corresponding to Array Factor Patterns in Figure 13.8

N	Location (λ)	Magnitude of R_p	Magnitude of I_n	Phase of R_p (degrees)	Phase of I_n (degrees)
−7			0.398		−39.48
−6	−3.32	0.398	0.215	−39.50	1.66
−5	−2.84	0.223	0.580	1.44	−12.31
−4	−2.35	0.626	0.510	−12.45	−9.96
−3	−1.79	0.673	0.828	−8.61	−5.49
−2	−1.25	1.020	0.778	−5.02	−4.76
−1	−0.607	1.218	1.00	−2.71	−1.72
0	0	1.18	0.881	0	0
1	0.607	1.218	1.00	2.71	1.72
2	1.25	1.020	0.778	5.02	4.76
3	1.79	0.673	0.828	8.61	5.49
4	2.35	0.626	0.510	12.45	9.96
5	2.84	0.223	0.580	−1.44	12.31
6	3.32	0.398	0.215	39.50	−1.66
7			0.398		39.48

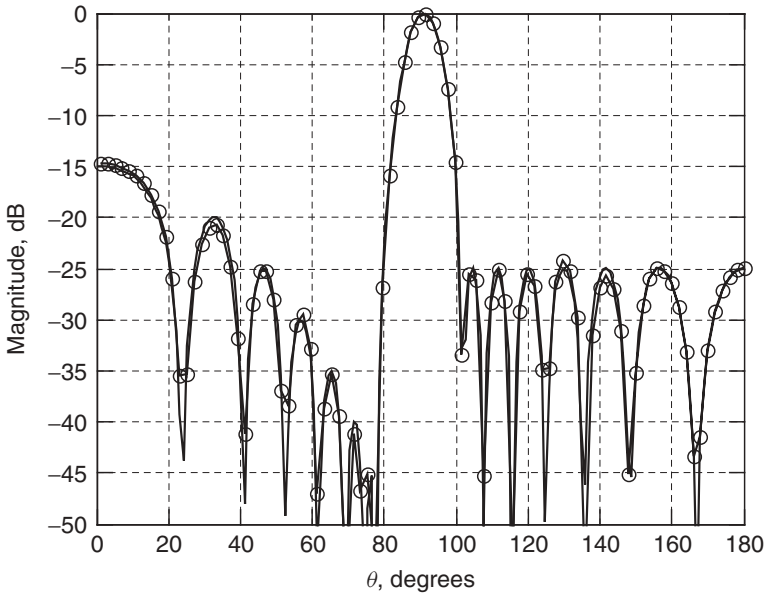


Figure 13.8 Array factor pattern computed using the pole–zero synthesis method (—**o**—) compared to Elliot's synthesis method (—).

13.5.5 Pole–Residue Antenna Modeling and Synthesis

Pole–residue antenna modeling can be approached using conventional antenna effective length. Licul and Davis [123] described a general approach for antenna modeling using a pole–residue representation via antenna realized effective length. The usual procedure is to measure the received signal of two identical antennas in the frequency domain, including the contributions of both the transmit and receive antennas, and then to perform an inverse transform to obtain the time-domain equivalent waveform. This section discusses how to obtain the response of a single antenna with realized effective length $\mathbf{h}_R(j\omega, \theta, \phi)$ in both the time and frequency domains from the frequency-domain s_{21} measurements with a vector network analyzer. The measured received waveform can be represented in the following form for a line-of-sight link [1, Sec. 9.1; 124; 125]:

$$s_{21}(j\omega, \theta, \phi) = 2\frac{j\omega\mu}{Z_0}\mathbf{h}_{Rr}(j\omega, \theta, \phi) \cdot \mathbf{h}_{Rt}(j\omega, \theta, \phi)\frac{e^{-j\beta r}}{4\pi r} \tag{13.73}$$

where $\mathbf{h}_{Rr}$ and $\vec{\mathbf{h}}_{Rt}$ represent realized effective lengths of the receive and transmit antennas, respectively. Realized effective length includes the impedance mismatch effects, which are not normally included in the effective length. If the angular dependence is not explicitly shown, Eq. (13.73) is

$$s_{21}(j\omega) = 2\frac{j\omega\mu}{Z_0}h_{Rt}(j\omega)h_{Rr}(j\omega)\frac{e^{-j\beta r}}{4\pi r} \tag{13.74}$$

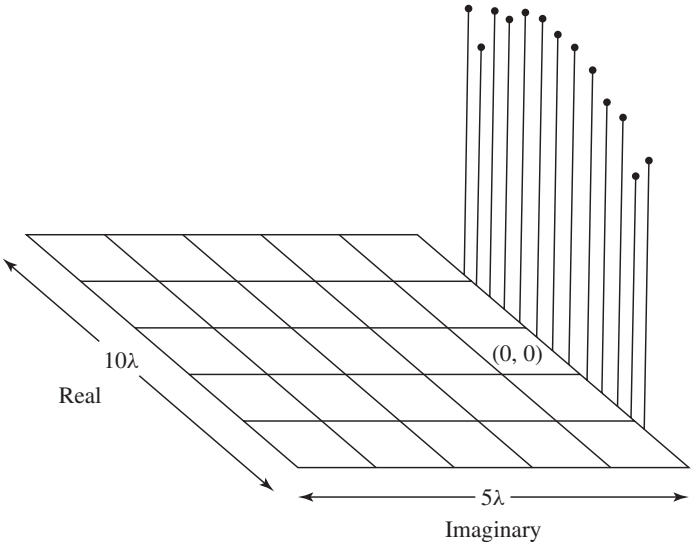


Figure 13.9 Source locations in the complex plane. (From Ref. 106.)

If the antennas are identical, the realized effective length of a single antenna is found from

$$h_R(j\omega) = \sqrt{\frac{Z_0}{j\omega\mu} 2\pi r e^{j\beta r} s_{21}(j\omega)} \tag{13.75}$$

The realized effective length of the antenna can also be obtained using the realized effective length of a known reference antenna, $h_{\text{ref}}(j\omega)$:

$$h_R(j\omega) = \frac{s_{21}(j\omega) Z_0 2\pi r e^{j\beta r}}{\mu h_{\text{ref}}(j\omega)} \tag{13.76}$$

The realized effective length includes the effects of the antenna impedance. In order to obtain the effective length one needs to use the following expression:

$$h(j\omega) = \frac{1}{(1 - s_{11}(j\omega))} h_R(j\omega) \tag{13.77}$$

Once the realized effective length is obtained in the frequency domain, one can perform an inverse fast Fourier transform (IFFT) to obtain the time-domain realized effective length of the single antenna.

For wideband antennas, the realized effective length is almost identical to the antenna effective length due to a very good impedance match across a wide range of frequencies (above the antenna fundamental resonance). For resonant structures, the realized effective length is an approximation of the antenna effective length because the impedance match can be poor at some frequencies of interest.

From the expression for the antenna radiated electric field in terms of the realized effective length

$$\mathbf{E}_{\text{rad}}(\theta, \phi, j\omega) = j\omega\mu \frac{e^{-j\beta r}}{4\pi r} I_t \mathbf{h}_R(\theta, \phi, j\omega) \tag{13.78}$$

where I_t is the transmit current. One can conclude that the antenna radiated electric field is proportional to the realized effective length and it is possible to extract the antenna frequency-domain pattern for each frequency of interest. Once all the residues for the azimuth and elevation plane are obtained, one can reconstruct the frequency-domain patterns for an arbitrary frequency as follows:

$$\mathbf{h}(\theta, \phi, j\omega) = \sum_m \frac{\mathbf{R}_m(\theta, \phi)}{-s_m + j\omega} \tag{13.79}$$

whereR_m and s_m are the complex residues and poles, respectively. The pattern is the combination of all the residues and poles. The poles used in pattern calculations are the poles obtained from the boresight direction signal. Usually, one needs to associate the poles for different directions in order to have a correct set of poles to represent the antenna system. This association is required due to experimental error. As an alternative, the poles obtained from the boresight direction can be used for all directions, with care taken not to miss some poles that may not be represented adequately at the boresight. The time-domain antenna pattern can be obtained from the following expression using the same set of complex residues and poles:

$$\mathbf{h}_R(\theta, \phi, t) = \sum_m \mathbf{R}_m(\theta, \phi) e^{s_m t} \tag{13.80}$$

The poles obtained from the boresight measurement using a Vivaldi antenna provided a good set. The authors obtained a 17-pole model of a Vivaldi antenna that models both its time- and frequency-domain characteristics. Figure 13.10 shows good agreement between the modeled and measured Vivaldi antenna patterns in the frequency range of 1–12 GHz using Eq. (13.79).

Because the model parameters (poles and residues) are directly related to antenna geometrical features, one can consider an antenna synthesis approach. The poles are independent of the antenna orientation while the residues are orientation dependent. Figure 13.11 illustrates the modeling and synthesis approach via realized effective length.

The wideband planar sectoral monopole shown in Figure 13.12 was investigated using the matrix-pencil method. Licul et al. [127] varied a single antenna geometry parameter as a function of realized effective length. Realized effective length was obtained using Eq. (13.75). The sector monopole antenna resonates at 5.1 GHz for $\alpha = 1^\circ$. Figure 13.13 shows the pole variation as a function of the antenna geometry parameter α. The antenna has three distinct pole–residue groups. The arrow indicates the pole movement for the values of α increasing from 1° to 40°. The first pole group, h_1, is directly related to the overall antenna size. This group represents antenna fundamental resonance. By increasing α the poles move further away from the $j\omega$-axis, indicating a broadband structure. Tesche [111] observed the same behavior by increasing the diameter of a wire. The second pole group, h_2, is related to an antenna slot radiating mechanism, the interaction between the ground plane, and the antenna outer edges. The third pole group, h_3, is directly related to

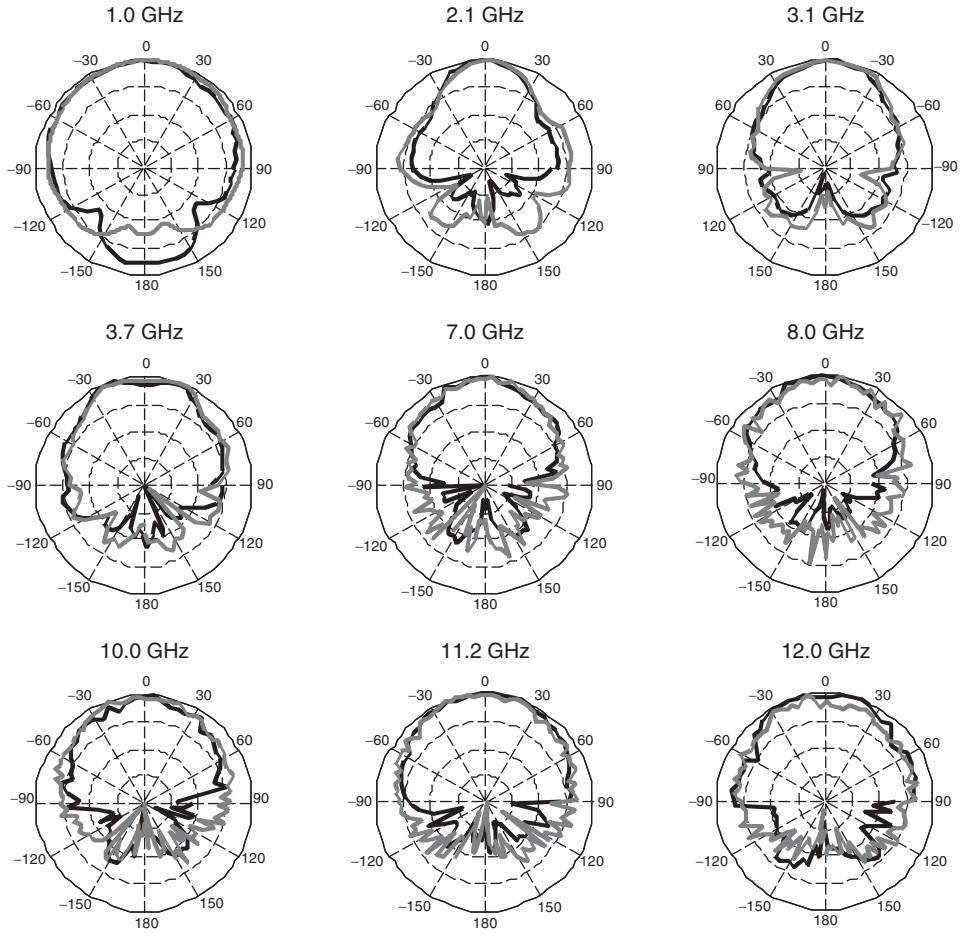


Figure 13.10 Patterns for the 17-pole model (black) of a Vivaldi antenna using Eq. (13.79) compared to the near-field measurements (gray) −10 dB/div. (From Ref. 126, with permission.)

antenna fundamental resonance. It represents the antenna third fundamental resonance. The antenna complete response, h_R, is a summation of all three pole groups.

This planar sectoral monopole antenna can be modeled with three parameters (six pole–residue parameters since they appear in complex conjugate pairs) for a particular α value in the range from 1° to 40°. The poles have distinct trajectories for a particular radiation mode. Conceivably, it should be possible to specify a radiation pattern, bounded to the sectoral monopole antenna, and obtain parameter α that best approximates the desired radiation pattern. This is a topic of a future research in pole–residue antenna modeling and synthesis.

13.5.6 Conclusions

SEM provides a reduced parameter description of the antenna radiation characteristics using only poles and residues. The poles are independent of the antenna orientation while the residues depend on the antenna orientation. The matrix-pencil method or

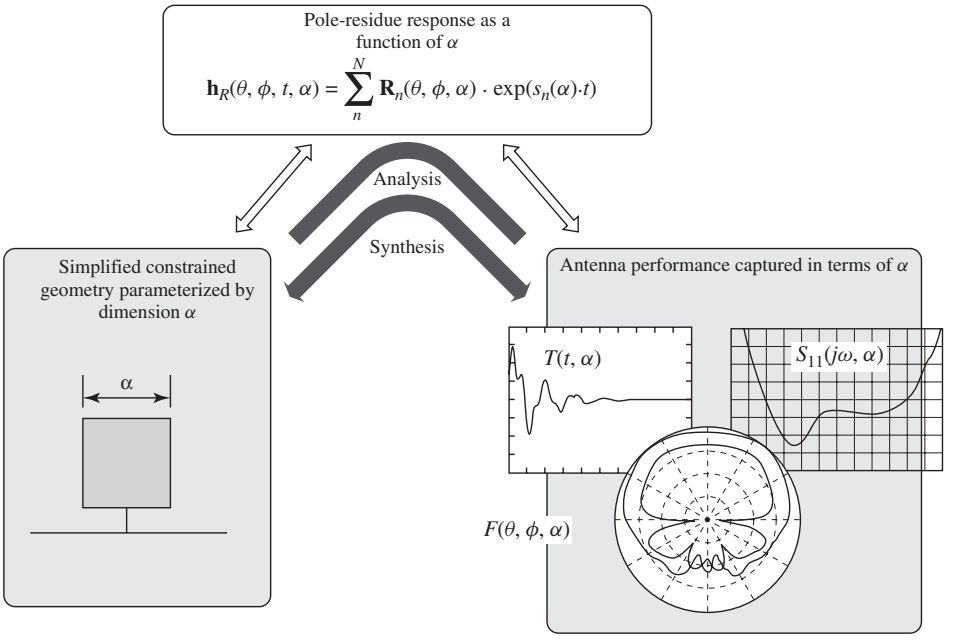


Figure 13.11 Synthesis and analysis approach using pole–residue modeling of the antenna effective length. (From Ref. 127.)

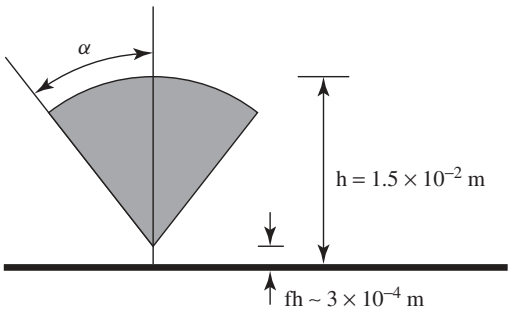


Figure 13.12 Planar sector monopole antenna dimensions. For $\alpha = 1^\circ$ the antenna resonates at 5.1 GHz. (From Ref. 127.)

Prony's method can be used to obtain a pole–residue representation of the antenna. The matrix-pencil method is a more elegant process for extracting poles and residues from a particular response because it is a one-step process. Prony's method is a two-step process, requiring calculation of polynomial roots to compute the poles.

Pole–residue linear array synthesis provides both source locations and strengths for a specified desired pattern and number of sources. Pole–residue antenna modeling provides a simplified representation that, via antenna realized effective length, can provide both time- and frequency-domain antenna radiation characteristics. The sector monopole antenna example of Figure 13.12 and 13.13 showed that the pole–residue model provides a framework for future work on antenna synthesis.

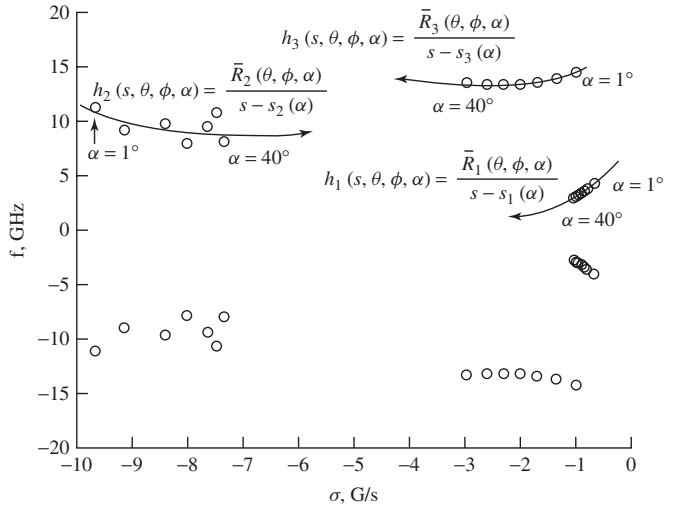


Figure 13.13 Poles of the planar sector monopole antenna of Figure 13.12. (From Ref. 127.)

REFERENCES

1. W. L. Stutzman and G. A. Thiele, *Antenna Theory and Design*, 2nd ed., John Wiley & Sons, Hoboken, NJ, 1998.
2. O. M. Bucci, G. D'Ella, G. Mazzarella, and G. Panariello, Antenna pattern synthesis: a new general approach, *Proc. IEEE*, Vol. 82, pp. 358–371, March 1994.
3. C. A. Balanis, *Antenna Theory: Analysis and Design*, 3 ed., John Wiley & Sons, Hoboken, NJ, 2005.
4. R. S. Elliott, *Antenna Theory and Design*, Prentice-Hall, Engelwood Cliffs, NJ, 1981; rev. ed., IEEE Press, Piscataway, NJ, 2003.
5. R. J. Mailloux, *Phased Array Antenna Handbook*, Artech House, Norwood, MA, 1994.
6. J. N. Sahalos, *Orthogonal Methods for Array Synthesis: Theory and the ORAMA Computer Tool*, John Wiley & Sons, Hoboken, NJ, 2006.
7. R. C. Hansen, Array pattern control and synthesis, *Proc IEEE*, Vol. 80, pp. 141–151, January. 1992.
8. P. M. Woodward, A method for calculating the field over a plane aperture required to produce a given polar diagram, *J. IEE*, Vol. 93, P. IIIA, pp. 1554–1558, 1946.
9. P. M. Woodward and J. D. Lawson, The theoretical precision with which an arbitrary radiation pattern may be obtained from a source of finite extent, *J. IEE*, Vol. 95,P. III, pp. 363–1370, September 1948.
10. D. R. Rhodes, A general theory of sampling synthesis, *IEEE Trans. Antennas Propag.*, Vol. 21, pp. 176–181, March 1973.
11. W. L. Stutzman, Synthesis of shaped-beam radiation patterns using the iterative sampling method, *IEEE Trans. Antennas Propag.*, Vol. 19, pp. 36–41, January 1971. Note: All the calculations in this chapter were made using the frequency-selective screen software first developed in this reference and further developed and validated over the years.
12. R. E. Collin and F. J. Zucker (Ed.), *Antenna Theory, Part 1* , McGraw-Hill, New York, 1969.
13. W. L. Stutzman, Sidelobe control of antenna patterns, *IEEE Trans. Antennas Propag.*, Vol. 20, pp. 102–104, January 1972.

14. W. L. Stutzman and E. L. Coffey, Radiation pattern synthesis of planar antennas using the iterative sampling method, *IEEE Trans. Antennas Propag.*, Vol. 23, pp. 764–769, November 1975.
15. H. P. Raabe, Antenna pattern synthesis of the most truthful approximation, *Wescon*, Vol. 2, pp. 178–183, August 1958.
16. N. Yaru, A note on super-gain antenna arrays, *Proc. IRE*, Vol. 39, pp. 1081–1085, September 1951.
17. D. R. Rhodes, The optimum line source for the best mean-square approximation to a given radiation pattern, *IEEE Trans. Antennas Propag.*, Vol. 11, pp. 440–446, July 1963.
18. T. Fong, On the problem of optimum antenna aperture distribution, *J. Franklin Inst.*, Vol. 283, pp. 235–249. March 1967.
19. W. Stutzman, Evaluation of the mean-squared error for optimum line source patterns, *IEEE Trans. Antennas Propag.*, Vol. 18, pp. 808–809, November 1970.
20. T. S. Fong and R. A. Birgenheier, Method of conjugate gradients for antenna pattern synthesis, *Radio Sci.*, Vol. 6, pp. 1123–1130, December 1971.
21. L. Godara (Ed.), *Handbook of Antennas in Wireless Communications*, CRC Press, Boca Raton, FL, 2002.
22. Y. Rahmat-Samii and E. Michielssen, (Eds.) *Electromagnetic Optimization by Genetic Algorithm*, John Wiley & Sons, Hoboken, NJ, 1999.
23. R. F. Hyneman and R. M. Johnson, A technique for the synthesis of shaped-beam radiation patterns with approximately equal-percentage ripple, *IEEE Trans. Antennas Propag.*, Vol. 15, pp. 737–743, November 1967.
24. C. H. Walter, *Traveling Wave Antennas*, McGraw-Hill, New York, 1965; available from SciTech.
25. S. Silver, *Microwave Antenna Theory and Design*, McGraw-Hill, New York, 1949; available from IET, www.theiet.org.
26. A. S. Dunbar, On the theory of antenna beam shaping, *J. Appl. Phys.*, pp. 847–853, August 1952.
27. R. A. Kallas, Synthetic apertures studied for satellite radar, *Microwave Systems News*, Vol. 8, pp. 56–67, September 1978.
28. H. Zhou, P. J. B. Clarricoats, G. T. Poulton, and S. G. Hay, Power pattern synthesis for mesh reflectors using the method of successive projections, *Electro. Lett.*, Vol. 28, pp. 1837–1839, 10 Sept 1992.
29. R. Guy, General radiation-pattern synthesis technique for array antennas of arbitrary configuration and element type, *IEE Proc.*, Vol. 135, Pt. H, pp. 241–248, August 1988.
30. L. Wu, A. Zielinski, and J. Bird, Synthesis of shaped radiation patterns using an iterative method, *Radio Sci.*, Vol. 30, pp. 1385–1394, September–October 1995.
31. D. F. Kelley and W. L. Stutzman, Array pattern modeling methods that include mutual coupling effects, *IEEE Trans. Antennas Propag.*, Vol. 41, pp. 1625–1632, December 1993.
32. I. Wolff, Determination of the radiating system which will produce a specified directional characteristic, *Proc. IRE*, Vol. 25, pp. 630–643, May 1937.
33. J. C. Simon, Application of periodic function approximation to antenna pattern synthesis and circuit theory, *IRE Trans. Antennas Propag.*, Vol. 4, pp. 429–440, July 1956.
34. P. A. Bricout, Pattern synthesis using weighted functions, *IRE Trans. Antennas Propag.*, Vol. 8, pp. 441–444, July 1960.
35. W. G. Jaeckle, Antenna synthesis by weighted Fourier coefficients, *IEEE Trans. Antennas Propag.*, Vol. 12, pp. 369–370, May 1964.
36. T. Chiba, On a pattern synthesis method for a linear array, *IEEE Int. Conv. Rec.*, Vol. 14, Pt. 5, pp. 172–179, March 1966.

37. G. A. Somers, A proof of the Woodward– Lawson sampling method for a finite linear array, *Radio Sci*, Vol. 28, pp. 481–485, July-August 1993.

38. H. E. Shanks, A geometrical optics method of pattern synthesis for linear arrays, *IRE Trans. Antennas Propag.*, Vol. 8, pp. 485–490, September 1960.

39. M. J. Maybell, The Woodward– Lawson method: a third opinion, *IEEE Antennas Propag. Soc. Newslett.*, pp. 49–50, June 1989.

40. D. H. Werner and A. J. Ferraro, Cosine pattern synthesis for single and multiple main beam uniformly spaced linear arrays, *IEEE Trans. Antennas Propag.*, Vol. 37, pp. 1480–1484, November 1989.

41. S. A. Schelkunoff, A mathematical theory of linear arrays, *Bell System Tech. J.,* Vol. 22, pp. 80–107, 1943.

42. T. T. Taylor and R. Whinnery, Application of potential theory to the design of linear arrays, *J. Appl. Phys*, Vol. 22, pp. 19–29, January 1951.

43. R. S. Elliott and G. J. Stern, A new technique for shaped beam synthesis of equispaced arrays, *IEEE Trans. Antennas Propag.*, Vol. 32, pp. 1129–1133, October 1984.

44. H. J. Orchard, R. S. Elliott, and G. J. Stern, Optimizing the synthesis of shaped beam antenna patterns, *IEE Proc.*, Vol. 132, Pt. H, pp. 63–68, Febuary 1985.

45. J. Fondevila-Gomez, et al., Synthesizing footprints for very large antenna arrays: Orchard-Elliot-Stern and Woodward-Lawson working hand-in-hand, *IEEE Trans. Antennas Propag. Society Int. Symp. Digest* (Albuquerque) pp. 4209–4212, July 2006.

46. M. Hoffman, The utility of the array pattern matrix for linear array computations, *IRE Trans. Antennas. Propag.*, Vol. 9, pp. 97–100, January 1961.

47. B. Ng, M. H. Er, and C. Kot, A flexible array synthesis method using quadratic programming, *IEEE Trans. Antennas Propag.*, Vol. 41, pp. 1541–1550, November 1993.

48. J. M. Johnson and Y. Rahmat-Samii, Genetic algorithms in engineering electromagnetics, *IEEE Antennas Propag. Magazine*, Vol. 39, pp. 7–21, August 1997.

49. R. J. Mitchell, B. Chambers, and A. P. Anderson, Array pattern synthesis in the complex plane optimized by a genetic algorithm, *Electron. Lett.*, Vol. 32, pp. 1843–1845, September 1996.

50. S. Blank, An algorithm for the empirical optimization of antenna arrays, *IEEE Trans. Antennas Propag.*, Vol. 31, pp. 685–687, July 1983.

51. R. A. Michelson and J. W. Schomer, A three parameter antenna pattern synthesis technique, *Microwave Journal*, Vol. 8, pp. 88–94, September 1965.

52. M. J. Buckley, Synthesis of shaped beam antenna patterns using implicitly constrained current elements, *IEEE Trans. Antennas Propag.,* Vol. 44, pp. 192–197, February 1996.

53. C. Christodoulou and M. Georgiopoulos, *Applications of Neural Networks in Electromagnetics*, Artech House, Norwood, MA, 2001.

54. M. M. Khodier and C. G. Christodoulou, Linear array geometry synthesis with minimum sidelobe level and null control using particle swarm optimization, *IEEE Trans. Antennas Propag.*, Vol. 53, pp. 2674–2679, August 2005.

55. D. W. Boeringer and D. H. Werner, Particle swarm optimization versus genetic algorithms for phased array synthesis, *IEEE Trans. Antennas Propag.*, Vol. 52, pp. 771–778, March 2004.

56. A. Ksienski, Equivalence between continuous and discrete radiating arrays, *Can. J. Phys*, Vol. 39, pp. 335–349, February 1961.

57. A. Ishimaru, Theory of unequally spaced arrays, *IRE Trans. Antennas Propag.*, Vol. 10, pp. 691–702, July 1962.

58. A. Ishimaru and J. H. Lahti, Unequally spaced arrays fed from traveling wave sources, *Wescon*, Vol. 7, Pt. 1, August 1963.

59. W. L. Stutzman, Shaped-beam synthesis of nonuniformly spaced linear arrays, *IEEE Trans. Antennas Propag.*, Vol. 20, pp. 449–501, July 1972.
60. S. J. Blank and M. F. Hutt, On the empirical optimization of antenna arrays, *IEEE Antennas Propag. Mag.*, Vol. 47, pp. 58–67, April 2005.
61. B. J. Strait and K. Hirasawa, Array design for a specified pattern by matrix methods, *IEEE Trans. Antennas Propag.*, Vol. 17, pp. 237–239, March 1969.
62. H. Steyskal and J. S. Herd, Mutual coupling compensation in small array antennas, *IEEE Trans. Antennas Propag.*, Vol. 38, pp. 1971–1975, December 1990.
63. P. J. B. Clarricoats, S. M. Tun, and R. C. Brown, Performance of offset reflector antennas with array feeds, *IEE Proc.*, Vol. 131, Pt. H, pp. 172–178, June 1984.
64. W. T. Smith and W. L. Stutzman, A pattern synthesis technique for array feeds to improve radiation pattern performance of large distorted reflector antennas, *IEEE Trans. Antennas Propag.*, Vol. 40, pp. 57–62, January 1992.
65. W. L. Stutzman and C. Buxton, Radiating elements for wideband phased arrays, *Microwave J*, Vol. 43, pp. 130–141, February 2000.
66. M. Uthansakul and M. E. Bialkowski, Fully spatial wide-band beamforming using a rectangular array of planar monopoles, *IEEE Trans. Antennas Propag.*, Vol. 54, pp. 527–532, February 2006
67. A. Ishimaru and Y.-S. Chen, Thinning and broadbanding antenna arrays by unequal spacings, *IEEE Trans. Antennas Propag.*, Vol. 13, pp. 34–42, January 1965.
68. D. D. King, R. F. Packard, and R. K. Thomas, Unequally-spaced, broad-band antenna arrays, *IRE Trans. Antennas Propag.*, Vol. 8, pp. 380–384, January 1960.
69. D. G. Shively and W. L. Stutzman, Wideband arrays with variable element sizes, *IEE Proc. Microwaves Antennas and Propag.*, Vol. 137, pp. 238–240, August 1980.
70. C. I. Coman, I. E. Lager, and L. P. Ligthart, The design of shared aperture antennas consisting of differently sized elements, *IEEE Trans. Antennas Propag.*, Vol. 54, pp. 376–383, February 2006.
71. C. V. Baker, A technique for the analytical synthesis of shaped beams for arrays, *IEEE Trans. Antennas Propag.*, Vol. 16, pp. 803–805, November 1969.
72. Y. T. Lo, S. W. Lee, and Q. H. Lee, Optimization of directivity and signal-to-noise ratio of an arbitrary antenna array, *Proc. IEEE*, Vol. 54, pp. 1033–1045, August 1966.
73. J. K. Butler and H. Unz, Beam efficiency and gain optimization of antenna arrays with nonuniform spacings, *Radio Sci.*, Vol. 2, pp. 711–720, July 1967.
74. K. Milne, Synthesis of power radiation patterns for linear arrays, *IEE Proc.*, Vol. 134, Pt. H, pp. 285–296, June 1987.
75. A. Ksienski, Synthesis of nonseparable two-dimensional patterns by means of planar arrays, *IRE Trans. Antennas Propag.*, Vol. 8, pp. 224–225, July 1960.
76. R. Elliott and G. Stern, Shaped patterns from a continuous planar aperture distribution, *IEE Proc.*, Vol. 135, Pt. H, pp. 366–370, December 1988.
77. J. Richie and H. Kritikos, Linear program synthesis for direct broadcast satellite phased arrays, *IEEE Trans. Antennas Propag.,* Vol. 36, pp. 345–348, March 1998.
78. R. E. Hodges and Y. Rahmat-Samii, On sampling continuous aperture distributions for discrete planar arrays, *IEEE Trans. Antennas. Propag.*, Vol. 44, pp. 1499–1508, November 1996.
79. L. Josefsson and P. Persson, Conformal array pattern synthesis, *Conformal Array Antenna Theory And Design*, Wiley, Hoboken, NJ, 2006, Chapter 10.
80. T. T. Taylor, Design of line-source antennas for narrow beamwidth and low side lobes, *IRE Trans. Antennas Propag.*, Vol. 3, pp. 16–28, January 1955.

81. R. C. Hansen Ed., *Microwave Scanning Antennas; Volume I: Apertures*, Academic Press, New York, 1964.
82. C. L. Dolph, A current distribution for broadside arrays which optimizes the relationship between beam width and side-lobe level, *Proc. IRE*, Vol. 34, pp. 335–348, June 1946.
83. A. T. Villeneuve, Taylor patterns for discrete arrays, *IEEE Trans. Antennas Propag.*, Vol. 32, pp. 1089–1093, October 1984.
84. R. J. Stegen, Excitation coefficients and beamwidths of Tschebysheff arrays, *Proc. IRE*, Vol. 41, pp. 1671–1674, November 1953.
85. A. Saffaai-Jazi, A new formulation for the design of Chebyshev arrays, *IEEE Trans. Antennas Propag.*, Vol. 42, pp. 439–443, March 1994
86. A. D. Bresler, A new algorithm for calculating the current distributions of Dolph–Chebyshev arrays, *IEEE Trans. Antennas Propag.*, Vol. 28, pp. 951–952, November 1980.
87. C. van der Mass, A simplified calculation for Dolph-Tschebyscheff arrays, *J. Applied Physics*, Vol. 25, pp. 121–124, Jan. 1954.
88. F. Ares and E. Moreno, New method for computing Dolph-chebyshev arrays, and its comparison with other methods, *IEE Proc.,* Vol. 135, Pt. H, pp. 129–131, April 1988.
89. R. S. Elliott, Beamwidth and directivity of large scanning arrays, *Microwave J.*, Part 1 Vol. 6 pp. 53–60, December 1963.
90. A. Saffaai-Jazi, Directivity of Chebyshev arrays with arbitrary element spacing, *Electronics Letters*, Vol. 31, pp. 772–774, May 1995.
91. R. C. Hansen, Dolph-Chebyshev array directivity versus spacing, *Electronics Letters*, Vol. 32, pp.1050–1056, 'June 1996.
92. R. A. Mucci, D. W. Tufts, and J. T. Lewis, Beam pattern synthesis for line arrays subject to upper and lower constraining bounds, *IEEE Trans. Antennas Propag.*, Vol. 23, pp. 732–734, September 1975.
93. R. F. Hyneman, A technique for the synthesis of line source antenna patterns having specified sidelobe behavior, *IEEE Trans. Antennas Propag.*, Vol. 16, pp. 430–435, July 1968.
94. R. S. Elliott, Design of line-source antennas for sum patterns with sidelobes of individually arbitrary heights, *IEEE Trans. Antennas Propag.*, Vol. 24, pp. 76–83, January 1976.
95. M. T. Ma, *Theory and Applications of Antenna Arrays*, John Wiley & Sons, Hoboken, NJ, 1974
96. Y. L. Chow, On grating plateaux of nonuniformly spaced arrays, *IEEE Trans. Antennas Propag.*, Vol. 13, pp. 208–215, March 1965.
97. H. Unz, Linear arrays with arbitrarily distributed elements, *IRE Trans. Antennas Propag.*, Vol. 8, pp. 222–223, March 1960.
98. R. E. Willey, Space tapering of linear and planar arrays, *IRE Trans. Antennas Propag.*, Vol. 10, pp. 369–377, July 1962.
99. C. A. Olen and R. T. Compton, A numerical pattern synthesis algorithm for arrays, *IEEE Trans. Antennas Propag.*, Vol. 38, pp. 1666–1676, October 1990.
100. Y. T. Lo and S. W. Lee, A study of space-tapered arrays, *IEEE Trans. Antennas Propag.*, Vol. 14, pp. 22–30, January 1966.
101. Y. T. Lo, A mathematical theory of antenna arrays with randomly spaced elements, *IEEE Trans. Antennas Propag.*, Vol. 12, pp. 257–268, May 1964.
102. R. J. Mailloux and E. Cohen, Statistically thinned arrays with quantized element weights, *IEEE Trans. Antennas Propag.*, Vol. 39, pp. 436–447, April 1991.
103. K. R. Carver, W. K. Cooper, and W. L. Stutzman, Beam-pointing errors of planar-phased arrays, *IEEE Trans. Antennas Propag.*, Vol. 21, pp. 199–202, March 1973.
104. S. W. Autrey, Approximate synthesis of nonseparable design responses for rectangular arrays, *IEEE Trans. Antennas Propag.*, Vol. 35, pp. 907–912, August 1987.

105. T. T. Taylor, Design of circular apertures for narrow beamwidth and low sidelobes, *IRE Trans. Antennas Propag.*, Vol. 8, pp. 17–22, January 1960.

106. R. C. Hansen, *Microwave Scanning Antennas; Volume II: Array Theory and Practice*, Academic Press, New York, 1996.

107. R. C. Hansen, Tables of Taylor distributions for circular aperture antennas, *IRE Trans. Antennas. Propag.*, Vol. 8, pp. 23–26, January 1960

108. R. C. Hansen, A one-parameter circular aperture distribution with narrow beamwidth and low side lobes, *IEEE Trans. Antennas. Propag.*, Vol. 24, pp. 477–480, July 1976.

109. C. E. Baum, The singularity expansion method, in *Transient Electromagnetic Fields*, Springer-Verlag, Berlin, pp. 129–179, 1976.

110. C. E. Baum, On the singularity expansion method for the solution of electromagnetic interaction problems, *Interaction Note 88*, December 1971.

111. F. M. Tesche, On the singularity expansion method as applied to electromagnetic scattering from thin-wires, *AFWL Interaction Note 102*, April 1972.

112. F. M. Tesche, On the analysis of scattering and antenna problems using the singularity expansion technique, *IEEE Trans. Antennas Propag.*, Vol. 21, pp. 53–62, January 1973.

113. J. T. Cordaro and W. A. Davis, Time-domain techniques in the singularity expansion method, *IEEE Trans. Antennas Propag.*, Vol. 29, pp. 534–538, May 1981.

114. M. L. Van Blaricum and R. Mittra, Problems and solutions associated with Prony's method for processing transient data, *IEEE Trans. Antennas Propag.*, Vol. 26, pp. 174–182, January 1978.

115. Y. Hua and T. K. Sarkar, Generalized pencil-of-function method for extracting poles of an EM system from its transient response, *IEEE Trans. Antennas Propag.*, Vol. 37, No. 2, pp. 229–234, February 1989.

116. T. K. Sarkar and O. Pereira, Using the matrix pencil method to estimate the parameters of a sum of complex exponentials, *IEEE Antennas Propag. Mag.*, Vol. 37, pp. 48–55, February 1995.

117. R. Prony, Essai experimental et analytique sur les lois de la dilatabilite de fluids elastiques et sur celles del la force expansive da la vapeur de l'alkool, a differentes temperatures, *Paris J. l'Ecole Poltechnique*, Vol. 1, No. 2, pp. 24–76, 1795.

118. R. Mittra, Integral equation methods for transient scattering, in *Transient Electromagnetic Fields*, L. Felsen (Ed.), Springer-Verlag, Berlin pp. 73–128, 1976.

119. E. K. Miller and D. M. Goodman, A pole– zero modeling approach to linear array synthesis: 1. The unconstrained solution, *Radio Sci.*, Vol. 18, No. 1, pp. 57–69, January-Febuary 1983.

120. E. K. Miller and D. L. Lager, Radiation field analysis and synthesis using Prony's method, *Electron. Lett.*, Vol. 14, No. 6, pp. 180–182, 16 March 1978.

121. E. K. Miller and G. J. Burke, Using model-based parameter estimation to increase the physical interpretability and numerical efficiency of computational electromagnetics, *Comput. Phy. Commun.*, Vol. 68, No. 1–3 pp. 43–75, 1991.

122. R. S. Elliott, On discretizing the continuous aperture distribution, *IEEE Trans. Antennas Propag.*, Vol. 25, pp. 617–621, September 1977.

123. S. Licul and W. A. Davis, Unified time and frequency antenna modeling and characterization, *IEEE Trans. Antennas Propag.*, Vol. 53, pp. 2882–2888, September 2005.

124. W. A. Davis, A reciprocity development of the relationship between antenna gain and antenna aperture, personal notes, 2002.

125. W. A. Davis and S. Licul, Ultra-wideband antennas, in *Introduction to Ultra-Wideband Communications,* J. H. Reed (Ed.), Prentice Hall, Englewood Cliffs, NJ, 2004.

126. S. Licul, Ultra-wideband antenna characterization and measurements, Ph.D. dissertation, Virginia Tech, Blacksburg, VA, 2004.

127. S. Licul, N. P. Cummings, W. A. Davis, and W. L. Stutzman, On a single element synthesis of planar monopoles, URSI, Albuquerque, June 2006.

128. Y. Hua and T. K. Sarkar, Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise, *IEEE Trans. Acoustics Speech Signal Processing*, Vol. 38, pp. 814–824, May 1990.

PART IV

STRUCTURES AND TECHNIQUES RELATED TO ANTENNAS

CHAPTER 14

Antenna Applications of Negative Refractive Index Transmission-Line (NRI-TL) Metamaterials

GEORGE V. ELEFTHERIADES and MARCO A. ANTONIADES

14.1 INTRODUCTION

In the 1960s Victor Veselago systematically examined hypothetical materials having simultaneously negative permittivity and negative permeability [1]. He described a number of unusual electromagnetic phenomena associated with such media. A characteristic property of these materials is that plane waves propagating in them would have their phase velocity antiparallel to the group velocity; hence these media would support backward waves. Likewise, the vectors describing the electric field, the magnetic field, and the propagation direction would follow the left-handed rule. For this latter reason, Veselago coined the term "left-handed" to describe these hypothetical media. Moreover, he associated this kind of "left-handedness" with the notion of negative refraction and he described several unusual focusing devices (e.g., lenses) based on negative refraction. Nevertheless, it should be pointed out that the notion of negative refraction at the interface between a normal dielectric and a dielectric supporting backward waves had been described as early as the beginning of the 20th century [2, 3]. It was only recently though that people devised methods of implementing these left-handed or negative refractive index (NRI) media. The first such implementation was produced at the University of California in San Diego and comprised a volumetric periodic array of straight metallic wires and split-ring resonators to synthesize negative effective permittivity and negative effective permeability, respectively, at microwave frequencies [4, 5].

Another way to implement artificial materials (metamaterials) that support the phenomenon of negative refraction was subsequently proposed based on the concept of loading planar transmission-line grids with reactive elements (see subsequent sections). This transmission-line (TL) approach does not rely on loosely coupled resonators to synthesize the negative permeability; rather, it depends on the electrical connections between the constituent NRI-TL unit cells to create tightly coupled resonators, thus leading to large bandwidths over which the refractive index remains negative. The explicit formulation of

Modern Antenna Handbook. Edited by Constantine A. Balanis

the coupled resonator approach and its relation to NRI-TLs can be found in Ref. 6. In this chapter we present a number of antenna-related applications that have been developed in the leading author's research group, based on the concept of NRI-TL metamaterials. This includes nonradiating phase-shifting lines that can produce either a positive or a negative phase shift while exhibiting a broadband linear phase response, compact and broadband series power dividers and associated series-fed printed dipole arrays with reduced beam squinting, backward and forward leaky-wave antennas, and electrically small ring antennas featuring vertical polarization and good radiation efficiency. Moreover, broadband baluns and antenna signal-monitoring couplers (reflectometers) are also described.

14.2 PLANAR NEGATIVE REFRACTIVE INDEX TRANSMISSION-LINE (NRI-TL) MEDIA

A two-dimensional (2D) negative refractive index (NRI) metamaterial (MTM) can be physically implemented by reactively loading a host transmission-line (TL) grid. A representative unit cell of such a periodic NRI-TL medium is depicted in Figure 14.1.

Specifically, a 2D host transmission-line medium (e.g., microstrip) is periodically loaded with discrete series capacitors and shunt inductors in a configuration that is dual (high pass) to that of a conventional transmission line (low pass) [6–8]. From the onset, the key observation is that there is a correspondence between negative permittivity and a shunt loading inductor (L_0), as well as between negative permeability and a series loading capacitor (C_0) [9, 10]. This allows the synthesis of artificial media (metamaterials) with a negative permittivity and a negative permeability, and hence a negative refractive index. When the unit cell dimension d (see Figure 14.1) is much smaller than a guided wavelength, the array can be regarded as a homogeneous effective medium, and as such can be described by effective constitutive parameters $\mu_N(\omega)$ and $\varepsilon_N(\omega)$, which are determined through a rigorous periodic analysis to be of the form shown in Eq. (14.1) (assuming 2D TM_y wave propagation in Figure 14.1):

$$\varepsilon_N(\omega) = 2\varepsilon_p - \frac{g}{\omega^2 L_0 d}, \qquad \mu_N(\omega) = \mu_p - \frac{1/g}{\omega^2 C_0 d} \tag{14.1}$$

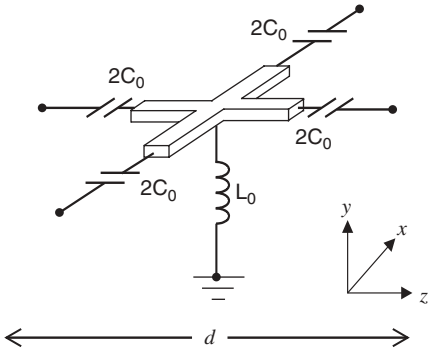


Figure 14.1 Unit cell for the 2D NRI-TL metamaterial. A host transmission line is loaded periodically with series capacitors and shunt inductors in a dual (high pass) configuration.

Here, ε_p and μ_p are positive constants describing the host transmission-line medium and they are proportional to the per-unit-length capacitance and inductance of this host transmission-line medium, respectively. On the other hand, the geometrical factor g relates the characteristic impedance of the transmission-line network to the wave impedance of the effective medium [11]. Moreover, the factor of 2 in front of the effective permittivity of the 2D medium is necessary to properly account for scattering at the edges of the unit cell (this factor becomes 1 for 1D media). These TL media support backward waves in which the phase and group velocities are antiparallel; hence, they implement the "left-handed" or NRI media envisioned by Veselago as was pointed out in Refs. 6–10. The relationship between 1D backward-wave lines (but in the ideal case without the contribution of the "right-handed" host medium) and "left-handed" lines was also pointed out in Ref. 12. Moreover, in Ref. 13 a NRI medium realization has been proposed based on a shielded stripline approach. A salient observation from inspecting Eq. (14.1) is that both the effective permittivity and permeability follow a Drude type of dispersion response, which naturally leads to wideband operation and low loss. Naturally, due to the host microstrip medium, the practically realizable unit cell of Figure 14.1 contains both a positive as well as a negative refractive-index response, as implied by Eq. (14.1) [8–10, 14]. For clarity, it should be pointed out that any periodic physical implementation of a "left-handed" metamaterial would inevitably contain a "right-handed" component. In other words, this is not unique to NRI-TL media. For example, in the split-ring-resonator/wire media the "right-handed" part is represented by the factor of 1 in front of the effective permeability and permittivity expressions (1) and (2) of Ref. 4 and corresponds to the wave propagation in the free space between adjacent unit cells.

The particular arrangement of the inclusions L_0 and C_0 in the NRI-TL media of Figure 14.1 provides the desired negative material contribution that diminishes with frequency ω and ensures compatibility with the Poynting theorem for dispersive media [1]. When the parameters are simultaneously negative, these structures exhibit a negative effective refractive index and have been utilized to experimentally demonstrate the predicted associated phenomena, including negative refraction and focusing [7, 8, 11, 15] as well as imaging beyond the diffraction limit [16, 17].

In practical realizations, the subwavelength unit cell of Figure 14.1 is repeated to synthesize artificial 2D materials with overall dimensions that are larger than the incident electromagnetic wavelength. Therefore the resulting structures are by definition distributed. However, the loading lumped elements could be realized either in chip [6, 8] or in printed form [9, 10, 14, 15, 18, 19].

It is worth mentioning that these transmission-line metamaterials have been generalized to three-dimensional (3D) isotropic media that can couple efficiently plane waves incident from free space [20, 21].

14.3 NRI-TL METAMATERIAL PHASE-SHIFTING LINES

In conventional positive refractive index (PRI) transmission lines, the phase lags in the direction of positive group velocity, thus incurring a negative phase. It therefore follows that phase compensation can be achieved at a given frequency by cascading a section of a NRI line (e.g., backward-wave line) with a section of a PRI line to synthesize positive, negative, or zero transmission phase over a short physical length (see Figure 14.2) [22].

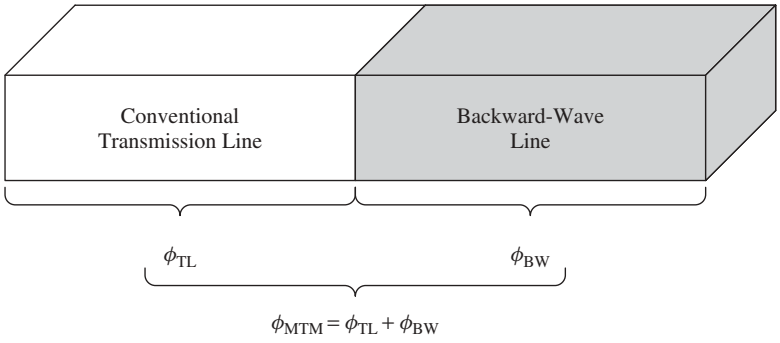


Figure 14.2 Method of phase compensation using a conventional TL and a NRI (backward-wave) line.

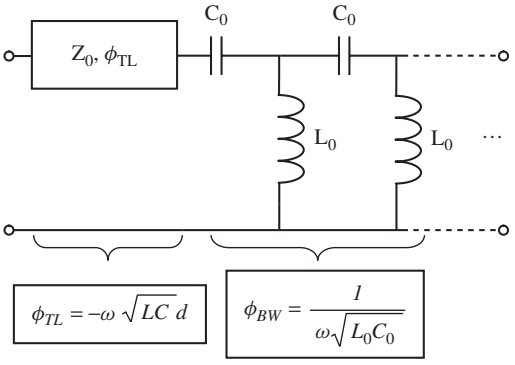


Figure 14.3 Phase compensating structure based on a conventional TL and a NRI (backward-wave) line.

This idea of phase compensation is inherent in Veselago's flat lens and was also proposed for implementing thin subwavelength resonators [23].

A physical implementation of this concept using transmission lines is shown in Figure 14.3. The structure of Figure 14.3 can be rearranged to form a series of symmetric metamaterial unit cells as proposed in Refs. 8 and 22. Such a unit cell is shown in Figure 14.4 and is simply a transmission line of characteristic impedance Z_0, periodically loaded with series capacitors C_0 and shunt inductors L_0 in a dual fashion. A representative dispersion diagram for a typical host transmission line loaded with series capacitors and shunt inductors is shown in Figure 14.5.

The metamaterial phase-shifting lines can then be constructed by cascading a series of these unit cells. The edges of the stopband f_{c1} and f_{c2} in Figure 14.5 are determined at the series resonance between the inductance of the transmission-line section L, and the loading capacitor C_0, and at the shunt resonance between the capacitance of the transmission-line section C, and the loading inductance L_0, respectively. Alternatively, these are the frequencies at which the effective permeability $\mu_N(\omega)$ and effective permittivity $\varepsilon_N(\omega)$ vanish; that is, $\varepsilon_N(\omega) = 0$ and $\mu_N(\omega) = 0$. Hence by setting the

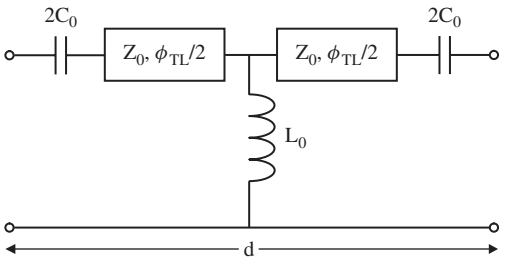


Figure 14.4 Unit cell of a metamaterial phase-shifting line comprising a host transmission line periodically loaded with series capacitors and shunt inductors.

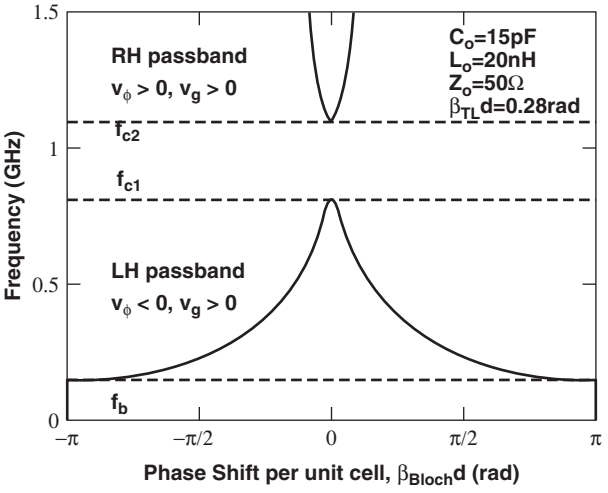


Figure 14.5 The dispersion diagram for the periodic structure of Figure 14.4 with typical line and loading parameters. The edges of the stopband are designated by f_{c1} and f_{c2}.

effective material parameters of Eq. (14.1) to zero, these cutoff frequencies are readily determined to be

$$f_{c1} = \frac{1}{2\pi}\sqrt{\frac{1/g}{\mu_p C_0 d}} \tag{14.2}$$

$$f_{c2} = \frac{1}{2\pi}\sqrt{\frac{g}{\varepsilon_p L_0 d}} \tag{14.3}$$

where the characteristic impedance of the host transmission line is $Z_0 = g\sqrt{\mu_p/\varepsilon_p} = \sqrt{L/C}$. By equating f_{c1} and f_{c2}, the stopband in Figure 14.5 can be closed, thus allowing access to phase shifts around the zero mark. The condition for a closed stopband is therefore determined to be

$$Z_0 = \sqrt{\frac{L_0}{C_0}} \tag{14.4}$$

This condition also implies that the PRI transmission line of Figure 14.2 is matched to the NRI line. The closed stopband condition, Eq. (14.4), was originally derived from Eq. (29) of Ref. 8 [Eq. (29)] and subsequently also reported in Ref. 14. Under this condition, and assuming that the unit cell size and phase shift per unit cell are small, an effective propagation constant, β_{eff}, can be derived and the approximate phase shift per metamaterial unit cell can be written as [22]

$$\phi_{\text{MTM}} = -\beta_{\text{eff}} d \approx -\omega\sqrt{LC}d + \frac{1}{\omega\sqrt{L_0 C_0}} \tag{14.5}$$

This expression can be interpreted as the sum of the phase incurred by the host L-C transmission line and a uniform backward C-L line as implied in Figures 14.2 and 14.3.

Various 1D phase-shifting lines were constructed in coplanar waveguide (CPW) technology at 0.9 GHz, as shown in Figure 14.6. The simulated and measured phase responses for a two-stage and a four-stage 0° phase-shifting line are shown in Figure 14.7, compared to the phase response of a conventional -360° TL and a -360° lowpass loaded line. Also shown is the magnitude response of the two-stage and four-stage 0° phase-shifting line.

It can be observed that the experimental results correspond very closely to the simulated results, highlighting the broadband nature of the phase-shifting lines and their small losses.

It can be concluded that the metamaterial phase-shifting lines offer some significant advantages when compared to conventional delay lines. They are compact in size, can be easily fabricated using standard etching techniques, and exhibit a linear and broadband phase response (with a large group velocity) around the design frequency. They can incur either a negative or a positive phase, as well as a 0° phase depending on the values of the loading elements, while maintaining a short physical length. In addition, the phase incurred is independent of the length of the structure. Due to their compact, planar design, they lend themselves easily toward integration with other microwave components and devices. The metamaterial phase-shifting lines are therefore well suited for broadband applications requiring small, versatile, linear devices.

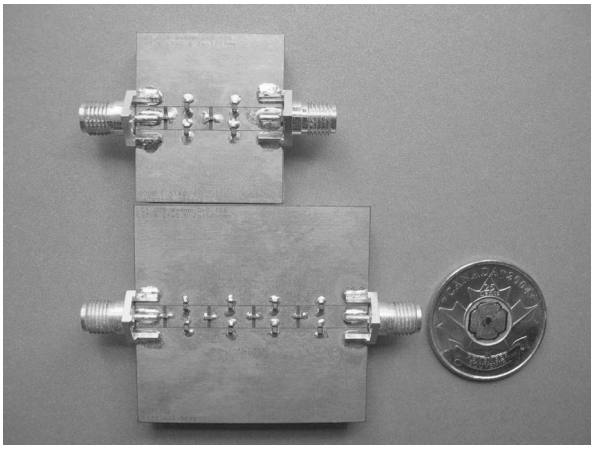


Figure 14.6 (Top) Two-stage phase-shifting line (16 mm). (Bottom) Four-stage phase-shifting line (32 mm) at 0.9 GHz. Note: Reference -360° TL line 283.5 mm (not shown).

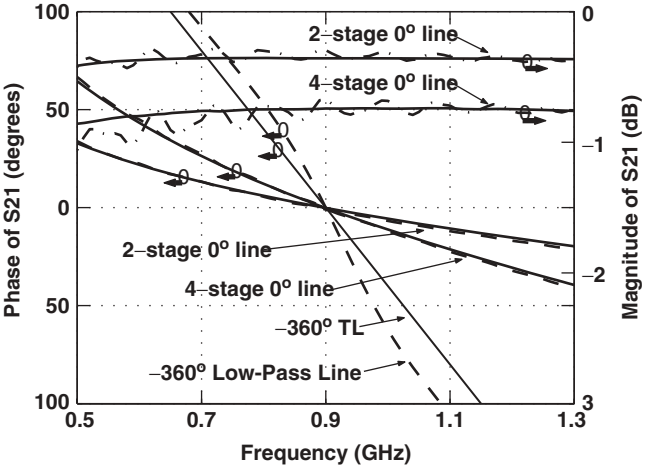


Figure 14.7 Phase responses of two- and four-stage 0° phase-shifting lines compared to a conventional -360° TL and a -360° lowpass loaded line at 0.9 GHz (---) measured and (—) Agilent-ADS simulation. Also shown is the amplitude response for the two- and four-stage devices (-·-) measured and (—) Agilent-ADS simulation.

It should be pointed out that these phase-shifting lines offer an advantage in terms of size and bandwidth, primarily when phase shifts about the zero-degree mark are needed. In this case, the proposed devices have a clear advantage when compared to a corresponding delay line about the one-wavelength mark (see Figure 14.7. Moreover, the same advantage is maintained against a slow-wave lowpass loaded line operating about the -2π mark, as also shown in Figure 14.7. This significant advantage arises from the short electrical length of the zero phase-shift lines, which also implies a broadband phase response (always when comparing to a one-wavelength delay line). Examples of harnessing these advantages in practical applications are discussed later. It should be pointed out that for electrically long PRI/NRI phase-shifting lines their broadband nature could be retained if the constituent NRI section is also designed to exhibit a negative group velocity, as was done in Ref. 24. In this case, not only the signs but also the slopes of the propagation constants (versus frequency) of the NRI and PRI lines compensate, thus leading to an inherently broadband response. Of course, the difficulty now becomes the issue of how to synthesize a negative group velocity over a broad bandwidth. Moreover the NRI lines of Ref. 24 are lossy and hence restoring amplifiers would need to be included for acceptable performance. A comprehensive review for some of these salient issues can be found in Ref. 25.

14.3.1 Nonradiating NRI-TL Metamaterial Phase-Shifting Lines

Any artificial transmission line that supports fast waves, that is waves whose phase velocity is greater than the speed of light, will tend to radiate into free space if its electrical length is sufficiently long. When the metamaterial (MTM) unit cells presented in the previous section are used to create zero-degree phase-shifting lines, the phase incurred by each unit cell, and therefore the propagation constant, is equal to zero at the design frequency. Since the phase velocity is defined as $v_\phi = \omega/\beta$, its value will be

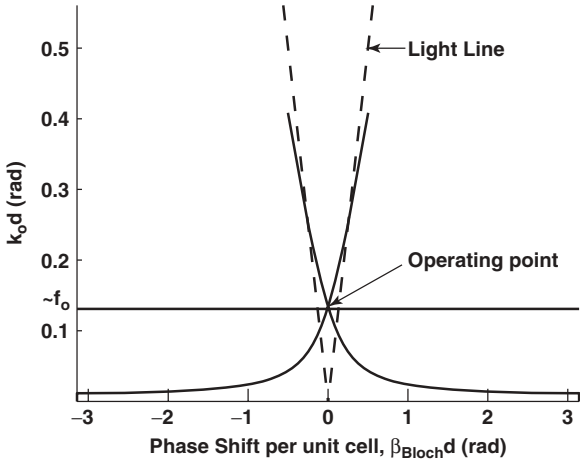


Figure 14.8 Dispersion diagram for a zero-degree MTM unit cell; the horizontal line designates the operating frequency.

infinite at the design frequency. Thus, the lines will support fast waves that will tend to radiate when they are long enough. A typical dispersion diagram with a closed stopband (see Eq. (14.4)) for a zero-degree MTM unit cell of the type shown in Figure 14.4 is depicted in Figure 14.8. If the MTM unit cells are designed to operate anywhere within the radiation cone of the Brillouin diagram, the lines will be prone to radiation.

In order to ensure that the MTM phase-shifting lines do not radiate, each MTM unit cell must be operated in the NRI backward-wave region, while simultaneously ensuring that the propagation constant exceeds that of free space. This will effectively produce a slow-wave structure with a positive phase shift per unit cell, ϕ_{MTM}. The Brillouin diagram for this scenario is shown in Figure 14.9. It can be observed that at the design frequency f_0, propagation occurs outside the radiation cone. The total insertion phase for an n-stage phase-shifting line is simply given by $\Phi_{\text{MTM}} = n(\phi_{\text{MTM}})$.

By cascading the MTM line of Figure 14.4 with a conventional (PRI) TL that inherently incurs a negative insertion phase, Φ_{TL}, a composite slow-wave metamaterial phase-shifting line is obtained that does not radiate, as shown in Figure 14.10. Furthermore, if Φ_{MTM} and Φ_{TL} are equal but opposite in sign, the structure will incur a zero insertion phase, given by Eq. (14.6):

$$\Phi_0 = \Phi_{\text{MTM}} + \Phi_{\text{TL}} = 0 \tag{14.6}$$

Thus it has been shown that it is possible to construct MTM phase-shifting lines that do not radiate, which can then be used for the design of antenna feed networks, without affecting the radiation patterns.

14.4 BROADBAND AND COMPACT NRI-TL METAMATERIAL SERIES POWER DIVIDERS

Applications that require equal, in-phase power division to a series of loads that are spaced less than a wavelength apart have traditionally used a one-guided-wavelength (λ_g) meandered

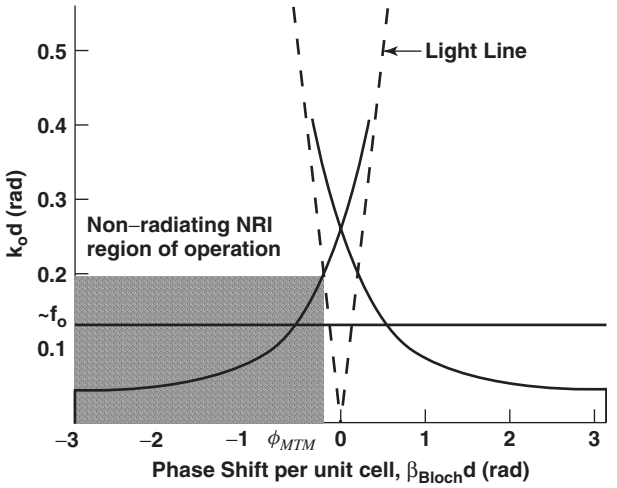


Figure 14.9 Dispersion diagram indicating regions of propagation outside the radiation cone; the horizontal line designates the operating frequency.

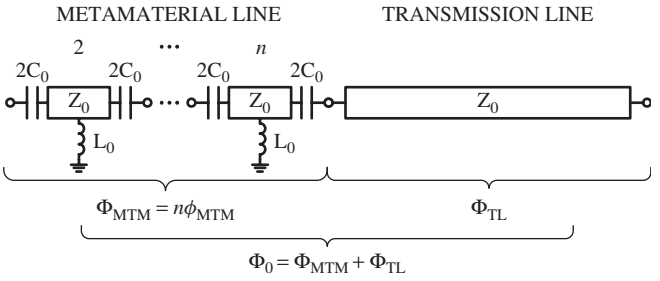


Figure 14.10 Nonradiating (slow-wave) metamaterial phase-shifting line.

TL to feed each of the loads. However, the resulting TL-based power dividers are large and narrowband due to the inherent frequency dependence of the one-wavelength TLs employed. By using the previously described nonradiating zero-phase shifting MTM lines, broadband and compact series power dividers can be implemented. For example, the 1:4 series power divider shown in Figure 14.11 provides equal power split to all four output ports over a significantly larger bandwidth compared to its conventional meander-line TL counterpart [26]. In addition, the MTM divider is very compact, scalable in size, and can be extended to an arbitrary number of ports. Therefore it is well suited for various applications including planar antenna feed networks [27] and power-combining amplifiers [28]. The MTM divider of Figure 14.11 consists of four series-connected nonradiating 0° MTM lines that feed four 200-Ω loads, spaced $\lambda_0/8$ apart. Thus at the design frequency of $f_0 = 1.92$ GHz, which is at the center of the PCS1900 band, the four loads appear in parallel and the circuit is matched to 50 Ω. Each 200-Ω load was realized by connecting two quarter-wavelength transformers to the test equipment 50-Ω port.

Both a meander-line TL and a MTM divider were fabricated, and the simulated and measured responses showed good agreement. The TL divider exhibited a measured

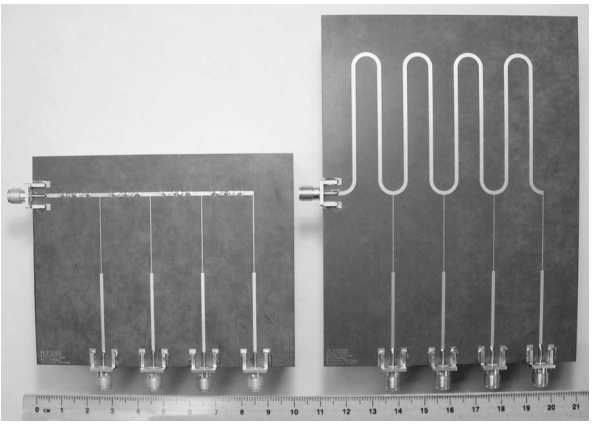


Figure 14.11 Photograph of the MTM divider (left) and the TL divider (right).

−10-dB return-loss bandwidth, $BW_{S_{11}}$, of 0.26 GHz, while the MTM divider had a $BW_{S_{11}}$ of 0.69 GHz, which corresponds to a 165% increase in the measured bandwidth. A representative through measurement to port 4 is shown in Figure 14.12. The through response for the TL divider is more narrowband, with a −3-dB through bandwidth to port 4, $BW_{S_{41}}$, of 0.54 GHz, compared to a $BW_{S_{41}}$ of 1.37 GHz for the MTM divider, which corresponds to a 154% increase in the measured bandwidth. The average measured material loss per port is 0.82 dB for the TL divider and 0.88 dB for the MTM divider, indicating that the losses for the two dividers are comparable and include both the material and the reflection losses. Moreover, the MTM divider is significantly more compact than its TL counterpart, occupying an area of 108 mm^2, which corresponds to a mere 2.6% of the 4098-mm^2 area that the TL meander-line divider occupies.

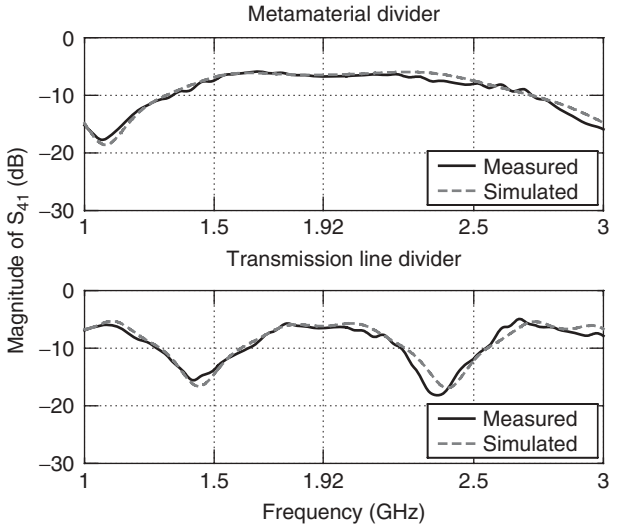


Figure 14.12 Measured versus simulated S_{41} magnitude responses for the MTM and TL meander-line dividers.

14.5 NRI-TL METAMATERIAL SERIES-FED ANTENNA ARRAYS WITH REDUCED BEAM SQUINTING

The nonradiating metamaterial phase-shifting lines described in Section 14.3.1 can be used to feed in phase the antenna elements within a series-fed linear antenna array, as was originally proposed in Ref. 27, and realized in Ref. 47. In this section, zero-degree nonradiating metamaterial phase-shifting lines are used to feed a fully printed four-element dipole array with an interelement spacing of $d_E = \lambda_0/4$. The metamaterial lines have a quasi true-time delay nature, which enables them to exhibit a broadband phase response. This, in turn, results in array patterns that experience less beam squinting with a change in frequency when compared to their conventional series-fed counterparts that employ meandered one-wavelength transmission lines (TLs) or lowpass loaded TLs that incur -2π rad phase shift.

14.5.1 Uniform Linear Arrays Employing True-Time Delay Phase Shifters

Phase shifters that have a phase response that is a linear function of frequency which also passes through the origin are said to exhibit true-time delay (TTD) characteristics [29]. True-time delay phase shifters have been implemented using many techniques including acoustic wave delay lines (including bulk acoustic wave (BAW) and surface acoustic wave (SAW) delay lines) [30–32], magnetostatic wave (MSW) delay lines [33, 34], traveling-wave tube (TWT) structures [35], optical implementations using heterodyne techniques [36], and MMIC implementations using constant-R networks [37]. However, by far the simplest and most common implementation of a true-time delay phase shifter is a simple section of a TEM transmission line, whose insertion phase can be written as

$$\phi_{\mathrm{TL}} = -\omega\sqrt{LC}d \tag{14.7}$$

which is by inspection a linear function of frequency. Here, L and C are the distributed inductance and capacitance of the TL, and d is the length of the TL.

True-time delay phase shifters have received considerable attention in the context of linear antenna array systems, because of their ability to eliminate *beam squinting*, the frequency dependence of the emerging beam's scan angle. In order to further understand the importance of a linear phase characteristic and its effect on beam squinting, it is instructive to consider a uniform linear array of N isotropic radiating elements arranged along the x-direction as shown in Figure 14.13.

Here, the numbers from 1 to N indicate the locations of the N radiating elements in the array and the angle γ indicates the location of the emerging beam from broadside. It is related to the azimuthal angle ϕ as follows: $\gamma = \pi/2 - \phi$. The interelement spacing is d_E, and the progressive phase shift between each of the elements is Φ_0. For generality, it is assumed that Φ_0 can take on positive or negative values corresponding to a current phase lead or lag, respectively. By applying suitable far-field approximations, the array factor (AF) for the N-element uniform linear array can be written as [38]

$$AF = \frac{1}{N}\frac{\sin((N/2)\psi)}{\sin\left(\frac{1}{2}\psi\right)} \tag{14.8}$$

where

$$\psi = k_0 d_E \sin(\gamma) + \Phi_0 \tag{14.9}$$

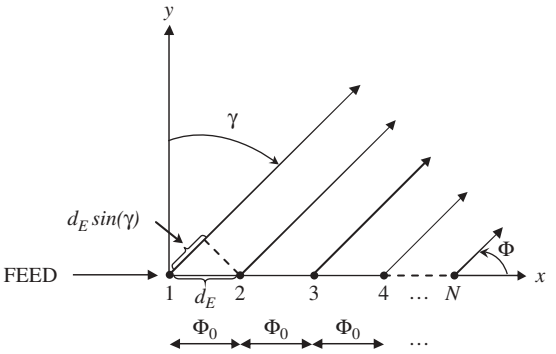


Figure 14.13 N-element uniform linear array of isotropic elements along the x-axis.

Here, k_0 is the wavenumber in free space, which can also be written $k_0 = \omega/c$, where ω is the angular frequency and c is the speed of light. In order for the array factor to attain its maximum value, ψ must equal zero. By setting Eq. (14.9) equal to zero and rearranging, the interelement phasing of the array elements that result in a maximum for the AF and therefore produce a main beam at the scan angle γ_{sc} is

$$\Phi_0 = -k_0 d_E \sin(\gamma_{sc}) = -\frac{\omega}{c} d_E \sin(\gamma_{sc}) \tag{14.10}$$

The scan angle for a uniform linear array can thus be written

$$\gamma_{sc} = \sin^{-1}\left(-\frac{\Phi_0}{k_0 d_E}\right) = \sin^{-1}\left(-\frac{c\Phi_0}{\omega d_E}\right) \tag{14.11}$$

In typical phased array systems, the interelement phasing is carried out through the use of phase shifters. If the phase shifters are independent of frequency, then a simple inspection of Eq. (14.11) reveals that the scan angle of the array will be frequency dependent, since c and d_E are constant. However, if Φ_0 is a linear function of frequency, then the frequency terms in the expression for γ_{sc} will cancel, therefore rendering the scan angle frequency independent. Thus for a true-time delay phase shifter, we can write the interelement phasing relationship as

$$\Phi_0 = \omega T_0 \tag{14.12}$$

where T_0 is a constant that is determined by the type of true-time delay phase shifter used. The resulting frequency-independent scan angle can be written

$$\gamma_{sc} = \sin^{-1}\left(-\frac{\Phi_0}{k_0 d_E}\right) = \sin^{-1}\left(-\frac{cT_0}{d_E}\right) \tag{14.13}$$

The term true-time delay originates from the fact that for an interelement phase relationship that is linear with frequency, like the one shown in Eq. (14.12), the time delay

between successive array elements, T_0, is independent of frequency. This can be seen by substituting Eq. (14.10) into Eq. (14.12) and solving for T_0:

$$T_0 = \frac{\Phi_0}{\omega} = -\frac{1}{c} d_E \sin(\gamma_{sc}) \tag{14.14}$$

Since Eq. (14.14) does not contain any frequency terms, if there is a successive time delay or advance between the array elements given by Eq. (14.14), then the emerging beam will be dependent only on the time shift T_0 and will not vary with frequency. Thus an array that uses time delay lines to excite the individual array elements will produce an emerging beam that does not squint with frequency. It follows that if a phase shifter can provide a linear phase characteristic with frequency, it essentially mimics the performance of a real time delay line, which is frequency independent. It is therefore said to exhibit "true-time delay" behavior and can thus be used to eliminate the generally undesirable phenomenon of beam squinting.

14.5.2 Transmission-Line-Fed Series Uniform Linear Arrays

As mentioned earlier, one of the most commonly used true-time delay phase shifters is a simple section of a transmission line (TL). If this type of phase shifter is used to provide the interelement phase shift Φ_0 in a series-fed uniform linear array like the four-element array shown in Figure 14.14, then Φ_0 can be written

$$\Phi_0 = \phi_{TL} = -\beta_{TL} d = -\omega\sqrt{LC}d = -\frac{\omega}{v_\phi} d = -\frac{\sqrt{\varepsilon_{eff}}}{c}\omega d \tag{14.15}$$

It should be noted that Φ_0 in Eq. (14.15) is now a negative quantity, to reflect the phase delay nature of transmission lines. In Eq. (14.15) we have used the following expression

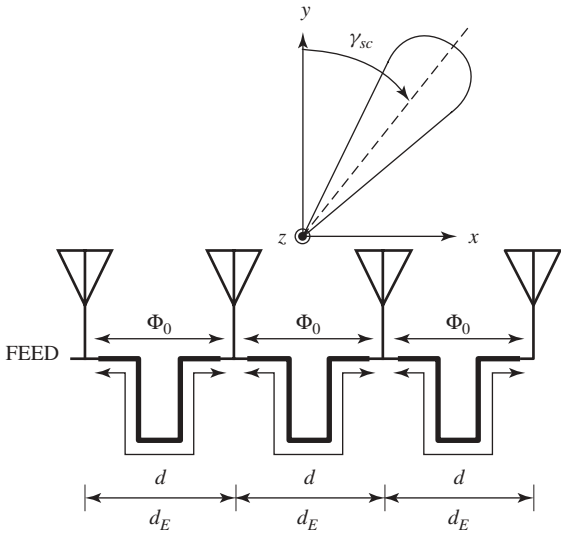


Figure 14.14 Transmission-line series-fed uniform linear array of four elements.

for the phase velocity of the transmission line:

$$v_\phi = \frac{c}{\sqrt{\varepsilon_{\text{eff}}}} \tag{14.16}$$

Here, ε_{eff} is the effective permittivity of the substrate on which the antenna array is printed. Therefore the scan angle of Eq. (14.11) can be expressed as

$$\gamma_{\text{sc,TL}} = \sin^{-1}\left(\frac{c\left(-\frac{\sqrt{\varepsilon_{\text{eff}}}}{c}\omega d\right)}{\omega d_E}\right) = \sin^{-1}\left(\frac{\sqrt{\varepsilon_{\text{eff}}}d}{d_E}\right) \tag{14.17}$$

The expression for the scan angle γ_{sc} from Eq. (14.17) is frequency independent and therefore it would appear that one could design squint-free series-fed linear arrays using conventional true-time delay TLs. However, it will be shown later that the physical constraints of a printed series-fed array actually eliminate the true-time delay nature of the TLs, therefore causing the beam to squint with frequency.

A convenient way to visualize the direction of the scan angle γ_{sc} is to construct a rectangular-to-polar graphical representation of the array factor, as shown in Figure 14.15 for an $N = 4$ element uniform array with an interelement spacing of $d_E = \lambda_0/4$ and a progressive phase shift of $\Phi_0 = -\pi/4$ rad. The radius R of the circle is given by

$$R = k_0 d_E \tag{14.18}$$

and the center of the circle with respect to the origin is given by

$$C = \Phi_0 \tag{14.19}$$

The visible region of the array pattern is given by [40]

$$\Phi_0 - k_0 d_E \leq \psi \leq \Phi_0 + k_0 d_E \tag{14.20}$$

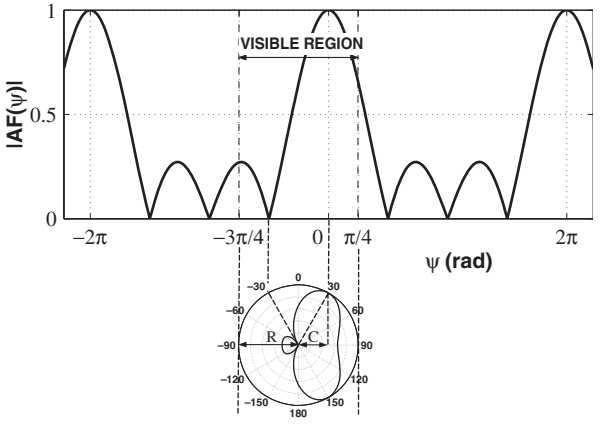


Figure 14.15 Rectangular-to-polar graphical representation of the array factor for a linear array with $N = 4$, $d_E = \lambda_0/4$, and $\Phi_0 = -\pi/4$ rad.

Thus, the radius of the circle for the example shown is $R = \pi/2$ rad and the center of the circle is located at $C = -\pi/4$ rad; that is, it is shifted to the left by $\pi/4$ rad. The visible region for this example is therefore $-3\pi/4 \leq \psi \leq \pi/4$ rad. It can be seen that within the visible region, the radiation pattern exhibits a major lobe at a scan angle of $\gamma_{sc} = 30^\circ$, a single sidelobe at $\gamma = -75^\circ$, and a null $\gamma = -30^\circ$. As Φ_0 is decreased from zero to $-\pi/2$ rad, the location of the main beam will move from broadside ($\gamma_{sc} = 0^\circ$) when $\Phi_0 = 0$ rad, through $\gamma_{sc} = 30^\circ$ when $\Phi_0 = -\pi/4$ rad, to endfire ($\gamma_{sc} = 90^\circ$) when $\Phi_0 = -\pi/2$ rad, thus covering the whole scanning range of the array.

Inspection of Eq. (14.17) reveals that if $\sqrt{\varepsilon_{eff}}d$ exceeds d_E, then the argument of the $\sin^{-1}(\ )$ function will exceed one, resulting in a physically unrealizable scan angle γ_{sc}. On the polar plot of Figure 14.15, this corresponds to the case where the distance between the center of the circle and the origin is physically larger than the radius of the circle, that is $|C| > |R|$. Therefore this creates a range of interelement phase shifts (Φ_0) for which a scan angle does not exist, because the main lobe can no longer be captured by the polar plot. Physically, this will result in a radiation pattern that contains some of the sidelobes, without capturing the peak of the main lobe. Thus the following condition should be satisfied if conventional transmission lines are to be used as true-time delay phase-shifting devices in a printed TL-fed series linear array:

$$\sqrt{\varepsilon_{eff}}d \leq d_E \tag{14.21}$$

The physical implication of Eq. (14.21) is that the antenna elements must be spaced a distance of at least the length of the interconnecting transmission lines, multiplied by the square root of the effective permittivity, which is typically larger than one for circuits printed on a dielectric substrate.

More generally, for any type of phase shifter, by following the same arguments outlined above, Eq. (14.11) yields the following condition that must be satisfied in order for the scanning angle γ_{sc} to have a physical meaning:

$$|\Phi_0| \leq k_0 d_E \tag{14.22}$$

The series configuration of Figure 14.14 imposes yet another important physical limitation on the choice of d and d_E. Since the sections of transmission lines must be long enough to physically connect the array elements, the length of the transmission lines cannot be less than the interelement spacing; that is

$$d \geq d_E \tag{14.23}$$

This condition, however, directly contradicts the condition imposed by Eq. (14.21), since as already stated, ε_{eff} is typically larger than one for antennas printed on a dielectric substrate. This seemingly contradictory situation can be resolved by adding a section of transmission line that is one guided wavelength λ_g long to the transmission line length, d. This is equivalent to shifting the visible region of the radiation pattern shown in Figure 14.15 to the left by -2π rad. Therefore d becomes d':

$$d' = d + \lambda_g \tag{14.24}$$

This new value of d now satisfies Eq. (14.23) for typical values of $d_E < \lambda_g$ which is imposed to avoid grating lobes. In addition, because d is now greater than d_E, this

implies that the interconnecting transmission line must be meandered in order for it to fit in between the antenna elements, as shown in Figure 14.14. The interelement phase shift Φ_0 remains effectively the same because of its periodicity in 2π, which has a modulo -2π property. Therefore Eq. (14.22) is also satisfied, implying that the series-fed array will produce an emerging beam at an angle γ_{sc}. However, because the center of the visible region has now been shifted to the left by -2π rad, Φ_0 undergoes the following transformation and becomes Φ_0':

$$\Phi_0' = -2\pi + \Phi_0 \tag{14.25}$$

By replacing Φ_0 with $-2\pi + \Phi_0$, this has the effect of eliminating the true-time delay characteristic that the delay lines would otherwise offer. This can be observed by considering the modified phase–frequency characteristic of the delay lines after the transformation of Eq. (14.25). This is shown graphically in Figure 14.16 for the representative case of $\Phi_0 = 0$ rad.

The transformation has the effect of shifting the phase characteristic up by a factor of 2π, while still maintaining an effective phase shift of Φ_0 at the design frequency of ω_0. Thus for frequencies below ω_0 the transmission line will effectively produce a positive phase shift and for frequencies above ω_0 it will produce a negative phase shift. The modified phase as a function of frequency for the TL is therefore given by the following equation:

$$\phi_{TL}' = (\Delta\phi/\Delta\omega)\omega + b \tag{14.26}$$

Here, $\Delta\phi/\Delta\omega$ is the slope of the line, and b is the ϕ-axis intercept point. For the particular example shown in Figure 14.16, and for all transmission lines that are augmented by a one-guided-wavelength λ_g section, $b = 2\pi$.

The true-time delay characteristic of the delay lines is eliminated because the transformation of Eq. (14.25) introduces a -2π constant term in the previously linear relationship between Φ_0 and ω and therefore the expression for the scan angle γ_{sc} from Eq. (14.11)

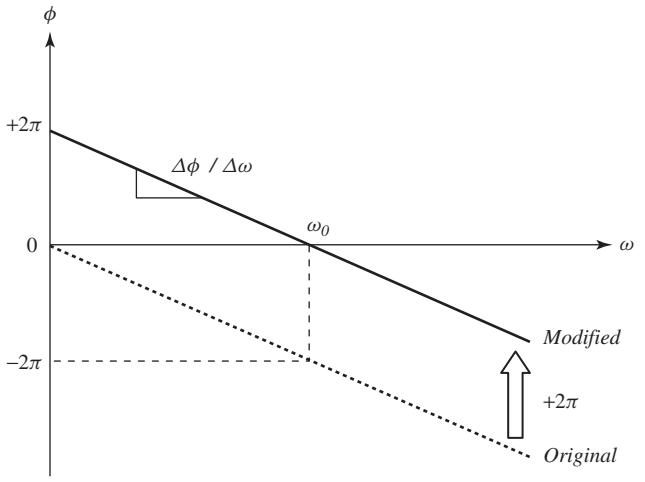


Figure 14.16 Modified phase versus frequency characteristic for a TL.

now contains an extra term that is inversely proportional to frequency:

$$\gamma_{\mathrm{sc,TL}} = \sin^{-1}\left(-\frac{-2\pi + \Phi_0}{k_0 d_E}\right) = \sin^{-1}\left(\frac{2\pi c}{d_E}\frac{1}{\omega} + \frac{\sqrt{\varepsilon_{\mathrm{eff}}}d}{d_E}\right) \tag{14.27}$$

Thus the TL-fed series uniform linear array can be physically implemented by adding a section of transmission line that is one guided wavelength λ_g long to the interconnecting transmission lines. This, however, results in a frequency-dependent scan angle γ_{sc}, where any variation from the design frequency will cause the emerging beam to squint.

It will be shown in the next section that, contrary to conventional delay lines, metamaterial lines possess an inherent quasi-TTD characteristic that can be exploited in series-fed arrays to effectively reduce the amount of beam squint that they experience with frequency.

14.5.3 Metamaterial-Fed Series Uniform Linear Arrays

In a typical series-fed linear array designed to radiate at broadside, the antenna elements must be fed in phase. In addition, an interelement spacing d_E of less than a half free-space wavelength ($d_E < \lambda_0/2$) is necessary to avoid capturing grating lobes in the visible region of the array pattern (see Figure 14.15. In order to achieve these design constraints, traditional designs employing TL-based feed networks have resorted to a meander-line approach, as described in Section 14.5.2 and shown in Figure 14.14. This allows the antenna elements to be physically separated by a distance of $d_E = \lambda_0/4$, while still being fed in phase with a one-guided-wavelength λ_g long meandered line that incurs a phase of -2π rad. Because the phase incurred by the TLs is linearly dependent on frequency with a constant offset (see Eq. (14.26)), a change in the operating frequency will cause the emerging beam to squint from broadside, which is generally an undesirable phenomenon. In addition, the fact that the lines are meandered causes the radiation pattern to experience high cross-polarization levels, as a result of parasitic radiation due to scattering from the corners of the meandered lines [39]. The feed networks shown in this section employ nonradiating metamaterial (MTM) phase-shifting lines within a series-fed linear array, as shown in Figure 14.17, to mitigate some of the problems encountered with conventional TL-based feed networks.

In order to investigate how the main beam of the uniformly excited MTM array shifts from the broadside direction as the frequency is varied, we must first consider the phase response of the MTM feed lines. For nonradiating MTM lines, the approximate phase incurred by the line is given by Eq. (14.6) and can be written as

$$\Phi_0 = \Phi_{\mathrm{MTM}} + \Phi_{\mathrm{TL}} = n\left(-\omega\sqrt{LC}d_{\mathrm{TL,H}} + \frac{1}{\omega\sqrt{L_0C_0}}\right) + (-\omega\sqrt{LC}d_{\mathrm{TL}}) \tag{14.28}$$

Here, L and C are the distributed inductance and capacitance correspondingly of the host transmission line, $d_{\mathrm{TL,H}}$ is the length of the host transmission line section ($d = d_{\mathrm{TL,H}}$ in Figure 14.4), L_0 and C_0 are the loading-element values, n is the number of metamaterial stages, d_{TL} is the length of the negative phase-compensating section of transmission line, and ω is the frequency of operation. Assuming that the same type of TL sections are used for $d_{\mathrm{TL,H}}$ and d_{TL}, then the characteristic impedance, Z_0, and L and C will be the same for both lines.

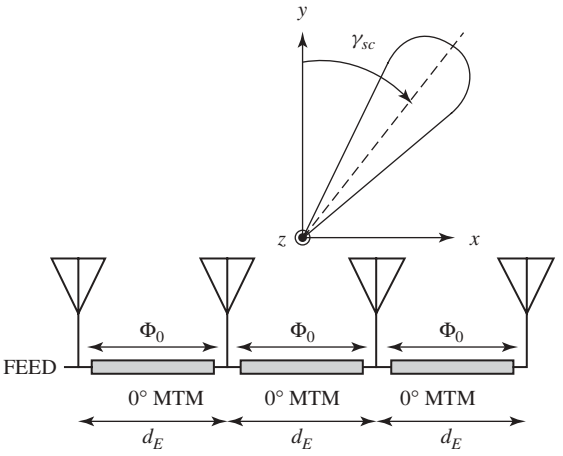


Figure 14.17 Metamaterial-fed series uniform linear array of four elements.

By substituting Eq. (14.28) into Eq. (14.11), the scan angle for the uniformly excited MTM array with the antenna elements equally distributed along the x-axis as shown in Figure 14.17 can therefore be written as

$$\begin{aligned}\gamma_{\text{sc,MTM}} &= \sin^{-1}\left(-\frac{\Phi_{\text{MTM}}+\Phi_{\text{TL}}}{k_0 d_E}\right)\\ &= \sin^{-1}\left(-\frac{cn}{d_E\sqrt{L_0C_0}}\frac{1}{\omega^2}+\frac{c}{d_E}\sqrt{LC}(nd_{\text{TL,H}}+d_{\text{TL}})\right)\end{aligned} \tag{14.29}$$

Inspection of Eq. (14.29) reveals that the first term in the $\sin^{-1}(\ )$ expression, $-cn/d_E\sqrt{L_0C_0}\omega^2$, is inversely proportional to the frequency squared and the second term, $c\sqrt{LC}\ (nd_{\text{TL,H}}+d_{\text{TL}})/d_E$, is a constant. Thus, if the first term that is frequency dependent can be made negligible compared to the second term, then the dependence of γ_{sc} on ω will diminish and the scan angle will become relatively insensitive to frequency variations. Due to the inverse-frequency-squared dependence of the $-cn/d_E\sqrt{L_0C_0}\omega^2$ term, this will become negligible compared to $c\sqrt{LC}(nd_{\text{TL,H}}+d_{\text{TL}})/d_E$ as the frequency increases. Thus at high frequencies the constant term $c\sqrt{LC}(nd_{\text{TL,H}}+d_{\text{TL}})/d_E$ will dominate, and γ_{sc} will remain relatively constant with frequency.

Therefore, it can be concluded that even though the metamaterial lines do not inherently possess a pure TTD characteristic, by operating at high enough frequencies, a quasi-TTD characteristic can be obtained, thus rendering the scan angle relatively frequency insensitive.

If we return to the scan angle characteristic of the TL-fed array given by Eq. (14.27), we can see that the frequency-dependent term in this expression, $-2\pi c/d_E\omega$, is inversely proportional to frequency, and not inversely proportional to the frequency squared as in the case of the metamaterial scan angle. Thus the scan angle of the TL-fed array will inherently be more frequency dependent, even at higher frequencies where the scan angle of the MTM-fed array is relatively frequency independent.

The theoretical scan angles from broadside for the TL-fed and MTM-fed linear arrays with $d_E = \lambda_0/4 = 15$ mm were calculated at $f_0 = 5$ GHz using Eq. (14.27) and (14.29),

respectively, and the scan angle characteristics are shown in Figure 14.18. The parameters used to calculate the phase responses of the TL feed lines were $d = \lambda_g = 52.44$ mm, $Z_0 = 170\ \Omega$ and of the 0° MTM feed lines, $n = 1$, $d_{\mathrm{TL,H}} = 3$ mm, $d_{\mathrm{TL}} = 12$ mm, $Z_0 = 170\ \Omega$, $C_0 = 0.13$ pF, $L_0 = 2.44$ nH.

It can be observed that the scan angle for the TL-fed array exhibits its full scanning range from endfire ($\gamma_{\mathrm{sc}} = +90^\circ$) to backfire ($\gamma_{\mathrm{sc}} = -90^\circ$) within a bandwidth of 2.67 GHz, while the corresponding scanning bandwidth for the metamaterial-fed array is 10.44 GHz, which amounts to a 291% increase in scan angle bandwidth. The larger scanning bandwidth of the MTM-fed array agrees favorably with the previous discussion relating to the scan angle expressions for the two arrays from Eq. (14.27) and (14.29). It is interesting to note that as the frequency increases above 5 GHz, the gradient of the slope of γ_{sc} versus frequency for the MTM-fed array is the smallest in the region from around 7 to 13.5 GHz, which is consistent with the discussion of Eq. (14.29) that the MTM lines exhibit quasi-TTD characteristics at higher frequencies.

Also shown in Figure 14.18 is the scan angle characteristic for a lowpass loaded (LPL) TL also of length $\lambda_0/4$ and designed around -2π rad. From Figure 14.18, it can be observed that the performance of this line is identical to that of the TL meandered feed line. Thus, although the lowpass loaded line can eliminate the need for meander lines, it does not provide the advantage of an increased scan angle bandwidth that the metamaterial feed lines offer.

14.5.4 Transmission-Line and Metamaterial Series-Fed Printed Dipole Arrays

In order to investigate further the radiation characteristics of a series-fed array as the frequency is varied, it is instructive to consider the total field of the array, which by the pattern multiplication principle is equal to the antenna element factor multiplied by the array factor. Even though an expression for the array factor of an N-element uniform linear array was already presented in Eq. (14.8), it is useful to derive a more general

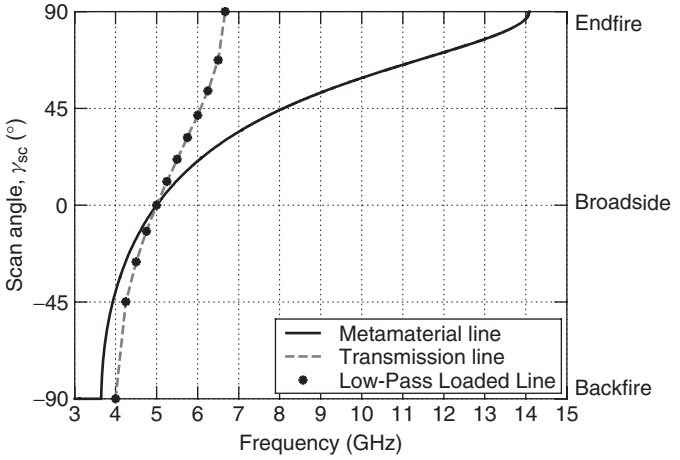


Figure 14.18 Comparison of the scan angle performance from broadside of a series-fed linear array with $d_E = \lambda_0/4$ using a metamaterial, a transmission-line, and a lowpass loaded line feed network.

expression for the array factor that does not assume a uniform excitation of the antenna elements. The general normalized array factor is therefore given by [38]

$$AF = \frac{1}{N}\sum_{n=1}^{N} I_n e^{jk\hat{\mathbf{r}}\cdot\mathbf{r}_n} \tag{14.30}$$

where N is the total number of antenna elements, I_n is the current excitation of the nth element, $\hat{\mathbf{r}}$ is the unit vector for the chosen coordinate system, and $\mathbf{r}_n$ is the position vector from the origin of the nth element. In general, the current excitation I_n has both an amplitude, a_n, and phase, Φ_0, associated with it and is given by:

$$I_n = a_n e^{j(n-1)\Phi_0} \tag{14.31}$$

Four-Element Series-Fed Printed Dipole Array The series-fed linear array considered herein is a four-element printed dipole array with the z-directed dipole elements arranged symmetrically along the x-axis with an interelement spacing of $d_E = \lambda_0/4$, as shown in Figure 14.19.

The position vectors for each of the dipoles are given by

$$\mathbf{r}_1 = -\tfrac{3}{8}\lambda_0\hat{\mathbf{x}}, \quad \mathbf{r}_2 = -\tfrac{1}{8}\lambda_0\hat{\mathbf{x}}, \quad \mathbf{r}_3 = \tfrac{1}{8}\lambda_0\hat{\mathbf{x}}, \quad \mathbf{r}_4 = \tfrac{3}{8}\lambda_0\hat{\mathbf{x}} \tag{14.32}$$

and the unit vector in the rectangular coordinate system is given by

$$\hat{\mathbf{r}} = \cos\phi\sin\theta\hat{\mathbf{x}} + \sin\phi\sin\theta\hat{\mathbf{y}} + \cos\theta\hat{\mathbf{z}} \tag{14.33}$$

The array factor from Eq. (14.30) therefore becomes

$$\begin{aligned} AF = \tfrac{1}{4}(&I_1 e^{-jk(3/8)\lambda_0\cos\phi\sin\theta} + I_2 e^{-jk(1/8)\lambda_0\cos\phi\sin\theta} \\ &+ I_3 e^{+jk(1/8)\lambda_0\cos\phi\sin\theta} + I_4 e^{+jk(3/8)\lambda_0\cos\phi\sin\theta}) \end{aligned} \tag{14.34}$$

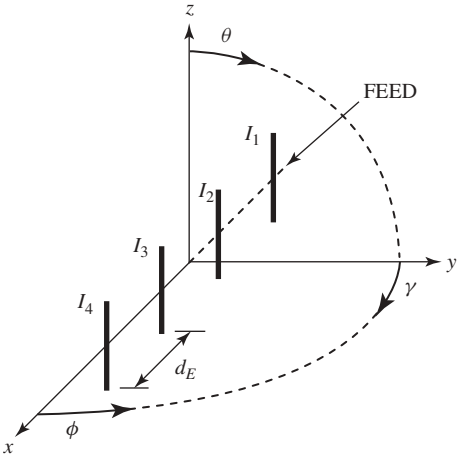


Figure 14.19 Four-element printed dipole array.

If we consider half-wavelength dipoles, then the electric field expression in the far field is given by [38]

$$E_\theta = j\eta \frac{I_0 e^{-jkr}}{2\pi r}\left[\frac{\cos((\pi/2)\cos\theta)}{\sin\theta}\right] \tag{14.35}$$

Therefore the total field pattern for the array is given by the pattern multiplication principle as

$$E_{\text{TOT}} = E_\theta \times AF \tag{14.36}$$

The main beam of the series-fed array shown in Figure 14.19 will scan in the xy-plane, which is perpendicular to the axis of the array. This is also the H-plane of the dipoles and is obtained by setting $\theta = \pi/2$. Considering Eq. (14.35) and setting $\theta = \pi/2$, we can observe the well-known fact that the field of a dipole does not vary in its H-plane. Since any variation of the total electric field of the array will be produced solely by the array factor, henceforth only the array factor will be considered when investigating the scanning characteristics of the array. If the dipoles are oriented in another direction or a different antenna element is used in the array, its field pattern would have to be considered together with the array factor as in Eq. (14.36).

The array factor in the xy-plane for the TL-fed array can be written using Eq. (14.31) and (14.34) as

$$\begin{aligned} AF = {} & \tfrac{1}{4}(a^0 e^{-jk(3/8)\lambda_0\cos\phi} + a^1 e^{j\Phi_{0,\text{TL}}} e^{-jk(1/8)\lambda_0\cos\phi} \\ & + a^2 e^{j2\Phi_{0,\text{TL}}} e^{+jk(1/8)\lambda_0\cos\phi} + a^3 e^{j3\Phi_{0,\text{TL}}} e^{+jk(3/8)\lambda_0\cos\phi}) \end{aligned} \tag{14.37}$$

where a is the amplitude factor per section of interconnecting TL. An amplitude taper along the array can be achieved by setting a to a value less than one. The phase response of the TL feed line is given by Eq. (14.26) as

$$\Phi_{0,\text{TL}} = (\Delta\phi/\Delta\omega)\omega + 2\pi \tag{14.38}$$

Similarly, the array factor in the xy-plane for the MTM-fed array can also be written using Eq. (14.31) and (14.34)

$$\begin{aligned} AF = {} & \tfrac{1}{4}(a^0 e^{-jk(3/8)\lambda_0\cos\phi} + a^1 e^{j\Phi_{0,\text{MTM}}} e^{-jk(1/8)\lambda_0\cos\phi} \\ & + a^2 e^{j2\Phi_{0,\text{MTM}}} e^{+jk(1/8)\lambda_0\cos\phi} + a^3 e^{j3\Phi_{0,\text{MTM}}} e^{+jk(3/8)\lambda_0\cos\phi}) \end{aligned} \tag{14.39}$$

where the phase response of the MTM feed line is given by Eq. (14.28) and can be rewritten as

$$\Phi_{0,\text{MTM}} = n\left(\frac{\Delta\phi_{\text{TL,H}}}{\Delta\omega}\omega + \frac{1}{\omega\sqrt{L_0 C_0}}\right) + \frac{\Delta\phi_{\text{TL}}}{\Delta\omega}\omega \tag{14.40}$$

When Eq. (14.37) and (14.39) are plotted as a function of the azimuthal angle ϕ (or equivalently, the angle from broadside $\gamma = \pi/2 - \phi$), the radiation pattern of the array is obtained. By varying the frequency one can observe how the location of the main beam shifts with frequency, as shown in Figures 14.20 and 14.21 for the three frequencies of 4, 5, and 6 GHz. The parameters used to calculate the phase responses of the TL feed lines were $\Delta\phi/\Delta\omega = -5 \times 10^{-9}$ s and of the 0° MTM feed lines, $n = 1$, $\Delta\phi_{\text{TL,H}}/\Delta\omega = -1.14 \times 10^{-11}$ s, $\Delta\phi_{\text{TL}}/\Delta\omega = -4.58 \times 10^{-11}$ s, $C_0 = 0.13$ pF, and $L_0 = 2.44$ nH.

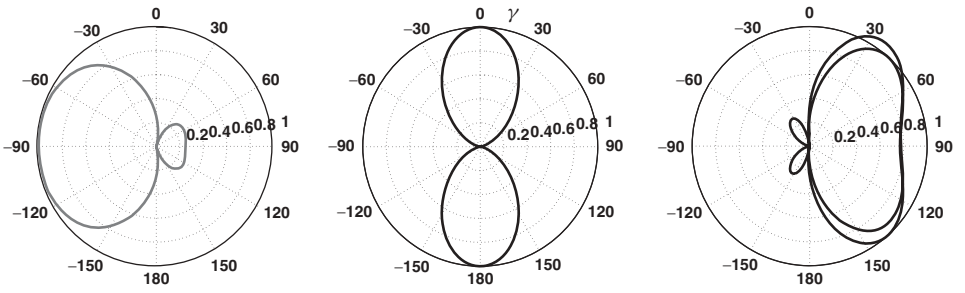


Figure 14.20 Normalized ideal array factor patterns in the xy-plane from Eq. (14.37) for the TL-fed dipole array at (a) 4, (b) 5, and (c) 6 GHz.

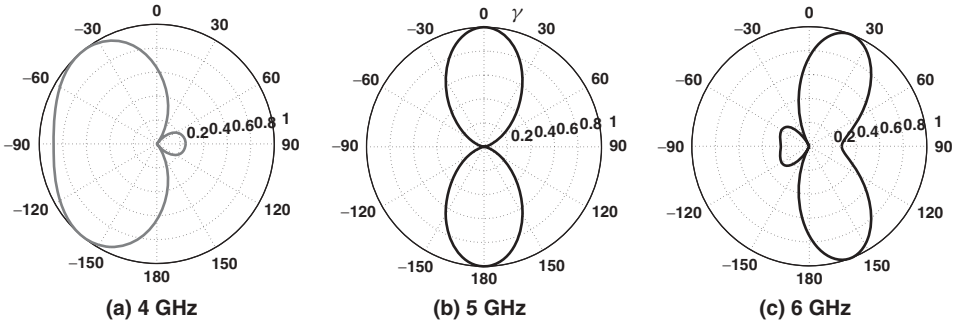


Figure 14.21 Normalized ideal array factor patterns in the xy-plane from Eq. (14.39) for the MTM-fed dipole array at (a) 4, (b) 5, and (c) 6 GHz.

From Figures 14.20 and 14.21, it can be observed that both the TL-fed and the MTM-fed arrays radiate at broadside ($\gamma = 0^{\circ}$) at 5 GHz; however, when the frequency is decreased the beam shifts to the left toward the backfire direction and when the frequency is increased the beam shifts to the right toward the endfire direction. It can also be observed that within the same frequency range of 4–6 GHz, the beam of the MTM-fed array shifts less to the left and right off broadside than the beam of the TL-fed array.

It was found that the scan angle characteristics for both the TL-fed and the MTM-fed arrays were identical to the ones presented in Figure 14.18 for the uniformly excited linear array, regardless of the value of the amplitude factor, a. This indicates that Eq. (14.11), which is valid for uniform linear arrays, is also valid for arrays that have an arbitrary amplitude taper along the axis of the array.

Figure 14.22 shows the amplitude of the maximum value of the array factor for the TL-fed array as a function of frequency for three representative cases of the amplitude factor a. According to Eq. (14.30), the data is normalized such that the value of the maximum array factor is equal to one when the beam is at broadside, that is, at $f_0 = 5$ GHz. By plotting the maximum value of the array factor, this is equivalent to the maximum amplitude of the main beam in the patterns shown, for example, in Figures 14.20

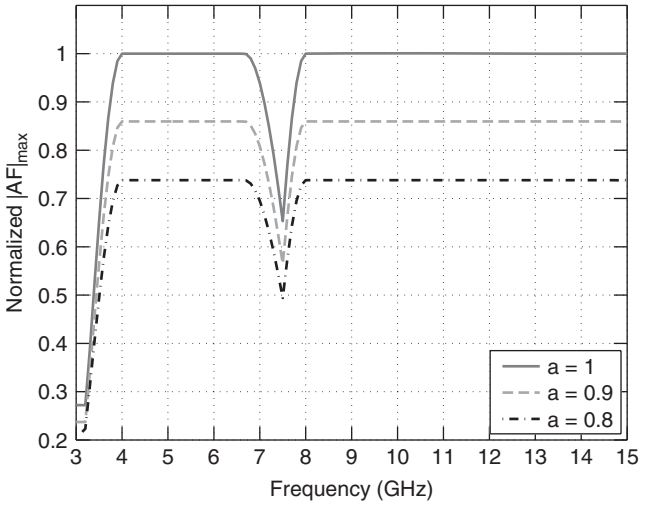


Figure 14.22 Normalized maximum array factor as a function of frequency for the TL-fed array for three representative amplitude factor values: $a = 1$, $a = 0.9$, and $a = 0.8$.

and 14.21. This can be considered a measure of the gain in the direction of maximum radiation.

When no amplitude taper is enforced along the array (i.e., $a = 1$), the maximum array factor remains constant at a value of one from 4 to 6.67 GHz, which is the full scanning range of the TL-fed array as shown in Figure 14.18. In the region from 6.67 to 7.5 GHz, the beam is directed toward the endfire direction and a variation in frequency causes the array factor maximum to drop because less of the endfire main lobe is being captured within the visible region of the array. From 7.5 to 8 GHz, the main beam reverses direction toward the backfire direction and increases in amplitude until 8 GHz. Above 8 GHz, the entire main lobe is captured again within the visible region of the array, and therefore the maximum amplitude of the array factor is restored to one. By observing the three responses shown in Figure 14.22, it can be seen that by enforcing an amplitude taper along the array, this reduces the value of the maximum array factor; however, the general shape of the curves is maintained the same as for the uniformly excited array.

Figure 14.23 shows the amplitude of the normalized maximum value of the array factor as a function of frequency for the MTM-fed array for the same three representative cases of the amplitude factor a.

By comparing Figures 14.22 and 14.23, it can be observed that the MTM-fed array has a clear advantage over the TL-fed array in as much as the amplitude of the main beam does not drop within the whole scanning range of the array, namely, from 3.65 to 14.1 GHz. This is because the peak of the main beam remains within the visible region over this entire frequency range. As in the case of the TL-fed array, enforcing an amplitude taper along the array reduces the value of the maximum array factor, but the shape of the curves is maintained.

Grounded Four-Element Series-Fed Printed Dipole Array In many applications it is desirable to have an antenna with a unidirectional radiation pattern. The printed dipole array described in the previous section will radiate a bidirectional broadside pattern at

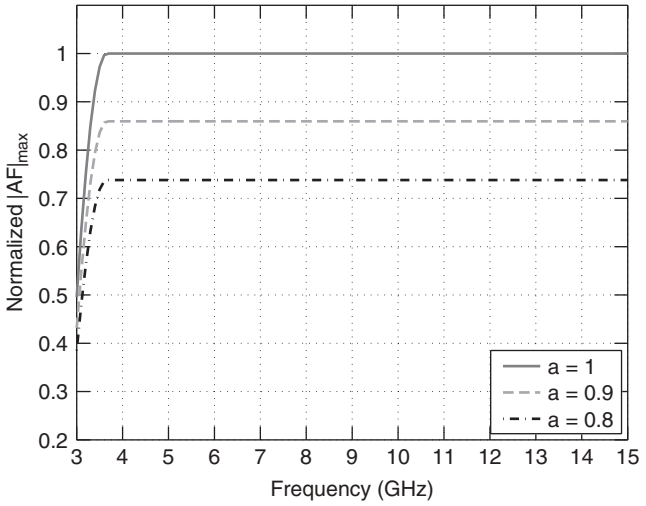


Figure 14.23 Normalized maximum array factor as a function of frequency for the MTM-fed array for three representative amplitude factor values: $a = 1$, $a = 0.9$, and $a = 0.8$.

its design frequency, as shown in Figures 14.20b and 14.21b. A well known technique to obtain a unidirectional pattern from a bidirectional antenna is to add a ground plane at a distance of $d_G = \lambda_0/4$ below the antenna, as shown in Figure 14.24.

Shown in Figure 14.24 is a four-element printed dipole array with $+z$-directed currents $I_1 - I_4$ at a distance of d_G in front of an infinite ground plane in the xz-plane. By image theory, the ground plane creates $-z$-directed image currents $I_5 - I_8$ at a distance of d_G behind the ground plane, thus effectively forming an eight-element antenna array. Thus for currents $I_1 - I_4$, each current I_n can be expressed using Eq. (14.31), while currents $I_5 - I_8$ can be expressed as $-I_n$.

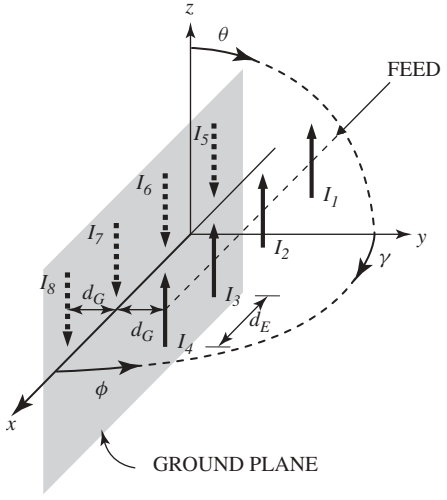


Figure 14.24 Grounded four-element printed dipole array.

The position vectors for each of the dipoles for $d_E = \lambda_0/4$ and $d_G = \lambda_0/4$ are given by

$$\begin{aligned} &\mathbf{r}_1 = -\tfrac{3}{8}\lambda_0\hat{\mathbf{x}} + \tfrac{1}{4}\lambda_0\hat{\mathbf{y}}, \qquad \mathbf{r}_2 = -\tfrac{1}{8}\lambda_0\hat{\mathbf{x}} + \tfrac{1}{4}\lambda_0\hat{\mathbf{y}}, \qquad \mathbf{r}_3 = \tfrac{1}{8}\lambda_0\hat{\mathbf{x}} + \tfrac{1}{4}\lambda_0\hat{\mathbf{y}}, \\ &\mathbf{r}_4 = \tfrac{3}{8}\lambda_0\hat{\mathbf{x}} + \tfrac{1}{4}\lambda_0\hat{\mathbf{y}} \end{aligned} \tag{14.41}$$

The position vectors for each of the image currents are given by

$$\begin{aligned} &\mathbf{r}_5 = -\tfrac{3}{8}\lambda_0\hat{\mathbf{x}} - \tfrac{1}{4}\lambda_0\hat{\mathbf{y}}, \qquad \mathbf{r}_6 = -\tfrac{1}{8}\lambda_0\hat{\mathbf{x}} - \tfrac{1}{4}\lambda_0\hat{\mathbf{y}}, \qquad \mathbf{r}_7 = \tfrac{1}{8}\lambda_0\hat{\mathbf{x}} - \tfrac{1}{4}\lambda_0\hat{\mathbf{y}}, \\ &\mathbf{r}_8 = \tfrac{3}{8}\lambda_0\hat{\mathbf{x}} - \tfrac{1}{4}\lambda_0\hat{\mathbf{y}} \end{aligned} \tag{14.42}$$

Using the unit vector from Eq. (14.33), the array factor from Eq. (14.30) therefore becomes

$$\begin{aligned} AF = \tfrac{1}{8}(&I_1 e^{jk(-(3/8)\lambda_0\cos\phi\sin\theta+(1/4)\lambda_0\sin\phi\sin\theta)} + I_2 e^{jk(-(1/8)\lambda_0\cos\phi\sin\theta+(1/4)\lambda_0\sin\phi\sin\theta)} \\ &+ I_3 e^{jk(+(1/8)\lambda_0\cos\phi\sin\theta+(1/4)\lambda_0\sin\phi\sin\theta)} + I_4 e^{jk(+(3/8)\lambda_0\cos\phi\sin\theta+(1/4)\lambda_0\sin\phi\sin\theta)} \\ &+ I_5 e^{jk(-(3/8)\lambda_0\cos\phi\sin\theta-(1/4)\lambda_0\sin\phi\sin\theta)} + I_6 e^{jk(-(1/8)\lambda_0\cos\phi\sin\theta-(1/4)\lambda_0\sin\phi\sin\theta)} \\ &+ I_7 e^{jk(+(1/8)\lambda_0\cos\phi\sin\theta-(1/4)\lambda_0\sin\phi\sin\theta)} + I_8 e^{jk(+(3/8)\lambda_0\cos\phi\sin\theta-(1/4)\lambda_0\sin\phi\sin\theta)}) \end{aligned} \tag{14.43}$$

Using the general expression for each current excitation from Eq. (14.31), we can rewrite the array factor in the xy-plane ($\theta = \pi/2$) from Eq. (14.43):

$$\begin{aligned} AF = \tfrac{1}{8}(&a^0 e^{jk(-(3/8)\lambda_0\cos\phi+(1/4)\lambda_0\sin\phi)} + a^1 e^{j\Phi_0} e^{jk(-(1/8)\lambda_0\cos\phi+(1/4)\lambda_0\sin\phi)} \\ &+ a^2 e^{j2\Phi_0} e^{jk(+(1/8)\lambda_0\cos\phi+(1/4)\lambda_0\sin\phi)} + a^3 e^{j3\Phi_0} e^{jk(+(3/8)\lambda_0\cos\phi+(1/4)\lambda_0\sin\phi)} \\ &- a^0 e^{jk(-(3/8)\lambda_0\cos\phi-(1/4)\lambda_0\sin\phi)} - a^1 e^{j\Phi_0} e^{jk(-(1/8)\lambda_0\cos\phi-(1/4)\lambda_0\sin\phi)} \\ &- a^2 e^{j2\Phi_0} e^{jk(+(1/8)\lambda_0\cos\phi-(1/4)\lambda_0\sin\phi)} - a^3 e^{j3\Phi_0} e^{jk(+(3/8)\lambda_0\cos\phi-(1/4)\lambda_0\sin\phi)}) \end{aligned} \tag{14.44}$$

Here, Φ_0 takes on the form of $\Phi_{0,\mathrm{TL}}$ from Eq. (14.38) for the TL-fed array and $\Phi_{0,\mathrm{MTM}}$ from Eq. (14.40) for the MTM-fed array.

Using the same parameters for the TL and MTM lines as before, the radiation patterns for the grounded TL-fed array and the MTM-fed array are shown in Figures 14.25 and 14.26 for the three representative frequencies of 4, 5, and 6 GHz.

As in the case of the ungrounded arrays, it can be observed that the main beam of the MTM-fed array squints less with frequency from the broadside location at 5 GHz. The complete scan angle characteristics for the two grounded arrays are shown in Figure 14.27.

The most striking observation from Figure 14.27 is that the scanning range of both the arrays is reduced by the addition of the ground plane. The TL-fed array exhibits a scanning range of $-54^\circ < \gamma_{\mathrm{sc}} < 65^\circ$ within a bandwidth of 3.9 GHz, from 3.5 to 7.4 GHz. The MTM-fed array exhibits a scanning range of $-52^\circ < \gamma_{\mathrm{sc}} < 72^\circ$ within a bandwidth of 8.5 GHz, from 3.1 to 11.6 GHz. The reduced scanning range of both the arrays can be attributed to the fact that at the edge of the scanning range the maximum of the array factor shifts to another lobe other than the main one. Thus when one of the

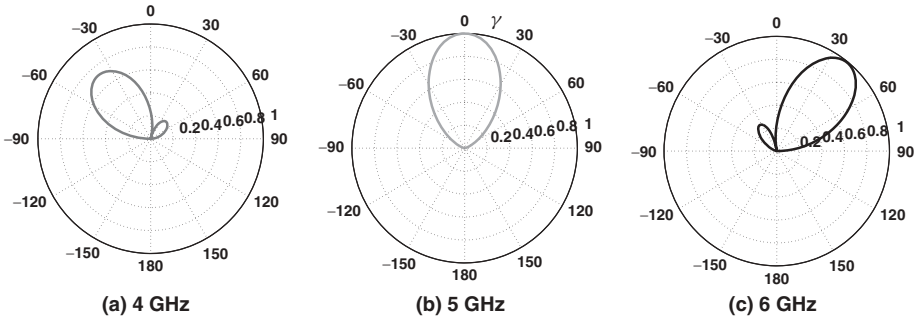


Figure 14.25 Normalized ideal array factor patterns in the xy-plane from Eqs. (14.38) and (14.44) for the grounded TL-fed dipole array at (a) 4, (b) 5, and (c) 6 GHz.

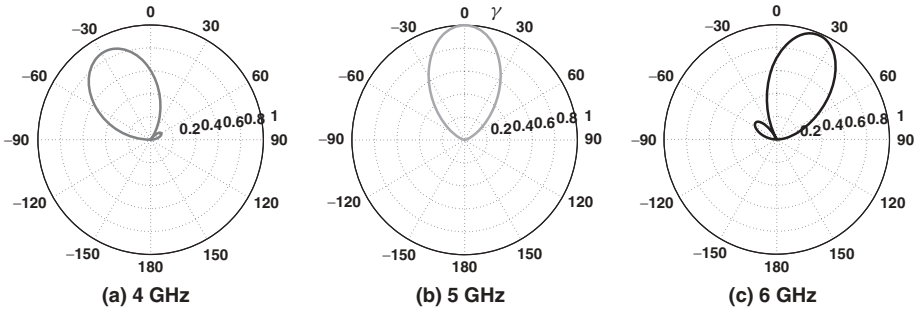


Figure 14.26 Normalized ideal array factor patterns in the xy-plane from Eqs. (14.40) and (14.44) for the grounded MTM-fed dipole array at (a) 4, (b) 5, and (c) 6 GHz.

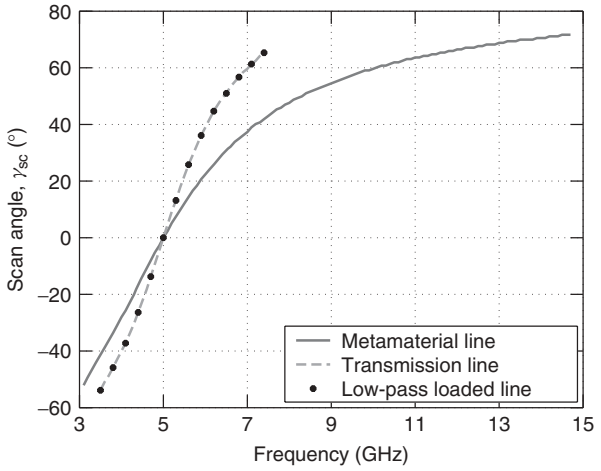


Figure 14.27 Scan angle performance from broadside of a grounded series-fed linear array with $d_E = \lambda_0/4$ using a metamaterial, a transmission-line, and a lowpass loaded line feed network.

backlobes becomes larger than the main lobe, this determines the limit on the scanning range. Another observation is that although the scanning range of the grounded arrays is decreased compared to the ungrounded arrays, the scanning bandwidth actually increases by 1.23 GHz for the TL-fed array and by 1.16 GHz for the MTM-fed array. As with the case of the ungrounded arrays, the scanning characteristics of the grounded arrays remained the same regardless of the value of the amplitude factor, a.

Figure 14.28 shows the normalized amplitude of the maximum value of the array factor as a function of frequency for the grounded TL-fed array for three representative cases of the amplitude factor a.

Around the design frequency of $f_0 = 5$ GHz, although the scanning bandwidth of the grounded TL-fed array from Figure 14.27 is 3.9 GHz, Figure 14.28 indicates that the frequency range over which the peak values of the main lobe remain at a maximum value is reduced to 1.3 GHz, from 4.7 to 6 GHz. This reduces the usefulness of the grounded TL-fed array, due to the limited frequency range that it can be operated in. At higher frequencies above 6 GHz the amplitude of the maximum array factor varies rapidly with frequency, and in the range of 9.5–11.5 GHz it drops to very low values. In the regions where the maximum array factor drops to such low values, the power is redistributed to other lobes in the pattern, thus reducing the amplitude of the main lobe. This is generally an undesirable condition, as a lot of the power in the array is lost to these other parasitic sidelobes or backlobes.

Figure 14.29 shows the amplitude of the normalized maximum value of the array factor as a function of frequency for the grounded MTM-fed array for the same three representative cases of the amplitude factor a.

By comparing Figures 14.28 and 14.29, it can be observed that the amplitude of the main beam of the grounded MTM-fed array fluctuates very little compared to the grounded TL-fed array. Thus the grounded MTM-fed array can effectively be operated within the frequency range of 4–14 GHz, without large variations in the amount of power delivered to the main lobe. This is because the peak of the main beam remains mainly within the visible region over this entire frequency range.

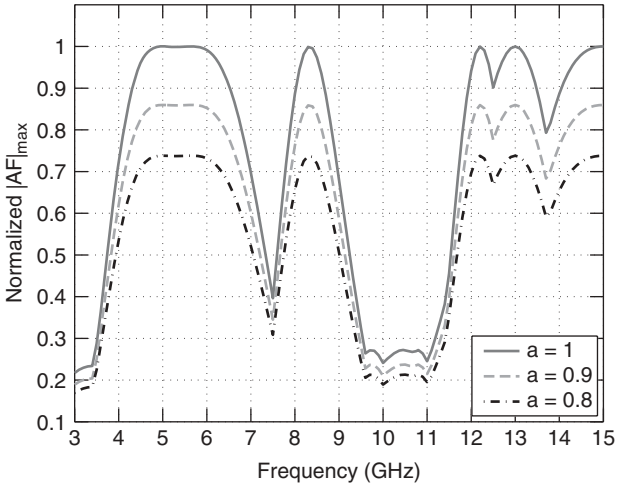


Figure 14.28 Normalized maximum array factor as a function of frequency for the grounded TL-fed array for three representative amplitude factor values: $a = 1$, $a = 0.9$, and $a = 0.8$.

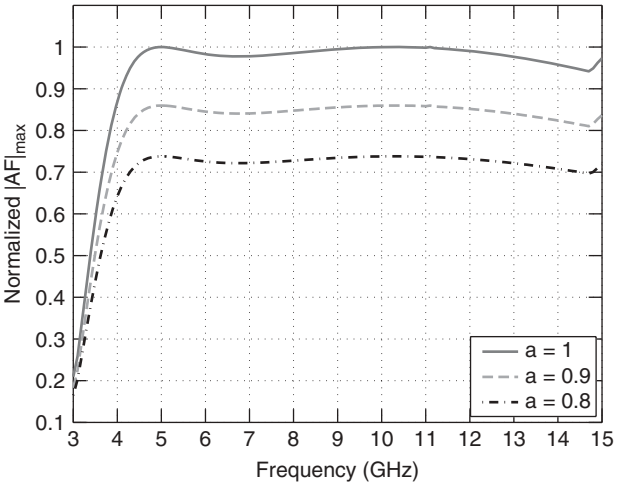


Figure 14.29 Normalized maximum array factor as a function of frequency for the grounded MTM-fed array for three representative amplitude factor values: $a = 1$, $a = 0.9$, and $a = 0.8$.

As in the case of the ungrounded arrays, by enforcing an amplitude taper along either of the arrays, this reduces the value of the maximum array factor; however, the general shape of the curves is maintained the same as for the uniformly excited array.

14.5.5 Physical Realizations of the Proposed Structures

With sight of a physical realization of the arrays discussed in the previous section, three types of feed lines were considered as potential candidates for feeding the antenna elements in a series-fed linear array: a 0° metamaterial line, a -360° meandered TL, and a lowpass loaded TL that also incurs -360°. Each of the feed lines was designed in coplanar strip (CPS) technology [41, 42], at a design frequency of $f_0 = 5$ GHz ($\lambda_0 = 60$ mm) on a Rogers RT5880 substrate with $\varepsilon_r = 2.2$ and dielectric height $h = 20$ mils. The 0° MTM line used to feed the MTM array is shown in Figure 14.30a and consisted of a single MTM unit cell of length $d_{\mathrm{TL,H}} = 3$ mm, and characteristic impedance $Z_0 = 170\ \Omega$, followed by a negative phase-compensating section of TL of length $d_{\mathrm{TL}} = 12$ mm. The loading element values used were $C_0 = 0.13$ pF for the series capacitors and $L_0 = 2.44$ nH for the shunt inductors. The -360° meandered TL used to feed the TL array is shown in Figure 14.30b and had a total length of $\lambda_g = 52.44$ mm and a characteristic impedance $Z_0 = 170\ \Omega$. The -360° LPL feed line used to feed the LPL array is shown in Figure 14.30c and consisted of five unit cells of length $d_{\mathrm{LPL}} = 3$ mm, and characteristic impedance $Z_0 = 170\ \Omega$, each incurring a phase of -72°. The loading element values used were $L_1 = 4.53$ nH for the series inductors and $C_1 = 0.156$ pF for the shunt capacitors.

The three feed lines were designed and simulated in Agilent's Momentum planar structure simulator at a center frequency of $f_0 = 5$ GHz, using interdigitated capacitors and meandered inductors to realize the loading element values. Both dielectric and conductor losses were taken into account by specifying a loss tangent of 0.0009 for the dielectric substrate and a thickness of 17 μm for the copper layer with a conductivity of 5.813×10^7 S/m.

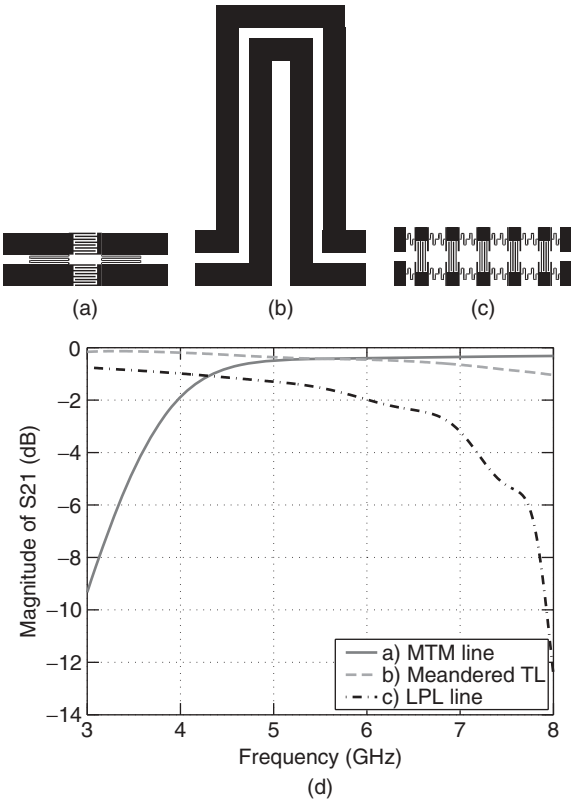


Figure 14.30 Proposed structures for (a) the CPS 0° MTM feed line, (b) the CPS -360° meandered TL, and (c) the CPS -360° lowpass loaded line. (d) Agilent Momentum simulated S_{21} magnitude responses for each of the feed lines shown in (a)–(c).

Figure 14.30d shows the magnitude of S_{21} for the three feed lines. At $f_0 = 5$ GHz the magnitude of S_{21} is -0.49 dB for the MTM line, -0.36 dB for the meandered TL, and -1.30 dB for the LPL line. It can be observed that the MTM line exhibits the expected highpass behavior with low insertion loss at higher frequencies, while the LPL line exhibits the expected lowpass behavior with significantly higher insertion loss at higher frequencies. The meandered TL exhibits similar insertion loss compared to the MTM line between 5 and 6 GHz; however, above 6 GHz its performance begins to degrade, exhibiting higher levels of insertion loss.

Although the insertion loss performance of the meandered TL was quite constant along the frequency band of interest, it was determined that the geometrical constraint of spacing the antennas in the array $\lambda_0/4$ apart prohibited the use of this type of feed line because of increased coupling between the feed line and the antenna elements in a uniplanar array design. It was therefore decided to implement the four-element printed series-fed linear dipole array using only the physically realizable MTM and LPL feed lines.

Both the four-element MTM dipole array and the four-element LPL dipole array were designed and simulated in Agilent's Momentum planar structure simulator at a center

frequency of $f_0 = 5$ GHz. The lengths of the dipole elements were initially designed to be approximately $\lambda_g/2$; however, these were subsequently increased in order to account for the mutual coupling between the antenna elements. Tables 14.1 and 14.2 summarize the geometrical details of the MTM and LPL antenna array components, respectively.

A prototype of the MTM dipole array was built and is shown in Figure 14.31. Figure 14.32 is a close-up view of the 0° nonradiating MTM line, offering a better view of the printed series capacitors and shunt inductors that are placed at the center of the line. Figure 14.33 shows the LPL dipole array structure, with Figure 14.34 showing the details of the lowpass unit cells, namely, the printed series inductors and shunt capacitors. A backing ground plane (10 cm × 5 cm) was added at a distance of $\lambda_0/4 = 15$ mm below both of the antennas in order to create unidirectional radiation patterns. In addition, both antennas were fed directly by a coaxial cable, whose center conductor was connected to the top CPS conductor and the outer sheath was connected to the bottom CPS conductor, as shown in Figures 14.31 and 14.33. In order to prevent parasitic currents from flowing on the outer surface of the coaxial feed line, caused by the inherent imbalance between the balanced CPS line and the unbalanced coaxial line, ferrite beads were placed around the coaxial line, acting as current chokes. The ferrite beads can also be seen in Figures 14.31 and 14.33.

TABLE 14.1 Geometrical Details of the MTM Printed Dipole Array Components (L = Length, W = Width, S = Separation)

CPS line	L = 15 mm	W = 2 mm	S = 0.9 mm
Dipole	L = 32.9 mm	W = 2 mm	S = 0.9 mm
Capacitor fingers	L = 1.884 mm	W = 0.133 mm	S = 0.1333 mm
Inductor	L = 27.8 mm	W = 0.1 mm	S = 0.1 mm

TABLE 14.2 Geometrical Details of the LPL Printed Dipole Array Components (L = Length, W = Width, S = Separation)

CPS line	L = 15 mm	W = 2 mm	S = 0.9 mm
Dipole	L = 34.9 mm	W = 2 mm	S = 0.9 mm
Capacitor fingers	L = 2.48 mm	W = 0.1 mm	S = 0.1 mm
Inductor	L = 5 mm	W = 0.1 mm	S = 0.1 mm

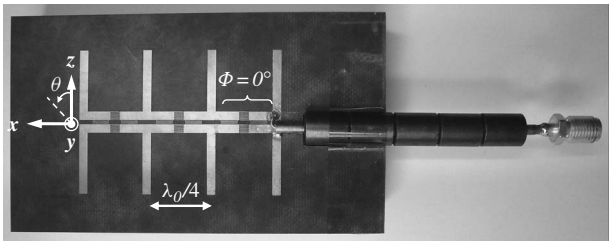


Figure 14.31 Photograph of the fabricated metamaterial-fed printed dipole array at 5 GHz, including the coaxial feedline surrounded by ferrite beads.

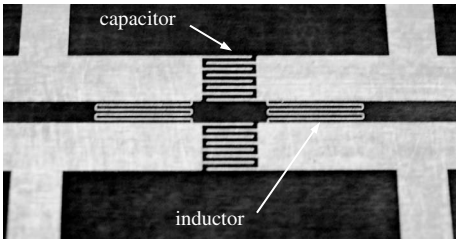


Figure 14.32 Close-up view of the printed components of the 0° MTM line, showing the series interdigitated capacitors and shunt meandered inductors.

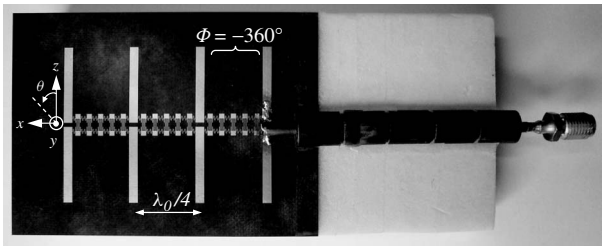


Figure 14.33 Photograph of the fabricated lowpass loaded TL-fed printed dipole array at 5 GHz, including the coaxial feedline surrounded by ferrite beads.

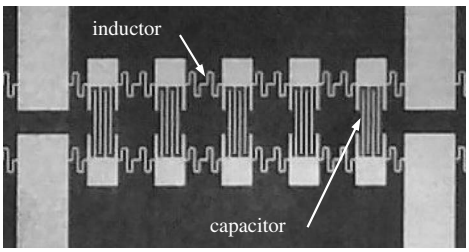


Figure 14.34 Close-up view of the printed components of the −360° LPL line, showing the series meandered inductors and the shunt interdigitated capacitors.

14.5.6 Simulation and Experimental Results

Figure 14.35 shows the measured and simulated scan angles versus frequency of the emerging beam from the broadside direction for the MTM and LPL arrays. Although both the arrays were designed to exhibit broadside radiation at 5 GHz, the main beam of the fabricated prototypes passed through broadside at 5.1 GHz and 5.2 GHz for the LPL and the MTM arrays, respectively. This shift in frequency can be attributed to the imperfections in the manufacturing of the printed array components. Specifically, it was observed that there was slight overetching of the copper strips that form the interdigitated capacitors and meandered inductors. As a result, the values of the loading elements of the MTL and LPL feed lines were slightly altered, leading to an increase in the phase incurred by the lines and thus resulting in a broadside beam at a slightly higher frequency.

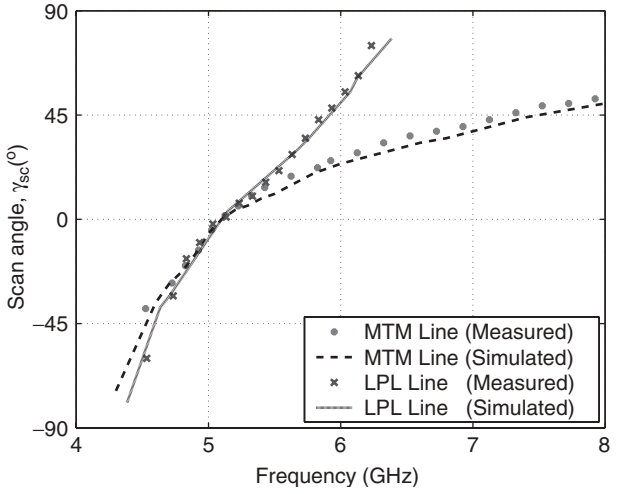


Figure 14.35 Measured and simulated scan angles from broadside of the MTM-fed and LPL-fed arrays.

In order to provide an equal comparison of the phase versus frequency characteristics of the two arrays, the appropriate data was offset in frequency such that the beam passed through broadside at exactly 5.1 GHz for all four of the cases shown in Figure 14.35.

It can be observed that the measured data matches the simulated data quite well. It is of interest to note that the measured and simulated scan angle plots of Figure 14.35 follow the same general patterns as the theoretical ideal scan angle plots of Figure 14.27 for both the MTM and the LPL arrays. It can be seen that at around 5 GHz the performance of the LPL array is similar to the MTM array; however, below 4.8 GHz and above 5.3 GHz their performances begin to diverge, with the MTM array exhibiting far more broadband scan angle characteristics at frequencies above 6 GHz. The LPL array exhibits a measured scan angle bandwidth of 2 GHz, from 4.33 to 6.33 GHz, while the MTM array has a measured scan angle bandwidth of 5.45 GHz, from 4.25 to 9.7 GHz. This corresponds to an increase of 173% in the scan angle bandwidth over the LPL array.

Figure 14.36 shows the measured normalized copolarization radiation patterns in the xy-plane ($\theta = \pi/2$) for the three selected frequencies of 4.9, 5.1, and 5.3 GHz for the MTM array.

It can be observed that at 4.9 GHz the main beam is directed off broadside at an angle of approximately -13°, while at 5.1 GHz the main beam is at broadside. When the frequency is further increased to 5.3 GHz, the main beam shifts to approximately -10° from broadside. In all three cases, the main beam is clearly distinguishable. Thus it has been demonstrated that the fabricated MTM dipole array exhibits nonlinear beam scanning characteristics with frequency about the broadside direction, which follows directly from the fact that the phase response of the 0° MTM feed lines has a nonlinear phase characteristic given by Eq. (14.28).

As a corollary, one can also observe that if the MTM array is operated at a higher frequency region, say, around 7 GHz, the linearity of the scan angle versus frequency response can be exploited to obtain an off-broadside main beam at approximately $+45^\circ$

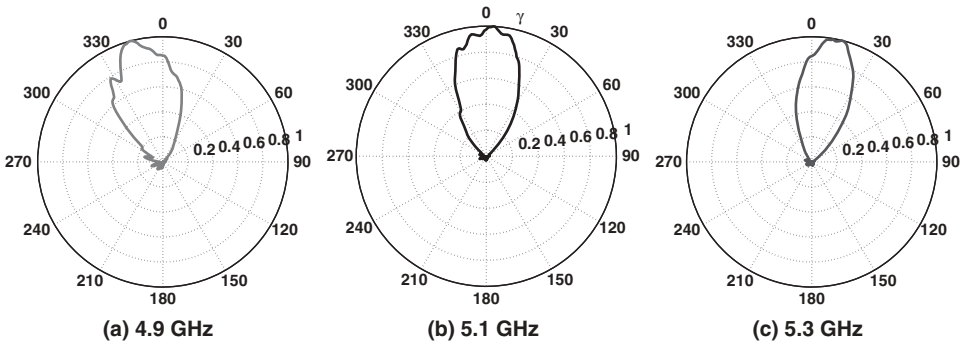


Figure 14.36 Measured normalized copolarization patterns in the xy-plane for the printed MTM-fed array at (a) 4.9, (b) 5.1, and (c) 5.3 GHz.

(see Figure 14.35), that will not experience significant beam squinting with a change in frequency. This can be deduced from the fact that the gradient of the scan angle with frequency in this region is quite small. Indeed, by inspection of Eq. (14.28) it can be observed that the effect of the second term $(1/\omega\sqrt{L_0\ C_0})$, which is inversely proportional to ω, diminishes as the frequency is increased. This results in a linear phase response for the MTM line, with a constant phase offset at dc, which is characteristic of a quasi-true-time delay line. This in turn results in a linear scan angle characteristic with frequency that has a small gradient, as outlined in Section 14.5.3.

In order to highlight the fact that the MTM-fed array does not exhibit significant beam squinting at higher frequencies, Figure 14.37 shows the measured normalized copolarization radiation patterns in the xy-plane ($\theta = \pi/2$) for 7.1, 7.3, and 7.5 GHz. It can be observed that in all three cases the main beam is directed off broadside at an angle of approximately $+45^\circ$. When the frequency is changed from 7.1 to 7.5 GHz the scan angle changes a mere 6°, from 43° to 49°, thus verifying that within this frequency range the MTM-fed array exhibits only modest beam squinting. It should be noted that even at the highest operating frequency of 7.5 GHz, the MTM feed lines still operate in the nonradiating (slow-wave) region (see Section 14.3.1). For the particular loading element values chosen for this design, the theoretical maximum operating frequency before propagation

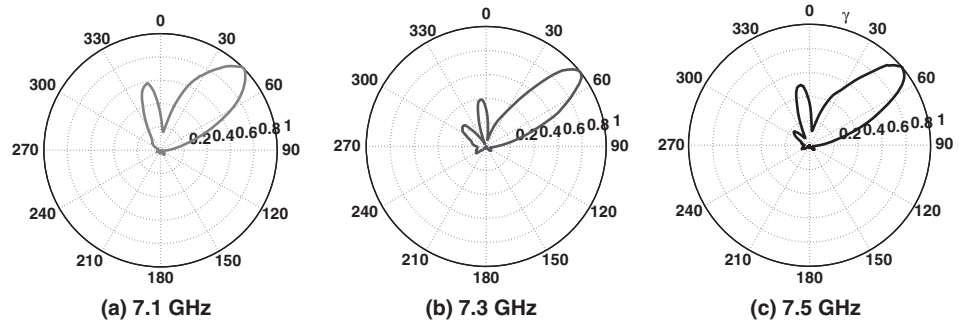


Figure 14.37 Measured normalized copolarization patterns in the xy-plane for the printed MTM-fed array at (a) 7.1, (b) 7.3, and (c) 7.5 GHz (observe the almost fixed beam direction).

begins to occur within the radiation cone is 7.76 GHz. Above this frequency the MTM lines could act as leaky-wave structures if their electrical length is made sufficiently long.

Figure 14.38 shows the measured normalized copolarization radiation patterns for the LPL-fed array in the xy-plane ($\theta = \pi/2$) for 7.1, 7.3, and 7.5 GHz. For the LPL-fed array, the main lobe is no longer captured in the visible region of the array pattern above 6.33 GHz; therefore a scan angle does not exist above this frequency (see Figure 14.35). This is verified by the plots of Figure 14.38, where the pattern changes shape rapidly with frequency, and in some cases (e.g., at 7.1 GHz) there is no clearly distinguishable main lobe, which is generally an undesirable condition. Therefore, at high frequencies the location and, in fact, the existence of a main beam cannot be accurately predicted for the LPL-fed array, which is a distinct disadvantage compared to the MTM-fed array.

Figures 14.39 and 14.40 show the measured and simulated S_{11} magnitude responses for the MTM-fed and LPL-fed arrays, respectively.

For the MTM-fed array, the measured results match quite closely with the simulated results, with a measured -10 dB $|S_{11}|$ bandwidth of 610 MHz centered around 5 GHz

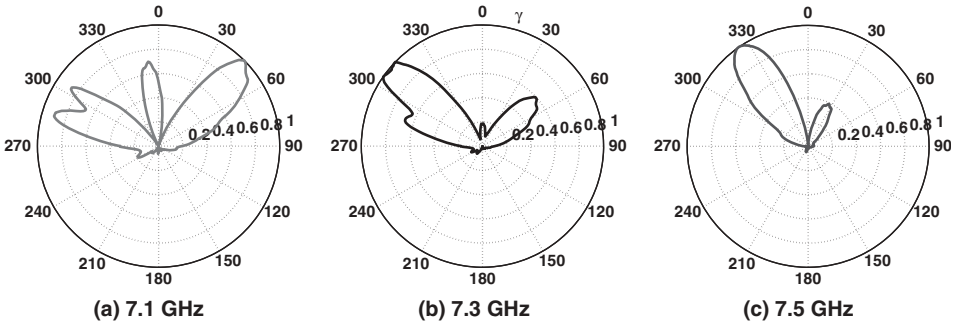


Figure 14.38 Measured normalized copolarization patterns in the xy-plane for the printed LPL-fed array at (a) 7.1, (b) 7.3, and (c) 7.5 GHz (observe the nonuniform beam direction and shape).

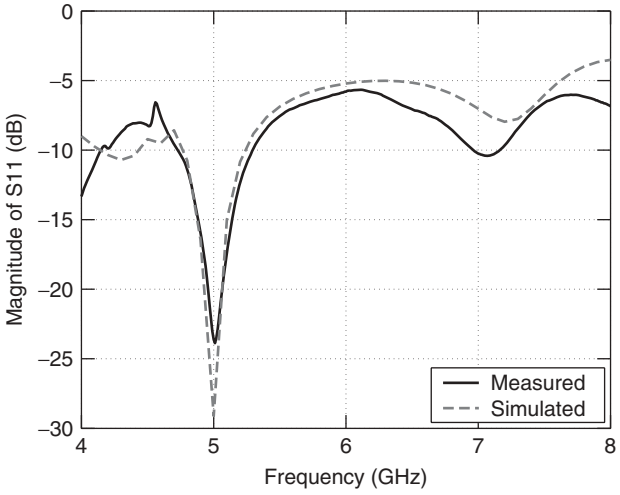


Figure 14.39 Measured and simulated S_{11} magnitude responses for the MTM-fed array.

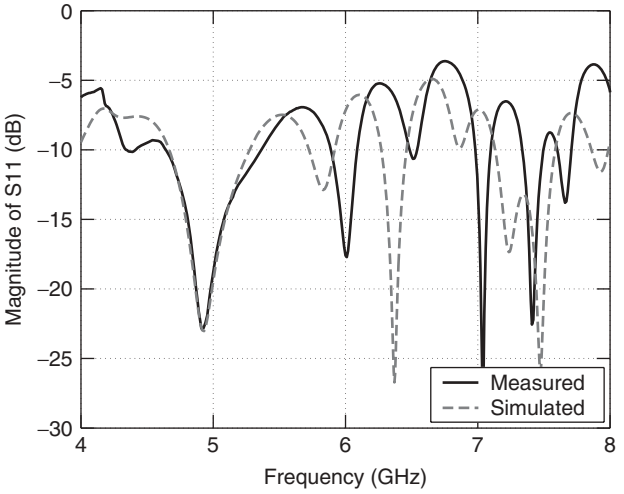


Figure 14.40 Measured and simulated S_{11} magnitude responses for the LPL-fed array.

and a minimum $|S_{11}|$ value of -23.9 dB at 5.01 GHz. The MTM array also exhibits reasonably good S_{11} performance around 7 GHz with a minimum $|S_{11}|$ value of -10.5 dB at 7.06 GHz, indicating that the array can be optimally designed to operate around this frequency, where the scan angle characteristic is quite broadband (see Figure 14.35). The corresponding measured results for the LPL-fed array match the simulated results reasonably well around 5 GHz; however, at higher frequencies the two diverge. The measured -10 dB $|S_{11}|$ bandwidth for the LPL-fed array is 720 MHz centered around 5 GHz, with a minimum $|S_{11}|$ value of -22.9 dB at 4.92 GHz. Finally, even though the S_{11} performance of the TL-fed array is good around 7 GHz, with a minimum $|S_{11}|$ value of -29.2 dB at 7.04 GHz, the resonance is very narrowband with a bandwidth of only 120 MHz. More importantly though, as has already been shown in Figure 14.38, the main beam in the frequency range around 7 GHz has a nonuniform direction and shape, which severely limits the usefulness of the LPL-fed array.

In summary, MTM NRI-TL feed networks can be used in series-fed linear arrays to effectively replace one-wavelength long conventional meandered transmission lines, thus achieving compact, broadside radiators, whose beam squints much less with frequency. In addition, the MTM feed networks can be used to create linear arrays whose main beam remains virtually fixed at a positive angle from broadside as the frequency is varied.

14.6 A NRI-TL METAMATERIAL LEAKY-WAVE ANTENNA WITH REDUCED BEAM SQUINTING

The preceding work employing MTM phase-shifting lines to reduce beam squinting in series-fed arrays prompted the investigation of the reduced beam squinting effect in other types of antennas as well. By operating the MTM lines in the radiation region of the dispersion curve, a leaky-wave antenna (LWA) can be created, which exhibits reduced beam squinting characteristics without the need for resonant elements [43, 44]. The fabricated LWA is shown in Figure 14.41 and consists of 20 fully printed MTM

unit cells realized in coplanar strip (CPS) technology, which are cascaded in series to create a leaky traveling-wave antenna. By appropriately choosing the size and the loading elements of each MTM unit cell, the antenna can be designed to radiate at a predefined location in space, which in this case was chosen to be 45° in the forward direction from broadside at 5 GHz. Most importantly though, in order to minimize the amount of beam squinting that the array experiences as the frequency is changed, the phase and group velocities of the MTM line should each be as close as possible to the speed of light. In order for the structure to be leaky, the phase velocity must approach the speed of light from above, and in order to satisfy causality, the group velocity should approach the speed of light from below. Both these criteria are met by operating the antenna in the upper RH radiating band, where the difference between the phase and group velocities is minimized, therefore minimizing the amount of beam squinting that the array experiences. This is consistent with the results presented in Figure 14.35 for the MTM-fed dipole array.

The fabricated MTM LWA prototype of Figure 14.41 exhibited a measured return loss bandwidth below −10 dB of 0.91 GHz (18.2%). Three representative measured radiation copolarization patterns in the xz-plane are shown in Figure 14.42 for the selected

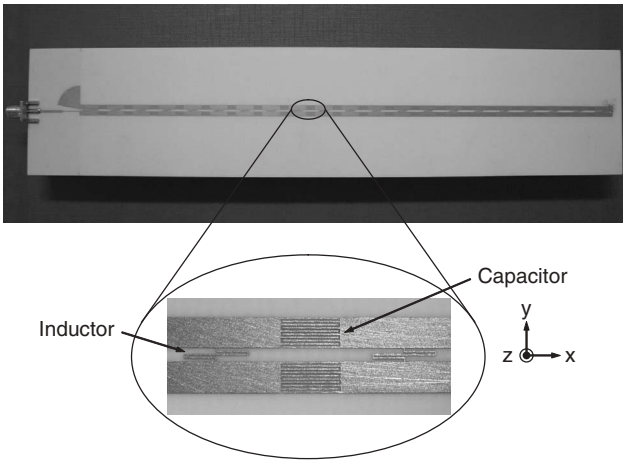


Figure 14.41 Photograph of the fabricated MTM LWA in CPS technology at 5 GHz.

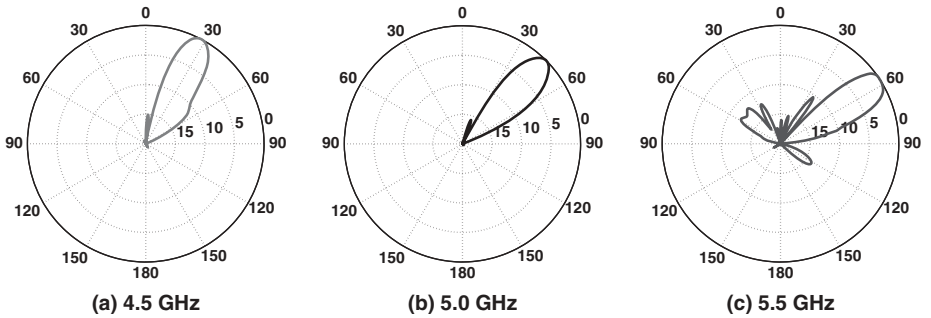


Figure 14.42 Measured normalized copolarization patterns (in dB) in the xz-plane for the printed MTM LWA at (a) 4.5, (b) 5.0, and (c) 5.5 GHz.

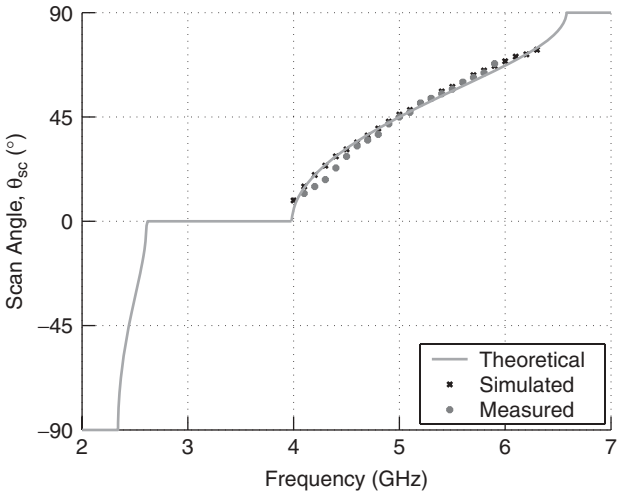


Figure 14.43 Scan angle versus frequency for the 20-element MTM LWA. The theoretical results were obtained by using $\theta_{sc} = \sin^{-1}(\beta/k_0)$, and the simulated results were obtained from Agilent's Momentum.

frequencies of 4.5, 5, and 5.5 GHz. It can be observed that within the 1-GHz bandwidth from 4.5 to 5.5 GHz, the main beam experiences modest beam squinting and remains close to the desired pointing direction of 45° from broadside. Figure 14.43 shows the complete theoretical, simulated, and measured scanning characteristics of the MTM LWA. The simulated LWA exhibited a scanning bandwidth of 2.3 GHz in which the scan angle varied a total of 65°, resulting in an average beam squint of 0.028°/MHz, and the measured LWA exhibited a scanning bandwidth of 1.8 GHz in which the scan angle varied a total of 56°, resulting in an average beam squint of 0.031°/MHz.

14.7 AN ELECTRICALLY SMALL NRI-TL METAMATERIAL RING ANTENNA WITH VERTICAL POLARIZATION

The MTM 0° phase-shifting lines were also used to create a compact ($\lambda_0/10 \times \lambda_0/10$ footprint) and low profile ($\lambda_0/20$ height) MTM ring antenna at 3.1 GHz. The design was first proposed in Ref. 45 and is based on wrapping around the two ends of a 0° phase-shifting line to construct a small-footprint ring antenna shown in Figure 14.44. The antenna consists of four 0° NRI-TL unit cells implemented in microstrip technology, arranged in a ring structure. The microstrip TLs were implemented by etching copper strips on a thin layer of Rogers RT5880 substrate (0.38 mm thick) suspended in air 5 mm above a small ground plane. The chip lumped-element series capacitors C_0 and shunt inductors L_0 were soldered to the top surface of the antenna, where the inductors were connected to the ground plane through vertical vias with a diameter of 0.51 mm. The series chip capacitors used were size 0402 Panasonic ECD-G0E capacitors with a self-resonant frequency (SRF) greater than 6 GHz and had a value of $C_0 = 0.8$ pF. The shunt chip inductors used were size 0302 Coilcraft CS inductors with an SRF greater than 9 GHz and had a value of $L_0 = 7.2$ nH. The vertical vias act as the main radiating

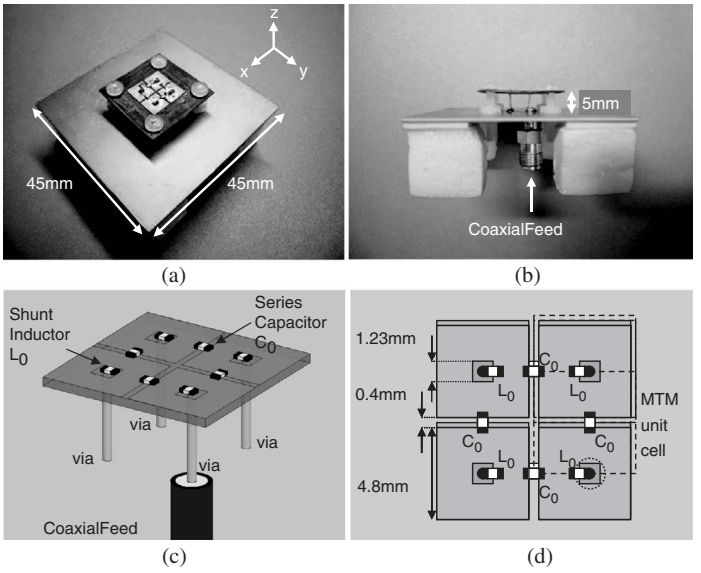


Figure 14.44 Electrically small NRI-TL metamaterial antenna: (a) Perspective view, (b) Side view (c) 3D diagram, and (d) Top view diagram.

elements of the antenna with an associated radiation resistance R_r and inductance L_{via}. The inductance of each of the vias was incorporated into the lumped-inductor value L_0 and the capacitance of the gap in the microstrip TL, C_{gap}, was also incorporated into the lumped-capacitance value C_0. The antenna was fed from the bottom of one of the vias by an SMA coaxial connector as shown in Figure 14.44b.

Each MTM unit cell was designed to incur an insertion phase of 0°, which allows the inductive posts (vias) within each of the unit cells to be fed in phase. Thus the MTM ring acts as an extremely compact feed network for the inductive posts, eliminating the need for long, bulky one-wavelength-long feed lines. In addition, the effective top loading of the vias by the microstrip line and the inherent monopole folding effect created by base feeding one of the vias enables the electrically small antenna to be matched to 50 Ω. The vias act as the main radiating elements, and since the currents in each of the vias are equal in magnitude and phase, this produces a field pattern with a vertical electric field polarization, similar to a short monopole antenna over a small ground plane. It should also be noted that not only is the antenna itself electrically small but also the corresponding truncated ground pane, which measures $0.465\lambda_0 \times 0.465\lambda_0$.

The antenna can be analyzed by decomposing the antenna current into a superposition of an even-mode (I_e) and an odd-mode (I_0). Since the antenna is symmetrical, analysis of a single unit cell will reveal the general characteristics of the entire antenna. Beginning with the even-mode excitation, we apply two equal voltages ($V/2$) to each end of the MTM unit cell of as shown in Figure 14.45a. Due to the symmetry of the unit cell, this effectively places an open circuit (O.C.) at the center of the circuit, thus producing two identical decoupled circuits. If we then replace the short TL sections with their equivalent series inductance $L' = Ld$ and capacitance $C' = cd$, we obtain the circuit of Figure 14.45b. No current will pass through the series resonator formed by $L'/2$ and $2C_0$ because it is open circuited on one end, therefore these components can be removed

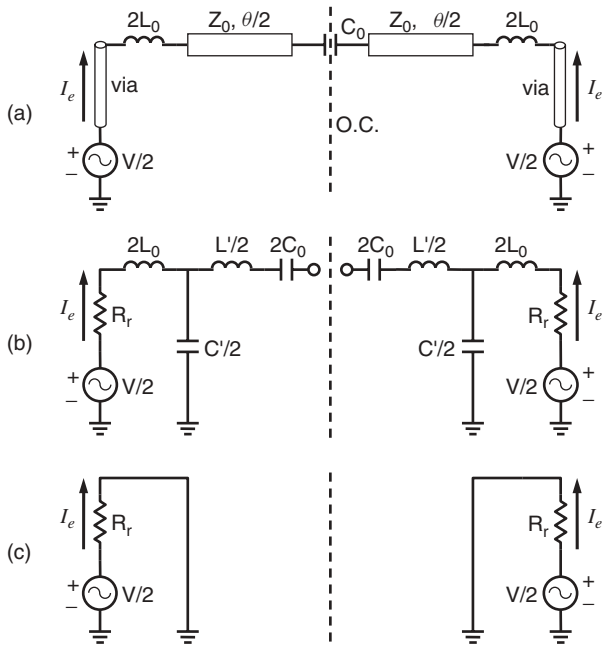


Figure 14.45 Even-mode equivalent circuits for a single unit cell of the MTM ring antenna: (a) TL-based MTM unit cell, (b) lumped-element equivalent circuit, and (c) simplified lumped-element circuit at resonance.

from the circuit. At resonance, the series resonator formed by $2L_0$ and $C'/2$ will become a short circuit, therefore resulting in the final simplified circuit of Figure 14.45c. Since no other impedances are present in the circuit other than R_r, maximum current will be delivered to the radiation resistance, given by $I_e = (V/2)/R_r$.

If we now consider the odd-mode excitation, we apply an equal and opposite voltage to each end of the MTM unit cell as shown in Figure 14.46a. Due to the symmetry of the unit cell, this effectively places a short circuit (S.C.) at the center of the circuit, thus producing two identical decoupled circuits with equal and opposite excitations. If we then replace the short TL sections with their equivalent series inductance L' and capacitance C', we obtain the circuit of Figure 14.46b. At resonance, the series resonator formed by $L'/2$ and $2C_0$ will become a short circuit, thus shorting out $C'/2$ and therefore resulting in the final simplified circuit of Figure 14.46c. It can now be observed that the odd-mode current will be given by $I_0 = (V/2)/(R_r + j\omega\ 2\ L_0)$. It can therefore be seen that the $2L_0$ inductor plays a pivotal role in adjusting the odd-mode current, and larger values of $2L_0$ will act to reduce the odd-mode current. In fact, for any value of $2L_0$ the even-mode current I_e will always be greater than the odd-mode current I_0.

If we consider some representative values, a short monopole with a uniform current distribution will have a radiation resistance of $R_r = 160\pi^2(h/\lambda)^2$, therefore for $h/\lambda = 1/20$, $R_r = 4\ \Omega$. Assuming $V = 1$ V, for a typical value of $L_0 = 20$ nH, at 3 GHz the magnitude of the odd-mode current is equal to $|I_0| = 0.66$ mA. The corresponding even-mode current is equal to $|I_e| = 127$ mA, which is approximately 150 times greater than $|I_0|$. Thus it has been shown that the magnitude of the odd-mode current is negligible

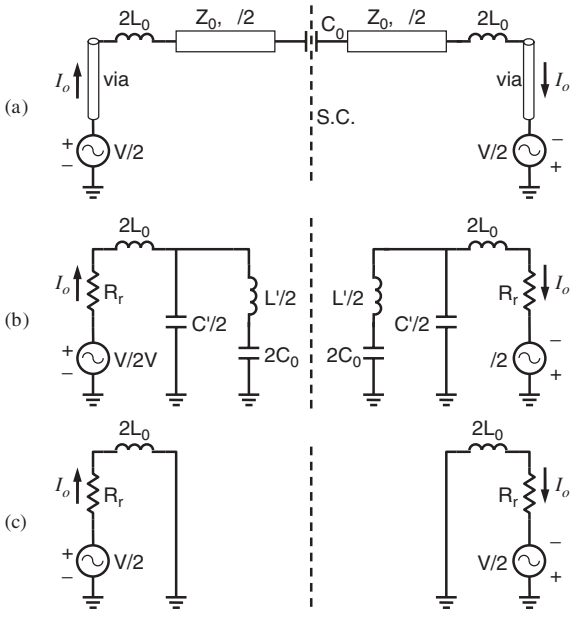


Figure 14.46 Odd-mode equivalent circuits for a single unit cell of the MTM ring antenna: (a) TL-based MTM unit cell from, (b) lumped-element equivalent circuit, and (c) simplified lumped-element circuit at resonance.

compared to the even-mode current. The majority of the current flowing on the antenna is even-mode current, therefore enabling the metamaterial structure to act as a good radiator.

Figure 14.47 shows the measured versus the simulated return loss obtained from Ansoft HFSS for the antenna. It can be observed that the antenna is well matched around

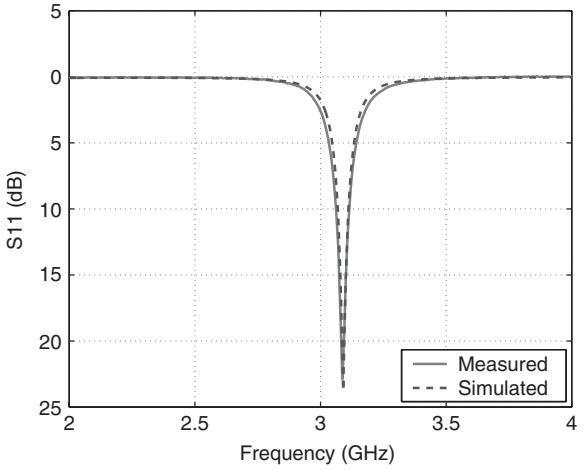


Figure 14.47 Measured and simulated return loss for the MTM ring antenna.

3.1 GHz, with a simulated return loss bandwidth (−10 dB) of 42 MHz, and a measured return loss bandwidth of 53 MHz, which corresponds to approximately 1.7%.

Figures 14.48 and 14.49 show the measured versus the simulated E-plane and H-plane patterns at 3.1 GHz. It can be observed that the antenna exhibits a radiation pattern with a vertical linear electric field polarization, similar to that of a short monopole on a small ground plane. The simulated directivity and gain from HFSS were 1.31 (1.18 dBi) and 0.95 (−0.24 dB), respectively, while the measured directivity and gain obtained using the gain comparison method were 1.78 (2.50 dBi) and 1.23 (0.90 dBi) respectively, resulting in a simulated efficiency of 72.3% and a measured efficiency of 69.1%. It should be noted that a similar implementation of this antenna using a two-unit-cell architecture is

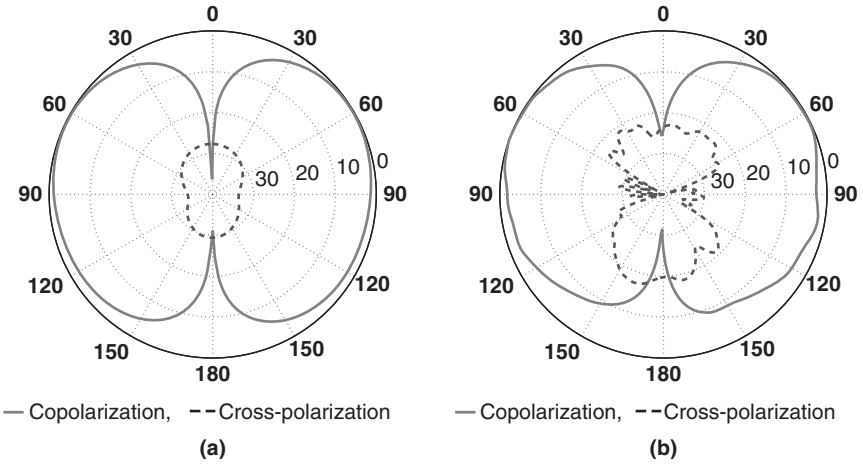


Figure 14.48 E-plane (xz-plane) patterns for the printed MTM ring antenna at 3.1 GHz: (a) simulated and (b) measured solid line: co-polarization, dashed line: cross-polarization.

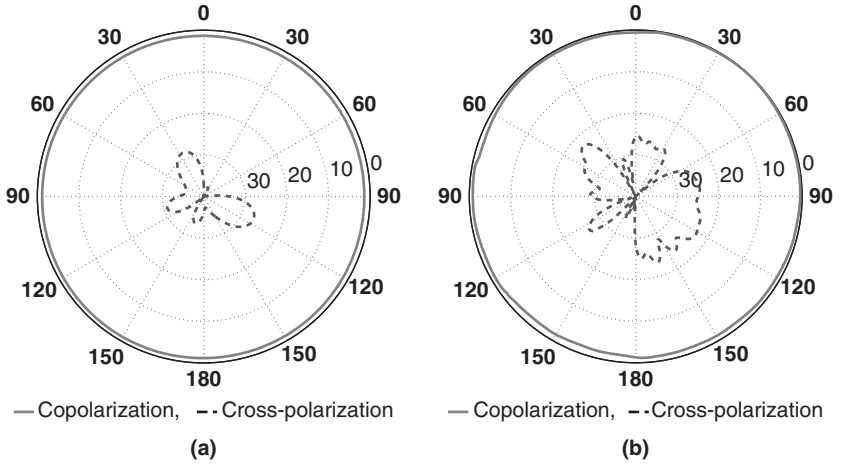


Figure 14.49 H-plane (xy-plane) patterns for the printed MTM ring antenna at 3.1 GHz: (a) simulated and (b) measured, solid line: co-polarization, dashed line: cross-polarization.

described in Ref. 46 using chip lumped-element components, and a version using fully printed lumped elements is described in Ref. 47.

These types of small, low profile antennas have sufficient bandwidth (~50 MHz) and good radiation efficiency (~70%) and are thus well suited for emerging high capacity multiple antenna (MIMO) wireless-access schemes. Other examples that have utilized metamaterial concepts to reduce the physical size or in general improve the performance of antennas can be found in Ref. 48–54.

14.8 BACKWARD/FORWARD LEAKY-WAVE ANTENNAS RADIATING IN THEIR FUNDAMENTAL SPATIAL HARMONIC

14.8.1 One-Dimensional NRI-TL Metamaterial Leaky-Wave Antennas

The transmission-line (TL) approach to synthesizing NRI metamaterials has led to the development of new kinds of leaky-wave antennas (LWAs). In Section 14.5.3, a LWA operating in the upper RH radiation region was shown to exhibit reduced beam squinting characteristics for a beam radiated in the forward direction. In this section it will be shown that by simply changing the circuit parameters of the dual TL model, a fast-wave structure can be designed that operates in the lower LH radiation region and therefore supports a fundamental spatial harmonic that radiates toward the backward direction (i.e., toward the feed) [9, 10].

A coplanar waveguide (CPW) implementation of such a backward leaky-wave antenna is shown in Figure 14.50. This can be recognized as the dual of the CPS structure presented in Section 14.5.3. The gaps in the CPW feedline serve as the series capacitors of the dual TL model, while the narrow lines connecting the center conductor to the coplanar ground planes serve as the shunt inductors (shorted stubs). The capacitive gaps are the radiating elements in this leaky-wave antenna and excite a radiating transverse magnetic (TM) wave. Due to the antiparallel currents flowing on each pair of the narrow

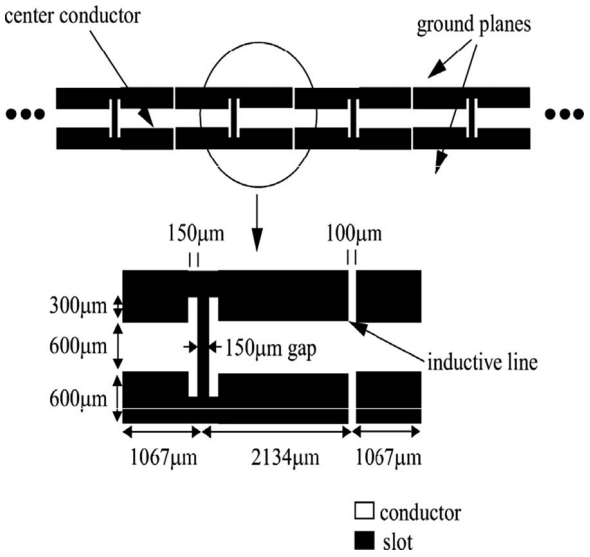


Figure 14.50 Backward leaky-wave antenna based on the dual TL model at 15 GHz.

inductive lines, they remain nonradiating. Simulated and experimental results for this bidirectional leaky-wave antenna were reported in Refs. 9 and 10. The corresponding unidirectional design is simply the leaky-wave antenna described in Ref. 10 backed by a long metallic trough as shown in Figure 14.51. Since the LWA's transverse dimension is electrically small, the backing trough can be narrow (below resonance). The trough used is a quarter-wavelength in height and width and covers the entire length of the antenna on the conductor side of the substrate. It acts as a waveguide below cut-off and recovers the back radiation, resulting in unidirectional far-field patterns.

A complementary forward unidirectional leaky-wave antenna was also reported in Ref. 55. The periodic structure of Ref. 55 also operates in the fundamental spatial harmonic and hence can be thought of as a metamaterial with a positive phase velocity. Here, we describe experimental results for the unidirectional design proposed in Ref. 9. A frequency shift of 3%, or 400 MHz, was observed in the experiments compared to the method-of-moments simulations of the LWA using Agilent's Advanced Design System (ADS). As a result, the experimental unidirectional radiation patterns were obtained at 14.6 GHz, while the simulation patterns were calculated at 15 GHz. The corresponding E-plane and H-plane patterns are shown in Figures 14.52 and 14.53 respectively. A gain improvement of 2.8 dB was observed for the unidirectional design over the bidirectional design, indicating that effectively all of the back radiation is recovered with the trough.

A similar leaky-wave antenna implemented in a microstrip configuration was also reported in Ref. 56. In that implementation varactor diodes were utilized to steer the beam from backward to forward directions in a design with a closed stopband (see Eq. (14.4).

14.8.2 Two-Dimensional NRI-TL Metamaterial Leaky-Wave Antennas

The above ideas can be extended to two-dimensional leaky-wave antennas in order to generate pencil rather than fan beams [57, 58]. The unit cell for the 2D implementations is shown in Figure 14.1. Of particular interest is the case in which the corresponding stopband between the fundamental backward band and the first forward band is closed. With

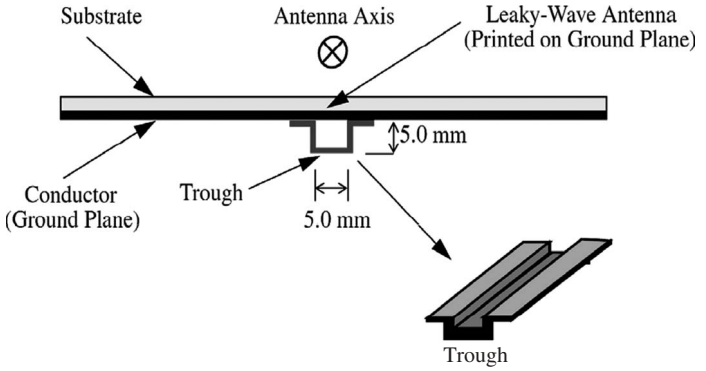


Figure 14.51 Unidirectional backward leaky-wave antenna design at 15 GHz.

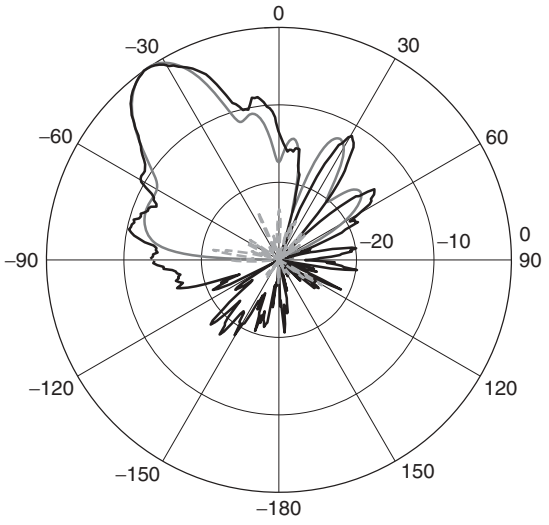


Figure 14.52 *E*-Plane pattern for the unidirectional leaky-wave antenna at 15 GHz: (—) experimental copolarization, (- ·-) experimental cross-polarization, and (___) simulated copolarization using Agilent's ADS (interior curve).

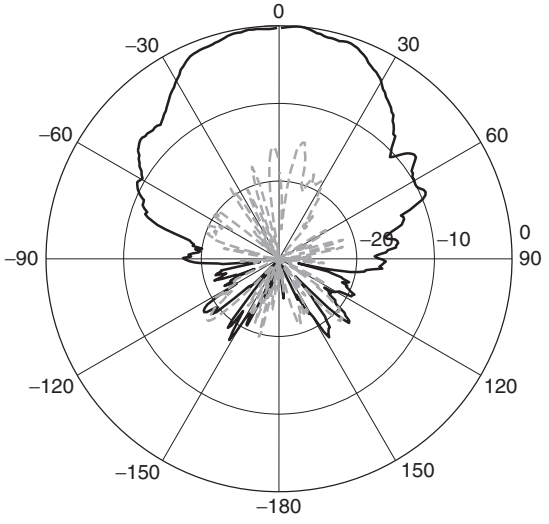


Figure 14.53 *H*-Plane pattern for the unidirectional leaky-wave antenna at 15 GHz: (—) experimental copolarization, and (-·-) experimental cross-polarization.

reference to Eq. (14.4), the corresponding closed stopband condition in two dimensions becomes

$$\frac{Z}{\sqrt{2}} = \sqrt{\frac{L_0}{C_0}} \tag{14.45}$$

where Z denotes the characteristic impedance of the host microstrip lines and L_0, and C_0 denote the loading elements in the dual configuration shown in Figure 14.1. In order to analyze and design these structures, a FDTD technique has been developed that is capable of predicting the phase and leaky attenuation constants from a single unit cell [59]. The corresponding dispersion diagram for a typical microstrip-based 2D structure is shown in Figure 14.54.

Since the stopband is closed, the radiated beam can be frequency scanned from the backward to the forward endfire direction [56]. The corresponding backward and forward radiation patterns are shown in Figure 14.55 [59].

The patterns in Figure 14.55 imply that the peak gain of the 2D antenna attains its maximum value at broadside, that is, when the effective phase constant vanishes. This is compatible with the behavior of the attenuation constant shown in Figure 14.54 which attains a minimum value around the frequency at which the phase constant becomes zero (2.72 GHz). At this frequency, all the unit cells resonate in unison, thus illuminating the corresponding NRI-TL aperture in phase. This uniform illumination property is illustrated in Figure 14.56, based on the 2D NRI-TL microstrip structure described in Ref. 11 and depicted in Figure 14.57.

Before finishing this section, it is perhaps interesting to point out that under the closed-stopband condition of Eq. (14.45) and exactly at the frequency where the phase constant is zero, the corresponding group velocity is also zero. This can be verified by visually inspecting the full-wave dispersion shown in Figure 14.54, where indeed the dispersion curve for β flattens out around 2.72 GHz, thus signifying a zero group

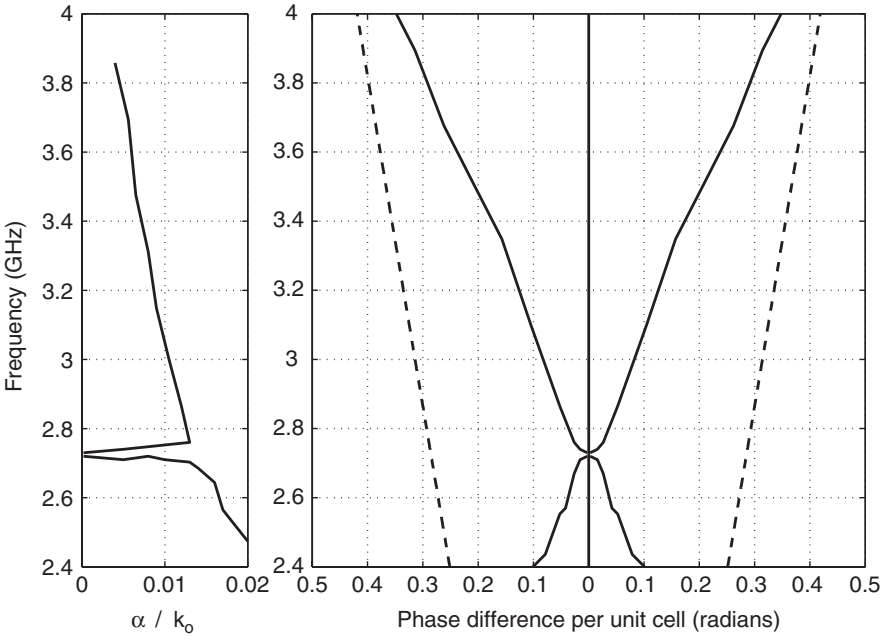
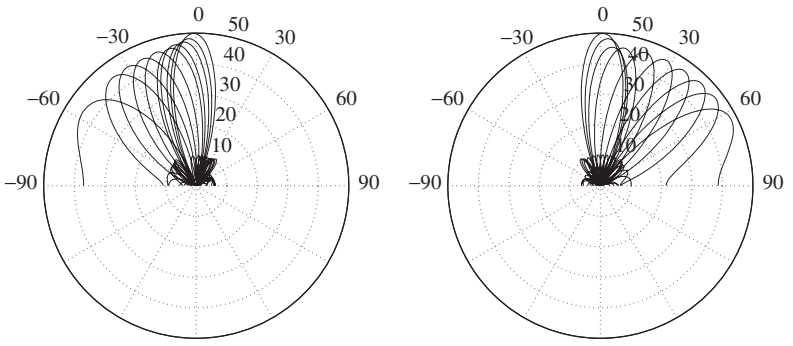


Figure 14.54 Complex propagation constant for the fast (radiating) modes supported by a typical 2D NRI-TL microstrip grid. The loading elements are $L_0 = 10.0$ nH and $C_0 = 1.0$ pF (closed-stopband case). The attenuation constant (left) and the phase constant (right) are plotted as a function of frequency. The dashed line represents the light line [59].



(a) Backward-radiation patterns (b) Forward-radiation patterns

Figure 14.55 (a) Backward and (b) forward radiation patterns (E-plane) as the frequency sweeps from 2.44 to 3.83 GHz for a 2D NRI-TL metamaterial surface with a closed stopband 59.

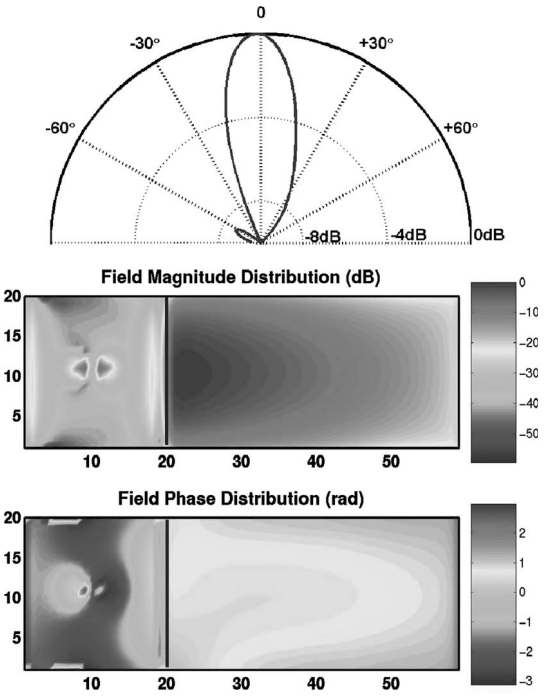


Figure 14.56 Simulated broadside E-plane pattern ($D = 12$ dB) and illumination of the NRI-TL surface (closed stopband).

velocity. This vanishing group velocity phenomenon is a consequence of the fact that the structure is radiating; the reader is referred to Ref. 59 for the complete discussion. This is in contrast to the ideal (nonradiating) case of Ref. 14, in which the closed-stopband condition of Eq. (14.45) results in a nonvanishing group velocity at the point where $\beta = 0$.

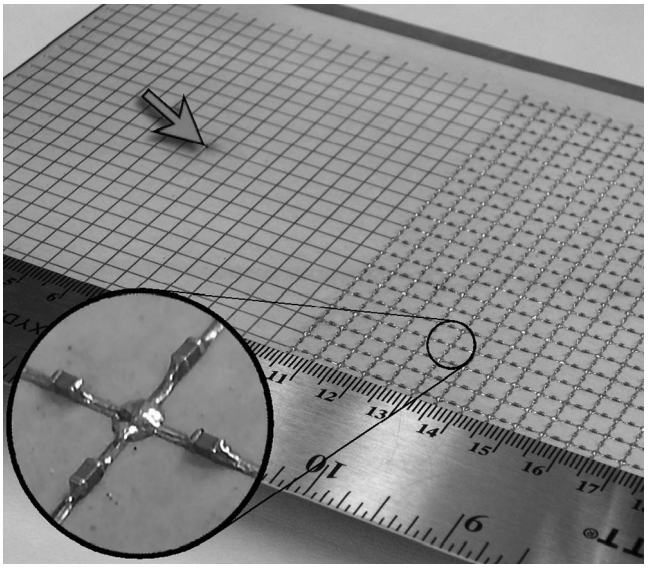


Figure 14.57 2D NRI-TL microstrip grid. The inset photograph shows the physical realization of the 2D NRI-TL unit cell of Figure 14.1.

14.9 A BROADBAND WILKINSON BALUN USING MICROSTRIP NRI-TL METAMATERIAL LINES

Baluns are particularly useful for feeding two-wire antennas, where balanced currents on each branch are necessary to maintain symmetrical radiation patterns with a given polarization. Two-wire antennas have input ports that are closely spaced; therefore their feeding structures should be chosen to accommodate for this requirement. This precludes the use of certain balun designs, whose output ports are spaced far apart [60].

Printed balun designs can generally be classified as distributed-TL or lumped-element type. Distributed-TL designs are inherently narrowband due to the frequency dependence of the TLs used. These can be made broadband; however, they usually require transmission lines that are at least several wavelengths long and are therefore not very compact [61]. Lumped-element designs, albeit compact, can suffer from a relatively narrowband differential output phase resulting from the inherent mismatch between the phase response of the lowpass/highpass output lines that they employ [62].

The proposed MTM balun, shown in Figures 14.58 and 14.59, consists of a Wilkinson power divider, followed by a $+90^\circ$ MTM phase shifting line along the top branch and a -90° MTM phase-shifting line along the bottom branch [63]. The design of the balun was based on the MTM unit cell shown in Figure 14.4 and was carried out by first selecting appropriate values for the loading elements of the $+90^\circ$ MTM line to produce a $+90^\circ$ phase shift at the design frequency f_0, while maintaining a short overall length. Then the pertinent parameters for the -90° MTM line were calculated such that the shape of the phase responses of the $+90^\circ$ and -90° MTM lines matched, thus maintaining a 180° phase difference over a large bandwidth.

In order to match the phase response of the -90° MTM line with that of the $+90^\circ$ MTM line, and therefore create a broadband differential output phase, the slopes of their

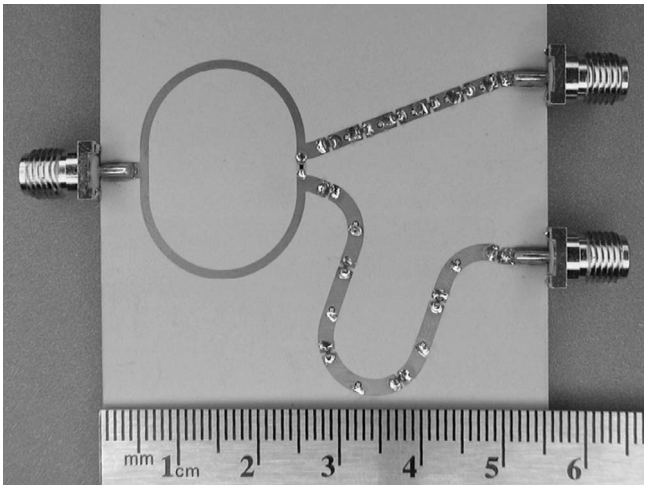


Figure 14.58 Photograph of the fabricated metamaterial balun (From Ref. 63.)

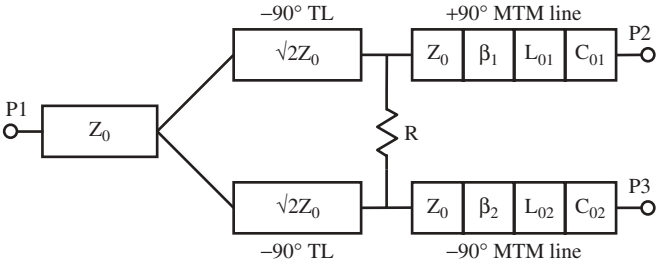


Figure 14.59 Architecture of the metamaterial balun.

phase characteristics must be equal at the design frequency, thus satisfying

$$\left.\frac{d\Phi_{\mathrm{MTM}+}}{d\omega}\right|_{\omega_0} = \left.\frac{d\Phi_{\mathrm{MTM}-}}{d\omega}\right|_{\omega_0} \tag{14.46}$$

Moreover, to ensure that the MTM phase-shifting lines do not radiate, each unit cell must be operated in the region outside the light cone on the Brillouin diagram. Thus the $+90^\circ$ MTM phase-shifting lines must be operated in the NRI backward-wave region, while simultaneously ensuring that the propagation constant of the line exceeds that of free space, resulting in a slow-wave structure with a positive insertion phase. Correspondingly, the -90° MTM phase-shifting lines must be operated in the positive refractive index (PRI) forward-wave region, while simultaneously ensuring that the propagation constant of the line also exceeds that of free space, resulting in a slow-wave structure with a negative insertion phase (see Section 14.3.1).

The MTM Wilkinson balun was implemented in microstrip technology on a Rogers RO3003 substrate with $\varepsilon_r = 3$ and height $h = 0.762$ mm at a design frequency of $f_0 = 1.5$ GHz. A five-stage design was chosen for the $+90^\circ$ MTM phase-shifting line as well

as the -90° MTM phase-shifting line. The experimental results were compared with the simulated results obtained using Agilent's ADS.

Figure 14.60 shows the measured versus the simulated return loss magnitude response for port 1, showing good agreement between the two, indicating that the device is well matched, especially around $f_0 = 1.5$ GHz. The measured and simulated return losses for ports 2 and 3 exhibit similar responses. Figure 14.61 shows excellent isolation for the device, as well as equal power split between the two output ports.

Figure 14.62 shows the measured versus the simulated phase responses of the two balun branches. Note the broadband phase difference between the two response curves in Figure 14.62. The experimental results agree very closely with the simulated results. It can be observed that the phase of S_{21} is exactly equal to $+90^\circ$ at $f_0 = 1.5$ GHz,

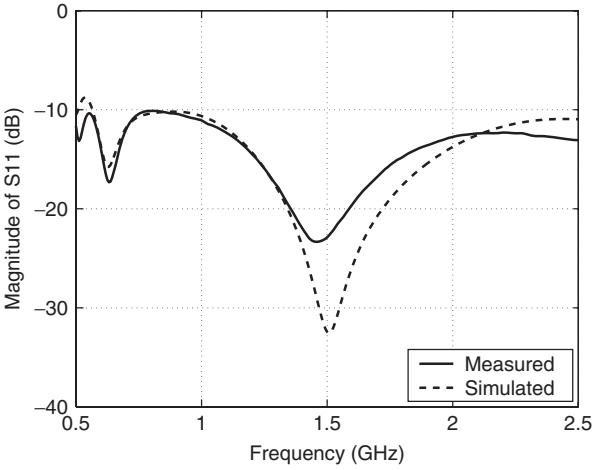


Figure 14.60 Measured and simulated return loss magnitude responses for port 1 of the MTM balun.

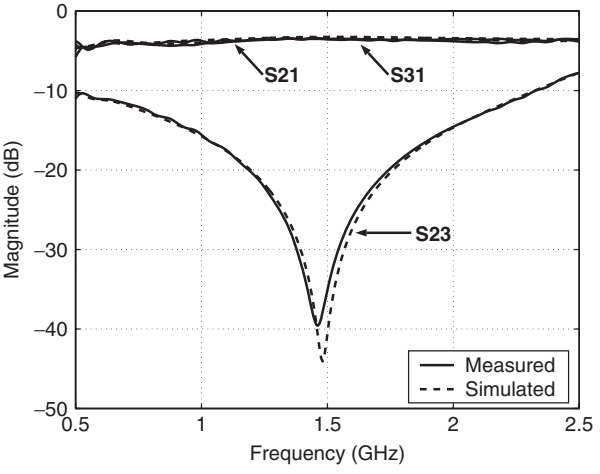


Figure 14.61 Measured and simulated isolation (S_{23}) and through (S_{21} and S_{31}) magnitude responses of the MTM balun.

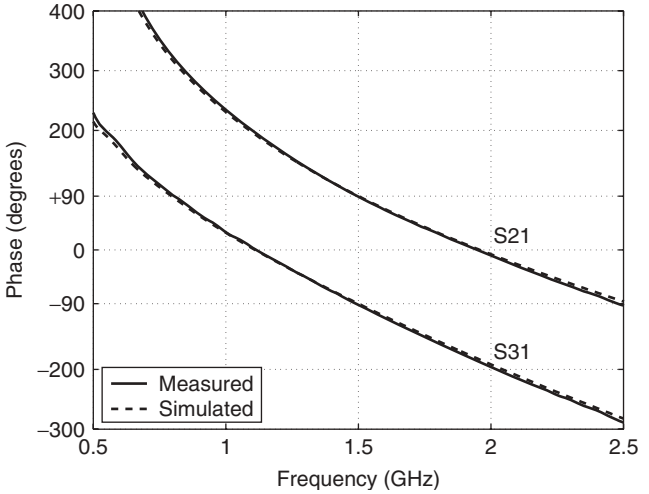


Figure 14.62 Measured and simulated phase responses of S_{21} ($+90^\circ$ MTM line) and S_{31} (-90° MTM line) of the MTM balun.

while the phase of S_{31} is exactly equal to -90° at $f_0 = 1.5$ GHz, and that the phase characteristics of the two branches are quite similar.

Figure 14.63 shows the measured and simulated differential output phase of the MTM balun, with excellent agreement between the two. It can be observed that the differential output phase remains flat for a large frequency band, which follows directly from the fact that the phase characteristics of the $+90^\circ$ and -90° lines correspond very closely. The flat differential output phase has a $180^\circ \pm 10^\circ$ bandwidth of 1.16 GHz, from 1.17 to 2.33 GHz. Since the device exhibits excellent return loss, and isolation and through characteristics over this frequency range, it can be concluded that the MTM balun can be used as a broadband single-ended to differential converter in the frequency range from 1.17 to 2.33 GHz.

For comparison, a distributed TL Wilkinson balun employing -270° and -90° TLs instead of the $+90^\circ$ and -90° MTM lines was also simulated, fabricated, and measured at $f_0 = 1.5$ GHz, and the differential output phase of the TL balun is also shown in Figure 14.63. It can be observed that the phase response of the TL balun is linear with frequency, with a slope equal to the difference between the phase slopes of the -270° and -90° TLs. Since the gradient of the resulting phase characteristic is quite steep, this renders the output differential phase response of the TL balun narrowband. Thus the TL balun exhibits a measured differential phase bandwidth of only 11%, from 1.42 to 1.58 GHz, compared to 77% exhibited by the MTM balun. In addition, the TL balun occupies an area of 33.5 cm^2 compared to 18.5 cm^2 for the MTM balun. Thus the MTM balun is more compact, occupying only 55% of the area that the conventional TL balun occupies. Furthermore, the MTM balun exhibits more than double the bandwidth compared to a lumped-element implementation using lowpass/highpass lines, which typically exhibits a bandwidth of 32% [62]. This can be attributed to the fact that the lowpass line has a linear phase response, while the response of the highpass line has a varying slope with frequency. Thus the shapes of the phase responses of the two lines do not match, resulting in a more narrowband differential output phase.

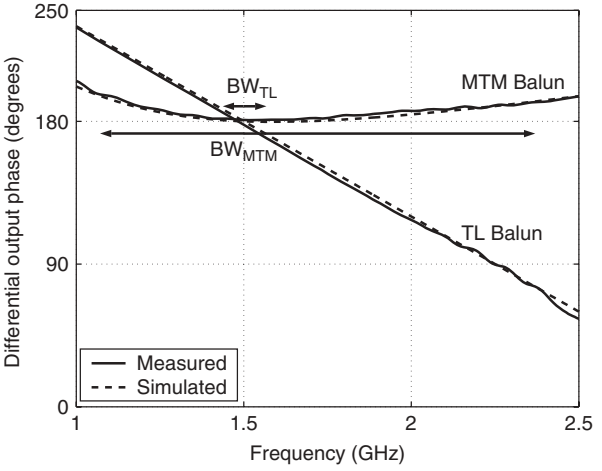


Figure 14.63 Measured and simulated differential phase comparison between the MTM balun and the TL balun.

14.10 A HIGH DIRECTIVITY BACKWARD NRI-TL/MICROSTRIP COUPLER

A useful device for antenna signal monitoring applications is a high directivity coupler that can be used to monitor reflections from, for example, cellular base-station antennas. Such reflectometer couplers should exhibit a low coupling level but a high isolation in order to detect weak reflection levels, without disturbing the main signal. Moreover, these couplers should exhibit a reasonable bandwidth and a low insertion loss and should be low cost. These requirements become possible by means of a coupled-line coupler that comprises a regular microstrip (MS) line edge coupled to a NRI-TL line (backward-wave line) [64–66]. The operation of such a coupler is shown schematically in Figure 14.64.

Such a MS/NRI coupler supports complex coupled modes at the tuning frequency, that is, the frequency at which the isolated propagation constants of the two lines are equal. These modes are characterized by codirectional phase but contradirection power flow on the two lines resulting in backward power coupling (see Figure 14.64). It has been demonstrated that such a coupler can be designed for arbitrary coupling levels [64, 65] and for high directivity [66].

The MS/NRI coupled-line coupler can be analyzed rigorously using the general coupled mode theory. It can be shown that the coupled modes with complex propagation constant γ satisfy the following eigenvalue equation [65]:

$$\begin{pmatrix} a-\gamma^2 & b \\ c & d-\gamma^2 \end{pmatrix}\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = 0$$
$$a = Z_1Y_1 + Z_mY_m \qquad b = Z_1Y_m + Z_mY_2 \tag{14.47}$$
$$c = Z_2Y_m + Z_mY_1 \qquad d = Z_2Y_2 + Z_mY_m$$

where Z_1 and Y_1 refer to the self-impedance and self-admittance per unit length, respectively, of line 1, and the terms with subscript 2 refer to line 2. Moreover, Z_m and Y_m

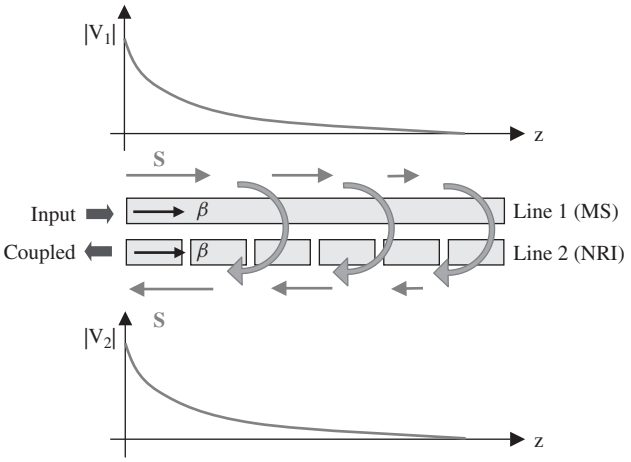


Figure 14.64 A coupled-line coupler comprising a microstrip (MS) line edge coupled to a negative refractive index (NRI) line. The arrows highlight the physical mechanism by which power continuously leaks from the microstrip line to the NRI line. Due to the excitation of complex modes, the voltage waves decay exponentially along the lines, thus leading to short coupling lengths.

describe the mutual coupling between the two lines. The various impedance/admittance terms in Eq. (14.47) are all purely imaginary as the system is assumed to be lossless. The decoupled eigenmodes of the individual lines are obtained by setting $Z_m = Y_m = 0$, and their isolated modal propagation constant β and characteristic impedance η can be expressed as $\beta_1^2 = -Z_1 Y_1$, $\eta_1^2 = Z_1/Y_1$, $\beta_2^2 = -Z_2 Y_2$ and $\eta_2^2 = Z_2/Y_2$.

Figure 14.65 shows a MS/NRI coupler with 4-mm long unit cells and 0.2-mm line spacing. The NRI-TL line consists of a periodically loaded MS line with 1.3-pF series

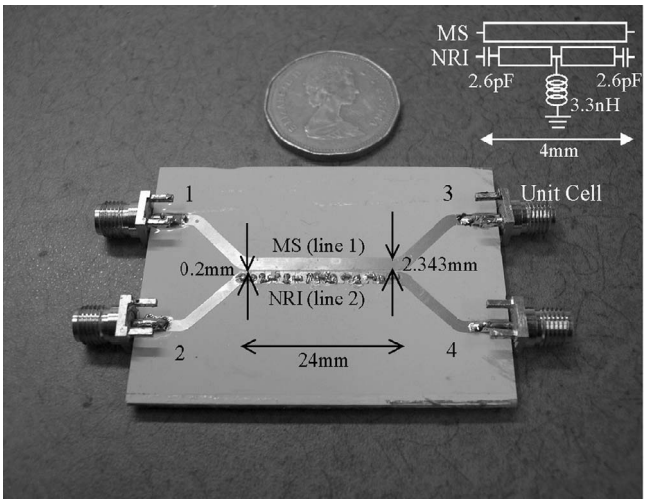


Figure 14.65 A 3- dB MS/NRI coupled-line coupler (six unit cells long) constructed on a 50-mil Rogers TMM4 ($\varepsilon_r = 4.5$) substrate.

capacitors and 3.3-nH shunt inductors to ground. The coupler was designed on a Rogers TMM4, 50-mil substrate of dielectric constant 4.5 and a linewidth of 2.34 mm was used. Agilent's ADS Microwave Circuit simulator was used to extract the various self- and mutual impedance/admittance terms appearing in Eq. (14.47) and the latter was solved numerically to obtain the dispersion relation shown in Figure 14.66.

It can be seen from Figure 14.66 that complex coupled modes are excited in the region where the two isolated propagation constants of the lines (β_1 and β_2) become equal. These modes decay and are characterized by oppositely directed Poynting vectors along the two lines such that the net power flow across every cross section remains constant. In this coupled-mode stopband, the power continuously leaks from the excited line to the coupled line, thus achieving arbitrary coupling levels, provided that the lines are made sufficiently long. This mechanism of power transfer has been illustrated in Figure 14.64.

The coupler in Figure 14.65 was made long enough to allow the power level in the input line (line 1) to decay to half its incident value (3-dB coupler). Using the negligible-loss and bounded-mode assumption, it was expected that the remaining power would appear in the coupled port (port 2 in Figure 14.64). Simulation and experimental results for the scattering parameters of this 3-dB coupler are shown in Figure 14.67. It can be seen that the measured power splits between the through and coupled ports at the levels of -3.03 and -3.68 dB, respectively, using this six-cell long coupler (24 mm) at the design frequency of 3 GHz. Moreover, the return loss and isolation are found to be below -20 dB whereas the directivity is about 20 dB.

The relative excitation of each eigenmode depends on the source and load terminations in each line. To this end, it is possible to derive simple closed-form expressions for the

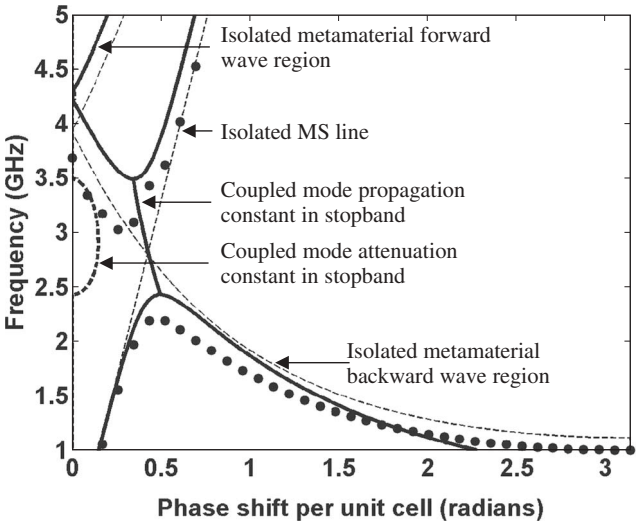


Figure 14.66 Coupled-mode dispersion diagram for the 3-dB MS/NRI coupled-line coupler of Figure 14.65: (bold line) coupled-mode theory; (thin line) isolated dispersion characteristic of individual lines; (dots) HFSS full-wave simulation. The attenuation constant in the contradirectional stopband formed around the frequency where the two dispersion diagrams of the isolated lines intersect is also shown.

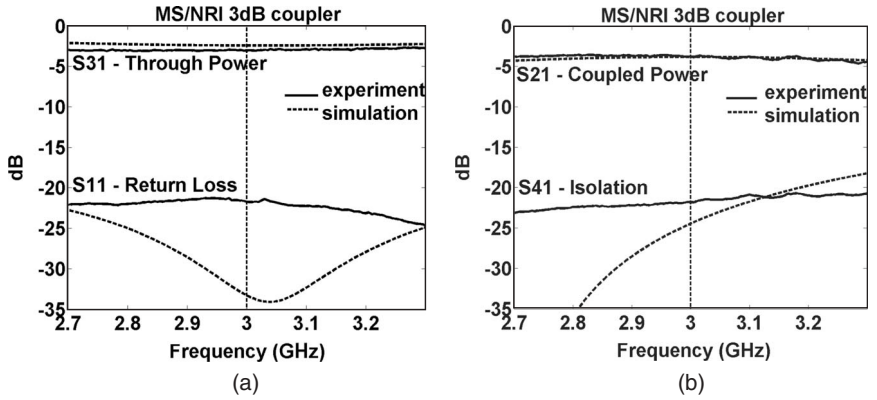


Figure 14.67 Simulation and experimental results for 3-dB MS/NRI coupled-line coupler: (a) return loss and through power and (b) coupled power and isolation.

scattering parameters of the coupler in the special case of operation at the center of the stopband [66].

$$S_{11} = j\frac{\Delta Z}{Z_0}\tanh(\alpha D) \tag{14.48}$$

$$S_{21} = j\tanh(\alpha D) \tag{14.49}$$

$$S_{31} = e^{-j\beta D}\sec h(\alpha D) \tag{14.50}$$

$$S_{41} = j\frac{\Delta Z}{Z_0}\sin(\beta D)\sec h(\alpha D) \tag{14.51}$$

In Eq. (14.48)–(14.51), D is the length of the coupler, ΔZ is the imaginary part of the mode impedance, and α is the attenuation constant of the mode, while β is the propagating component (phase constant). The excitation is assumed to be at port 1 and it is assumed that all ports are resistively terminated with the real part of the mode impedance Z_0 at the center of the stopband.

Eq. (14.49) demonstrates that the coupled power (S_{21}) can be made arbitrarily large by increasing the length of the coupler. This is accompanied by a corresponding decrease (depletion) in the power level at the through port (see Eq. (14.50)). An interesting observation is that when the coupler is made half as long as its guide wavelength, the isolation (S_{41}) vanishes, thus resulting in very high directivity (see Eq. (14.51)).

The preceding observation motivates the design of a high directivity reflectometer for antenna monitoring applications, as was described at the beginning of this section. A fully printed implementation of such a coupler was fabricated on a 50-mil Rogers TMM4 substrate ($\varepsilon_r = 4.5$ tan $\delta = 0.002$) with the equivalent unit cell shown in Figure 14.68. It was designed for operation at 2 GHz and comprises of four 9.76-mm long unit cells. The 1.6-pF capacitors were realized with 12 interdigital fingers (4.643 mm long and 0.295 mm wide) separated by 0.155-mm gaps. The 2.9-nH end inductors were made with 0.2-mm wide shorted stubs of length 2.5 mm. Shorted stubs 0.2 mm wide and 0.7 mm long were used as central inductors possessing half the inductance value of the end ones. The 30-Ω lines were realized using 5.185-mm wide microstrip lines.

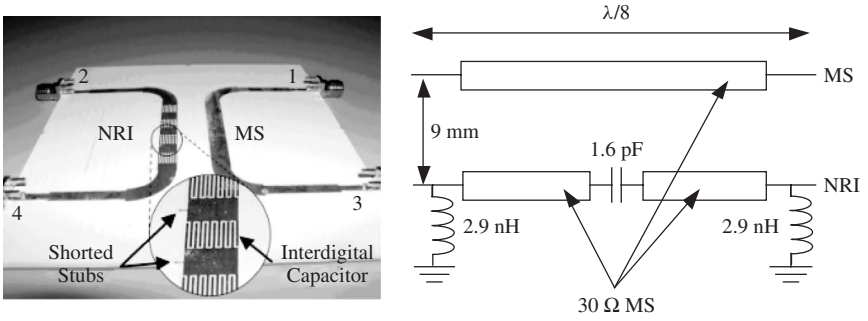


Figure 14.68 Fully printed MS/NRI highdirectivity coupler (reflectometer) and the equivalent circuit of one of its unit cells. (Adapted from Ref. 66).

Figure 14.69 shows theoretical, Agilent's Momentum full-wave simulation, and measured results for the MS/NRI coupler. The theoretical curve was obtained by solving the system of coupled-mode equations numerically using precalculated eigenvalues of the system and applying the port boundary conditions. The insertion loss is lower than 0.3 dB and the return loss is better than −25 dB. The measured coupling level at 2.04 GHz is −27 dB and the isolation is −72 dB yielding a directivity of 45 dB at the design frequency (with better than 30-dB directivity over a 40-MHz bandwidth). The coupler was designed to be λ/2 long at 2 GHz, and the results show that the isolation is indeed the best for this length, as predicted by theory. The frequency shift between theory and full-wave simulation arises from the inability to accurately represent the printed components with lumped-element models. The circles in Figure 14.69 correspond to the predictions of the approximate closed-form expressions in Eqs. (14.48)–(14.51).

In conclusion, the demonstrated performance of the printed MS/NRI coupled-line couplers makes them well suited for providing low cost solutions for applications requiring arbitrary power division and high directivity.

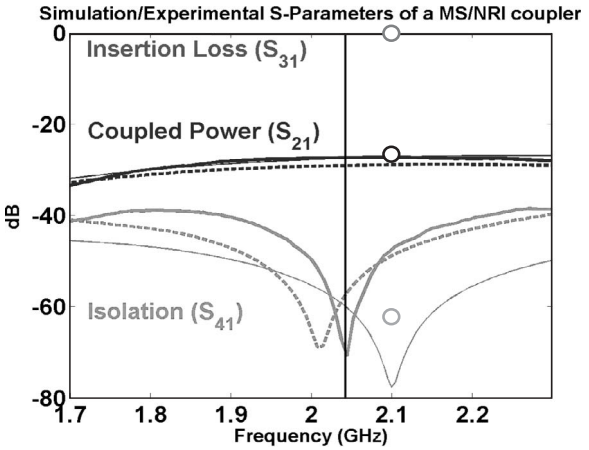


Figure 14.69 Full-wave simulation (dashed line), experimental (solid line), and coupled-mode theory (thin line) *S*-parameters of the MS/NRI coupler. Circles represent approximate Eqs. (14.48)–(14.51).

14.11 SUMMARY

One-dimensional (1D) and to a lesser extent two-dimensional negative refractive index transmission-line (NRI-TL) metamaterials and corresponding antenna applications have been presented. An expression for the total phase shift per-unit-cell of a 1D NRI-TL has been derived, and a technique to ensure that these metamaterial lines operate in the slow-wave regime, and therefore do not radiate, has been presented. Subsequently, various antenna feeding devices have been designed and implemented based on NRI-TL metamaterials. Specifically, a compact and broadband series power divider and a corresponding series-fed linear antenna array with reduced beam squinting have been demonstrated. A leaky-wave antenna radiating in the forward direction has also been shown to exhibit reduced beam squinting. Moreover, an electrically small/low profile ring antenna radiating vertical polarization has also been described. One-dimensional and two-dimensional leaky-wave antennas capable of producing forward or backward beams have also been outlined. Other antenna related NRI-TL devices have been presented including a broadband Wilkinson balun and a coupled-line metamaterial coupler featuring high directivity and low coupling levels for antenna signal monitoring applications. Some of these and other developments in NRI-TL metamaterials, as well as recent developments in other NRI metamaterial technologies, can be found in Ref. 67.

ACKNOWLEDGMENTS

One of the authors (GVE) would like to thank Ashwin Iyer and Rubaiyat Islam for assisting him with writing certain parts of this chapter. Moreover, he is grateful to the Natural Sciences and Engineering Research Council of Canada (NSERC) for financial support and also thanks Nortel Networks for additional financial support.

REFERENCES

1. V. G. Veselago, The electrodynamics of substances with simultaneously negative values of permittivity and permeability, *Sov. Phys. Usp*., Vol. 10, No. 4, pp. 509–514, January 1968.
2. H. Lamb, On group-velocity, *Proc. London Math. Soc.*, Vol. 1, pp. 473–479, 1904.
3. A. Shuster, *An Introduction to the Theory of Optics*, Arnold, London, 1904.
4. R. A. Shelby, D. R. Smith, and S. Schultz, Experimental verification of a negative index of refraction, *Science*, Vol. 292, pp. 77–79, April 6, 2001.
5. J. B. Pendry, A. J. Holden, D. J. Robbins, and W. J. Stewart, Magnetism from conductors and enhanced nonlinear phenomena, *IEEE Trans. Microwave Theory Tech.*, Vol. 47, No. 11, pp. 2075–2084, November 1999.
6. G. V. Eleftheriades, Analysis of bandwidth and loss in negative-refractive-index transmission-line (NRI–TL) media using coupled resonators, *IEEE Microwave Wireless Components Lett.*, Vol. 17, No. 6, pp. 412–414, June 2007.
7. A. K. Iyer and G. V. Eleftheriades, Negative refractive index metamaterials supporting 2-D waves, in *Proceedings of the IEEE MTT-S International Microwave Symposium, Seattle*, Vol. 2, pp. 1067–1070, June 2002.
8. G. V. Eleftheriades, A. K. Iyer, and P. C. Kremer, Planar negative refractive index media using periodically L-C loaded transmission lines, *IEEE Trans. Microwave Theory Tech.*, Vol. 50, No. 12, pp. 2702–2712, December 2002.

9. A. Grbic and G. V. Eleftheriades, A backward-wave antenna based on negative refractive index L-C networks, in *Proceedings of the IEEE International Symposium on Antennas and Propagation*, San Antonio, TX, Vol. 4, pp. 340–343, June 2002.
10. A. Grbic and G. V. Eleftheriades, Experimental verificiation of backward-wave radiation from a negative refractive index metamaterial, *J. Appl. Phys.*, Vol. 92, No. 10, pp. 5930–5935, November 2002.
11. A. K. Iyer, P. C. Kremer, and G. V. Eleftheriades, Experimental and theoretical verification of focusing in a large, periodically loaded transmission line negative refractive index metamaterial, *Optics Express*, Vol. 11, pp. 696–708, April 2003. Available at http://www.opticsexpress.org/abstract.cfm?URI = OPEX-11-7-696.
12. C. Caloz, H. Okabe, T. Iwai, and T. Itoh, Transmission line approach of left-handed (LH) materials, in *Proceedings of the IEEE-AP-S USNC/URSI National Radio Science Meeting*, San Antonio, TX, Vol. 1, p. 39, June 2002.
13. A. A. Oliner, A periodic-structure negative-refractive-index medium without resonant elements, in *Proceedings of the IEEE-AP-S USNC/URSI National Radio Science Meeting*, San Antonio, TX, Vol. 1, p. 41, June 2002.
14. A. Sanada, C. Caloz, and T. Itoh, Characteristics of the composite right/left-handed transmission lines, *IEEE Microwave Wireless Components Lett.*, Vol. 14, pp. 68–70, Febraury 2004.
15. A. Sanada, C. Caloz, and T. Itoh, Planar distributed structures with negative refractive index, *IEEE Trans. Microwave Theory Tech*., Vol. 52, No. 4, pp. 1252–1263, April 2004.
16. A. Grbic and G. V. Eleftheriades, Overcoming the diffraction limit with a planar left-handed transmission-line lens, *Phys. Rev. Lett.*, Vol. 92, No. 11, p. 117403, March 19, 2004.
17. J. B. Pendry, Negative refraction makes a perfect lens, *Phys. Rev. Lett.*, Vol. 85, No. 18, pp. 3966–3969, October 2000.
18. G. V. Eleftheriades, Planar negative refractive index metamaterials based on periodically L-C loaded transmission lines. *Workshop of Quantum Optics, Kavli Institute of Theoretical Physics*, University of Santa Barbara, July 2002. Available at http://online.kitp.ucsb.edu/online/qo02/eleftheriades/(slide #12)
19. A. Grbic and G. V. Eleftheriades, Dispersion analysis of a microstrip based negative refractive index periodic structure, *IEEE Microwave Wireless Components Lett.,* Vol. 13, No. 4, pp. 155–157, April 2003.
20. A. Grbic and G. V. Eleftheriades, An isotropic three-dimensional negative-refractive-index transmission-line metamaterial, *J. Appl. Phys.,* Vol. 98, No. 4 p. 043106, August 2005.
21. W. J. R. Hoefer, P. P. M. So, D. Thompson, and M. M. Tentzeris, Topology and design of wideband 3D metamaterials made of periodically loaded transmission line arrays, *Proceedings of the IEEE MTT-S International Microwave Symposium*, Long Beach, CA, June 2005, pp. 313–316.
22. M. A. Antoniades and G. V. Eleftheriades, Compact, linear, lead/lag metamaterial phase shifters for broadband applications, *IEEE Antennas Wireless Propag. Lett.*, Vol. 2, No. 1, pp. 103–106, July 2003.
23. N. Engheta, An idea for thin, subwavelength cavity resonators using metamaterials with negative permittivity and permeability, *IEEE Antennas Wireless Propag. Lett.,* Vol. 1, No. 1, pp. 10–13, 2002.
24. O. Siddiqui, M. Mojahedi, and G. V. Eleftheriades, Periodically loaded transmission line with effective negative refractive index and negative group velocity, *IEEE Trans. Antennas Propag.* (Special Issue on Metamaterials), Vol. 51, No. 10, pp. 2619–2625, October 2003.
25. I. S. Nefedov, and S. A. Tretyakov, On potential applications of metamaterials for the design of broadband phase shifters, *Microwave Optical Technol. Lett.,* Vol. 45, No. 2, pp. 98–102, 2005.

26. M. A. Antoniades and G. V. Eleftheriades, A broadband series power divider using zero-degree metamaterial phase-shifting lines, *IEEE Microwave Wireless Component Lett.,* Vol. 15, No. 11, pp. 808–810, November 2005.

27. G. V. Eleftheriades, M. A. Antoniades, A. Grbic, A. K. Iyer, and R. Islam, Electromagnetic applications of negative-refractive-index transmission-line metamaterials, in *Proceedings of the 27th ESA Antenna Technology Workshop on Innovative Periodic Antennas*, Santiago de Compostela, Spain, pp. 21–28, March 2004.

28. M. A. Antoniades and G. V. Eleftheriades, A broadband 1:4 series power divider using metamaterial phase-shifting lines, in *Proceedings of the 35th European Microwave Conference*, Paris, France, pp. 717–720, October 2005.

29. B. Bhat and S. K. Koul, *Microwave and Millimeter Wave Phase Shifters*, Artech House, Norwood, MA, 1991.

30. F. A. Olson, Today's microwave acoustic (bulk wave) delay lines, *Microwave J.*, Vol. 13, pp. 67–76, March 1970.

31. E. G. Lean and A. N. Broers, Microwave surface acoustic delay lines, *Microwave J.*, Vol 13, pp. 97–101, March 1970.

32. B. R. McAvoy, Microwave acoustic devices and systems, *IEEE Trans. Microwave Theory Tech.*, pp. 557–559, 1986.

33. J. P. Castera, State of the art in design and technology of MSW devices, *J. Appl. Phy.*, Vol. 55, pp. 2506–2511, March 1984.

34. L. R. Adkins, Dispersion control in magnetostatic wave delay lines, *Circuits Systems Signal Processing*, Vol. 4, No. 1–2, pp. 137–156, 1985.

35. J. R. Pierce, *Travelling Wave Tubes*, Van Nostrand, New York, 1950.

36. L. Xu, R. Taylor, and S. R. Forrest, True time-delay phased-array antenna feed system based on optical heterodyne techniques, *Photonics Technol. Lett.*, Vol. 8, pp. 160–162, January 1996.

37. J. G. Willms, A. Ouacha, L. de Boer, and F. E. van Vliet, A wideband GaAs 6-bit true-time delay MMIC employing on-chip digital drivers, *Proceedings of the 30th European Microwave Conference*, pp. 1–4, October 2000.

38. C. A. Balanis, *Antenna Theory: Analysis and Design*, 2nd ed., John Wiley & Sons, Hoboken, NJ, 1997.

39. M. Qiu, M. Simcoe, and G. V. Eleftheriades, High-gain meander-less slot arrays on electrically thick substrates at millimeter-wave frequencies, *IEEE Trans. Microwave Theory Tech.*, Vol. 50, No. 2, pp. 517–528, February 2002.

40. W. L. Stutzman and G. A. Thiele, *Antenna Theory and Design*, 2nd ed., John Wiley & Sons, Hoboken, NJ, 1998.

41. F. Elek and G. V. Eleftheriades, A two-dimensional uniplanar transmission-line metamaterial with a negative index of refraction, *New J. Phys. Focused Issue on Negative Refraction*, Vol. 7, No. 163, pp. 1–18, August 2005.

42. T. Crepin, J. F. Lampin, T. Decoopman, X. Melique, L. Desplanque, and D. Lippens, Experimental evidence of backward waves on terahertz left-handed transmission lines, *Appl. Phys. Lett.,* Vol. 87, pp. 104–105, September 2005.

43. M. A. Antoniades and G. V. Eleftheriades, A negative-refractive-index transmission-line (NRI-TL) leaky-wave antenna with reduced beam squinting, in *Proceedings of the IEEE International Symposium Antennas and Propagation*. Honolulu, HI, June 2007, pp. 5817–5820.

44. M. A. Antoniades and G. V. Eleftheriades, A CPS leaky-wave antenna with reduced beam squinting using NRI-TL metamaterials, *IEEE Trans. Antennas Propag.,* Vol. 56, No. 3, pp. 708–721, March 2008.

45. G. V. Eleftheriades, A. Grbic, and M. A. Antoniades, Negative-refractive-index transmission-line metamaterials and enabling electromagnetic applications, in *Proceedings of the IEEE International Symposium on Antennas and Propagation*, Monterey, CA, June 2004, Vol. 2, pp. 1399–1402.

46. F. Qureshi, M. A. Antoniades, and G. V. Eleftheriades, A compact and low-profile metamaterial ring antenna with vertical polarization, *IEEE Antennas Wireless Propag. Lett.,* Vol. 4, pp. 333–336, 2005.

47. G. V. Eleftheriades, M. A. Antoniades, and F. Qureshi, Antenna applications of negative-refractive-index transmission-line structures, *IET Trans. Microwaves Antennas Propag.,* Special Issue on Metamaterials LHM, Vol. 1, No. 1, pp. 12–22, February 2007.

48. R. W. Ziolkowski and A. D. Kipple, Application of double negative materials to increase the power radiated by electrically small antennas, *IEEE Trans. Antennas Propag.,* Vol. 51, No. 10, pp. 2626–2640, October 2003.

49. M. Schussler, J. Freese, and R. Jakoby, Design of compact planar antennas using LH-transmission lines, in *Proceedings of the IEEE MTT-S International Microwave Symposium Fort Worth, TX*, Vol. 1, pp. 209–212, June 2004.

50. A. Rennings, S. Otto, C. Caloz, and P. Waldow, Enlarged half-wavelength resonator antenna with enhanced gain, in *Proceedings of the IEEE International Symposium Antennas and Propagation.*, Vol. 3A, pp. 683–686, Washington DC, July 2005.

51. A. Alu, F. Bilotti, N. Engheta, and L. Vegni, Subwavelength, compact, resonant patch antennas loaded metamaterial, *IEEE Trans. Antennas and Propag.*, Vol. 55, No. 1, pp. 13–25, January 2007.

52. K. Z. Rajab, R. Mittra, and M. T. Lanagan, Size reduction of microstrip antennas using metamaterial, *Proceedings of the IEEE International Symposium Antennas and Propagation*, Washington DC, Vol. 2B. pp. 296–299, July 2005,

53. J. L. Volakis et al., Miniaturization methods for narrowband and ultrawideband antennas, in *Proceedings of the IEEE International Workshop on Antenna Technology: Small Antennas and Novel Metamaterial*, pp. 119–121, March 2005.

54. S. A. Tretyakov and M. Ermutlu, Modeling of patch antennas partially loaded with dispersive backward-wave materials, *IEEE Antennas Wireless Propag. Lett.,* Vol. 4, pp. 266–269, 2005.

55. A. Grbic and G. V. Eleftheriades, Leaky CPW-based slot antenna arrays for millimeter-wave applications, *IEEE Trans. Antennas Propag.,* Vol. 50, No. 11, pp. 1494–1504, November 2002.

56. S. Lim, C. Caloz, and T. Itoh, Metamaterial-based electronically controlled transmission-line structure as a novel leaky-wave antenna with tunable radiation angle and beamwidth, *IEEE Trans. Microwave Theory Tech.,* Vol. 51, pp. 161–173, January 2005.

57. A. K. Iyer, and G. V. Eleftheriades, Leaky-wave radiation from a two-dimensional negative-refractive-index transmission-line metamaterial, in *Proceedings of the 2004 URSI EMTS International Symposium on Electromagnetic Theory*, Pisa, Italy, May 2004.

58. A. K Iyer, and G. V. Eleftheriades, Leaky-wave radiation from planar negative-refractive-index transmission-line metamaterials, *Proceedings of the IEEE International Symposium Antennas and Propagation*, Monterey, CA, pp. 1411–1414, June 2004.

59. T. Kokkinos, C. D. Sarris and G. V. Eleftheriades, Periodic FDTD analysis of leaky-wave structures and applications to the analysis of negative-refractive-index leaky-wave antennas, *IEEE Trans. Microwave Theory Tech.,* Vol. 54, No. 4, pp. 1619–1630, June 2006.

60. J.-S. Lim, H.-S. Yang, Y.-T. Lee, S. Kim, K.-S. Seo, and S. Nam, E-band Wilkinson balun using CPW MMIC technology, *IEE Electron. Lett.*, Vol. 40, No. 14, pp. 879–881, July 2004.

61. M. Basraoui and S. N. Prasad, Wideband, planar, log-periodic balun, *Proceedings of the IEEE MTT-S International Microwave Symposium*, Baltimore, Vol. 2, pp. 785–788, June 1998.

62. H. S. Nagi, Miniature lumped element 180° Wilkinson divider, *Proceedings of the IEEE MTT-S International Microwave Symposium*, Philadelphia, Vol. 1, pp. 55–58, June 2003.

63. M. A. Antoniades and G. V. Eleftheriades, A broadband Wilkinson balun using microstrip metamaterial lines, *IEEE Antennas Wireless Propag. Lett.*, Vol. 4, No. 1, pp. 209–212, 2005.

64. R. Islam and G. V. Eleftheriades, A planar metamaterial co-directional coupler that couples power backwards, *Proceedings of the IEEE MTT-S International Microwave Symposium*, Philadelphia, Vol. 1, pp. 321–324, June 2003.

65. R. Islam, F. Elek, and G. V. Eleftheriades, Coupled-line metamaterial coupler having codirectional phase but contradirectional power flow, *IEE Electron. Lett.*, No. 5, pp. 315–317, March 2004.

66. R. Islam and G. V. Eleftheriades, Printed high-directivity metamaterial MS/NRI coupled-line coupler for signal monitoring applications, *IEEE Microwave Wireless Component Lett.*, Vol. 16, No. 4, pp. 164–166, April 2006.

67. G. V. Eleftheriades and K. G. Balmain (Eds.), *Negative-Refraction Metamaterials: Fundamental Principles and Applications,* John Wiley & Sons/ IEEE Press, Hoboken, NJ, 2005.

CHAPTER 15

Artificial Impedance Surfaces for Antennas

DANIEL F. SIEVENPIPER

15.1 INTRODUCTION

Engineered electromagnetic surface textures can be used to alter the properties of metal surfaces to perform a variety of functions. For example, specific textures can be designed to change the surface impedance, to manipulate the propagation of surface waves, or to control the reflection phase. These surfaces provide a way to design new boundary conditions for building new kinds of electromagnetic structures, such as to control the radiation pattern of a small antenna. These surfaces can also be tuned, enabling electronic control of their electromagnetic properties. Tunable impedance surfaces can be used as simple steerable reflectors, or as steerable leaky wave antennas. Using holographic methods, artificial impedance surfaces can provide detailed control over the scattering properties of metal structures and can be used to build conformal antennas on complex shapes.

The simplest example of a textured electromagnetic surface is a metal slab with quarter-wavelength deep corrugations [1–4], as shown in Figure 15.1a. This is often described as a soft or hard surface [5] depending on the polarization and direction of propagation. It can be understood by considering the corrugations as quarter-wavelength transmission lines, in which the short circuit at the bottom of each groove is transformed into an open circuit at the top surface. This provides a high impedance boundary condition for electric fields polarized perpendicular to the grooves, and low impedance for parallel electric fields. Regardless of the polarization of the wave, the surface is described as "soft" for waves propagating perpendicular to the grooves, and "hard" for waves propagating parallel to the grooves, by analogy to the corresponding boundary conditions in acoustics. Soft and hard surfaces are used in various applications such as manipulating the radiation patterns of horn antennas or controlling the edge diffraction of reflectors. These structures are described in detail in another chapter. The same technique has also been used to build structures with two-dimensional periodicity, such as shorted rectangular waveguide arrays [6], or the inverse structures, often known as pin-bed arrays [7]. These textured surfaces are typically one-quarter wavelength thick in order to achieve a high impedance boundary condition.

Modern Antenna Handbook. Edited by Constantine A. Balanis

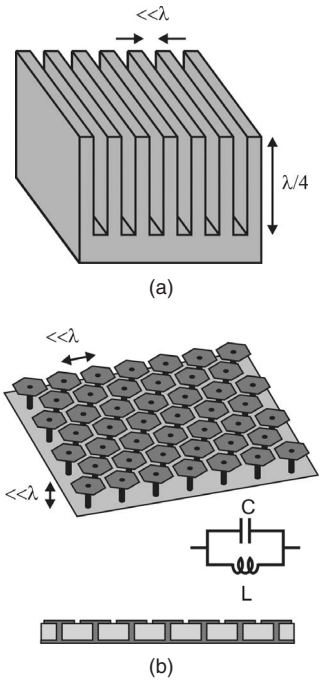


Figure 15.1 (a) A traditional corrugated surface consists of a metal slab with narrow quarter-wavelength long slots. The boundary condition at the top surface depends on the polarization of the incoming wave. The surface impedance is low for electric fields parallel to the slots, and high for electric fields perpendicular to the slots. (b) A high impedance surface is built as a thin two-dimensional lattice of plates attached to a ground plane by metal plated vias. The plates provide capacitance and inductance, and it has high electromagnetic impedance near its *LC* resonance frequency.

Other structures known as artificial impedance surfaces have been built that can also alter the electromagnetic boundary condition of a metal surface, but which are much less than one-quarter wavelength thick [8, 9]. They are typically fabricated as subwavelength mushroom-shaped metal protrusions, as shown in Figure 15.1b. They can be analyzed as resonant *LC* circuits, and the reduction in thickness is achieved by capacitive loading. Near the *LC* resonance frequency, these materials have two important properties: (1) They provide a high impedance boundary condition for both polarizations, and for all propagation directions. (2) They reflect with a phase shift of 0, rather than ρ as with an electric conductor. They are sometimes known as artificial magnetic conductors, because the tangential magnetic field is zero at the surface, just as the tangential electric field is zero at the surface of an electric conductor. In addition to their unusual reflection phase properties, these materials have a surface wave bandgap, within which they do not support bound surface waves of either TM or TE polarization. They may be considered as a kind of photonic crystal [10, 11] for surface waves [12]. Although bound surface waves are not supported, leaky TE waves can propagate within the bandgap, which can be useful for certain antenna applications.

Although most of this chapter is focused on the mushroom-type high impedance surface shown in Figure 15.1b, other types of high impedance surfaces have also been

studied which have their own unique properties. For example, the uniplanar compact photonic bandgap structure [13] achieves similar reflection phase properties by using a Jerusalem cross type frequency selective surface on a grounded dielectric substrate. Other kinds of frequency selective surfaces have also been studied, including multilayer structures [14]. Specific structures have also been designed to provide multiple resonances [15] or to provide greater stability of the resonance frequency with angle or polarization [16]. These structures will not be covered in detail here, but many of the concepts presented here for the mushroom structures apply to other types of high impedance surfaces.

By incorporating tunable materials or devices into textured surfaces, their capabilities are expanded to include active control of electromagnetic waves. This can be accomplished using mechanical structures such as movable plates [17] or electrical components such as varactor diodes. With a tunable textured surface, one can build devices such as programmable reflectors that can steer or focus a reflected microwave beam [18]. These can provide a low cost alternative to traditional electrically scanned antennas (ESAs), where phase shifters and complicated feed structures are replaced by a planar array of varactor diodes and a free-space feed. Despite being low cost, these steerable reflector antennas are ruled out for some applications because they require a space feed and are thus not entirely planar.

Steerable leaky waves provide an alternative approach to electronic beam steering, without requiring a space feed [19]. The surface is programmed with a periodic impedance function that scatters the surface wave into free space. This steering method allows the scattered radiation to be steered over a wide scan range in both the forward and backward directions. Backward leaky waves can also be understood as resulting from bands of negative dispersion, similar to those in other negative index materials.

Finally, borrowing concepts from holography, impedance surfaces can be patterned to control the scattering properties of a metal surface to produce radiation patterns that would otherwise be difficult to obtain. For example, a small monopole antenna surrounded by a thin artificial impedance surface can produce a narrow, high gain beam in an arbitrary direction. By covering a metal body with such a surface texture, it is also possible to allow a small antenna on one side of the object to produce significant gain on the opposite side of the object, in a direction that would normally be shadowed. By offering a high degree of control over the scattering properties of metal objects, these holographic artificial impedance surfaces provide the potential to obtain arbitrary radiation patterns from small antennas on arbitrarily shaped objects.

15.2 SURFACE WAVES

A surface wave is an electromagnetic wave that is bound to an interface between two materials. The materials can include metals, dielectrics, free space, electromagnetic bandgap structures, or other kinds of metamaterials. The surface wave decays with distance from the interface, as shown in Figure 15.2. Surface waves can be classified as either transverse magnetic (TM) or transverse electric (TE). For TM surface waves, the magnetic field is entirely transverse to the direction of propagation. Referring to Figure 15.2, a TM surface wave propagating in the x-direction has nonzero magnetic field only in the y-direction. Similarly, a TE surface wave propagating in the x-direction has nonzero electric field only in the y-direction.

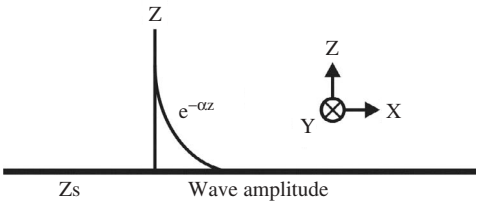


Figure 15.2 A surface wave is a wave that is bound to a surface and decays into the surrounding space.

By applying a texture to a metal surface, we can alter its surface impedance, and thereby change its surface wave properties. The behavior of surface waves on an impedance surface depends on whether the surface is inductive or capacitive, and on the magnitude of the reactance. The complete derivation is beyond the scope of this chapter, and the reader is referred to several texts in which it is described [8, 20]. In brief, the derivation proceeds by assuming a surface having an impedance Z_s, and a wave that decays exponentially away from a surface with decay constant α, as shown in Figure 15.2. For TM waves, we apply Maxwell's equations to determine the relationship between the surface impedance and the surface wave properties. It can be shown that TM waves occur on an inductive surface, in which the surface impedance is given by the following expression:

$$Z_s = \frac{j\alpha}{\omega\varepsilon} \tag{15.1}$$

Conversely, transverse electric (TE) waves can occur on a capacitive surface, with the following impedance:

$$Z_s = \frac{-j\omega\mu}{\alpha} \tag{15.2}$$

In the above expressions, ε and μ are the permittivity and permeability of the space surrounding the surface, which may be vacuum, and ω is the angular frequency of the wave. We see that TM waves require a positive imaginary impedance, or an inductive surface, while TE waves require a negative imaginary impedance, or a capacitive surface.

Ordinary metals are slightly inductive, due to the self-inductance of the thin layer of current that forms within a skin depth of the surface. Being inductive, they support TM surface waves. At optical frequencies these are often called surface plasmons [21]. At microwave frequencies, they are simply the ordinary surface currents, and they are only very weakly bound to the surface. A diagram of a TM surface wave is shown in Figure 15.3a. While bare metals do not support TE surface waves, dielectric coated metals can support TE waves above a cutoff frequency that depends on the thickness and dielectric constant of the layer. Photonic crystals, frequency selective surfaces, textured surfaces, and other interfaces can also support TE waves if the effective surface impedance is capacitive. The surface impedance of the textured metal surface described here is characterized by a parallel resonant LC circuit. As will be described in the next section, at low frequencies the surface is inductive and supports TM waves. At high frequencies it is capacitive and supports TE waves, as depicted in Figure 15.3b. Near the LC resonance frequency, the surface impedance is very high. In this region, waves

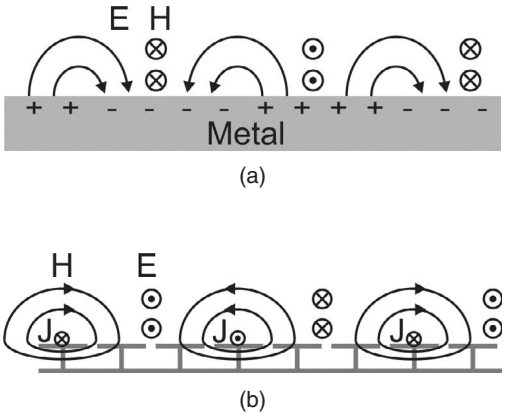


Figure 15.3 (a) In a transverse magnetic (TM) surface wave, shown here on a flat metal surface, the electric field arcs out of the surface, and the magnetic field is transverse to the surface. (b) The fields take the opposite form in a transverse electric (TE) surface wave, shown here on a high impedance surface.

are not bound to the surface; instead, they radiate readily into the surrounding space as leaky waves.

15.3 HIGH IMPEDANCE SURFACES

High impedance surfaces consist of an array of metal protrusions on a flat metal sheet. The protrusions are arranged in a two-dimensional lattice and can be visualized as mushrooms or thumbtacks protruding from the surface. High impedance surfaces are typically constructed as printed circuit boards, where the bottom side is a solid metal ground plane, and the top contains an array of small ($\ll \lambda$) metal patches, as shown in Figure 15.1b. The plates are connected to the ground plane by metal plated vias to form a continuous conductive metal texture. It can be considered as a two-dimensional version of the corrugated ground plane, where the quarter-wavelength resonant corrugations have been folded up into small resonant circuits and distributed on a two-dimensional lattice. For greater capacitance, multilayer circuit boards with overlapping plates can be used.

When the period is small compared to the wavelength of interest, we may analyze the material as an effective medium, with its surface impedance defined by effective lumped-element circuit parameters that are determined by the geometry of the surface texture. A wave impinging on the material causes electric fields to span the narrow gaps between the neighboring metal patches, and this can be described as an effective sheet capacitance C. As currents oscillate between the neighboring patches, the conducting paths through the vias and the ground plane provide a sheet inductance L. These form a parallel resonant circuit that dictates the electromagnetic behavior of the material, as shown in Figure 15.4. Its surface impedance is given by the following expression:

$$Z_s = \frac{j\omega L}{1 - \omega^2 LC} \tag{15.3}$$

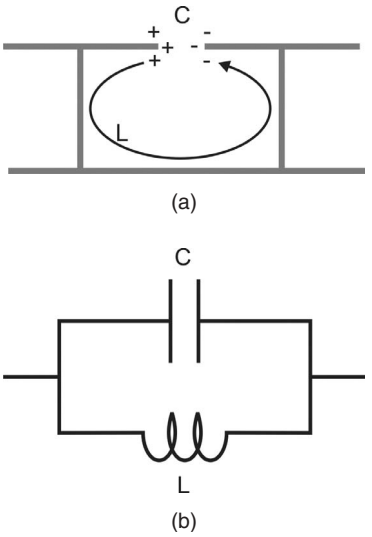


Figure 15.4 (a) The capacitance in a high impedance surface is due to the proximity of the neighboring plates. The inductance comes from the current loops that are formed between the plates and the ground plane, through the vias. (b) The impedance of the surface can be modeled as a parallel resonant LC circuit.

The resonance frequency of the circuit is given by

$$\omega_0 = \frac{1}{\sqrt{LC}} \tag{15.4}$$

It is evident from Eq. (15.3) that, below resonance, the surface is inductive. As described in the previous section, inductive surfaces support TM waves. Above resonance, the surface is capacitive and supports TE waves. Near ω_0, the surface impedance is much higher than the impedance of free space, and the material does not support bound surface waves.

In addition to its unusual surface wave properties, the high impedance surface also has unusual reflection phase properties. In the frequency range where the surface impedance is very high, the tangential magnetic field is small, even with a large electric field along the surface. Such a structure is sometimes described as an artificial magnetic conductor. Because of this unusual boundary condition, the high impedance surface can function as a new type of ground plane for low profile antennas. The image currents in the ground plane are in-phase with the antenna current, rather than out-of-phase, allowing radiating elements to lie directly adjacent to the surface, while still radiating efficiently. For example, a dipole lying flat against a high impedance ground plane is not shorted as it would be on an ordinary metal sheet.

15.4 SURFACE WAVE BANDS

Many of the important properties of the high impedance surface can be explained using an effective surface impedance model. The surface is assigned an impedance equal to

that of a parallel resonant LC circuit, as described previously. The use of lumped circuit parameters to describe electromagnetic structures is valid as long as the wavelength is much longer than the size of the individual features. The effective surface impedance model can predict the reflection properties and some features of the surface wave band structure, but not the bandgap itself, which by definition must extend to large wavevectors.

The wavenumber, k, is related to the spatial decay constant, α, and the frequency, ω, by the following dispersion relation:

$$k^2 = \mu_0 \varepsilon_0 \omega^2 + \alpha^2 \tag{15.5}$$

For TM waves we can combine Eq. (15.5) with Eq. (15.1) to find the following expression for k as a function of ω, in which η is the impedance of free space and c is the speed of light in vacuum:

$$k = \frac{\omega}{c}\sqrt{1 - \frac{Z_s^2}{\eta^2}} \tag{15.6}$$

We can find a similar expression for TE waves by combining Eq. (15.5) with Eq. (15.2):

$$k = \frac{\omega}{c}\sqrt{1 - \frac{\eta^2}{Z_s^2}} \tag{15.7}$$

Inserting Eq. (15.3) into Eqs. (15.6) and (15.7), we can plot the dispersion diagram for surface waves, in the context of the effective surface impedance model. An example of the complete dispersion diagram, calculated using the effective medium model, is shown in Figure 15.5.

Below resonance, TM surface waves are supported. At low frequencies, they lie very near the light line, indicated in Figure 15.5 by the dotted line with a slope equal to the speed of light, c. The fields extend many wavelengths beyond the surface, as they do

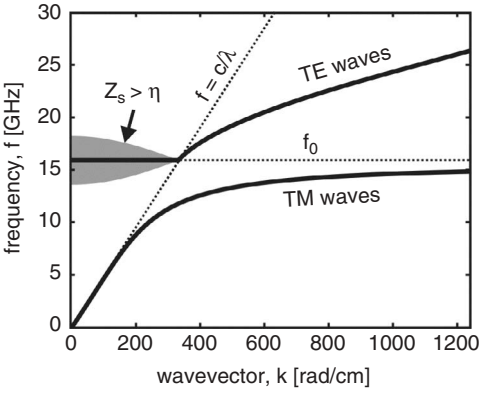


Figure 15.5 The effective surface impedance model can determine many of the properties of the high impedance surface, including the shape and polarization of the surface wave bands. This is the predicted surface wave dispersion diagram for a surface with sheet capacitance of 0.05 pF and sheet inductance of 2 nH. The surface supports TM waves below the resonance frequency and TE waves at higher frequencies. This model does not predict the bandgap, but it does predict a region of radiative loss.

on a flat metal sheet. Near the resonant frequency, the surface waves are tightly bound to the surface and have a very low group velocity. The dispersion curve is bent over away from the light line. In the effective surface impedance limit, there is no Brillouin zone [22] boundary, and the TM dispersion curve approaches the resonance frequency asymptotically. Thus this approximation does not predict the bandgap.

Above the resonance frequency, the surface is capacitive, and TE waves are supported. The lower end of the dispersion curve is close to the light line, and the waves are weakly bound to the surface, extending far into the surrounding space. As the frequency is increased, the curve bends away from the light line, and the waves are more tightly bound to the surface. The slope of the dispersion curve indicates that the waves feel an effective index of refraction that is greater than unity. This is because a significant portion of the electric field is concentrated in the capacitors.

The TE waves that lie to the left of the light line exist as leaky waves that are damped by radiation, which can be modeled as a resistor in parallel with the high impedance surface. The damping resistance is the impedance of free space, projected onto the surface at the angle of radiation. This blurs the resonance frequency, so the leaky waves actually radiate within a finite bandwidth. Small wavevectors represent radiation perpendicular to the surface, while wavevectors near the light line represent radiation at grazing angles. In place of a bandgap, the effective surface impedance model predicts a frequency band characterized by radiation damping.

In the effective impedance surface model described earlier, the properties of the textured surface are summarized into a single parameter—the surface impedance. This model correctly predicts the shape and polarization of the surface wave bands, and also the reflection phase, to be described later. However, it does not predict the bandgap itself. For a more accurate picture of the surface wave properties, we can use full-wave numerical electromagnetic techniques. The metal and dielectric regions are discretized on a grid, and the electric field at all points on the grid is described in terms of an eigenvalue equation, which may be solved numerically. A single unit cell is simulated, and Bloch boundary conditions [23] are used. The calculation yields the allowed frequencies for each wavevector. An example of a high impedance surface is shown in Figure 15.6, along with the calculated dispersion diagram. The lowest band is TM, the second band is TE, and both have a similar shape to that predicted by the effective surface impedance model. A bandgap, within which the surface does not support bound surface waves of either polarization, extends from the top of the TM band to the point where the TE band crosses the light line. The finite element model also predicts additional higher order bands that are not predicted by the simple effective surface impedance model.

Surface wave modes can be measured by recording the transmission between a pair of small coaxial probes placed near the surface. Depending on their orientation, the probes will excite surface waves with TM, TE, or both polarizations. An example of a high impedance surface and the measured surface wave transmission across a 12-cm sample is shown in Figure 15.7 for both TM and TE polarizations. As predicted by both models described previously, TM waves are supported at low frequencies, and TE waves are supported at high frequencies. The TM and TE bands are separated by a bandgap within which bound surface waves of either polarization are not supported. For comparison, an electric conductor of the same size exhibits nearly flat transmission for TM waves at microwave frequencies, at around -30 dB, and very low transmission for TE waves, at around -60 dB.

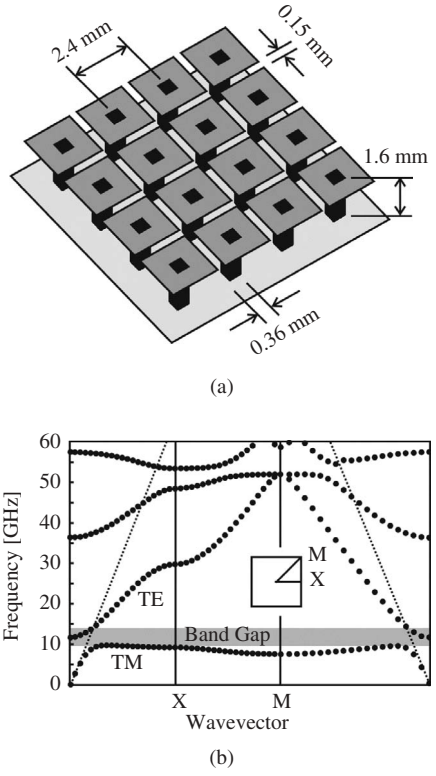
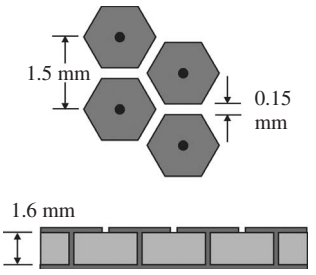


Figure 15.6 (a) The complete dispersion diagram can be obtained accurately using numerical methods, and square lattices are often easier to simulate. The substrate, which is not shown, has a relative dielectric constant of 2.2. (b) The lowest bands are qualitatively similar to that of the effective surface impedance model. The finite element model also predicts a bandgap where bound surface waves of neither polarization are supported, between the first two bands. It also predicts several higher bands.

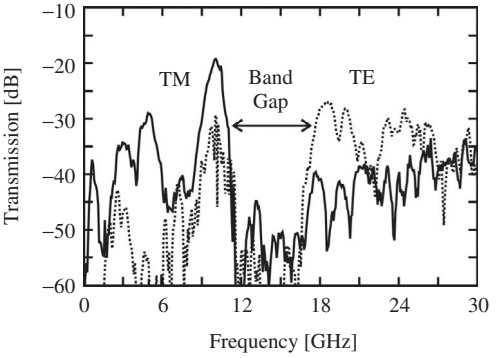
15.5 REFLECTION PHASE

The surface impedance defines the boundary condition at the surface for the standing wave formed by incident and reflected waves. For a low impedance surface, such as an electric conductor, the ratio of electric field to magnetic field is small. The electric field has a node at the surface, and the magnetic field has an antinode. Conversely, for a high impedance surface, the electric field has an antinode at the surface, while the magnetic field has a node. Another term for such a surface is again an artificial magnetic conductor. Using the effective surface impedance model described previously, we can determine the reflection phase for the resonant textured surface described above. For a normally incident wave, the reflection phase for a plane wave in free space with impedance η reflecting from a surface with impedance Z_s is given as follows:

$$\Phi = \mathrm{Im}\left[\ln\left(\frac{Z_s - \eta}{Z_s + \eta}\right)\right] \tag{15.8}$$



(a)



(b)

Figure 15.7 (a) Measurements were performed on a triangular lattice of hexagons, built on a substrate with a relative dielectric constant of 2.2. (b) The high impedance surface supports TM surface waves (solid curve) at low frequencies and TE waves (dashed curve) at high frequencies. Between these two bands is a gap, within which waves of neither polarization are supported. Variations in transmission magnitude within the surface wave bands are due to multipath propagation on the surface.

In the above expression, Z_s is given by Eq. (15.3), and η is the impedance of free space. At very low frequencies, the reflection phase is π, and the structure behaves like a smooth metal surface. At higher frequencies, the reflection phase slopes downward and eventually crosses through zero at the resonance frequency, where it behaves as a magnetic conductor. Above the resonance frequency, the phase returns to $-\pi$. The phase falls within $\pi/2$ and $-\pi/2$ when the magnitude of the surface impedance exceeds the impedance of free space. The behavior of the reflection phase predicted by Eq. (15.8) is identical to the measured result shown in Figure 15.8. This reflection phase curve was measured using the surface shown in Figure 15.7a. It is worth noting that, for a wide range of geometries, the edges of the surface wave bandgap occur at the same frequencies where the reflection phase crosses through $\pi/2$ and $-\pi/2$. However, this is only the case for the mushroom-type surface described here. Other types of high impedance surfaces, particularly those that do not include vertical vias, do not necessary have the same relationship between the reflection phase and a surface wave bandgap.

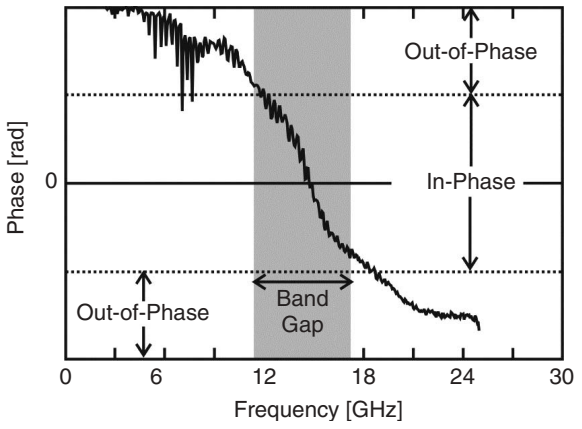


Figure 15.8 The reflection phase was measured for the same surface as shown in Figure 15.7. The phase is zero at the resonance frequency, but it approaches π for frequencies far from the resonance. The phase crosses through $\pi/2$ and $-\pi/2$ near the edges of the surface wave bandgap.

15.6 BANDWIDTH

The bandwidth of textured surfaces can be derived in two ways. The first method involves calculating the bandwidth of a small antenna near the textured surface. It is instructive for developing a physical intuition for the behavior of these surfaces and how they affect nearby antennas. However, additional factors often play a role in the antenna bandwidth, such as how the antenna is fed, and the use of impedance matching structures. Furthermore, outside the bandgap, the antenna can couple into bound surface wave modes, and this often causes the apparent bandwidth as determined by the input match to be greater than the useful bandwidth as determined by the gain and radiation pattern. Therefore the following derivation should be considered as the maximum practical radiation bandwidth for an antenna on an artificial impedance surface. Nonetheless, the surface itself has an intrinsic bandwidth, and understanding what contributes to it is important for comparing different designs of artificial impedance surfaces. In the second method for calculating bandwidth, we dispense with the need for an antenna and simply assume that the surface behaves as a resonant cavity that is damped by radiation into the surrounding free space. Both of these methods predict the same bandwidth result. The following analysis is for mushroom-type structures where the unit cells are small compared to the wavelength. However, the more general solution also applies to the reflection phase bandwidth of other types of artificial impedance surfaces that can be modeled as a parallel LC circuit.

An antenna lying parallel to a textured surface will see the impedance of free space on one side, and the impedance of the surface on the other side. Where the textured surface has low impedance, far from the resonance frequency, the antenna current is mirrored by an opposing current in the surface. Since the antenna is shorted out by the nearby conductor, the radiation efficiency is very low. Within the bandgap near resonance, the textured surface has much higher impedance than free space, so the antenna is not shorted out. In this range of frequencies, the radiation efficiency is high.

The textured surface can be modeled as an LC circuit in parallel with the antenna, and the radiation into free space can be modeled as a resistor with a value of the impedance of free space. The amount of power dissipated in the resistor is a measure of the radiation efficiency of the antenna. The maximum radiation efficiency occurs near the LC resonance frequency of the surface, where its reactance is much greater than the impedance of free space. At very low frequencies, or at very high frequencies, currents in the surface cancel the antenna current, and the radiated power is reduced. In this model, the frequencies where the radiation drops to half of its maximum value occur when the magnitude of the surface impedance is equal to the impedance of free space, as described by the following equation:

$$|Z_s| = \eta \tag{15.9}$$

In the above expression, η is the impedance of free space and is given by the following:

$$\eta = \sqrt{\mu_0/\varepsilon_0} \tag{15.10}$$

Substituting Z_s from Eq. (15.3) into Eq. (15.9), we can solve for ω to determine edges of the operating band:

$$\omega^2 = \frac{1}{LC} + \frac{1}{2\eta^2 C^2} \pm \frac{1}{\eta C}\sqrt{\frac{1}{LC} + \frac{1}{4\eta^2 C^2}} \tag{15.11}$$

The terms in $1/(nC)^2$ are typically small compared to the terms in $1/LC$, so we will neglect them. This approximation yields the following equation for the edges of the operating band:

$$\omega = \omega_0\sqrt{1 \pm \frac{Z_0}{\eta}} \approx \omega_0\left(1 \pm \frac{1}{2}\frac{Z_0}{\eta}\right) \tag{15.12}$$

In the above expression, Z_0 can be considered as a kind of characteristic impedance of the surface.

$$Z_0 = \sqrt{L/C} \tag{15.13}$$

The two frequencies designated by the $\pm$ signs in Eq. (15.12) delimit the range over which an antenna would radiate efficiently on such a surface. The total bandwidth (BW) is roughly equal to the characteristic impedance of the surface divided by the impedance of free space:

$$BW = \frac{\Delta\omega}{\omega_0} \approx \frac{Z_0}{\eta} = \frac{\sqrt{L/C}}{\sqrt{\mu_0/\varepsilon_0}} \tag{15.14}$$

This is the bandwidth over which the phase of the reflection coefficient falls between $\pi/2$ and $-\pi/2$, and image currents are more in-phase than out-of-phase. As noted in the previous section, this range often coincides with the surface wave bandgap for mushroom-type structures. It also represents the maximum usable bandwidth of a flush-mounted antenna on a resonant surface of this type.

The intrinsic bandwidth of the surface can be calculated in a more general way by considering the surface as a two-dimensional resonator that stores electromagnetic energy, which is radiated away into free space as a plane wave normal to the surface. Imagine

an electromagnetic mode on the surface that oscillates at the LC resonance frequency. The mode consists of a uniform electric field across the capacitive gaps at the top of the surface, and the associated currents that flow on the ground plane below, corresponding to the inductive part of the structure. The surface and the oscillating mode are both assumed to be infinite in extent, to reduce the model to one dimension. The radiation into surrounding space is represented as a resistor equal to the impedance of free space, $R = \eta$ as described by Eq. (15.10).

The quality factor, Q, of any damped resonator is defined as follows:

$$Q = 2\pi \frac{\text{energy stored}}{\text{energy lost per cycle}} \tag{15.15}$$

For a parallel resonant RLC circuit, the Q as defined by Eq. (15.15) is given by the following expression, in which V is the voltage across the circuit, and I is the current through the inductor:

$$Q = 2\pi \frac{\frac{1}{2}CV^2 + \frac{1}{2}LI^2}{\frac{V^2}{R}\frac{1}{f}} \tag{15.16}$$

We can substitute the impedance of the inductor to eliminate I, and obtain the following:

$$Q = \omega \frac{CV^2 + \frac{V^2}{\omega^2 L}}{2\frac{V^2}{R}} = \frac{\omega RC}{2} + \frac{R}{2\omega L} \tag{15.17}$$

Since we are operating near the resonance frequency, we can substitute Eq. (15.4) to eliminate ω and obtain an expression for the Q of the textured surface.

$$Q = R\sqrt{\frac{C}{L}} \tag{15.18}$$

Substituting the impedance of free space η for R, we see that this equation for Q is simply the inverse of the equation for bandwidth given by Eq. (15.14). Thus both methods for calculating bandwidth produce the same value.

$$BW = \frac{1}{Q} = \frac{\sqrt{L/C}}{\sqrt{\mu_0/\varepsilon_0}} \tag{15.19}$$

We can use Eq. (15.4) to substitute for C in the above expression to obtain

$$BW = \frac{\omega_0 L}{\sqrt{\mu_0/\varepsilon_0}} \tag{15.20}$$

It will be shown in the next section that the inductance of the surface L is equal to the product of the permeability, μ and the thickness, t. Substituting for L, we can obtain the following:

$$BW = \omega_0 t\sqrt{\mu_0\varepsilon_0} \tag{15.21}$$

Next, we use two expressions for the speed of light, c:

$$c = \frac{1}{\sqrt{\mu_0 \varepsilon_0}} = \frac{\lambda_0 \omega_0}{2\pi} \tag{15.22}$$

Using Eq. (15.22) in Eq. (15.21), we obtain a more useful expression for the bandwidth of a thin ($t \ll \lambda_0$), nonmagnetic ($\mu = \mu_0$), resonant textured ground plane:

$$BW = \frac{2\pi}{\lambda_0} t \tag{15.23}$$

In the above expression, λ_0 is the free-space wavelength at the resonance frequency. This result is significant because it proves that the bandwidth is determined entirely by the thickness of the surface, with respect to the operating wavelength. Note that the dielectric constant of the substrate has no direct effect on the bandwidth, and dielectric loading cannot be used to reduce the thickness, except at the expense of bandwidth.

Equation (15.23) can also be written

$$BW = \frac{t}{r_0} \tag{15.24}$$

The term r_0 is the radianlength as defined by Wheeler [24]:

$$r_0 = \frac{\lambda_0}{2\pi} \tag{15.25}$$

Just as Eq. (15.24) shows that the bandwidth of a resonant planar source (a one-dimensional problem) is limited by its thickness, a similar limitation exists for a resonant point source such as an electrically small antenna (a three-dimensional problem) [24–26]. In that case, the bandwidth is determined by the volume of the antenna, compared to a normalized volume.

$$BW \leq \frac{V}{V_0} \tag{15.26}$$

In the above expression, V is the volume of the sphere enclosing the antenna, and V_0 is the volume of the radiansphere [24], which is a sphere with a radius of one radianlength. Its volume is given by the following expression, in which r_0 is the same term as in Eq. (15.25):

$$V_0 = \frac{4\pi}{3} r_0^3 \tag{15.27}$$

Thus both the small resonant antenna and the planar resonant surface are governed by their size and thickness, respectively, normalized by a common length $\lambda_0/2\pi$.

For electrically small antennas, Eq. (15.26) represents the upper bound for the achievable bandwidth, when the electromagnetic fields are uniformly distributed throughout a spherical volume. The bandwidth is reduced from this upper limit by the degree to which the fields do not uniformly fill that volume, so an antenna with highly localized fields will have narrower bandwidth. The same rule may be applied to a resonant textured ground plane, and structures with uneven field distributions tend to have narrower bandwidths. It is known that the geometry of the resonant elements can have a similar effect on the performance of reflectarrays [32].

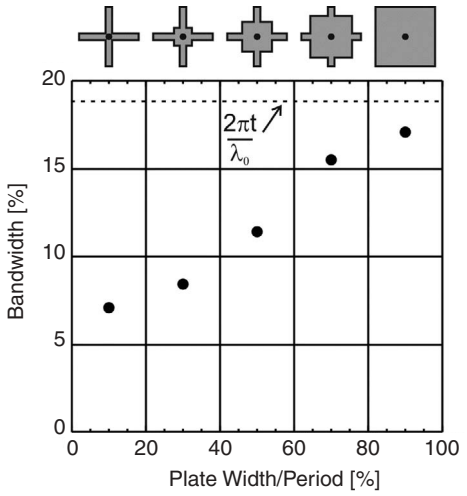


Figure 15.9 The instantaneous bandwidth of a tunable impedance surface as a function of the shape of the metal pattern in each unit cell. Broad square shapes produce a bandwidth that approaches the theoretical bandwidth limit for a high impedance surface. Narrow wire structures result in a narrower bandwidth.

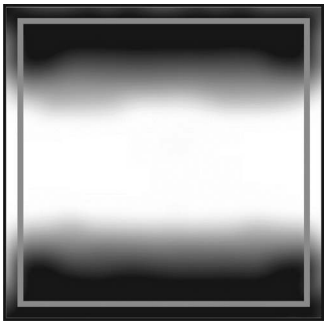
The effect of the unit cell on the bandwidth of artificial impedance surfaces is illustrated in Figure 15.9. For the structures with broad plates, the bandwidth approaches the theoretical maximum bandwidth described by Eq. (15.24). For structures that resemble narrow wires, the bandwidth is narrower. This dependence of the bandwidth on the plate geometry can be explained by examining the electric fields within the unit cell. The field inside the substrate is shown for two cases in Figure 15.10. For the narrow wire grid structure, the electric field is primarily concentrated beneath the wires, while for the lattice of square plates, it is more evenly distributed across the plate edges. As discussed previously for electrically small antennas, greater concentration of the fields in a small space within the available volume reduces the bandwidth. The same concept also applies to artificial impedance surfaces.

15.7 DESIGN PROCEDURE

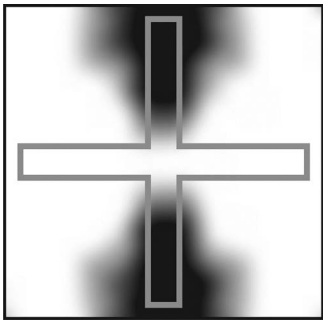
The following is a general procedure for designing mushroom-type high impedance surfaces for a required frequency and bandwidth. In the two-layer geometry shown in Figure 15.4, the capacitors are formed by the fringing electric fields between adjacent metal patches. For fringing capacitors, the capacitance can be approximated [8] as follows:

$$C_{\text{fringe}} \approx \frac{w(\varepsilon_1 + \varepsilon_2)}{\pi} \cosh^{-1}\left(\frac{a}{g}\right) \tag{15.28}$$

In the above expression, a is the lattice constant, g is the gap between the plates, w is the width of the plates, and ε_1 and ε_2 are the dielectric constants of the substrate and the material surrounding the surface, which may be free space. More accurate expressions for the fringing field capacitance exist [27], but Eq. (15.28) is adequate for first order designs.



(a)



(b)

Figure 15.10 The relationship between bandwidth and unit cell geometry can be understood by examining the fields within each unit cell. (a) For a broad square unit cell, the electric field is more evenly distributed within the substrate. (b) For a unit cell consisting of narrow wires, the electric field is concentrated beneath the wires. This reduces the effective size of the resonant structure and narrows the bandwidth. Similar behavior is seen in electrically small antennas.

A three-layer design shown in Figure 15.11 achieves a lower resonance frequency for a given thickness by using capacitive loading. In this geometry, parallel-plate capacitors are formed by the top two overlapping layers. The capacitance can be calculated with the well known equation.

$$C_{\text{parallel}} \approx \frac{\varepsilon A}{d} \tag{15.29}$$

In this case, ε is the dielectric constant of the material between the plates, A is the area of the plates, and d is their separation.

Figure 15.11 Thin high impedance surfaces with a low resonance frequency can be built by using greater capacitive loading, such as overlapping plates, as shown in this three-layer structure. For a given resonance frequency, thinner structures have smaller fractional bandwidth.

In either case, the sheet capacitance is determined by the value of the individual capacitors and a geometrical factor F that depends on the choice of lattice:

$$C = C_{\text{individual}} \times F \tag{15.30}$$

The geometrical factor takes into account the number of capacitors in series or parallel, to convert the value of the individual capacitors into the sheet capacitance per square. For a square lattice $F = 1$, for a triangular lattice $F = \sqrt{3}$, and for a hexagonal grid of capacitors $F = 1/\sqrt{3}$. Examples of various lattices, for a three-layer design with overlapping capacitors, are shown in Figure 15.12.

The inductance of a high impedance surface is determined entirely by its thickness. This can be understood by considering a solenoid of current that includes two rows of plates, and their associated vias. Current flows up one row of vias, across the capacitors, and down the next set of vias, to return through the ground plane. The length and width of the solenoid are canceled to obtain the sheet inductance:

$$L = \mu t \tag{15.31}$$

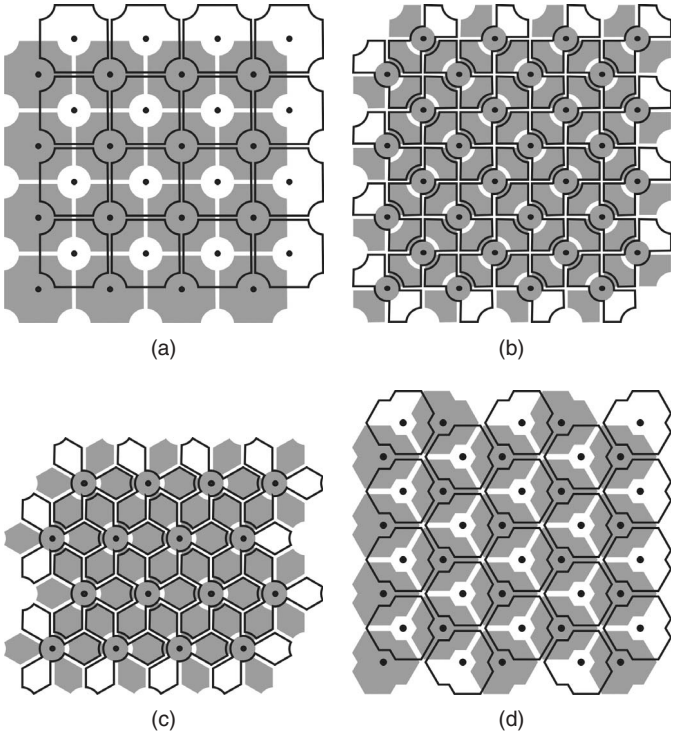


Figure 15.12 The electromagnetic properties of the high impedance surface depend primarily on the surface capacitance and inductance and do not significantly depend on the geometry. Shown here are several three-layer structures, including (a) a square lattice with a completely overlapping layer, (b) a square lattice with two similar layers, (c) a triangular lattice, and (d) a hexagonal lattice, which is another form of the triangular lattice but with two vias per unit cell. In all cases, the shaded regions represent the lower metal layer, and the outlined regions represent the upper metal layer. The solid dots represent conductive vias.

To design a surface for a desired frequency ω_0 and bandwidth BW, we combine Eqs. (15.4), (15.23), and Eq. (15.31). This procedure yields an equation for the required thickness:

$$t = \frac{c \cdot BW}{\omega_0} \tag{15.32}$$

It also provides an equation for the required sheet capacitance.

$$C = \frac{1}{\omega_0 \eta \cdot BW} \tag{15.33}$$

Finally, using either Eq. (15.28) or Eq. (15.29), together with Eq. (15.30), an appropriate geometry for the capacitors can be found. In order for the effective surface impedance approximation to be valid, the lattice constant should be small compared to the wavelength, and this often dictates whether a two-layer or three-layer structure should be used. Note that aside from the effects of the geometrical factor F, the choice of lattice and the shape and material composition of the capacitors has no effect on the electromagnetic properties of the surface, as long as their value and arrangement follow the guidelines given above.

For an accurate design, numerical electromagnetic software should be used to refine the estimate given by this procedure. A single unit cell can be simulated with minimal computing resources. Electric and magnetic conducting boundaries are used on opposing walls of the unit cell. Simulations of the reflection phase, the geometry, and the materials can be adjusted to provide the desired resonance frequency and bandwidth. Nonetheless, it is useful to have an intuitive solution as given above, to more rapidly converge on the correct design.

15.8 ANTENNA APPLICATIONS

The high impedance surface can be used to provide several advantages for antenna applications, using either the suppression or enhancement of surface waves, or using its unusual reflection phase. Manipulation of surface wave effects can be demonstrated with a simple vertical monopole, shown in Figure 15.13a. It is fabricated by feeding a coaxial cable through a hole in the ground plane. The center conductor is extended through the other side to form a radiating wire, and the outer conductor is shorted to the ground plane.

On a finite metal ground plane, currents generated by the monopole are scattered at the edges of the ground plane. This can be seen as radiation in the backward direction, and also as ripples in the forward portion of the radiation pattern because the scattered radiation interferes with the direct radiation from the monopole. Figure 15.13b shows the radiation pattern of a 3-mm monopole on a 5-cm square metal ground plane, measured at 35 GHz.

If the metal ground plane is replaced with a high impedance surface designed to resonate near 35 GHz, surface waves are suppressed, and the radiation pattern is changed. While driven currents can exist on any reflective surface, they do not propagate on the high impedance ground plane. Any induced currents are restricted to a localized region around the antenna and never reach the edges of the ground plane. The absence of radiation from the edges results in a smoother radiation pattern, with less power in

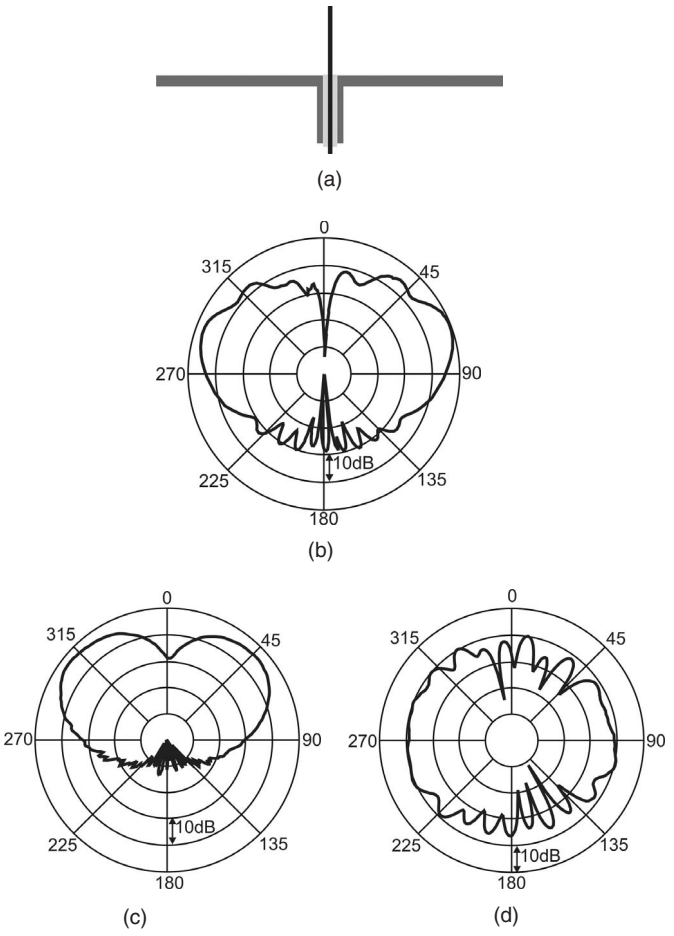


Figure 15.13 (a) A monopole antenna can be built by feeding a coaxial cable through a ground plane. The outer conductor is attached to the ground plane, and the inner conductor is extended to the other side to form the antenna. (b) On a flat metal ground plane, the monopole produces the expected radiation pattern. (c) On a high impedance ground plane, at a frequency within the bandgap, the antenna produces a smooth pattern with reduced radiation in the backward direction. (d) Outside the bandgap, the antenna produces a complex pattern with significant power in the backward direction.

the backward direction, as shown in Figure 15.13c, due to suppression of scattering from the ground plane edges. This could be used to reduce effects of nearby objects or discontinuities in the ground plane.

Two additional features are apparent in Figure 15.13c. First, the center null is diminished because of asymmetry in the local geometry of the antenna wire and the surrounding metal patches. With more careful construction, the null could be recovered. Second, the received power is lower with the high impedance ground plane, especially at the horizon. This is because the image currents on the high impedance ground plane are reversed with respect to their direction on a metal ground plane. For a vertical monopole, this tends to cancel the radiation from the antenna current, particularly along the horizontal directions.

If the antenna is operated outside the bandgap of the high impedance surface, where surface waves are supported, the radiation pattern is significantly different. Figure 15.13d shows the radiation pattern of the same antenna at 26 GHz, within the TM surface wave band. The vertical monopole couples strongly into the surface wave modes, and a high density of states at the upper TM band edge also increases the amount of energy in the surface wave modes. Because of the presence of surface waves, the pattern contains many lobes and nulls, and a significant amount of power in the backward direction. Such a pattern could be useful for applications requiring nearly omnidirectional radiation in environments where significant shadowing would otherwise occur.

While the vertical monopole illustrates the application of high impedance surfaces for the suppression or enhancement of surface currents, it does not explain the advantage of the unusual reflection phase properties. The benefits of an artificial magnetic conductor can be seen by using a horizontal wire antenna, as shown in Figure 15.14a. A simple wire antenna is fed through the back of the surface by a coaxial cable, in a manner similar to the monopole; and it is bent over across the surface. The wire is typically about one-half wavelength long at the resonance frequency of the surface. On a flat metal ground plane, a horizontal wire is shorted out, and most of the power transmitted to the feed is reflected back. However, on the high impedance surface, a horizontal wire antenna is well matched if operated within the bandgap, as shown by the return loss in Figure 15.14b. The radiation pattern in Figure 15.14c indicates that the antenna produces gain, despite being roughly 1 mm above the ground plane. This is because the reflection phase of the surface is 0, rather than π, as with an ordinary conductor. Thus currents in the high impedance surface reinforce the currents in the wire, instead of canceling them as a smooth metal surface does. This effect can be used to build a variety of low profile antennas that can lie directly adjacent to the artificial magnetic ground plane, such as antennas with various polarizations, including circular, as well as various directive radiation patterns.

The small wire antennas described in this section are useful because they demonstrate how one can achieve a variety of unusual effects using high impedance surfaces. However, the gain of the simple horizontal wire antenna is about 7 dBi, which is easily achievable with a standard patch antenna. Furthermore, as with many types of antennas, as the size of the surrounding high impedance surface or ground plane is reduced, the gain diminishes and the radiation toward the backward direction increases. If the whole antenna including the surrounding structure is smaller than about $\lambda/2\pi$, it is nearly impossible to achieve appreciable directivity. Thus the most significant advantage of these surfaces is not that they can be used to build small antennas, but rather that they provide a new kind of boundary condition that can be used for things like control of surface wave propagation, control over the diffraction properties of metal surfaces, and electronic beam steering.

15.9 DIFFRACTION CONTROL

Surface textures such as artificial impedance surfaces can be used to control the diffraction properties of a metal object for one or both polarizations [28]. One way to visualize this is through the support or suppression of surface waves. The high impedance surface described earlier suppresses TM, TE, or both polarizations, depending on the resonance frequency of the surface and the bandgap, in relation to the frequency of interest.

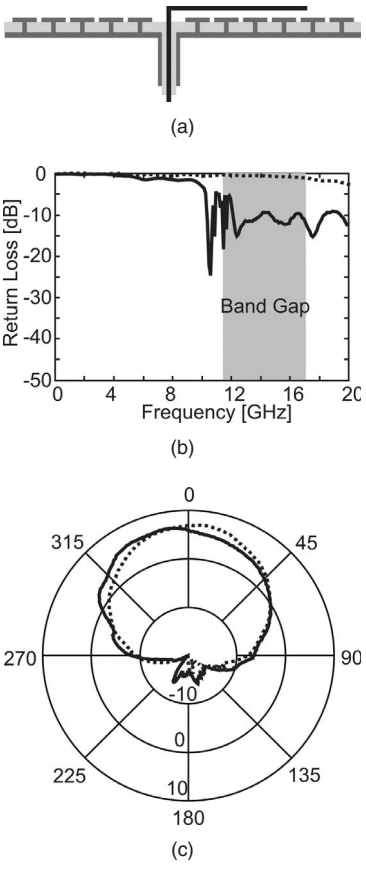


Figure 15.14 (a) Low profile antennas can be built on high impedance surfaces, such as a horizontal bent wire antenna that is a small fraction of a wavelength above the surface. (b) The measured return loss of the horizontal wire antenna on the high impedance surface is low within the bandgap (solid line). On a smooth metal ground plane, the antenna is shorted and does not radiate (dashed line). (c) The radiation pattern of the horizontal wire antenna on the high impedance ground plane is symmetrical, and the E-plane pattern (solid curve) is very similar to the H-plane pattern (dashed curve).

One example of where diffraction control can be useful is in the generation of circular polarization. The task of designing an antenna that transmits or receives in circular polarization over a wide range of angles is often complicated by the presence of the metallic structures on which the antenna is mounted. For example, antennas on a flat metal ground plane will tend to emit in vertical polarization at angles near grazing, because horizontal fields are shorted by the metal surface, while vertical fields can propagate along the metal. We often describe the polarization purity of a wave in terms of its axial ratio, which is the ratio of the major axis to the minor axis of the polarization ellipse.

A simple aperture antenna, shown in Figure 15.15, demonstrates the use of the high impedance surface as a means of improving the symmetry of the radiation pattern. The aperture was the open end of a standard Ku-band rectangular waveguide, which was attached to a similarly sized rectangular hole in the center of a high impedance ground

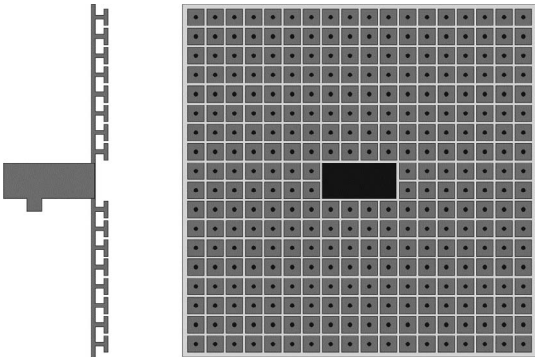


Figure 15.15 The side view (left) and front view (right) of our aperture antenna in a high impedance surface. The unit cells of the high impedance surface measure 3.7 mm, and the size of the ground plane (not shown to scale) is 12.7 cm. The aperture is fed by a coax to Ku-band rectangular waveguide transition.

plane. In this particular case, the surface has a bandgap that spans from 12 to 18 GHz. The results from this surface can be compared to an identical aperture antenna with a metal ground plane of the same size.

While the radiation pattern is somewhat affected by the shape of the aperture, it is primarily determined by the geometry of the surrounding ground plane, and the electromagnetic boundary condition of that surface. The flat metal ground plane supports the propagation of TM polarized waves, because in these waves the electric field is perpendicular to the metal surface. Waves with this polarization can propagate for long distances in close proximity to a metal surface. For this reason, the E-plane radiation pattern in Figure 15.16a is quite broad. TE waves, on the other hand, cannot propagate at grazing angles to a metal surface because their transverse electric field is shorted by the conducting surface. The H-plane is therefore much narrower. This is the expected radiation pattern for an aperture antenna in a conducting ground plane.

On the textured ground plane, the pattern is much more symmetrical, as shown in Figure 15.16b. This can be attributed to the suppression of both TM and TE surface waves near the resonance frequency. The gain is also higher in the forward direction, and this can be attributed to standing waves that occur at the resonance frequency and surround the aperture, which cause a slight increase in the effective aperture area. The radiation patterns shown here are for 13 GHz, within the bandgap. The antenna produces a similar pattern throughout most of the bandgap region. However, as the frequency is increased toward the upper edge where leaky TE waves are supported, the H-plane actually becomes broader than the E-plane, as shown in Figure 15.16c. If one had a ground plane that behaved as a magnetic conductor, one would expect a broad H-plane and a narrow E-plane—the opposite of the electric conductor. Thus, in this way, the textured surface mimics a magnetic conductor.

15.10 TUNABLE IMPEDANCE SURFACES

The resonance frequency and the reflection phase of a high impedance surface can be tuned by changing its effective capacitance, the inductance, or both. However, without

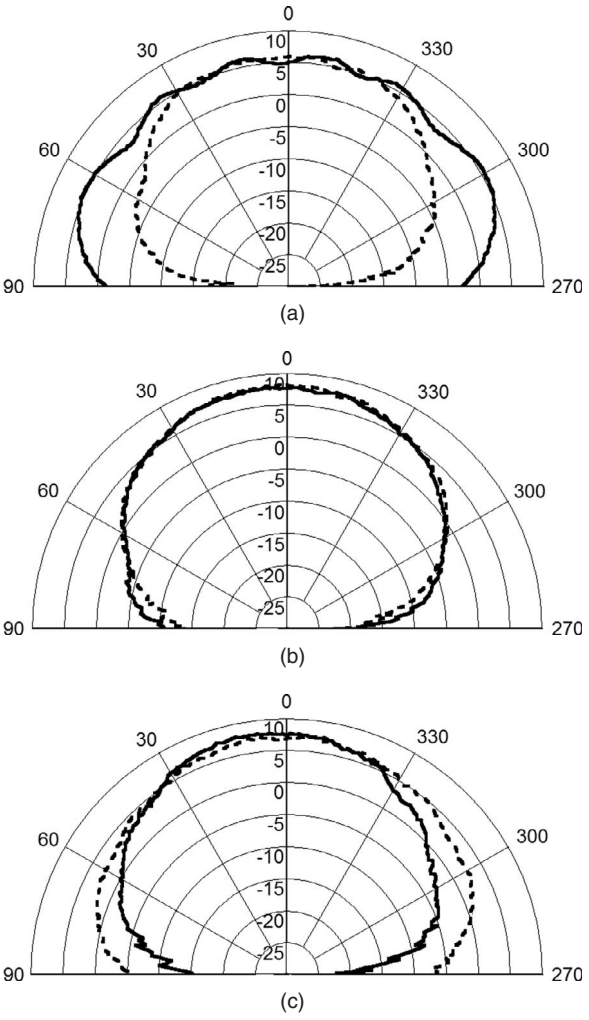


Figure 15.16 The radiation pattern of an aperture antenna in a conventional metal ground plane. The E-plane (solid line) has a much broader pattern than the H-plane (dashed line) because horizontally polarized waves are shorted out by the metal surface, while vertically polarized waves are allowed. (b) The radiation pattern of the aperture antenna in the high impedance surface. The E-plane and H-plane are nearly identical at the resonance frequency of the surface, where waves of both polarization are suppressed. (c) The radiation pattern of the aperture antenna in the high impedance ground plane near the edge of the TE band. The H-plane is broader than the E-plane, as would be expected if the surface were a magnetic conductor.

magnetically active materials, the inductance is determined entirely by the thickness of the surface and is difficult to tune. On the other hand, the capacitance can be controlled by changing the geometry and arrangement of the metal plates, or by adding tunable lumped capacitors. Because the reflection phase is determined by the frequency of the incoming wave with respect to the resonance frequency, such a surface can perform as a distributed phase shifter. As the resonance frequency is swept from low to high values,

the curve in Figure 15.8 is shifted from left to right, so the reflection phase at any fixed frequency varies from $-\pi$ to π.

An electrically tunable impedance surface can be built by connecting neighboring cells with varactor diodes. Changing the bias voltage on the diodes adjusts the capacitance and tunes the resonance frequency. In order to supply the required voltage to all of the varactors, we alternately bias half of the cells and ground the other half in a checkerboard pattern, as shown in Figure 15.17. At the center of each biased cell, a metal via passes through a hole in the ground plane and connects to a control line located on a separate circuit layer on the back of the surface. The varactors are oriented in opposite directions in each alternate row, so that when a positive voltage is applied to the control lines, all the diodes are reverse biased. By individually addressing each cell, the reflection phase can be programmed as a function of position across the surface.

The reflection phase for various bias conditions is shown in Figure 15.18. As the voltage across the varactors is increased, the capacitance decreases, and the resonance frequency increases. For a fixed frequency, the reflection phase increases with bias voltage. The impedance and reflection phase are governed by Eqs. (15.3) and (15.8), respectively. For frequencies within the tuning range, nearly any reflection phase can be obtained by choosing the correct bias voltage. A series of measured data relating the reflection phase to frequency and voltage forms the basis of a calibration table that can be used to steer a reflected beam at any frequency within the tuning range.

15.11 REFLECTIVE BEAM STEERING

If the reflection phase is programmed as a function of position across the surface, it can be used for beam steering. A linear phase gradient $\partial\phi(x, y)/\partial x$ will reflect a normally

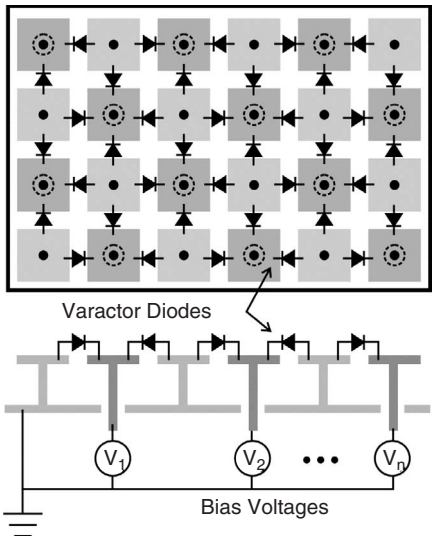


Figure 15.17 A tunable impedance surface consists of a high impedance surface in which adjacent cells have been connected by varactor diodes, which have voltage-tunable capacitance. Half of the vias are grounded, but the other half are attached to a voltage control network on the back of the surface. The grounded and biased plates are arranged in a checkerboard pattern.

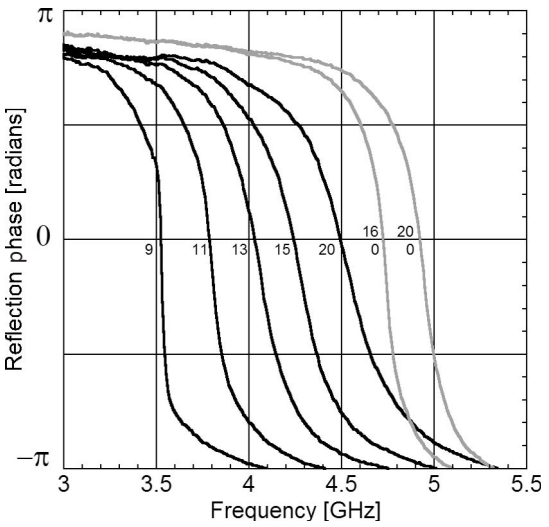


Figure 15.18 The reflection phase of the surface can be tuned electronically by varying the bias voltage on the varactors. Numbers by each curve represent voltages. The two grey curves are for alternating voltages on every other row. For frequencies within the tuning range, nearly any reflection phase can be created by the appropriate choice of bias voltage.

incident microwave beam to an angle θ that depends on the magnitude of the gradient:

$$\theta = 2\tan^{-1}\left(\frac{\lambda}{2\pi}\frac{\partial\phi(x,y)}{\partial x}\right) \tag{15.34}$$

This can be derived by calculating the reflection angle from a plane reflector that is oriented at an angle that would provide the same phase gradient. Other phase functions can be used for other tasks, such as a parabolic phase function for focusing. These concepts have been demonstrated previously using arrays of various resonant elements ranging from dipoles to patches, and beam-forming structures employing this technique are commonly known as reflectarrays [29–34]. Tunable reflectarrays using varactor diodes and related devices known as grid arrays [35, 36] have also been built. The tunable impedance surface [18] has the advantage, when compared to other kinds of tunable reflectarrays, that the bias lines do not interfere with the microwave fields on the front side, and two-dimensional steering is possible.

To create an electronically steerable reflector, the tunable impedance surface is illuminated with a microwave beam, and a phase gradient is created electronically, as shown in Figure 15.19. To steer the beam into a particular angle, we calculate the required reflection phase gradient, as described by Eq. (15.34); select a frequency; and then calculate the corresponding voltages for each bias line based on a previously measured calibration table. The radiation patterns for several sets of control voltages corresponding to several beam steering angles are shown in Figure 15.20. Since each of the cells is individually addressable through the bias lines in the back, the surface can steer in two dimensions. For this example, the surface is about 3.75 wavelengths square and operates at about 4.5 GHz. The surface can steer a reflected beam over $\pm 40^\circ$ for both polarizations. Wider steering angles would be possible with a larger surface.

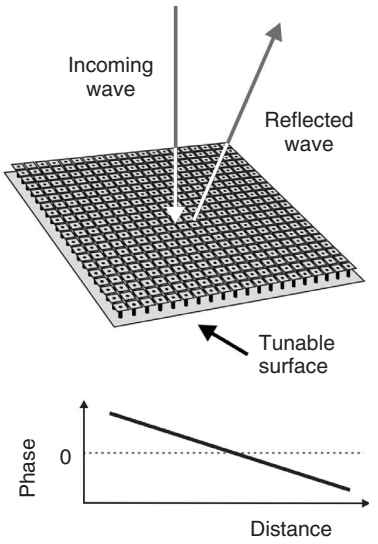


Figure 15.19 The tunable surface can be used as an electronic beam steering reflector by programming the surface to have a reflection phase gradient. A reflected microwave beam will be steered to an angle that depends on the phase gradient. This can serve as a simple electronically scanned antenna.

It should be noted that the decrease in gain with steering angle follows the standard cosine dependence that is typical for phased arrays. There is not a separate specular reflection component due to the ground plane. This is because the reflection phase calculated in Eq. (15.8) and plotted in Figure 15.18 is for the entire structure, including the metal patches, vias, ground plane, and dielectric substrate. Furthermore, the lattice period is typically much smaller than one-half wavelength at the operating frequency, so there is not a separate area of ground plane between the cells to provide a separate specular reflection component to the radiation pattern.

Despite their advantages for cost, size, and weight compared to traditional phased arrays, tunable artificial impedance surfaces do have their practical limitations. For example, the varactor tuning range limits the achievable phase range. Using a surface with a steeper phase curve can mitigate this problem, but at the expense of bandwidth. There is ultimately a trade-off between intrinsic surface bandwidth, varactor tuning range, and allowable phase error. These three parameters affect the sidelobe levels and usable bandwidth of the antenna. Furthermore, since the varactors respond not only to the applied bias but also to the incoming RF field, these antennas can support only limited power levels. The effect of high incident power is to cause an increase in sidelobe levels. However, this can be mitigated by using varactors that support high tuning voltage.

15.12 LEAKY WAVE BEAM STEERING

Despite being simple and low cost, steerable reflectors based on tunable impedance surfaces are ruled out for some applications because they require a free-space feed and thus are not entirely planar. An alternative is to use a leaky wave design [37–40],

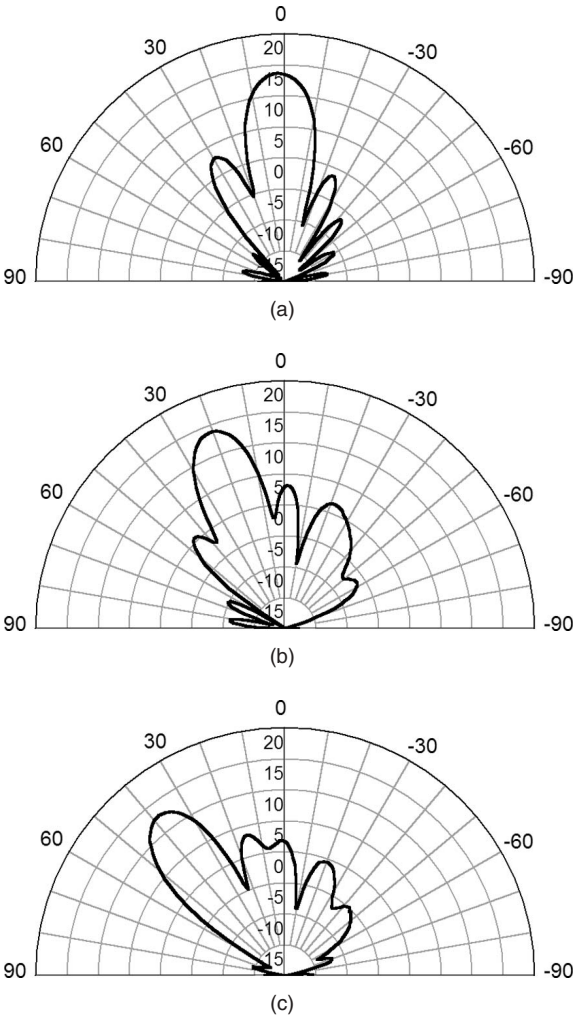


Figure 15.20 The beam can be steered over a range of $\pm 40^{\circ}$, and greater steering would be possible with a larger structure. Example radiation patterns are shown for (a) broadside, (b) 20°, and (c) 40°.

where a surface wave is excited directly in the surface and then radiates energy into the surrounding space as it propagates. This method involves programming the surface with a periodic impedance function that scatters the surface wave into free space. The period of the surface impedance can be varied to change the phase matching condition between the surface wave and the space wave, and thus steer the radiated wave. The beam can be steered electronically over a wide range in both the forward and backward directions. The decay rate of the surface waves can also be controlled independently of the beam angle to allow adjustment of the aperture profile.

To build a steerable leaky wave antenna [19], a feed structure is integrated into the tunable surface, such as the flared notch antenna shown in Figure 15.21. It can be as close as a small fraction of a wavelength from the surface, but it should not be close

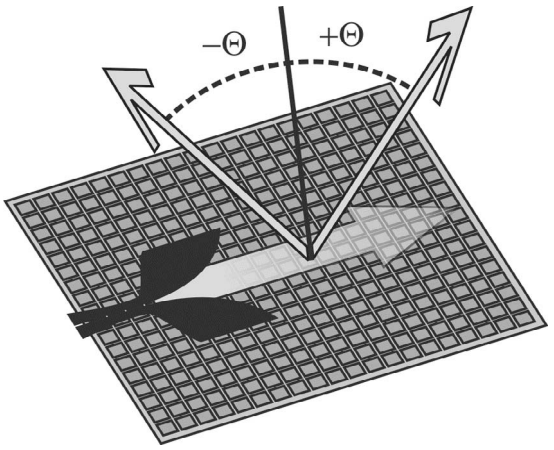


Figure 15.21 The tunable impedance surface can be used as an electronically steerable leaky wave antenna by incorporating a conformal feed, such as a flared notch antenna. The surface wave propagates away from the antenna, but the radiation can propagate in either the forward or backward direction, depending on the phase matching condition at the surface.

enough to detune the capacitance between the plates below it. A flared notch antenna will generate TE waves, or a wire antenna can be used for TM waves.

A periodic pattern of voltages is applied to the tunable surface to create a periodic surface impedance function. When waves propagate across the surface, they are scattered by the nonuniform surface impedance. The scattered energy radiates at an angle determined by the wavevector of the surface wave, and the periodicity of the surface impedance. The radiation angle may be determined by assuming that a wave launched into the surface feels an effective refractive index of n_{eff}. Its wavevector is $k_0 n_{\text{eff}}$, where $k_0 = 2\pi/\lambda$ is the free-space wavevector. The surface impedance has period p, corresponding to a wavevector $k_p = 2\pi/p$. The scattered radiation in free space must have a total wavevector of k_0, and phase matching requires that it have a component parallel to the surface that is equal to the sum of the wavevectors of the surface wave and the surface impedance function. As illustrated in Figure 15.22, the radiation is scattered into the forward direction if $k_p < k_0 n_{\text{eff}}$ and it is scattered backward if $k_p > k_0 n_{\text{eff}}$. In general, the radiation angle is given by the following expression:

$$\theta = \sin^{-1}\left(\frac{k_0 n_{\text{eff}} - k_p}{k_0}\right) \tag{15.35}$$

For backward leaky waves, the energy still travels outward from the feed, so its group velocity is in the forward direction, but its phase velocity, which determines the radiation angle, is in the backward direction. Leaky wave structures capable of backward or broadside radiation have been studied extensively [41–43], but tunable impedance surfaces are novel because they can be electronically reconfigured to steer continuously from the forward to the backward direction at a single frequency. Recently, antennas based on other types of planar metamaterials have demonstrated similar capabilities [44, 45]. Figure 15.23 shows examples of radiation in both the forward and backward directions.

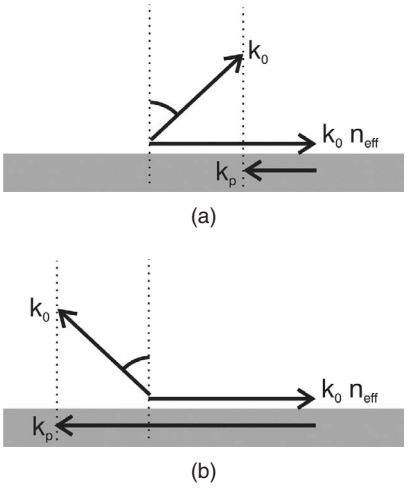


Figure 15.22 The direction of radiation is determined by phase matching at the surface. The tangential component of the wavevector of the space wave must match the difference between that of the surface wave and that of the periodic surface impedance. (a) Forward leaky waves are generated when the surface impedance has a period that is greater than the wavelength of the surface waves, corresponding to a shorter wavevector. (b) Backward leaky waves are generated when the period of the surface impedance is shorter than the wavelength of the surface wave.

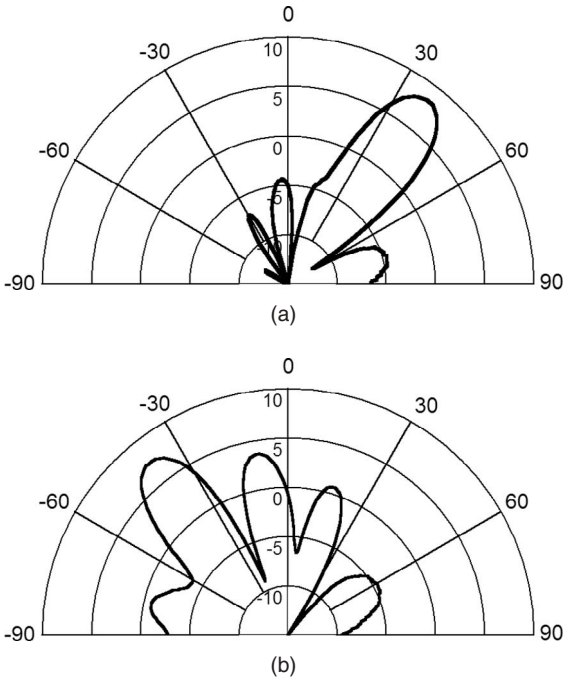


Figure 15.23 The surface can be configured for either forward or backward leaky wave radiation. This shows examples of the leaky wave radiation patterns when the surface is programmed for (a) forward radiation and (b) backward radiation.

15.13 BACKWARD BANDS

The existence of backward leaky waves can be analyzed in terms of backward bands, similar to those produced in other kinds of bulk metamaterials [46–49]. In this section, we explore the properties of backward bands on textured surfaces and study their behavior through reflection measurements. Consider a tunable impedance surface in which alternate rows of plates are biased at two different voltages, thus creating rows of alternating capacitance values. For a TE wave, the electric field is transverse to the direction of propagation, so it sees alternating capacitance as it propagates from row to row. The effective lattice period is doubled, and the Brillouin zone is halved, as shown in Figure 15.24. The upper half of the TE band is folded into a reduced Brillouin zone, labeled BZ. In the upper part of the TE band, the sign of the phase velocity ω/k is opposite to that of the group velocity $d\omega/dk$, so we may describe this as a backward band. The group velocity corresponds to the direction of energy propagation along the surface, which is always outward from the feed. The phase velocity, which determines the direction of radiation, progresses backward toward the feed.

Using mode analysis, the direction and relative strength of the electric field in the capacitors can be determined for various points on the band diagram. Groups of small arrows in Figure 15.24 illustrate the electric field in four adjacent rows. At the bottom of the TE band, the electric field is parallel throughout the entire surface. In the mode at the top of the backward band, the fields are antiparallel in each adjacent row of capacitors. Because the capacitors have alternating values on every other row, the period of this mode matches that of the surface, with alternating capacitance values, and it lies at $k = 0$. For modes that occur at the edge of the folded zone, one-half wavelength fits in each period of two capacitors. Thus the field is zero in alternate rows of capacitors and antiparallel in every other alternate row.

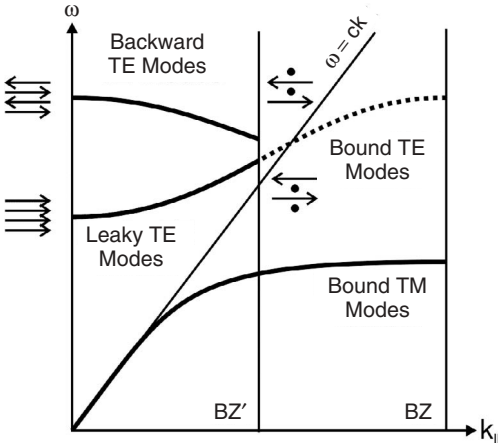


Figure 15.24 Backward leaky wave radiation can be understood in terms of backward bands. If the surface is tuned so that every other row has alternate voltages, the TE waves will see a surface with a period that is twice as large. The Brillouin zone will be reduced by half, and the upper portion of the TE band will be folded into the reduced zone, labeled BZ′. The phase velocity and group velocity in this band will have opposite signs, corresponding to a backward wave. The fields in each row can be deduced by mode analysis for points at the edges of the bands.

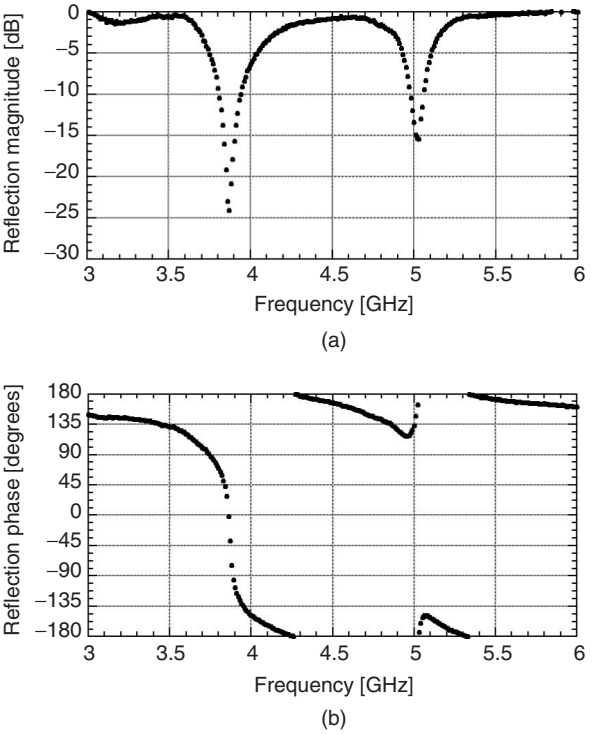


Figure 15.25 The presence of a backward band, as depicted in Figure 15.24, can be measured from the normal incidence reflection properties of the surface. Modes at zero wavevector are visible as (a) dips in the magnitude and (b) corresponding curves in the phase. The second mode corresponds to the top of the backward band. It disappears if a uniform voltage is applied to the entire surface.

It is possible to detect the presence of the backward band using reflection measurements. Modes at $k = 0$ are standing waves that support a finite tangential electric field at the surface, and they can be identified by frequencies where the reflection phase is zero, and by decreased reflectivity due to losses in the varactor diodes. Figure 15.25 shows the reflection magnitude and phase when adjacent rows were biased at 10 and 20 volts. When two different voltages are applied to alternate rows, two modes are visible, corresponding to the lower edge of the forward TE band and the upper edge of the backward TE band. The presence of the second mode is experimental evidence of the backward band. It is not present when a uniform voltage is applied to all of the varactors.

15.14 HOLOGRAPHIC ARTIFICIAL IMPEDANCE SURFACES

A common problem for antenna designers is the integration of low profile or conformal antennas onto complex metallic objects, while maintaining the desired radiation pattern. Simple shadowing by the surrounding structure can result in deep, wide nulls in the radiation pattern, so it can be nearly impossible to achieve appreciable radiation to some angles. Furthermore, the antenna excites currents in the surrounding structure, and these

currents contribute to the radiation pattern in unexpected ways. Artificial impedance surfaces can provide a way to control the electromagnetic scattering and the propagation of surface currents around complex metallic shapes, to simultaneously enable greater flexibility in the design and position of conformal antennas, and more control over the radiation pattern.

Artificial impedance surfaces do not need to contain all elements of the high impedance surfaces discussed above; they can be created with a variety of simple planar metal patterns. Vertical vias can be used if necessary to adjust the band structure beyond what is possible with planar structures, or to create bandgaps to completely block propagation in certain regions. The material used here consists of a grounded dielectric layer covered with a pattern of square conductive patches, avoiding the use of vertical vias for simplicity. The patches are small compared to the wavelength of interest, so their scattering properties can be described in terms of their effective surface impedance. By varying the geometry of the patches, one can control the surface impedance as a function of position.

Holographic artificial impedance surfaces [50, 51] combine concepts from three fields: (1) leaky waves on modulated impedance surfaces, (2) holographic antennas, and (3) artificial electromagnetic media. The study of leaky wave antennas dates back as far as the mid-1950s [52]. Many early leaky wave antennas involved dielectric structures in which the thickness was modulated [53, 54] or tapered [55] to vary the effective surface impedance. Oliner and Hessel [56] developed a comprehensive model of leaky waves on modulated impedance surfaces. This model describes how the propagation and radiation of leaky waves are controlled by the surface reactance, modulation depth, and period, and it serves as a foundation for the thin printed leaky wave antennas described here.

The surface impedance is designed using the concept of holography. Originally demonstrated at optical frequencies, holography involves producing an interference pattern using two waves, one representing the object to be imaged, as illuminated by a laser beam, and the other representing a reference plane wave from the same laser. The interference pattern formed by these two waves is recorded on photographic film. When the reference wave illuminates the developed film, it is scattered to produce a copy of the original image wave.

Microwave holography [57, 58] works in much the same way. A source antenna produces the reference wave, and the desired radiation pattern corresponds to the image wave. The microwave hologram is a collection of scatterers arranged on the interference pattern produced by these two waves. Like the optical analog, the microwave hologram scatters the reference wave to produce the image wave. The reference wave can either be a surface wave [59] or a plane wave propagating parallel [60] or perpendicular [61] to the surface. Most traditional holographic antennas involve metal strips or dipoles [62, 63] that are printed on a dielectric layer. Designs based on alternating regions of electric and magnetic conductors have also been studied numerically [67]. Silicon-based steerable holographic antennas have also been studied [64, 65]. Recent work shows that holographic antennas are capable of high efficiency if the surface and feed are properly designed [66].

Various kinds of artificial impedance surfaces have been studied in the past, such as pin-bed structures [68], and high impedance surfaces [9]. These materials can be combined with the concept of tunable impedance surfaces [18] to allow for completely arbitrary electronic reconfigurability of the surface scattering properties. Because the metallic patterns in these structures are small compared to the wavelengths of interest, they can be described in terms of effective medium properties such as the effective

surface impedance. This permits the use of fast numerical methods that can analyze structures that are many wavelengths in size, without modeling the detailed structure of the material, which would render the problem impractically large.

Traditional holographic antennas are built using discrete scatterers, such as metallic strips or dipoles. Artificial impedance surfaces provide a continuously variable surface impedance. Implemented as a lattice of subwavelength metallic patterns, they produce an effective surface impedance that depends on the geometry of the metal shapes. Compared to the previously studied binary holograms, artificial impedance surfaces allow greater control over the radiation properties of the antenna. An example of a microwave hologram is shown in Figure 15.26. The interference pattern that is formed by a point source in a two-dimensional plane, and a plane wave propagating 60° from normal to that plane appears as a pattern of ellipses.

One implementation of an artificial impedance surface is a lattice of subwavelength metal patches on a grounded dielectric substrate. Figure 15.27 shows an example of a single unit cell of such a surface, and the surface impedance range as a function of the gap between the metal patches, as calculated by Ansoft HFSS electromagnetic software. The impedance range is plotted for the case of a phase difference of 72° between neighboring cells. The impedance is defined here as the ratio of average electric to magnetic field above the surface, for the eigenmode that satisfies the periodic boundary conditions on the cell walls. For this example, with gaps ranging from 1 to 0.2 mm, the effective impedance varied from $161j$ to $234j$ ohms at a frequency of about 17 GHz. An alternative way to calculate the surface impedance is simply to solve for the eigenfrequency for a given wavevector, as determined by the phase difference across the unit cell, and use Eq. (15.6) to find the surface impedance. These two methods produce equivalent results.

The holographic pattern shown in Figure 15.26 is combined with the impedance versus geometry data shown in Figure 15.27, to determine the required geometry versus position on the surface, and generated a pattern of squares. A small section of the nonuniform pattern of metal squares is shown in Figure 15.28. The surface is fed by a small monopole antenna inserted from the back, at the focus of the ellipses shown in Figure 15.26. Currents generated by the monopole are scattered by the holographic impedance surface to produce a narrow pencil beam at 60°, as shown in Figure 15.29, with a gain of about 20 dBi. Thus the holographic pattern has transformed the initial wave (the currents from

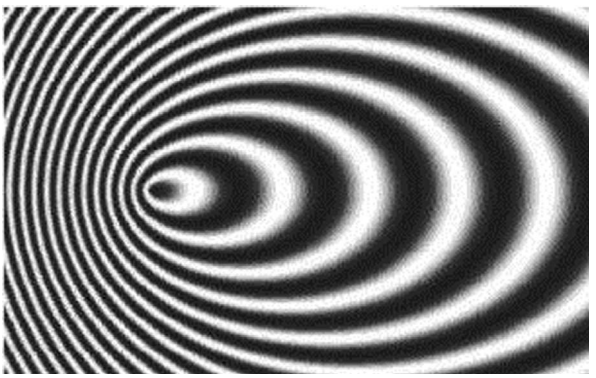


Figure 15.26 A microwave hologram generated as the interference pattern between a point source and a plane wave propagating at 60° to the surface.

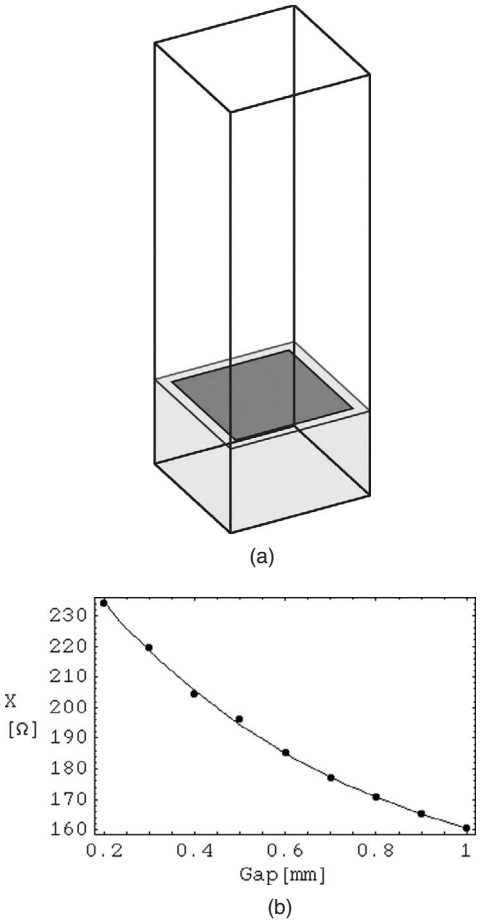


Figure 15.27 (a) The effective surface impedance of the lattice of metal regions is simulated with a single unit cell, consisting of a small metal patch on a grounded dielectric slab. Periodic boundary conditions are imposed on the opposing pairs of vertical walls. (b) Calculated reactance of the square unit cell with a period of 3 mm, on a 1.57-mm thick slab of dielectric constant 2.2, as a function of the gap between neighboring unit cells.

the monopole) into the desired wave (a plane wave at 60°). For comparison, a monopole on a similar smooth metal surface produces the expected pattern, shown on the same plot.

Holographic artificial impedance surfaces can be used to design antennas on curved shapes to have radiation properties that would ordinarily be impossible with a smooth metal surface. To illustrate this concept, a holographic pattern was designed to enable a monopole antenna on a conducting cylinder to produce a narrow beam on the other side of the cylinder, toward a direction that is otherwise shadowed. The monopole antenna generates surface currents that propagate around the cylinder, and a holographic impedance surface was designed using the interference pattern formed by those surface currents, and a plane wave at 135° on the opposite side of the cylinder, as shown in Figure 15.30. As the surface currents propagate around the cylinder, they are scattered by the modulated impedance surface. The scattered radiation experiences constructive interference at 135°,

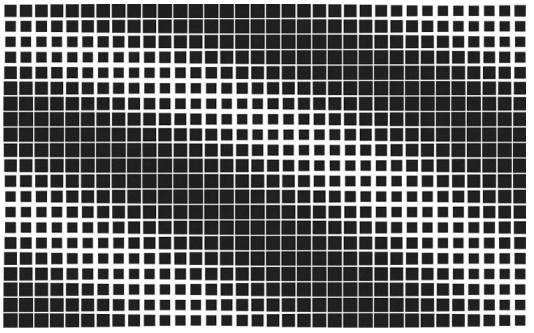


Figure 15.28 A small section of the artificial impedance surface showing a nonuniform lattice of metal squares. The squares are small compared to the wavelength of interest, so the wave only sees the effective surface impedance of the artificial medium.

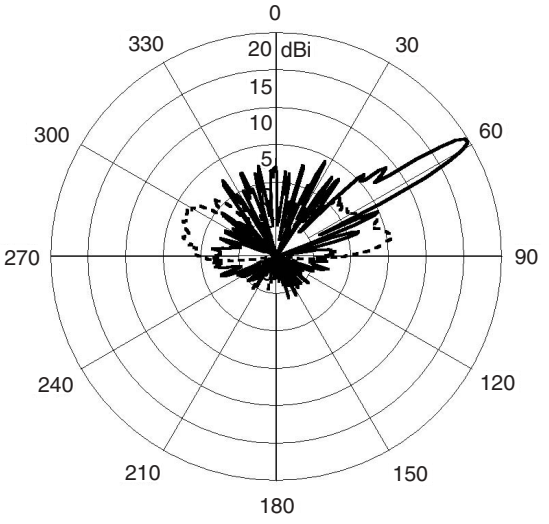


Figure 15.29 Radiation pattern of a monopole antenna on the holographic artificial impedance surface (solid curve) and on a flat conducting surface of the same size (dashed curve). The patterned surface transforms the monopole into a highly directive antenna.

the angle for which the surface was designed, to produce a narrow beam on the other side of the cylinder. This is shown in Figure 15.31, along with the radiation pattern of a similar monopole antenna on an ordinary conducting cylinder. Thus holographic artificial impedance surfaces enable conformal antennas on complex shapes to produce radiation patterns that would otherwise be difficult to achieve.

In order to design antennas for large, complex objects, it is necessary to have tools that can calculate the radiation pattern using the effective surface impedance, rather than the details of the individual unit cells. To do otherwise would require large computational resources that would make problems of useful sizes impractical to calculate. Radiation patterns from artificial impedance surfaces have been computed using software such as

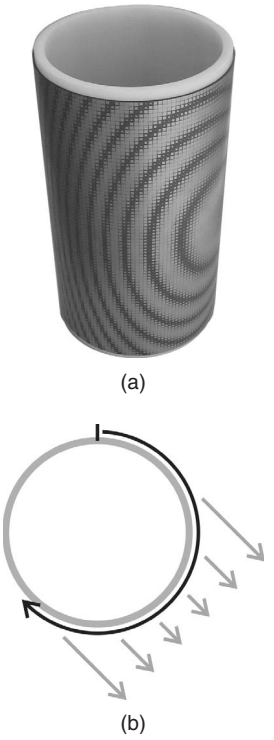


Figure 15.30 (a) A cylinder covered with an artificial impedance surface that is designed to produce radiation on the other side of the cylinder. (b) A diagram showing the propagation of the surface wave and the radiation produced by the surface wave.

the FastScat electromagnetic scattering code [69, 70]. FastScat is a frequency domain boundary integral code using high order Nyström discretizations and the fast multipole method. Impedance boundary conditions for varying impedance surfaces are handled by a suitable extension of the usual boundary integral operators; for example, the extended combined field integral equation has two additional integral operator terms that include the effects of arbitrary surface impedance profiles. Surface impedance profiles are specified by functional form; hence a holographic impedance surface is fully characterized by the functional forms of the desired radiation and surface waves. The modeled surfaces are then the smoothly varying approximations (as shown in Figure 15.26) to the constructed patched surfaces (as shown in Figure 15.28). The FastScat simulation results for the cylinder shown in Figure 15.30 are plotted in Figure 15.32. A simple point dipole above the cylindrical surface is used to model the monopole excitation. Good agreement is found between the simulated and measured results for both holographic impedance and conducting cylinders. This is important for two reasons: (1) it shows that the effective surface impedance is sufficient to describe the surface properties to allow for accurate prediction of the antenna performance, and (2) because this method does not require the simulation of the detailed structure of the surface, it can be used to simulate large objects with practical computing resources. Thus the holographic artificial impedance surface method can be used to design thin printed antennas on large complicated objects with a high degree of confidence.

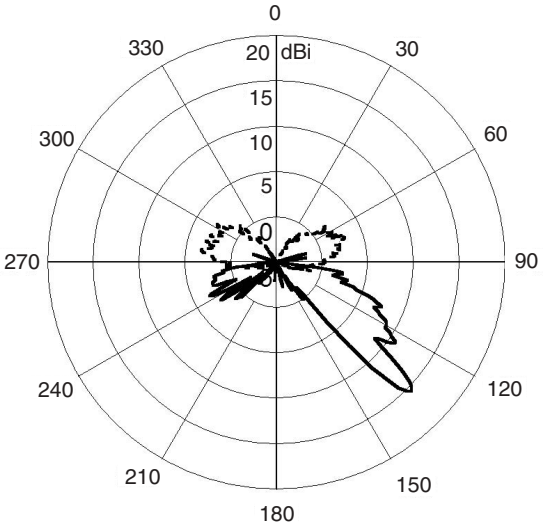


Figure 15.31 Measured radiation pattern of a point dipole antenna above the patterned cylinder (solid curve) and on a smooth conducting cylinder (dashed curve). On the holographic artificial impedance surface, radiation from the monopole antenna couples into a surface wave that wraps around the cylinder and radiates in a controlled manner to produce a highly directive beam on the other side, at 135°.

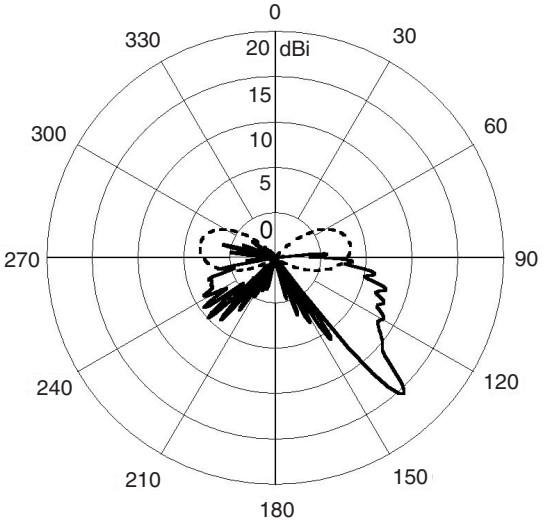


Figure 15.32 Calculated radiation pattern of a point dipole antenna above the patterned cylinder (solid curve) and on a smooth conducting cylinder (dashed curve). The calculated result agrees with the measurement shown in Figure 15.31.

15.15 SUMMARY

Subwavelength textures can be applied to metal surfaces to change their electromagnetic properties. These include corrugated structures to produce hard or soft boundary conditions and a variety of two-dimensional structures. Thin coatings containing lattices of grounded metal plates can behave as a high impedance surface and can be analyzed using a simple lumped circuit parameter model. These surfaces have two important properties: (1) they suppress the propagation of surface waves within a bandgap, and (2) they provide a reflection phase of zero at the resonance frequency. The bandwidth of these properties is related to the thickness of the surface. High impedance surfaces can be used for a variety of antenna applications, such as to suppress scattering of surface waves by nearby structures, or to build various kinds of low profile antennas.

Electronically tunable impedance surfaces can be built by incorporating varactor diodes into the lattice. These can be used as electronically steerable reflectors, for low cost beam steering applications. They can also be used as steerable leaky wave antennas, by incorporating a conformal feed. Leaky waves can be steered over a wide range of angles in both the forward and backward directions. Backward leaky waves can be understood in terms of backward bands, the presence of which can be measured directly. These backward bands are similar to those produced by other means in various other kinds of metamaterials.

The concept of holography can also be applied to artificial impedance surfaces to allow antennas with novel radiation properties to be formed into curved shapes. Some of the important capabilities of these antennas include the ability to achieve high gain toward angles that are otherwise shadowed. This technique can be used to build high performance conformal antennas with arbitrary radiation patterns on complex objects.

REFERENCES

1. L. Brillouin, Wave guides for slow waves, *J. Appl. Phys*., Vol. 19, No. 11, pp. 1023–1041, November 1948.
2. W. Rotman, A study of single-surface corrugated guides, *Proc. IRE*, Vol. 39, pp. 952–959, August 1951.
3. R. Elliot, On the theory of corrugated plane surfaces, *IRE Trans. Antennas Propag*., Vol. 2, pp. 71–81, April 1954.
4. A. Harvey, Periodic and guiding structures at microwave frequencies, *IRE Trans*., Vol. 8, pp. 30–61, June 1959.
5. P.-S. Kildal, Artificially soft and hard surfaces in electromagnetics, *IEEE Trans. Antennas Propag*., Vol. 38, pp. 1537–1544, June 1990.
6. S. Lee, and W. Jones, Surface waves on two-dimensional corrugated surfaces, *Radio Sci*., Vol. 6, pp. 811–818, August 1971.
7. R. King, D. Thiel, and K. Park, The synthesis of surface reactance using an artificial dielectric, *IEEE Trans. Antennas Propag*., Vol. 31, pp. 471–476, May 1983.
8. D. Sievenpiper, High-impedance electromagnetic surfaces, Ph.D. dissertation, Department of Electrical Engineering, University of California, Los Angeles, 1999.
9. D. Sievenpiper, L. Zhang, R. Broas, N. Alexopolous, and E. Yablonovitch, High-impedance electromagnetic surfaces with a forbidden frequency band, *IEEE Trans. Microwave Theory Tech*., Vol. 47, pp. 2059–2074, November 1999.

10. E. Yablonovitch, Inhibited spontaneous emission in solid-state physics and electronics, *Phys. Rev. Lett*., Vol. 58, pp. 2059–2062, May 1987.
11. J. Joannopoulos, R. Meade, and J. Winn, *Photonic Crystals: Molding the Flow of Light*, Princeton University Press, Princeton, NJ, 1995.
12. W. Barnes, T. Priest, S. Kitson, and J. Sambles, Physical origin of photonic energy gaps in the propagation of surface plasmons on gratings, *Phys. Rev. B*, Vol. 54, pp. 6227–6244, September 1996.
13. F. Yang, K. Ma, Y. Qian, and T. Itoh, A novel TEM waveguide using uniplanar compact photonic-bandgap (UC-PBG) structure, *IEEE Trans. Microwave Theory Tech.*, Vol. 47, No. 11, pp. 2092–2098, November 1999.
14. A. Monorchio, G. Manara, and L. Lanuzza, Synthesis of artificial magnetic conductors by using multilayered frequency selective surfaces, *IEEE Antennas Wireless Propag. Lett.*, Vol. 1, No. 11, 196–199, 2002.
15. D. J. Kern, D. H. Werner, A. Monorchio, L. Lanuzza, and M. J. Wilhelm, The design synthesis of multiband artficial magnetic conductors using high impedance frequency selective surfaces, *IEEE Trans. Antennas Propag.*, Vol. 53, No. 1, Jan 2005, pp. 8–17. January 2005.
16. C. R. Simovski, P. De Maagt, and I. V. Melchakova, High-impedance surface having stable resonance with respect to polarization and incidence angle, *IEEE Trans. Antennas Propag.*, Vol. 53, No. 3, pp. 908–914, March 2005.
17. D. Sievenpiper, J. Schaffner, R. Loo, G. Tangonan, S. Ontiveros, and R. Harold, A tunable impedance surface performing as a reconfigurable beam steering reflector, *IEEE Trans. Antennas Propag.*, Vol. 50, pp. 384–390, March 2002.
18. D. Sievenpiper, J. Schaffner, H. J. Song, R. Loo, and G. Tangonan, Two-dimensional beam steering reflector using an electrically tunable impedance surface, *IEEE Trans. Antennas Propag*., Vol. 51, pp. 2713–2722, October 2003.
19. D. Sievenpiper, Forward and backward leaky wave radiation with large effective aperture from an electronically tunable textured surface, *IEEE Trans. Antennas Propag*., Vol. 53, pp. 236–247, January 2005.
20. S. Ramo, J. Whinnery, and T. Van Duzer, *Fields and Waves in Communication Electronics*, 2nd ed., John Wiley & Sons, Hoboken, NJ, 1984.
21. H. Raether, *Surface Plasmons on Smooth and Rough Surfaces and on Gratings*, Springer-Verlag, New York, 1988.
22. L. Brillouin, *Wave Propagation in Periodic Structures*, McGraw-Hill, New York, 1946.
23. N. Ashcroft, and N. Mermin, *Solid State Physics*, Saunders College Publishing, Orlando, FL, 1976.
24. H. Wheeler, Fundamental limitations of small antennas, *Proc. IRE*, Vol. 35, pp. 1479–1484, December 1947.
25. L. Chu, Physical limitations of omni-directional antennas, *J. Appl. Phys*., Vol. 19, pp. 1163–1175, December 1948.
26. J. McLean, A re-examination of the fundamental limits on the radiation Q of electrically small antennas, *IEEE Trans. Antennas Propag*., Vol. 44, pp. 672–675, May 1996.
27. S. Tretyakov, and C. Simovski, Dynamic model of artificial reactive impedance surfaces, *J. Electromagnetic Waves Applications*, Vol. 17, pp. 131–145, 2003.
28. D. Sievenpiper, J. Schaffner, and J. Navarro, Axial ratio improvement in aperture antennas using a high-impedance ground plane, *Electron. Lett*., Vol. 38, No. 23, pp. 1411–1412, November 2002.
29. D. G. Berry, R. G. Malech, and W. A. Kennedy, The reflectarray antenna, *IEEE Trans. Antennas Propag*., Vol. 11, pp. 645–651, November 1963.

30. R. D. Javor, X.-D. Wu, and K. Chang, Design and performance of a microstrip reflectarray antenna, *IEEE Trans. Antennas Propag.*, Vol. 43, No. 9, pp. 932–939, November 1995.

31. D. C. Chang, and M. C. Huang, Multiple polarization microstrip reflectarray antenna with high efficiency and low cross-polarization, *IEEE Trans. Antennas Propag.*, Vol. 43, pp. 829–834, August 1995.

32. D. M. Pozar, S. D. Targonski, and H. D. Syrigos, Design of millimeter wave microstrip reflectarrays, *IEEE Trans. Antennas Propag.*, Vol. 45, No. 2, pp. 187–296, February 1997.

33. M. E. Bialkowski, and H. J. Song, Investigations into a power combining structure using a reflect array of dual-feed aperture coupled microstrip patch antennas, *IEEE Trans. Antennas Propag.*, Vol. 50, pp. 841–849, June 2002.

34. R. Waterhouse, and N. Shuley, Scan performance of infinite arrays of microstrip patch elements loaded with varactor diodes, *IEEE Trans. Antennas Propag.*, Vol. 41, pp. 1273–1280, September 1993.

35. W. Lam, C. Jou, H. Chen, K. Stolt, N. Luhmann, and D. Rutledge, Millimeter wave diode grid phase shifters, *IEEE Trans. Microwave Theory Tech.*, Vol. 36, pp. 902–907, May 1988.

36. L. B. Sjogren, H. X. Liu, X. Qin, C. W. Domier, and N. C. Luhmann, Phased array operation of a diode grid impedance surface, *IEEE Trans. Antennas Propag.*, Vol. 42, pp. 565–572, April 1994.

37. G. Broussaud, Un nouveau type d'antenne de structure plane, *Anna. Radioelectricite*, Vol. 11, pp. 70–88, January 1956.

38. J. W. Lee, J. J. Eom, K. H. Park, and W. J. Chun, TM-wave radiation from grooves in a dielectric-covered ground plane, *IEEE Trans. Antennas Propag.*, Vol. 49, No. 1, pp. 104–105, January 2001.

39. C.-N. Hu, and C. -K. C. Tzuang, Analysis and design of large leaky-mode array employing the coupled-mode approach, *IEEE Trans. Microwave Theory Tech.*, Vol. 49, pp. 629–636, April 2001.

40. P. W. Chen, C. S. Lee, and V. Nalbandian, Planar double-layer leaky wave microstrip antenna, *IEEE Trans. Antennas Propag.*, Vol. 50, pp. 832–835, June 2002.

41. T. Tamer, and F. Kou, Varieties of leaky waves and their excitation along multilayered structures, *IEEE J. Quantum Electron.*, Vol. 22, pp. 544–551, April 1986.

42. M. Guglielmi, and D. Jackson, Broadside radiation from periodic leaky-wave antennas, *IEEE Trans. Antennas Propag.*, Vol. 41, pp. 31–37, January 1993.

43. S.-G. Mao, and M.-Y. Chen, Propagation characteristics of finite-width conductor-backed coplanar waveguides with periodic electromagnetic bandgap cells, *IEEE Trans. Microwave Theory Tech.*, Vol. 50, pp. 2624–2628, November 2002.

44. L. Liu, C. Caloz, and T. Itoh, Dominant mode leaky wave antenna with backfire-to-endfire scanning capability, *Electron. Lett.*, Vol.38, No. 23, pp. 1414–1416, November 2002.

45. A. Grbic and G. V. Eleftheriades, Experimental verification of backward-wave radiation from a negative refractive index metamaterial, *J. Appl. Phys.*, Vol. 92, No. 10, pp. 5930–5935, November 2002.

46. D. Smith, W. Padilla, D. Vier, S. Nemat-Nasser, and S. Schultz, Composite medium with simultaneously negative permeability and permittivity, *Phys. Rev. Lett.*, Vol. 84, pp. 4184–4187, May 2000.

47. D. Smith, and N. Kroll, Negative refractive index in left-handed materials, *Phys. Rev. Lett.*, Vol. 85, pp. 2933–2936, October 2000.

48. G. Eleftheriades, A. Iyer, and P. Kremer, Planar negative refractive index media using periodically loaded *L-C* transmission lines, *IEEE Trans. Microwave Theory Tech.*, Vol. 50, pp. 2702–2712, December 2002.

49. L. Liu, C. Caloz, and T. Itoh, Dominant mode leaky-wave antenna with backfire to endfire scanning capability, *Electron. Lett.*, Vol. 38, pp. 1414–1416, November 2002.
50. D. Sievenpiper, J. Colburn, B. Fong, J. Ottusch, and J. Visher, Holographic artificial impedance surfaces for conformal antennas, in *Antenna Applications Symposium Digest*, Monticello, Illinois, 21, September 2005.
51. D. Sievenpiper, J. Colburn, B. Fong, J. Ottusch, and J. Visher, Holographic artificial impedance surfaces for conformal antennas, in *IEEE APS Symposium Digest*, Washington DC, 5, July 2005.
52. J. Simon, and V. Biggi, Un noveau type d'aerien et son applicatoin a la transimssion de television a grand distance, *L'Onde Electrique*, No. 332, pp. 1–16, November 1954.
53. A. Thomas, and F. J. Zucker, Radiation from modulated surface wave structures I, *IRE Natl. Conv. Rec.*, Pt. 1, pp. 153–160, March 1957.
54. R. Pease, Radiation from modulated surface wave structures II, *IRE Natl. Conv. Rec.*, Pt 2. pp. 161–165, March 1957.
55. L. Felsen, Radiation from a tapered surface wave antenna, *IRE Trans. Antennas Propag.*, Vol. 8, No. 6, pp. 577–586, November 1960.
56. A. Oliner, and A. Hessel, Guided waves on sinusoidally-modulated reactance surfaces, *IRE Trans. Antennas Propag.*, Vol. 7, No. 5, pp. 201–208, December 1959.
57. R. Dooley, X-band Holography, *Proc. IEEE*, Vol. 53, No. 11, pp. 1733–1735, November 1965.
58. W. Kock, Microwave holography, *Microwaves*, Vol. 7, No. 11, pp. 46–54, November 1968.
59. P. Checcacci, V. Russo, and A. Scheggi, Holographic antennas, *IEEE Trans. Antennas Propag.*, Vol. 18, No. 6, pp. 811–813, November 1970.
60. K. Iizuka, M. Mizusawa, S. Urasaki, and H. Ushigome, Volume-type holographic antenna, *IEEE Trans. Antennas Propag.*, Vol. 23, No. 6, pp. 807–810, November 1975.
61. D. Sazonov, Computer aided design of holographic antennas, *Antennas Propag. Soc. Int. Symp.*, Vol. 2, pp. 738–741, July 1999.
62. K. Levis, A. Ittipiboon, A. Petosa, L. Roy, and P. Berini, Ka-band dipole holographic antennas, *IEE Proc. Microwaves Antennas Propag.*, Vol. 148, No. 2, pp. 129–132, April 2001.
63. M. ElSherbiny, A. Fathy, A. Rosen, G. Ayers, and S. Perlow, Holographic antenna concept, analysis, and parameters, *IEEE Trans. Antennas Propag.*, Vol. 52, No. 3, pp. 830–839, March 2004.
64. N. Farhat, Holographically steered millimeter wave antennas, *IEEE Trans. Antennas Propag.*, Vol. 28, No. 4, pp. 476–480, July 1980.
65. A. Fathy, A. Rosen, H. Owen, F. McGinty, D. McGee, G. Taylor, R. Amantea, P. Swain, S. Perlow, and M. ElSherbiny, Silicon-based reconfigurable antennas—concepts, Analysis, implementation and feasibility, *IEEE Trans. Microwave Theory Tech.*, Vol. 51, No. 6, pp. 1650–1661, June 2003.
66. S. Thingvold, A. Ittipiboon, A. Sebak, and A. Petosa, Holographic antenna efficiency, *Antennas Propag. Soc. Int. Symp.*, Vol. 3, pp. 721–724, June 2003.
67. V. Sanchez, and R. Diaz, unpublished work.
68. R. King, D. Thiel, and K. Park, The synthesis of surface reactance using an artificial dielectric, *IEEE Trans. Antennas Propag.*, Vol. 31, No. 3, pp. 471–476, May 1983.
69. L. F. Canino, J. J. Ottusch, M. A. Stalzer, J. L. Visher, and S. M. Wandzura, Numerical solution of the helmholtz equation in 2D and 3D using a high-order nystrom discretization, *J. Comput. Phy.*, Vol. 146, pp. 627–663, 1998.
70. J. J. Ottusch, M. A. Stalzer, J. L. Visher, and S. M. Wandzura, Scalable electromagnetic scattering calculations on the SGI Origin 2000, in *Proceedings of Supercomputing Conference*, 1999.

CHAPTER 16

Frequency-Selective Screens

THOMAS CWIK

16.1 INTRODUCTION

Frequency-selective screens play an important role in a wide range of antenna systems. The fundamental characteristic of a frequency-selective screen is that it will transmit nearly all energy at some specified frequency band while reflecting nearly all energy at another band. Based on the requirements of a design, the frequency bands may be tailored to be quite wide or narrow, have a fast roll-off, have multiple pass bands or stopbands, or be polarization dependent, among other parameters. The screen consists of a two-dimensional array of aperture elements in a metal sheet or an array of metal patch elements fabricated on dielectric layers. The size, shape, and spacing of the apertures or patches dictate the frequency response of the screen. The layers can be sandwiched in the third dimension to further tailor the frequency response of the screen. Aperture and patch screens generally give complementary frequency responses—the field will be nearly totally transmitted at the resonant frequency of the aperture while a screen comprised of patches will be nearly totally reflective at the resonant frequency of the patches. Frequency-selective screens are resonant structures—they operate because the fields across the two-dimensional aperture or metal element become resonant, adding perfectly in phase at a specific frequency, to reflect all energy (for a patch screen) or transmit all energy (for an aperture screen). This powerful filtering property is what allows a patch screen, that may have only a small fraction of the screen covered with metal, to reflect all energy at the resonant frequency.

Similarly for an aperture screen, all energy can be transmitted even though only a small fraction of the metal screen is comprised of apertures. This fundamental resonant property is used in the design of frequency-selective screens to match the frequency response to the desired application.

The screen can operate in the microwave portion of the spectrum—the lower frequency of operation is only dictated by the physical size of the screen needed to allow the fields across it to be resonant. Similarly for high frequencies, the only limitation is the manufacturing needed to fabricate the apertures or patches on a layer accurately enough to be resonant at the required frequency. Shown in Figure 16.1 is the reflection response for a dipole patch element array shown on the left-hand side of the figure.

Modern Antenna Handbook. Edited by Constantine A. Balanis

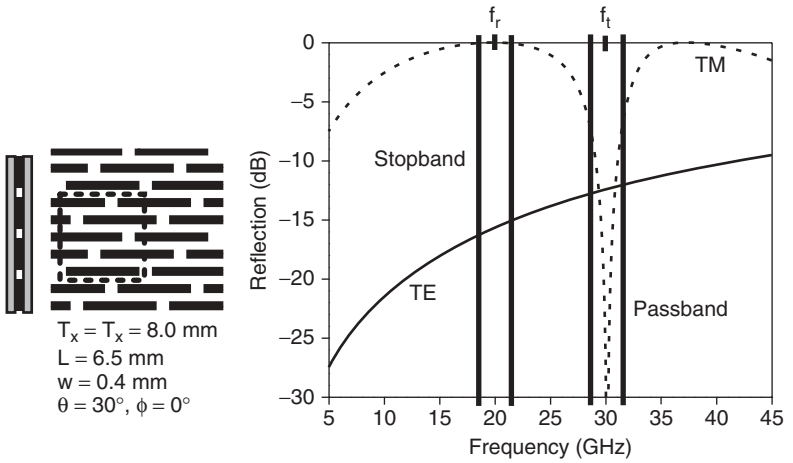


Figure 16.1 Frequency response of dipole array described in Section 16.3. The passband centered at the reflection frequency near 20 GHz and the passband centered at the transmit frequency near 30 GHz are shown for the TE (perpendicular) and TM (parallel) polarizations. (TE and TM are transverse to the screen normal.) The dielectric layers surrounding the dipole patch array are 10 mils thick with relative dielectric constant 2.2.

The reflection response for the electric field polarized parallel and perpendicular to the screen is shown. This figure outlines the general properties of a screen: there are passbands and stopbands centered around a transmit frequency f_t and a reflection frequency f_r. Bandwidths are defined for the passbands and stopbands, and typically the ratio f_t/f_r is a driving parameter in the design.

In this chapter the term frequency-selective screen is used to describe the complete patch or aperture array, or arrays if they are sandwiched, and all supporting dielectric layers. The term frequency-selective surface is used to describe a patch array and supporting dielectric layer (e.g., a thin kapton layer below and above the array). Typically, a screen is then a frequency-selective surface with a thick supporting dielectric layer, or multiple frequency-selective surfaces separated by thicker dielectric spacer layers. This difference is defined since it is natural to break a design down into the response of a frequency-selective surface with its array spacing, element size, and shape, and then refine the design to include supporting layers or multiple surfaces. The term dichroic plate is commonly used for an aperture array in a metal surface, perhaps with some appreciable electrical thickness.

Section 16.2 in this chapter outlines the wide variety of applications of frequency-selective screens. The applications result in a set of requirements that will drive the design. They are varied and full, extending across the electromagnetic spectrum, and require a range of materials to be used. Section 16.3 examines the design parameters of a frequency-selective screen. This section breaks down the parameters available for design of a screen and how they are used to produce the required spectral response. Array spacing, array element types, effects of dielectric layers, and the design of multilayer surfaces are described. Section 16.4 gives a brief overview of the design and analysis methods used in frequency-selective screens. Specific elements of the methods that impact the designer are pointed out. Finally, in Section 16.5, a few common design objectives

that extend standard practice are outlined. Described are the performance modifications of the screen due to a nonuniform (feedhorn) field, curved subreflector screens, and operation in the submillimeter and infrared portion of the spectrum.

16.2 SYSTEM REQUIREMENTS FOR CLASSES OF FREQUENCY-SELECTIVE SCREENS

The design of a frequency-selective screen is fundamentally driven by requirements of the system where it is applied. The screen is used to combine or filter multiple frequency bands across the spectrum. Depending on the specific application, the system design will include one or more frequency-selective screens that must meet a range of requirements including bandwidth (reflecting or transmitting a prescribed amount of energy across the band), dissipative loss, polarization, angular range of incident and reflected energy (such as from a feedhorn), environmental effects (space or terrestrial), amplitude and phase variation (as in polarizers), power handling, out-of-band performance (e.g., in stealth RCS applications), multimode use (as in diffraction gratings), and manufacturability among others. The range of applications extends across the electromagnetic spectrum, from the microwave to applications in the infrared where screens are fabricated using micromachining capabilities, from telecommunications to applications in sensitive scientific measurements, as antenna radomes, and for controlling antenna radiation. Generally, frequency-selective screens can be classified by application that dictates the requirements used to design the screen. The following sections categorize a range of applications and requirements.

16.2.1 Reflector Antenna Frequency Reuse

Traditionally, the most common frequency-selective screen application is derived from the need to reuse a common reflective antenna aperture across multiple bands. A frequency-selective screen is a simple alternative to a multifrequency feedhorn or can be used in conjunction with multifrequency feedhorns to combine a larger number of bands. The geometry of the antenna system dictates the arrangement of the aperture antenna, screens, and feed systems. Usually, a complex trade study is completed to design an optimum configuration given the system center frequencies and bandwidths, amount of insertion and dissipative losses allowed at the different bands, and space available to the antenna system. Two fundamental configurations are (1) a beam-waveguide system using planar frequency-selective screens in the optical path to combine or filter frequency bands, and (2) a curved frequency-selective subreflector with feed systems behind (typically near the reflector focus) and transmitting energy through the subreflector, and a second set of feeds in front (typically near the reflector apex) and reflecting energy off the subreflector.

16.2.1.1 Beam-Waveguide System An example of a microwave beam-waveguide system is found in the NASA Deep Space Network. Shown in Figure 16.2 is the NASA DSS 13 beam-waveguide antenna at Goldstone, California. This antenna—and later systems [1]—was designed for efficient multifrequency use of the 34-m Cassegrain system, guiding energy reflected from the solid subreflector to a pedestal room built underground to house further reflectors and electronics. Being used for deep space

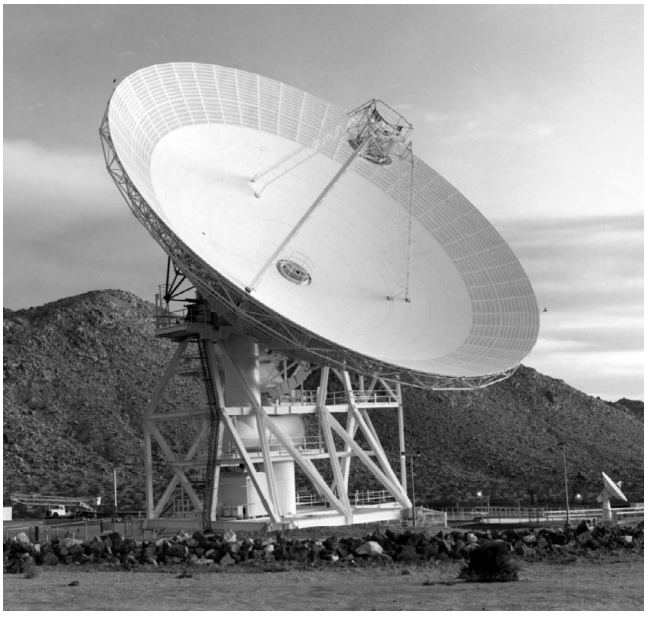
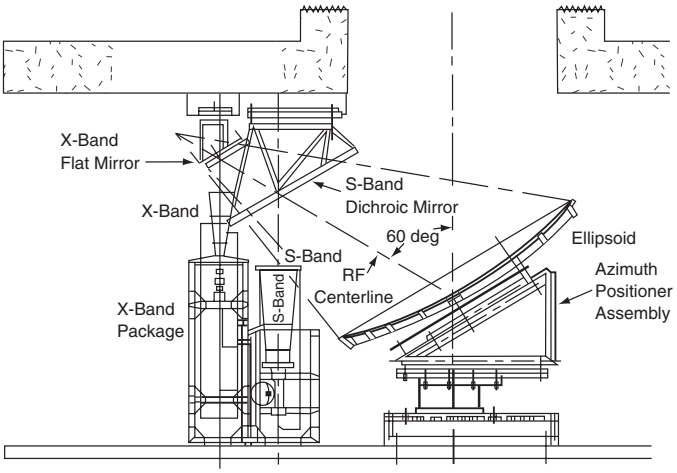


Figure 16.2 NASA DSS 13 beam-waveguide antenna at Goldstone, California. The pedestal room is underground, below the antenna. (From Ref. 1.)

communications, the requirements were driven by reuse of the S (2–2.3 GHz), X (7.1–8.5 GHz), and Ka (31–35 GHz) bands, high transmit gain, and very low dissipative loss in the system to maximize the system G/T needed to receive extremely weak spacecraft signals. Shown in Figure 16.3 is the geometry of the underground pedestal room beam-waveguide mirrors for the S/X configuration with the feed system and dichroic plate used to reflect the S-band signal to the S-band feed, and transmit the X-band signal to a secondary flat mirror and to the X-band feed. The requirements for this dichroic plate were defined to be transparent to the X-band uplink and downlink bands and simultaneously be reflective to S-band uplink and downlink bands. The plate must also be able to transmit 100kW X-band uplink power without arcing. These requirements are summarized in Table 16.1 [2].

This plate had to meet the above requirements for perpendicular and parallel polarizations at 30° nominal incidence, and over a range of angles from 15° to 45° to accommodate the effect of the feedhorn pattern. To maintain mechanical stability and low dissipative loss, a metal plate design with thick rectangular holes was used in DSS 13 and later antennas. The rectangular apertures allowed two degrees of freedom to compensate the 30° incidence geometry (nonuniform E-field in the hole aperture) [2] and followed the heritage of a circular hole design [3]. This design was then extended to S/X- and X/Ka-band beam-waveguide systems utilizing two dichroic plates at different clock rotations and optical paths in the pedestal room (S/X—reflecting S-band, transmitting X-band as outlined above; and X/Ka—reflecting X-band, transmitting Ka-band along a separate optical path) [4].

A second example of a beam-waveguide system, this time used in a science spaceflight instrument that measures thermal microwave emission from the Earth's limb in order to



(a)

(b)

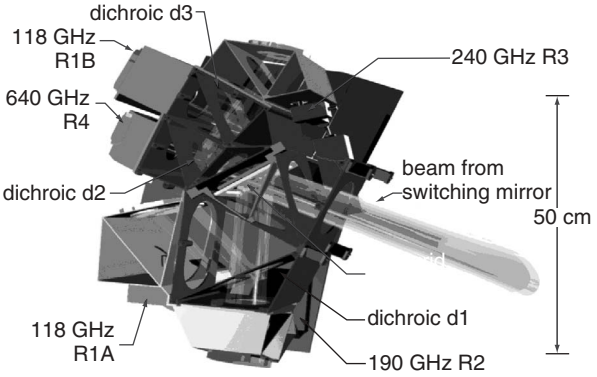
Figure 16.3 (a) DSS 13 S/X-band optics and feed system geometry in pedestal room. (b) Feed system installed in the beam-waveguide antenna showing the dichroic plate, flat reflector, and feed systems. (From Ref. 1.)

TABLE 16.1 Summary of the Design Requirements for S- and X-bands of the S/X Dichroic Plate Used in DSS 13

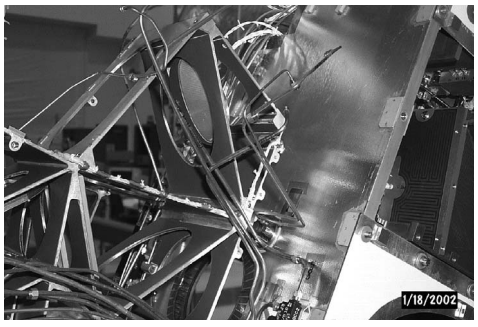
Band (GHz)	Insertion Loss (dB)	Dissipative Loss (dB)
2.090–2.320 S-band	>45 reflective	N/A
7.145–7.190 X-band uplink	<0.04	N/A
8.400–8.450 X-band downlink	<0.04	<0.02

Source: From Ref. 2.

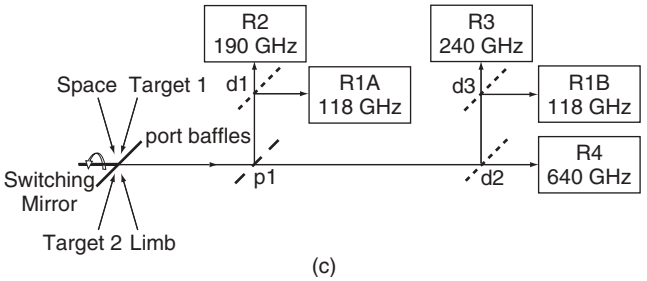
determine composition and temperature of the atmosphere, is shown in Figure 16.4. Here, a set of three dichroic plates on the NASA Microwave Limb Sounder instrument is used to separate 118-GHz, 190-, 240-, and 640-GHz bands of energy reflected from a 1.6 m $\times$ 0.8 m elliptical projection main reflector [5, 6]. The driving requirements were efficient separation of the bands (minimizing the insertion loss to maximize the received thermal emissions of the atmosphere), polarization sensitivity since a wire grid polarizer is also used to separate the signals, and because of the relatively small wavelengths



(a)



(b)

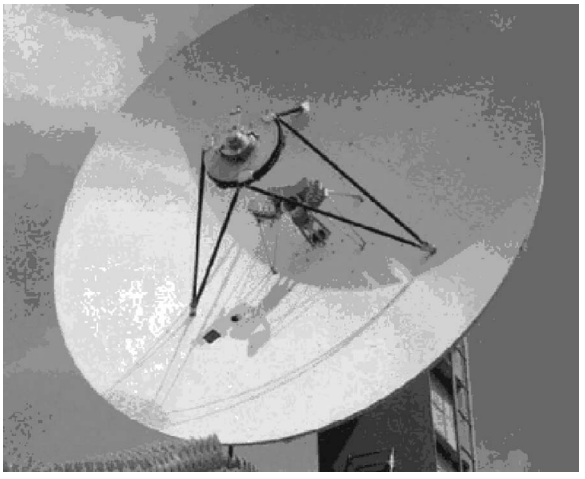


(c)

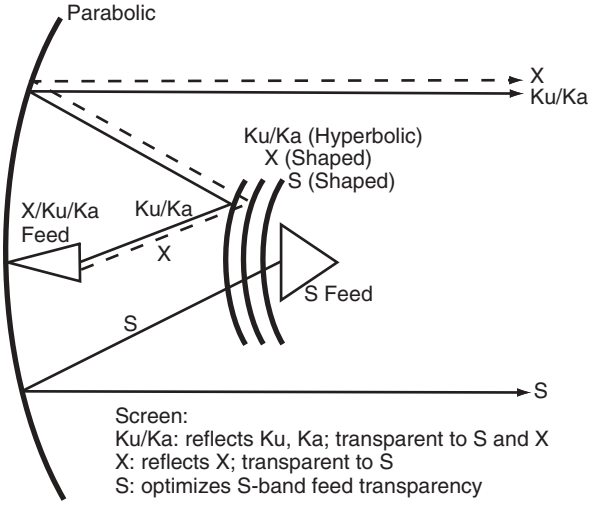
Figure 16.4 (a, b) Instrument layout for Microwave Limb Sounder showing dichroic 1 (separating 180 and 190 GHz), dichroic 2 (separating 118/240 and 640GHz), and dichroic 3 (separating 118 and 240 GHz). R indicates a receiver, mounting fixtures for section of optical path. (c) Schematic of optical path corresponding to instrument layout. (From Ref. 5.)

of operation (2.5–0.4 mm), manufacturability and material selection that are appropriate both for the space environment and withstanding the forces of launch. Once again electrically thick dichroic plates with holes properly designed in low loss metals were fabricated using micromachining and electroforming. The frequency-selective screens, even at these small wavelengths, could be self-supporting and meet the electrical and mechanical requirements of the mission.

16.2.1.2 Curved Frequency-Selective Reflectors A relatively complex example of a frequency-selective screen subreflector is shown in Figure 16.5. This subreflector is part of the NASA Cassini spacecraft antenna system that arrived at



(a)



(b)

Figure 16.5 (a) NASA Cassini spacecraft reflector, feed systems, and frequency-selective screen subreflector mounted on antenna range. (b) Schematic showing optical paths and application of frequency subreflector for Cassini antenna system. (From Ref. 7.)

Saturn in 2004 [7]. The antenna system required S/X/Ku/Ka-band operation for a number of science and telecommunication modes including telecommunications to Earth, relay communications with the Huygens probe that landed on the moon Titan, radio science experiments, and radar mapping science. The Cassegrain antenna system was constrained by the launch vehicle faring volume, launch mechanical constraints, and thermal requirements at Saturn. These constraints, along with the band allocation to the various modes of operation and relative prioritization of modes, led to a four-band subreflector that reflected X/Ku/Ka-bands while transmitting S-band from a feed at the prime focus of the main reflector. A three-frequency feed (X/Ku/Ka) was located at the Cassegrain focus of the system, with a separate Ku-band array feed at the Cassegrain focal plane for radar science. These complex requirements and design options resulted in a three-layer frequency-selective screen reflecting and transmitting the various bands as shown in Figure 16.5. The three surfaces were comprised of layers of ring elements separated by a layer with varying thickness to compensate for the changing incident angle of transmitted energy due to the subreflector curvature. Shaping of the subreflector for optimum antenna system performance at the various bands was also rolled into the design. The measured results of this multiobjective design are tabulated in Table 16.2, where the various losses are shown for the frequency bands.

16.2.2 Radomes

A second common class of frequency-selective surfaces involves radomes that protect an antenna from the environment while presenting a favorable shape to the attached system (like an airplane) and impact the performance of the antenna as little as possible in the band of operation. The radomes are typically constructed from high strength materials chosen for the application, with metalized frequency-selective surfaces embedded within the materials. Key system drivers can include extreme environmental performance as in a high speed missile or jet, where temperature and mechanical stability are essential, aerodynamic shape such as in being conformal to the vehicle or presenting minimum wind resistance on ground-based antenna terminals, and low radar cross-section signature when the radome operates in a mode to be invisible to the antenna system in-band while blending smoothly into the structure for out-of-band operation. With the radome being in

TABLE 16.2 Summary of Design Requirements for Bands Used in Cassini Antenna System

Band Center Frequency (GHz)	Ohmic + T/R Loss[a] (dB)	Amplitude and Phase Losses[b] (dB)	Total (dB)
2.040	0.26	0.14	0.4
2.298	0.16	0.0	0.16
7.175	0.51	0.08	0.59
8.425	0.23	0.16	0.39
13.7765	0.65	—	0.65
32.028	0.55	—	0.55
34.316	0.89	—	0.89

[a] T/R indicates transmission/reflection losses.

[b] Amplitude and phase losses are those of the primary antenna pattern calculated from using measured subreflector pattern in the primary antenna calculation.

Source: From Ref. 7.

the near vicinity of the antenna, mutual interactions of the antenna and frequency-selective screen complicate the design. In reduced radar cross-section applications, reflection from the enclosed antenna system in the band of operation must be considered as part of the overall system requirements. Key to radome designs that involve frequency-selective screens is the need to consider the effects of curved surfaces. This includes the proper packing geometry on a curved substrate as well as the effect of the varying incidence angle due to the surface curvature. A thorough overview of radome design and performance can be found in Ref. 8 and will not be detailed in this chapter.

16.2.3 Diffraction Gratings and Frequency-Scanned Antennas

Historically, frequency-selective surfaces followed from the diffraction gratings widely used in optics to disperse different colors of an incident beam. The simplest gratings are made from single periodic structures of period T and some profile as shown in Figure 16.6. Different surface profiles are chosen based on the application and manufacturing process. The reflective grating will reflect light into its diffractive orders according to the grating equation

$$\sin\theta_m = \sin\theta_{\text{inc}} + m(\lambda/T) \tag{16.1}$$

where θ_m is the angle of reflection of the mth diffracted order, θ_{inc} is the angle of incidence, and λ is the wavelength of the assumed incident plane wave. Depending on the actual extent of the grating, the shape of the periodic profile, and the range of wavelengths relative to the spacing, the reflected fields will have a diffraction efficiency (intensity in one order relative to that of the incident wave) and spectral response. A similar transmissive grating can be fabricating using dielectric materials.

This concept has been extended to the microwave and millimeter wave spectrum to design a frequency-scanned antenna [9]. In a simplified two-dimensional antenna explained using the above grating equation, a surface can be designed with periodic profile and period T to reflect energy into only the $m = -1$ diffracted order, minimizing energy that reflects into the specular direction ($m = 0$). For this specific design the grating equation shows that a change in wavelength will scan the beam according to the relation

$$\theta_{-1} = \sin^{-1}[\sin\theta_{\text{inc}} - (\lambda/T)] \tag{16.2}$$

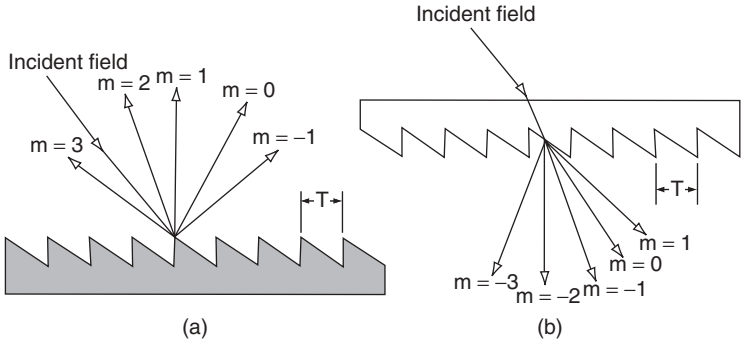


Figure 16.6 (a) Reflective metal grating and (b) transmission dielectric grating with incident single frequency incident field and diffracted harmonics.

In Ref. 9, this approach was developed for a three-dimensional antenna system comprising a two-dimensional planar frequency-selective surface made of concentric rings backed by a reflective groundplane. A feedhorn shining on the surface is used to radiate the structure. The period of the frequency-selective surface slowly varies away from the central region to provide a common focal point distance mimicking the geometry of a parabolic reflector. Here the system required a complex design with precise element spacing, operation over a range of incident angles, and frequency band to produce the appropriate range of antenna scan.

16.2.4 Meanderline Polarizers

Yet another variation of the use of periodic screens is to use the relative phase difference between polarizations transmitted through a screen to transform a linear polarized field into one being circularly polarized. A "meandering" thin metal trace can shift the phase of one polarization relative to the opposite polarization with little reflection for either polarization. By cascading several surfaces appropriately separated, one polarization is shifted 90° relative to the other.

Shown in Figure 16.7 is a meanderline polarizer used to convert linear polarization into circular polarization for a waveguide slot array-fed waveguide lens array used to illuminate a satellite dual reflector system [10]. It is intended for the frequency band of 12.2–12.7 GHz. The specified polarization purity is greater than 25 dB and the specified return loss should be greater than 25 dB also. The size of the polarizer outline measures

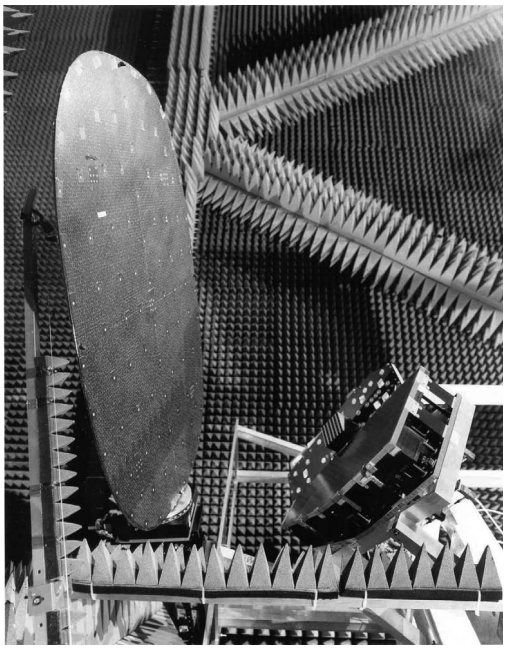


Figure 16.7 Meanderline polarizer mounted on feed array in near-field range with subreflector. The reflector is shown to the left and the meanderline polarizer and feed array are to the lower right. (From Ref. 12.)

approximately 30 by 15 inches, with a pair of 5-inch square cutouts providing clearance for dual-polarized horns to "look through" the polarizer.

The three meanderline sheets shown in Figure 16.8 were optimized using a genetic algorithm with a cost function included in the return loss and polarization purity. Each patterned surface rotated one polarization relative to the other a fixed amount. The combination of three surfaces along with the spacing layers then gave the 90° phase shift for this application. In this example the transmitted amplitude and phase were evaluated using general design and analysis software [11, 12].

16.2.5 Active Surfaces and Frequency Screening

The frequency-filtering property of frequency-selective screens can naturally lead to requirements on time-varying or reconfigurable surfaces that allow the spectral response to be rapidly switched in time as needed. The most direct implementation of this involves the use of semiconductor switches integrated with the elements to change their electrical shape and thus the spectral response. An example is shown in schematic form in Figure 16.9, where PIN diodes are placed in the path of square loop elements. The bottom and top portions of the loops are biased with a voltage [13]. By appropriately

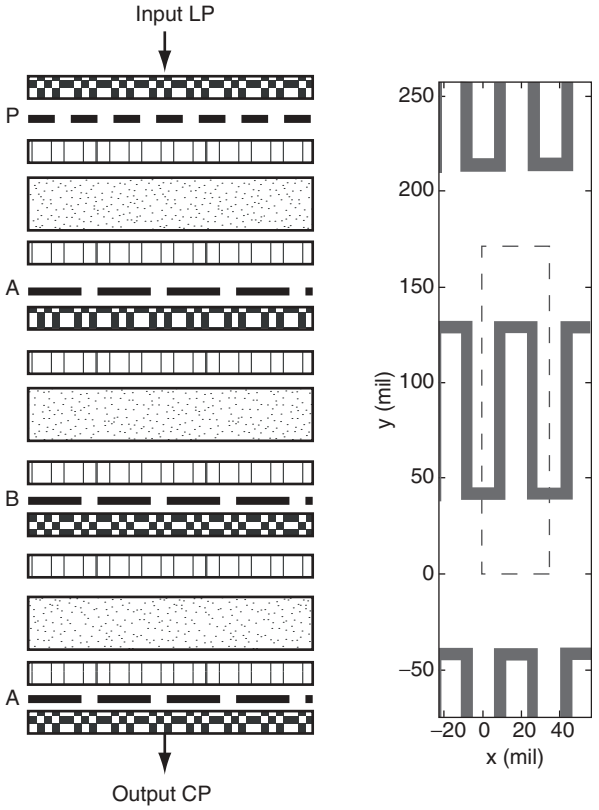


Figure 16.8 Schematic of showing meanderline polarizer surfaces (A and B) with metal trace (shown on right) and dielectric spacer and face layers. Surfaces A and B have slightly different dimensions. The top surface (P) indicates a linear polarizing strip grid.

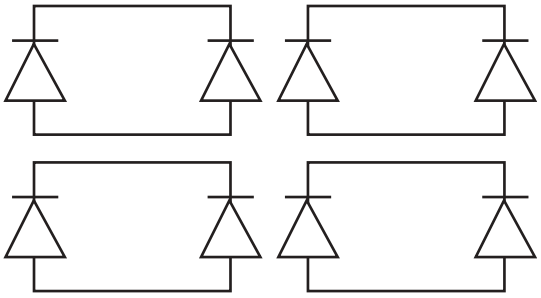


Figure 16.9 Schematic of rectangular ring array with PIN diodes in arms to allow upper and lower arms to be connected electrically or open circuited. Transmission response can be switched by biasing upper and lower arms to turn on and off the diodes. (Adapted from Ref. 13.)

changing the voltage on the loops, the PIN diodes can be made to conduct and present a short circuit or be biased off and present an open circuit to the attached metal arms. The screen now requires bias lines to all elements, fabricated in a way that does not disturb the electromagnetic response of the frequency-selective surface. Additional losses, introduced due to the semiconductor materials attached to the elements, and power and cooling requirements must be considered in the design. Microelectromechanical system (MEMS) switches can also be used in the fabrication process to provide an integrated solution [14].

Another driving requirement for active surfaces comes from the need to control the transmission of unwanted radiation into or out of a room or building. As in the household microwave oven that has a frequency-selective door that is transparent to visible light while reflecting the microwave energy inside, frequency-selective walls coated with frequency-selective "wallpaper" can be used to selectively block the transmission of signals that interfere with electronic equipment. These active surfaces can be used to allow real-time reconfigurability of the transmission properties if manufacturing requirements can be met [15].

16.2.6 Microwave Absorber Arrays

Another traditional use of periodic structures is the microwave absorber used in anechoic chambers. Sensitive antenna and scattering measurements lead to requirements that the wall material must absorb energy across wide-frequency bands and at wide-ranging incident angles. These requirements produced designs of carbon-loaded foam structures with wedge (single periodicity) or pyramid (doubly periodic) arrays that cover the chamber walls. Rather than seeking total transmission or reflection across a band, these requirements lead to total absorption across the usable frequencies of the chamber and for the class of scattering or antenna measurements to be performed in the chamber [16].

In another absorbing application, radar-absorbing material (RAM) is used to coat a low-observable vehicle to reduce the radar cross-section signature. The requirement again is to absorb as much of the energy as possible over a prescribed frequency band and for a range of incident angles. Traditionally, the RAM can be fabricated from of a number of carbon-fiber layers, but it has been shown that the performance can be improved by embedding frequency-selective screens into the layers, where the screen is tuned to the band of operation to further enhance absorption over the band of interest [17].

16.2.7 Photonic Bandgap Structures: Controlling Microwave Radiation and Propagation

The above examples draw from frequency-selective structures that consist of one or a few surfaces in a layered configuration. If there are multiple surfaces in the layered configuration, they are designed to provide stability across a range of incidence angles, to extend the passband or stopband, or to allow multiple bands of operation. The designer could extend the frequency-selective screen by periodically adding many layers in the third dimension, creating a type of photonic bandgap structure, also called an electromagnetic bandgap structure [18]. In general, the photonic bandgap material consists of three-dimensional periodic structures that exhibit spectral bands where all propagation is prohibited, at any incident angle and for any polarization—the so-called complete bandgap. By engineering these materials appropriately, they can be combined with the groundplane of planar antennas to increase their efficiency [19] or to tailor the bandwidth of microstrip circuits [20]. In practice, being finite in the third dimension, they cannot provide a complete bandgap and so must be optimized for the application.

16.3 DESIGN PARAMETERS FOR FREQUENCY-SELECTIVE SCREENS

With the application defined and a set of requirements on the frequency-selective screen set, the screen parameters can be designed to produce the required spectral response. Given a type of screen needed to meet the application, the key parameters in the design are the element geometry, the periodic spacing, and multiple layers and dielectrics. This section will first consider the periodic array spacing—something that is designed independent of the element geometry. The section then considers properties of the element geometry independent of the dielectric layering, then the effect of dielectric layers and multiple surfaces separated by layers.

16.3.1 Periodic Array Spacing and the Incident Field

The first parameter that must be chosen in a frequency-selective screen design is the periodic array or lattice spacing. The spacing must be based on the frequency of operation—typically the highest frequency in the band will dictate the spacing such that grating lobes that reduce energy in the main transmitted or reflected harmonics will not appear. (An exception is found in the frequency-scanned antenna in Section 16.2.3.) Figure 16.10 shows the geometry of a general frequency-selective surface. The periodic cell is that section of surface that repeats after translation of mT_x, and $nT_y/\sin\Omega$ along the skewed coordinate axes for all (m, n). The periodic element may be fully contained within the periodic cell or may overlap cells. Any versatile design software must allow elements within the cell to run across the cell boundaries to handle all geometries of interest.

The incident field is defined as a plane wave incident at angles $(\theta^{\text{inc}},\phi^{\text{inc}})$. The field is polarized with the electric vector parallel to the surface or TE_z in the plane of incidence, or with a component perpendicular to the surface or TM_z.

Because of the periodic nature of the surface, the currents and reflected and transmitted fields can be represented by an infinite set of spatial harmonics. These plane waves consist of the main or dominant harmonic and the higher-order harmonics—called grating lobes—if propagating. The mathematical form of the harmonics is entirely dependent

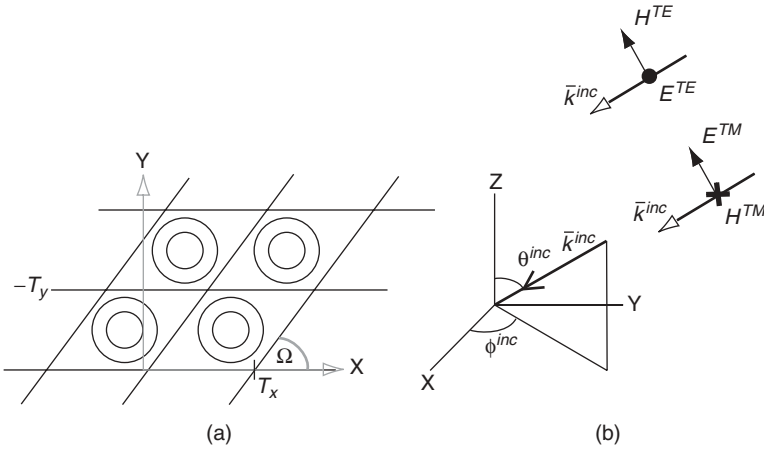


Figure 16.10 (a) Geometry of frequency-selective surface showing concentric ring elements. (b) Incident field coordinate system with TE and TM definitions.

on the periodic array spacing, incidence angle, and wavelength of operation. Due to this periodic and infinite nature of the surface, all fields and currents must satisfy Floquet's condition that states the fields or currents in one periodic cell must be identical to those in an adjacent cell differing only by the incident field phase shift. These observations are summarized by the set of Floquet harmonics used to mathematically and physically characterize the periodic fields and currents on and around the surface:

$$\Psi_{mn}(x, y) = \frac{1}{\sqrt{T_x T_y}} e^{j(k_{x_m} x + k_{y_{mn}} y)} \tag{16.3}$$

where the harmonics are normalized to the periodic cell area, and the propagation constants for a periodic structure are

$$\begin{aligned} k_{x_m} &= \frac{2\pi m}{T_x} + k_0 \sin\theta \cos\phi \\ k_{y_{mn}} &= \frac{2\pi n}{T_y} - \frac{2\pi m}{T_x} \cot\Omega + k_0 \sin\theta \sin\phi \end{aligned} \tag{16.4}$$

These harmonics are simply the transverse component of periodic plane waves that are incident on and reflected and transmitted from a planar periodic surface. The time harmonic plane wave will then have the form

$$\Psi_{mn}(x, y) e^{\mp j k_{z_{mn}} z} \tag{16.5}$$

for propagation in the $\pm z$-direction, where the plane waves are propagating or are evanescent depending on the z-component of the propagation constant:

$$k_z = \begin{cases} \sqrt{k_0^2 - (k_{x_m}^2 + k_{y_{mn}}^2)}, & k_0^2 \geq (k_{x_m}^2 + k_{y_{mn}}^2) \text{ propagating} \\ -j\sqrt{(k_{x_m}^2 + k_{y_{mn}}^2) - k_0^2}, & k_0^2 < (k_{x_m}^2 + k_{y_{mn}}^2) \text{ evanescent} \end{cases} \tag{16.6}$$

A geometric description of these plane waves—one that shows whether they are propagating or not for a given wavelength, and at what incident angle higher-order harmonics start propagating for that wavelength—is the well known grating lobe diagram shown in Figure 16.11 for rectangular spacing ($\Omega = 90^\circ$) and for the cell spacing $T_x, T_y < \lambda/2$. The geometry is found from setting $k_z = 0$ in Eq. (16.6). The region within the circle at the origin (0,0) is termed the visible or propagating region and contains all angles of incidence out to grazing ($\theta^{\text{inc}} = 90^\circ$) for all ϕ^{inc}, whereas the region outside the circle indicates the invisible or evanescent waves as computed from Eq. (16.6). For the normalized coordinates shown, these circles are of unit radius and are repeated for all harmonics with spacing directly related to the array spacing; this diagram is also called the reciprocal lattice diagram [21]. Since no circles overlap, there are no angles of incidence that will excite a grating lobe. The indicated point within the (0,0) circle represents the specific angle ($\theta^{\text{inc}}, \phi^{\text{inc}}$) of the incident wave. For any incident angle, at the frequency specific to this diagram, this point lies within the center circle. Shown in Figure 16.12 is a grating lobe diagram for the NASA Deep Space Network antenna dichroic plate development (see Section 16.2.1.1) [22, 23]. This design results in the grating lobes overlapping for the cell spacing shown in the inset. The regions where circles overlap correspond to angles of incidence that excite a higher-order harmonic. For an angle of incidence that lies within this region, the dominant harmonic as well as the higher-order harmonic specific to the overlapping circle will propagate. The angle of propagation for both the transmitted and reflected higher-order harmonic can again be calculated from Eq. (16.6). An alternate display of this information is shown in Figure 16.13, one that combines feedhorn contours incident on the dichroic plate, and the onset of grating lobes for two different frequencies. The contours are those of the feedhorn projected onto the incident angles centered in a coordinate system on the dichroic plate. Feedhorn boresight is at an angle of 30° on the plate; hence the contours are centered on this angle. Since the feedhorn propagates energy within an angular pattern and with a prescribed edge taper,

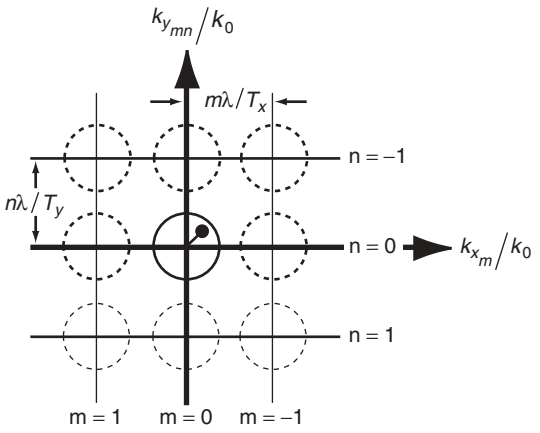


Figure 16.11 Grating lobe diagram for rectangular geometry and with the periodic spacing less than one-half wavelength in both x- and y-directions. The dominant (0,0) propagating harmonic for any incident angle will geometrically be a point within the unit circle at the origin (shown in the center circle). As the frequency increases the unit cells squeeze to overlap, and some incident angles will lie within the overlap regions allowing propagation of higher-order harmonics (see Figure 16.12.)

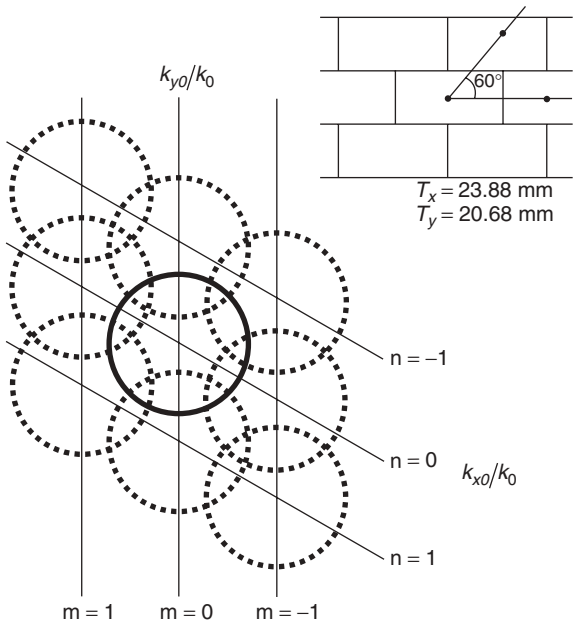


Figure 16.12 Grating lobe diagram for the DSS 13 dichroic plate [22, 33]. Grating lobes will propagate when the incident angle lies within the regions of overlapping circles.

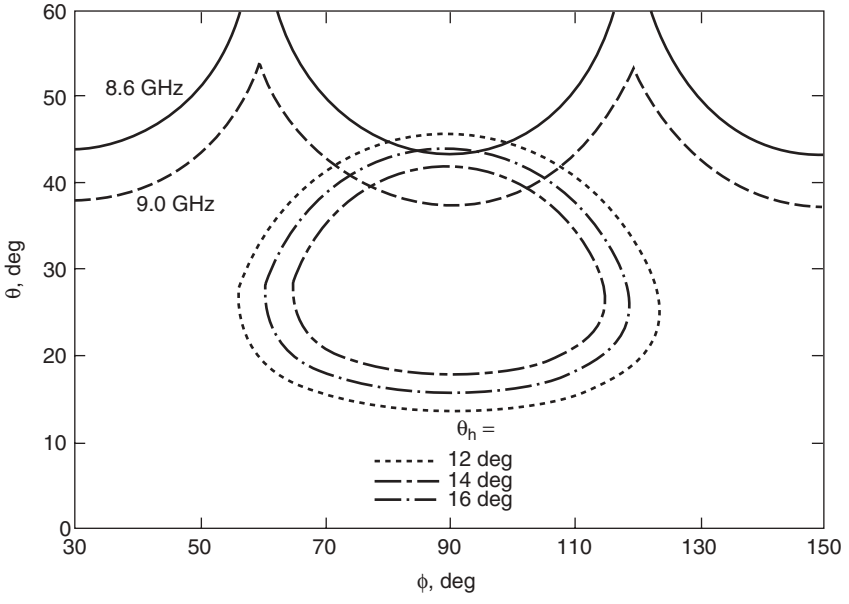


Figure 16.13 Contours of feedhorn projected on the NASA DSS 13 S/X dichroic plate [22, 33]. At a given frequency, energy within the feedhorn contour will cause a grating lobe to propagate.

these contours represent energy incident at those angles on the dichroic plate. When one of these contour lines crosses one of the frequency curves, it indicates a higher-order harmonic will begin propagating. This type of plot combines a range of useful design information including frequency of operation and a practical tapered incident field for a given periodic spacing.

The array spacing and skew angle are used to appropriately pack elements to avoid higher-order harmonics from propagating. Common array geometries include rectangular, brick (square periodic cell with $\Omega = 63.4^\circ$), and a general isosceles triangle ($\Omega = 60^\circ$). From the grating lobe diagram, the largest spacing allowed for a given frequency and maximum incident elevation angle can be found. For rectangular spacing ($\Omega = 90^\circ$) it is

$$T_x(\text{cm}), T_y(\text{cm}) < \frac{30}{f(\text{GHz})}\left(\frac{1}{1+|\sin\theta_{\max}|}\right) \tag{16.7}$$

where the geometry is measured in units of centimeters and frequency in gigahertz. For the common array geometries (brick or isosceles triangle), a general expression is

$$\frac{T_y}{\lambda} < \frac{1/\sin\Omega}{1+|\sin\theta_{\max}|} \tag{16.8}$$

For the common geometries, the isosceles triangle ($\Omega = 60^\circ$) grid spacing gives the tightest packing to push out grating lobes, followed by the brick and then the rectangular spacing. The maximum spacing calculated in Eq. (16.8) is for the theoretical onset of propagation of the higher-order harmonic. In practice, a design must allow for the beginning of radiation at angles somewhat less than these—the shoulder of the grating lobe; therefore some margin needs to be incorporated in the design to lessen sensitivity to the onset of grating lobes [24].

16.3.2 Element Types

The array geometry (skew angle) and spacing must be chosen to place the band of operation within the region where grating lobes are not propagating, but the actual performance of the frequency-selective screen will be heavily dependent on the shape of the element within the periodic cell. The performance characteristics affected by the element shape are (1) stability of the resonant frequency with angle of incidence; (2) bandwidth shape—narrow or broadband and rate of roll-off; (3) separation of reflection and transmission center frequencies; and (4) cross-polarization levels.

It is this combination of the array spacing and element shape (as well as dielectric layers considered in the next section) that will define the screen performance. The bandwidth characteristic of the screen follows from general filter terminology and is loosely broken into lowpass, highpass, and bandpass or bandstop classes. To achieve a bandstop or bandpass filter response, the element must be resonant at or near the operating frequency. This is the fundamental characteristic of a frequency-selective screen. To achieve a lowpass or highpass design, the screen is typically nonresonant.

Classically, the lowpass and highpass screens consist of rectangular gratings of negligible electrical thickness as shown in Figure 16.14 (although circular or other shapes are perfectly feasible). When the aperture (grid) grating becomes electrically thick, the effect of the waveguide thickness becomes a design parameter allowing a bandpass filter response. The transmission response of the highpass screen is shown in Figure 16.15.

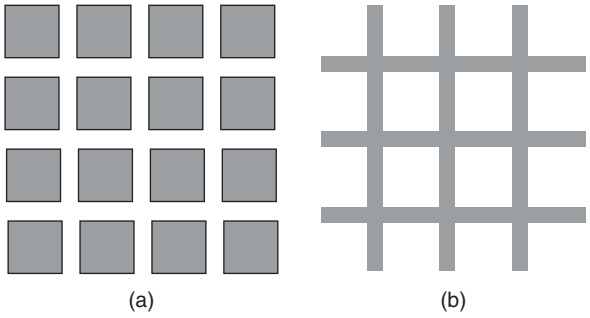


Figure 16.14 Lowpass and highpass screens consisting of (a) metal patches and (b) a metal grid.

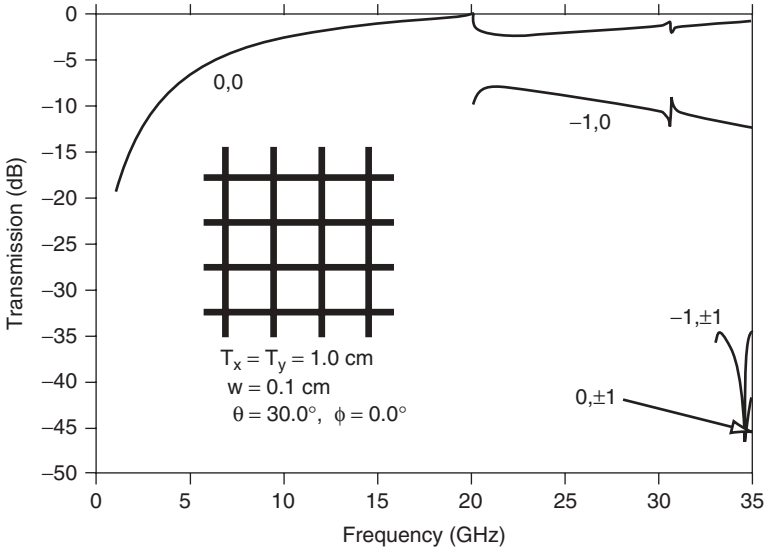


Figure 16.15 Transmission response of the freestanding grid shown in the inset. The parameter w is the metal-trace width. The transmission coefficients are shown for the listed harmonics. Due to symmetry, the $(-1, +1)$ and $(-1, -1)$ harmonics have identical transmission responses and overlap on the plot. The $(0, +1)$ and $(0, -1)$ harmonics also have identical responses and just begin to propagate past 34 GHz, barely visible on the plot.

This screen consists of a metallic grid, freestanding or self-supporting in a frame, and of negligible electrical thickness. The incident plane wave is at 30° incidence and the electric field is polarized TE_z or parallel to the surface. It is seen that there is little transmission for the lowest frequencies, with the transmission approaching unity at 20 GHz close to the onset of the first grating lobe. At the onset of grating lobes, the transmission of the dominant (0,0) harmonic suddenly transitions, and one higher-order harmonic begins propagating as shown on the plot. The (0,0) harmonic continues being transmitted at the incident angle (30°), the reflected (0,0) harmonic at the Snell's law reflection angle ($\theta = 30^\circ$, $\phi = 180^\circ$), while higher-order harmonics first propagate along the surface, and as the frequency increases, the angle of the transmitted and reflected angles moves

toward the surface normal. For 30° incidence, the (−1, 0) harmonic first propagates in the $\phi = 0^\circ$ plane almost at 20 GHz, then the (−1, ±1) harmonics at 32.9 GHz and the (0, ±1) harmonics at 34.6 GHz. The ± harmonics act symmetrically due to the incident vector being in the $\phi = 0^\circ$ plane, so their reflection plots are superimposed on the plot. If we continued to increase the frequency, the (1, ±1) harmonics would begin to propagate next and so on, capturing propagating higher-order harmonics of the reciprocal lattice diagram for this surface (geometrically, the circles of these higher-order harmonics on the reciprocal lattice chart all begin overlapping). We note a narrow resonance near 31 GHz that is related to the width of the strips of the grid.

The complementary structure of metallic patches with negligible electrical thickness would have a complementary filter response (interchanging transmission and reflection coefficients due to Babinet's principle). This type of lowpass filtering screen is used as a solar filter, for example, to block sunlight, while allowing infrared light from planets to be received.

16.3.2.1 Resonant Elements A useful classification of resonant elements is adapted from Ref. 8, as shown in Figure 16.16. The "plate-type" elements of Group 3, seen to be nonresonant elements, were outlined in the previous section, whereas the other groups all have a resonant characteristic. To generally achieve a bandpass or

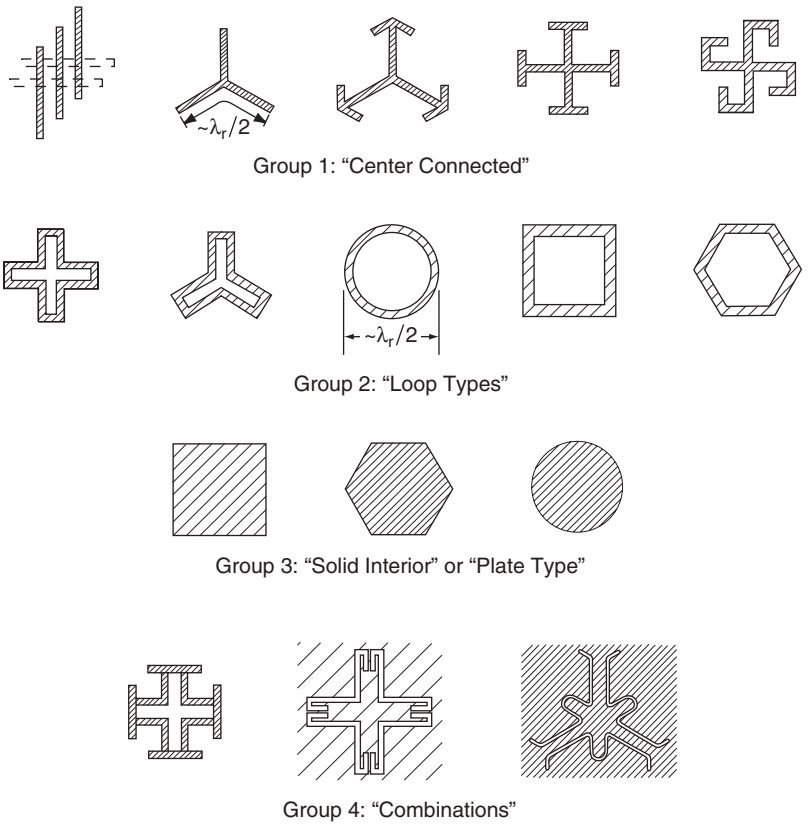


Figure 16.16 Classification of elements according to resonance characteristics. (From Ref. 8.)

bandstop filter response, and to allow transmission at one or more frequencies while other frequencies—perhaps closely spaced—are reflected, a resonant element must be used in the periodic array. The simplest resonant element consists of the metallic dipole (Group 1 in the figure) etched onto a dielectric substrate. The incident electric field, when polarized along the direction of the dipole, will resonate when the dipole is close to multiples of one-half wavelength. The exact resonant wavelength will depend on the periodic array geometry and the dielectric substrate. If the electric vector remains parallel to the dipole as the angle falls off from normal, the performance will change little with incident angle, making the element stable for that polarization. As will be seen, the reflection bandwidth around the resonance can be increased by packing elements tighter in the periodic array. For this narrow element, there is little current directed across the dipole so the cross-polarization levels generated are relatively small. Finally, the highly resonant nature of the dipole, with resonances occurring every half-wavelength, allows for a narrow separation of the transmission and reflection bands f_t/f_r. In general, for any resonant element, the total reflection resonances will always bracket a frequency of total transmission—a consequence of Foster's reaction theorem [25]. Hence, in the plots of frequency response shown in this section, one way to reduce the ratio f_t/f_r is to bring the first two resonances of the element within the periodic cell closer together.

Shown in Figure 16.17 is the response for narrow dipoles etched on a dielectric substrate of dielectric constant 2.2, with a similar dielectric cap, 30° incidence for both TE_z (parallel) and TM_z (perpendicular) polarization, and the periodic array spacing shown. (For all elements considered in this section, thin dielectric substrate and superstrates of thickness 0.254 mm (10 mils) of relative $\varepsilon = 2.2$ will be used. The full effect of the dielectric as well as using dielectrics as design parameters are considered in the next section. All calculations in this chapter were made using the frequency-selective screen software first developed in Refs. 26 and 27, further developed and validated over the years, and available from the author. The dipole array follows from Refs. 8 and 28 for both the TE_z (perpendicular) and TM_z (parallel) polarizations. The response for the TM_z polarization (electric field parallel to the dipole) shows a resonance indicating

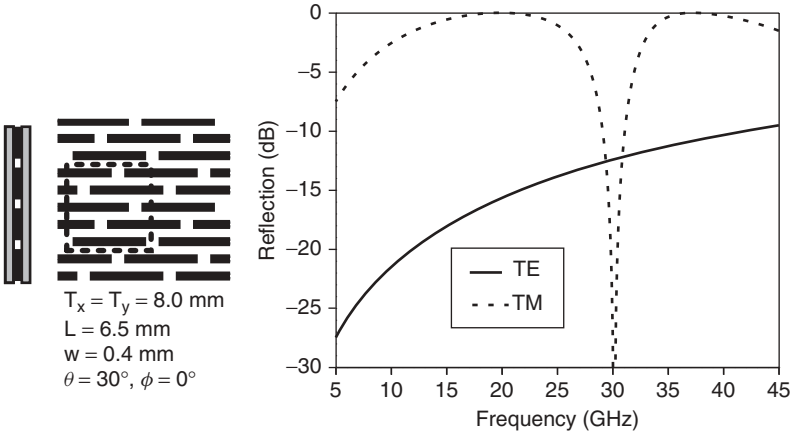


Figure 16.17 Reflection response of dipole array shown on the left in the figure. Dipoles are x-directed and cross periodic cell boundaries. The dielectric layers are 10 mils thick with relative dielectric constant 2.2. The periodic cell is shown with the dashed line.

total reflection near 20 GHz, a second resonance near 37 GHz, and an antiresonance at 30 GHz. With a dipole length of 6.5 mm, a half-wavelength and full-wavelength resonance for a single dipole with no dielectric loading would perfectly exist at 23 GHz and 46 GHz, respectively; we see the modification of these due to a combination of the infinite array, dielectric sheets, finite width of the dipole, and incident field variation. It is clear that when the electric field is polarized transverse to the direction of the dipole (TE_z) the reflected field is low, as expected.

The dipole is a very simple element but is clearly limited by reflecting one polarization at resonance while allowing the opposite polarization to be largely transmitted. This can be a useful property for certain bandpass, polarization sensitive surfaces but is limited for general polarized systems. Two surfaces orthogonally oriented can be sandwiched to provide reflection at both polarizations, but at the expense of additional fabrication and multiple-layer design issues. The class of elements that circumvent this issue, while still allowing good packing density to push out grating lobes, is the class of loop elements. With a loop element, the circumference of the metal trace defines the resonant frequency—the element will reflect when the circumference is roughly one wavelength around. Higher-order resonances then occur when the circumference is multiple wavelengths around. Shown in Figure 16.18 is the reflection response for the crossed-slot element (also termed the four-legged loaded element [8]) on a square grid with the dimensions shown. Again, the first resonance is chosen to occur near 20 GHz, and both polarizations are stable over a range of incidence angles. The second resonance occurs at roughly twice this frequency, somewhat shifted higher, and again an antiresonance then falls in-between these two points. The 20-GHz resonance is somewhat less than what would be expected from the element circumference (measuring the metal trace around all lengths of the crossed slot—a distance that is approximately four times the slot length). The resonant frequency is modified by the array geometry and dielectric substrate. The ratio f_t/f_r here is about 1.85, somewhat higher than that of the dipole. The sharp resonance near 41 GHz for the TE polarization is due to the onset of trapped harmonics in the attached dielectric layers—a topic discussed in Section 16.3.3.1.

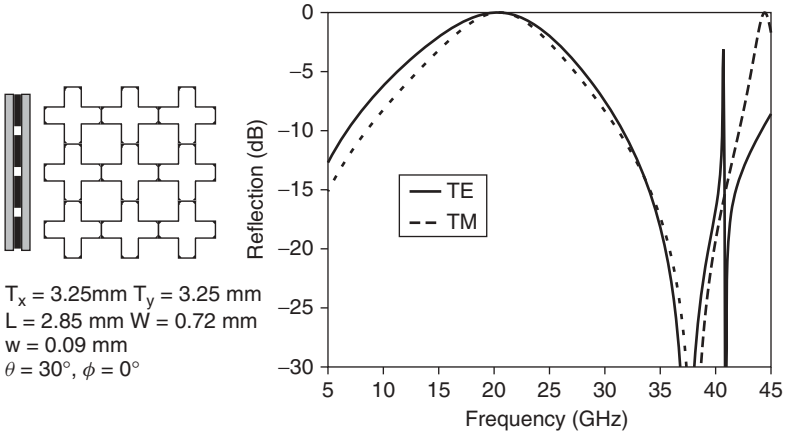


Figure 16.18 Reflection response of crossed-slot array shown on the left in the figure (also termed four-legged loaded element). L and W are the length and width of the slots; w is the metal trace width. The dielectric layers are 10 mils thick with relative dielectric constant 2.2.

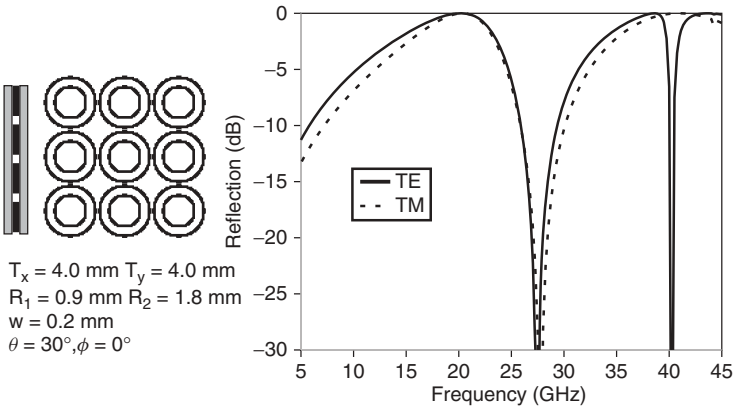


Figure 16.19 Reflection response of concentric ring array shown on the left in the figure. R_1 and R_2 are the inner and outer ring radii and w is the metal trace width. The dielectric layers are 10 mils thick with relative dielectric constant 2.2.

A second loop element is the concentric ring frequency-selective surface shown in Figure 16.19 [29]. Once again the circumference of the rings defines the resonant wavelengths. The outer ring with the larger circumference defines the first (lower) reflection resonance near 20 GHz. The circumference is near one full wavelength, once again modified by the array spacing and dielectric layers. The inner ring will resonant at a higher frequency corresponding to the length of its circumference and is also used as a design parameter to push the second resonance of the frequency-selective surface lower, also modifying the antiresonance located near 27.5 GHz. This allows the ratio f_t/f_r to be modified for a given design. For this design it is about 1.4. Again, the TE response shows a sharp resonance near 41 GHz due to the onset of the first trapped harmonic in the attached dielectric layers, as will be discussed in Section 16.3.3.1.

The elements described above outline the essential properties of resonant frequency-selective surfaces for bandpass or bandstop filtering. A large number of elements are available, all with slightly varying characteristics. Their properties can be matched to the design requirements as needed. In the centered connected class are Jerusalem crosses [30–32], and the related crossed dipole [33, 34] and tripoles [8, 35]. Additional loop elements include loaded tripoles [8], square rings [36], and hexagonal rings [8]. Surfaces that combine these elements as well as others can also be constructed.

16.3.2.2 Element Bandwidth and Stability Once the element is chosen for a given requirement set, the length (for center connected elements) or the circumference (for loop elements) will locate the reflection and transmission resonances. It is also necessary to fix the array periodicity. The periodicity defines when grating lobes will begin to propagate, but it also affects the bandwidth and stability with respect to incidence angle. Generally, as the packing is made tighter, the bandwidth around the first reflection resonance will increase. As the packing is made tighter, the element is also more stable to increasing incidence angle. Shown in Figure 16.20 is the change in bandwidth around the reflection frequency as the element spacing is increased from a tightly packed rectangular array ($T_x = T_y = 3.25$ mm) of the crossed-slot element shown in the inset. The bandwidth is defined as the full 0.5-dB bandwidth about the resonant reflection frequency. Plotted is the resonant frequency (solid line at the center of the shaded region) as a function of the

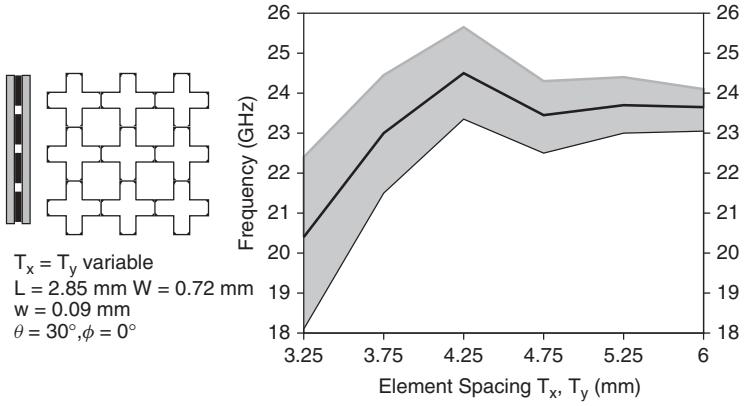


Figure 16.20 Bandwidth of crossed-slot geometry shown on the left in the figure (same as Figure 16.18). Bandwidth is full 0.5-dB bandwidth around the reflection center frequency shown as a solid line at the center of the shaded region. Upper and lower edges of the shaded region indicate upper and lower 0.5-dB points on the resonant reflection curve.

element spacing. The upper and lower edges of the shaded region indicate the upper and lower 0.5-dB points of the resonant curve. As is seen, the resonant bandwidth decreases as the element spacing is increased—over a factor of 2 decrease in the bandwidth over the spacing shown. It is also noted that the center resonance frequency shifts with a change of spacing, an effect that must be accounted for in the design.

For slot-type elements, the slot width is another parameter that influences the bandwidth. In complete analogy to the width of metal dipoles, the bandwidth will decrease as the slot width decreases. Figure 16.21 is a plot of the 0.5-dB bandwidth for the same crossed-slot geometry as in the previous figure and with rectangular element spacing of

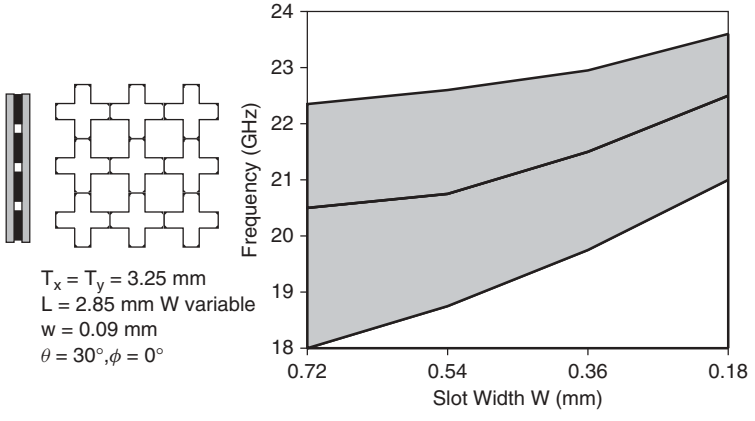


Figure 16.21 Bandwidth of crossed-slot geometry as a function of slot width W for geometry shown on the left in the figure. Bandwidth is full 0.5-dB bandwidth around the reflection center frequency shown as a solid line at the center of the shaded region. Upper and lower edges of the shaded region indicate upper and lower 0.5-dB points on the resonant reflection curve. The dielectric layers are 10 mils thick with relative dielectric constant 2.2.

3.25 mm. As the slot width is decreased from the "fat" slot of 0.72 mm to the "skinny" slot of 0.18 mm, the bandwidth decreases roughly by a factor of 2. The combination of element spacing and slot width is a potent means to control the bandwidth of this element, tuning it to the requirements.

Finally, the stability of the element to changes of incidence angle is shown in Figure 16.22 for the "fat" crossed-slot in the previous figures in a tightly packed array configuration. Plotted is the reflection versus frequency for normal, 30°, and 60° incidence angle for both TE_z (parallel) and TM_z (perpendicular) polarizations. It is clear that this configuration produces a relatively stable reflection coefficient versus angle of incidence, the roll-off for the TM_z (perpendicular) polarization is relatively stable, while the TE_z (parallel) polarization is less stable around the antiresonance. For this tightly packed array, the grating lobes are pushed out—for the above case the first grating lobe begins propagation above 46 GHz at grazing incidence. If the array spacing must be larger, grating lobes can cause problems in the design. For the same element, but at 4.75-mm spacing, the first grating lobe begins to propagate at 33.8 GHz when the incidence angle is 60°. One solution is to move from a rectangular spacing to an isosceles triangle or brick spacing to pack the elements tighter, pushing out grating lobes. As described in Section 16.3.1, Eq. (16.7) the grating lobes are pushed out by a factor of $1/(\sin 60^\circ) = 1.15$ relative to that of rectangular spacing, and for the crossed-slot geometry, elements can be packed tighter as well, thus reducing the element spacing at the same time [23].

Resonant elements are at the heart of meeting the frequency-selective surface design requirements. As noted previously, the choice of element, sizing the element to the resonant frequency, and the element array spacing give first-order design parameters to be tuned. The slot width, or width of the metal elements on center connected elements, affects the resonant bandwidth to second order. A combination of the element type and spacing can affect the f_t/f_r ratio for a design. All these parameters are used in a trade study to meet the design requirements in an optimal fashion. Although a rule of thumb is to quickly focus a design, it is essential to employ software design tools to accurately predict the response and vary the parameters to reach an optimum design.

16.3.3 Multilayer Surfaces

To expand the range of design parameters available, surfaces can be sandwiched and combined with dielectric layers to provide more flexibility. This is especially important

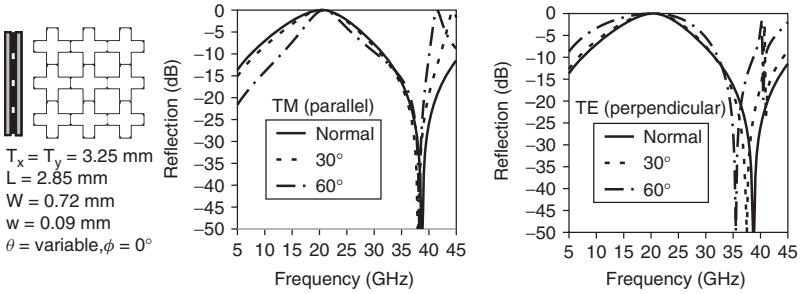


Figure 16.22 Stability for crossed-slot geometry shown on the left in the figure as a function of incidence angle for both polarizations. The dielectric layers are 10 mils thick with relative dielectric constant 2.2.

when the bandwidth needs to be better controlled (either wider or narrower), the roll-off of the passband or stopband needs to be sharpened, multiple bands of operation are needed, or when mechanical support and construction require a range of various materials to be used.

16.3.3.1 Effect of Dielectric Layers Plotted in Figure 16.23 is the resonant frequency as a function of dielectric layer thickness for the crossed-slot geometry shown on the left in the figure. This is the same element from Figure 16.18, with the same relative dielectric constant, but for varying thickness of a layer on one side (single-sided) or layers of equal thickness on both sides of the surface (double-sided). For zero thickness, the surface is freestanding and has a resonant frequency f_{r_0} of 28.4 GHz. For layers on both sides, the resonant frequency drops as the layers increase, finally flattening out at a value of $f_{r_0}/\sqrt{\varepsilon_r}$, as would be expected for an element embedded in a medium of dielectric ε_r. As seen on the plot, it only takes about a 0.8-mm layer to "look infinite" to the element, and as seen in the scale on top of the chart, this is less than one-tenth of the nonloaded resonant wavelength λ_{r_0}. For a single-sided surface (no layer on the opposite side), the resonant frequency drops to a value that is in-between that of no dielectric loading and full loading on both sides. A rule of thumb is that the effective dielectric constant is the average of both layers, or in this case $(1 + \varepsilon_r)/2$. Again, for a single-sided screen, the effect is seen for a thin layer, where f_r levels off for just a 0.6-mm layer or about $0.05\,\lambda_{r_0}$. These effects are shown for a range of dielectric constants in Figure 16.24 for both single-sided and double-sided screens for the same crossed-slot element. Again, similar effects of the loading are seen as the relative dielectric constant is increased from 2.2 to 9. These effects of dielectric loading are common for any elements printed or etched on the dielectric layers.

The effect of the layers is to necessitate the modification of the element size from what it would be in free space. For linear, center connected elements, the length corresponding to the first resonance will be shortened from the free-space half-wavelength value. For loop-type elements, the circumference size of one wavelength will be similarly shortened due to the dielectric loading. The reduction in size will correspond to

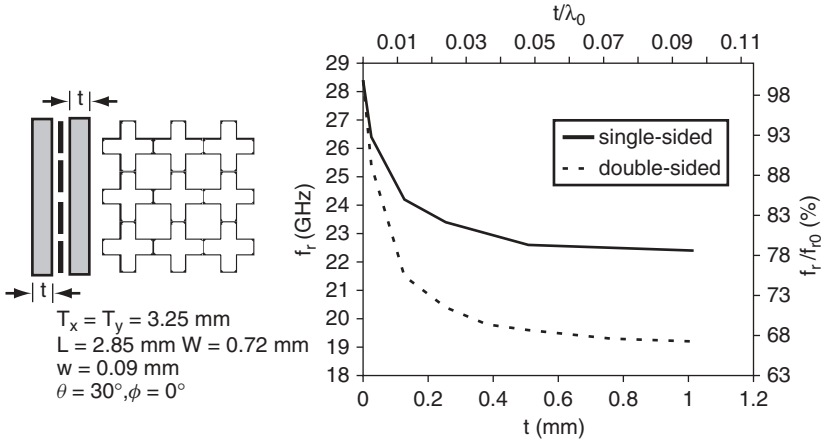


Figure 16.23 Change of resonant frequency as a function of the dielectric layer thickness for a single-sided layer (below the periodic array) or for a double-sided design (below and above the periodic array). The relative dielectric constant is 2.2.

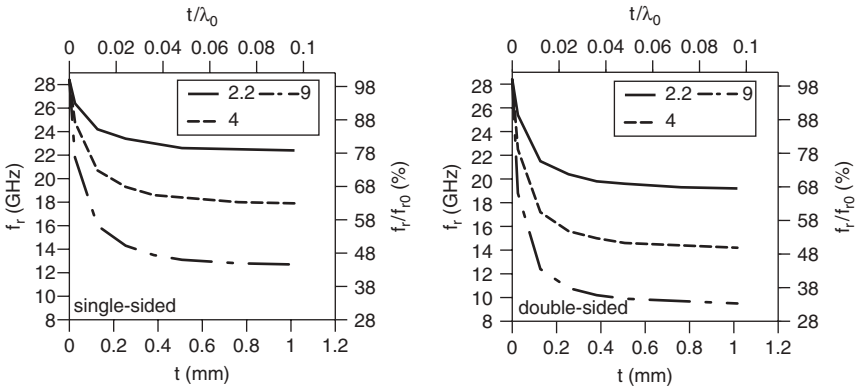


Figure 16.24 Variation of resonant frequency as a function of the thickness for differing relative dielectric constant. Shown are plots for a single-sided screen (dielectric layer below the periodic surface) and a double-sided screen.

that shown on the plots in Figures 16.23 and 16.24 for the various dielectric constants and layer thicknesses.

The effect of the dielectric layer, besides loading the element and reducing its size, is to modify the stored or evanescent energy within and near the surface. From Section 16.3.1.1, it was seen that the grating lobes begin to propagate at frequencies entirely dependent on the array period and incidence angle. This is independent of whether or not there are dielectric layers surrounding the surface. Although the higher-order harmonics of Eq. (16.6) would propagate within the dielectric at lower frequencies (lesser by $1/\sqrt{\varepsilon_r}$) if the dielectric region were infinite, the propagation of the harmonics in the free-space region outside the layers is independent of the layer material. This physical model is represented by harmonics within the layers (governed by Eq. (16.6) with k_0 modified to $\sqrt{\varepsilon_r}k_0$) and matched with the harmonics outside the layer at the layer boundary. Within the layers, the harmonics will become trapped and appear at frequencies calculated according to Eq. (16.6) with k_0 modified by $\sqrt{\varepsilon_r}k_0$. The trapped harmonic represents stored energy within the layer and accounts for the quick swings in the transmission or reflection coefficients at the frequency at which these trapped harmonics begin to appear. An example of this is seen in Figures 16.18 and 16.19 for the crossed-slot and ring elements, respectively, at roughly 41 GHz for the TE polarization. This is the frequency at which the first higher harmonic would propagate if the array were embedded in an infinite medium with the dielectric constant 2.2. The effect is polarization dependent and also depends on the angle of incidence, layer dielectric constant, and array spacing. If more than one layer is present, the quick swings in the coefficients would depend on the effects of all the layers combined.

16.3.3.2 Multisurface Frequency-selective Screens From Figures 16.23 and 16.24 the element looks as if it is in an infinite dielectric medium for a relatively thin layer surrounding the surface. This results from the observation that energy in the higher-order harmonics has decayed to negligible values when the layer is about one-tenth wavelength, and only the dominant harmonic is propagating. With only a single harmonic propagating, it becomes reasonable to apply the practice of microwave circuits to cascade surfaces to further modify the reflection and transmission coefficients of the screen

to meet a given set of requirements. For the metal-etched elements considered in this chapter, the immediate use of multiple cascaded surfaces is to extend the reflection bandwidth of the surface, creating a controlled bandstop filter. In analogy to microwave circuit filters, the surfaces can be placed apart an appropriate distance to combine the equivalent impedances representing the surface in a manner that extends the screen's bandwidth. The model of this approach follows from an equivalent circuit comprised of frequency-dependent elements on a transmission line used to represent each surface and supporting dielectric layers [8]. The circuit has a spectral response modeling that of the surface; for example, see Refs. 37 and 38. The equivalent elements for each surface are placed a distance apart on a transmission line with characteristic admittance modeling the dielectric layer. The design method is to then choose an element and spacing as a starting point for the desired reflection response. The resonant reflection frequency, desired stability with incidence angle, and polarization response for the single screen is first chosen and simulated. For broadband response, the element shape and spacing for wide bandwidth are chosen. This surface will then be cascaded with a second surface, separated by a realizable dielectric layer. The spacing of the dielectric layer will be chosen to transform the admittance of the second screen such that its interaction with the first surface will result in a wider bandwidth. Using the analogy to microwave circuits, the circuit model that transforms the admittance back into itself is propagation through one-half wavelength in the dielectric. Hence choosing the dielectric spacer to be one-half wavelength will transform the admittance back onto itself, and the parallel combination with the admittance of the first surface will result in an in-phase response of the combination. (Indeed, any $(n + 1)\lambda_d/2$ layer will accomplish this if thicker layers are needed.)

A design, following this process, to increase the reflection bandwidth of a single surface is shown in Figure 16.25. The crossed-slot element geometry is used for each surface in the cascade screen. A honeycomb dielectric spacer layer, modeled with a dielectric constant of 1.1, is placed between the surfaces. Since each surface has a reflection resonance at about 20 GHz, one-half wavelength in the dielectric is about 7.1 mm. The design in Figure 16.25 shows the response of this cascade screen for the two surfaces with a 7-mm spacing layer. As seen from the plot, the 0.5-dB bandwidth has increased from about 4 GHz around the center frequency of 20 GHz at 30° incidence to over 11 GHz centered around 20 GHz at 30° incidence. The resonant frequency remains the same and the bandwidth has increased by nearly a factor of 3. As seen on the plot a narrowband glitch appears near the 20-GHz resonant frequency (0° and 30° plots). This null, a result of combining the surfaces through the half-wavelength matching process, is very narrow and can be varied depending on the exact spacing layer size and dielectric constant. In practice, the null is reduced by dielectric and metal losses and tolerances of the materials and elements. In can also be moved in frequency if it falls in a critical band. As seen from the plot, the bandwidth is reduced for TE incidence as the incident angle approaches 60°. This is dominantly dependent on the individual element response (crossed-slot geometry with given dimensions and array spacing) and should be factored into the design. If a specific broad angle design is required, an individual element with the required response should be used.

This change in response as a function of incident angle points out a limitation of the circuit analogy to the design process. The equivalent circuit model is obviously a function of frequency but is also a function of the incident angle and polarization. It is not possible to accurately capture these parameters into the circuit model and use the

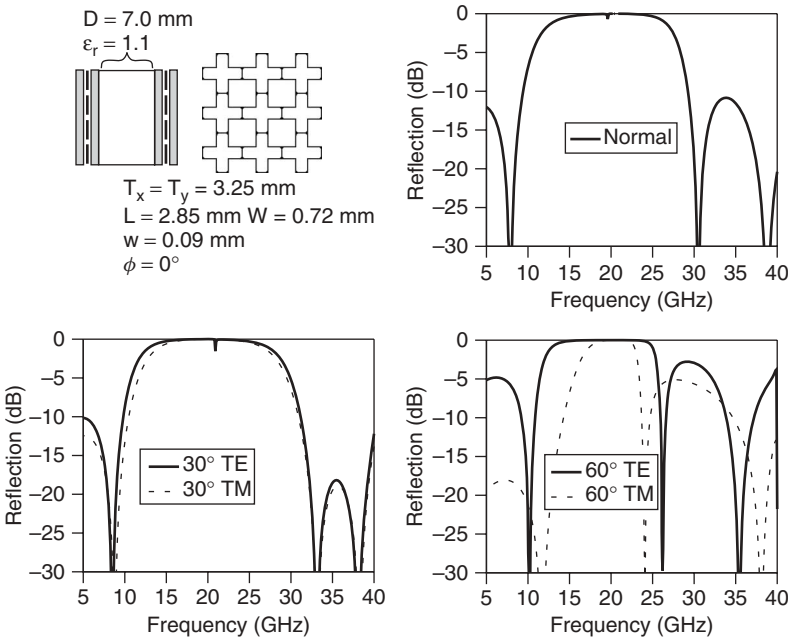


Figure 16.25 Two-layer design used to extend the bandwith of the surface. Both frequency-selective surfaces are identical and shown in the inset. Plots are for normal, 30°, and 60° incidence. The dielectric layers above and below the periodic surfaces are 10 mils thick with relative dielectric constant 2.2. The spacer layer in between these layers is shown.

microwave circuit analogy for designing the multiple-surface screen. However, it does allow a very good starting point for the design and, when used in conjunction with accurate modeling software to optimize the parameters, provides excellent results.

A third surface, as well as additional surfaces, can similarly be combined with the above two-surface screen to extend the bandwidth. One added design method is to stagger-tune the surfaces. Shown in Figure 16.26 is a third surface, consisting of a crossed-slot element tuned to a resonant frequency of 25 GHz, added to the screen of Figure 16.25. The periodicity of the element is identical to the first two surfaces, but the element is smaller, allowing it to resonate at the higher frequency. The equivalent circuit model of the new surface (resonating at 25 GHz) is now somewhat different from that of the other two surfaces (resonating at 20 GHz). Therefore the design process of rotating the admittance through one-half wavelength in the dielectric to match the admittance of the second layer does not exactly hold for the new geometry. The one-half wavelength in the dielectric, however, is a good starting point for the design, and after multiple design iterations, it is concluded that a dielectric layer of 7.5 mm produces an enhanced bandwidth design. Once again a honeycomb dielectric spacer modeled with relative dielectric constant 1.1 is used. From the plots in Figure 16.26 it is seen that the 0.5-dB bandwidth is now nearly 16 GHz, centered near 22 GHz for a 30° incident angle. As in the two-surface screen, a narrow null appears near the 20-GHz resonance of the larger crossed-slot element surface. And as in the two-surface screen, the response decays at 60° incident angle for the TM polarization. With a stagger-tuned screen, ripples in the reflection response within the stopband will appear depending on how far apart

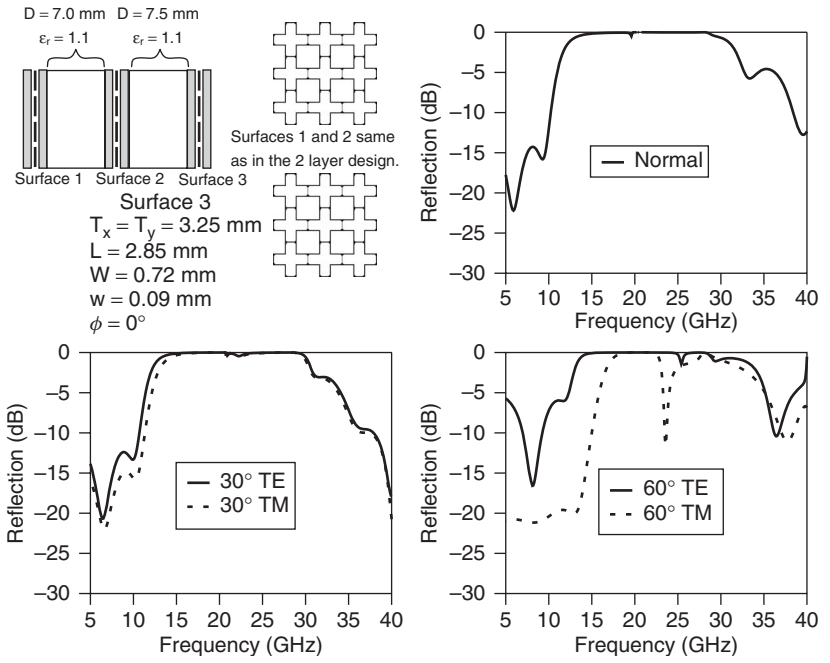


Figure 16.26 Three-layer design used to further extend the bandwith of the surface. The first two frequency-selective surfaces are identical as in the two-layer design. The third surface resonates at a higher frequency and is shown in the inset. Plots are for normal, 30°, and 60° incidence. The dielectric layers above and below the periodic surfaces are 10 mils thick with relative dielectric constant 2.2. The spacer layer in between these layers is shown on the figure.

the individual surface resonant frequencies are placed. These ripples are analogous to those in microwave filters, but once again will vary with angle and polarization for these full-wave electromagnetic frequency-selective screens.

16.4 ANALYSIS AND DESIGN METHODS

An accurate full-wave electromagnetic software application is key to frequency-selective screen design. The software must consider the full range of element geometries, multiple frequency-selective surface layers, and multiple dielectric layers of varying thickness—those quite electrically thin as in glue layers and those electrically thick as in honeycomb spacers. Various mathematical approaches have been developed for modeling screens. The most common uses an integral equation formulation to solve for unknown currents on the metal traces [39–42]. Modeling of multiple dielectric layers is handled by exactly incorporating their characteristics into the Green's function. The integral equation formulation can be developed with spectral or spatial formulations, the key practical difference being the accurate and efficient numerical summations that must be completed in the spatial approach [43, 44]. A second approach couples the differential equation finite-element model within the periodic cell to Floquet harmonics in the regions above and below the screen [45–50]. This approach is useful

when a complicated thick geometry is used, as in the thick plates outlined earlier. Similarly, a finite-difference time-domain (FDTD) approach can be constructed and used [51, 52]. Either approach requires attention to the meshing within the periodic cell and especially at the periodic cell walls to appropriately satisfy the periodic boundary condition. Finally, as noted in Section 16.3.3.2 the equivalent circuit models provide useful design avenues [53–55]. Each method has its strengths and weaknesses and is suited to the geometric configuration and computational efficiency. Modern software tools will be joined with optimization methods to help zero in on specified designs and allow easy tolerance studies to be performed. Common to any formulation is the process of breaking the large surface down to a single periodic cell, exploiting the infinite (or electrically large) structure to allow a tractable computational solution.

Independent of the modeling approach, an accurate solution for the transmission and reflection coefficients depends on the formulation satisfying Maxwell's equations, the periodic boundary condition, boundary conditions at the surfaces of different materials, and the radiation condition on the transmitted and reflected waves. The software must model any losses in the dielectrics and metals, handle a range of elements comprised of thin or fat metal traces, and be computationally efficient to allow the optimization necessary in any iterative design process. All calculations in this chapter were made using the frequency-selective screen software first developed in Refs. 26, 56 and 57 further developed and validated over the years and available from the author. Additionally, general optimization methods can be integrated with the design software to assist the designer [58–60]. The following sections highlight the key elements of the analysis, giving special emphasis to differences between methods that impact the designer.

16.4.1 Integral Equation Methods

The geometry of a multilayered screen is shown in Figure 16.27. In general, there are any number of dielectric layers and frequency-selective surfaces consisting of a general periodic metal element geometry. The dielectric layers are described by a complex relative permittivity and permittivity. An optional conducting groundplane can be placed at the bottom of the screen when a scanned screen or reflectarray is considered. Shown in the figure are the set of Floquet harmonics described in Section 16.3.1 that completely characterize the periodic structure. The forward and backward traveling harmonics are labeled $a^i_{m,n}$ and $b^i_{m,n}$ for each ith layer. If the groundplane is present below the screen, $b^I_{m,n}$ is identically zero. The goal in the analysis is to accurately compute the coefficients $a^0_{m,n}$ and $b^I_{m,n}$ for an incident harmonic $b^0_{m,n}$. Because the surfaces are planar, it is natural to construct the solution with fields TE and TM to the $\hat{\mathbf{z}}$-direction. There is then a set of coefficients $a^i_{m,n}$ and $b^i_{m,n}$ for each polarization, and any other incident and scattered field polarization can be derived from these. The reflection and transmission coefficients are directly related to these harmonic coefficients. The following discussion outlines the scattering problem for the electric field with the induced current as the fundamental unknown formulated in the spectral domain. These harmonics are derived directly from Maxwell's equations and satisfy the radiation condition for propagating waves. It is only necessary to match fields at all boundary interfaces to complete the solution.

To build a solution for the harmonic coefficients, the tangential fields are expanded in the set of Floquet harmonics defined in Section 16.3.1. (The $\hat{\mathbf{z}}$-components can then be

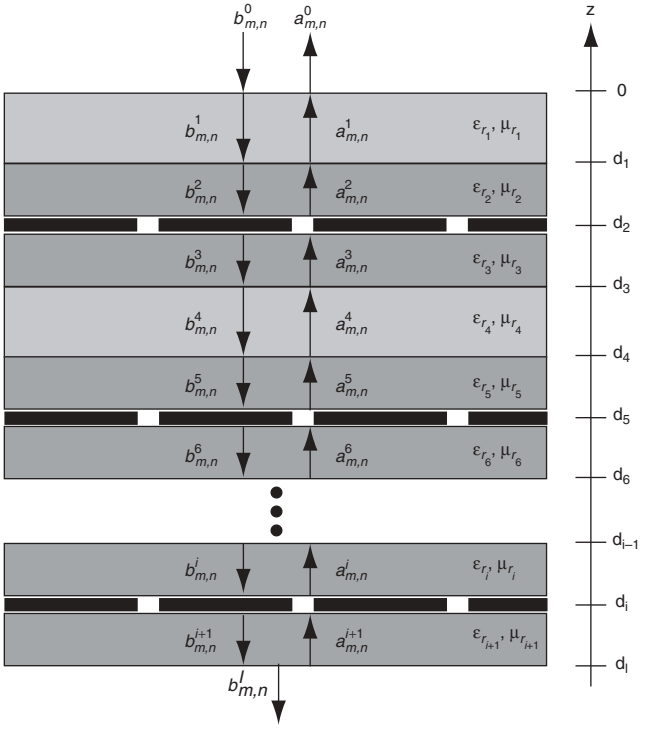


Figure 16.27 Geometry of layered frequency-selective screen showing multiple dielectrics, multiple metal surfaces, and harmonics definitions within the layers. An optional metal groundplane can be placed below the screen, resulting in b^I being identically zero. Any number of dielectric layers or metal periodic surfaces can be layered in the screen, and their thicknesses are arbitrary.

easily found from the tangential components.) This is the starting point for the spectral formulation. Both TE and TM polarizations (p) in each layer (i) have the same form:

$$\begin{aligned}\mathbf{E}^i_p(x,y,z) &= \sum_{m,n}(b^i_{p|m,n}e^{jk_{z_{m,n}}z} + a^i_{p|m,n}e^{-jk_{z_{m,n}}z})\mathbf{e}^i_{p|m,n}\psi_{m,n}(x,y)\\ \mathbf{H}^i_p(x,y,z) &= -\hat{\mathbf{z}}\times\sum_{m,n}(b^i_{p|m,n}e^{jk_{z_{m,n}}z} + a^i_{p|m,n}e^{-jk_{z_{m,n}}z})\mathbf{e}^i_{p|m,n}Y^i_{p|m,n}\psi_{m,n}(x,y)\end{aligned} \tag{16.9}$$

where the TE and TM vector components are

$$\begin{aligned}\mathbf{e}_{\mathrm{TE}_{m,n}} &= \hat{\mathbf{x}}k_{y_{m,n}} - \hat{\mathbf{y}}k_{x_m}\\ \mathbf{e}_{\mathrm{TM}_{m,n}} &= \hat{\mathbf{x}}k_{x_m} + \hat{\mathbf{y}}k_{y_{m,n}}\end{aligned} \tag{16.10}$$

the TE and TM wave impedances are

$$Y_{\mathrm{TE}_{m,n}} = \frac{k_{z_{m,n}}}{\omega\mu}, \qquad Y_{\mathrm{TM}_{m,n}} = \frac{\omega\varepsilon}{k_{z_{m,n}}} \tag{16.11}$$

the propagation constants $k_{x_m}, k_{y_{m,n}}$, and $k_{z_{m,n}}$ are defined in Section 16.3.1 and the summations are theoretically over all harmonics. The total tangential fields in each region are then

$$\begin{aligned} \mathbf{E}^i(x, y, z) &= \mathbf{E}^i_{\mathrm{TE}}(x, y, z) + \mathbf{E}^i_{\mathrm{TM}}(x, y, z) \\ \mathbf{H}^i(x, y, z) &= \mathbf{H}^i_{\mathrm{TE}}(x, y, z) + \mathbf{H}^i_{\mathrm{TM}}(x, y, z) \end{aligned} \tag{16.12}$$

With the fields defined in each layer, a straightforward process of matching tangential $\mathbf{E}$ and $\mathbf{H}$ can be used to relate the $a^i_{m,n}$ and $b^i_{m,n}$ coefficients across each of the surfaces above and below the frequency-selective surfaces. At each dielectric boundary surface the tangential electric and magnetic fields are matched, repeatedly relating a set of the coefficients $(a^i_{p|m,n}, b^i_{p|m,n})$ to those of the next layer $(a^{i+1}_{p|m,n}, b^{i+1}_{p|m,n})$. At the frequency-selective surface, the tangential magnetic field is directly related to the induced surface current $\mathbf{J}(x,y)$ on the metal patch in a periodic cell,

$$\mathbf{J}^i(x, y) = \hat{\mathbf{z}} \times [\mathbf{H}^{i+1}(x, y) - \mathbf{H}^i(x, y)] \tag{16.13}$$

where $\mathbf{H}^i(x,y)$ is the total magnetic field at the ith layer. This process of matching fields through the layer boundaries and relating the magnetic field difference at the boundary with the metal array constructs an equation for the scattered field with the unknown now being the induced current on the metal patches. The problem has been reduced to modeling the current within a periodic cell (on the metal patch) for each frequency-selective metal surface. There may, in general, be any number of dielectric layers and metal surfaces.

The next step in the development of the method is to exploit the orthogonality of the Floquet harmonics to form a set of equations for the unknown current. Key to this orthogonality is the relation

$$\int_{\partial s} \mathbf{e}_{p|m,n} \psi_{m,n} \cdot \mathbf{e}_{p'|m',n'} \psi^*_{m',n'}\, ds = k^2_{\rho_{m,n}} \delta_{p'p} \delta_{m'm} \delta_{n'n} \tag{16.14}$$

where $k^2_{\rho_{m,n}} = k^2_{x_m} + k^2_{y_{m,n}}$, the integral is over the periodic cell, and $\delta_{i,j}$ is the Kronecker delta function. By matching fields at each boundary layer, matching the total magnetic field to the induced current through Eq. (16.13), and repeatedly applying the orthogonality property (Eq. (16.14)), an equation for the scattered electric field on the surface of the metal patches is found:

$$\begin{bmatrix} E^s_x(x, y) \\ E^s_y(x, y) \end{bmatrix} = -\sum_{m,n} \begin{bmatrix} Z_{xx_{m,n}} & Z_{xy_{m,n}} \\ Z_{yx_{m,n}} & Z_{yy_{m,n}} \end{bmatrix} \begin{bmatrix} \tilde{J}_{x_{m,n}} \\ \tilde{J}_{y_{m,n}} \end{bmatrix} \psi_{m,n}(x, y) \tag{16.15}$$

where the impedance matrices Z can be found in Refs. 26 and 41, and due to the use of the orthogonality condition (Eq. (16.14)), it is the Fourier transform of the current that is used in the analysis ($\tilde{J}$). At this point in the development, all boundary conditions have been satisfied except matching the scattered field to the impedance boundary condition at the metal surfaces. In general, this equation is written

$$\mathbf{E}^{\mathrm{inc}} + \mathbf{E}^s = Z\mathbf{J} \tag{16.16}$$

where Z is the surface impedance and $\mathbf{J}$ is the spatial (nontransformed) current from Eq. (16.15). The surface impedance Z is an approximation used to model nonperfect conducting metal surfaces and is useful when loss is purposely introduced, or when the metals become lossy—especially at higher frequencies. Models for this term can be found in Ref. 26, and are important at higher frequencies in the infrared and submillimeter spectra (see Section 16.5.2).

The incident field is similarly broken into TE and TM components:

$$\mathbf{E}^{\text{inc}}(x, y, z) = b_{\text{TE}}\hat{\mathbf{u}}_{\text{TE}} + b_{\text{TM}}\hat{\mathbf{u}}_{\text{TM}} \tag{16.17}$$

where the coefficients b_{TE} and b_{TM} refer to the TE and TM (0,0) harmonics in region 1 above the screen.

To compute the reflection and transmission coefficients of the screen, the unknown current in Eq. (16.16) is solved by writing a method of moments matrix equation and solving the resulting system. The current is expanded in a set of basis functions,

$$\mathbf{J}(x, y) = \hat{\mathbf{x}} \sum_i c_i B_i(x, y) + \hat{\mathbf{y}} \sum_{i'} c_{i'} B_{i'}(x, y) \tag{16.17a}$$

where the coefficients c_i are complex unknown coefficients of the basis set $B(x, y)$. Eq. (16.15) requires the Fourier transform of the current, and hence the set $B(x, y)$. Using the usual method of moments, Eq. (16.17a) can be inserted into Eq. (16.16) and a Galerkin procedure used to generate a matrix equation [26, 41]:

$$Ac = b \tag{16.17b}$$

where the matrix A is the method of moments matrix, b are the coefficients related to the known incident field, and c are the unknown current coefficients to be solved. Different approaches can be used to solve the system—direct matrix or iterative methods, although for most screens of interest direct matrix methods are accurate and efficient. With the solution of the coefficients, the amplitude of the current is known, and the scattered fields—both reflected and transmitted—is derived. From Eq. (16.9), the TE and TM reflected and transmitted fields for each harmonic are written and both propagating and evanescent harmonics are calculated. The coefficients are defined by a set of scattering parameters for the transmitted and reflected harmonics:

$$\begin{bmatrix} T_{\text{TE}} \\ T_{\text{TM}} \end{bmatrix} = \begin{bmatrix} S^{21}_{\text{TE,TE}} & S^{21}_{\text{TE,TM}} \\ S^{21}_{\text{TM,TE}} & S^{21}_{\text{TM,TM}} \end{bmatrix} \begin{bmatrix} b_{\text{TE}} \\ b_{\text{TM}} \end{bmatrix} \tag{16.17c}$$

$$\begin{bmatrix} R_{\text{TE}} \\ R_{\text{TM}} \end{bmatrix} = \begin{bmatrix} S^{11}_{\text{TE,TE}} & S^{11}_{\text{TE,TM}} \\ S^{11}_{\text{TM,TE}} & S^{11}_{\text{TM,TM}} \end{bmatrix} \begin{bmatrix} b_{\text{TE}} \\ b_{\text{TM}} \end{bmatrix} \tag{16.17d}$$

where the scattering coefficients correspond to the transmission coefficient (21 superscript) or reflection coefficient (11 superscript), and the polarization order refers to the (incident,scattered) subscript. These matrices contain the co- and cross-polarized coefficients for each incident polarization.

16.4.1.1 Current Basis Functions Equation (16.15) is an exact representation for the scattered field under the infinite surface approximation. The numerical formulation has been reduced from the infinite screen to that over a single periodic cell although the metal traces that make up an element can extend across periodic cell boundaries. Once the current is solved in Eq. (16.16), reflection and transmission coefficients of the screen are easily calculated. Since it is the Fourier transform of the current that is used in the formulation, picking current basis functions with known transforms is important and commonly used. The choice of these basis functions is key to accurate solutions for the transmission and reflection coefficients. Because frequency-selective screens are resonant structures, and the only unknown quantity in the analysis is the amplitude and spatial form of the induced current, a good choice of these basis functions is essential for an efficient and accurate solution of the transmission and reflection coefficients as a function of frequency. In general, the choice of basis functions can limit the accuracy of the center frequency of the response as well as that of the bandwidth.

The basis functions can be broken into subdomain and entire domain sets, the difference being the generality of frequency surface elements that each can model (Figure 16.28). In subdomain basis functions, the current is broken into pieces that are typically less than one-tenth wavelength in size, and hence can model the most arbitrary element shapes. Rooftop functions have been used and integrated with an efficient fast Fourier transform solution method to create an especially efficient solution method [26]. The widely used Rao–Wilton–Glisson (RWG) triangular basis functions have also been used to provide somewhat more generality in the solution since they more easily model arbitrary element shapes due to the triangularization meshing [56]. Piecewise sinusoids can also be used for thin elements [8]. A key accuracy issue arises when periodic surface elements are used that have some thickness (even as small as one-tenth wavelength in size), where the current across the thin element as well as edge current singularities must be modeled accurately to produce high-fidelity solutions. The rooftop functions and RWG basis set allow this modeling, whereas the piecewise sinusoids that model current directed only along the thin trace generally do not.

Entire-domain basis functions can lead to a very efficient solution if the right set can be chosen to model the element geometry. For example, in Figure 16.28 sinusoid basis functions model current on the dipole trace. Current is directed only along the

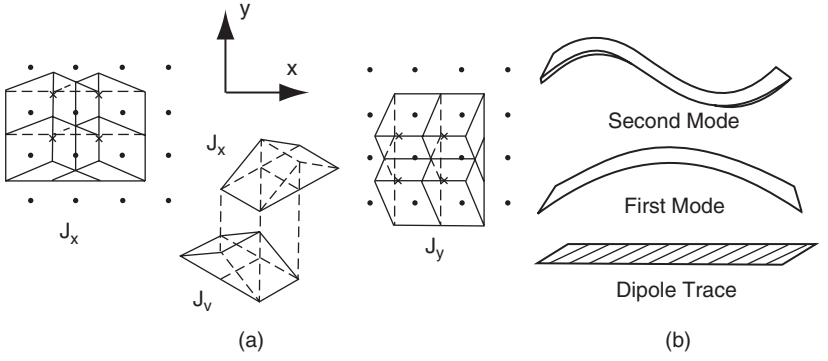


Figure 16.28 Current basis functions used in models. (a) Subdomain rooftop basis showing x- and y-directed currents. (b) Entire-domain basis set specific to narrow dipole element. Shown are the first and second modes of current distribution. Current is uniform across the narrow dipole.

dipole—the aspect ratio is large, 10:1 in this example. Convergence of the solution for the transmission or reflection spectrum can be found using upwards of 10–20 basis functions and a similar matrix size. Hence the numerical solution for the frequency response can be quite rapid. As the element type becomes more general, the basis set is more difficult to define, and for elements with smaller aspect ratios along a metal trace, the current transverse to the long direction of the periodic element is difficult to model. As in the piecewise sinusoid basis set, this limits the solution's accuracy.

Using any type of basis function, the proper current representation at the edges of the metal trace must be maintained to accurately predict the resonant response. For example, the current must reduce to zero for the vector component of current perpendicular to an edge. For vector components parallel to an edge, the current can become singular—hence the current basis function must allow large values along this edge. Because frequency-selective screens are highly resonant structures, the proper modeling of these current components can ultimately limit accuracy of the solution. It has been found that errors in the resonant frequency can be traced to limitations in the basis set used to model the current.

16.4.1.2 Multiple Surfaces in a Screen For screens comprised of multiple frequency-selective surfaces, separated by dielectric layers, the current defined by the vector sets in Eq. (16.15) consists of the sum of all current basis functions on all the surfaces. The matrix size is proportional to the number of basis functions and the solution time and memory needs are also nonlinearly proportional to this size. The matrix solution exactly accounts for all interactions between currents on the surfaces. When the surfaces are closely spaced, the interaction between current sheets is strong, with interactions occurring due to the dominant propagating harmonic as well as some set of evanescent higher-order harmonics. It is essential then that the direct matrix solution includes all basis functions on all surfaces, solved simultaneously. As the electrical distance between surfaces increases, the evanescent harmonics will decay, and at some point only the propagating harmonics will interact between surfaces. For this situation, it becomes possible to break the structure into individual frequency-selective surfaces and attached dielectric layers and to solve for the harmonic amplitudes of this building block independently, cascading the harmonics using a scattering parameter formulation [57]. This approach greatly reduces the computational requirements and allows design flexibility by building up libraries of element responses and optimizing the distance between building blocks. For example, it is commonly used for meanderline polarizers. From Section 16.3.3 it is noted that a surface with attached dielectric layer appears to look as it is in an infinite medium when the layer depth is more than about one-tenth wavelength. In practice, using a separation of one-quarter wavelength allows a minimal number of harmonics (the dominant and a small set of evanescent) to be calculated for each building block in the screen and in the cascade formulation. It is also practical to model the building block as an individual frequency-selective surface with dielectric layers attached to each side, separating the building blocks in the middle of a dielectric layer if feasible.

The integral equation approach has been formulated with the induced current as the unknown. For surfaces with apertures in a metal layer, an entirely analogous approach using the field in the aperture can be developed [41]. The resultant formulation is written with the equivalent magnetic current as an unknown, and the same choice of basis functions as described above applies. Although both formulations will yield identical

solutions, when the aperture area is less than half that of the periodic cell, the aperture formulation will be more computationally efficient. An exception occurs when the metal is lossy, and the electric current formulation with the impedance boundary condition must be used independent of the geometry.

16.4.2 Finite-Element and Finite-Difference Models

The integral equation formulation provides an efficient and accurate model for the frequency-selective screen made up of planar dielectric layers and metal periodic surfaces of negligible electrical thickness. If screens are comprised of thick metallic plates or if the metal periodic elements have some thickness, a formulation that combines a differential equation solution of Maxwell's equations coupled to the Floquet harmonics can be used. Above and below the screen, the Floquet harmonic expansion described in the previous sections is used to model the fields. Within the screen a finite-element solution is used. A finite-element mesh must be generated within the screen and coupled to the Floquet harmonics at apertures above and below the screen. The volume within the mesh region can now be comprised of arbitrary, nonplanar dielectric layers or metal elements. It is only necessary to be able to mesh the region using commercially available meshing tools. For the meshed region, it is essential that the mesh on one of the periodic walls (say, the $+x$ wall) match the surface mesh on the opposite ($-x$) wall. This requirement on the mesh can be used to force the periodic Floquet boundary condition to hold, an essential ingredient in an accurate solution. (The identical meshing procedure is required on the y walls.) The meshing tool must force this boundary matching on opposite walls, a feature not available in all meshing software. A formulation that allows decoupled meshes and does not require the wall matching can also be used [46]. A process of matching the elements at the top and bottom surfaces to the Floquet harmonics is then applied to generate a solution for the reflection and transmission coefficients. In general, the method is less efficient than the integral equation formulation when used to model identical screens, but provides greater flexibility in the type of screens that can be modeled. A modified formulation that eliminates the need for a full Floquet expansion above and below the screen uses approximate absorbing boundary conditions to enforce the radiation condition. In this modification, the mesh is generally extended above and below the screen a distance such that the propagating dominant harmonic exists and the higher-order harmonics have decayed to a negligible strength. The approximate absorbing boundary condition generally becomes less accurate as the angle of incidence (and hence the angle of the reflected and transmitted harmonics) moves off normal and approaches grazing.

An alternative approach to using finite elements is to apply a finite-difference time-domain approach in the meshed region and couple this to an approximate absorbing boundary condition or time-domain version of the Floquet harmonic expansion. This approach is useful when non-time-harmonic waves (such as pulsed signals) are incident on the structure. Care must be taken in meshing the structure since, generally, uniform Cartesian meshes are used and fidelity in accurately modeling the geometry can lead to very dense meshes and correspondingly long computational times. Methods using conformal meshes can be applied to reduce this issue. Finally, being highly resonant structures, the time-domain solutions typically need to be calculated over long time ranges to allow "ringing" to diminish and the solution to converge.

16.5 DESIGN OBJECTIVES

In some circumstances, the standard design practice must be augmented to better predict performance in the system. Two common differences are in the incident field that strikes the screen, and in the screen being curved rather than planar. Another difference is found when the screen is built for operation in the terahertz and infrared portions of the spectrum.

16.5.1 Designing for Finite Extent and Curvature

Fundamentally, for a frequency-selective screen to be totally transparent at one band while being totally reflective within another band, it must be infinite in size, and the incident field must have a planar phase front at the surface of the screen. In practice, the screen is obviously finite and, in general, it is placed in the path of a propagating field that will have some amount of phase and amplitude variation. The effects of the finite size and nonplanar phase variation must be addressed in the design of the screen. Physically, a planar periodic screen resonates to totally reflect or transmit a signal when the phase front across the elements is uniform. It is not essential that the amplitude be uniform, although standard frequency-selective screen analysis methods require uniform amplitude and phase to reduce the problem definition from the infinite screen to the analysis of a single periodic cell. When the screen is curved, as in a frequency-selective subreflector, the phase front in general does not match the screen curvature and there will be degradation in the field transmission and reflection from the ideal.

In practice, the fields incident on frequency-selective screens used in antenna systems are confined in space, having an edge taper, and because they are typically in focusing systems, have somewhat uniform phase fronts within the beam path where the screen is placed. This will then limit the extent of the screen needed in the beam path and not require specific design to limit edge effects. When the fields are not greatly diminished at the edge of the screen, for example, when the screen is used as a radome to cover an antenna, and the scattering response of the screen due to external fields is needed, the edge effects must be understood in the design. The systems can be categorized as follows:

1. Screens used with an incident field that has nearly uniform phase front and amplitude taper that will limit the size of the frequency-selective screen and exhibit minimal edge effects. This can include typical curved subreflectors and feed systems or beam-waveguide systems in Cassegrain or Gregorian antennas. Degradation in the transmission or reflection response will depend on the amount of phase variation across the screen. The screen's design can still be handled using standard methods augmented by proper expansion of the incident field.
2. Screens used with an incident field that either has a large nonuniform phase variation across the screen or significant field amplitude striking the edge of the screen. Although the screen can still generally maintain transmission and reflection bands, the edge effects will disrupt the response depending on the electrical size of the screen. Design methods that break down the screen to one that models a single periodic cell are generally not accurate since the entire screen must be modeled. This case is not addressed in this chapter.

The key design considerations when using the screen with nonuniform field excitation are its size, location in the beam path, and the performance degradation from ideal. Typically, there is little leeway in location since physical considerations will dictate the location, as in the placement of a subreflector in a multireflector antenna. The location of the frequency-selective screen is also traded against its physical size when placed in a diverging beam, although minimizing the amount of wavefront phase variation from uniform is also essential.

16.5.1.1 Performance in a Nonuniform Field: Effect of the Feedhorn Shown in Figure 16.29 is the placement of the X/Ka-band frequency-selective screen in the path

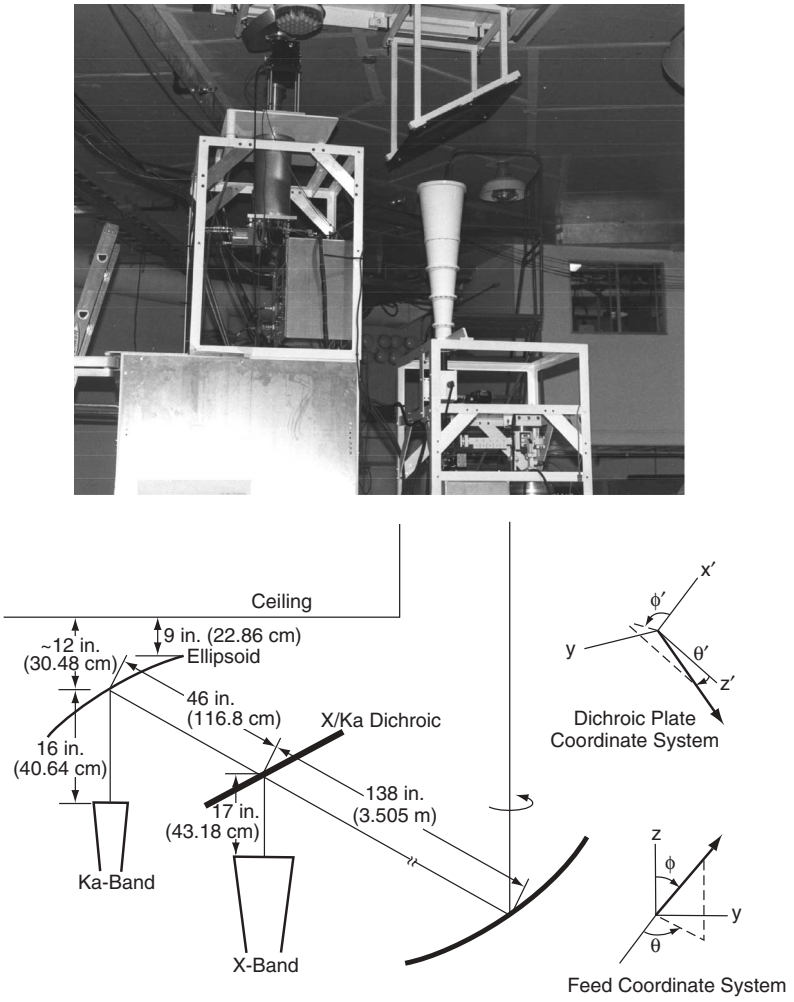


Figure 16.29 Layout of NASA DSS 13 beam-waveguide antenna pedestal room showing X- and Ka-band optics. Upper photograph is the physical layout; lower left shows the schematic with feedhorns, Ka-band curved reflector to upper left, X/Ka-band dichroic plate and main ellipsoid at lower center. In the lower right are the feed (bottom) and dichroic plate (top) coordinate systems. (From Ref. 1.)

of the beam-waveguide system outlined in Section 16.2 [61]. Also shown in the figure are coordinate systems of the X-band horn and that of the dichroic plate whose normal is set at an angle of 30° from the ray path to the feedhorn. Nominally, this is then a frequency-selective screen with 30° incidence that will transmit Ka-band from the horn to the left (reflecting off a curved mirror) and reflect X-band from the horn to the right below the screen.

The Ka-band beam incident on the screen consists of a 26.1-dB gain horn pattern reflected from the curved mirror and transmitted through the screen. The amplitude of the horn tapers to as much as −35 dB at the edge of the frequency-selective portion of the plate. The phase distribution across the screen is that of a near spherical phase front centered at the horn phase center. For the segment of the horn pattern containing the dominant portion of the power amplitude, the phase discrepancy from spherical is within 10° for this as well as other well designed systems. The plate is oversized to present an electrically larger reflecting surface at X-band; the apertures machined into the center portion of the plate need only cover enough of the plate to transmit the Ka-band energy to the required specification. The energy reflected from the oversized frame of the aperture area must be low enough so as not to contribute appreciably to the noise figure of the system if it is a constraint [62]. Generally, the screen, can operate as a frequency-selective screen for sizes as small as 12–15 wavelengths across, although the amount of spillover past the edge can be limiting when considering the system noise performance. Hence the aperture section is made large enough to meet the transmission specification, reduce the spillover noise energy to a negligible level, and meet accommodation constraints.

To calculate the effect of the feed pattern on the frequency-selective screen performance, the pattern must be represented by the plane wave spectrum outlined in Section 16.4, and a transformation of the feed pattern coordinates to the screen coordinates must be used [63]. The feed pattern is decomposed into Floquet harmonics with amplitude and phase sampled from the feed pattern as a function of incidence angle. Once the sampled feed pattern is expressed in the frequency-selective screen coordinates, the reflection and transmission coefficients for that Floquet harmonic are calculated as usual, weighted by the amplitude and phase of the sampled feed pattern. The feed pattern transmitted through the screen is then reassembled in the feed coordinate system.

A linearly polarized feed pattern can be represented in the feed coordinate system as

$$\begin{aligned}\mathbf{E}^{\text{feed}}(\theta,\phi) &= E_\theta(\theta)\sin\phi\hat{\mathbf{u}}_\theta + E_\phi(\theta)\cos\phi\hat{\mathbf{u}}_\phi \\ &= [E^{\text{feed}}]\begin{bmatrix}\hat{\mathbf{u}}_r\\ \hat{\mathbf{u}}_\theta\\ \hat{\mathbf{u}}_\phi\end{bmatrix}\end{aligned} \tag{16.18}$$

where $E_\theta(\theta)$ and $E_\phi(\theta)$ are the E- and H-plane amplitude and phase patterns. These can be calculated or measured patterns. From Section 16.4, the field incident on the frequency-selective screen, expressed in the screen coordinate system, is

$$\begin{aligned}\mathbf{E}^{\text{inc}}(\theta',\phi') &= b_{\text{TE}}\hat{\mathbf{u}}_{\text{TE}} + b_{\text{TM}}\hat{\mathbf{u}}_{\text{TM}} \\ &= [b]\begin{bmatrix}\hat{\mathbf{u}}'_x\\ \hat{\mathbf{u}}'_y\\ \hat{\mathbf{u}}'_z\end{bmatrix}\end{aligned} \tag{16.19}$$

where the coefficients b_{TE} and b_{TM} are the complex amplitudes of the TE and TM polarized fields. A coordinate transformation is used to express the feed pattern in the

screen coordinate system:

$$\begin{aligned}\mathbf{E}^{\text{feed}}(x', y', z') &= [E^{\text{feed}}][R_{\alpha\beta\gamma}]\begin{bmatrix}\hat{\mathbf{u}}'_x\\ \hat{\mathbf{u}}'_y\\ \hat{\mathbf{u}}'_z\end{bmatrix}\\ &= b_{\text{TE}}\hat{\mathbf{u}}_{\text{TE}} + b_{\text{TM}}\hat{\mathbf{u}}_{\text{TM}}\end{aligned} \tag{16.20}$$

where $R_{\alpha\beta\gamma}$ is the Eulerian transformation matrix from the feed spherical coordinate system to the screen Cartesian system and can be found in Ref. 63. The feed pattern is now transformed to the frequency-selective surface coordinate system and decomposed into TE and TM components. For the example shown above, using a 26-dB feedhorn pattern, the amplitude distribution at the screen surface is shown in Figure 16.30.

The feed pattern is sampled at a fine resolution (in the example above, at every 1° in θ and similarly in ϕ) to generate the equivalent set of Floquet harmonics incident on the screen. The transmitted field for each incident harmonic is then found by multiplying the TE and TM coefficients in Eq. (16.20) by the scattering matrix of the screen (described in Section 16.4)

$$\begin{bmatrix}T_{\text{TE}}\\ T_{\text{TM}}\end{bmatrix} = \begin{bmatrix}S^{21}_{\text{TE,TE}} & S^{21}_{\text{TE,TM}}\\ S^{21}_{\text{TM,TE}} & S^{21}_{\text{TM,TM}}\end{bmatrix}\begin{bmatrix}b_{\text{TE}}\\ b_{\text{TM}}\end{bmatrix} \tag{16.21}$$

where S are the calculated scattering matrices for the screen for a given incident angle. The total transmitted field for each (θ,ϕ) harmonic modeling the feed pattern is then

$$E^{\text{trans}}(x', y') = [T_{\text{TE}} \quad T_{\text{TM}}]\begin{bmatrix}\hat{\mathbf{u}}_{\text{TE}}\\ \hat{\mathbf{u}}_{\text{TM}}\end{bmatrix} \tag{16.22}$$

and a reverse coordinate system transformation can be applied to find the field in the original feed E- and H-plane representations [63]. This transmitted field can be used

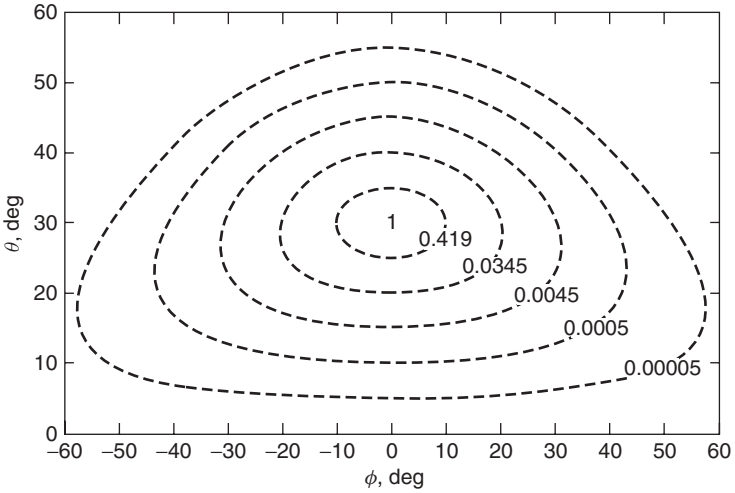


Figure 16.30 Incident horn pattern at 34.5 GHz in the frequency-selective screen coordinates. The values shown are power amplitudes at rings of constant theta (in the feed coordinate system) of 0°, 5°, 10°, 15°, 20°, and 25°. (From Ref. 63.)

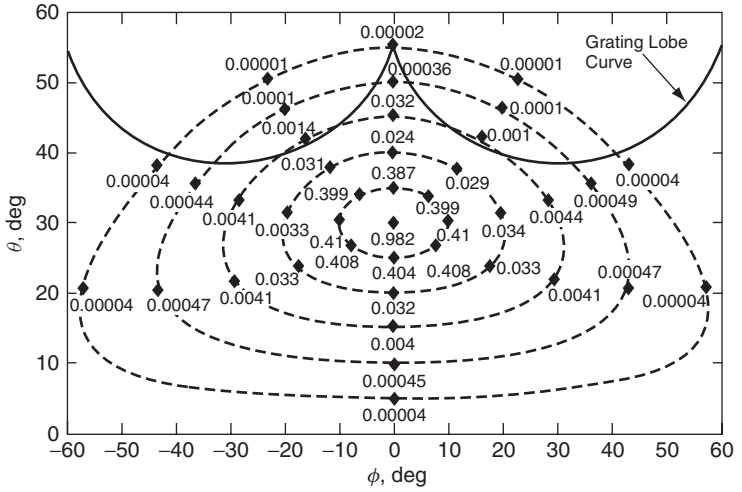


Figure 16.31 Transmitted power through the frequency-selective screen for the horn pattern at 34.5 GHz. The values shown are power amplitudes at the same contours of Figure 16.30. Angles at which grating lobes begin to propagate are also displayed for this frequency. (From Ref. 3.)

in the antenna design or analysis—in the above example this field is propagated to the ellipse and used to calculate the field through the beam-waveguide system. Shown in Figure 16.31 is the field distribution transmitted through the screen for the same power contours of Figure 16.30. The difference between the values on these two charts is due to the reflection loss.

As shown in the above example, effects of the screen on the transmitted field can be made relatively small. The pattern of the incident field is modified slightly as it is transmitted through the screen; its effect can be calculated, and typically there is a sight decrease in antenna gain. There can also be a slight beam shift in the transmitted field leading to a beam shift in the antenna system if the screen has some electrical thickness. There will be a slight phase shift through the screen relative to that without a screen. Similarly, the field reflected from the screen will not have a perfect uniform 180° phase shift across the screen, especially if the interior portion of the screen is different from the exterior portion as described in the example above. Again the effects are small, and this phase shift can be compensated for as described in Ref. 64.

16.5.1.2 Extension to a Curved Subreflector The example above considered a planar frequency-selective screen used in a beam-waveguide antenna system. The same procedure can be used for modeling the effects of a frequency-selective subreflector. The subreflector will consist of metal patches layered on a curved reflector as described in Section 16.2 for the Cassini subreflector. Again, the feed pattern will be tapered at the edge of the subreflector to eliminate spillover, and for a focused system the phase front of the feed will be relatively uniform across large portions of the reflector. Once again the feed can be sampled and used to generate the reflected and transmitted field at the frequency-selective subreflector. One key difference compared to the planar screen example above is the need to account for the curvature of the reflector. The locally planar approximation is used to generate the reflection and transmission coefficients

[65–67]. For each incidence angle of the sampled feed pattern, the local normal of the subreflector is calculated and used in the coordinate system transformation of Eq. (16.20). The reflected or transmitted complex field is calculated. This field can then be used in standard reflector analysis and design programs to propagate fields to the main reflector and then to the far field [68].

16.5.2 Design Considerations for Submillimeter and Infrared Screens

Frequency-selective screens have a long history of being fabricated and used in the microwave and millimeter wave bands. Equivalently, when operating in the optical region, diffraction gratings are used for filtering. In this regime the period of the grating may be several wavelengths, producing multiple orders of light in reflection or transmission (see Section 16.2.3). In regions between these bands, frequency-selective filters consisting of patterned metal layers fabricated using electron beam lithography techniques have been developed. The operation is completely analogous to screens made in the microwave and millimeter wave region except for the choice of materials used and the fabrication process. In addition, the lithography process allows an arbitrary pattern to be machined—a process that can lead to unique designs in special applications. The design of these terahertz and infrared filters follows the same design methods used in the microwave region. Exacting modal matching, integral equation or finite-element methods can be used for design. A major difference, however, is the introduction of material parameters and thicknesses that may not be as important in longer wavelength designs. A screen operating in the terahertz region can be found in Ref. 69; what follows outlines frequency-selective screens built for infrared filtering.

16.5.2.1 Fabrication Using Microlithography Frequency-selective screens operating in the infrared spectral range are fabricated by patterning thin metal films using electron beam lithography techniques. Typically, arrays of apertures are manufactured with varying dimensions on a substrate; for example, manufacturing can be accomplished by using a gold layer deposited on a calcium fluoride substrate. This technique can produce large area arrays (up to 1.5 cm $\times$ 1.5 cm), offering high reproducibility and allowing arbitrary tailoring of the bandpass using the design methods in this chapter. The manufacturing process is not detailed here. Early developments using electron beam lithography used direct-write electron beam lithography on polymethyl methacrylate (PMMA) [70]. Later methods used focused-on-beam (FIB) lithography processes [71, 72] that allow high resolution fabrication but can be limited in the overall size of structure that is manufacturable. Newer processes were developed that allow resolution necessary for frequency-selective screens operating in the infrared (resonance wavelengths in the 2–5-μm range or longer) while still fabricating area arrays greater than 1 cm square [73, 74]. Shown in Figure 16.32 is an intermediate step in the processing for a crossed-slot screen designed to have a resonance (total transmission) at a wavelength of 4 μm. A calcium fluoride (CaF_2) substrate is used to hold the various layers used in processing. At the stage of processing shown in Figure 16.32, the pattern of the aperture has been deposited on the substrate. After this stage, a 0.08-μm gold layer is deposited on the structure. Further processing then removes the layers shown in Figure 16.32 leaving the patterned, gold frequency-selective screen on a CaF_2 substrate as shown in Figure 16.33. It is noted that the corners are not perfect in this fabrication, but high uniformity across the screen is achieved. This example produced a wide area screen (1 cm $\times$ 1 cm) in

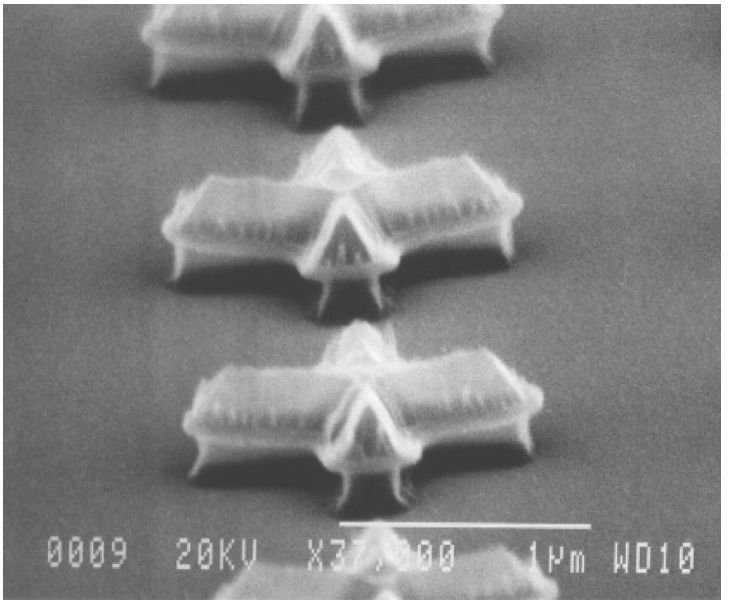


Figure 16.32 Micrograph of intermediate processing step in fabricating the frequency-selective screen. Shown is a calcium fluoride layer holding layers of polymethyl methacrylate and e-beam resist that comprise the crossed-slot apertures. In the next step of processing, a layer of gold is deposited on this foundation.

size—quite large for 4-μm center wavelength operation (there are thousands of unit cells in each dimension). The wide area screen allows differing size elements to be fabricated on different portions of the substrate. Full-sized frequency-selective screens have been patterned on portions of the substrate, allowing filtering at a range of wavelengths tailored to the application. Depending on the incident beam location on the screen, different wavelengths can be filtered and detected past the screen. This has application in a range of imaging spectroscopy applications [74, 75].

16.5.2.2 Modeling Material Losses A key parameter in accurately modeling screens in the terahertz and infrared spectra is the proper accounting of loss in the materials. At short wavelengths the materials allow some penetration into the metal and show ohmic loss. Ideally, a full-wave analysis modeling the fields in the metals is needed and the finite-element or finite-difference time-domain methods described in Section 16.4 would be used. Most metals in the terahertz and infrared ranges, however, are very good conductors and can be modeled using integral equation methods using an approximate boundary condition. From Section 16.4, the final equation to be solved in the integral equation model is

$$\mathbf{E}^{\text{inc}} + \mathbf{E}^{s} = Z\mathbf{J} \tag{16.23}$$

where Z is the approximate surface impedance, the other terms are developed in Section 16.4, and the equation holds for tangential components of the fields over the support of the current J. The complex impedance Z depends on the material parameters and electrical thickness of the metal. In the high frequency region, the materials are

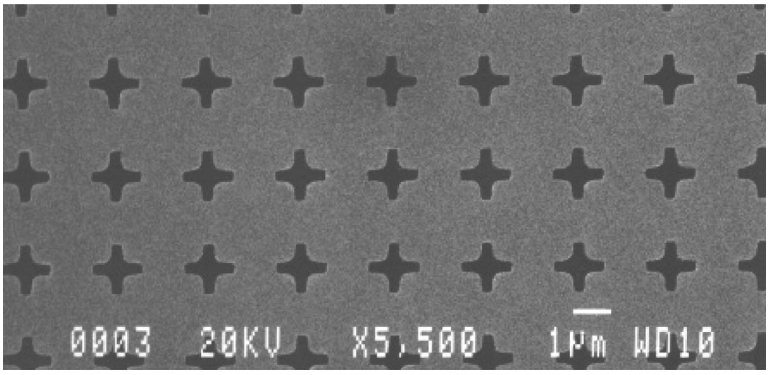


Figure 16.33 Final micrograph of filter after gold layer is deposited in Figure 16.32 and the existing layers comprising the crossed-slot apertures have been removed. The dark slot shapes are the apertures within the gold layer. From the scale in the lower right, the slot length is approximately 1 μm for this design. The slot aperture corners are not exactly square due to processing.

characterized by the optical parameters (n, k) of the metal and are frequency dependent. The relationship between the optical parameters and the complex relative permittivity is

$$\varepsilon_r = (n - jk)^2 = (n^2 - k^2) - j2nk \tag{16.24}$$

and the materials used are nonmagnetic so the relative permeability is unity. A simple impedance boundary condition can be applied if the fields do not greatly penetrate into the metal, satisfying the relation $k_0 t|\varepsilon_r|^{\frac{1}{2}} \ll 1$, where t is the thickness of the metal layer [76]. Most usable metals for application satisfy this criterion, greatly attenuating the field as it penetrates the medium, and hence Z is the medium impedance:

$$Z = \eta = \eta_0 \sqrt{1/\varepsilon_r} = \eta_0 \frac{(n + jk)}{(n^2 + k^2)} \tag{16.25}$$

For the screen fabricated in the example outlined above, the gold layer is 0.08 μm thick, has optical parameters $(n, k) = (1.0, 27.82)$, and the above criterion applies. Shown in Figure 16.34 is the calculated and measured transmission response of the crossed-slot screen shown in Figure 16.33 at normal incidence. The measurement was made using a Beckman IR 4250 spectrophotometer and Micro Scale Test Adjunct setup [77]. We note that the dominant transmitted harmonic is plotted, with higher-order harmonics beginning to propagate for wavelengths less than 3.6 μm. The substrate used in the fabrication is an approximately 1-mm thick CaF_2 layer. At 1 mm, the substrate is greater than 250 wavelengths thick at a wavelength of 4 μm. Impurities in the material, as well as nonperfectly planar surfaces of the substrate do not allow the fields in the CaF_2 substrate to be coherent, acting as if they are in a perfect dielectric layer reflecting in phase or out of phase at the surfaces and propagating uniformly. Rather, they propagate incoherently in this layer. To model this large medium, a half-space approximation beneath the gold layer was used. Instead of modeling a finite-sized dielectric layer in air, the half-space below the layer was a half-space with dielectric constant of CaF_2 $(n = 1.41)$. It is seen that the resonant wavelength is predicted well, although the bandwidth is broader than

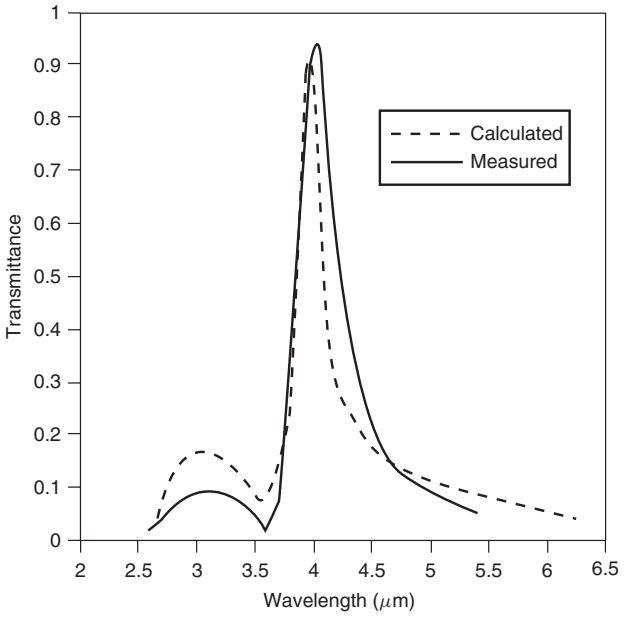


Figure 16.34 Calculated and measured response of gold crossed-slot filter. The dimensions of the crossed slot are 1.434 μm long, 0.159 μm wide, with a square periodic cell of size 2.55 μm. The substrate is approximately 1-mm thick calcium fluoride.

measured. The transmission peak is less than unity due to the CaF_2 substrate acting as a half-space region, as well as the ohmic losses.

To further understand metal losses, a single slot (not a crossed slot as above) in an aluminum metal layer has been modeled. The loss in aluminum is almost 10 times larger than gold ($(n, k) = (6.43, 39.8)$) at 4 μm. Shown in Figure 16.35 is the transmission response of the single slot on the thick CaF_2 layer. Plotted is the TE transmitted field for TE incident polarization at normal incidence—the cross-polarized field is negligible. The transmitted, reflected, and ohmic loss values are shown for the aluminum screen. Also shown is a model of a perfect (lossless) metal screen. It is noted that the effect of loss is to decrease the transmission, shift the center wavelength, and broaden the bandwidth of the passband. The percentage of power is also shown, and it peaks at resonance as would be expected since the induced currents are strongest at these wavelengths.

ACKNOWLEDGMENTS

I would like to acknowledge the long collaboration in this work with Professor Chi Chan. I also thank Peter Simon for his collaboration and material on meanderline polarizers, Rick Cofield for information on the MLS diplexer systems, and Dan Hoppe for discussions related to the finite-sized screens. I thank the reviewers for their comments and needed additions, Professor Balanis for his patience, and Cindy Copeland for her skillful and artistic completion of the manuscript and figures. This research was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA.

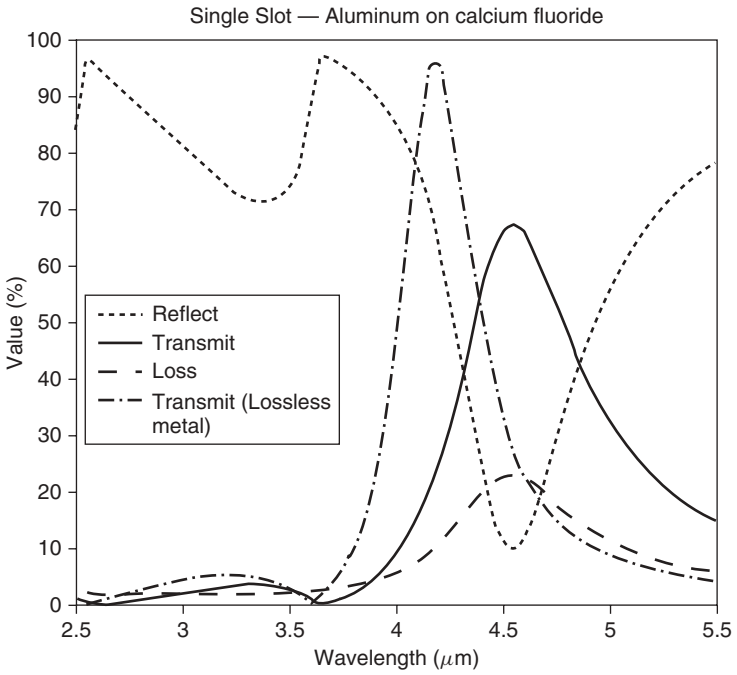


Figure 16.35 Effect of loss on a single slot in an aluminum layer deposited on a calcium fluoride substrate. The slot is 1.4 μm × 0.156 μm, and the periodic cell size is 2.5 μm square. Shown are the transmitted, reflected, and percentage of loss for the screen. Shown for comparison is the transmission curve for lossless metal. Grating lobes begin to propagate for wavelengths less than 3.6 μm.

REFERENCES

1. W. A. Imbriale, *Large Antennas of the Deep Space Network*, Deep Space Communications and Navigation Series, J. H. Yuen (Ed.), John Wiley & Sons, Hoboken, NJ, 2003, p. 316.
2. J. C. Chen, Analysis of a thick dichroic plate with rectangular holes at arbitrary angles of incidence, *TDA PR 42-104*, JPL, Pasadena, CA, February 15, 1991, pp. 9–16.
3. T. Y. Otoshi, and M. M. Franco, *Dual Passband Dichroic Plate for X-Band*, JPL, Pasadena, CA, 1988, pp. 110–134.
4. T. Cwik, and J. Chen, *DSS 13 Phase II Pedestal Room Microwave Layout*, JPL, Pasadena, CA, 1991, pp. 45–106.
5. R. E. Cofield, and P. C. Stek, Design and field-of-view calibration of 114–660-GHz optics of the Earth Observing System Microwave Limb Sounder, *IEEE Trans. Geosci. Remote Sensing*, Vol. 44, No.5, pp. 1166–1181, 2006.
6. P. H. Siegel, R. J. Dengler, and J. C. Chen, THz dichroic plates for use at high angles of incidence, *IEEE Microwave Guided Wave Lett.*, Vol. 1, No.1, pp. 8–9, 1991.
7. R. Mizzoni, The Cassini high-gain antenna subsystem, in *Spaceborne Antennas*, W. A. Imbriale (Ed.), John Wiley & Sons, Hoboken, NJ, 2006, pp. 257–297.
8. B. Munk, *Frequency-Selective Surfaces: Theory and Design*, John Wiley & Sons, Hoboken, NJ, 2000, p. 410.

9. F. S. Johansson, L. G. Josefsson, and T. Lorentzon, A novel frequency-scanned reflector antenna, *IEEE Trans. Antennas Propag.*, Vol. 37, No.8, pp. 984–989, 1989.
10. P. Simon, A meanderline polarizer system, personal communication, Pasadena, CA, 2006.
11. P. S. Simon, Efficient Green's function formulation for analysis of frequency selective surfaces in stratified media, in *IEEE Antennas and Propagation Society International Symposium*, Boston, MA, 2001.
12. P. S. Simon, Modified RWG basis functions for analysis of periodic structures, in *IEEE MTT-S International Microwave Symposium Digest*, Seattle, WA, 2002.
13. T. K. Chang, R. J. Langley, and E. A. Parker, Active frequency-selective surfaces, *IEEE Proc. Microwaves Antennas Propag.*, Vol. 143, No.1, pp. 62–66, 1996.
14. B. Schoenlinner, et al., Switchable low-loss RF MEMS Ka-band frequency-selective surface, *IEEE Trans. Microwave Theory Tech.*, Vol. 52, No. 11, pp. 2474–2481, 2004.
15. B. M. Cahill, and E. A. Parker, Field switching in an enclosure with active FSS screen, *Electron. Lett.*, Vol. 37, No. 4, pp. 244–245, 2001.
16. W. M. Sun, K. F. Liu, and C. A. Balanis, Analysis of singly and doubly periodic absorbers by frequency-domain finite-difference method, *IEEE Trans. Antennas Propag.*, Vol. 44, No. 6, pp. 798–805, 1996.
17. Y. Sha, et al., Experimental investigations of microwave absorber with FSS embedded in carbon fiber composite, *Microwave Optical Technol. Lett.*, Vol. 32, No. 4, pp. 245–249, 2002.
18. J. D. Joannopoulos, R. D. Meade, and J. N. Winn, *Photonic Crystals: Molding the Flow of Light*, Princeton University Press, Princeton, NJ, 1995, p. 137.
19. A. R. Weily, et al., A planar resonator antenna based on a woodpile EBG material, *IEEE Trans. Antennas Propag.*, Vol. 53, No. 1, pp. 216–223, 2005.
20. V. Radisic, et al., Novel 2-D photonic bandgap structure for microstrip lines, *IEEE Microwave Guided Wave Lett.*, Vol. 8, No. 2, pp. 69–71, 1998.
21. N. Amitay, V. Galindo, and C. P. Wu, *Theory and Analysis of Phased Array Antennas*, Wiley-Interscience, Hoboken, NJ, 1972.
22. J. C. Chen, X-/Ka-band dichroic plate design and grating lobe study, *TDA PR 42-105*, JPL, Pasadena, CA, May 15, 1991, pp. 21–30.
23. L. W. Epp, et al., Experimental-verification of an integral-equation solution for a thin-walled dichroic plate with cross-shaped holes, *IEEE Trans. Antennas Propag.*, Vol. 42, No. 6, pp. 878–882, 1994.
24. T. K. Wu, *Frequency-Selective Surface and Grid Array*, Wiley Series in Microwave and Optical Engineering, John Wiley & Sons, Hoboken, NJ, 1995.
25. R. F. Harrington, *Time-Harmonic Electromagnetic Fields*, McGraw-Hill Texts in Electrical Engineering, McGraw-Hill, New York, 1961.
26. T. A. Cwik, and R. Mittra, Scattering from a periodic array of freestanding arbitrarily shaped perfectly conducting or resistive patches, *IEEE Trans. Antennas Propag.*, Vol. 35, No. 11, pp. 1226–1234, 1987. Note: All the calculations in this chapter were made using the frequency-selective screen software first developed in this reference and further developed and validated over the years
27. C. H. Chan, and R. Mittra, On the analysis of frequency-selective surfaces using subdomain basis functions, *IEEE Trans. Antennas Propag.*, Vol. 38, No. 1, pp. 40–50, 1990.
28. T. W. Kornbau, Analysis of periodic arrays of rotated linear dipoles, rotated crossed dipoles, and of biplanar dipole arrays in dielectric, Ph.D. dissertation, Dept. of Electrical Engineering, Ohio State University, Columbus, 1984.
29. E. A. Parker, and J. C. Vardaxoglou, Plane-wave illumination of concentric-ring frequency-selective surfaces, *IEE Proc. Microwaves Antennas Propag.*, Vol. 132, pt. H, pp. 176–180, 1985.

30. R. Cahill, and E. A. Parker, Concentric ring and jerusalem cross arrays as frequency-selective surfaces for a 45-degrees incidence diplexer, *Electron. Lett.*, Vol. 18, No. 8, pp. 313–314, 1982.
31. E. A. Parker, S. M. A. Hamdy, and R. J. Langley, Modes of resonance of the jerusalem cross in frequency-selective surfaces, *IEE Proc. Microwaves Antennas Propag.*, Vol. 130, No. 3, Pt. H, pp. 203–208, 1983.
32. C. H. Tsao, and R. Mittra, Spectral domain analysis of frequency-selective surfaces comprised of periodic arrays of cross dipoles and jerusalem crosses, *IEEE Trans. Antennas Propag.*, Vol. 32, No. 5, pp. 478–486, 1984.
33. E. A. Parker, A. N. A. Elsheikh, and A. C. D. Lima, Convoluted frequency-selective array elements derived from linear and crossed dipoles, *IEE Proc. Microwaves Antennas Propag.*, Vol. 140, No. 5, Pt. H, pp. 378–380, 1993.
34. E. A. Parker, et al., GA optimisation of crossed dipole FSS array geometry, *Electron. Lett.*, Vol. 37, No. 16, pp. 996–997, 2001.
35. J. C. Vardaxoglou, and E. A. Parker, Performance of 2 tripole arrays as frequency-selective surfaces, *Electron. Lett.*, Vol. 19, No. 18, pp. 709–710, 1983.
36. T. K. Wu, Cassini frequency-selective surface development. *J. Electromagnetic Waves Applications*, Vol. 8, No. 12, pp. 1547–1561, 1994.
37. S. W. Lee, G. Zarrillo, and C. L. Law, Simple formulas for transmission through periodic metal grids or plates, *IEEE Trans. Antennas Propag.*, Vol. 30, No. 5, pp. 904–909, 1982.
38. J. A. Arnaud, and F. A. Pelow, Resonant-grid quasi-optical diplexers, *Bell Syst. Tech. J.*, Vol. 54, No. 2, pp. 263–283, 1975.
39. C. C. Chen, Scattering by a 2-dimensional periodic array of conducting plates, *IEEE Trans. Antennas Propag.*, Vol. 18, No. 5, p. 660, 1970.
40. S. Lee, Scattering by dielectric-loaded screen, *IEEE Trans. Antennas Propagation*, Vol. 19, No. 5, p. 656, 1971.
41. R. Mittra, C. H. Chan, and T. Cwik, Techniques for analyzing frequency-selective surfaces—a review, *Proc. IEEE*, Vol. 76, No. 12, pp. 1593–1615, 1988.
42. C. Scott, *The Spectral Domain Method in Electromagnetics*, Artech House, Norwood, MA, 1989.
43. R. E. Jorgenson, and R. Mittra, Scattering from structured slabs having 2-dimensional periodicity, *IEEE Trans. Antennas Propag.*, Vol. 39, No. 2, pp. 151–157, 1991.
44. R. C. Hall, R. Mittra, and K. M. Mitzner, Scattering from finite thickness resistive strip gratings, *IEEE Trans. Antennas Propag.*, Vol. 36, No. 4, pp. 504–510, 1988.
45. T. F. Eibert, et al., Hybrid FE BI modeling of 3-D doubly periodic structures utilizing triangular prismatic elements and an MPIE formulation accelerated by the Ewald transformation, *IEEE Trans. Antennas Propag.*, Vol. 47, No. 5, pp. 843–850, 1999.
46. T. Cwik, Coupling finite-element and integral-equation solutions using decoupled boundary meshes, *IEEE Trans. Antennas Propag.*, Vol. 40, No. 12, pp. 1496–1504, 1992.
47. T. Cwik, Coupling finite element and integral equation methods to model frequency selective screens or dichroic plates, in *The Fourth Biennial IEEE Conference on Electromagnetic Field Computation*, Toronto, Canada, 1990.
48. E. W. Lucas, and T. P. Fontana, A 3-D hybrid finite element/boundary element method for the unified radiation and scattering analysis of general infinite periodic arrays, *IEEE Trans. Antennas Propag.*, Vol. 43, No. 2, pp. 145–153, 1995.
49. M. Bozzi, et al., Design, fabrication, and measurement of frequency-selective surfaces, *Optical Eng.*, Vol. 39, No. 8, pp. 2263–2269, 2000.
50. Z. Lou, and J.-M. Jin, Analysis of 3D frequency-selective structures using a high-order finite-element method, *Microwave Optical Technol. Lett.*, Vol. 38, No. 4, pp. 259–263, 2003.

51. H. Rogier, D. De Zutter, and F. Olyslager, Rigorous analysis of frequency-selective surfaces of finite extent using a hybrid finite difference time domain–boundary integral equation technique, *Radio Sci.*, Vol. 35, No. 2, pp. 483–494, 2000.
52. W. Yu, S. Dey, and R. Mittra, On the modeling of periodic structures using the finite-difference time-domain algorithm, *Microwave Optical Technol. Lett.*, Vol. 24, No. 3, pp. 151–155, 2000.
53. R. J. Langley, and E. A. Parker, Equivalent-circuit model for arrays of square loops, *Electron. Lett.*, Vol. 18, No. 7, pp. 294–296, 1982.
54. E. A. Parker, S. M. A. Hamdy, and R. J. Langley, Modes of resonance of the jerusalem cross in frequency-selective surfaces, *IEE Proc. Microwaves Antennas Propag.*, Vol. 130, No. 3, Pt. H, pp. 203–208, 1983.
55. R. J. Langley, and E. A. Parker, Double-square frequency-selective surfaces and their equivalent-circuit, *Electron. Lett.*, Vol. 19, No. 17, pp. 675–677, 1983.
56. C. H. Chan, and R. Mittra, On the analysis of frequency-selective surfaces using subdomain basis functions, *IEEE Trans. Antennas Propag.*, Vol. 38, No. 1, pp. 40–50, 1990.
57. T. Cwik, and R. Mittra, The cascade connection of planar periodic surfaces and lossy dielectric layers to form an arbitrary periodic screen, *IEEE Trans. Antennas Propag.*, Vol. 35, No. 12, pp. 1397–1405, 1987.
58. C. Zuffada, and T. Cwik, Synthesis of novel all-dielectric grating filters using genetic algorithms, *IEEE Trans. Antennas Propag.*, Vol. 46, No. 5, pp. 657–663, 1998.
59. T. Cwik, et al., Multi-bandwidth frequency selective surfaces for near infrared filtering: design and optimization, in *IEEE Antennas and Propagation Society International Symposium*, Orlando, FL, 1999.
60. D. S. Weile, and E. Michielssen, Design of doubly periodic filter and polarizer structures using a hybridized genetic algorithm, *Radio Sci.*, Vol. 34, No. 1, pp. 51–63, 1999.
61. W. A. Imbriale, *Large Antennas of the Deep Space Network*, Deep Space Communications and Navigation Series, J. H. Yuen (Ed.), John Wiley & Sons, Hoboken, NJ, 2003, p. 316.
62. J. C. Chen, Analysis of a thick dichroic plate with rectangular holes at arbitrary angles of incidence, *TDA PR 42-104*, JPL, Pasadena, CA, February 15, 1991, pp. 9–16.
63. J. C. Chen, Computation of reflected and transmitted horn radiation patterns for a dichroic plate, *TDA PR 42-119*, JPL, Pasadena, CA, 1994, pp. 236–254.
64. N. Miyahara, et al., An influence on radiation characteristics by FSS on a beam-waveguide feed with FSS, *Electron. Commun. Japan Part I—Commun.*, Vol. 86, No. 5, pp. 21–31, 2003.
65. T. Cwik, and R. Mittra, The effects of the truncation and curvature of periodic surfaces: a strip grating, *IEEE Trans. Antennas Propag.*, Vol. 36, No. 5, pp. 612–622, 1988.
66. B. Philips, E. A. Parker, and R. J. Langley, Finite curved frequency-selective surfaces, *Electron. Lett.*, Vol. 29, No. 10, pp. 882–883, 1993.
67. B. Philips, E. A. Parker, and R. J. Langley, Ray-tracing analysis of the transmission performance of curved FSS, *IEE Proc. Microwaves Antennas Propag.*, Vol. 142, No. 3, pp. 193–200, 1995.
68. Rahmat-Samii Y. and A. N. Tulintseff, Diffraction analysis of frequency-selective reflector antennas, *IEEE Trans. Antennas Propag.*, Vol. 41, No. 4, pp. 476–487, 1993.
69. M. Bozzi, et al., Design, fabrication, and measurement of frequency-selective surfaces, *Optical Eng.*, Vol. 39, No. 8, pp. 2263–2269, 2000.
70. D. M. Byrne, et al., Infrared mesh filters fabricated by electron-beam lithography, *J. Vacuum Sci. Technol. B*, Vol. 3, No. 1, pp. 268–271, 1985.
71. J. C. Wolfe, et al., A proximity ion beam lithography process for high density nanostructures, *J. Vacuum Sci. Technol. B*, Vol. 14, No. 6, pp. 3896–3899, 1996.
72. M. D. Morgan, et al., Application of optical filters fabricated by masked ion beam lithography, *J. Vacuum Sci. Technol. B*, Vol. 14, No. 6, pp. 3903–3906, 1996.

73. G. M. Atkinson, et al., Self-aligned high electron-mobility transistor gate fabrication using focused ion-beams, *J. Vacuum Sci. Technol. B*, Vol. 9, No. 6, pp. 3506–3510, 1991.

74. A. Ksendzov et al., Wedge filters for spectral imaging in the near IR using metal grids, *Proc. Infrared Astro. Instrum.*, p. 3354, 1998.

75. T. Cwik, et al., Design of multi-bandwidth frequency-selective surfaces for near infrared filtering, in *SPIE's 43rd Annual Meeting on Optical Science, Engineering, and Instrumentation*, San Diego, CA, 1998.

76. K. M. Mitzner, Effective boundary conditions for reflection and transmission by an absorbing shell of arbitrary shape, *IEEE Trans. Antennas Propag.*, Vol. 16, No. 6, p. 706, 1968.

77. C. Labaw, Measuring spatially varying transmittance of a filter, *NASA Tech Briefs*, p. 66, 1992.

CHAPTER 17

MEMS Integrated and Micromachined Antenna Elements, Arrays, and Feeding Networks

BO PAN, JOHN PAPAPOLYMEROU, and MANOS M. TENTZERIS

17.1 INTRODUCTION

With the ever increasing demand for intelligent, high performance, low cost, and low power-consumption wireless communication systems, the need for integrating multifunctional modules (digital, RF, optical, etc.) on a single chip or into a single package becomes more pronounced.

Microelectromechanical systems (MEMSs) and the application of this technology to radiofrequency (RF) systems enable the production of tunable components with low power consumption, high linearity, and high performance. The tunable characteristics of RF-MEMSs can easily be exploited in their integration with antennas, providing numerous advantages such as reconfigurability in the polarization, frequency, and radiation pattern. Furthermore, the monolithic fabrication of the antenna together with these tunable components reduces drastically the power losses and parasitic effects compared to the integration of discrete components.

Currently, RF-MEMS switches have shown immense commercial and defense potential as a key component for intelligent wireless front-end integration. MEMS switches were first reported in the early 1990s [1]. MEMS devices quickly surpassed the RF performance of their solid-state equivalents. Even early MEMS switches had an insertion loss of 0.15 dB at 20 GHz, compared to an on-state insertion loss of approximately 1 dB for a typical GaAs-FET or PIN-diode switch at the same frequency. Today, MEMS switches, in particular, can be purchased with insertion losses as low as 0.1 dB up to 50 GHz with the potential for operating more than 100 billion cycles and handling multiwatt power levels, enabling the development of numerous types of reconfigurable antennas using this state-of-the-art technology [2].

Meanwhile, micromachining technology also has demonstrated a great potential for the successful implementation of intelligent, high performance, low cost, and low

Modern Antenna Handbook. Edited by Constantine A. Balanis

power-consumption wireless communication systems. Compact circuit designs are typically achieved in high index materials, which is in direct contrast to the low index substrates imposed by antenna performance requirements. The ideal solution requires the capability to integrate the planar antenna on electrically thick low index regions while the circuitry remains on the high index regions in the same substrate. In the past, this requirement was satisfied by selecting the substrate that offers optimum component performance; unfortunately, this has led to hybrid integration schemes and high development costs. As the frequency increases, however, this approach becomes increasingly difficult and costs are prohibitively high. To solve this problem, micromachining technology has been investigated and found to significantly help build high performance planar and three-dimensional antennas using modern semiconductor compatible technology to meet the integration requirement for antennas and other function modules.

This chapter reviews the latest advances in RF-MEMS integrated antennas, as well as various micromachining technologies used in high performance antenna design.

17.2 RF-MEMS SWITCHES AND MICROMACHINING TECHNOLOGIES

17.2.1 RF-MEMS Switches

RF-MEMS switches' functions are similar to traditional PIN and FET switches. The operating mechanism is different, however. MEMS switches are based on a technology that has made use of advantages of both solid-state and electromechanical systems and has overcome most of their disadvantages. MEMS switches are electromechanical components, while their fabrication is based on standard solid-state technologies. The electrostatic actuation mechanism seems to be the most common method for MEMS switches because of its low power consumption [3].

Currently, two types of MEMS switches have gained popularity: the *cantilever switch* and the *air-bridge switch* [4].

The cantilever structure consists of a thin metal strip fixed at one end, while being suspended over the transmission line with a gap of a few micrometers. The cantilever is connected in series with the transmission line as shown in Figure 17.1 [4]. Between the transmission line and the fixed end of the cantilever, there is an electrode for direct current (dc) biasing. The actuation voltage will pull down the membrane to close the gap and make a conducting path (Figure 17.2).

In the air-bridge type configuration, shown in Figure 17.3 [4], a thin metal membrane is suspended over the free space in the middle with the two ends anchored. This type of MEMS switch is usually configured in shunt to hang over the signal path in coplanar waveguide transmission lines. When an actuation voltage is applied to the control electrode, the electrostatic force causes the membrane to deform and be pulled downward. A dielectric layer is deposited on top of the center transmission line to prevent dc contact. In radiofrequencies, when the membrane is pulled down and touches the dielectric layer, a RF short-circuit condition can be achieved (Figure 17.4). Otherwise, the suspended membrane will not affect the signal transmission and will act as an RF open-circuit condition [3, 4].

RF-MEMS switches have experienced significant growth since they were first reported in the 1990s. The comparison with traditional electronic switches in Table 17.1 demonstrates the great advantages MEMS switches have offered.

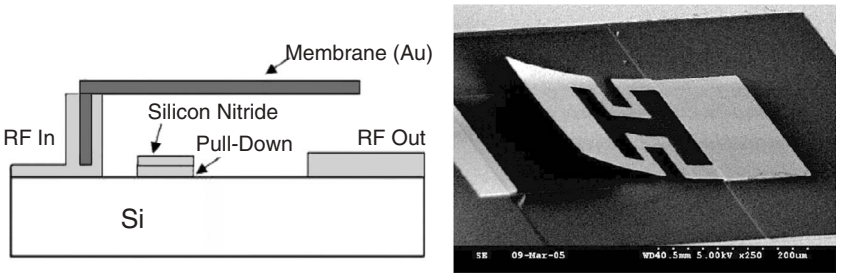


Figure 17.1 Cross section of cantilever RF-MEMS switches and SEM picture of such MEMS switch at up state. (From Ref. 4.)

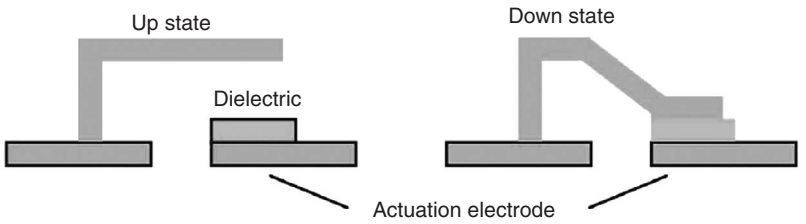


Figure 17.2 Cross section of cantilever RF-MEMS switches operating at up state and down state. (From Ref. 4.)

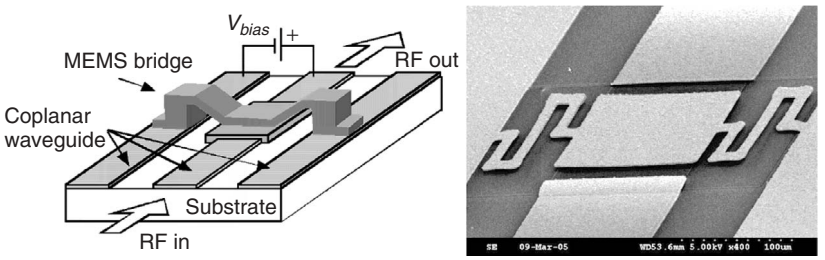


Figure 17.3 Illustration of an air-bridge RF-MEMS switches and SEM picture of an air-bridge MEMS switch. (From Ref. 4.)

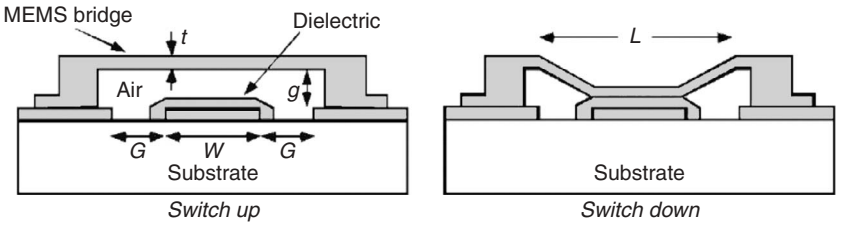


Figure 17.4 Cross section of air-bridge RF-MEMS switches operating at up/down state. (From Ref. 4.)

TABLE 17.1 Performance Comparison of FET, PIN Diode, and RF-MEMS Switches

Parameter	RF-MEMS	PIN	FET
Voltage (V)	20–80	3–5	3–5
Power consumption (mW)	0	3–20	0
Switching time	1–300 μs	1–100 ns	1–100 ns
Cutoff frequency (THz)	20–80	1–4	0.5–2
Loss (1–100 GHz) (dB)	0.05–2	0.3–1.2	0.4–2.5
Isolation (10–40 GHz)	Very high	Medium	Low
Isolation (60–100 GHz)	High	Medium	None

Source: From Ref. 1.

17.2.2 Micromachining Technologies

Silicon micromachining has been a key factor in the vast progress of MEMSs. Silicon micromachining is comprised of two technologies: *bulk micromachining*, in which structures are etched into silicon substrate, and *surface micromachining*, in which the micromechanical layers are formed from layers and films deposited on the surface [3].

Bulk micromachining allows for the selective removal of silicon from the substrate to form membranes on one side of a wafer, a variety of trenches, holes, or other structures [3] (Figure 17.5). The bulk micromachining technique can be divided into wet etching and dry etching. Both of them are used to realize micromechanical structures within the bulk of a single-crystal silicon wafer by selectively removing wafer material.

On the other hand, surface micromachining builds structures on the surface of the silicon by depositing thin films of "sacrificial layers" and "structural layers" and by removing eventually the sacrificial layers to release the mechanical structures (Figure 17.6). The prime advantage of surface-micromachined structures is their easy integration with integrated components on the same wafer.

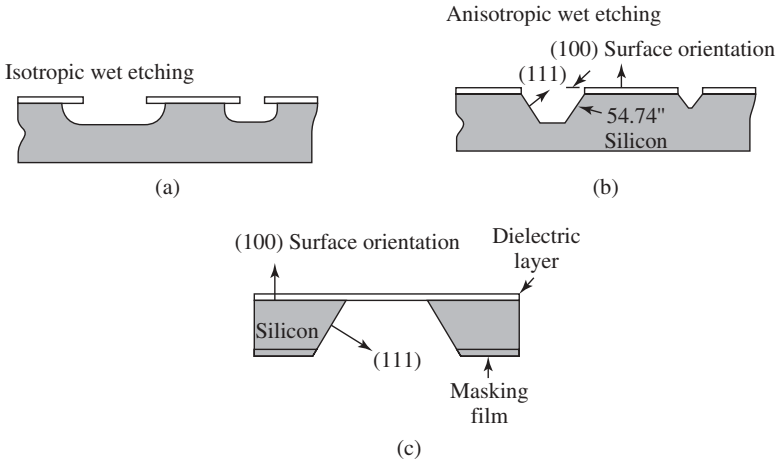


Figure 17.5 Bulk silicon micromachining: (a) isotropic wet etching, (b) anisotropic wet etching, and (c) dielectric membrane released by back-sidebulk etching. (From Ref. 5.)

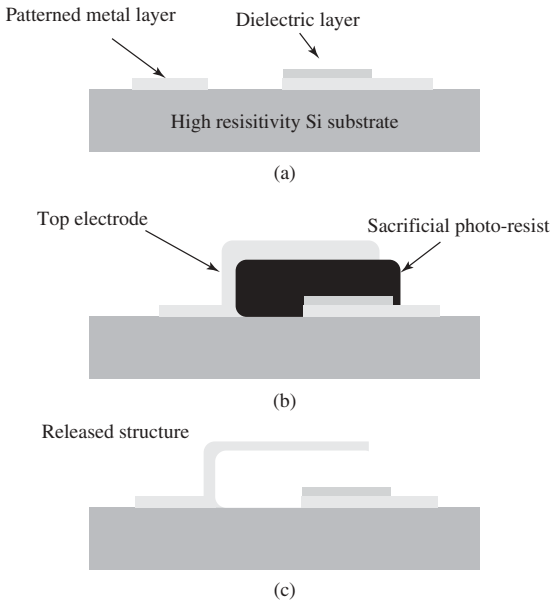


Figure 17.6 Processing steps of typical surface micromachining for a MEMS switch. (From Ref. 4.)

MEMS switches as well as micromachining technologies have been used in various applications. The remaining sections in this chapter introduce recent advances in the following areas:

- Antenna reconfigurability
- Mechanically movable antennas
- Phase shifters in antenna feed networks
- Antenna micromachining

17.3 RECONFIGURABLE ANTENNAS USING RF-MEMS SWITCHES

In general, reconfigurable antennas also include phased array antennas since they provide radiation pattern reconfigurability. A RF-MEMS switch-based phased array shows a drastic dc power reduction due to the use of ultralow power-consumption phase shifters. Compared with the designs using GaAs-FET switches or PIN diodes, the average loss of state-of-the-art 3-bit MEMS phase shifters can feature a 3–4-dB improvement [1]. In this section, we focus on a variety of designs of simple antennas with RF-MEMS switches capable of frequency/pattern/polarization reconfigurability, rather than phase shifters [6]. Phase shifters with MEMS switches are discussed in the next section.

17.3.1 Frequency Reconfigurable MEMS Antennas

The simplest way to control the resonant frequency of a microwave antenna is to change the length of the radiating wire/slot/patch using RF-MEMS switches. Three

examples are provided here, including a frequency reconfigurable half-wavelength dipole, a half-wavelength slot antenna, and a microstrip patch antenna. MEMS switches are placed symmetrically on the radiating arm of the strip dipole and slot to change the lengths of the arms with the MEMS switches turned "on" or "off." For the microstrip patch antenna, they are used to connect/disconnect floating patches to increase/decrease the patch length.

Figure 17.7 gives several designs for frequency reconfiguration [1]. Real designs are more complicated after including the required dc biasing for the switch.

Similar to other active switches, actuation voltages are applied via dc bias lines [11]. However, in order to prevent RF leakage into the dc path, careful attention needs to be given to dc bias lines themselves. Usually, the high resistivity line is used for dc biasing to prevent the radiation from the biasing directly. The following examples are included to provide more details for these design concepts with experimental verifications.

17.3.1.1 Frequency Reconfigurable Dipole Antenna Using a similar idea to that plotted in Figure 17.7, four cantilever switches were placed on each arm of the half-wave dipole antenna [1] and thus five resonating frequencies were observed when the switches were turned on successively [8]. This reconfigurable antenna can cover the whole X-band from 8 to 12 GHz. Four switches in total are connected in each arm to provide the whole shift of the bandwidth from 12 to 8 GHz (25%). When the first switch is turned "on," while the others are kept at their initial "off" state, the antenna resonates at a frequency 900 MHz lower than its previous state. The geometry is shown in Figure 17.8. (See also Figure 17.9.)

17.3.1.2 Frequency Reconfigurable Microstrip Patch Antennas For a microstrip antenna [1], using the cantilever MEMS switches to connect the patch radiating edge with adjacent floating strips can increase the length of the microstrip patch. Microstrip antennas with integrated RF-MEMS switches [9] designed in this way can operate at dual frequencies. Figure 17.10 shows such an antenna with circular polarization. The switches are incorporated into the diagonally fed square patch for controlling the operating frequency, and a rectangular stub attached to the edge of the patch acts as the perturbation to produce the circular polarization at 6.69 and 7.06 GHz.

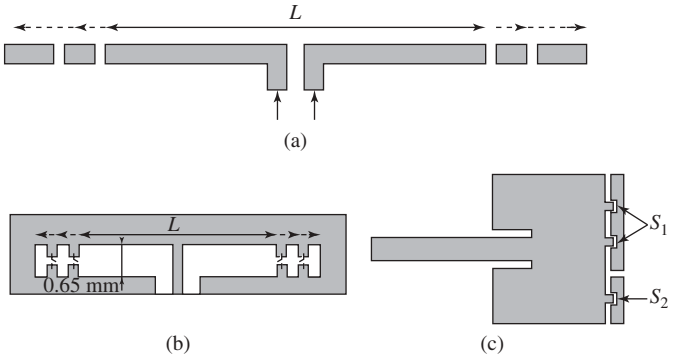


Figure 17.7 (a) Frequency reconfigurable half-wavelength dipole, (b) half-wavelength slot antenna, and (c) MEMS tunable microstrip patch antenna. (From Ref. 1.)

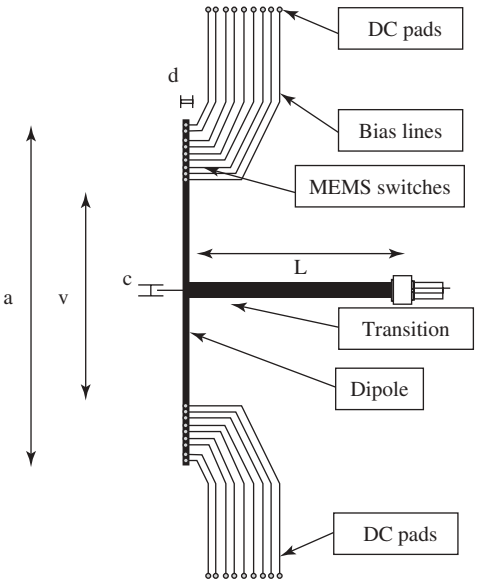


Figure 17.8 Topside view of the reconfigurable planar dipole antenna design. (From Ref. 8.)

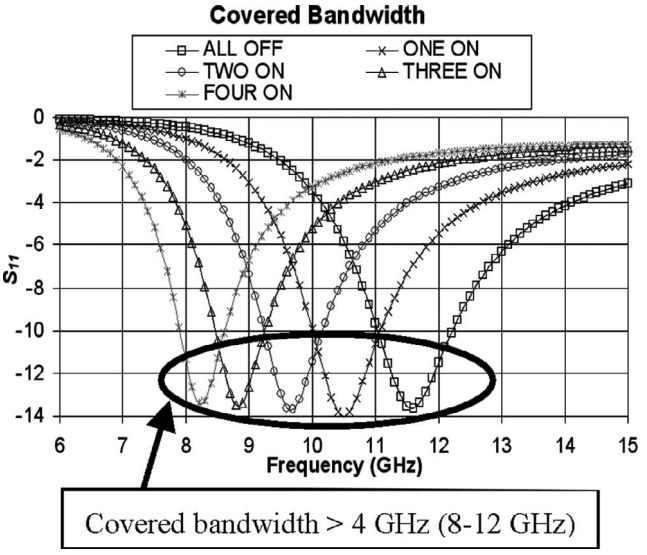


Figure 17.9 Return losses when the MEMS switches are successively activated. (From Ref. 8.)

17.3.1.3 Frequency Reconfigurable Slot Antenna Reconfigurable slot antennas (RSAs) can also be implemented with RF-MEMS switches [10]. This can nest an inner slot ring with an outer one. As shown in Figure 17.11, when the MEMS switch S1 is not activated (i.e., the microstrip segments are disconnected), the outer slot radiates ($f_{1ow} = 2.4$-GHz operation); when S1 and two other MEMS switches, S2 and S3 (double

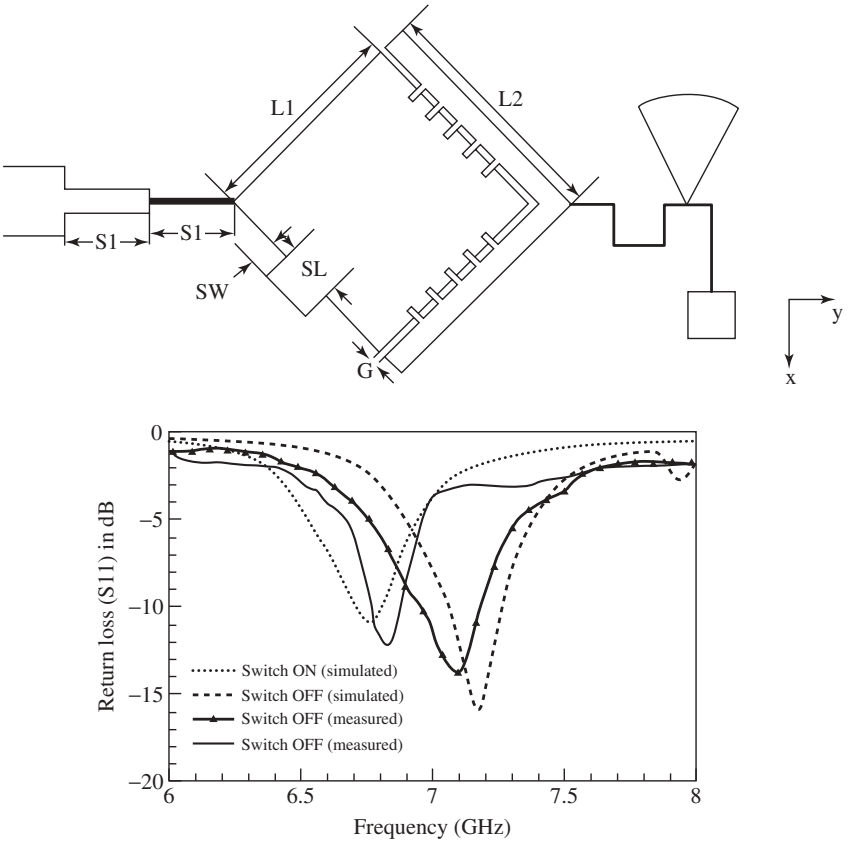
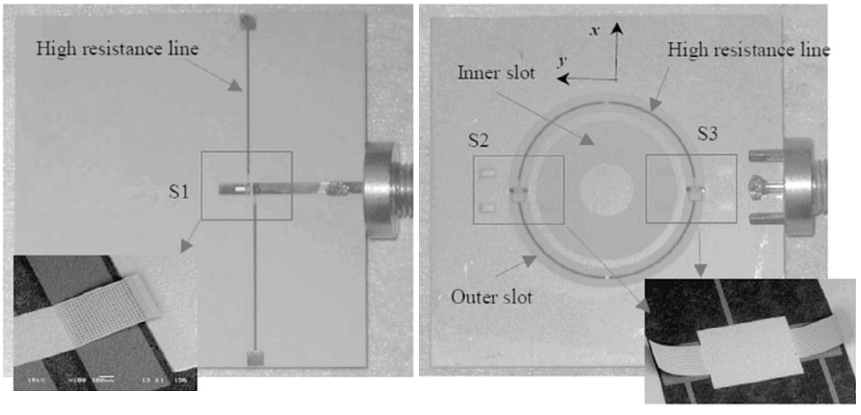


Figure 17.10 Circularly polarized MEMS frequency reconfigurable patch antenna and measured performances. (From Ref. 9.)



(a) (a)

Figure 17.11 Photographs of the MEMS reconfigurable slot ring antennas. (From Ref. 10.)

arm switch; see the inset of Figure 17.11b) spanning the outer slot, are all activated, the outer slot is shorted to the RF ground and the inner slot is fed since the backside microstrip segments are connected; the inner slot ring radiates ($f_{high} = 5.3$-GHz operation). This antenna finds application in WLAN for different standards.

17.3.1.4 Frequency Reconfigurable Fractal Antennas Frequency reconfigurability can also be designed using fractal antennas. The fractal antenna has the important property that radiation patterns remain almost constant due to the self-similarity of the antenna. Two design examples are given here. In the first example [11], ohmic contact cantilever RF-MEMS switches are integrated with self-similar planar antennas to provide a reconfigurable fractal antenna system that radiates with similar patterns over a wide range of frequencies (Figure 17.12). High resistivity dc biasing lines are used for this design. The antenna's simulated and measured reflection losses are shown in Figures 17.13 and 17.14. When the dc probes come down and touch the biasing pads, but

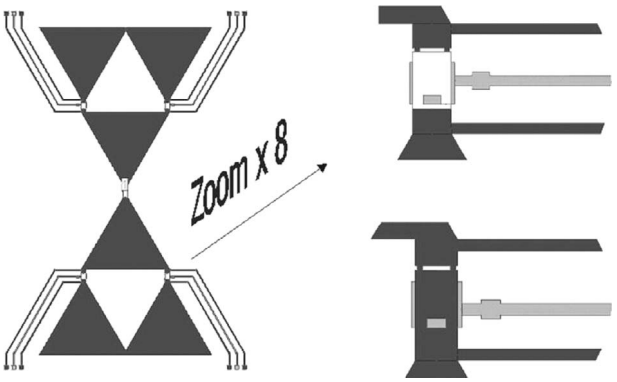


Figure 17.12 A second-order Sierpinski fractal triangular patch antenna with MEMS switches. (From Ref. 11.)

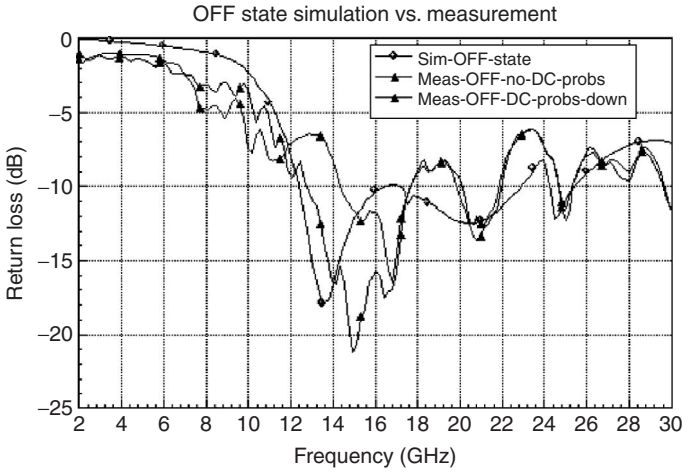


Figure 17.13 Simulated and measured return loss when MEMS switches are off and only one triangular patch closest to the feeding point radiates. (From Ref. 11.)

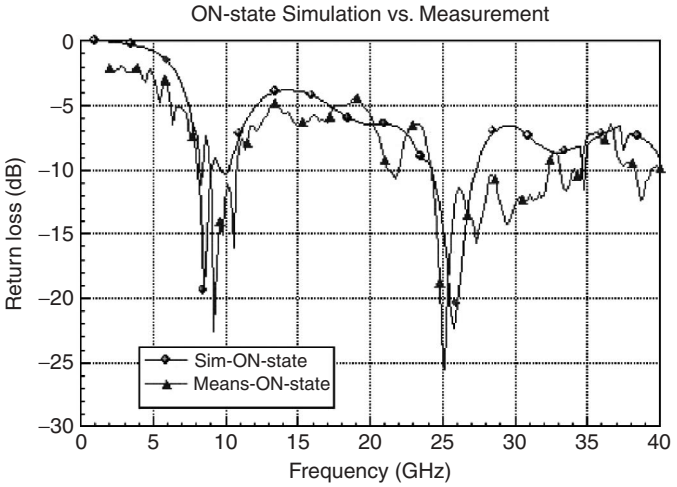


Figure 17.14 Simulated and measured return loss when MEMS switches are on and all four triangular patches radiate. (From Ref. 11.)

no voltage is applied to them, the antenna resonates at 15 GHz, with a bandwidth from 12.75 up to 17.5 GHz. When all switches are in the "on" state, the antenna resonates at two frequencies, very close to the simulated ones. The first occurs at 9.2 GHz, with a bandwidth of 1.6 GHz, and the second at 25.2 GHz, with a bandwidth from 24.3 up to 28.2 GHz. The second resonance is harmonic. In contrast to the next example that has no need for dc biasing lines, MEMS switches in this example have identical electrical and mechanical characteristics. Separate biasing lines are included for different switches. The switches are controlled independently.

When the order of the fractal antenna goes higher, more switches are needed. This will introduce a tremendous interference from dc biasing lines. The solution is to use multiple switches with different actuation voltages [7]. This can add another level of reconfigurability to fractal antenna design. This type of MEMS switch was used to sequentially activate and deactivate parts of a multiband antenna between 2.4 and 18 GHz while maintaining its radiation characteristics. Figure 17.15 shows four different cantilever switch geometries. The stationary posts are shown in black. The dotted areas show the electrostatic regions. With the different membrane geometries, the switches can be actuated from lowest actuation voltage (lowest spring constant) to the highest actuation voltage (highest spring constant). Table 17.2 shows the designed and measured pull-down voltages for the four geometries.

Since different MEMS switches use sequentially increased actuation voltages, all of the switches can share a common dc feed. This technique eventually provides reconfigurability without the need for dc bias lines (See Figures 17.16–17.18.) This is advantageous since dc bias lines take up space, add loss, and can reduce the bandwidth of a device. Four resonances can be observed from the measurement to be 2.4 GHz, 5.7 GHz, 9.4 GHz, and 18 GHz, respectively, for different applications. Radiation patterns remain almost constant for all four frequencies.

Other configurations can also function similarly [12]. Figure 17.19 shows a reconfigurable maze antenna based on fractal loop structures, where switching devices are used

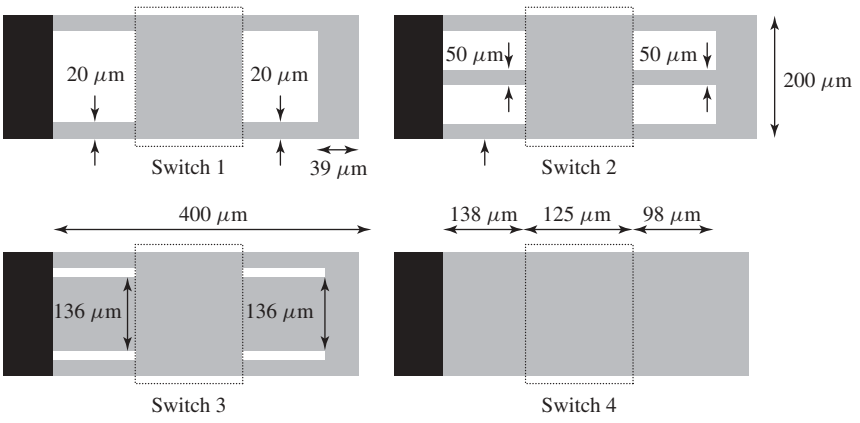


Figure 17.15 Four different cantilever switch geometries (top view) with different activation voltages (from the lowest to the highest). (From Ref. 7.)

TABLE 17.2 Calculated and Measured Pull-Down Voltages

Switch	Calculated Voltage	Measured Voltage	Difference	Percent Error
1	17.45 V	18 V	0.55 V	3.15%
2	26.19 V	28 V	1.81 V	6.91%
3	36.54 V	38 V	1.46 V	4.00%
4	39.21 V	42 V	2.79 V	7.12%

Source: From Ref. 7.

to alter the interconnectivity of the loop to build the different lengths of the resonance. Four different frequency bands (GSM/WCDMA/WLAN 802.11 a, b, g) are separately covered while their radiation patterns and input impedances remain almost the same.

17.3.2 Radiation Pattern Reconfigurable Antennas

To provide spatial diversity for applications in multiple-input multiple-output (MIMO) systems, radiation pattern reconfigurable antennas using MEMS switches have attracted lots of attention. Designs of this type of antenna usually fall into three categories. The first design connects several directional antennas together, with each one pointing in a different direction. When a certain feeding path is activated by MEMS switches, the radiation beam is pointed to a specified direction. The second design activates parasitic radiators using MEMS switches to make different combinations of the radiators to point the radiation beam at different angles. The third design plays with the different operation modes for a single antenna. Several antennas intrinsically can be operated for different radiation modes when the physical dimensions change. Several design examples are provided next to demonstrate these three ideas.

17.3.2.1 Activating Different Feeding Paths Using MEMS Switches For this type of reconfigurable antenna, MEMS switches feed one antenna at a time. As shown in Figure 17.20 [13], two coplaner waveguide (CPW)-fed slot antennas are activated by

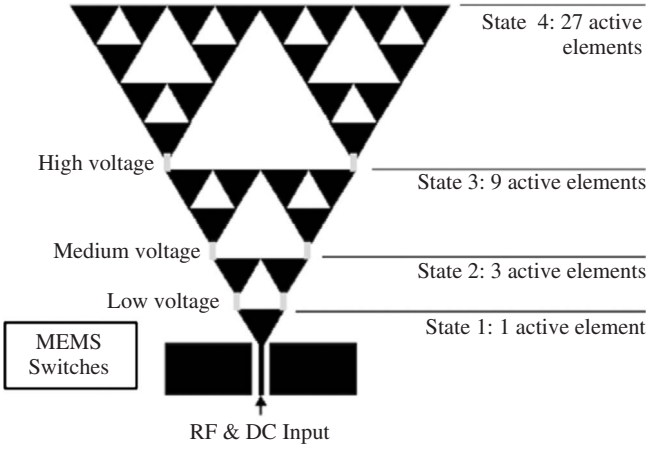


Figure 17.16 A fourth-order Sierpinski fractal triangular patch antenna with MEMS switches without using dc biasing lines. (From Ref. 7.)

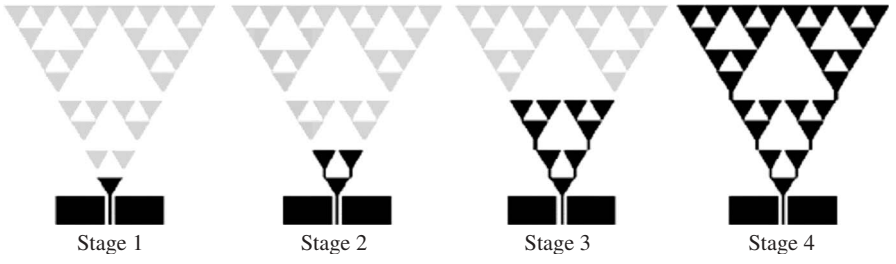


Figure 17.17 Sequentially switching between different operating states for the fourth-order Sierpinski fractal patch antenna with MEMS switches without using dc biasing lines. (From Ref. 7.)

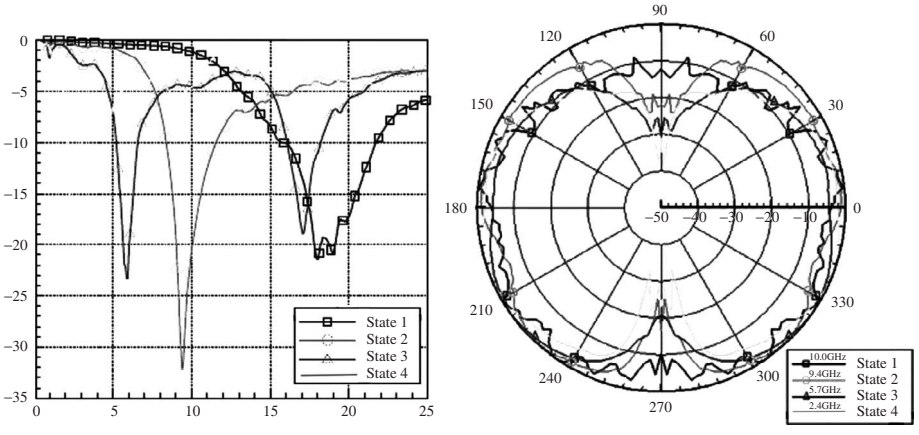


Figure 17.18 Measured return loss for the fourth-order Sierpinski fractal triangular patch antenna with MEMS switches without using dc biasing lines. (From Ref. 7.)

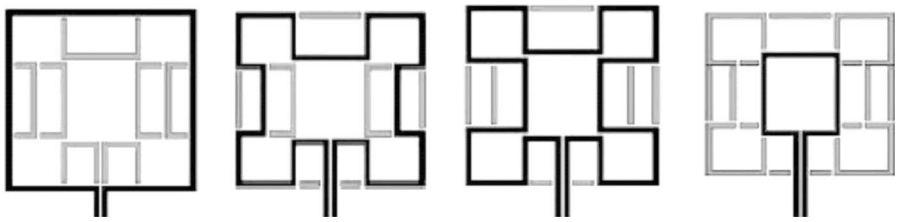


Figure 17.19 Maze antenna configurations at (a) 2.4 GHz, (b) 2.0 GHz, (c) 1.8 GHz, and (d) 5.2 GHz. (From Ref. 12.)

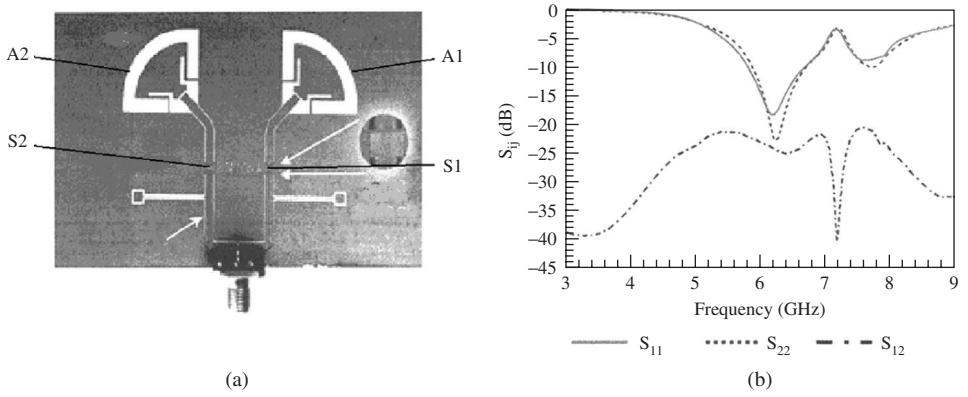


Figure 17.20 (a) Photograph of the monolithic RF-MEMS switched diversity antenna. (b) Measured return loss and mutual coupling for the RF-MEMS switched diversity antenna. (From Ref. 13.)

MEMS switches on the feeding line separately. The left antenna A2 is controlled by the MEMS switch S2 and the right antenna A2 is controlled by S1. The input impedance matching remains unchanged when the different radiating paths are in use because of the symmetry. It is important to maintain low mutual coupling between different radiating elements; thus when one feeding path is activated, the isolation to the other feeding paths should be high. As can be seen in Figure 17.21, the radiation pattern changed dramatically.

This design can be extended to include more radiating elements if the individual antenna has a narrower beamwidth. A design example using three CPW-fed cactus slot antennas is also shown here. A more complicated switching network needs to be designed, as shown in Figure 17.22 [14]. (See also Figure 17.23.)

17.3.2.2 Different Parasitic Radiator Combinations Using MEMS Switches

In the previous design, although the unused antenna is not fed when the MEMS switch on its feeding path is at the "off" position, it can still disturb the radiation pattern, resulting from the electromagnetic wave coupling. This generates the relatively high sidelobes and the radiation pattern is not completely reconfigured to the desired direction.

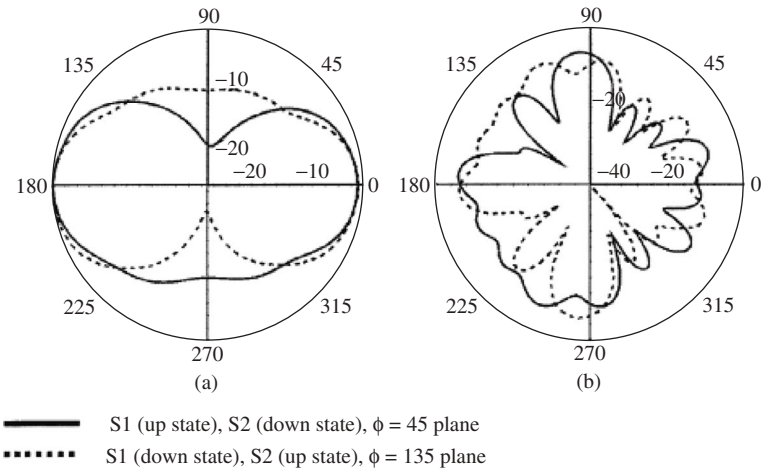


Figure 17.21 Measured copolarization radiation patterns of the RF-MEMS switched diversity antenna at 6.15 GHz. (a) In elevation planes, $\phi = 45^\circ$ and $\phi = 135^\circ$ planes. (b) In azimuth plane (xy-plane). (From Ref. 13.)

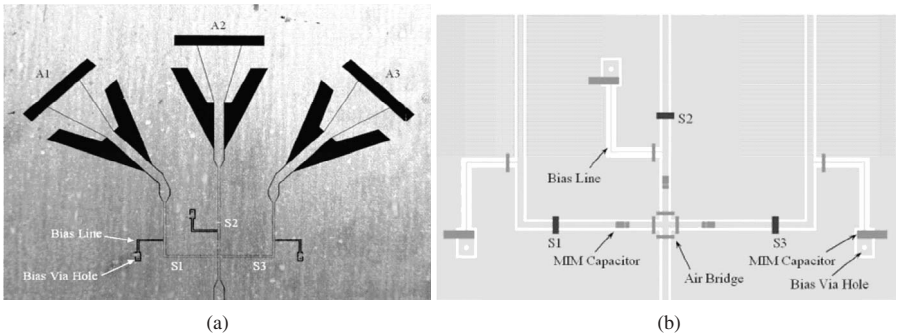


Figure 17.22 (a) Photograph of the radiation pattern antenna integrated with RF-MEMS switches for selection combining. S1, S2, and S3 are RF-MEMS switches. A1, A2, and A3 are cactus antennas. (b) Detailed schematic of the feeding structure. (From Ref. 14.)

A design to overcome this limitation is to completely shut down the parasitic radiation from these unwanted antennas by changing their physical geometries using MEMS switches, not just the feeding lines for them. One design example is in Figure 17.24 [15]. Slot 1 is fed directly by a microstrip line. Slot 2 and slot 3 are parasitic antennas and act as directors in this three-antenna array. These two slots are loaded by CPW stubs connected in the middle of each slot. The lengths of these CPW stubs can be reconfigured using MEMS switches spanning these stubs. By changing the states of the MEMS switches on two parasitic slots, only one slot radiates at a time; the other one does not radiate at all. Furthermore, shielding the bias line allows a neat radiation pattern. The two bias lines are covered by a thin adhesive dielectric layer and adhesive copper (not shown in Figure 17.24), to shield these slots at radiofrequencies and avoid any degradation of the radiation pattern. Experimental results show good agreement with simulations with

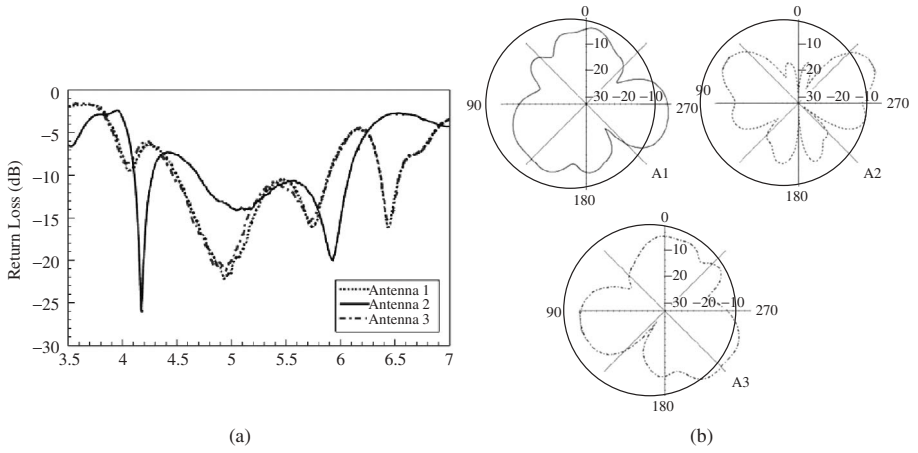


Figure 17.23 (a) Measured return loss of individual selected antenna. (b) Measured copolar radiation patterns at 5 GHz of each antenna element in the *xy*-plane when the three switches are sequentially activated. (From Ref. 14.)

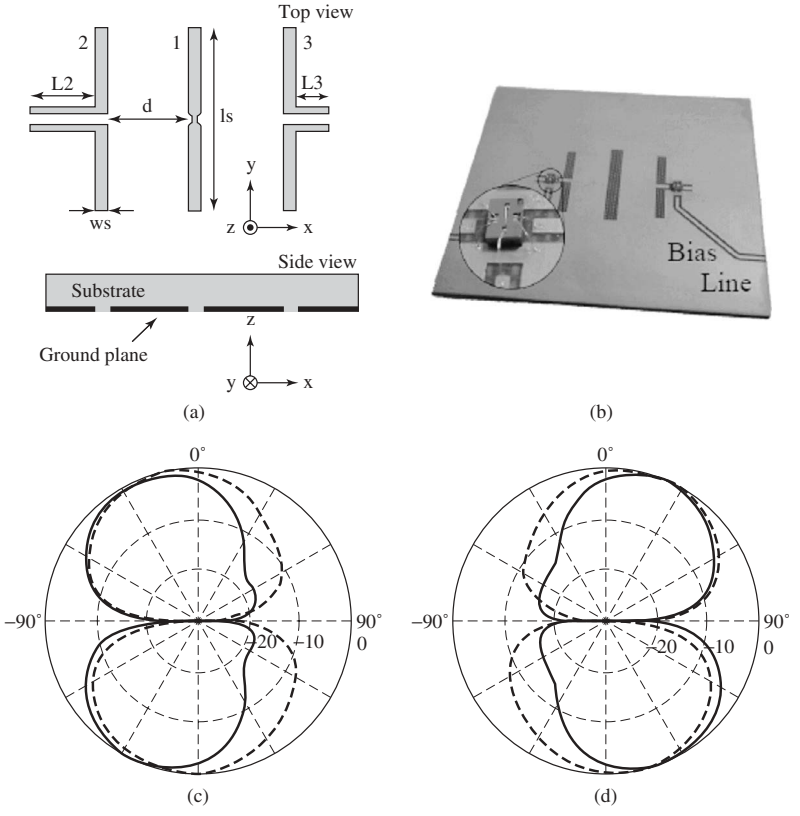


Figure 17.24 (a) Schematic of radiation pattern configurable antenna with parasitic slots, (b) photography of the prototype, and (c, d) simulated (solid) and measured (dashed) radiation patterns at 5.6 GHz for (c) on–off and (d) off–on states. (From Ref. 15.)

two switched beams at $\pm 25^{\circ}$. Compared with the results in the previous section, a very clear radiation pattern can be observed.

17.3.2.3 Multimode Radiators with MEMS Switches The third design uses a multimode antenna whose physical geometry can be reconfigured by MEMS switches to operate at different modes. Different radiation patterns are associated with different modes. A spiral antenna is a typical multimode radiator. It can generate either broadside or endfire radiation [16].

The antenna is a single-turn square microstrip spiral, shown in Figure 17.25 [16]. The total linear dimension of the antenna is 81 mm, and it is fabricated on Duroid 5880 substrate with $\varepsilon_r = 2.2$ and $h = 3.175$ mm. To excite this resonant antenna, a vertical SMA probe is connected to the interior end of the spiral. The outer end of the spiral is shorted to ground with a via. In the endfire configuration ($\theta = 90^{\circ}$, $\phi = 0^{\circ}$ with respect to the z-axis), switch 1 is closed and switch 2 is open, creating the single-turn square spiral geometry. For the broadside configuration ($\theta = 0^{\circ}$ with respect to the z-axis), switch 1 is open and switch 2 is closed. In this design, commercially available packaged MEMS switches are used [17]. The switches are glued to the substrate using epoxy and then wire-bonded to the microstrip lines of the antenna.

Figures 17.26–17.28 present voltage standing wave ratio (VSWR) and radiation patterns for the two primary elevation cut-planes ($\phi = 0^{\circ}$ and $\phi = 90^{\circ}$) in both pattern configurations (broadside and endfire). The resonating frequencies for both configurations remain almost unchanged.

Similar results were reported with on-wafer integrated MEMS switches and interested readers can refer to Ref. 18. to find more details.

17.3.3 Polarization Reconfigurable Antennas

MEMS switches can also be used to change antenna radiation polarization. An extruding metal piece from the edge or cuts into the patch can change the radiation polarization

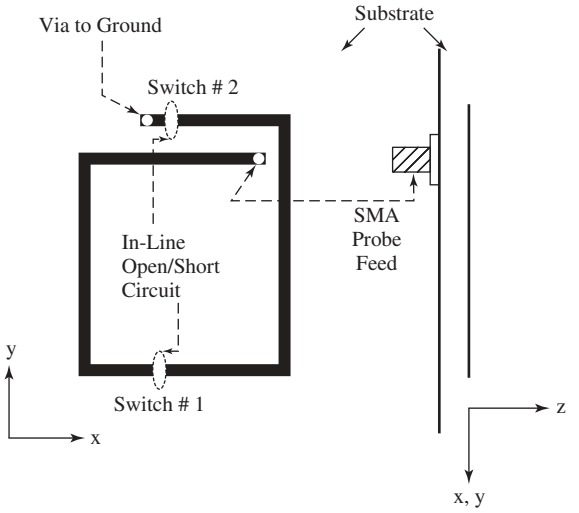


Figure 17.25 Pattern reconfigurable antenna geometry including switch locations. (From Ref. 16.)

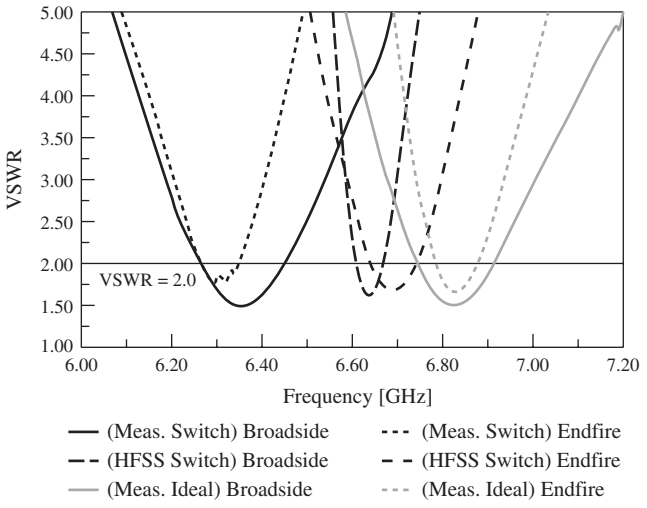


Figure 17.26 Measured and simulated VSWR of both antenna configurations. The overall shared 2:1 VSWR (impedance) bandwidth is 80.0 MHz. (From Ref. 16.)

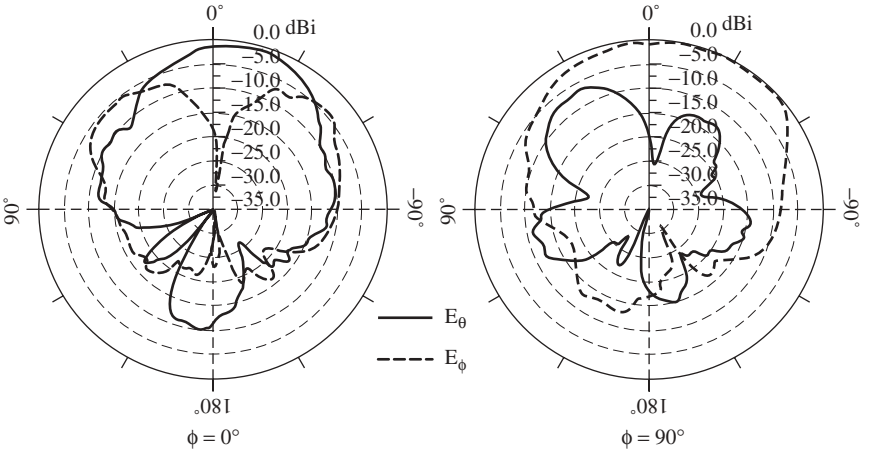


Figure 17.27 Measured broadside configuration radiation patterns for the two primary elevation cut-planes $\phi = 0^\circ$ (left) and $\phi = 90^\circ$ (right) using Radant RF-MEMS switches 17. (From Ref. 16.)

of a microstrip antenna. MEMS switches are used to control the connectivity of these mode-perturbing pieces or slots.

A nearly square patch antenna with MEMS switches is shown in Figure 17.29 [19]. The switch consists of an air bridge suspended over a metal stub. The overpass is supported at both ends by metalized vias, which are electrically connected to the nearly square patch antenna. The metal overpass is actuated by an electrostatic force of attraction set up by a voltage applied between the overpass and the metal stub. A dielectric film deposited over the metal stub prevents friction when the surfaces come in contact. The nearly square patch antenna with notches illustrated in Figure 17.29 is designed to support two

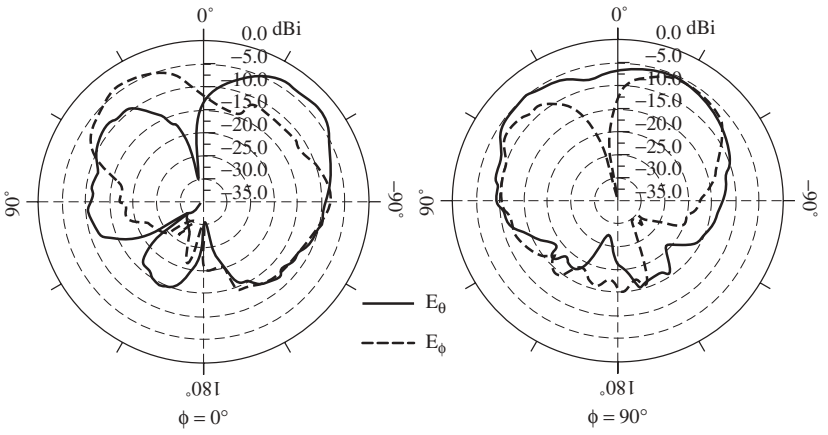


Figure 17.28 Measured endfire configuration radiation patterns for the two primary elevation cut-planes $\phi = 0^\circ$ (left) and $\phi = 90^\circ$ (right) using Radant RF-MEMS switches 17. (From Ref. 16.)

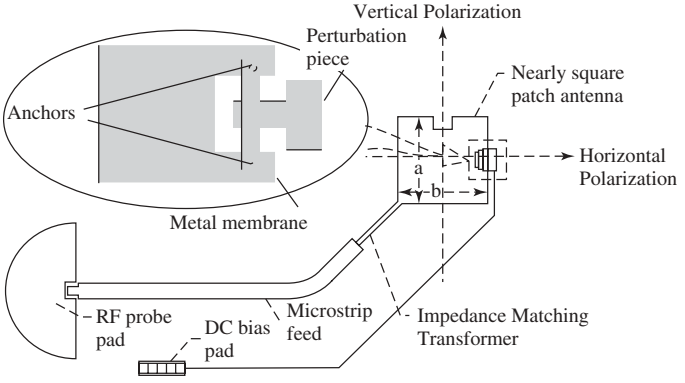


Figure 17.29 Polarization reconfigurable patch antenna element with integrated MEMS switches. Inset shows details of the MEMS switched integrated perturbation piece. (Adapted from Ref. 19.)

degenerate orthogonal modes when excited. When the MEMS actuator is in the "off" state, the perturbation of the modes is negligible and hence the patch radiates a circularly polarized (CP) wave. When an electrostatic force resulting from the application of a bias pulls down the overpass, the MEMS actuator is in the "on" state. This action perturbs the phase relation between the two modes, causing the patch to radiate dual linearly polarized (LP) waves. In measurement, the patch is well matched to the 50-ohm feed line and resonates at a frequency of 26.7 GHz. The measured radiation patterns along the two orthogonal planes are shown in Figure 17.30. The measured axial ratio at boresight is about 2.0 dB. In the "on" state the patch is also well matched to the 50-ohm feed line and resonates at a frequency of 26.625 GHz. In the "on" state, the patch radiates dual linearly polarized waves. The measured E- and H-plane radiation patterns for the vertical polarization are shown in Figure 17.30. Similar radiation patterns are observed for the horizontal polarization.

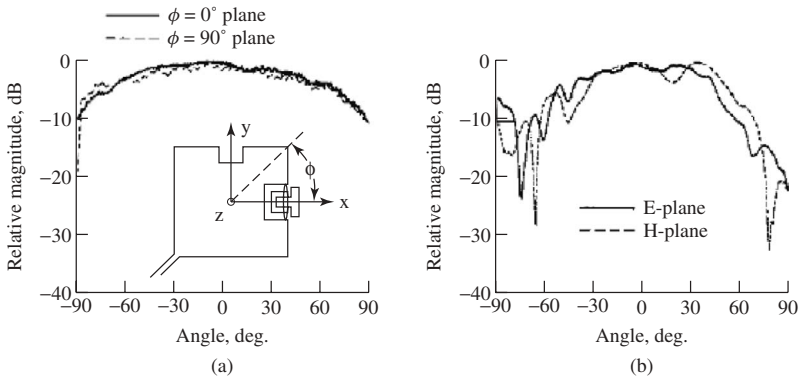


Figure 17.30 (a) Measured circular polarized radiation patterns. (b) Measured linearly polarized radiation patterns for vertical polarization. (From Ref. 19.)

Spiral antennas can also be used for polarization reconfiguration. A spiral antenna can alter radiation between right-hand circular polarization (RHCP) and left-hand circular polarization (LHCP), depending on spiral winding directions. The following example introduces a multipixel design for polarization reconfigurable antennas [20]. A metal patch matrix (pixels) is patterned and the connection between the adjacent pixels is controlled by MEMS switches. Each pixel can be connected with four adjacent pixels. By connecting different paths with MEMS switches, the pixel matrix can be configured to either a clockwise or counterclockwise spiral antenna and thus radiates RHCP or LHCP waves. Biasing voltages are configured as in Figure 17.31a, b to get the equivalent spiral winding directions shown in Figure 17.31c, d.

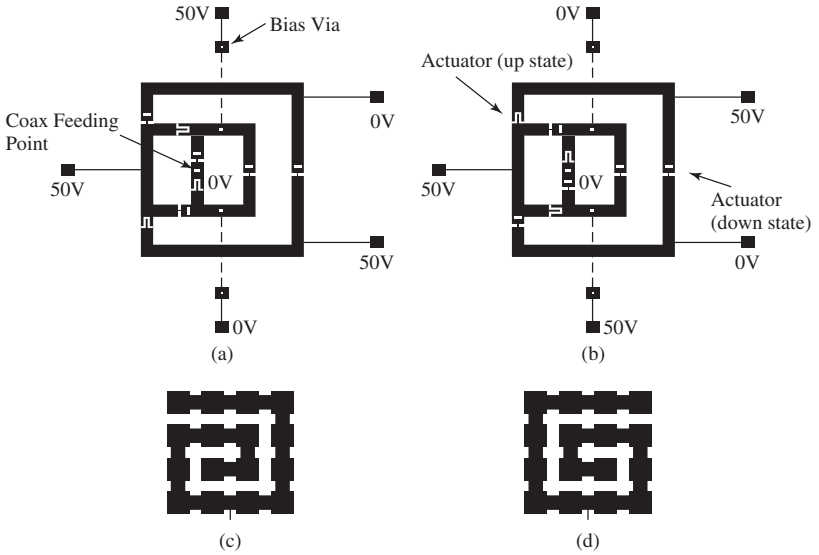


Figure 17.31 Top view schematics (c, d) of an MRA spiral corresponding to each reconfigurable mode of operation: (a) left-hand circular polarization and (b) right-hand circular polarization. (From Ref. 20.)

17.4 MECHANICALLY MOVABLE ANTENNAS

Besides electrically steering the antenna radiation beam, the possibility of steering the antenna radiation beam in millimeter wave applications using an electrostatic force has also been exploited. The section briefly reviews three examples based on this design concept.

17.4.1 Microelectromechanical Reconfigurable Vee Antenna Capable of Beam Scans

In the first example, MEMS actuators are used to adjust the angles of two arms of a Vee antenna operative at 17.5 GHz (Figure 17.32). When both arms are rotated together in one direction, the beam is scanned. A prototype antenna has been built and the experimental results proved the concept [21].

17.4.2 Two-Dimensional (2D) Mechanical Steering Patch Antenna Array for V-Band Applications [22]

The second example is a mechanical beam-steering antenna with two degrees of freedom of motion working at the V-band, centered at 60 GHz. Figure 17.33a shows the schematic view of the proposed antenna. High resistivity silicon is used as an actuating plate on top of which microstrip patch antennas are patterned. The actuating plate is located at the center of the whole structure and rotates around the two torsional hinge pairs. Magnetic actuation is used as the driving mechanism instead of electrostatic actuation. In the latter case, a very small gap between the electrodes is required to generate a large electrostatic force for a reduced driving voltage. However, the small gap limits the maximum rotating angle. In this work, magnetic material is formed on the backside of the structure for the actuation. Magnetic materials are placed on the edges of both the center plate and outer frame to independently control the motion in both directions.

Figure 17.33b shows the measured radiation patterns when the antenna structure is rotated by magnetic force to -25°, 0°, and 10°, respectively. The asymmetric radiation patterns after rotating are caused by reasons such as interference of the patterns with the surrounding frame structure and parasitic reflection from the measurement apparatus mounted on the backside of the sample (e.g., the big chunk of driving coils).

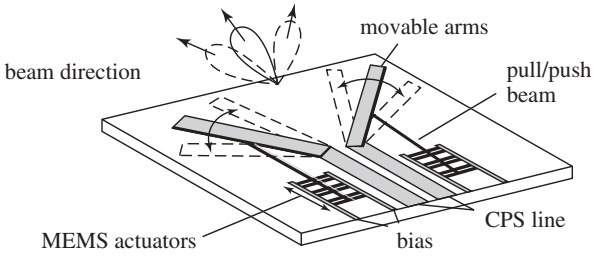
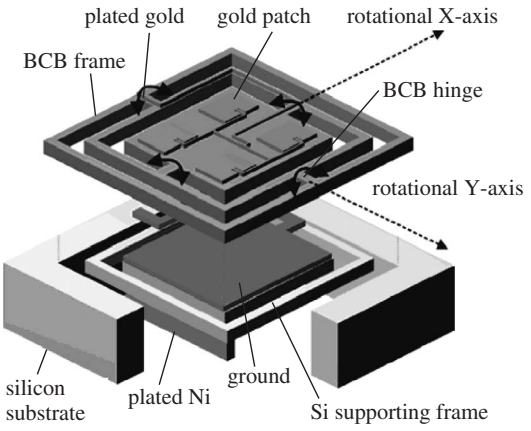
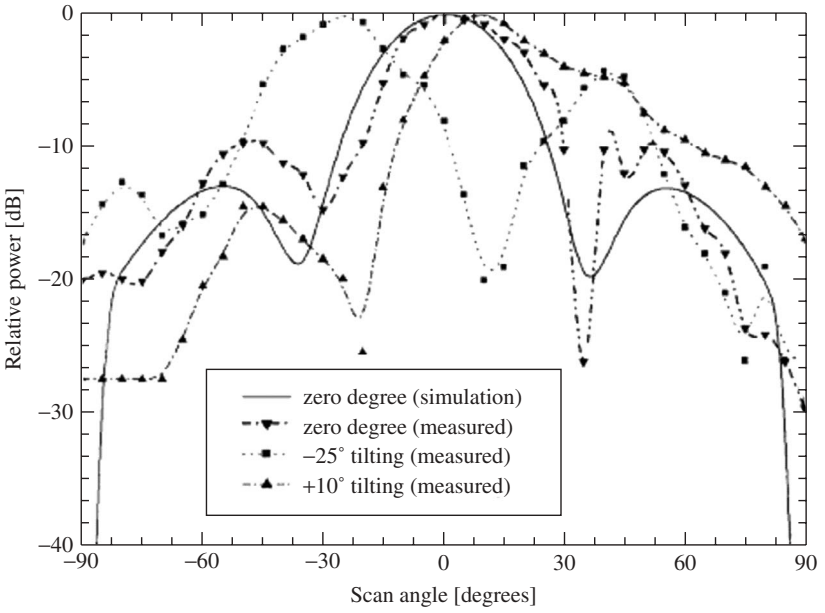


Figure 17.32 Micromachined Vee antenna for beam shaping and beam steering. (From Ref. 21.)



(a)



(b)

Figure 17.33 (a) Architecture of 2D beam-steering array. (b) Simulated and measured radiation patterns that were steered between -25° and $+10^{\circ}$. (From Ref. 22.)

17.4.3 Deployable Microstrip Patch Antenna by MEMS Actuator

The third example is a mechanical movable antenna for frequency reconfigurable applications [23], as shown in Figure 17.34. This MEMS antenna consists of a fixed microstrip patch antenna and a wrapped metallization on the edges. Wrapped metal is deployed back to the substrate surface after applying voltage between the conductive substrate and the patch. The patch area extends after the wrapped metal is pulled down. The antenna operates at 23.8 GHz unbiased, and 12.4 GHz with 150 volts applied.

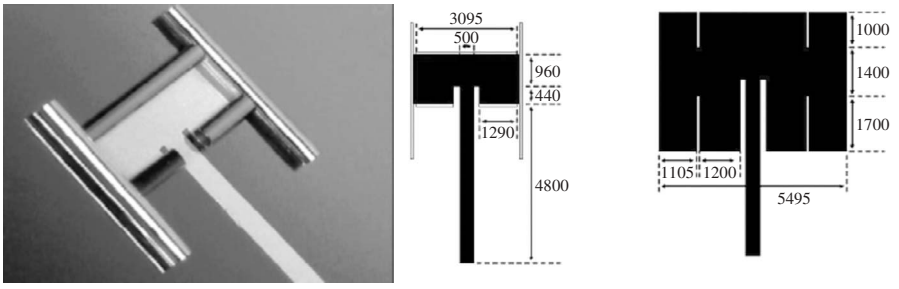


Figure 17.34 (a) Photograph of a foldable microstrip patch antenna with wrapped metal on the edges. (b) Geometries and dimensions of the patch before and after unfolding metal strips. (From Ref. 23.)

17.5 MEMS PHASE SHIFTERS IN ANTENNA FEED NETWORKS

Developing low loss and low cost phase shifters is essential to implementation of phased array antennas for beam-steering applications [24]. In phased array antennas, each radiating element is connected with a feeding line with a certain electrical length to get the required phase delay. If the combined radiation beam needs to be pointed in another direction, the phase delay provided by each feeding line ought to be changed accordingly. Since physical lengths of feeding lines are fixed, one way to realize the required dynamic phase change without physically rotating the array is by replacing fixed feeding lines with electrically tunable transmission lines—phase shifters, which can provide several phase states or continuous variation of phase delays controlled by the dc biasing.

As mentioned previously, MEMS switches are low loss control components. By replacing PIN diodes or FET switches with MEMS switches, the power consumption of phase shifters can be greatly reduced and this will lead to a simpler and lower cost system due to the smaller number of transmitters/receivers (T/R) to drive the antenna element [25].

There exist several different types of MEMS phase shifters, including the *switched-line phase shifter*, the *distributed phase shifter*, and the *reflect-line phase shifter*. The first two types are the dominant current designs and more details on them are given in this section.

Integrated MEMS phase shifters with superior performances have been reported for X-band, Ka-band, V-band, and W-band applications. Various substrates have been explored for MEMS phase shifter integration, including silicon, GaAs, quartz, and laminated organic substrate such as liquid crystal polymer (LCP) [2, 25–30]. Switched-line MEMS phase shifters have been demonstrated with a measured on-wafer loss of approximately 0.3 dB/bit at 10 GHz and 0.7 dB/bit at 35 GHz [25, 26]. A distributed MEMS phase shifter was first developed by Barker and Rebeiz [27]. Its performance is competitive and even better than the switched-line approach at 40 GHz and above. Measured results indicate a loss of 0.6 dB/bit at 14 GHz, 0.7 dB/bit at 38 GHz, 0.7 dB/bit at 65 GHz, and 0.9 dB/bit of loss at 75–110 GHz [28–30]. The data indicate that MEMS-based phase shifters are undoubtedly suitable for high performance phased array antenna applications.

17.5.1 Switched-Line MEMS Phase Shifters

The earlier MEMS phase shifter designs use MEMS switches to choose between various signal paths with different physical lengths. MEMS switches are ideal for switched-line topologies because of the excellent isolation they can provide. Figure 17.35 plots the basic circuit for switched-line architecture. Two different paths with a phase delay difference of $\Delta\phi$ are chosen by MEMS switches on the paths.

The phase shifter in Figure 17.35 has only two different phase delay states, so the phased array can only be reconfigured to two directions. To use it in a large-scale array, it should be cascaded to provide multiple phase delays for more beam angles. Figure 17.36 shows two examples of 4-bit cascaded switched-line MEMS phase shifters [31, 32]. In Figure 17.36b, delay lines are curved to get better impedance matching.

Multibit operation can also be realized using a series-shunt configuration. This configuration is claimed to have better performance such as low loss and compactness. This performance level is achieved by using single-pole four-throw (SP4 T) MEMS switches (Figure 17.37a), which reduces the number of switches on any signal path by half; for

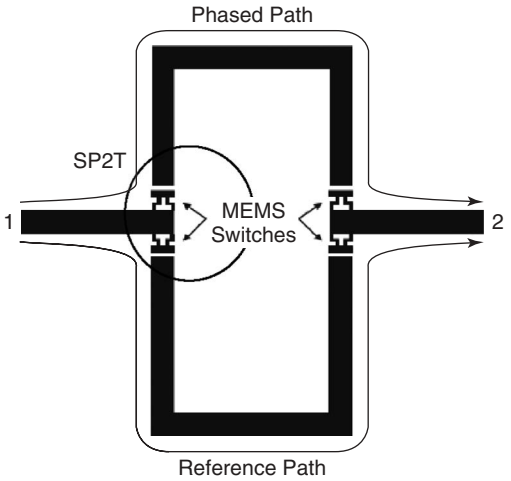
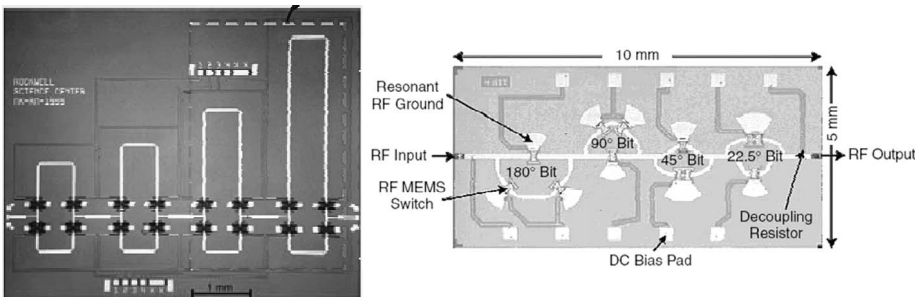


Figure 17.35 Illustration of the basic concept of switched-line phase shifter with MEMS switches.



(a) (b)

Figure 17.36 (a) Rockwell Science 4-bit dc 40 GHz MEMS phase shifter [31]. (b) Raytheon 4-bit Ka-band MEMS phase shifter [32].

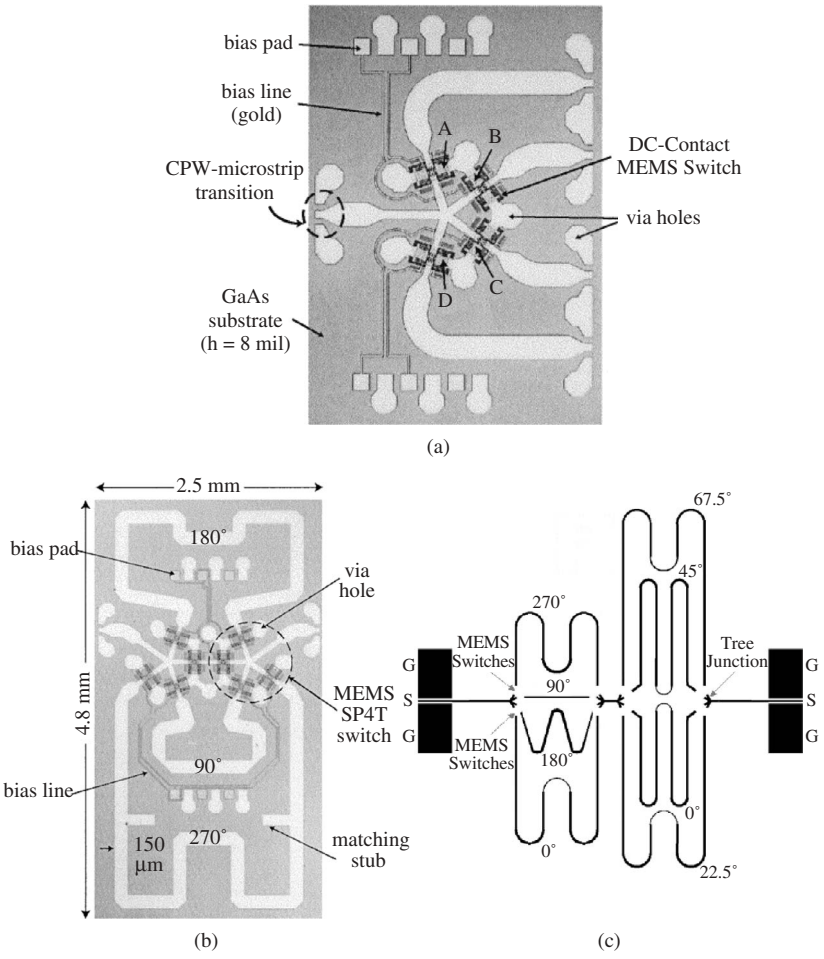


Figure 17.37 (a) Circuit layout of a SP4T MEMS switch [25]. (b) A 2-bit phase shifter with SP4T MEMS switches (From Ref. 25). (c) A series-shunt configuration for 4-bit phase shifter on LCP substrate. (From Ref. 2).

the 2-bit with cascaded single-pole two-throw (SP2T) configurations, each path consists of two delay lines and four switches (Figure 17.36). Two design examples using this configuration are shown in Figure 17.37b, c.

Both series and shunt MEMS switches can be used on the phase delay path. For a series configuration, turning on the MEMS switches on both ends of a line will choose that delay path. The MEMS switches can be placed right after the power dividing junction, as indicated in Figure 17.37c; when the switches are turned off, an open-circuit condition will not affect the impedance matching at the junction. For a shunt configuration, turning on the MEMS switches on the path will short the signal path with the ground, as indicated in Figure 17.36b. (Note that radial stubs are used here. Since a quarter-wavelength radial stub transforms an open-circuit condition on the edge to a short-circuit boundary condition at the root of the stub, the MEMS switch is equivalently shorted to the ground when pulled down.) When MEMS switches are placed a quarter-wavelength away from the

junction, after the transformation, the impedance presented to the junction will still be open circuit and this will improve the isolation between different paths.

However, switched-path type phase shifters are inherently digital ones and quickly become cumbersome if a large number of phase states are desired. Devices such as phased antenna arrays often require high resolution in phase control, which would lead to a large and lossy digital phase shifter system. This leads to the invention of the distributed phase shifter.

17.5.2 Distributed MEMS Phase Shifters

An alternative phase shifter design uses MEMS tunable capacitors to periodically load a high impedance transmission line [3], thus allowing for phase velocity and, consequently, phase shift to be continuously varied (Figure 17.38). Such a phase shifter design has been well documented by Barker and Rebeiz [27, 33] and is referred to as a distributed MEMS transmission line (DMTL).

Because the electrical length of a line is frequency dependent, the switched-line configuration is usually good for narrowband application. In contrast, periodic loading is intrinsically a broadband solution and has drawn more and more attention.

Tuning range before *pull-in* (the phenomenon in which a membrane cannot be restored after overbiasing) is limited in traditional flat-bridge MEMS capacitor structures that were widely used in distributed MEMS phase shifters. A flat bridge can only be deflected a maximum of approximately 33% of its initial height when force is applied at the center [34], resulting in a maximum 50% capacitance tuning range, or a tuning factor of 1.5:1. This is because at equilibrium the force of the electrostatic attraction and the restoring force generated from the deformed beam are equal in magnitude. Beyond this point, the top membrane will be snapped onto the bottom and cannot be returned to the original position after the biasing voltage is turned off [35].

Several modifications on MEMS switches used in the distributed phase shifter can be made to increase the phase tuning range. These modifications focus on the MEMS

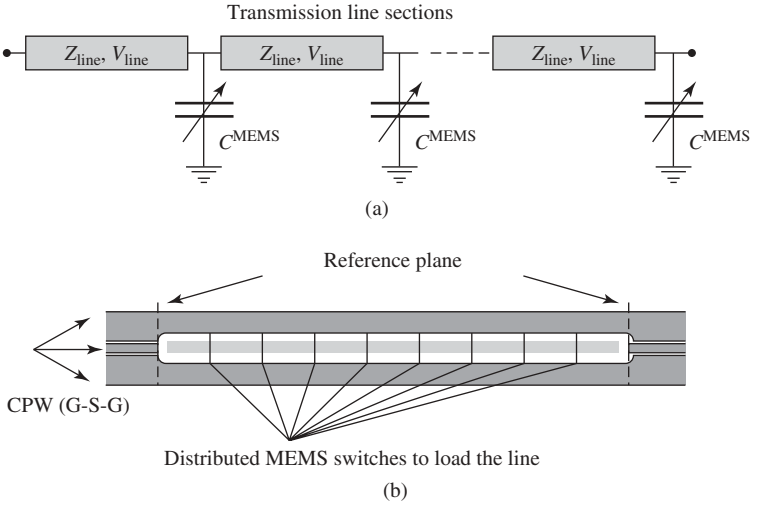


Figure 17.38 (a) Equivalent circuit of periodically loaded transmission line with shunt MEMS switches. (From Ref. 3.) (b) Implementation of periodic loading for a CPW line.

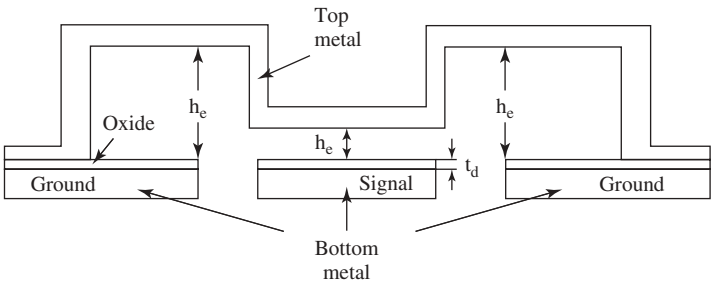


Figure 17.39 Cross section of two-level MEMS switch of increased tuning range for a distributed phase shifter. (From Ref. 34).

switches' architecture and do not increase the physical dimensions of the transmission line. In Refs. 34 and 35, the two-level structure shown in Figure 17.39 was proposed. The capacitance between the ground and the signal is mainly controlled by h_c while the maximum relative deflection is mainly controlled by h_e separately. Since h_e is larger than h_c, the relative tuning range of h_c can be magnified before h_e reaches its mechanical tuning critical point. The phase shifter's performance employing MEMS switches is usually described by the maximal phase shift relative to the insertion loss at a given frequency—$\Delta\phi$/dB. This design leads to a $\Delta\phi$/dB ratio that is increased to 170°/dB at 40 GHz.

Another modification approach is to use a slow-wave cell that has movable beams for the CPW ground [36]. By introducing two ground movable beams, the signal can be routed to operate in a slow-wave mode. The operating mechanism and equivalent circuit models are shown in Figure 17.40. Good impedance matching as well as an increased $\Delta\phi$/dB value can be achieved at the same time. The measured $\Delta\phi$/dB at 12, 50, and 110 GHz are 429°/dB, 358°/dB, and 150°/dB, respectively.

17.5.3 Other Types of MEMS Phase Shifters

The reflect-line phase shifter shown in Figure 17.41a [37] can also be implemented with MEMS switches. In a reflect-line phase shifter, signals travel first out of, then back into, a hybrid coupler. So only $\Delta\phi/2$ electrical length is required for a transmission line to get a $\Delta\phi$ phase delay at the output port. Ideally, this leads to a miniaturized design. However, this type of phase shifter needs an integrated hybrid coupler, which usually occupies additional space and introduces a major source of loss into the module. To solve this problem, couplers can also be fabricated using micromachining technologies [38]. A fully micromachined reflect-line phase shifter is plotted in Figure 17.41b. The hybrid coupler is based on an overlap coplanar waveguide design.

17.6 MICROMACHINING TO IMPROVE ANTENNA PERFORMANCE

In high frequency applications, the challenge of integrating a low cost planar microstrip antenna with other components is to choose the optimum dielectric constant for the substrate [3]. From the circuit's point of view it is always preferred to use materials with high values of dielectric constant. This reduces the overall size, as well as helping to

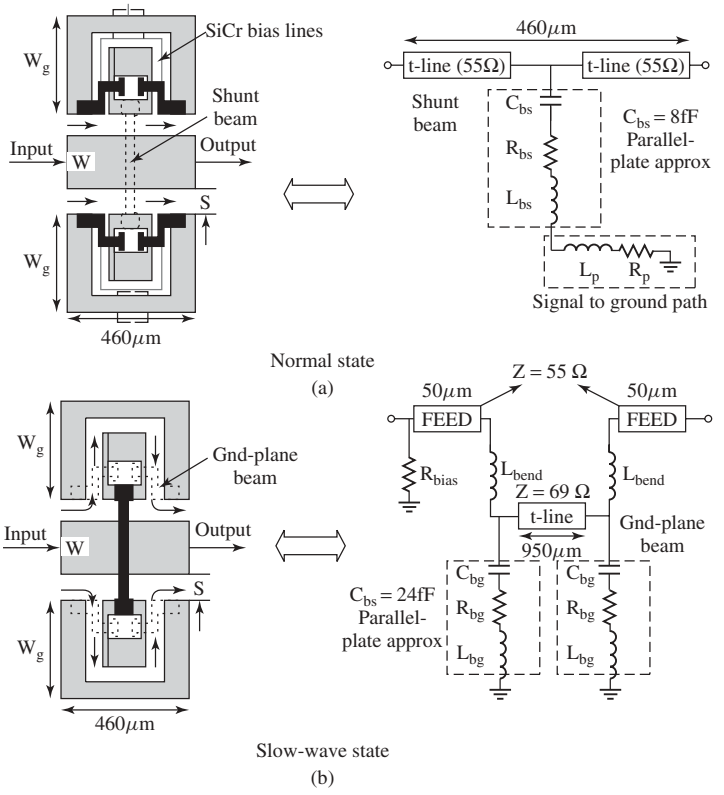


Figure 17.40 (a) Normal mode: ground beams actuated and shunt beam spanning over the signal not actuated; ground current takes a shorter path. (b) Slow-wave mode: ground beams not actuated and shunt beam spanning over the signal line actuated; ground current takes a longer path around the slot. (From Ref. 36).

reduce losses such as those from radiation. On the contrary, if an antenna were to be integrated into a system, one may have to cope with pattern distortion caused by surface waves in electrically thick high-ε_r substrates. One possibility is to use two different substrates—one for circuits and another for antennas. In general, monolithic substrates are pursued to minimize overall cost. Micromachining offers an alternative scheme that satisfies the demands of the antenna and circuit perspectives, using the same substrate. Micromachining technology is ideal for solving this problem. This involves partially removing the substrate underneath the microstrip antenna, or even completely elevating the microstrip antenna into the air by thick-film surface micromachining technology.

As the antenna size becomes smaller in millimeter wave applications, it becomes possible to integrate three-dimensional high performance antennas onto the wafer using modern semiconductor-compatible processes. Horn antennas, vertical monopoles, Yagi–Uda antennas, and helix antennas can be integrated onto the wafer using advanced micromachining technologies such as thick-film surface micromachining and stereolithography [39]. These technologies provide advantages such as high manufacturing precision, compatibility with modern semiconductor technology, and monolithic integration. This section introduces advances in this direction.

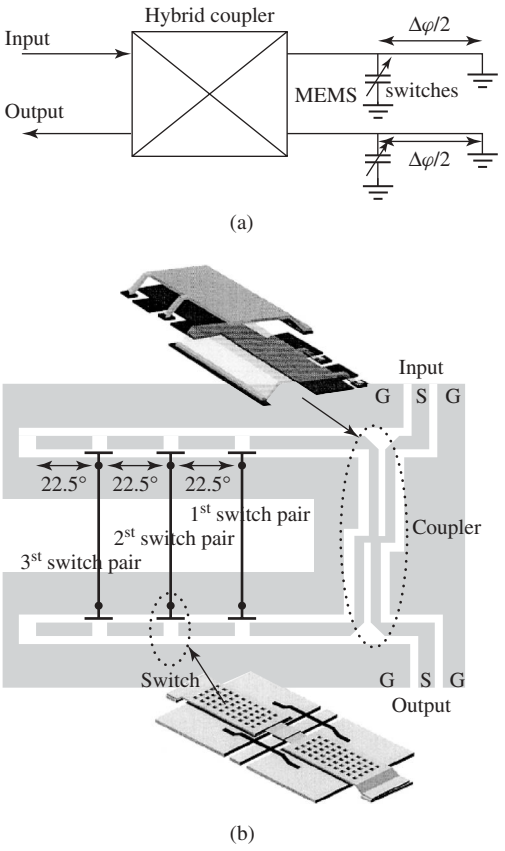


Figure 17.41 (a) Structure layout of a reflect-line phase shifter [37]. (b) Schematic view of the 2-bit reflection-type phase shifter with air-gap overlay CPW coupler and series MEMS switches. (From Ref. 38).

17.6.1 Antennas Fabricated by Si Micromachining

17.6.1.1 Micromachined Patch Antennas The first attempt used a substrate with closely drilled holes [40]. The presence of these air-filled holes primarily reduces the effective dielectric constant of the medium surrounding the antenna. A simpler way to lower the substrate dielectric constant was introduced by Papapolymerou et al. [41]. Shown in Figure 17.42, the silicon substrate underneath the patch antenna is partially removed by wet etching. This reduces the effective dielectric constant of the substrate and results in significant improvements in the antenna bandwidth as well as radiation efficiency. The antenna bandwidth is improved by over 60%, compared with conventional substrate (Figure 17.43). The cavity below the patch is designed to have a resonant frequency close to that of the patch in an effort to improve the bandwidth.

A similar approach can also be used for other antenna designs. In Ref. 42, instead of forming a cavity below the patch, trenches were formed just below the radiating edges so that the conductor of the patch was overhanging to improve the antenna characteristics. A 40% improvement in bandwidth and a marked improvement in radiation efficiency are achieved for an antenna designed at 13.8 GHz by this approach. Extension of the

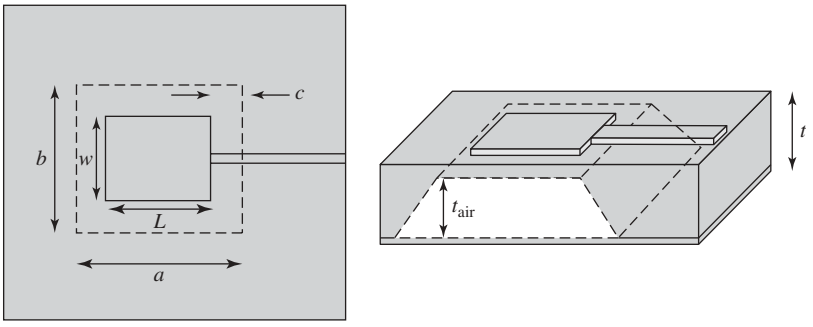


Figure 17.42 Micromachined microstrip antenna with a portion of substrate material below the patch removed by backside etching. (From Ref. 41.)

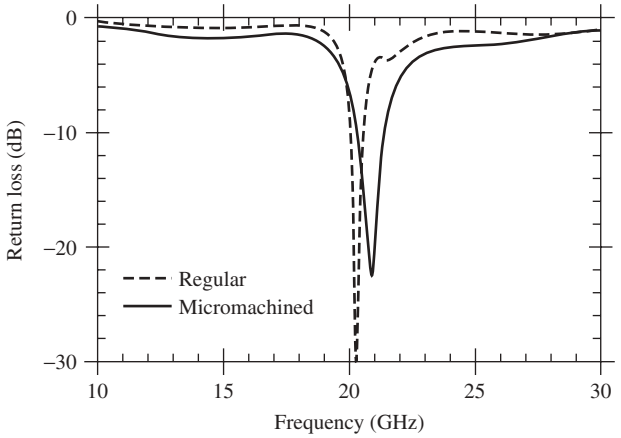


Figure 17.43 Performance improvement of micromachined microstrip antenna compared with one on a regular substrate. (From Ref. 41.)

approach to a microstrip antenna with a coplanar waveguide (CPW) feed for 94 GHz is presented in Ref. 43.

17.6.1.2 Micromachined Horn Antennas Bulk micromachining technology can be used to create horn antennas for high gain applications. Silicon can be wet or dry etched to form the half-cut of a waveguide (cuts along the E-plane or H-plane) and two identical metalized silicon cuts can be bonded together to create a single silicon waveguide. Various micromachined silicon horn antennas have been studied [44–46]. Different cross-section shapes can be created by isotropic/anisotropic etching along different crystal planes of the silicon wafer. These horn antennas have achieved comparable performance to their metallic counterparts. Usually, this type of antenna is operated at very high frequencies, such as terahertz applications, due to the very limited thickness of the silicon wafer. In order to use it at relatively lower frequency applications, such as the W-band (75–110 GHz), several silicon wafers need to be etched through and bonded

together to get the required height. Interested readers can find more information in the literature.

17.6.2 Antenna Fabricated by Thick-Film Surface Micromachining

By using thick-film surface micromachining, it is possible to elevate the microstrip patch antenna into the air. In Ref. 47, a Ka-band elevated patch antenna demonstrates a 10.5% 10-dB fractional bandwidth and 9.5-dBi directivity for an elevation of 600 μm, as shown in Figure 17.44. Although the substrate used in this design has a relatively high dielectric constant of 8.1, the measured radiation pattern still demonstrates a clear main beam in the z-direction without a sidelobe (Figure 17.45). This eliminates the surface wave by elevating the whole structure into the air.

Another design [48] utilizes via-hole electroplating with a thick-film sacrificial layer to create 200-μm high supporting metal posts underneath the middle line of the patch where the value of E_z is zero. A 60-GHz CPW-fed post-supported 2 × 1 patch antenna array using this micromachining technology was also reported, as shown in Figure 17.46. The fabricated antenna shows broadband characteristics such as 10-dB bandwidth of 4.3 GHz from 58.7 to 64.5 GHz in the single patch antenna and 8.7 GHz from 56.3 to 65 GHz in the 2 × 1 patch array. Micromachining techniques have provided the advantages including an increased operating bandwidth and increased antenna gain.

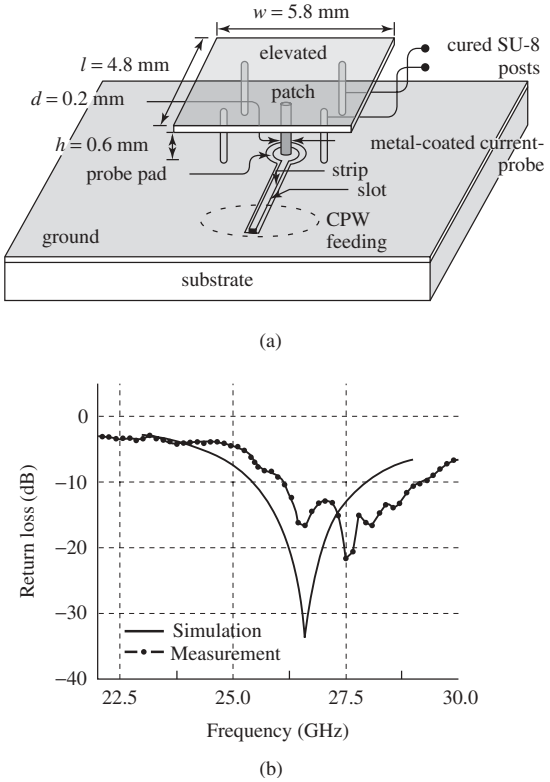


Figure 17.44 (a) Elevated Ka-band patch antenna using thick-film surface micromachining. (b) Simulated and measured return loss of the elevated antenna. (From Ref. 47.)

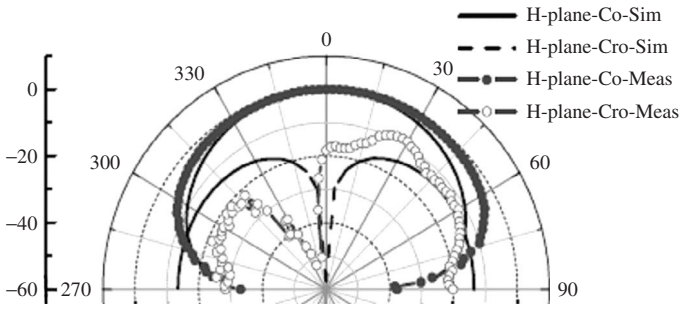


Figure 17.45 Simulated and measured H-plane pattern for the elevated patch. (From Ref. 47.)

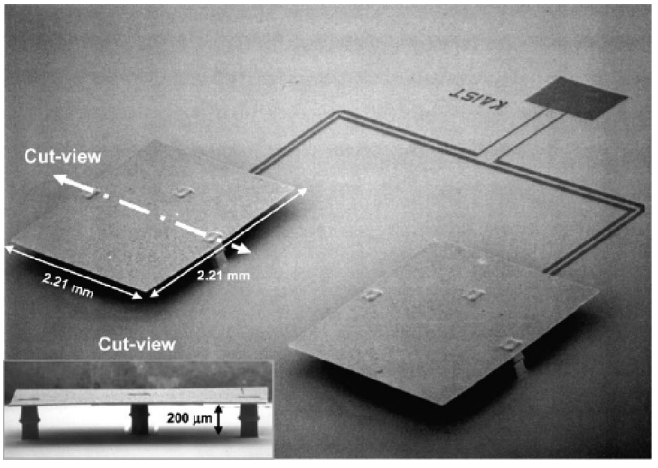
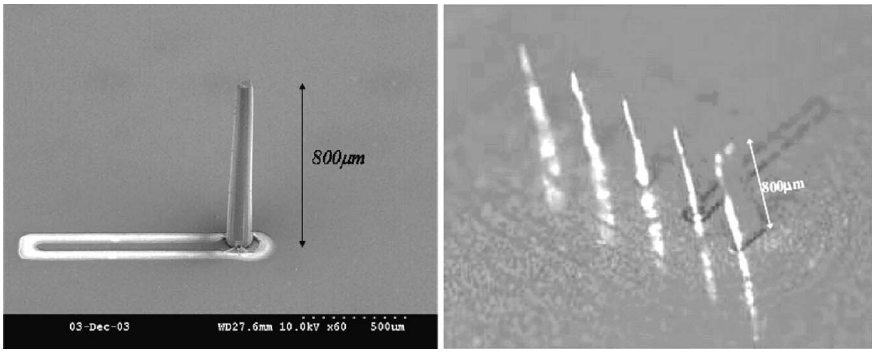


Figure 17.46 SEM picture of a V-band 2×1 micromachined patch array. (From Ref. 48.)

17.6.2.1 W-Band Monopole and Yagi–Uda Array For high frequency applications, such as for the W-band, the operating wavelength is only several millimeters and the physical dimension for a quarter-wavelength monopole is on the order of several hundred micrometers. This is within processing limits of current micromachining technologies. Several simple wire antennas that are widely used in low frequency applications can be fabricated on the wafers with the capability to be integrated with other planar microwave components. Two such examples are shown in Figure 17.47 [49, 50]. These two vertical integrated antennas provide superior performance because of the micromachining technology, including easy integration with planar components and improved impedance bandwidth and radiation efficiency.

17.7 DESIGN CONSIDERATIONS AND DEVELOPMENT FLOW

This section covers the typical development procedure for MEMS-integrated reconfigurable antennas. Compared with other microwave circuit designs, additional attention should be paid to the mechanical/electrical codesign and optimization of biasing schemes.



(a) (b)

Figure 17.47 W-band coplanar-waveguide-fed vertical (a) monopole and (b) Yagi–Uda array. (From Refs. 49 and 50.)

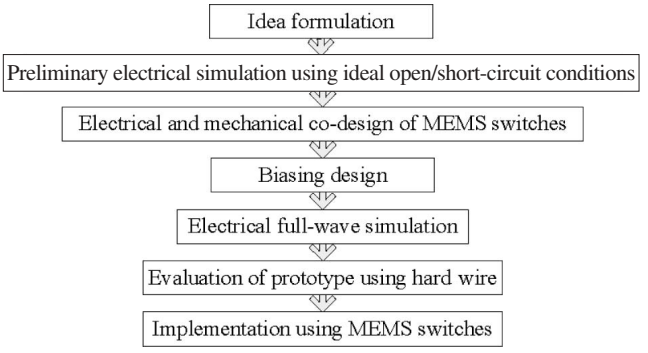


Figure 17.48 Development flow for MEMS-integrated modules.

Figure 17.48 illustrates a common development flow for MEMS-integrated modules. Detailed considerations for each step are explained in the following paragraphs.

17.7.1 Design Idea Formulation and Preliminary Electrical Simulation with the Assumption of Ideal Open/Short-Circuit Conditions

For reconfigurable antennas, MEMS switches are often used to change the length, area, and connections of separate pieces to realize frequency, radiation pattern, and polarization reconfigurations. Readers can form their own design ideas for specific applications. A simple schematic or full-wave simulation should be performed first, using ideal circuit conditions for MEMS switches (i.e., open circuit for "off" state and short-circuit for "on" state).

17.7.2 Electrical/Mechanical Codesign of MEMS Switches

For electrical and mechanical codesign, designers should be familiar with first-order equivalent circuit models for different types of MEMS switches. Reference 1 provides

a good resource for MEMS modeling and also has a comprehensive library of MEMS switches to start with. With knowledge of the equivalent circuit model, the designer will have a general idea of how the physical architecture and dimensions affect MEMS switches' electrical performance. This can narrow the choices and give a more realistic design. Membrane dimensions and beam structures need to be carefully designed and evaluated. The designer should also know how the physical architecture and dimensions affect the mechanical response. For example, the designer should know that the pull-down voltage can be lowered by using low-spring-constant support beam structures; and that adding holes on the beam can reduce the actuation voltage.

At this stage, real MEMS switch structures should be modeled and simulated with full-wave electrical simulators and mechanical simulation tools. There are lots of powerful tools available. For instance, Ansoft HFSS is a good tool for full-wave electrical simulation and COMSOL FEMLAB is famous for mechanical modeling. Modifications are made to satisfy requirements from both sides. The most important parameters that need to be carefully chosen include membrane material, membrane thickness, and membrane height and geometry.

17.7.3 Biasing Design

The goal of a good dc biasing design is to provide high isolation between the RF path and dc path, as well as to minimize the parasitic radiation from dc biasing lines [2]. To prevent RF leakage into the dc path, careful attention needs to be given to dc bias lines themselves. This can be implemented in different ways [7]:

1. By using a quarter-wavelength transmission line connected to a quarter-wavelength open-circuit radial stub. Following this approach, every MEMS switch would require a different dc bias line, and for an antenna integrated with a lot of switches, that would require many lengthy metal lines being added. This would have a pronounced effect on the antenna performance. This solution is not suitable for broadband applications.
2. Another option is to use high resistance lines to provide a wider bandwidth alternative. Aluminum-doped zinc oxide (AZO) is one such example used for biasing. Thin films of this kind are generally deposited using combustion chemical vapor deposition (CCVD), which uses very high temperatures. This is not a problem for materials like silicon, but the temperatures used in CCVD surpass the melting points for most organic substrates and thus these materials cannot be used.

The layout for DC biasing should be planned wisely to save the chip area, as well as reduce the interference to the antenna.

17.7.4 Electrical Full-Wave Simulation

After inserting MEMS switches and dc biasing lines into the circuit, the ideal response might be changed significantly. At this stage, full-wave electrical simulations and optimizations need to be performed to make the necessary adjustments for the initial design. Some ideas may not work and the designer may have to start over from basics.

17.7.5 Hard-Wire Concept-Proofing

For those designs that have passed full-wave verifications, building a hard-wire concept-proofing prototype cannot be skipped. It is very useful to further verify the design idea. Some important factors may be neglected in simulations and will emerge at this stage. Necessary modifications are made for final success.

17.7.6 Implementation Using MEMS Switches and Experiment Verification

At this stage, prototypes with MEMS are built and high quality fabrication yield control is critical to get the desired measurement results.

17.8 SUMMARY

In this chapter, we briefly introduce the exciting advances in MEMS-integrated antennas with reconfigurability, as well as several high performance millimeter antennas created using modern micromachining technologies. There are lots of excellent designs in this area. However, we can only select several examples to represent the current state-of-the-art progress. Interested readers can find more MEMS or micromachined antennas in the literature.

The most important benefit of introducing RF-MEMS into antenna design is to reduce the power consumption of the switch itself. This will help create the power-efficient wireless front-end for next-generation intelligent communication systems. We believe RF-MEMS switches will be used in more and more antenna applications in the near future when more advances are introduced in the MEMS area.

REFERENCES

1. G. M. Rebeiz, *RF MEMS Theory, Design, and Technology*, John Wiley & Sons, Hoboken, NJ, 2003.
2. N. Kingsley, Development of miniature, multilayer, integrated, reconfigurable RF MEMS communication module on liquid crystal polymer (LCP) substrate, Ph.D. Dissertation, Georgia Institute of Technology, Atlanta, GA, May 2007.
3. V. K. Varadan, K. J. Vinoy, and K. A. Jose, *RF MEMS and Their Applications*, John Wiley & Sons, Chichester, UK, 2003.
4. G. Wang, RF MEMS switches with novel materials and micromachining techniques for Soc/SoP RF front ends, Ph.D. Dissertation, Georgia Institute of Technology, Atlanta, GA, December 2006.
5. C. L. Goldsmith, Z. M. Yao, S. Eshelman, and D. Dennston, Performance of low loss RF MEMS capacitive switches, *IEEE Microwave Guided Wave Lett.*, Vol. 8, No. 8, pp. 269–271, 1998.
6. G. M. Rebeiz, G.-L. Tan, and J. S. Hayden, RF MEMS phase shifters: design and applications, *IEEE Microwave Mag.*, Vol. 3, No. 2, pp. 72–81, June 2002.
7. N. Kingsley, D. Anagnostou, M. M. Tentzeris, and J. Papapolymerou, RF MEMS sequentially reconfigurable Sierpinski antenna on a flexible, organic substrate with novel DC-biasing tuning, *IEEE J. MEMS*, Vol. 16, No. 5, pp. 1185–1192, October 2007.
8. D. Anagnostou, G. Zheng, S. Barbin, M. Chryssomallis, J. Papapolymerou, and C. Christodoulou, An X-band reconfigurable planar dipole antenna, in *Proceedings of the*

2005 SBMO/IEEE MTT-S International Conference on Microwave and Optoelectronics, pp. 654–656.

9. S. Liu, M. Lee, C.-W. Jung; G.-P. Li, and F. De Flaviis, A frequency-reconfigurable circularly polarized patch antenna by integrating MEMS switches, in *Proceedings of the 2005 IEEE Antennas and Propagation Society International Symposium*, Vol. 2A, pp. 413–416.
10. B. A. Cetiner, Q. Xu, and L. Jofre, Frequency reconfigurable annular slot antenna, in *Proceedings of the IEEE Antenna and Propagation Symposium*, June 2007, pp. 5845–5848.
11. D. Anagnostou, G. Zheng, M. Chryssomallis, J. C. Lyke, G. E. Ponchak, J. Papapolymerou, and C. Christodoulou, Design, fabrication, and measurements of an RF-MEMS-based self-similar reconfigurable antenna, *IEEE Trans. Antennas Propag.*, Vol. 54, No. 2, Part 1, pp. 422–432, February 2007.
12. S. Yang, H. K. Pan, A. E. Fathy, S. El-Ghazaly, and V. K. Nair, A novel reconfigurable mini-maze antenna for multi-service wireless universal receiver using RF MEMS, in *Digest of 2006 IEEE MTT-S International Microwave Symposium*, pp. 182–185.
13. B. A. Cetiner, J. Y. Qian, H. P. Chang, M. Bachman, G. P. Li, and F. De Flaviis, Monolithic integration of RF MEMS switches with a diversity antenna on PCB substrate, *IEEE Trans. Microwave Theory Tech.*, Vol. 51, No. 1, Part 2, pp. 332–335, 2003.
14. B. A. Cetiner, L. Jofre, J. Y. Qian, S. Liu, G. P. Li, and F. De Flaviis, A compact broadband MEMS-integrated diversity system, *IEEE Trans. Vehicular Technol.*, Vol. 56, No. 2, pp. 436–444, March 2007.
15. L. Petit, L. Dussopt, and J. M. Laheurte, MEMS-switched parasitic-antenna array for radiation pattern diversity, *IEEE Trans. Antennas Propag.*, Vol. 54, No. 9, pp. 2624–2631, September 2006.
16. G. H. Huff and J. T. Bernhard, Integration of packaged RF MEMS switches with radiation pattern reconfigurable square spiral microstrip antennas, *IEEE Trans. Antennas Propag.*, Vol. 54, No. 2, Pt. 1, pp. 464–469, February 2006.
17. http://www.radantmems.com/radantmems/index.html.
18. C. Jung, M. Lee, G. P. Li, and F. De Flaviis, Reconfigurable scan-beam single-arm spiral antenna integrated with RF-MEMS switches, *IEEE Trans. Antennas Propag.*, Vol. 54, No. 2, Pt. 1, pp. 455–463, February 2006.
19. R. N. Simons, D.-H. Chun, and L. P. B. Katehi, Polarization reconfigurable patch antenna using microelectromechanical systems (MEMS) actuators, in *Proceedings of the 2002 IEEE International Symposium of Antennas and Propagation*, Vol. 2, pp. 6–9.
20. B. A. Cetiner, E. Akay, E. Sengul, and E. Ayanoglu, A MIMO system with multifunctional reconfigurable antennas, *IEEE Antennas Wireless Propag. Lett.*, Vol. 5, No. 1, pp. 463–466, December 2006.
21. J. C. Chiao, Y. Fu, I. M. Chio, M. DeLisio, and L.-Y. Lin, MEMS reconfigurable Vee antenna, in *Digest of 1999 IEEE MTT Symposium*, pp. 1515–1518.
22. C.-W. Baek, S.-H. Song, J.-H. Park, S.-H. Lee, J.-M. Kim, W. Choi, C. Cheon, Y.-K. Kim, and Y.-W. Kwon, A V-band micromachined 2-D beam-steering antenna driven by magnetic force with polymer-based hinges, *IEEE Trans. Microwave Theory Tech.*, Vol. 51, No. 1, Pt. 2, pp. 325–331, January 2003.
23. P. Blondy, D. Bouyge, A. Crunteanu, and A. Pothier, A wide tuning range MEMS switched patch antenna, in *Digest of 2006 IEEE MTT Symposium*, pp. 152–154.
24. C. Balanis, *Antenna Theory: Analysis and Design*, 2nd ed., John Wiley & Sons, Hoboken, NJ, 1997.
25. G.-L. Tan, R. E. Mihailovich, J. B. Hacker, J. F. DeNatale, and G. M. Rebeiz, Low-loss 2- and 4-bit TTD MEMS phase shifters based on SP4T switches, *IEEE Trans. Microwave Theory Tech.*, Vol. 51, No. 1, Pt. 2, pp. 297–304, January 2003.

26. J. B. Hacker, R. E. Mihailovich, M. Kim, and J. F. DeNatale, A Ka-band 3-bit RF MEMS true-time-delay network, *IEEE Trans. Microwave Theory Tech.*, Vol. 51, No. 1, pp. 305–308, Pt. 2, January 2003
27. N. S. Barker and G. M. Rebeiz, Optimization of distributed MEMS transmission-line phase shifters—U-band and W-band designs, *IEEE Trans. Microwave Theory Tech.*, Vol. 48, No. 11, Pt. 1, pp. 1957–1966, November 2000.
28. J. S. Hayden, and G. M. Rebeiz, Very low-loss distributed X-band and Ka-band MEMS phase shifters using metal–air–metal capacitors, *IEEE Trans. Microwave Theory Tech.*, Vol. 51, No. 1, Pt. 2, pp. 309–314, January 2003.
29. H.-T. Kim, J.-H. Park, S. Lee, S. Kim, J.-M. Kim, Y.-K. Kim, and Y. Kwon, V-band 2-b and 4-b low-loss and low-voltage distributed MEMS digital phase shifter using metal–air–metal capacitors, *IEEE Trans. Microwave Theory Tech.*, Vol. 50, No. 12, pp. 2918–2923, December 2002.
30. J.-J. Hung, L. Dussopt, and G. M. Rebeiz, Distributed 2- and 3-bit W-band MEMS phase shifters on glass substrates, *IEEE Trans. Microwave Theory Tech.*, Vol. 52, No. 2, pp. 600–606, February 2004.
31. M. Kim, J. B. Hacker, R. E. Mihailovich, and J. F. DeNatale, A DC-to-40 GHz four-bit RF MEMS true-time delay network, *IEEE Microwave Wireless Components Lett.*, Vol. 11, No. 2, pp. 56–58, February 2001. (See also *IEEE Microwave and Guided Wave Letters*.)
32. B. Pillans, S. Eshelman, A. Malczewski, J. Ehmke, and C. Goldsmith, Ka-band RF MEMS phase shifters, *IEEE Microwave Guided Wave Lett.*, Vol. 9, No. 12, pp. 520–522, December 1999. (See also *IEEE Microwave and Wireless Components Letters*.)
33. S. Barker and G. M. Rebeiz, Distributed MEMS true-time delay phase shifters and wide-band switches, *IEEE Trans. Microwave Theory Tech.*, Vol. 46, No. 11, Pt. 2, pp. 1881–1890, November 1998.
34. J. Zou, C. Liu, J. Schutt-Aine, J. Chen, and S.-M. Kang, Development of a wide tuning range MEMS tunable capacitor for wireless communication systems, in *Technical Digest of 2000 IEDM,* December 2000, pp. 403–406.
35. G. McFeetors and M. Okoniewski, Distributed MEMS analog phase shifter with enhanced tuning, *IEEE Microwave Wireless Components Lett.*, Vol. 16, No. 1, pp. 34–36, January 2006.
36. B. Lakshminarayanan and T. M. Weller, Design and modeling of 4-bit slow-wave MEMS phase shifters, *IEEE Trans. Microwave Theory Tech.*, Vol. 54, No. 1, pp. 120–127, January 2006.
37. D. Pozar, *Microwave Engineering*, 3rd ed., John Wiley & Sons, Hoboken, NJ, 2004.
38. J.-H. Park, H.-T. Kim, W. Choi, Y. Kwon, and Y.-K. Kim, V-band reflection-type phase shifters using micromachined CPW coupler and RF switches, *J. Microelectromechanical Systems*, Vol. 11, No. 6, pp. 808–814, December 2002.
39. X. Gong, B. Liu, L. P. B. Katehi, and W. J. Chappell, Laser-based polymer stereolithography of vertically integrated narrow bandpass filters operating in K band, in *Proceedings of the IEEE MTT-S International Microwave Symposium Digest*, June 2004, pp. 425–428.
40. G. P. Gauthier, A. Courtay, and G. M. Rebeiz, Microstrip antennas on synthesized low dielectric constant substrates, *IEEE Trans. Antennas Propag.*, Vol. 4, pp. 1310–1314, 1997.
41. J. Papapolymerou, R. F. Drayton, and L. P. B. Katehi, Micromachined patch antennas, *IEEE Trans. Antennas Propag.*, Vol. 46, No. 2, pp. 275–283, February 1998.
42. Q. Chen, V. F. Fusco, M. Zheng, and P. S. Hall, Micromachined silicon antennas, in *Proceedings of the 1998 International Conference on Microwave and Millimeter Wave Technology*, pp. 289–292.
43. G. P. Gauthier, J. P. Raskin, L. P. B. Katehiand, and G. M. Rebeiz, A 94-GHz aperture-coupled micromachined microstrip antenna, *IEEE Trans. Antennas Propag.*, Vol. 47, No. 12, pp. 1761–1766, December 1999.

44. J. W. Digby, C. E. McIntosh, G. M. Parkhurst, B. M. Towlson, S. Hadjiloucas, J. W. Bowen, J. M. Chamberlain, R. D. Pollard, R. E. Miles, D. P. Steenson, L. S. Karatzas, N. J. Cronin, and S. R. Davies, Fabrication and characterization of micromachined rectangular waveguide components for use at millimeter-wave and terahertz frequencies, *IEEE Trans. Microwave Theory Tech.*, Vol. 48, No. 8, pp. 1293–1302, August 2000.

45. B. A. Shenouda, L. W. Pearson, and J. E. Harriss, Etched-silicon micromachined W-band waveguides and horn antennas, *IEEE Trans. Microwave Theory Tech.*, Vol. 49, No. 4, Pt. 1, pp. 724–727, April 2001.

46. V. Douvalis, Y. Hao, and C. G. Parini, A monolithic active conical horn antenna array for millimeter and submillimeter wave applications, *IEEE Trans. Antennas Propag.*, Vol. 54, No. 5, pp. 1393–1398, May 2006.

47. B. Pan, Y.-K. Yoon, G. E. Ponchak, M. G. Allen, J. Papapolymerou, and M. M. Tentzeris, Analysis and characterization of a high-performance Ka-band surface micromachined elevated patch antenna, *Antennas Wireless Propag. Lett.*, Vol. 5, No. 1, pp. 511–514, December 2006.

48. J.-G. Kim, H.-S. Lee, H.-S. Lee, J.-B. Yoon, and S. Hong, 60-GHz CPW-fed post-supported patch antenna using micromachining technology, *IEEE Microwave Wireless Components Lett.*, Vol. 15, No. 10, pp. 635–637, October 2005.

49. B. Pan, Y. Yoon, P. Kirby, J. Papapolymerou, M. M. Tentzeris, and M. Allen, A W-band surface micromachined monopole for low-cost wireless communication systems, in *Proceedings of the 2004 IEEE-MTT Symposium*, pp. 1935–1938.

50. Y.-K. Yoon, B. Pan, J. Papapolymerou, M. M. Tentzeris, and M. G. Allen, Surface-micromachined millimeter-wave antennas, in *Proceedings of the 13th International Conference on Solid-State Sensors, Actuators and Microsystems (Transducers 2005)*, June 2005, Vol. 2, pp. 2986–2989.

CHAPTER 18

Feed Antennas

TREVOR S. BIRD

18.1 INTRODUCTION

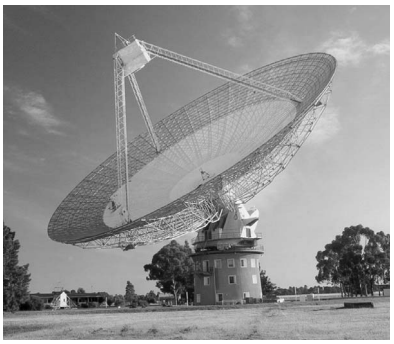
A "feed antenna" is a means of supplying energy to (or receiving energy from) a "secondary antenna," such as a reflector, lens, or beam waveguide, via a transmission line or waveguide. Therefore a feed antenna is often referred to as a "primary feed." Everyday applications of antennas with feeds include satellite communications, radar, radio telescopes, deep space probes, and terrestrial microwave and millimeter-wave radio links. Some examples are shown in Figures 18.1 and 18.2. Depending on the application, secondary antennas can be designed to produce beams that are highly directional to minimize the effects of interference and noise or can be shaped to cover a specific region or angular segment. For example, shaped beams allow reuse of the frequency spectrum in another region as in the case of satellite coverage on the Earth's surface. In other applications, such as broadcasting, secondary antennas and feeds can be used to produce beams tailored to suit the demographics around the transmitter. The secondary antenna can be used to illuminate another device such as in a beam waveguide system for a reflector or a particle accelerator.

Feed antennas come in many different shapes and sizes, ranging from simple dipoles and waveguides to horns, printed antennas such as microstrip patches, arrays, and traveling-wave antennas. The choice of feed in any situation is often a complex decision and depends on the type of application and secondary antenna. Factors to be considered when selecting a feed antenna include the frequency of operation, bandwidth, polarization requirements, impedance properties, the transmission line connecting to the feed, mechanical aspects such as the size, volume, and weight, available materials, and cost.

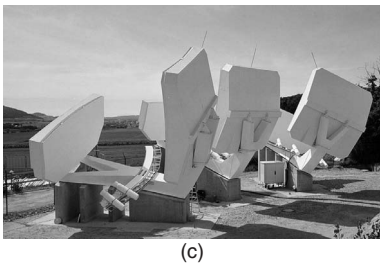
Secondary antennas are typically capable of extremely wide bandwidth performance and are, in practice, severely restricted by the bandwidth limitations of the feed antenna and its associated network, the feed system. Over the past few years, there have been significant advances in improving the performance of feed systems, especially in regard to wideband and multiband operation.

The most common type of secondary antenna is the parabolic reflector (Figure 18.2a). This has a point focus and, when illuminated by a feed emitting a spherical wave, it produces a narrow pencil beam. Dual reflectors (Figure 18.1b) are frequently used

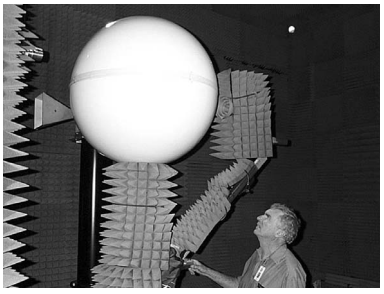
Modern Antenna Handbook. Edited by Constantine A. Balanis



(a)

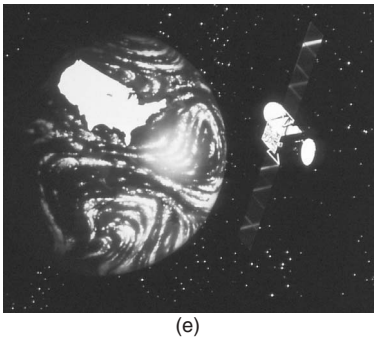


(c)

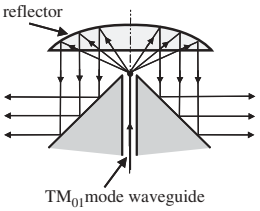


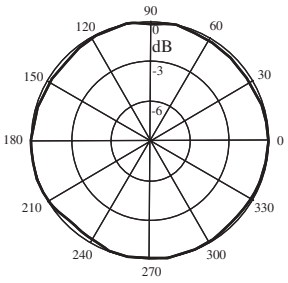
(d)

(b)



(e)





(f)

Figure 18.1 Some applications of feed antennas: (a) 64-m radio telescope at Parkes, Australia; (b) dual-shaped Cassegrain reflector Earth station; (c) three offset-fed reflector 4.5-m multibeam Earth stations at SES-ASTRA, Luxembourg, give over 110° coverage of the geostationary satellite arc; (d) 0.9-m spherical Luneburg lens; (e) shaped-beam satellite antenna; and (f) omnidirectional reflector antenna.

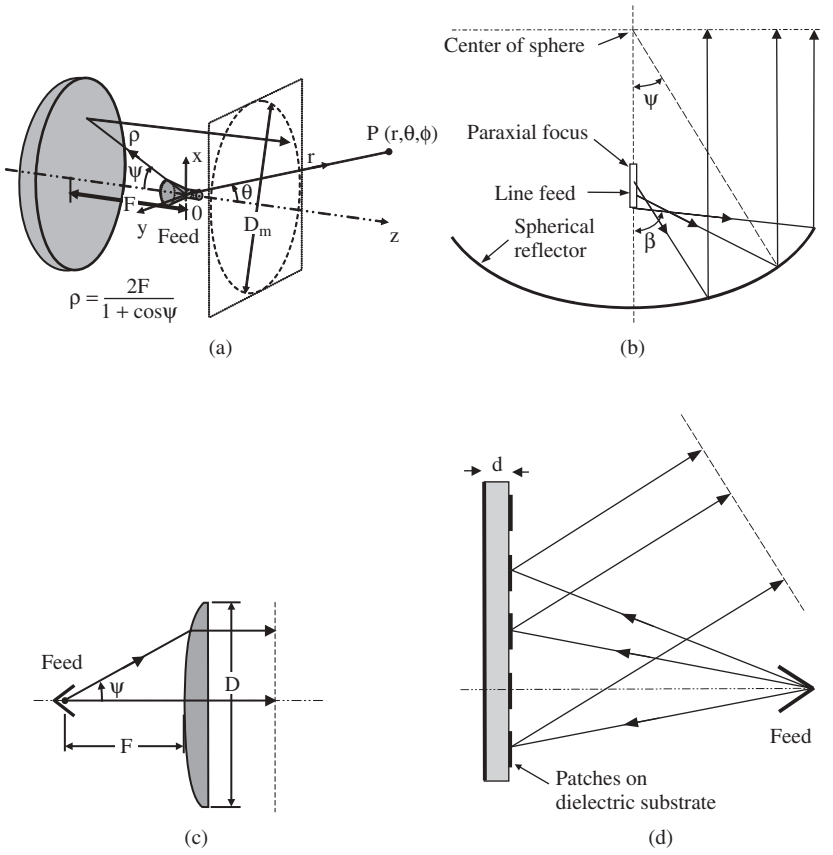


Figure 18.2 Examples of secondary antennas: (a) parabolic reflector, (b) spherical reflector, (c) Two-dimensional lens, and (d) reflectarray.

in professional applications. The reduced blockage compared with a single reflector, the fact that the electronics can be placed close to the feed without long cables, the better noise performance, and the narrow angle of power radiated from the feed all give much greater flexibility to the feed designer. Other secondary antennas are cylindrical or spherical reflectors (Figure 18.2b), which require a line source feed to produce a pencil beam, lenses (Figures 18.1d and 18.2c), and reflectarrays (Figure 18.2d). Reflectarrays combine the techniques of reflectors and printed circuit antennas. A standard reflectarray consists of an array of microstrip patches or dipoles printed on a thin dielectric substrate. The elements of the printed array are suitably phased to convert a spherical wave from a conventional feed into a plane wave. The advantage is that the antenna is planar, has low volume, and is potentially lightweight. Phasing of the printed array is achieved with patches that have different sizes, different length delay lines, or varying angles of rotation.

In some applications, the aim is to shape the secondary radiation pattern and the role of the antenna feed is simply to give the correct illumination. For example, an array feed or a corrugated horn can be used to illuminate a shaped reflector system for a satellite

system that will produce a shaped secondary pattern to cover selected Earth coverage regions (see Figure 18.1e). Another application is a mobile base station antenna that requires an omnidirectional pattern. A suitable feed antenna and a shaped reflector are able to produce a 360° coverage as shown in Figure 18.1f. Secondary antennas with elliptical beam shapes can be used to maximize the radiated field intensity within an angular region, which has different beamwidths in two orthogonal planes. Such antennas are used in radar, Earth stations, and satellite applications to provide an elliptical footprint on the Earth.

18.1.1 Historical Perspective

The first feed antennas date back to the pioneering work of Hertz during his demonstration of the consequences of James Clerk Maxwell's equations, which were published a decade earlier. In 1888, Hertz employed a half-wavelength dipole in front of a cylindrical parabolic reflector to demonstrate the optics-like properties of electromagnetic waves at about 60 cm [1]. Other feeds and antennas now used as basic feeds were discovered prior to 1900 by Lodge (circular waveguide), Bose (dielectric rod and pyramidal horn), and Rayleigh [2]. From the early 1900s to the 1930s, little work on feeds appears to have occurred, mainly due to the outstanding success of long range low frequency communications, where directly radiating antennas were the natural choice.

The rising interest in microwave techniques in the 1930s resulted in major developments in feed antennas. The then new applications using ultrahigh frequencies and microwaves, such as radar and point-to-point communications, required many new antennas, and reflectors, in particular, became extremely important because of their versatility, effectiveness, and practicability. In short, feeds became an area of significant interest and research. The MIT Radiation Laboratory series of books published in the late 1940s, and especially the volume edited by Silver [3], did much to establish the basis of reflector and feed antenna design. It is significant that Silver's book was the primary reference for a generation of feed designers and is still a great source of information. With radar, and the deployment of terrestrial microwave systems for communications, reflectors came into common use. These used simple waveguides or horns, rear-radiating feeds (e.g., Cutler [4]), or arrays such as the Yagi–Uda or log-periodic arrays [5]. The creation of new applications in the 1950s, such as radio astronomy and satellite communications, added new dimensions to feed design with requirements for low noise and low cross polarization. These demands resulted in new feed antennas, including the dual-mode Potter horn [6], the corrugated horn (see Thomas [7] for an historical summary of this important feed), multimode horns [8], stepped horns, and dielectric-lined feeds. Many of the most influential papers on horns written prior to 1976 have been collected by Love [9] and a more recent survey is given by Olver et al. [10]. Array feeds were investigated first in the early 1960s [11] and over the next two decades were adopted in radar and satellite communications for producing scanned and shaped beams with reflectors. In the 1980s and early 1990s, compact corrugated horns covering octave bandwidths were developed, while easy to manufacture, high performance feeds, such as axially corrugated horns, were widely adopted. Major advances in electromagnetic modeling have allowed the development of horns whose performance can be tailored by computer optimization. In recent years, improvements in the theoretical understanding of printed antennas has resulted in microstrip and printed slot or Vivaldi arrays being used as feeds.

18.1.2 Design of Feeds

A feed antenna should be designed to suit the secondary antenna to obtain the best overall performance. This is usually expressed in terms of the signal level, whether received or transmitted, immunity to noise or interference, and frequency of operation. There are essentially two ways of designing the overall antenna. The first way, called the transmit mode of operation, specifies the combined radiation pattern of the feed and the secondary antenna that meets the system requirements when both are transmitting signals. Figure 18.2a depicts a parabolic† reflector operating in the transmit mode. The electromagnetic field radiated by the feed is called the primary field and the radiation pattern the primary pattern. The illuminated antenna is the secondary antenna and the associated radiated field is the secondary field.

In the second approach, called the receive mode of operation, the antenna feed is designed to match the fields produced in the secondary antenna due to a specific illumination; for example, plane wave illumination is used for a directive beam. In most common situations the two approaches are equivalent (due to the principle of reciprocity [5]) and are often used interchangeably. If the fields are matched in the focal region of a reflector or lens, this is called focal-plane matching. Optimum antenna gain and minimum cross polarization are achieved simultaneously if the feed aperture field (represented by the complex vector $\mathbf{E}_a$) is conjugate matched to the focal field of the reflector or lens ($\mathbf{E}_f$). This is expressed approximately by

$$\mathbf{E}_{a,\mathrm{cp}} \approx \mathbf{E}^*_{f,\mathrm{cp}}, \qquad \mathbf{E}_{a,\mathrm{xp}} \approx -\mathbf{E}^*_{f,\mathrm{xp}} \tag{18.1}$$

where cp and xp are the copolar and cross-polar components of the fields and the asterisk indicates the complex conjugate. This approach is applicable to single feeds or arrays in the focal region.

The design of the feed antenna's primary or aperture field depends on the geometry of the secondary antenna. For example, the geometry of a symmetrical parabolic reflector is defined simply by its diameter D_m and its focal length F (see Figure 18.2a). However, in the symmetrical case there is usually some blockage by the feed and the supporting struts close to where the energy is strongest. For a given F/D_m ratio, the half-subtended angle ψ_c at the feed is given by

$$\psi_c = 2 \arctan\left(\frac{1}{4F/D_m}\right) \tag{18.2}$$

In typical reflectors, F/D_m has a value from 0.3 to 0.6, which corresponds to ψ_c ranging from about 80° down to 45°. The offset parabolic reflector geometry (Figure 18.3) is slightly more complex, but the advantage is the ability to avoid blockage by the feed or to shift this blockage toward a region of the aperture where the field is weak. In an offset parabolic reflector the feed is rotated through an angle ψ_o relative to the axis of the parent paraboloid. Three parameters instead of two now determine the cone angle of

†This refers to a paraboloid or part of a paraboloidal surface as for an offset-fed parabolic reflector.

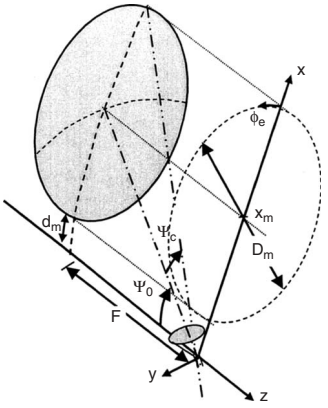


Figure 18.3 Geometry of offset parabolic reflector.

the feed. The half-subtended angle in this case is

$$\psi_c = 2\arctan\left[\frac{1}{2\sin^2(\psi_o/2)}\left(\sqrt{\left(\frac{4F}{D_m}\right)^2 + \sin^2\psi_o} - \frac{4F}{D_m}\right)\right]$$

$$\cong 2\arctan\left[\frac{\cos^2(\psi_o/2)}{4F/D_m}\right], \quad F/D_m > 0.8 \tag{18.3}$$

The largest cone angle at any given F/D_m is for the symmetrical case.

Particular applications require specific feed illuminations and the choice of illumination for the secondary antenna is a compromise between gain, sidelobes, beamwidth, and spillover, which are determined by the characteristics of the primary feed antenna. We return to this compromise at the end of the next section.

18.1.3 Characteristics of Primary Feed Antennas

It is usual to describe the performance of an antenna in terms of its far-field radiation characteristics and its terminal impedance or reflection coefficient into a specified transmission line. Some important terms used to describe feed antennas are:

1. Radiation pattern
2. Gain, directivity, and efficiency
3. Phase center
4. Polarization
5. Input match
6. Bandwidth
7. Spillover
8. Edge illumination and edge taper

These terms are defined next.

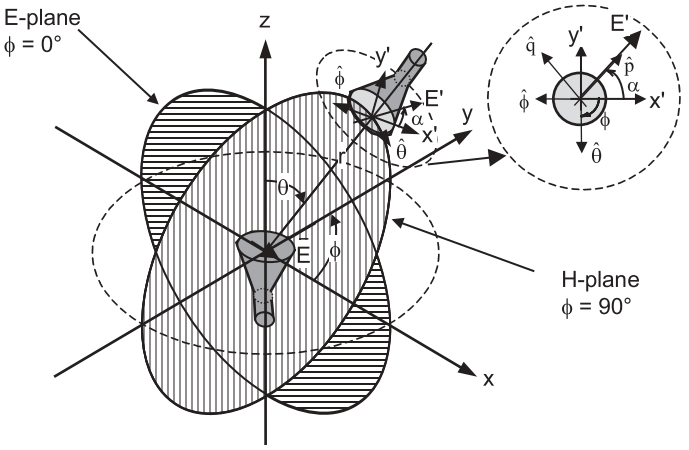


Figure 18.4 Geometry for radiated fields from feeds.

18.1.3.1 Radiation Pattern In the far field of an antenna, the amplitude of the electromagnetic fields radiated is proportional to $1/r$, where r is the distance from the antenna (see Figure 18.4). Plots of the magnitude of the radiated power at a constant distance r are called radiation power patterns. The power pattern, P, in a given direction (θ, ϕ) is

$$P(\theta, \phi) = 4\pi r^2 P_r(r, \theta, \phi) \tag{18.4}$$

where P is the power density per unit solid angle, P_r is the power density (in W/m^2), and r, θ, and ϕ are the spherical polar coordinates (see Figure 18.4). The power density is the radial component of the Poynting vector given by

$$P_r(r, \theta, \phi) = \frac{1}{2\eta}|\mathbf{E}|^2 \tag{18.5}$$

where $\mathbf{E}$ is the electric field intensity (V/m) transverse to the direction of wave propagation. The quantity $\eta = \sqrt{\mu_0/\varepsilon_0}$ is the wave impedance in free space (μ_0 and ε_0 are the permeability and permittivity of free space) and is approximately 120π ohms.

The power pattern of a feed usually consists of a maximum beam in the main direction and a series of subsidiary maxima called sidelobes. These are numbered consecutively from the main beam. Usually the power pattern (Eq. (18.4)) is normalized to the maximum value and plots are made of this function as contours or intensity plots in two dimensions, or one-dimensional pattern cuts. The magnitude of the pattern is usually expressed in decibels (dB). The most common type of pattern cut for feeds is a plot of P at a constant angle ϕ, the so-called azimuth cut.

The half-power beamwidth (HPBW) is the angle subtended at the half-power or -3-dB points of the normalized power pattern. The beamwidth between first nulls (BWFN) and the -12-dB power beamwidth are also used.

18.1.3.2 Gain, Directivity, and Efficiency The absolute power gain [5] of an antenna in a given direction (θ, ϕ) is defined as the ratio of the power intercepted by a large sphere enclosing the antenna when that same power density at (θ, ϕ) is radiated in

all directions (i.e., isotropically) and the total power input to the antenna†. The power density at (θ, ϕ) is given by Eq. (18.5). If this power were radiated isotropically, the power that would be radiated on a large sphere of radius r is given by Eq. (18.4). Therefore, if P_{in} is the power incident on the antenna, the absolute power gain function, G, is

$$G(\theta, \phi) = \frac{P(\theta, \phi)}{P_{\text{in}}} \tag{18.6}$$

The maximum value of Eq. (18.6) is the maximum power gain and is often used to describe the performance of an antenna. Usually when the term "gain" is used, maximum power gain is being referred to.

Directivity is defined similarly to gain except that the reference power is now in the radiated field. Thus the directivity function, D, is defined as

$$D(\theta, \phi) = \frac{P(\theta, \phi)}{P_T} \tag{18.7}$$

where P_T is the total radiated power and is found by integrating the power density, P_r, over a surface enclosing the antenna. For aperture antennas, it is common to integrate the power density over the aperture, rather than the far-field sphere, as this is easier to do. A special case of Eq. (18.7) is where the antenna is excited by sources with uniform amplitude and phase distribution (or illumination) over its aperture area A. The maximum directivity in this case is $D_0 = 4\pi A/\lambda^2$.

The aperture efficiency, η_a, is the ratio of the maximum directivity, $D_{\max}$, given by Eq. (18.7), and the directivity of a uniformly illuminated aperture

$$\eta_a = \frac{D_{\max}}{D_0} \tag{18.8}$$

Most practical feeds have aperture efficiencies in excess of 60%, and close to 100% can be achieved over limited bandwidths.

Gain and directivity are related through the antenna efficiency, which is the ratio of gain and directivity or equivalently the ratio of total radiated power and the power incident at the antenna input. Antenna efficiency (η_{ae}) is equal to the product of the mismatch efficiency (η_m), conversion efficiency (η_c), and the aperture efficiency η_a. In a lossless, matched feed, the antenna efficiency equals the aperture efficiency.

18.1.3.3 Phase Center The phase center is the apparent origin of the spherical wave radiating from an antenna. Because phase varies significantly over the full sphere, this center is not unique for all applications because it depends on the solid angle over which the radiated field is incident on the secondary antenna. The phase center is obtained experimentally by rotating the antenna about its axis and determining the position on the axis where the measured phase function variation is minimized over the desired angle for illuminating the secondary antenna. This can be time consuming as the phase center is usually different for each pattern cut. An alternative approach is to calculate the phase

†This compares with the IEEE definition (Std. 145—1993 §2.165) for gain, which uses the power *accepted* by the antenna. The present definition is more convenient for feeds and differs from the IEEE definition by the factor $1/(1 - |\Gamma_{\text{in}}|^2)$ for single-mode excitation, where Γ_{in} is the input reflection coefficient.

center from measured or numerically obtained data. One approach is to average the phase center from locations d_n of N pattern cuts. Thus the phase center, d_{pc}, is given by

$$d_{\text{pc}} = \frac{1}{N}\sum_{n=1}^{N} d_n \tag{18.9}$$

The phase center can be estimated by the method described in Ref. 12 and in the nth azimuth plane it is given by

$$kd_n = \frac{a_0 b_1 - a_1 b_0}{a_0 a_2 - a_1^2} \tag{18.10}$$

where $k = 2\pi/\lambda$, $a_0 = \sum_{i=1}^{M} \theta_i$, M is the number of data points of the phase function $\Phi(\theta_i, \phi_n), \theta_i$ is the ith angle of the nth pattern cut ϕ_n, $a_1 = \sum_{i=1}^{M} \cos\theta_i$, $a_2 = \sum_{i=1}^{M} \cos^2\theta_i$, $b_0 = \sum_{i=1}^{M} \Phi(\theta_i, \phi_n)$, and $b_1 = \sum_{i=1}^{M} \Phi(\theta_i, \phi_n)\cos\theta_i$. The simplest case of Eq. (18.9) corresponds to a single cut ($N = 1$). A better estimate is to average the phase centers of the two principal plane patterns ($N = 2$). With circularly symmetrical feeds, an average of the two principal planes plus the intercardinal plane (45° and 135°) phase centers ($N = 4$) is usually sufficiently accurate for most purposes.

18.1.3.4 Polarization Although the fields radiated by feed antennas are usually complicated, they can be resolved into two components on any chosen surface, such as the far-field sphere (see Figure 18.4) or on the secondary antenna. Without loss of generality, we consider here only the former. The two components are the desired polarization or the copolarized component (denoted by the vector $\hat{\mathbf{p}}$) and the usually undesired cross-polarized component (given by $\hat{\mathbf{q}}$). Furthermore, there are two main types of polarization—linear and circular. There is no unique way of defining the components of these polarizations. For feeds, the common method is to define copolarization and cross polarization in the manner they are measured in practice. In the case of predominantly linearly polarized feeds, the copolar component is obtained as for conventional far-field measurement, by aligning the test antenna on boresight with the polarization of the distant source antenna. Maintaining this alignment, the test antenna is then rotated about the phase center. The signal received at the test antenna is the copolar pattern. If the polarization of the distant antenna is now rotated through ±90° and the radiation pattern is remeasured, the received signal is the cross-polar pattern of the test antenna. If the test antenna is rotated in the plane of the principal electric field component, the pattern cut is called the E-plane (see Figure 18.4). The orthogonal plane is called the H-plane cut. For feed antennas it is common also to measure the pattern in the 45° and 135° planes. The copolarized and cross-polarized field components are obtained from the spherical field components by means of

$$\begin{bmatrix} E_p \\ E_q \end{bmatrix} = \begin{bmatrix} \cos(\phi - \alpha) & -\sin(\phi - \alpha) \\ -\sin(\phi - \alpha) & -\cos(\phi - \alpha) \end{bmatrix} \cdot \begin{bmatrix} E_\theta \\ E_\phi \end{bmatrix} \tag{18.11}$$

where E_θ and E_ϕ are the spherical components in the θ and ϕ directions, ϕ is the azimuth angle, and α is the polarization angle referenced to the x-axis of the coordinate system (Figure 18.4).

Cross polarization becomes important in antenna design when two orthogonal polarizations are used to transmit two separate channels. Frequency reuse by dual polarization

effectively doubles the available system bandwidth. The antenna, and the feed in particular, is the means by which isolation is maintained between the channels. Signals may be sent or received in either linearly polarized vertical and horizontal components or circularly polarized, where the signals rotate in space and time in a right- or left-hand sense.

18.1.3.5 Input Match This is the input reflection coefficient of the feed antenna when it is excited at its input. Usually this is measured with the feed alone; however, care must be taken that reflection from the secondary antenna does not significantly affect the match. If this occurs, a matching vertex plate can be added to the reflector or lens to reduce the reflection.

18.1.3.6 Bandwidth The bandwidth of the feed antenna is the continuous frequency range over which the feed performance is acceptable. The parameter measured is often the input reflection coefficient and the acceptable level depends on the application. In satellite ground stations and radio astronomy, a commonly used maximum level is −17.7 dB (VSWR = 1.3). The percentage bandwidth is used to compare the performance of different feeds. This is calculated from the minimum and maximum frequencies, $f_{\min}$ and $f_{\max}$, over which the parameter meets the criterion, thus

$$\% \text{ Bandwidth} = 200\frac{(f_{\max} - f_{\min})}{(f_{\max} + f_{\min})} \tag{18.12}$$

18.1.3.7 Spillover In transmit mode, the energy from a feed antenna that is not incident on the secondary antenna is called spillover. In receive mode, received spillover is an important contributor to the overall antenna noise temperature. Considerable effort goes into minimizing the amount of spillover in systems requiring low antenna noise temperatures such as radio astronomy and deep space probes. Spillover efficiency, η_{spill}, is the ratio of the power intercepted by the secondary antenna to the total power radiated by the feed.

18.1.3.8 Edge Illumination and Edge Taper Edge illumination is the ratio of the field strength radiated by the feed in the direction of the edge of the secondary antenna and its level at the center of the secondary antenna, where field strength is measured on a sphere of radius equal to the distance to the center of the secondary antenna. Edge taper is a related quantity and gives the ratio of the feed field strength at the center of the secondary antenna relative to the level at the edge. The difference between edge illumination and edge taper (apart from the sign) is the free-space loss between the sphere and the edge. In a symmetrical reflector antenna, the edge is the reflector rim and the intersection point is the vertex. The situation in the case of an offset parabolic reflector is illustrated in Figure 18.3. In the general case of an offset parabolic reflector with focal length F and diameter D_m,

$$\text{Edge taper (dB)} = 20\log_{10}\left(\frac{\cos^2(\psi_o 2)}{F}\sqrt{x_m^2 + x_m D_m \cos\phi_e + F^2\left(1 + \frac{1}{4(F/D_m)^2}\right)^2}\right) - \text{Edge illumination (dB)} \tag{18.13}$$

TABLE 18.1 Types of Feed Antennas

Aperture	Linear	Traveling-Wave	Compound
Waveguide	Dipole	Dielectric rod	Array
Horn	Yagi–Uda	Profiled slot	Beam waveguide
Microstrip patch	Log-periodic	Vivaldi	Splash plate
Slot	Zig zag	Line source	Dichroic reflector
Rear-radiating			Reflector
Resonant-cavity			

where $x_m = 2F \tan(\psi_o/2)$ and ϕ_e is the azimuthal angle on the rim (both are shown in Figure 1.3). The highest edge taper occurs when $\phi_e = 0^\circ$ and the lowest taper is when $\phi_e = 180^\circ$. Clearly, for a paraboloid ($\psi_o = 0^\circ$) the edge taper is constant at all angles ϕ_e.

The choice of primary feed illumination for a secondary antenna (e.g., reflector F/D_m and ψ_0) is often a compromise between spillover efficiency and the secondary antenna aperture efficiency (η_{sae}). As the feed beamwidth is increased, the edge taper decreases and η_{sae} increases. At the same time, the spillover loss increases (η_{spill} decreases), causing the overall antenna efficiency to decrease. Consequently, there is an optimum edge taper that maximizes the overall antenna efficiency and, therefore, gain for a given secondary antenna. A -12-dB edge illumination is recommended for a paraboloid as a compromise between gain, sidelobes, and spillover (noise). With this edge illumination, a paraboloid with $F/D_m = 0.4$ has an edge taper of 14.9 dB. Further increase in edge taper results in lower sidelobes and wider beamwidth in the secondary pattern. A similar argument applies for a lens, except that a lower edge taper can be used in receive-only applications because the feed spillover radiation contributes to the sidelobes close to the main beam rather than the rear lobes of the antenna.

18.1.4 Types of Feed Antennas

Feed antennas can be grouped into four basic categories as shown in Table 18.1. The class of aperture antennas is the most extensive and has the property that the area normal to the direction of propagation of the radiation from the antenna determines the gain and beamwidth. Linear antennas are distinguished by their physical appearance and how the dominant currents are supported on line elements. Traveling-wave antennas derive essentially from fields or currents that propagate along the antenna and radiate in the endfire direction with gain that is approximately proportional to the length of the antenna. Most linear arrays belong to this category also. Compound feeds are made up of several antenna types including aperture or traveling antennas with reflectors, lenses, or grids. In the following sections, different feed types are described based on this classification.

18.2 APERTURE FEED ANTENNAS

Aperture antennas radiate from a surface that is the interface between the circuit attached to the transmitter or receiver and the radiation field. On this surface actual or equivalent

electric and magnetic currents flow and from these the radiated fields and impedance properties can be derived. The equivalence principle says there is no unique surface; although for horns, the aperture is commonly taken as the opening and the external walls. For slot antennas, the aperture is usually the opening and the surrounding metallic surface, while for rectangular microstrip antennas the aperture is sometimes taken as a pair of rectangular "slots" formed between the edges of the patch and the ground plane. Examples of some aperture antennas are shown in Figure 18.5.

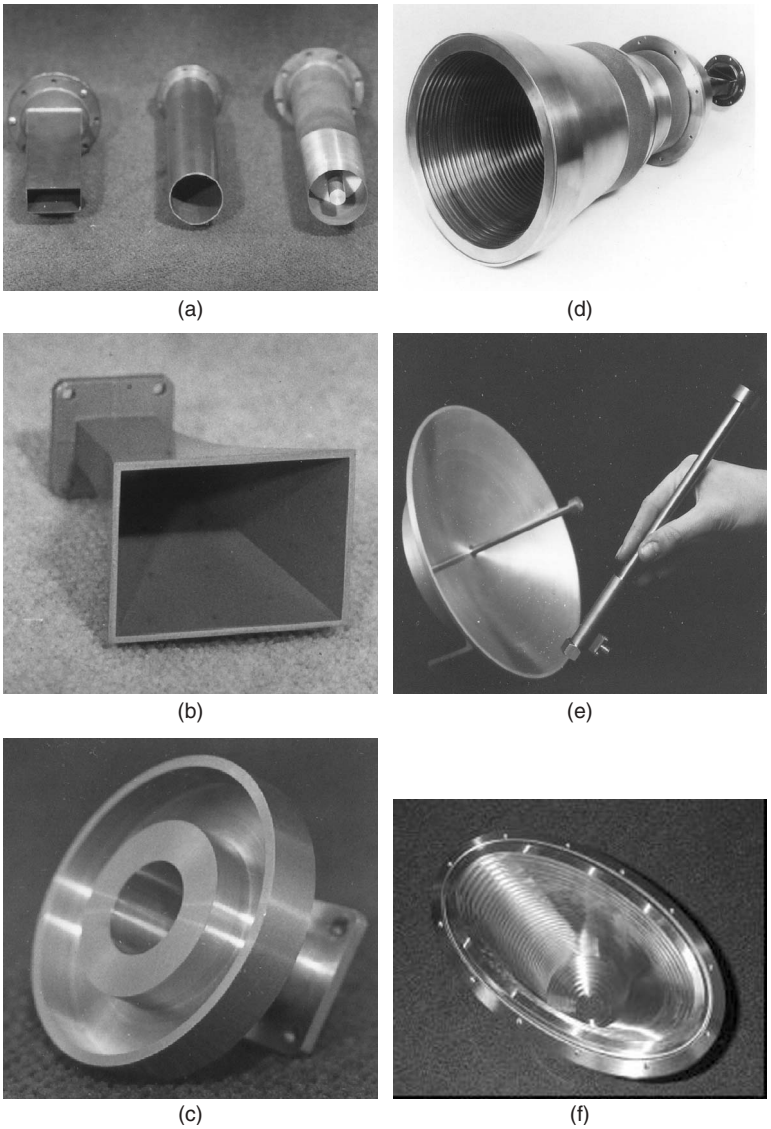


Figure 18.5 Examples of aperture antennas: (a) waveguide feeds—left to right, rectangular, circular, and coaxial; (b) pyramidal horn; (c) circular waveguide with choke ring; (d) corrugated horn; (e) cup feed; and (f) profiled superelliptical corrugated horn.

Some of the simplest and most commonly used aperture feeds are based on smooth metallic waveguides and their extension by flaring into horns. The main types of waveguide and horn feed have rectangular or circular cross sections. In feed applications, the radiation in the forward direction is the fundamental concern and, to a reasonable approximation, radiation is due to the dominant or first waveguide mode.

A commonly used model for radiation from aperture feeds is the E-field model. It is so called because the field is derived from the electric fields in the aperture and corresponds physically to an aperture opening out into an infinite ground plane. The model is useful for predicting the field in the forward direction and, of course, it predicts zero fields behind the conductor. The fields radiated by simple waveguide and horn feeds are a spherical wave of the form

$$\mathbf{E} = jk\frac{\exp(-jkr)}{2\pi r}(\hat{\boldsymbol{\theta}}A(\theta,\phi)\cos\phi - \hat{\boldsymbol{\phi}}\cos\theta\ B(\theta,\phi)\sin\phi), \quad |\theta| \le 90^\circ \tag{18.14a}$$

$$\mathbf{H} = \frac{1}{\eta}\hat{\mathbf{r}} \times \mathbf{E} \tag{18.14b}$$

where $k = 2\pi/\lambda$, λ is the free-space wavelength, and the functions $A(\theta, \phi)$ and $B(\theta, \phi)$ depend on the aperture geometry. The radial field components are small in the far field and are neglected in Eq. (18.14), although they are important when the secondary antenna is in the near field of the feed. Assuming the waveguide aperture is in a ground plane or the waveguide wall is thick, A and B are approximately given by the E-field model and are listed in Table 18.2 for rectangular, circular, coaxial, and corrugated waveguides. Other waveguide cross sections, such as the elliptic and polygonal shapes, have the same form as Eq. (18.14), although the functions are not expressible in terms of simple functions and must be computed by other methods. The effect of a finite-size ground plane or flange can be included in the E-field model simply by the method of geometrical or physical theory of diffraction [5, 13]. The main effect on the main beam of the finite ground plane is to change the multipliers (known as Huygens factors) in Eq. (18.14), respectively 1 and $\cos\theta$ of the $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\phi}}$ vector components. In this model $|A(\theta, 0)|$ and $|B'(\theta, \pi/2)|$ is the H-plane pattern, where $B'(\theta, \phi) = \cos\theta B(\theta, \phi)$.

When the pattern functions $A(\theta, \phi)$ and $B'(\theta, \phi)$ are independent of ϕ, the radiation pattern is axisymmetric; furthermore, if A and B' are identical the cross-polarized field is zero. Ideal corrugated horns have these properties if the "balanced hybrid condition" is satisfied. For smooth-wall circular horns, A and B' are generally different, although the phase of these functions may be almost the same. Equation (18.14) shows that for feeds with axisymmetric patterns the peak cross polarization is in the intercardinal planes and equals $|A(\theta) - B'(\theta)|/2$; that is, the cross polarization depends on the difference of the E- and H-plane pattern functions.

The copolar pattern in the intercardinal planes is $|A(\theta) + B'(\theta)|/2$, that is, the average of the E- and H-plane pattern functions. The directivity is obtained from Eq. (18.6) as

$$D(\theta,\phi) = \frac{(|A(\theta,\phi)|^2\cos^2\phi + |B(\theta,\phi)|^2\cos^2\theta\sin^2\phi)}{\int_0^{2\pi}\int_0^{\pi/2}(|A(\theta',\phi')|^2\cos^2\phi' + |B(\theta',\phi')|^2\cos^2\theta'\sin^2\phi')\sin\theta'\,d\theta'\,d\phi'} \tag{18.15}$$

TABLE 18.2 Simple Waveguide Feeds

Type of Feed Aperture	$A(\theta, \phi)$	$B(\theta, \phi)$	Notes
TE_{01} mode rectangular waveguide	$\frac{2ab}{\pi} S\left(U\frac{b}{2}\right) C\left(V\frac{a}{2}\right)$	$A(\theta, \phi)$	a and b are the width and height of the waveguide
TE_{11} mode circular waveguide	$2\pi a J_1(k_c a)\frac{J_1(wa)}{k_c wa}$	$2\pi a J_1(k_c a)\frac{k_c J_1'(wa)}{k_c^2 - w^2}$	a is the waveguide radius and $k_c a = 1.8412$
TE_{11} mode coaxial waveguide	$\frac{2\pi a[Z_1(k_c a, k_c b)J_1(wa) - Z_1(k_c b, k_c b)J_1(wb)]}{k_c wa}$	$\frac{2\pi a k_c\left[Z_1(k_c a, k_c b)J_1'(wa) - \frac{b}{a}Z_1(k_c b, k_c b)J_1'(wb)\right]}{k_c^2 - w^2}$	a and b are outer and inner conductor radii. $k_c a \approx 2/(1 + b/a)$ and is the first root of $J_1'(k_c a)Y_1'(k_c b) - J_1'(k_c b)Y_1'(k_c a) = 0$
HE_{11} mode corrugated waveguide	$2\pi a J_1(k_\rho a)\frac{k_\rho J_0(wa)}{k_\rho^2 - w^2}$	$A(\theta, \phi)$	a is the waveguide radius and $k_\rho a = 2.4048$

$U = w\cos\phi, V = w\sin\phi, w = k\sin\theta, J_p$ is a Bessel function of order p and J_p' is its first derivative. Y_1 is a Bessel function of the second kind and Y_1' is its first derivative. $S(x) = (\sin x)/x$, $C(x) = (\cos x)/(1 - (2x/\pi)^2)$, $Z_1(x, y) = J_1(x) - [J_1'(y)/Y_1'(y)]Y_1(x)$.

18.2.1 Feeds with Circular and Related Cross Section

Circular aperture antennas find wide application as feeds due to their geometrical symmetry, low cross polarization, and ease of manufacture. Examples of this type of feed are illustrated in Figure 18.6.

A circular waveguide (Figure 18.6a) is an efficient feed for moderately deep reflectors or lenses where the half-cone angle is around 60°. It has an almost equal E- and H-plane pattern (pattern symmetry) for pipe diameters in the range 0.7–1.2 wavelengths depending on the size of the flange at the aperture. Higher efficiencies may be obtained by the addition of parasitic rings as shown in Figure 18.6b. A coaxial waveguide (Figure 18.6c) operating predominantly in the TE_{11} mode has potentially greater flexibility of available radiation patterns because of the extra degree of freedom given by the internal conductor. Figure 18.6d shows a self-supporting radiating coaxial waveguide (sometimes called a "tomato can") feed for a reflector. Here the inner conductor extends all the way to the vertex of the reflector, allowing the feed to be driven from a transmission line or waveguide inside this conductor.

Flaring the circular or coaxial waveguide produces a conical horn (Figure 18.6e) that has a more directive radiation pattern. The quality of the pattern depends on many factors including the size of the output aperture, the internal structure of the flare, and

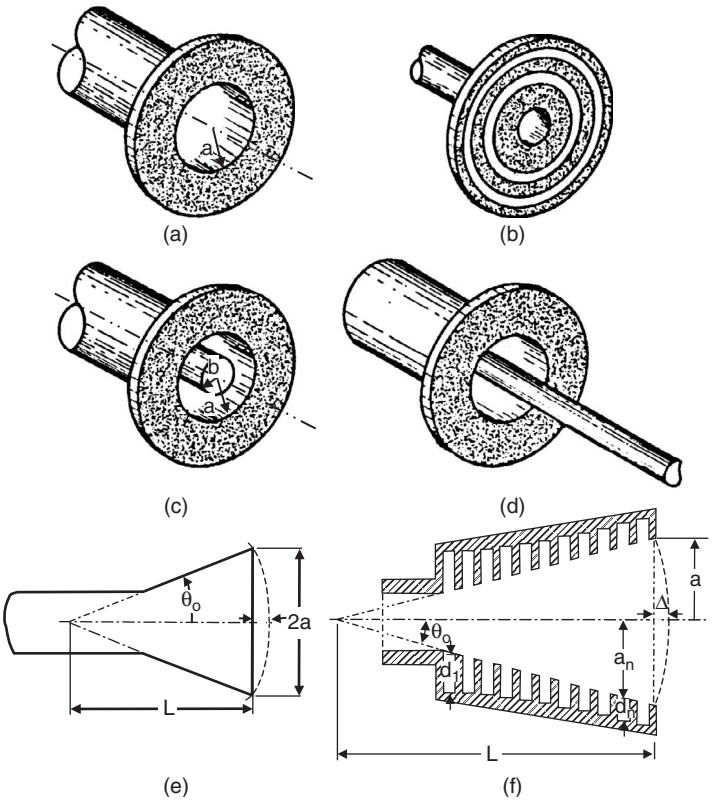


Figure 18.6 Circular waveguides and horn antennas: (a) circular waveguide, (b) circular waveguide with parasitic rings, (c) coaxial waveguide, (d) rear-radiating coaxial waveguide, (e) conical horn, and (f) corrugated horn, (parts (b)–(d) from Ref. 14, courtesy of IEEE.)

the means of exciting the flared section. As an example, a horn with transverse internal corrugations (Figure 18.6f), the so-called corrugated horn, is very commonly used in satellite communications and radio astronomy because of its low cross polarization. When the depth of the corrugations is about a quarter of a wavelength, almost pure polarized radiation patterns result. Several other types of feeds with circular or elliptical cross section are discussed in the following sections.

18.2.1.1 Circular and Coaxial Waveguide Smooth wall circular or coaxial waveguides and horns are most frequently excited in the first waveguide mode, the TE_{11} mode (see Figure 18.7a). The transverse field components of this mode in a circular waveguide are

$$\mathbf{E}_t = c_0 \left[\hat{\boldsymbol{\rho}} \frac{J_1(k_c\rho)}{k_c\rho} \cos\phi - \hat{\boldsymbol{\phi}} J_1(k_c\rho) \sin\phi \right] e^{-j\beta z} \tag{18.16a}$$

and

$$\mathbf{H}_t = Y_w \hat{\mathbf{z}} \times \mathbf{E}_t \tag{18.16b}$$

where $k_c = 1.8412/a$ is the cutoff wavenumber, a is the waveguide radius, $Y_w = \beta/k\eta$ and $\beta = \sqrt{k^2 - k_c^2}$, J_1 is the Bessel function of first order, c_0 is a normalization constant related to power in the mode, and (ρ, ϕ, z) are the cylindrical polar coordinates.

When the TE_{11} mode is incident on the open end as in Figure 18.7a, an external field is produced and the TE_{1n} and TM_{1n} $(n = 1, 2, \ldots)$ modes are excited in order to satisfy the boundary conditions. Part of the power in these modes is radiated and another part is reflected from or excited at the aperture and returns back down the waveguide; the dominant reflected component is due to the TE_{11} mode. The resulting excitation and coupling between modes in the aperture is often ignored as a first approximation, but for accurate design it should be included. To first order, the aperture fields ($z = 0$ in Eq. (18.16)) expressed in terms of rectangular components are

$$\mathbf{E}_a = \hat{\mathbf{x}} \frac{c_0}{2} [J_0(k_c\rho) + J_2(k_c\rho) \cos 2\phi] + \hat{\mathbf{y}} \frac{c_0}{2} J_2(k_c\rho) \sin 2\phi \tag{18.17a}$$

and

$$\mathbf{H}_a = Y_w \hat{\mathbf{z}} \times \mathbf{E}_a \tag{18.17b}$$

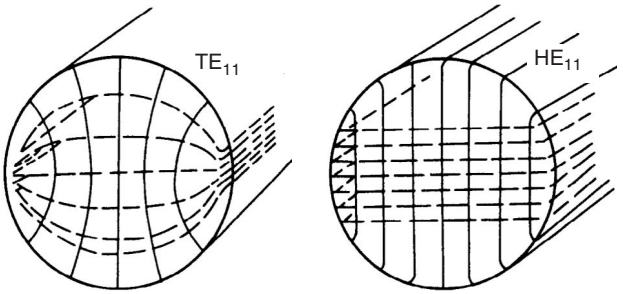


Figure 18.7 Electric and magnetic fields in (a) smooth-wall circular waveguide and (b) corrugated waveguide (— — — electric; ------ magnetic).

Equations (18.17a) and (18.17b) show that the TE_{11} mode has both x- and y-field components and therefore so do the radiated fields. The principal aperture field polarization is maximum on axis and the cross-polarized component is maximum in the intercardinal (45° and 135°) planes. When a center conductor of radius b is introduced into the circular waveguide, as in Figure 18.6c, the basic properties of a TE_{11}-mode excited coaxial waveguide are similar, but there are significant differences, as shown in Figures 18.8–18.10. The fields radiated by circular and coaxial waveguides are given in Table 18.2. The −12-dB semiangle of the principal radiation patterns for circular ($b = 0$) and coaxial waveguides ($b\rangle 0$) is shown in Figure 18.8 and the maximum cross-polar level is given in Figure 18.9. The E- and H-plane patterns of the circular waveguide are approximately the same when $ka \sim 3.4$ ($a \sim 0.55\lambda$), and this results in a cross-polar minimum for the same ka value (Figure 18.9). The coaxial waveguide also has a cross-polar minimum, as Figure 18.9 shows, which depends on the inner-to-outer conductor ratio (b/a).

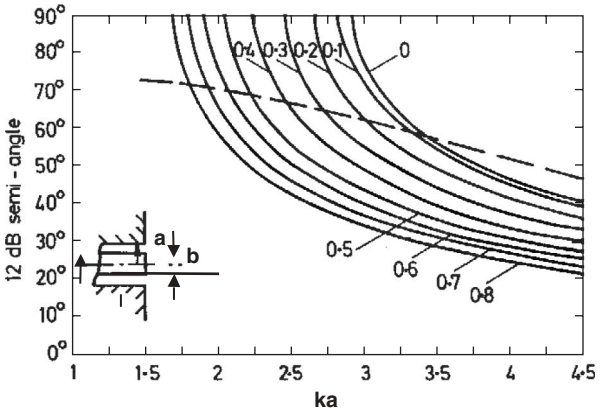


Figure 18.8 The 12-dB semiangle of the far-field radiation pattern of a coaxial waveguide. Solid line E-plane. Dash line H-plane. Parameter is b/a. (From Ref. 14, courtesy of IEEE).

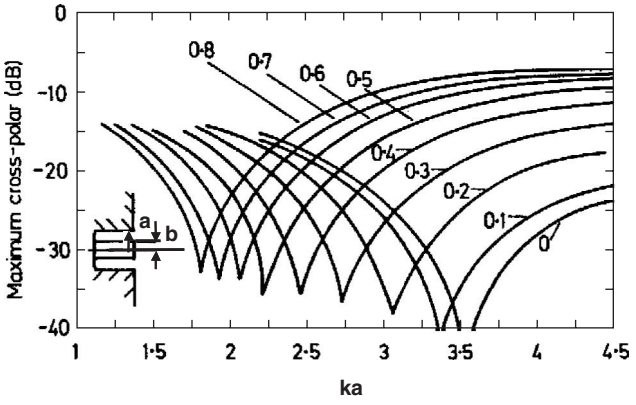


Figure 18.9 Maximum cross-polar level relative to peak copolar level within a 75° semiangle versus normalized frequency ka. Parameter is b/a. (From Ref. 14, courtesy of IEEE).

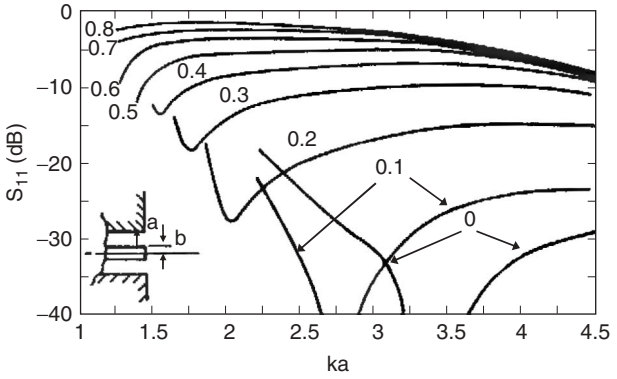


Figure 18.10 Reflection coefficient at the aperture of a coaxial waveguide. Parameter is b/a. (From Ref. 15, courtesy of IEEE).

The reflection coefficient, S_{11}, of the TE_{11} mode in the circular waveguide due to the aperture mismatch is approximately

$$S_{11} = \left(\frac{1 - y_{11}}{1 + y_{11}}\right) \tag{18.18}$$

where the admittance of the TE_{11} mode in the circular waveguide is given by

$$y_{11} = 9.617\frac{k}{\beta}\int_0^\infty \left(\frac{w\sqrt{1-w^2}}{(ka)^2}\frac{J_1'^2(kaw)}{[w^2 - (1.84118/ka)^2]^2} + 0.087\frac{J_1^2(kaw)}{w\sqrt{1-w^2}}\right) dw \tag{18.19}$$

As shown in Figure 18.10, S_{11} has a minimum when $ka \sim 3.4$ ($a \sim 0.55\lambda$) which is coincident with the condition for minimum cross polarization. Providing $a > \lambda$, the TE_{11}-mode reflection coefficient is less than -30 dB. The coaxial waveguide has a large mismatch, as Figure 18.10 shows, when the ratio of the inner-to-outer conductor diameters exceeds about 0.3. By means of a pair of capacitive and inductive coaxial irises, it is possible to match coaxial feeds with inner-to-outer conductor ratios of up to 0.6 over $>17\%$ bandwidth. This is achieved by stepping down to an inner-to-outer conductor diameter ratio of 0.3 and matching into this guide using the technique described by Bird et al. [15].

The maximum gain predicted by the E-field model for the circular waveguide is

$$G_{\max} = 0.209(ka)^2\frac{k}{\beta}\left(1 + \frac{\beta}{k}\right)^2\left(\frac{|1 + S_{11}|^2}{1 - |S_{11}|^2}\right) \tag{18.20}$$

where the mode is above cutoff. For frequencies well above cutoff ($\beta \approx k$), $G_{\max} \approx 0.837(ka)^2$ and the maximum aperture efficiency is 83.7%.

Circular horns are desirable for dual-polarized applications because of their geometrical symmetry. However, this does not ensure low cross polarization and thereby high polarization isolation. We have seen that the TE_{11} mode of the circular waveguide has a radial directed electric field at the wall, and therefore cross-polarized fields arise to

satisfy the boundary conditions. Cross polarization is transferred from the aperture to the radiation field but this can be reduced in some situations by exciting other modes.

Cross-polar patterns of rationally symmetric feeds have a characteristic null on axis. If the symmetry is broken and other modes are excited it is possible to have cross polarization on axis. Equality of the E- and H-plane patterns may be difficult or impossible to achieve in practice over any reasonable bandwidth. In feed applications, reasonably low cross polarization can be achieved in the secondary pattern by selecting the geometry so that the E- and H-plane patterns cross over at around the -10-to -14-dB level. Choke rings either inside the horn or at the aperture (Figure 18.6b) are useful for tailoring the radiation pattern of small horns to minimize cross polarization over a moderate bandwidth [16]. Figure 18.8 shows that up to a certain ka value, the E-plane beamwidth is wider than the H-plane. By placing a 0.2λ to 0.26λ deep choke ring concentric with the aperture, the E-plane beamwidth can be reduced to almost the same as the H-plane beamwidth.

18.2.1.2 Conical Horn If a circular aperture is flared into a conical horn as shown in Figure 18.6e, the field at the aperture has a spherical wavefront. In simple terms, when the TE_{11} mode reaches the flare, it expands outward in order to satisfy the wall conditions and forms a spherical wave, which propagates to the aperture. Therefore the distance between a planar and a spherical wavefront is approximately given by [5]

$$\delta \approx \frac{1}{2}\frac{\rho'^2}{L} \tag{18.21}$$

where ρ' is the radial distance from the axis to the field point on the aperture and L is the distance from the horn apex to the aperture. The result is that Eq. (18.17) is modified by the application of the phase factor $\exp(-jk\delta)$ and the radiated fields are now given approximately by the functions

$$A(\theta, \phi) = L_0(\theta) - L_2(\theta) \quad \text{and} \quad B(\theta, \phi) = L_0(\theta) + L_2(\theta) \tag{18.22}$$

where

$$L_m(\theta) = 2\pi \int_0^a J_m(1.841184\rho'/a) J_m(k \sin\theta \ \rho') \exp(-jk\rho'^2/2L)\rho' \, d\rho. \tag{18.23}$$

Equation (18.22) gives reasonable results providing the semicone angle $\theta_o = \arctan(a/L)$is less than about 30°.

The radiation pattern of the conical horn is strongly dependent on the flare angle. In Figure 18.11 the E- and H-plane patterns are plotted as a function $\Delta = (a/\lambda)\tan(\theta_o/2)$, which is the distance in wavelengths between the spherical wavefront and the aperture (see Figure 18.6e). As the angle is increased, the patterns broaden and lose distinctive sidelobes. The phase center in these two principal planes has been calculated from Eq. (18.8) over the angular range defined by the -12-dB semiangles. The results given in Figure 18.12 show when Δ is small, the phase center is near the aperture. However, as Δ increases, the phase centers in the two planes move toward the apex at a different rate. When $\Delta > 0.8$, the phase center is almost at the apex.

The radiation performance of the conical horn can be improved by changing the surface of the horn, with corrugations, for example, or by introducing discontinuities

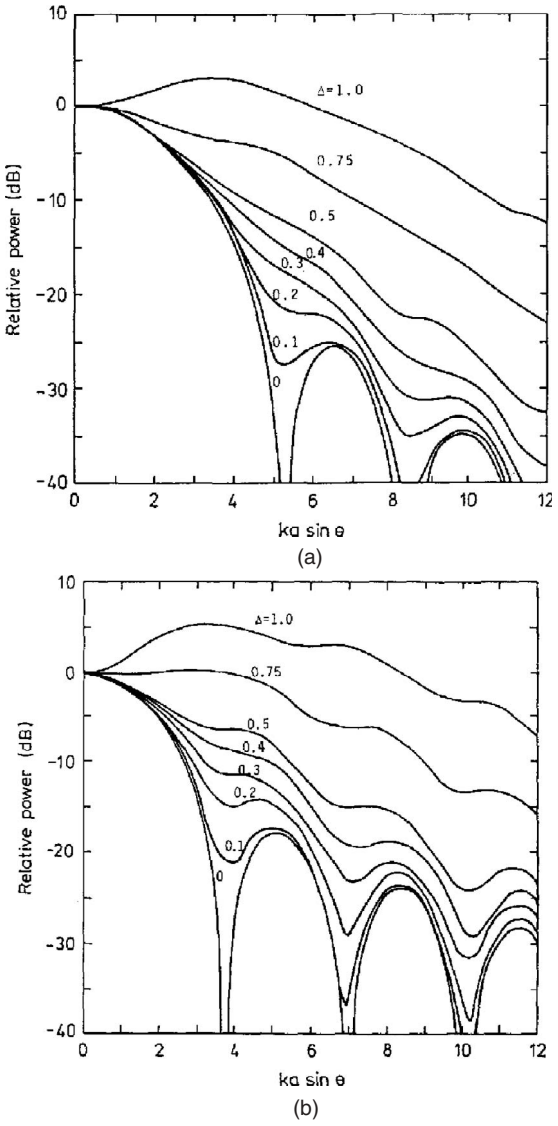


Figure 18.11 Normalized radiation patterns of conical horn where $\Delta = (a/\lambda)\tan(\theta_0/2)$ is the parameter. (a) E-plane; and (b) H-plane.

to excite high-order modes, particularly the TM_{11} mode. The Potter horn discussed in Section 18.2.1.6 uses steps placed near the throat to produce such improvements.

18.2.1.3 Corrugated Horns The corrugated horn is widely used as a feed antenna for reflector antennas and lenses because it can radiate fields with a high degree of axial symmetry. Smooth-wall circular waveguides and horns have, in general, different E- and H-plane patterns because the electric and magnetic field components in the pipe satisfy different types of boundary conditions on the walls (see Figure 18.7). In simple terms, the tangential electric field of the TE_{11} mode experiences a short circuit at the wall

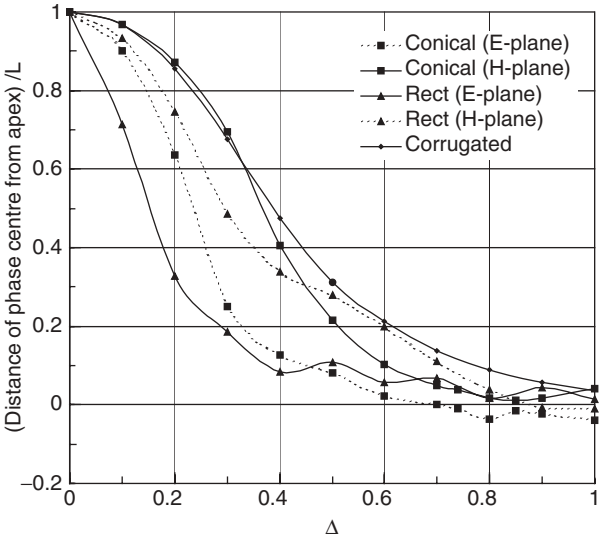


Figure 18.12 Phase center location for the principal planes of conical, corrugated, and rectangular horns.

while the tangential magnetic field experiences an open circuit. The end result is that the field lines in the waveguide are curved. By modifying the wall-surface conditions, such as corrugating the surface, covering it with dielectric, or inserting strips, the wall impedance experienced by the mode can be altered and the field lines made almost unidirectional, which results in low cross polarization.

The corrugated surface was one of the first used to create alternative wall impedances and it has proved most successful in circular waveguide and horns. When there are many corrugations per waveguide (typically greater than about five per wavelength) and the corrugation depth is approximately $\lambda/4$ (where λ is the free-space wavelength), the radial transmission line formed by the slots transforms the short circuit for the electric field into an open circuit while having only a second-order effect on the transverse magnetic field. This creates a nonzero axial electric field at the corrugated surface, allowing a zero circumferential component. Consequently, the modes in the corrugated waveguide or horn are no longer TE or TM to the longitudinal (propagation) direction and, because both axial field components are generally nonzero, they are called hybrid modes. The principal mode used for feed horns is the HE_{11} mode and this is shown in Figure 18.7b. An axisymmetric radiation pattern is produced if the HE_{11} mode operates at the "balanced hybrid condition," which is achieved when the corrugations are approximately $\lambda/4$ deep. Under this condition, the horn has low cross polarization, pattern symmetry, and low sidelobes, which are almost ideal for many applications and give rise to the popularity of these hybrid-mode feed horns [17, 18].

At the balanced hybrid condition, the aperture electric field of the HE_{11} mode is approximately

$$\mathbf{E}_a = \hat{\mathbf{x}} c_0 J_0(k_\rho \rho) \tag{18.24}$$

and the magnetic field is given by Eq. (18.17b). J_0 is the zero-order Bessel function, c_0 is a constant, $\rho(\leq a)$ is the radical distance, and $k_\rho a = 2.4048$. The far-zone field radiated by

the HE_{11} mode is provided in Table 18.2. The maximum gain of the corrugated waveguide under these conditions is $G_{\max} \approx 0.692(ka)^2$ and the maximum aperture efficiency is 69.2%.

The conical corrugated horn may be modeled by the same method described earlier for the smooth-wall conical horn. The radiated fields are given approximately by Eq. (18.14), where

$$A(\theta, \phi) = L_0(\theta) = B(\theta, \phi) \tag{18.25}$$

where $L_0(\theta)$ is given by Eq. (18.23), wherein $m = 0$ and k_c is replaced by k_ρ.

The radiation pattern of a corrugated horn operating in the balanced HE_{11} mode is given by the set of universal curves shown in Figure 18.13 plotted as a function of $\Delta = (a/\lambda)\tan(\theta_o/2)$. When $\Delta > 0.4$, the shape of the horn's pattern changes very little with Δ, indicating the pattern and gain become virtually independent of aperture size. This is called "gain saturated" and means the horn is suitable for wideband operation. For $\Delta < 0.4$, the horn is said to be narrowband.

By virtue of the high pattern symmetry, the phase center for the corrugated conical horn is the same in all planes. Figure 18.12 gives the phase center location as a function of Δ based on the pattern over the -12-dB beamwidth.

The reflection coefficient of the HE_{11} balanced hybrid mode at the aperture is relatively low in most practical cases. If required it can be calculated from Eq. (18.17) and the self-admittance

$$y_{11} = \frac{5.783}{(ka)^2}\frac{k}{\beta}\int_0^\infty \frac{w(2-w^2)}{\sqrt{1-w^2}}\frac{J_0^2(kaw)}{[w^2-(2.4048/ka)^2]^2}dw \tag{18.26}$$

Equation (18.26) predicts that when $a > 1.1\lambda$, the HE_{11}-mode reflection from the aperture is less than -30 dB.

The corrugated horn equations given in Table 18.2 and Eq. (18.24) are useful only for narrowband feed design near the balanced hybrid condition. In a typical conical corrugated horn, there are five to ten slots per wavelength, the slot width-to-pitch ratio

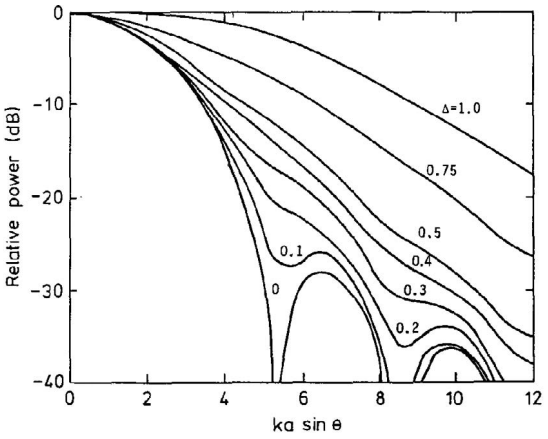


Figure 18.13 Normalized radiation pattern of corrugated horn with balanced hybrid mode. Δ is the parameter.

(w/p) is >0.75, and the slot depth is approximately [19]

$$d_n \approx \frac{\lambda}{4} \exp[2/(5ka_n)] \tag{18.27}$$

where a_n is the radius of the corrugated surface at the nth slot (see Figure 18.26f). Wideband feed design requires more accurate design methods and to achieve best performance attention needs to be paid to the input throat region, the flare from the input, and the profile to the aperture. The corrugated horn is limited to a continuous bandwidth performance of about 2.2:1. Thomas et al. [20] discuss matching of corrugated horns and wideband design in detail.

The standard design of a corrugated horn starts from the smooth-wall input circular waveguide with a section of about a wavelength long, where the depth of the corrugations are gradually decreased from a half-wavelength to a quarter-wavelength deep as the horn is widened. Following this input transition, a linear taper can be used for the horn profile, although some other profiles, such as "hourglass" (sine-squared), have been found more effective. To achieve the widest bandwidth operation, the first four or five slots of the input section may need to terminate in a short-circuited coaxial waveguide—known as ring-loaded slots. These slots present a capacitive reactance over a considerably wider bandwidth than that of a conventional constant width slot. The design of ring-loaded slots is described by James and Thomas [19]. The final part of the design is tapering the profile from the initial transition to the final output diameter. Corrugated horns with a profiled corrugated internal surface [21] produce a more compact design while maintaining high performance. A number of different profiles have been tried with the aim of achieving low sidelobes and cross polarization. The hyperbolic [10] or the identical Gaussian [22] profile has been found very effective overall for designing compact horns. Thus if a_i is the radius of the input section, a_o is the output or aperture radius, and L is the length of the profile section, the horn profile as a function of axial distance (z) is given by

$$a(z) = a_i \sqrt{1 + \left(\frac{z}{L}\right)^2 \left[\left(\frac{a_o}{a_i}\right)^2 - 1\right]} \tag{18.28}$$

18.2.1.4 Horns with Axial Corrugations Horns with internal axial corrugations (or internal short-circuited ring slots) are widely used as feeds for direct broadcast by satellite applications. As well as having good performance, they are relatively inexpensive to manufacture, especially by die-casting. There are several different types of axially corrugated horn. One particular configuration is the choked aperture feed shown in Figure 18.14a. It has choke rings in the aperture plane and also internal ring slots [23]. Although a simple prime-focus feed, it has an excellent performance over nearly an octave bandwidth.

Through the selection of a number of internal axial corrugations, a combination of TE_{11}, TM_{11}, and TE_{12} modes is produced in the aperture. These modes are maintained by a compensating action of the TE_{11} mode in the coaxial corrugations, which vary in length from about 0.37λ at the throat to about 0.25λ at the aperture. At least two corrugations are needed to achieve reasonable performance but little is gained from using more than four corrugations.

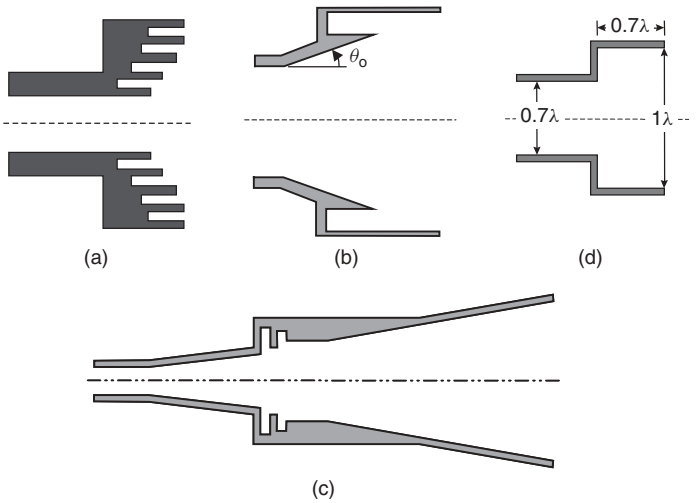


Figure 18.14 Circular horn feeds: (a) circular horn with axial corrugations, (b) scrimp horn, (c) potter horn, and (d) stepped horn.

18.2.1.5 Scrimp Horn Originally devised by Wolf [24], the scrimp horn (Figure 18.14b) is a compact and lightweight feed element for satellite antenna feed arrays. The scrimp horn has high aperture efficiency (typically $>80\%$), operating bandwidth of about 20%, low cross polarization, and better sidelobes than most similar sized horns. The scrimp horn has three main sections: a conical horn input section, a shorted coaxial waveguide section, and a cylindrical output waveguide section that terminates in the aperture. A TE_{11} mode in a circular waveguide of between 0.75λ and 0.95λ is input to the conical horn section and this generates a standing wave in the coaxial waveguide section, which is dominated by a TE_{11} mode. The semiangle of the flare (θ_o in Figure 18.14b) of the conical horn is typically 15° to 25° and the length of the coaxial waveguide is about 0.25λ. This input TE_{11} mode in the conical horn oscillates in odd phase to the TE_{11} mode in the output section and to ensure continuity of the fields additional higher-order modes, mainly TM_{11} and TE_{12}, are generated. These modes are superimposed with the fundamental TE_{11} mode to produce a field distribution in the aperture that is low in cross-polarized components. The aperture diameter can be from about 1.2λ to about 1.8λ depending on the required edge taper.

18.2.1.6 Multimode and Stepped Circular Horns The basic idea of multimode and stepped horns is to excite modes in addition to the fundamental mode in such a way as to improve the aperture efficiency, shape the radiation pattern, or reduce cross polarization. An early feed antenna that was designed for low cross-polarization characteristics is the dual-mode horn (Figure 18.14c) described by Potter [6]. In this horn, a combination of TE_{11} and TM_{11} modes that are approximately in the ratio 1:0.3 in the aperture modes are excited in a smooth-wall horn. The modes are brought into the correct phase relationship over the length of the horn. This concept has been extended by several workers including Takabayashi et al. [25], where a double-band dual-mode conical horn design is described. The band separation is not given and, as is typical with this type of design, the individual bandwidths are quite narrow (approximately $<5\%$).

Another approach due to Satoh [26] uses a thin dielectric ring in the flare of the horn, producing a rotationally symmetric beam and low sidelobe levels over a bandwidth greater than 25%.

Instead of inserting discontinuities inside the horn to generate higher-order modes, the horn profile can also be stepped. A horn of this type, which is useful for closely packed arrays, is the one-wavelength diameter dual-mode stepped horn in Figure 18.14d [27]. This feed is excited by a traveling TE_{11} mode and an evanescent TM_{11}, which is produced at the step, is brought into phase over about a 0.7λ uniform section to the aperture.

An example of multistepped circular horns are the elements of the multibeam array feed for the Parkes radio telescope [28], which has been in operation since early 1997 (the horns and receiver cryogenics are shown in Figure 18.15). Bandwidths of about 20% can be achieved with these types of horns.

The trimode horn of Rudge and Adatia [29] employs both steps and inserts to excite modes in a circular horn to cancel cross polarization in a single offset parabolic reflector via (Eq. (18.1)). In this "matched" feed approach, the horn aperture field is a conjugate match of the field distributions at the focus of an offset parabolic reflector system, thereby compensating for the cross-polarized fields introduced by the reflector. The TM_{11} and TE_{21} modes are produced with just the right amount of amplitude and phase to cancel the inherent cross-polar fields due to the single offset. The fundamental TE_{11} mode is complemented by the TM_{11} mode, which is excited with a stepped section, and the TE_{21} mode, which is created with two orthogonal coplanar posts. Other means of exciting the TE_{21} mode can be used. The amplitudes of the modes are determined by ratios of the diameters and the probe depths. Matched feeds are narrowband in performance and limited to only a few percent in operating range.

Aperture efficiencies in excess of 90% can be obtained with circular horns by exciting the right combination of modes using steps. For example, Bhattacharyya and Goyette

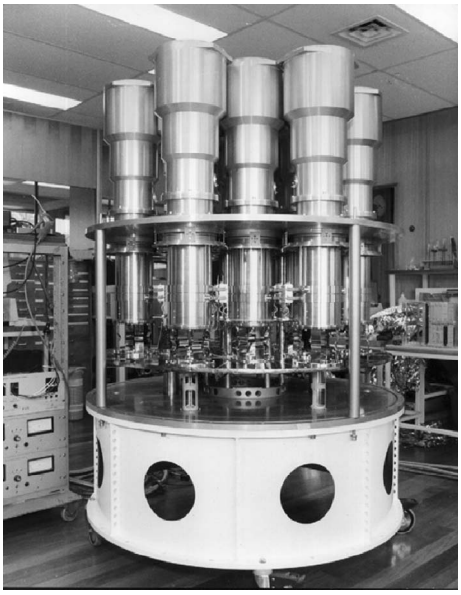


Figure 18.15 The 13-element multibeam feed system for the Parkes (Australia) radio telescope.

[30] achieve both high aperture efficiency and low cross polarization by exciting mainly TE-type modes in appropriate amplitudes and suppressing undesired TM modes in the aperture, although the TM modes can help to excite the TE modes of the correct amplitude.

Excitation of two or more hybrid modes in corrugated horns can also be used to improve aperture efficiency over a narrowband. For example, Thomas [31] showed that if the HE_{11} and HE_{12} hybrid modes were excited, the feed gave an increased efficiency of 12% over a single-mode design. Thomas and Bathkar [32] were able to achieve very low sidelobes with a multimode corrugated horn.

18.2.1.7 Profiled and Optimized Circular Horns The profile of smooth-wall or corrugated horns can be optimized to achieve a desired performance. This is done either by using known information on the effect of conventional profiles, by combining some of these profiles [22, 33], or through direct optimization [34]. The former approach is illustrated by the use of Gaussian/hyperbolic profiles (Eq. (18.28)), which produce a beam with low cross polarization over very wide bandwidths. Direct optimization is achieved using the profile and/or slot depths of each corrugation as parameters. An alternative approach that is more efficient is to fix the slot width-to-pitch ratio and optimize the profile through a spline representation. An example of a horn designed by this method for a Ku-band transmit/receive feed application is shown in Figure 18.16 [34]. Through these means, excellent performance can be realized with very compact structures.

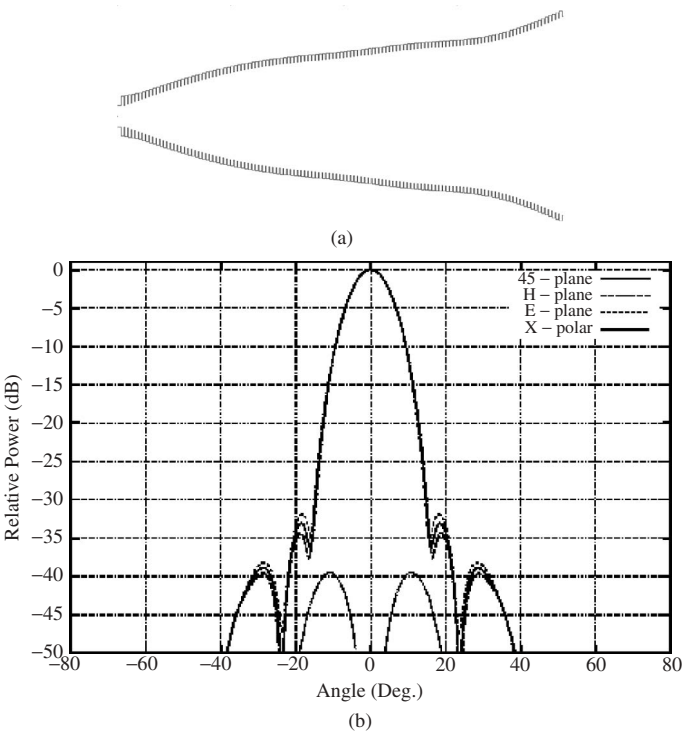


Figure 18.16 Profiled corrugated horn: (a) geometry, and (b) radiation pattern at 11.7 GHz. (Ref. 34.)

18.2.1.8 Dual-Band and Multiband Feeds Some applications require feeds that operate in two or more frequency bands simultaneously. Two basic ways to achieve this are to multiband the feed or to use dichroic reflectors (see Section 18.5.4). A common dual-band feed is a coaxial horn structure where the high frequency band is diverted into an inner (usually) circular horn and the low frequency band is handled by an outer coaxial waveguide. The two bands often share a tapered section that leads to the aperture. The match of the outer coaxial horn is strongly affected by the diameter of the inner pipe and, to keep this at an acceptable level, this waveguide is sometimes loaded with dielectric or has an aperture choke ring. To improve the match and the pattern, the dielectric loading sometimes protrudes into the common horn. The concentric coaxial feed concept can be extended to three or more bands with increasing difficulty for matching and inputting signals.

Multistep Potter-type horn designs can be used for coaxial structures, as demonstrated by James et al. [35] in providing a simultaneous dual-band/dual circularly polarized feed for a Cassegrain reflector antenna. Figure 18.17 shows a cross-sectional view of the complete feed design to provide S/X-band operation and satellite tracking. The inner section is a Potter horn for X-band (including provision of tracking) and the outer S-band system is a coaxial waveguide geometry. In both horns, multiple steps are used in the throat region of the horn to provide a symmetrical low cross-polarization radiation pattern over the $\sim$5% bandwidth of each band. The input to the dual-polarized coaxial horn is through four orthogonal probes (two excited in antiphase to couple the TE_{11} mode) and the X-band horn is input through a compound bandpass/bandstop filter in a circular waveguide, where both irises and ring-loaded slots were utilized in the design to achieve the required performance. Another example of a coaxial feed is the compact dual-band feed of Henderson and Richards [36] suitable for small front-fed parabolic reflectors. This feed is integrated with a diplexer and septum polarizer to provide dual-circular polarization in both bands.

An alternative approach to multiband horn design is shown in Figure 18.18 [37]. This dual-depth corrugated horn (see inset to Figure 18.18) operates in three relatively narrow frequency bands to provide transmit and receive in X-band and receive-only in Ka-band. The X-band receive signal is obtained from a junction of four waveguides that are arranged radially at 90°. A similar junction is used to inject the transmit signal at X-band. Ka-band is extracted from the rear of the corrugated horn. Another example is

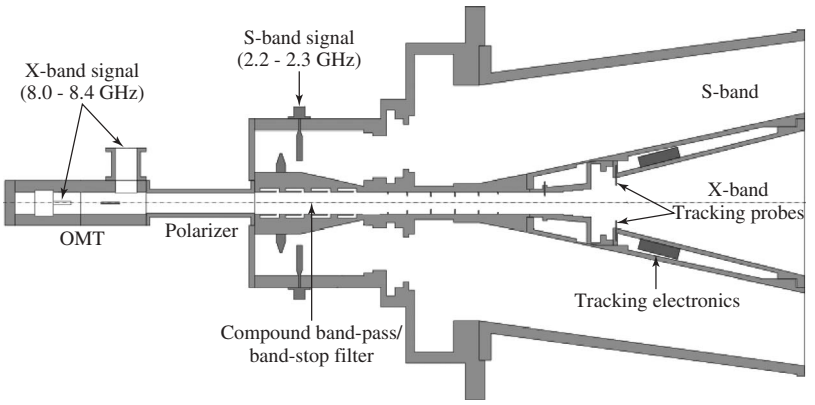


Figure 18.17 Dual-band coaxial horn for S- and X-band operation. (Ref. 35.)

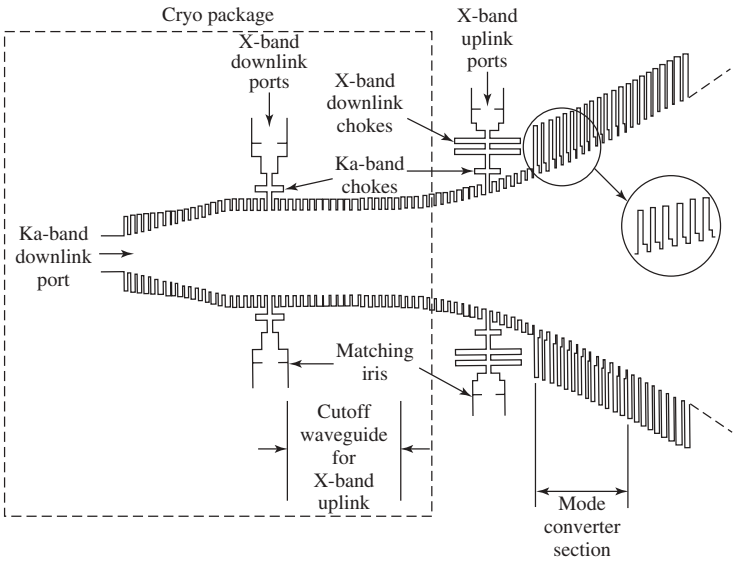


Figure 18.18 A triband X/X/Ka horn. (From Ref. 37, courtesy of IEEE).

given by Green and Smith [38] for a dual-band corrugated horn design where operation is required in Ku/Ka-bands separated by a ratio of 1.5:1.

18.2.1.9 Elliptical Horns Horns with elliptical cross section are used to feed reflectors or lenses that are noncircularly symmetric, particularly when low cross polarization or circular polarization is required. The field in the aperture of an elliptical-shaped horn is tapered in the two principal planes, resulting in lower sidelobes in the E-plane than for a rectangular horn. The dominant modes of the air-filled horn are the TE_{c11} (even) and TE_{s11} (odd) modes, corresponding to the electric field maximum aligned with the minor and the major axes, respectively.

In an elliptical waveguide or horn (Figure 18.19), the fields are expressed in terms of Mathieu functions, which are readily calculated via a computer but which, unlike circular geometries, preclude closed-form expressions in terms of known functions for the radiated fields (Eq. (18.14)). When the flare angle is small ($<10^\circ$), the components of the fields (Eq (18.41a)) for an elliptical aperture radiating in the TE_{c11} mode (polarized

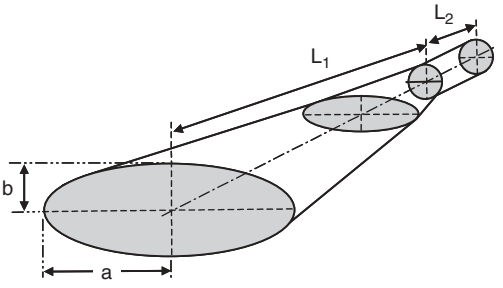


Figure 18.19 Geometry of an elliptical horn.

in the x-direction, Figure 18.14) are given by

$$E_\theta(\theta, \phi) = N_x \cos\phi + N_y \sin\phi \tag{18.29a}$$

and

$$E_\phi(\theta, \phi) = \cos\theta(-N_x \sin\phi + N_y \cos\phi) \tag{18.29b}$$

where

$$N_x = c_0 \int_0^{\xi_0} \int_0^{2\pi} [-Ce_1(\xi)ce_1'(\eta)\cosh\xi \sin\eta + Ce_1'(\xi)ce_1(\eta)\sinh\xi\cos\eta] \times \exp[jkh\sin\theta(\cosh\xi\cos\eta\cos\phi + \sinh\xi\sin\eta\sin\phi)]\sqrt{\cosh^2\xi - \cos^2\eta}\, d\eta\, d\xi \tag{18.30a}$$

$$N_y = c_0 \int_0^{\xi_0} \int_0^{2\pi} [Ce_1(\xi)ce_1'(\eta)\sinh\xi\cos\eta + Ce_1'(\xi)ce_1(\eta)\cosh\xi\sin\eta] \times \exp[jkh\sin\theta(\cosh\xi\cos\eta\cos\phi + \sinh\xi\sin\eta\sin\phi)]\sqrt{\cosh^2\xi - \cos^2\eta}\, d\eta\, d\xi \tag{18.30b}$$

where $h = ae = \sqrt{a^2 - b^2}$, a is the length of the semimajor axis (see Figure 18.9) and c_0 is a constant. The Mathieu functions $ce_1(\eta, q)$ and $Ce_1(\xi, q)$ are the ordinary and modified even type of order 1, respectively, where a prime on the Mathieu functions denotes the first derivative, q is the first zero of $Ce_1'(\xi_0, q) = 0$ and is implied in Eq. (18.30), e is the eccentricity of the aperture, and $\cosh\xi_0 = 1/e$. The aperture mismatch for the TE_{c11} mode is less than -20 dB for $a > \lambda$ providing $e < 0.92$.

The patterns in the two principal planes are generally different, although for some geometries and eccentricities it can be shown that the beam can be made symmetric. For example, a horn of length 5.09λ along which the ellipticity varies, commencing from an elliptical input waveguide having a semimajor axis 0.42λ and eccentricity 0.422, has good beam symmetry for an aperture with semimajor axis 2.65λ when $e \approx 0.67$ [39].

Elliptical corrugated horns have been designed and used in several applications [40, 41], although care is required in the transition region to ensure maximum bandwidth (about 45%), otherwise high-order modes will be excited.

18.2.1.10 Dielectric-Loaded Horns Horns that are partially loaded with a central dielectric cone or core are capable of very wideband operation (Figure 18.20). Between the core of permittivity ε_{r2} and the metal wall is a material of thickness t with a permittivity ε_{r1} close to 1. An estimate of the optimum gap thickness for circular, dielectric-lined horns, which is valid for large apertures, is [42]

$$t \approx \frac{\lambda}{4\pi\sqrt{\varepsilon_{r2} - \varepsilon_{r1}}} \ln\left(\frac{\sqrt{\varepsilon_{r2}} + \sqrt{\varepsilon_{r1}}}{\sqrt{\varepsilon_{r2}} - \sqrt{\varepsilon_{r1}}}\right) \tag{18.31}$$

However, the horn performance is not sensitive to the gap thickness. Dielectric-lined horns support a hybrid mode similar to a corrugated horn [43], but which is not band limited by the slot depth of the corrugations. The potential of the dielectric-loaded horn has been made possible by the advent of accurate design software and also high quality low loss dielectric materials. For example, Clark and James [44] showed that a partially

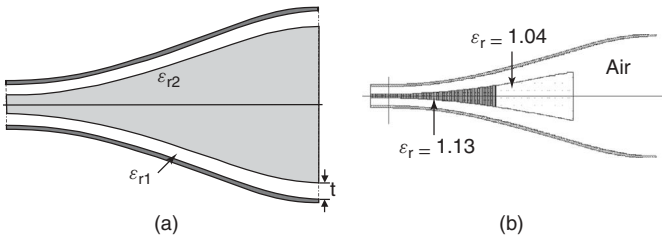


Figure 18.20 Dielectric-lined horns: (a) cone-loaded and (b) profiled ultrawideband.

filled circular dielectric-loaded horn (Figure 18.20b) could have high performance with respect to return loss, cross polarization, pattern symmetry, and phase center stability for a bandwidth ratio of over 30:1. However, as frequency increases, the main frequency dependency is in the general reduction in beamwidth.

Many different types of dielectric-loaded horns have been developed. Some use a solid dielectric with permittivity of about 2.5 and others use foam materials with permittivity close to 1. The wider bandwidth horns use a dielectric permittivity of between 1.1 and 1.2 for the inner core and also a thin air gap.

Other types of dielectric-lined horns have been shown to give excellent performance [45]. An estimate for gap thickness is given by Eq. (18.31) but is often profiled to suit the horn profile. In a rectangular dielectric-lined horn, if the air gap is small, the field in the horn appears approximately uniform in the principal direction, compared with an air-filled guide that has a cosine dependence. The result is a radiation pattern in both principal planes with a sinc-function dependence. The elliptical dielectric-lined horn has lower sidelobes compared to the rectangular horn for the same beamwidth, indicating that the latter has a slightly lower efficiency when used as a feed.

A practical difficulty with the dielectric-loaded horn is obtaining suitable dielectric materials with a permittivity of $\sim$1.2 with an acceptable low loss (loss tangent $< 5 \times 10^{-4}$). One approach is to simulate the required dielectric by forming alternate layers of low loss materials such as PTFE and low loss, low density, expanded polystyrene foam in a sandwich construction. With several layers of foam and PTFE per wavelength, and adjusting the relative thickness of the two materials or cutting holes in the material, a simulated material of the required permittivity with low loss can be achieved [46].

18.2.2 Rectangular Feeds

Rectangular horns that operate predominantly in the fundamental TE_{10} mode[†] radiate fields that have significantly different E- and H-plane patterns. Equation (18.14) and Table 18.2 give the radiated field. The quantity a is the aperture width in the H-plane and b is the height in the E-plane. The different patterns can be an advantage in some array feed applications or where the secondary structure is not axisymmetric, as in a pillbox reflector. The reason for this difference is seen through the TE_{10}-mode aperture fields. The E-plane of this mode's field is almost uniform while in the H-plane the field is cosine tapered. The radiation by single-mode rectangular and pyramidal horns is detailed

[†]The TE_{10} and TE_{01} modes are fundamental modes of a rectangular waveguide with the electric field aligned in the y- and x-directions, respectively. The latter is obtained from the former by rotating the waveguide clockwise by 90°.

in many texts on antennas and is not repeated here [5, 10]. In addition, a design procedure for so-called standard gain horns is provided in these references. We note, however, that although the standard gain horn design maximizes gain for given aperture dimensions, this is usually not a good design strategy for feed antennas because the pattern shape is often unsuitable, and high sidelobes occur in the E-plane causing significant spillover loss.

The radiation pattern of a rectangular horn radiating in the TE_{10} mode is given by the set of universal curves shown in Figure 18.21 as a function of $\Delta = (w/\lambda)\tan(\theta_o/2)$, wherein $w = b$ in the E-plane and w = a in the H-plane. An open-ended rectangular waveguide or small flare-angle horn is an effective feed antenna if the aperture dimensions are suitably chosen. For a symmetrical reflector or lens the aperture dimensions are usually chosen to equalize the E- and H-plane patterns across the main beam, although the cross-polar level in the intercardinal planes remains high. The rectangular horn dimensions for maximum gain with a given reflector F/D_m have been determined by Truman and Balanis [47]. The dimensions are shown in Figure 18.22 for various horn slant lengths (infinity corresponds to uniform waveguide).

The phase center in the two principal planes of a flared rectangular horn is given in Figure 18.12, which was obtained from the patterns over the angular range defined by the −12-dB semiangles in the E- and H-planes. In this figure, L is the distance from the apex to the aperture in each plane. Similar to the conical horn, the phase centers are significantly different in the two planes, except when the aperture size is large and they are close to the horn apex.

The reflection coefficient of the TE_{10} mode at the aperture is given by Eq. (18.18) and the mode self-admittance

$$y_{11} = \frac{2jk^2}{\pi ab\beta}\left(\int_0^{\alpha_0} d\alpha \int_0^{a\sec\alpha} dx + \int_{\alpha_0}^{\pi/2} d\alpha \int_0^{b\,\mathrm{cosec}\,\alpha} dx\right)\exp(-jkx)(b - x\sin\alpha)$$
$$\times \left\{\frac{1}{k_c}\left[1 + \left(\frac{k_c}{k}\right)^2\right]\sin(k_c x\cos\alpha) + \left[1 - \left(\frac{k_c}{k}\right)^2\right]\right.$$
$$\left.\times\,(a - x\cos\alpha)\cos(k_c x\cos\alpha)\right\} \tag{18.32}$$

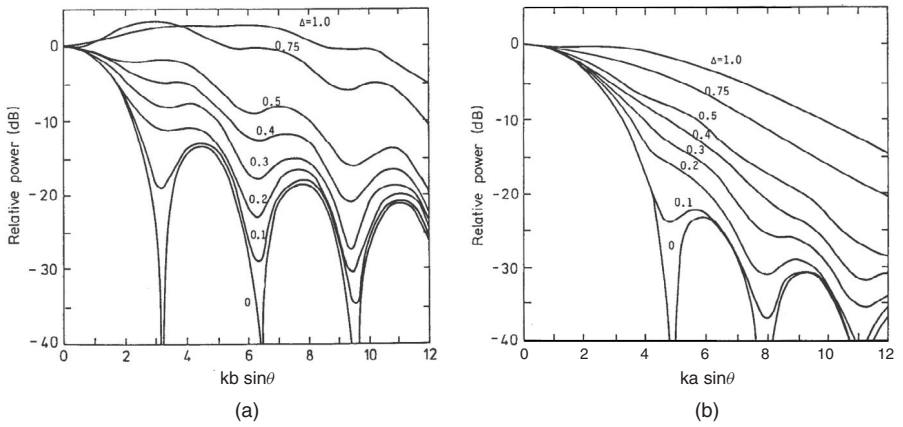


Figure 18.21 Normalized pattern of rectangular horn: (a) E-plane and (b) H-plane.

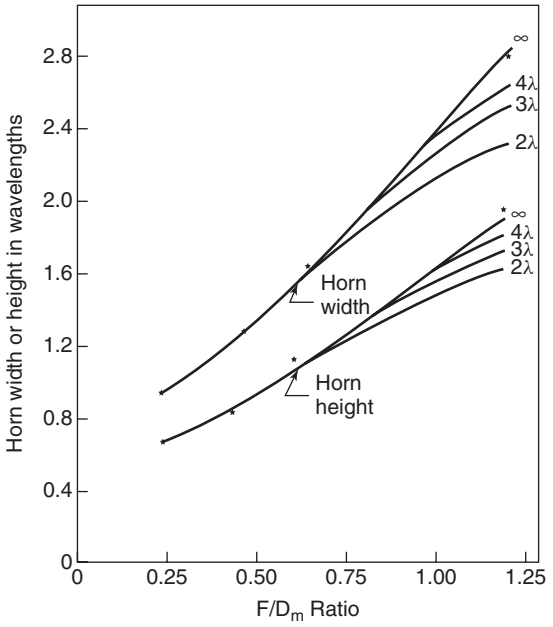


Figure 18.22 Optimum horn dimensions versus reflector F/D_m for various rectangular horn lengths. (From Ref. 47, courtesy of IEEE).

where $k_c = \pi/a$ and $\alpha_0 = a\tan(b/a)$. The aperture mismatch is less than -25 dB for $a > 2\lambda$ and $a < b < 2a$. When the height (b) is small compared with the width (a), as with some sectoral horns, the aperture mismatch can be significant (approximately -20 dB).

18.2.2.1 Multimode Feeds A way to improve the pattern of rectangular horns is to excite one or more modes additional to the TE_{10} mode to make the patterns in the two planes more identical or to reduce sidelobes. Beam shaping in rectangular horns can be achieved in the same way as with circular horns. As described by Cohn [8], changes of flare angle introduced along the length of the horn can be used to excite the required mixture of high-order modes, which must be brought into correct phase at the aperture. In particular, flare-angle changes can generate the TE_{12}/TM_{12} mode pair and relative to the TE_{10} mode the amplitude of the TE_{12}/TM_{12} modes is

$$A_{TE_{12}-TM_{12}} \approx \frac{2}{3} j \frac{(\theta_1 - \theta_2)s}{\lambda} \tag{18.33}$$

providing the flare angle change $|\theta_1 - \theta_2|$ is small ($<25^\circ$), where s is the width of waveguide at the transition. Since the mixture of the TE_{12}/TM_{12} modes leads the phase of the TE_{10} mode by 90°, this phase must be removed with a length of waveguide that gives a 270° phase shift. Despite this complication, especially in designing several flares to avoid violating the above condition, good pattern symmetry can be achieved over moderate bandwidths (about 30%).

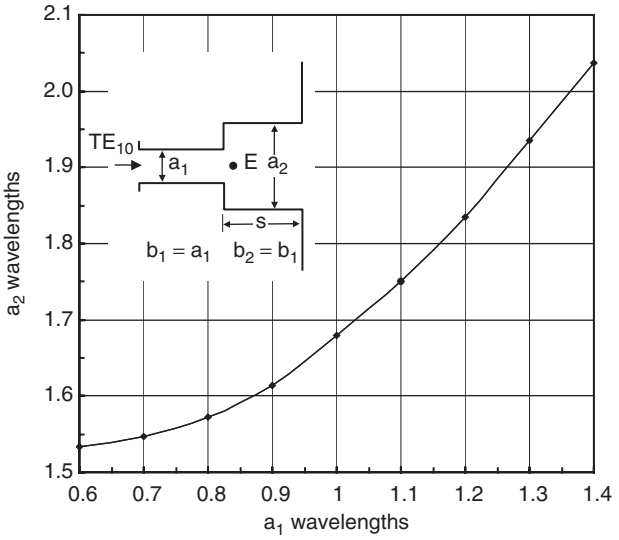


Figure 18.23 Width a_2 in the H-plane required to produce a TE_{30} mode with amplitude one-third that of the TE_{10} mode in the aperture of a dual-mode rectangular horn, which is excited from a square waveguide with side length a_1.

Rectangular horns are often used as feed elements in arrays to produce shaped beams from secondary reflectors for satellite communications. In such cases, the efficient illumination of the reflector is crucial and this can be achieved with multimode horns. Multimode rectangular horns employ TE_{m0} ($m = 1, 3, 5, \ldots$) modes with approximate amplitude of $1/m$ to improve the aperture field uniformity. This results in higher aperture efficiencies than horns with the fundamental mode alone, typically $>90\%$. The aperture dimension in the H-plane of these horns needs to be at least $m\lambda/2$ for efficient excitation and for these modes to propagate. However, significant improvement in performance is achieved by exciting the TE_{30} mode in addition to the fundamental TE_{10} mode [3]. The TE_{30} mode is produced with an H-plane step and Figure 18.23 shows the dimension needed for a step from a square waveguide to produce a TE_{30} mode with an amplitude one-third that of the TE_{10} mode in the output waveguide [48]. This step excites the TE_{30} mode approximately 180° out of phase with the transmitted TE_{10} mode, and a length of waveguide, s, is required to equalize the mode phases at the aperture.

18.2.2.2 Diagonal Horn Another type of dual-mode horn is the diagonal horn due to Love [49, 50]. Two spatially orthogonal modes TE_{10} and TE_{01} are excited with equal amplitude and phase in a square waveguide. This waveguide is then flared into the required aperture dimensions. The principal electric field of the diagonal horn is along one of the diagonals. While principal plane patterns are the same, in the intercardinal plane the copolar pattern is different and the cross polarization is unacceptably high for most dual-polarized applications. The horn has best performance when the aperture size is relatively large and therefore is mainly suitable for long focal length or dual-reflector applications.

18.2.2.3 Dielectric-Loaded Rectangular Horns Another form of dielectric-loaded rectangular horn has a dielectric lining the wall parallel to the E-plane as shown in Figure 18.24. Dielectric loading of feed arrays of rectangular waveguides and pyramidal horns provides an extra degree of freedom in the design of shaped beams for satellite antennas. It improves the illumination efficiency of the array and can be applied to both small and large aperture horns. The aperture field is almost uniform in the H-plane if the dielectric thickness, t, is chosen to be approximately [51]

$$t \approx \frac{\lambda}{4\sqrt{\varepsilon_{r1} - \varepsilon_{r2}}} \tag{18.34}$$

where ε_{r1} and ε_{r2} are the dielectric constants of the regions shown in Figure 18.24 At the center frequency when t is given by Eq. (18.34), the radiation pattern of the dielectric-loaded horn is obtained from Eq. (18.14) wherein $A = B$ and A is given by

$$A(\theta, \phi) = 2bd\ S\left(U\frac{b}{2}\right)\left[S(Vd) + \frac{2t}{\pi d}\frac{\cos\left(V\frac{a}{2}\right) + \frac{2t}{\pi}V\sin(Vd)}{1 - \left(\frac{2t}{\pi}V\right)^2}\right] \tag{18.35}$$

where S, U, and V are defined in Table 18.2 while a, b, and d are defined in Figure 18.24. The factor in the square brackets determines the H-plane pattern and the first term dominates if t/d is small. Consequently, for thin dielectric layers, the E- and H-plane radiation patterns have almost the same shape.

An important property of dielectric-loaded horns is that their gain can be greater than the corresponding unloaded horn. The relative aperture efficiency is

$$\frac{\eta_{\text{load}}}{\eta_{\text{unload}}} = \frac{\pi^2 d^2}{2a(a-t)}\left(1 + \frac{2t}{\pi d}\right)^2\sqrt{1 - \left(\frac{\lambda}{2a}\right)^2} \tag{18.36}$$

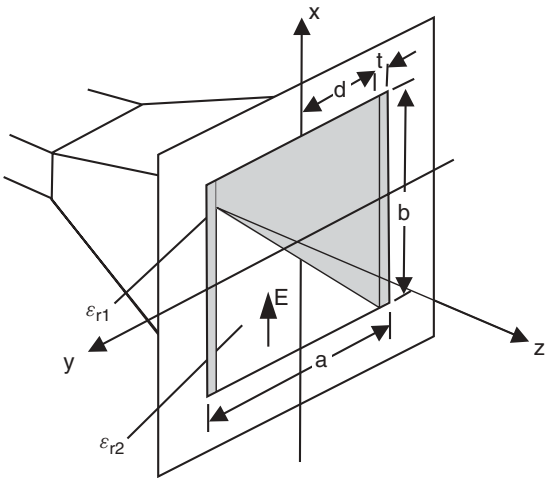


Figure 18.24 Dielectric-lined rectangular horn.

where "load" and "unload" subscripts refer to loaded (i.e., satisfy Eq. (18.34) with $\varepsilon_{r2} = 1$) and unloaded waveguide cases. Equation (18.36) shows that, for most commonly available dielectrics, an increase occurs only if $a > \lambda$. With dielectric loading, the H-plane beamwidth is usually less than in the unloaded case. This is important in array feed applications for satellites because the reflector edge is sometimes overilluminated, and therefore dielectric loading can help improve antenna efficiency. However, ohmic loss in the dielectric decreases the gain, although thin dielectric layers have low loss.

18.2.2.4 Rectangular Corrugated Horns Rectangular corrugated horns have received attention as feeds since the mid-1960s [18, 52]. Corrugations in the two walls parallel to the H-plane can be used to taper the field in the E-plane and thereby reduce the sidelobes in that plane. Horns with corrugations on four faces were described first by Bryant [53] with the aim of producing circularly polarized radiation with low sidelobes [54]. The main advantage of the rectangular corrugated horn is that the pattern can be shaped in both planes for feeding reflectors that are required to give dual-polarized rectangular or elliptical shaped beams [18, 54, 55]. The disadvantages compared with a circular corrugated horn are its comparatively narrow bandwidth, higher cross polarization, difficulty of manufacture, and higher weight. Detailed analysis is required to obtain the accurate mode combination in the aperture for the general case and use of computer packages mentioned in Section 18.7 is highly recommended for design. In brief, the corrugations are chosen to be close to a quarter-wavelength deep near the aperture, with a gradual taper of the initial slots at the throat from about a half-wavelength for a good reflection match. Sometimes pattern control is difficult. This can be improved by filling the corners where the corrugations meet [54, 56]. The bandwidth and poor cross polarization of the rectangular corrugated horn are due to the generation of a higher-mode slow wave, which is predicted in the rigorous analysis of the horn [56]. The bandwidth is typically $<40\%$ and this compares with the bandwidth of a circular corrugated horn of over 70%.

18.2.3 Rear-Radiating Feeds

In prime-focus fed symmetrical reflectors, blockage by the feed and feed support struts causes reduced gain and increased sidelobe levels. The loss of performance due to strut blockage can be eliminated by employing a self-supporting or axial feed antenna. Mechanical support is provided in this case by the transmission line, which extends from the reflector vertex to the feed. Such feeds have a long history and many different types have been reported. Some examples are shown in Figure 18.25.

All rear-radiating feeds depend for their operation on the efficient transition from the feeder line in the axial conductor to the radiator. Usually it is this feature that limits the usable bandwidth to only a few percent of the operating frequency. For example, the match from the resonant dipole disk feed [3] in Figure 18.25a is narrowband and this is even worse for the feed in Figure 18.25b, which incorporates a resonant ring to improve the radiation pattern [57]. In the Cutler feed [4] (Figure 18.25c), two radiating slots are excited by a cavity driven from the waveguide transmission line. A tuning screw in the cavity improves the match, although the design remains essentially narrowband. The cup feed (an example is shown in Figure 18.25d [58]) is distinguished by its axial symmetry, and a number of designs have been tried in order to achieve an effective transition from the circular central conductor to the cup region. An example of a cup feed design is a

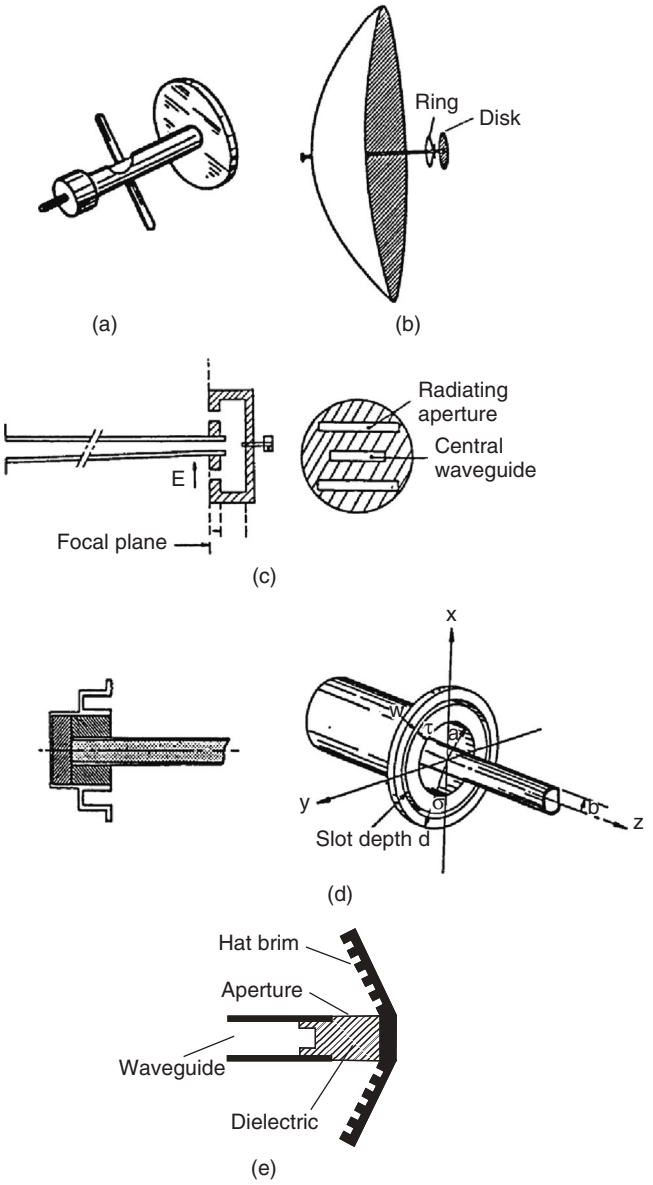


Figure 18.25 Rear-radiating feed antennas: (a) dipole and disk fed by transmission line (adapted from Ref. 3); (b) the same as (a) with ring; (c) rectangular waveguide with radiating slots (adapted from Ref. 4); (d) cup feed [58] (From Ref. 61, courtesy of IEEE); and (e) hat feed 63).

circular waveguide version of the Cutler feed using an annular slot, while another design by Griffin [59] uses coaxial waveguide sections at the end of the cup to improve the match. Both of these designs are relatively narrowband (approximately 5%). However, Schwerdtfeger [60] was able to achieve up to 20% bandwidth by exciting dipoles within the cup with a thin TEM transmission line.

The cup feed developed by Poulton and Bird [58], shown in Figure 18.25d, uses a simple transition from circular waveguide to excite the cup. This produces a bandwidth of about 20% and its symmetry allows dual polarization to be used. The waveguide to cup transition is most effective when the inner-to-outer conductor radius ratio (b/a) is about 0.5. Radiation from the cup is predominantly due to the TE_{11} coaxial waveguide mode in the cup. This radiates in the presence of the supporting central conductor, which has a significant effect when the central conductor diameter is even moderately large. To keep cross-polar levels less than -20 dB, the central conductor radius should be kept relatively small, typically $b/a < 0.3$ and $ka > 2.5$. A cup feed excited with an air-filled circular waveguide has a relatively poor performance with poor pattern symmetry and, as a result, has high levels of cross polarization. Much improved performance is obtained by dielectric loading the central conductor and this also provides mechanical support for the cup. The main disadvantage of dielectric loading is matching into the feeder waveguide. Further improvement in the pattern is possible by adjusting the size of the flange surrounding the aperture and by inclusion of a concentric ring slot. The ring-slot flange considerably improves the radiation performance over a narrow band (typically 5% to 10%) [61].

The "hat" feed of Kildal [62, 63], shown in Figure 18.25e, uses a circular waveguide transmission line, a circumferential slot, and a corrugated reflecting brim. An iris-matching network in the waveguide gives a good match over about a 15% bandwidth. The feed is excited by a propagating TE_{11} mode in the waveguide and this excites radial waveguide modes in the dielectric aperture. The $\lambda/4$-deep corrugations ensure the radiation pattern is symmetrical with little spillover radiation. The beamwidth is relatively broad for a flat brim but this can be reduced by angling the brim as shown in Figure 18.25e.

18.2.4 Microstrip Feeds

Microstrip patch antennas are a possible alternative to horns in feed applications or as a feeder for a horn. The advantages of a microstrip as a feed are its ease of integration with the system electronics, compact size, reduced weight, and lower volume compared with a horn. Multilayer patch antennas can provide a return loss of 10 dB or better over a 50% or more bandwidth. Microstrip feed antennas are used singly or in arrays and can be incorporated in a cavity or a horn to improve the radiation performance for feed applications where low noise temperature is required [64, 65].

Microstrip antennas are suitably shaped printed circuits that are designed to radiate at discontinuities [66, 67]. The most common types are rectangular and circular patches. Two simple models are used to describe these antennas. These are the cavity and transmission-line models. In the cavity model, the antenna, which is excited by a probe or a slot in the ground plane, sets up resonances in the cavity formed between the patch and the ground plane. The antenna radiates through currents set up around the edge of the patch. In the case of a circular patch, for example, the radius of the patch, a, is determined by the desired mode and its resonance frequency. The TM_{11} mode has suitable characteristics for feed antennas and its resonance occurs at the first root of $J_1(k\sqrt{\varepsilon_r}a) = 0$, where J_1 is the first-order Bessel function and ε_r is the dielectric constant of the substrate, which gives $a \approx 0.61\lambda/\sqrt{\varepsilon_r}$. In the transmission-line model of a rectangular microstrip patch antenna, the radiation appears to occur at two rectangular apertures on either side of the patch, and these apertures are connected by a parallel-plate

transmission line, which is the length ℓ of the patch. Now assume a patch of width w lies in the yz-plane on a substrate of thickness h. The radiated fields are given by Eq. (18.14) wherein A is approximately

$$A(\theta, \phi) = wh \; S\left(\frac{kw}{2} \sin\theta \sin\phi\right) \cos\left(\frac{k\sqrt{\varepsilon_r}\ell}{2} \cos\theta\right) \tag{18.37}$$

where the first function on the right side of Eq. (18.37) is defined in Table 18.2 and $A(\theta, \phi) = B(\theta, \phi)$. The first factor in Eq. (18.37) is due to the radiation pattern of the $w \times h$ apertures at the ends of the patch while the second function is a consequence of the path length between these apertures. In this geometry, the E-plane is the $\phi = 0$ plane and the H-plane corresponds to $\theta = 90^\circ$. Further discussion of the design of microstrip antennas is beyond the scope of this chapter and for further information the reader should consult specialized texts [10, 66, 67].

The patterns given by Eq. (18.14) and (18.37) are asymmetric and produce high cross polarization. By designing the ground plane size (typically a diameter of about 0.8λ) and incorporating a choke at the edge of the ground plane, the pattern symmetry, and hence cross-polar performance, can be improved considerably over about a 5% to 10% bandwidth. The chokes also assist to reduce the spillover energy and rear-lobes, which is important in low noise applications. An example of a two-layer microstrip patch feed is shown in Figure 18.26 [64, 65].

18.2.5 Short-Backfire Antenna

A useful narrowband feed is the short-backfire antenna, particularly in low frequency applications when compactness and light weight are important [68]. It has a relatively directive pattern, low sidelobes, and low rear lobes. The short-backfire antenna consists of two parallel-plate reflectors, spaced about $\lambda/2$ apart, and a driven element, an approximately $\lambda/2$ dipole shown in the example in Figure 18.27, that is placed about midway between the reflectors. In this arrangement the backfire antenna is basically a modified endfire array. One of the reflectors is small and the energy concentrated between the reflectors radiates through this reflector. The larger rear reflector has a skirt of about $\lambda/4$ to reduce the sidelobes. Due to the resonant structure, the antenna has only about a 3% bandwidth. Rectangular waveguide versions of this antenna have also performed successfully.

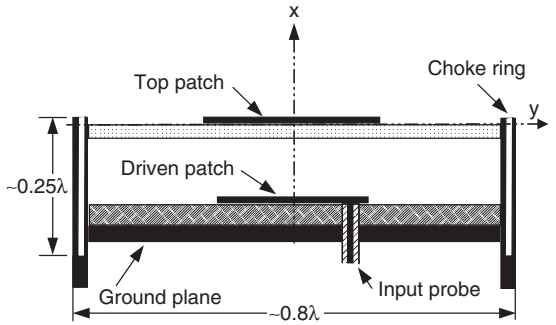


Figure 18.26 Two-layer circular microstrip antenna in a cavity with a choke ring.

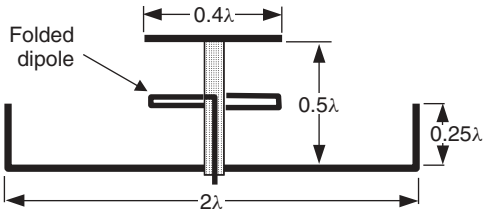


Figure 18.27 Geometry of a circular short-backfire antenna that is fed by a folded dipole.

18.3 LINEAR FEED ANTENNAS

Historically, one of the earliest feeds for a reflector was a half-wave dipole, which was fed from a transmission line via a balun. The dipole is backed by a circular disk set about $\lambda/4$ away to produce a unidirectional radiation pattern (see Figure 18.25a) [3, Chap. 8]. The disk diameter is from 0.5λ to 1λ, depending on the required edge taper. Alternative to the circular disk, one or more parasitic elements can achieve a similar effect. This type of feed is relatively narrowband.

The extension of the parasitic array is the Yagi–Uda array and the log-periodic antennas (Figure 18.28) [5, 50, 69–70]. These are often used as feeds especially at low frequencies for practical considerations, such as low weight, and also because the latter can operate over relatively wide bandwidths. Other practical advantages are that they are low profile, which minimizes blockage, and they have relatively simple input requirements. A major disadvantage in wideband applications is the defocusing caused by axial phase center movement with frequency. The optimum feed design for maximum antenna gain changes with the angle subtended by the secondary antenna (i.e., ψ_c given in Eq. (18.2) or (18.3)). In narrowband applications, the Yagi–Uda or log-periodic is placed at the focal point of the secondary antenna and defocusing loss is negligible.

The Yagi–Uda array consists of a center-driven dipole element with a number of parasitic elements arranged in the same plane. These parasitic elements increase the gain of the antenna over that of a single dipole. A longer parasitic element is a reflector and is opposite the main radiation direction. The shorter elements lying in the direction of the main radiation are called directors. Varying the length and spacing of the elements relative to the driven element optimizes the radiation pattern. The input match is sensitive to the first reflector element and usually only one or possibly two reflectors are needed to control the rear radiation and hence the pattern front-to-back ratio. The typical dimensions of a Yagi–Uda feed for a reflector with $F/D_m = 0.4$ are given in Figure 18.28a. The classical Yagi–Uda array has a limited bandwidth of about 2% and the directivity increases approximately with the number of directors.

The bandwidth of a linear antenna may be increased significantly by means of a log-periodic geometry that feeds all elements in the array instead of a single dipole (Figure 18.28b). The log-periodic antenna consists of dipole elements arranged in a plane that have a constant ratio (τ) between the element's distance from the apex (x_n) and the dipole lengths (l_n). Successive dipole elements are excited by reversing the phase. The pattern depends on the apex angle (α) and element spacing (d_n) (Figure 18.28b), usually expressed in terms of the parameter $\sigma = d_n/2l_n$. Higher gains occur with large τ and σ and small α. For broadband designs, log-periodic antennas with large α are favored, as

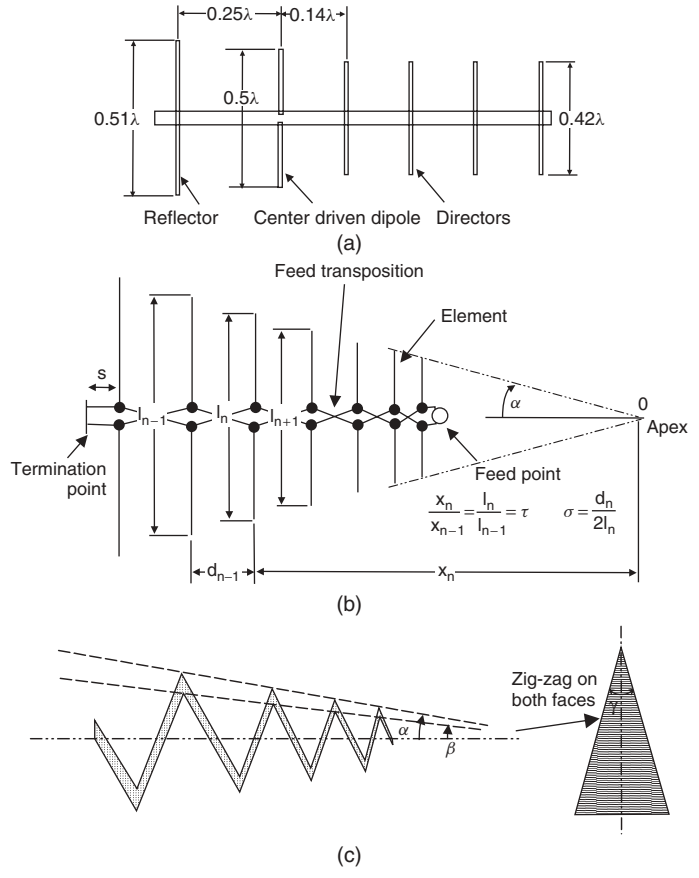


Figure 18.28 Linear feed antennas: (a) Yagi–Uda, (b) log-periodic, and (c) zigzag.

the phase center is closer to the apex and suffers less from defocusing loss. Details of the optimum design of log-periodic feeds are given by Imbriale [71].

The tapered zigzag antenna (Figure 18.28c) has similar properties to the classical linear log-periodic and can be more convenient to fabricate. It has found use as a wideband feed for a reflector with F/D_m ratios from about 0.4 to 0.5. To ensure correct scaling is achieved, the wire is replaced with a plane metal sheet that tapers along the antenna. As well as the parameter τ, the design of the log-periodic zigzag antenna depends on the two angles α and β shown in Figure 18.28c as described by Lo and Lee [70, Chap. 9]. Of importance in feed design is that the E- and H-plane half-power beamwidths are similar when $\gamma = 2\alpha$.

18.4 TRAVELING-WAVE FEEDS

Traveling-wave antennas of the surface-wave type support a wave propagating along the antenna structure with a phase velocity $v < c$, producing an endfire beam. The distinctive feature of this type of antenna is that the directivity is proportional to the

length. With traveling-wave antennas, the maximum directivity is achieved by meeting the Hansen–Woodyard condition [5], which requires that the phase difference between the surface wave and the free-space wave be 180° at the termination of the antenna. Under this condition, the gain is $10\log(10L/\lambda)$ dB and the beamwidth is $55\sqrt{\lambda/L}$ degrees, where L is the length. Other excitation conditions give lower gain and sidelobe levels that are often more suitable for feed antennas.

18.4.1 Dielectric Rod

The dielectric rod antenna is excited frequently with an open-ended waveguide and has the property of a single lobe along the axis of the antenna. The wave traveling along the rod consists of hybrid modes whose fields exist inside and outside the dielectric. To improve the match and prevent standing waves from occurring, the rod is often tapered either along its length or near its end. The radiation from a waveguide-fed tapered dielectric rod has three main contributions: radiation from the waveguide–rod interface, along the rod, and from the end of the rod. The choice of rod material is a compromise between size, directivity, and bandwidth. Both circular and rectangular dielectric rod antennas are used [71–73]. Rods can be used in arrays and for ease of integration with electronics and printed launchers have proved effective over a modest bandwidth (Figure 18.29a) [74]. Excitation efficiencies approaching 100% are possible over a 20% bandwidth by realizing a magnetic ring current equivalent to a ring diameter equal to 0.6 of the rod diameter.

18.4.2 Traveling-Wave Slot

This type of antenna consists of a tapered slotline cut in a thin layer of metal, which is usually supported by a thin dielectric substrate on one side (Figure 18.29b) [75, 76]. The slotline is narrowed at one end to efficiently couple to the feeding network and is tapered outward at the other end to improve radiation effectiveness. A traveling wave excited by the feed network propagates in the slot and radiates in an endfire direction at the tapered end. The energy in the traveling wave is tightly bound to the slotline when the separation distance is small compared to the free-space wavelength. An electric field is created across the slot, which produces a linearly polarized beam in the plane of the antenna (E-plane). As the slotline width increases, radiation occurs most efficiently when the slotline width is about a half-wavelength. Tapered slot antennas have been found to have very wide bandwidths and the ability to generate a symmetric beam pattern despite their planar geometry.

The Vivaldi antenna (Figure 18.29b) is one of a class of slot traveling-wave antennas and is characterized by an exponential curve for the taper in the slotline [77]. It belongs also to the class of aperiodic continuously scaled antenna structures that have the property of a theoretically infinite bandwidth. The Vivaldi antenna can produce a symmetric endfire beam (in both the E- and H-planes) with moderately high gain and low sidelobes. Figure 18.29b shows the basic geometry of the Vivaldi antenna and how the exponential taper is defined. Vivaldi antennas lend themselves well to integration with different types of circuits due to their planar construction. The radiating portion of the antenna is well separated and isolated from the nonradiating portion, allowing circuits to be integrated onto the dielectric at the nontapered end of the slot. It is common for the tapered slot to continue into a slotline circuit, which can then be fed by a microstrip/stripline to slotline transition or a finline. There are many variations of the Vivaldi antenna, including antipodal [78] and balanced configurations [79], which have, respectively, advantages of

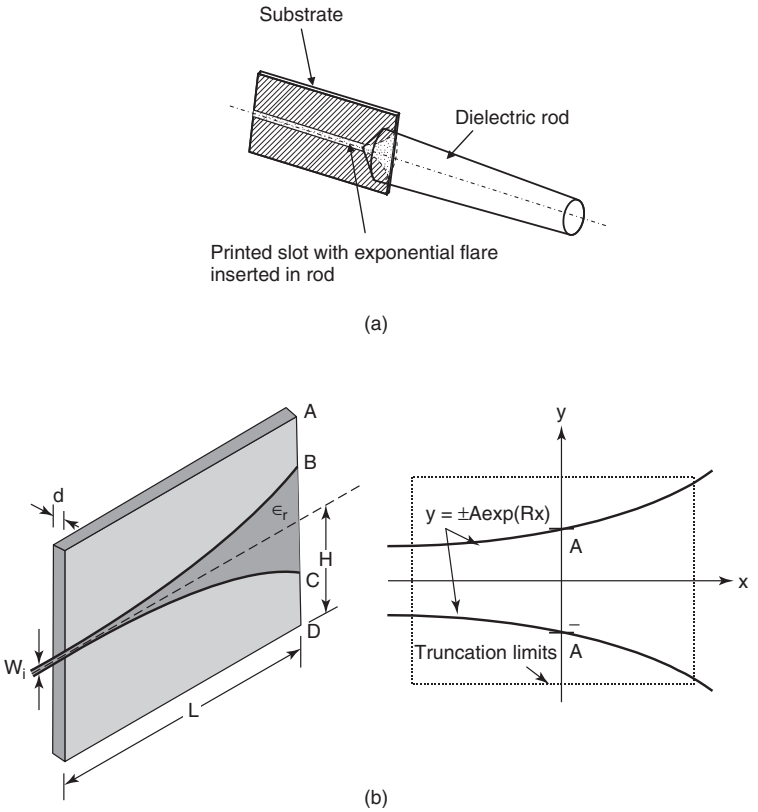


Figure 18.29 Traveling-wave feed antennas: (a) dielectric rod (from Ref. 74) and (b) Vivaldi antenna.

better feeding and lower cross polarization than the conventional Vivaldi. Arrays of Vivaldi antennas are used also as feeds for reflectors and lenses (Figure 18.30).

18.4.3 Line-Source Feeds

Some secondary antennas, unlike the parabolic reflector, do not have a unique focus. Examples of such antennas are spherical and cylindrical reflectors. In a spherical reflector (Figure 18.2b), only the axial ray has a unique point focus on the axis and the remainder suffer from spherical aberration. There are several ways to correct for spherical aberration and one approach is to use a phased line feed, which adds the incident radiation in correct phase at a single point. Rumsey [80] detailed the condition for exciting a spherical reflector for maximum gain. This is a conjugate field match condition (Eq. (18.1)) where the feed produces a tangential electric field at the surface of a mathematical cylinder that just encloses it, equal to the conjugate of the field that would be produced by the reflector currents if the reflector were illuminated by a plane wave.

Line feeds can be divided into two main types: those that use waveguides and those that use cables for phasing. A waveguide can support phase velocities higher than the velocity of light and this gives the required aberration correction. Examples include a channel waveguide, dielectric rod antennas, a combination of the previous two antennas

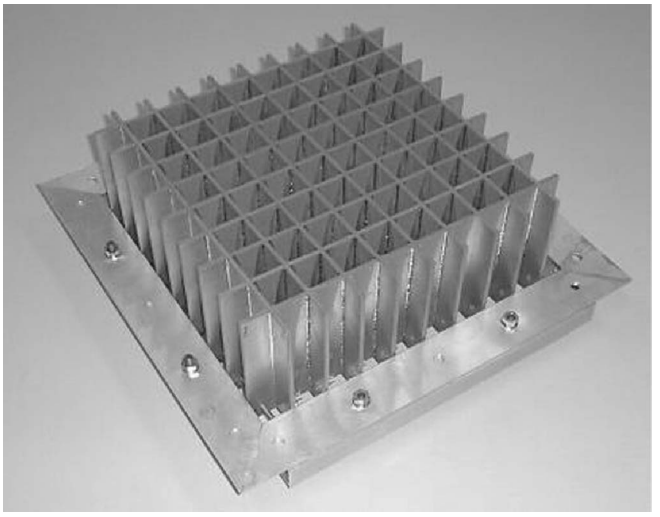


Figure 18.30 Dense-packed array of Vivaldi antenna elements for a focal plane array (courtesy of ASTRON).

as described by Love [81], and a slotted waveguide. In the latter, varying the slot length while maintaining a fixed slot width and constant slot spacing throughout controls the rate of radiation. Using the TE_{10} mode in a waveguide of rectangular cross section enables the phase velocity to be controlled by varying the cutoff wavelength, which depends on the wide dimension of the waveguide. The second approach is to use a linear array of radiating elements such as dipoles, microstrip patches, or Vivaldi elements and achieve the correct phasing by adjusting the transmission-line lengths between the elements. The main limitation is that aberration correction can be maintained only over a very narrow bandwidth with conventional transmission lines. When the frequency is varied, phase errors arise and the gain of the antenna is reduced. This problem may be overcome by means of a correcting subreflector and a point feed in which the spherical aberration is corrected by introducing a correcting path length in air.

An example of a circularly polarized line feed was a feed developed for the 300-m diameter spherical reflector antenna at Arecibo [82, 83]. This feed used slotted circular cylinders (Figure 18.31a) with circumferential fins between each slot, which acted as radial waveguides to provide the correct phasing between the two orthogonal fields within the fins. The slots and fins provided also a radiation pattern with the correct azimuthal dependence. To improve the illumination, a few of the upper feed elements of the azimuthally directed slot fields were closed off to compensate for the finite length and the spacing between the upper elements was reduced to lower spillover.

Another feed for spherical reflectors is the sectoral hoghorn shown in Figure 18.31b [84]. This consists of a primary feed and a section of subreflector surface, which are enclosed between parallel plates spaced $<\lambda/2$ apart. The aperture of the hoghorn is placed parallel to the incident rays in the receiving case. Energy from the spherical reflector is incident on the aperture of the hoghorn and an appropriate path length is included in the parallel-plate region to produce in-phase addition of the incident energy at the primary feed.

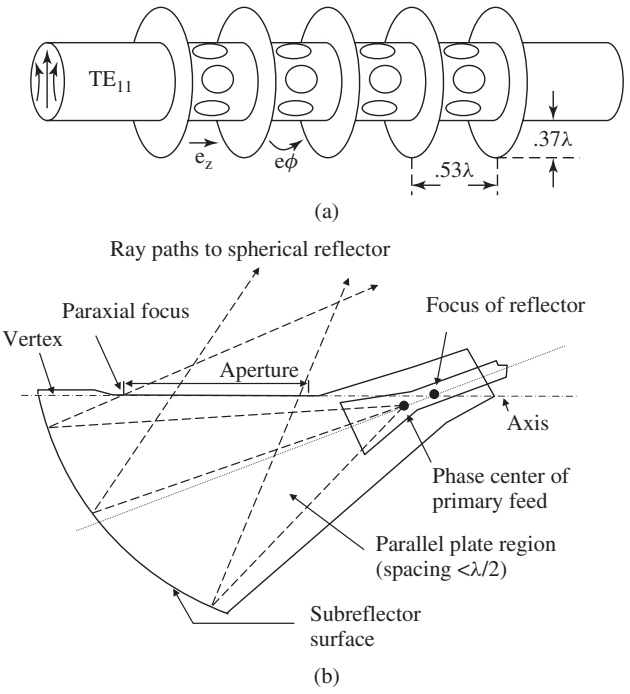


Figure 18.31 Line-source feeds: (a) slotted cylinder (from Ref. 82, courtesy of IEEE) and (b) sectoral hoghorn (from Ref. 84).

A parabolic cylinder reflector antenna, which is a cylindrical reflector with a parabolic cross section, has a line feed along the focal line of the reflector. An example of a feed for this antenna is a linear array of crossed dipoles. The dipoles are located parallel and normal to the focal line (i.e., longitudinal and transverse polarization). The two polarizations are fed separately through two completely branched transmission-line systems, exciting each dipole with identical amplitude and phase. The beam may be steered by adjusting the phasing of the elements. Another suitable feed for a parabolic cylinder is a pillbox horn [3]. These are usually single polarization although two polarizations can be obtained by stacking two orthogonally polarized horns.

18.5 COMPOUND FEED ANTENNAS

Feeds for secondary antennas are sometimes made from a combination of the other types of feeds, reflectors, or resonant structures. The most common types of compound feeds are the array, beam waveguide, splash-plate, and dichroic reflector.

18.5.1 Array Feeds

Single antenna feeds can be combined in arrays to create properties that are unattainable with more elementary antennas. Arrays of feeds are used for a variety of applications including the creation of shaped beams, multiple pencil beams, beam scanning over a limited region of space, or correction of deficiencies in the reflector system arising from

surface errors or misalignments. Single feeds or clusters are used to excite multiple beams. Overlapping clusters can produce beams closer together than arrays of single feeds. In pencil beam applications, where the array essentially compensates for the field distortion through the reflector, a natural design technique is conjugate-field matching (Eq. (18.1)). In this approach, the network connected to the array input excites the feeds so that the field in the aperture is the complex conjugate of the fields in the focal region that would be produced by a beam from the required direction. If there is more than one beam, then the array excitation is the conjugate of the superposition of focal fields of these beams.

The most common elements of feed arrays are waveguide horns, typically circular or rectangular cross section (see Figures 18.15 and 18.32). The circular geometry is convenient for hexagonal array arrangements, while rectangular elements have advantages for producing shaped beams from satellites. Other types of array elements used for array feeds are tapered slot, dielectric rod, Vivaldi, and microstrip patches.

Array feeds can also be phased to allow beam scanning over a limited angular range, the limit being the creation of undesirable grating lobes. If the spacing between elements is s, the maximum angular range the beam can be scanned before grating lobes appear is given by $\sin\theta_{\max} = \lambda/2s$. In typical practical arrays, $s \sim \lambda$ due to the physical size of the elements, and therefore the scan range of the feed is less than 30°. Grating lobes may not be of concern in some feeding applications, providing the associated lower efficiency is acceptable; these lobes do not illuminate part of the secondary antenna or a noise source. Close spacing ($s \leq \lambda/2$) is needed in wide-scan angle applications. The Vivaldi element allows close packing and therefore has good scan capability. Coaxial waveguides operating in the TE_{11} mode (see Table 18.2) can also be used in arrays to achieve closer packing than circular horns in applications requiring moderate bandwidths [85]. In closely packed arrays of TE_{11}-mode coaxial apertures with inner and outer conductor radii b and a, respectively, the array spacing s (in wavelengths) should satisfy

$$\frac{s}{2\lambda} \geq \frac{a}{\lambda} \geq \frac{1}{\pi}[1 + (b/a)] \tag{18.38}$$

For example, with $b/a > 0.6$ a spacing of $s < \lambda/2$ can be maintained over a bandwidth $> 20\%$.

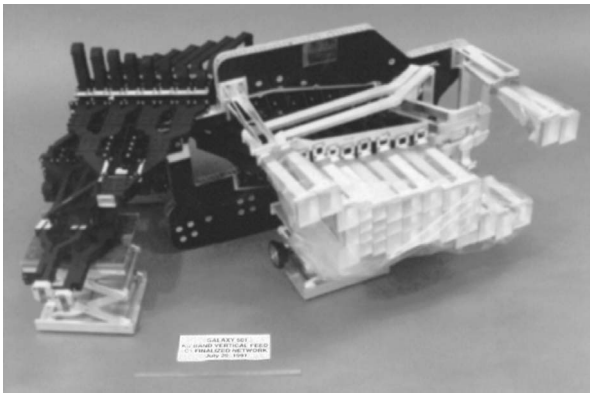


Figure 18.32 Array feed and beamforming network for the Galaxy IV satellite (courtesy of Hughes Aircraft Co.).

Another concern is the design of the individual elements in the array. For example, Amitay and Gans [86] showed that blind spots can exist due to forced aperture resonances in arrays of tapered rectangular horn elements with oversized (overmoded) apertures. The dominant contributor to these blind spots is the resonance of the equivalent transverse magnetic waveguide mode (TM_{12}). By appropriate design of the taper these resonances can be avoided.

Finally, the effect of mutual coupling between the elements should be included in an array design. Mutual coupling affects both the input match and the radiated fields, particularly cross polarization. Mutual coupling in finite arrays has been studied in detail for many different horn types, including rectangular [47, 87], circular [88], and coaxial [89] apertures and also tapered-slot arrays [90]. The effect can be represented by a scattering matrix that is combined with the scattering matrices of the different elements comprising the array. The results obtained from a multimode analysis of mutual coupling are very accurate, and computer programs based on mode matching methods have been widely used in satellite antenna and radio telescope feed array design (e.g., see Refs. 48 and 85). In large arrays, element patterns and array performance can be predicted with success using the Floquet periodic boundary conditions in combination with finite-element or finite-difference time-domain numerical methods (see Section 18.7).

18.5.2 Beam Waveguide

A beam waveguide feed uses a sequence of focusing elements to translate a beam pattern from a conventional feed horn to the secondary antenna. An example is shown in Figure 18.33 for the NASA Deep Space Network [91]. The focusing elements may be reflectors or lenses, and sometimes polarizing grids are used to combine or separate two beams with orthogonal polarizations. The design technique commonly used is based on Gaussian beam propagation [92, 93]. In this approach, the field radiated by the feed horn, usually a transversely corrugated horn, is represented by a sum of Gaussian beam modes. These modes have the characteristic that the amplitude has a Gaussian function dependence with distance from the axis of propagation. The distance from the axis to the point where the Gaussian beam has decayed to $1/e$ of its value is called the beam waist (w) and this is a function of distance (z) along the beam. Thus

$$w = w_0 \sqrt{1 + \left(\frac{\lambda z}{\pi w_0^2} \right)^2} \tag{18.39}$$

where w_0 is the minimum beam waist radius, which occurs where the radius of curvature of the wavefront is infinite, corresponding to a plane wave. The path through the sequence of reflectors, lenses, and grids can be represented step-by-step as a series of spreading transverse Gaussian beams. Each transformation from input to output beam through the focusing element can be represented by an ABCD matrix and the complete system can be found by concatenating these matrices. The ABCD matrices of different focusing elements is described by Kitsuregawa [92] and also by Goldsmith [93], who also give many more details than there is space to record here. Ultimately, the method leads to output Gaussian beams that can be summed to give the field emanating from the beam waveguide. Kitsuregawa [92] provides a design procedure from feed horn to output reflector and design equations. If the focusing elements are moderate in size (say, greater

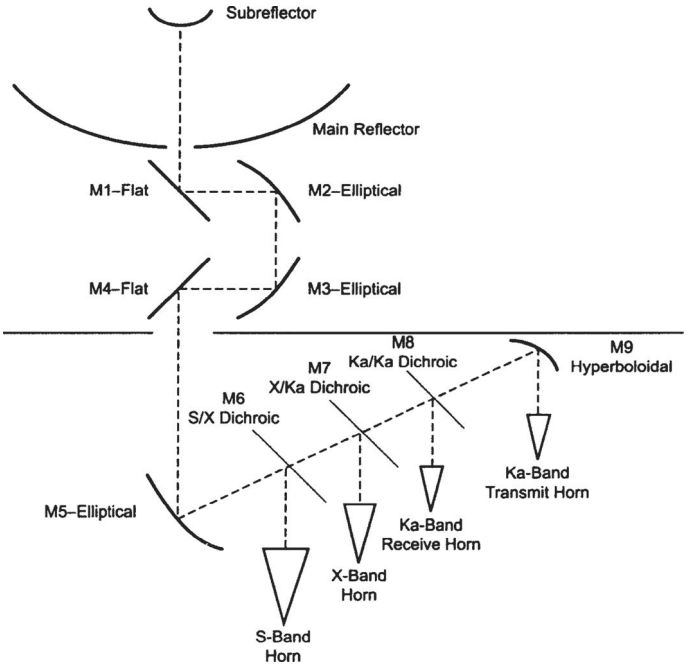


Figure 18.33 Geometry of beam waveguide system for the NASA Deep Space Network antenna DSS-25. (From Ref. 91.)

than 10 wavelengths, the minimum size for a beam waveguide element), it is prudent to calculate the effect of diffraction through the system using a more rigorous analysis method than Gaussian beams before completing the system design.

18.5.3 Splash-Plate Feeds

A splash-plate feed is a version of a self-supporting feed (Figure 18.34). An open-ended waveguide irradiates a small reflector, the "splash plate," that is usually located a short distance s away. The splash plate then illuminates the secondary antenna, commonly a parabolic reflector. From image theory, the radiation pattern is approximately equal to that of the waveguide situated a distance s behind the plate and directed at the reflector, plus the edge diffraction field due to the rim of the splash plate [94]. Struts or a dielectric insert in the waveguide are often used to support the plate, which is usually about a wavelength in diameter a knob or cone (see Figure 18.34) is often fitted at the center to minimize the reflections back into the feed. The reflection coefficient in the waveguide is approximately given by [12]

$$\Gamma \approx \Gamma_{\text{wg}} + \frac{G_{\text{wg}}\lambda}{8\pi s}\left[\frac{R_1 R_2}{(R_1+s)(R_2+s)}\right]^{1/2} e^{-j(2ks+\theta_{\text{wg}})} \tag{18.40}$$

where Γ_{wg} is the reflection coefficient of the waveguide alone, s is the distance from the waveguide phase center to the splash plate, R_1 and R_2 are the principal radii of curvature at the center of the splash plate, G_{wg} is the maximum gain of the waveguide, and θ_{wg} is

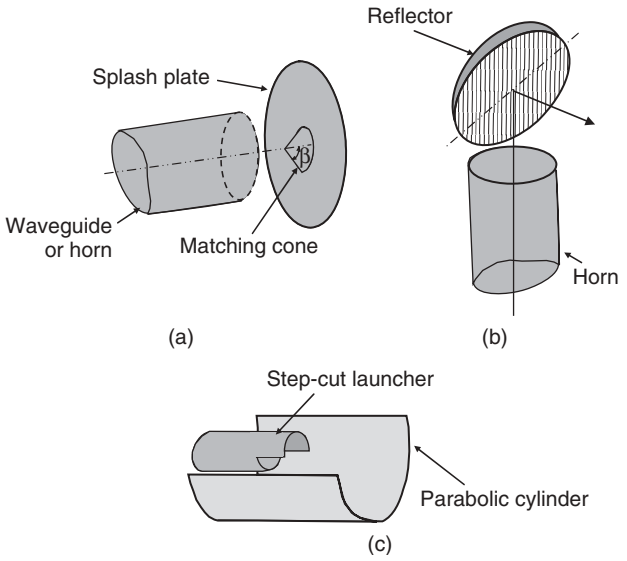


Figure 18.34 Splash-plate feeds: (a) vertex plate with matching cone, (b) periscope, and (c) step-cut Vlasov mode converter.

a phase constant that depends on the reference plane in the waveguide; $\theta_{wg} = 0$ at the waveguide aperture.

The periscope feed in Figure 18.34b is used to convert the beam shape to provide a more appropriate edge illumination for a secondary antenna. A third type of splash plate, the Vlasov mode converter, shown in Figure 18.34c, converts zero-azimuthal index modes from high power microwave and millimeter-wave sources into TE_{11} or HE_{11} modes for use in plasma heating and conventional antenna systems.

18.5.4 Dichroic Reflector

In the usual arrangement, a dichroic reflector (or frequency-selective surface) consists of an array of resonant slots or patches on a surface that is illuminated by signals in two frequency bands (see Figure 18.35). The resonant elements are designed to reflect the fields of one feed in one frequency band and to transmit the fields of the second feed in the second band. In this way, a dual-band feed can be designed for a secondary reflector. The resonant patch approach uses a grid of crossed dipoles, resonant rings, concentric loops, or dumbbell-shaped patches printed on a dielectric surface, which has the profile of the desired subreflector for the primary reflector [93]. These patches resonate at the frequency of the feed at the secondary focus of the surface, usually the higher of the two bands. The second feed is placed in the prime focus of the primary reflector, and at low frequencies the grid of patches and the dielectric is almost transparent. The complementary approach to resonant patches is to use a metallic reflector with resonant slots or perforations. Here the reflector reflects the signals at frequencies below the cutoff frequency of the slots and is transparent above this cutoff frequency. More frequency bands can be accommodated using more than one dichroic reflector as shown in Figure 18.33.

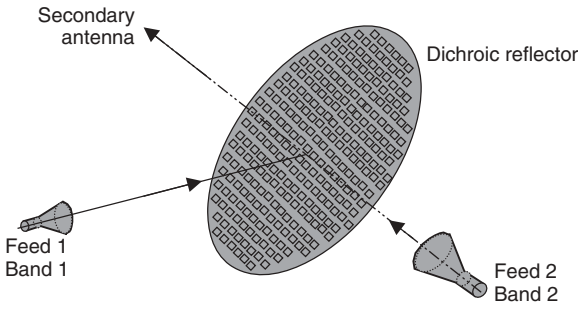


Figure 18.35 Dichroic reflector.

18.6 FEED SYSTEMS

Before the feed antenna is usually a circuit consisting of waveguide, printed circuits, finline, or coaxial cables and the design of these is a separate topic in itself [95]. However, an overview of feed antennas would be incomplete without some mention of the circuit following the feed, which is commonly called the feed system. There are several major types of feed systems and these depend on the application. Three types of feed system will be briefly mentioned—namely, diplexers, beamforming networks, and comparator networks for monopulse tracking.

18.6.1 Diplexers

After the feed antenna, the orthomode transducer (OMT) and (for circular polarization) a polarizer are basic components of the feed system. Uher et al. [95] detail design techniques for waveguide components used in feed systems. To date, the maximum continuous bandwidth achieved by both polarizers and OMTs is about 50% and this limits the feed system configurations possible when considering multiband operation over wide frequency ranges. The means of diplexing several frequency bands can complicate considerably the feed system design. One of the "simpler" diplexing techniques is to use a series of dichroic plates to deflect selectively a given band to an appropriate feed system (Figure 18.33). This type of diplexing can be designed using Gaussian beam optics analysis [93].

For dual-frequency operation a number of feed system configurations are possible. If the individual bandwidths are small (typically less than 10%) and the overall bandwidth ratio to cover the two bands is less than ~40%, then diplexing can take place at the OMT ports before the signals are combined through the OMT into a common waveguide and through the polarizer. In that case we need consider only a single polarized signal in the diplexer and associated filters. These are usually built in a rectangular waveguide to maintain polarization purity. A feed system for reception of dual linear and circularly polarized signals is shown in Figure 18.36.

For larger bandwidth requirements, it has been necessary to develop more sophisticated feed configuration schemes. An example of a feed system for both circular and linear polarization is shown in Figure 18.37. The feed is a very wideband dielectric-loaded horn. The lower frequency band has a bandwidth of 24% while the upper band has a bandwidth of 19%, with an interband separation ratio of ~3.7:1 [96]. The bands of this

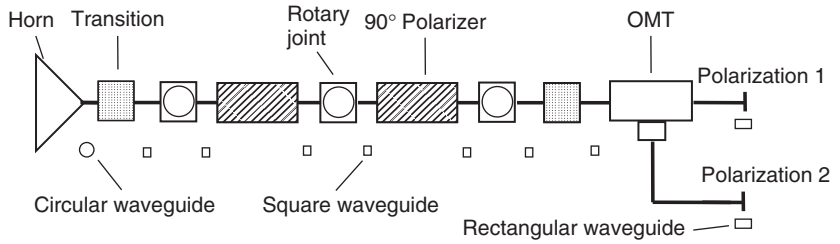


Figure 18.36 Schematic of a feed system for a dual linear/circular receive-only antenna.

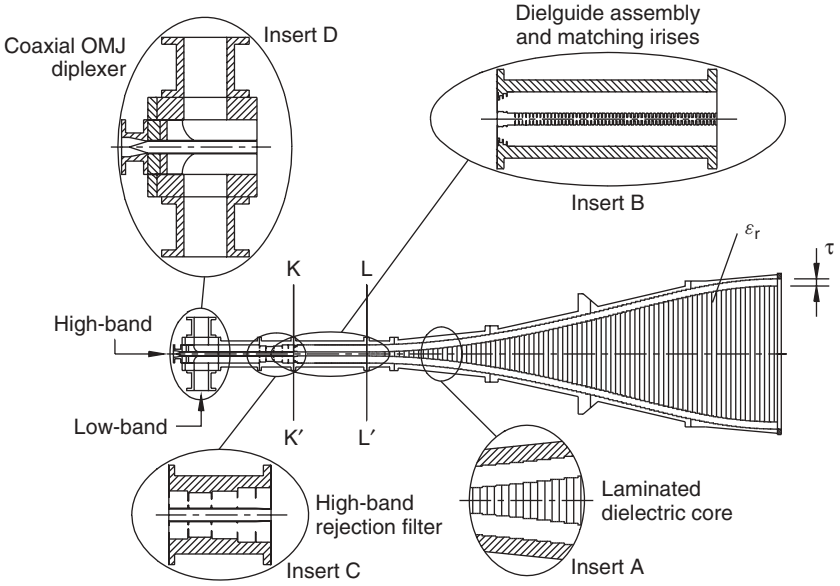


Figure 18.37 Dielectric cone-loaded feed system for dual-band applications. (From Ref. 96, courtesy of IEEE.)

feed system are separated in a coaxial orthomode-junction diplexer, where the center core of the coaxial waveguide is a loaded circular waveguide, which uses a low loss dielectric with a permittivity of 2.4. This coaxial junction is matched to the horn with a dielectric of 1.15 permittivity by a dielguide section. The horn, dielguide section through to the coaxial junction, and beyond the high-band rejection filter in the coaxial waveguide were completely analyzed by mode-matching methods, which was essential to obtain a workable design.

18.6.2 Beamforming Networks

A beamforming network (BFN) is required for an array feed to produce either a fixed illumination, as for the case of a shaped beam for a satellite, or variable illumination, if the feed is phased for radar systems. In satellite applications, the standard approach for producing contoured beams is with a parabolic reflector and an array feed of which the

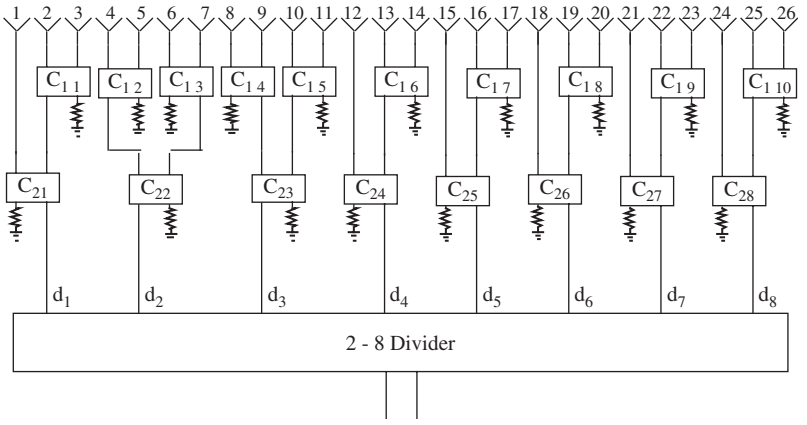


Figure 18.38 Beamforming network layout for the vertical transmit beam [98]. C_{mn} is the coupling value.

amplitude and phase at the input of each feed is supplied by a BFN, which must be low loss, low volume, and lightweight [97]. Typically, for Ku- and Ka-bands these networks are realized in waveguide technology and for C-band and lower frequencies in TEM line. The BFN may need to be reconfigured into a number of fixed beams for different satellite orbital locations. The main components are power dividers, often 1-to-2 way 1-to-4 way, or 1-to-8 way that may be fixed or variable lengths of waveguide for phasing, variable phase shifters, and microwave switches. To achieve the best performance, the power divider couplers need to be accurately known and made. Accurate computer-aided design software is used for this purpose and, with software based on methods such as mode matching, BFNs can be designed and built for flight without any trimming. An example of a BFM for a reflector providing a beam covering the continental United States, Puerto Rico, Alaska, and Hawaii is shown in Figure 18.38 [98]. The array itself is shown in Figure 18.32. When the beam is scanned, rather than fixed, variable phase shifters are usually required. There are several ways of doing this, including uniform amplitude illumination over the array with variable phasing to steer the beam and use of Butler matrices with intervening phase shifters to produce multiple simultaneous beams. More details are given in Lo and Lee [70] and Mailloux [99].

18.6.3 Monopulse Feeds

A monopulse feed for a reflector or lens allows searching and tracking in radar and communications. The aim of such a feed is to produce sum, azimuth difference, and elevation difference signals from which tracking information can be obtained [100]. Among the available designs that use a secondary antenna there are three main types. The first type uses four or five feeds arranged orthogonally in the focal plane of the secondary antenna. One arrangement uses four multimode rectangular feed horns and a suitable feed network to balance the efficiency of the sum and difference beams. A practical implementation of this design is given by Lee and Chu [101] (see Figure 18.39a). These multi-mode horn designs suffer from different phase centers and different beamwidths between the sum and difference beams, which results in poor efficiency for the sum beam and high spillover for the difference beam. The second method uses a single

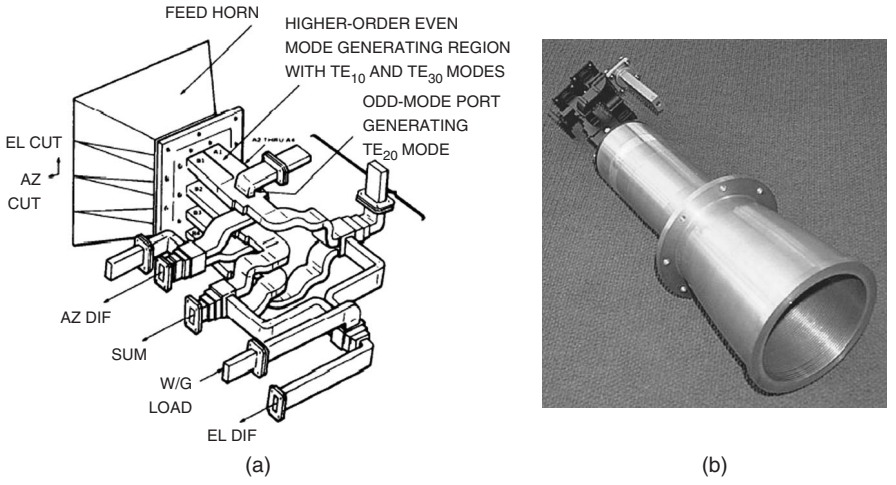


Figure 18.39 Monopulse tracking feeds: (a) multifeed array (from Ref. 101, courtesy of IEEE) and (b) multimode monopulse feed.

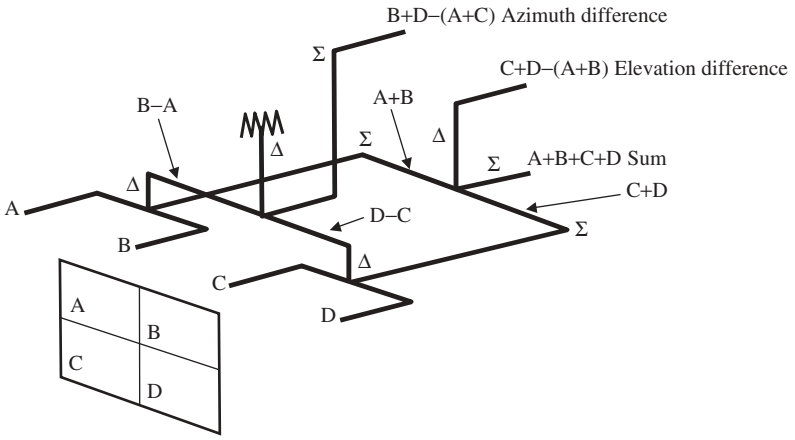


Figure 18.40 Schematic of a monopulse comparator (adapted from Ref. 104). Σ and Δ indicate sum and difference ports of four-port magic "T" junctions.

circular waveguide input to a corrugated horn from a beamforming network, which produces appropriate modes to give the three tracking signals. In one implementation of this, due to Watson et al. [102], the beamforming network produces a combination of TM_{01} and TE_{21} modes. The third method employs a single multimode horn and four input waveguides that control the modes excited in the horn by means of a comparator network (see Figs. 6.4b and 18.40). As shown in Figure 18.41, a corrugated horn can be used in this case where the HE_{11} mode provides the sum beam while combinations of orthogonal HE_{21} modes with H_{01} and E_{02} modes of corrugated waveguide provide azimuth and elevation difference beams [103, 104]. The challenge with this type of monopulse feed design is balancing the diameter of the horn at the junction, which gives

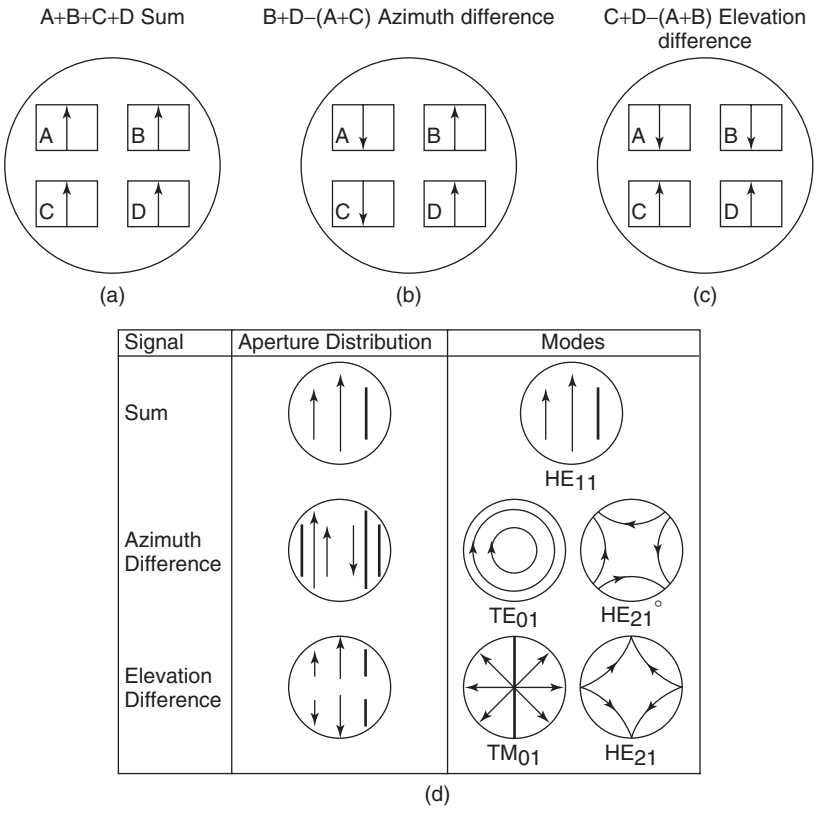


Figure 18.41 Mode excitations used to create (a) sum, (b) azimuth difference, and (c) elevation difference patterns; (d) desired aperture field distribution in horn and associated modes (adapted from Ref. 104).

the correct mode excitation, with the size of the input waveguides and also phasing the modes at the aperture. A rectangular waveguide can be used, but a square waveguide allows better performance. While greater beam efficiency is obtained for all beams at the center frequency, the bandwidth for best performance is typically narrow ($\sim$2%) and the match is often poor for the difference beam (reflection coefficient approximately -10 dB).

18.7 DESIGN METHODS

A feature of modern antenna feed design is the extent to which accurate computer software is now used. Electromagnetic software is used to analyze both the antenna and the feed system. While many analysis methods are available, there are five principal analysis methods on which the main software packages are based. These are finite-difference time-domain (FDTD), the finite-element method (FEM), mode matching (MM), method of moments (MoM), and the transmission-line matrix (TLM) method [105]. A brief outline of these methods now follows; for details the reader should consult the references [105–110]. The FDTD method is based on a discretization of Maxwell's equations in

differential form and the fields are solved iteratively in alternating time steps for a selected time period. The FEM discretizes the geometry by representing the region as a large number of subdomains, commonly tetrahedral or rectangular bodies in three dimensions on which variables are represented usually as polynomials with unknown coefficients. These are obtained by minimizing functions that incorporate Maxwell's equations and the boundary conditions. On the other hand, MM uses known closed-form solutions for parts of the structure (i.e., already satisfy Maxwell equations) that meet some of the boundary conditions, but not all, and therefore have unknown coefficients. The boundary conditions are enforced on these modal solutions plus the use of orthogonality conditions results in matrix equations in the unknown coefficients, which can be solved by usual matrix methods. The MoM represents the surface currents, called shape functions, on subdomains such as triangular functions for line currents [107] and the "roof top" tiles of Rao et al. [110]. These shape functions are introduced into integral equations for the fields, usually expressed as a summation of elementary radiating current sources, called a Green's function integral equation. Unknown coefficients on the shape functions are found by solving a set of equations, which are obtained by enforcing the boundary conditions. In the TLM method, the field equations are expressed as a network of transmission lines where the voltages and currents are related to the fields. Voltage pulses at each time step are incident on the network and the scattering of these pulses creates a new set of pulses, which become incident pulses on adjacent parts of the network at the next time step. The incident and scattered pulses are related through a scattering matrix. Additional elements, such as active components, and different materials can easily be added to the network. To handle radiation problems in FDTD, FEM, and TLM methods, in particular, absorbing boundary conditions (ABCs) are used, which is a special element or region that provides high loss for the numerically derived fields in the region and low reflection from the surface. Radiation is readily included in the MM method and the MoM through selected functions or Green's functions.

For this reason, in the past, the most widely used methods for feed antennas were the MoM and MM. However, with improvements in ABCs, the accuracy of the more general finite methods such as FDTD and FEM is approaching that of the traditional methods. Hybrid methods are also used and these can give additional flexibility. Methods used to design components relevant to feed antennas are reviewed in articles by Steer et al. [108] and Arndt et al. [109].

The most accurate method currently available for aperture-type feeds, and therefore radiation problems, employs a combination of MM and MoM. The former is used to analyze the internal structure while the MoM is used for the outer structure. To analyze a feed by MM, the scattering matrix solution of the transverse modal coefficients is required at each discrete boundary surface that describes the geometry of the structure. By cascading these matrices, the device can be completely characterized [10, 109]. This technique has been used with great success in analyzing numerous waveguide horns, arrays, and components. While a common elemental or building-block solution required to analyze complex structures in circular waveguide components is the junction between two circular waveguides, other scattering matrix solutions are available for junctions between coaxial waveguides, partially filled circular waveguides, rectangular waveguides, elliptical waveguides, and ridged waveguides as well as junctions between rectangular and circular waveguides. In all cases, however, the basic procedure remains the same. Radiation from the outer structure is handled by the MoM. The MM and MoM equations are solved together to give the mode amplitudes in the aperture and the external currents.

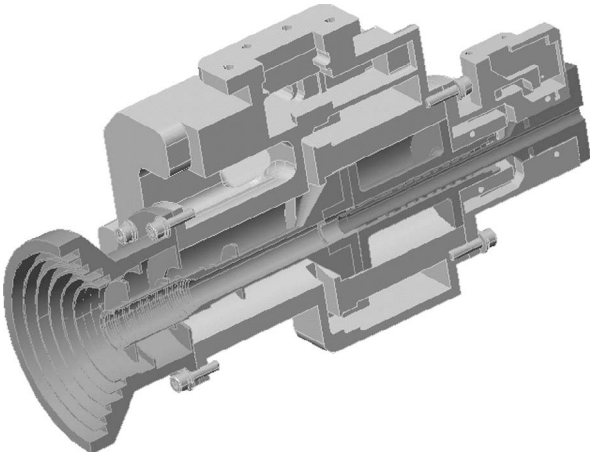


Figure 18.42 Cut-away drawing of a quad-band feed system that is analyzed completely with computer software (© CSIRO).

Once these are known, the radiated field is found by summing the contributions from all modes and sources. It is important to include all high-order terms of the radiated field in the calculation, especially when the feed and reflector are within the Fresnel zone, as ignoring these terms can lead to significant errors in predicting cross polarization and phase center. This particular approach can predict the performance of horn antennas very accurately, frequently over a 60-dB range of the radiation pattern.

Proprietary electromagnetics software packages, such as Ansoft's HFSS [111], which uses FEM, and CST Microwave Studio [112], which is based on another technique called finite integration that uses an integral form of Maxwell's equations, are now achieving impressive results that more specialized software attained in the past. As an example, Figure 18.42 shows the three-dimensional picture of a quad-band feed system that has been analyzed with CST Microwave Studio. The results for the input reflection coefficient for all bands show excellent agreement with accurate mode-matching software as well as with measured data for the fabricated feed system.

Another aspect of design is optimization for performance against specified criteria. Until recently, computer speed limited the degree that this was possible. Now, for example, horns are regularly designed by MM to suit particular performance criteria by optimizing the profile [33]. Optimizers are also employed in many electromagnetics software packages to adjust the geometry to peak the performance [108].

18.8 FUTURE

The developments in feed antennas proceed in parallel with new applications and improvements in software, materials, devices, and manufacturing methods. The unending requirement for higher levels of integration in future systems should see active devices, both receivers and transmitters, fully integrated with the feed antenna. At the moment, the lower bandwidth capability and higher initial costs of these units are limiting wider use of integrated feeds. The ability to manufacture a feed antenna has, in the past, limited some of the antennas that are used. With further improvements in software,

materials, and computer-aided manufacture, some of these limitations should be reduced. Of the feed antennas themselves, we expect to see more multiband and wideband feeds, which will allow system designers to provide greater flexibility and functionality. Some examples of these wideband feeds that will be more widely used are dielectric-lined horns and arrays of traveling-wave slots. The means of extracting the different frequency bands or wideband signals from these antennas remain a challenging problem. Another influence on feed development is the use of higher frequencies than in the past to achieve wider bandwidths and create smaller subsystems. The result is smaller feed antennas, which have higher tolerances and are more difficult to manufacture.

Feed antennas are by definition used in conjunction with other antennas, forming systems. In the past, these antennas were designed separately, and the effect of the interactions between them was ignored as insignificant or calculated approximately. Effects such as scattering from reflectors, feeds, struts, or the supporting structure itself can be important contributors to cross-channel interference in communication systems. It is anticipated that future computers and software, will allow the design and optimization of complete antenna systems as a matter of course. Encouraging steps have been made in this direction in recent years. While future developments depend much on improvements in computers, software, and display technology, the search will continue for the most accurate and fastest electromagnetics methods for the analysis and optimization of feed antennas.

REFERENCES

1. J. H. Bryant, *Heinrich Hertz—The Beginning of Microwaves*, IEEE Press, Piscataway, NJ, 1988.
2. J. F. Ramsay, Microwave antenna and waveguide techniques before 1900, *Proc. IRE*, Vol. 46, pp. 405–415, 1958.
3. S. Silver, *Microwave Antenna Theory and Design*, McGraw-Hill, New York, 1949.
4. C. C. Cutler, Directional microwave antenna, U.S. Patent 2,422,184, 1947.
5. C. A. Balanis, *Antenna Theory: Analysis and Design*, 2nd ed., John Wiley & Sons, Hoboken, NJ, 1997.
6. P. D. Potter, A new horn antenna with suppressed sidelobes and equal beamwidths, *Microwave J.*, Vol. 6, pp. 71–78, June 1963.
7. B. MacA. Thomas, A review of the early developments of circular-aperture hybrid-mode corrugated horns, *IEEE Trans. Antennas Propag.*, Vol. 34, pp. 930–935, 1978.
8. S. B. Cohn, Flare-angle changes in a horn as a means of pattern control, *Microwave J.*, Vol. 13, pp. 41–46, October 1970.
9. A.W. Love (Ed.), *Electromagnetic Horn Antennas*, IEEE Press, Piscataway, NJ, 1976.
10. A. D. Olver, P. J. B. Clarricoats, A. A. Kishk, and L. Shafai, *Microwave Horns and Feeds*, Peter Peregrinus Ltd., London, 1994.
11. P. C. Loux and R. W. Martin, Efficient aberration correction with a transverse focal plane array technique, *IRE Int. Conv. Rec.*, Vol. 12, pp. 125–131, 1964.
12. W. V. T. Rusch and P. D. Potter, *Analysis of Reflector Antennas*, Academic Press, New York, 1970.
13. S. G. Hay, F. R. Cooray, and T. S. Bird, Accurate modelling of edge diffraction in arrays of circular and rectangular horns, *J. Int. Nice Antennes–JINA*, pp. 645–648, 1996.
14. T. S. Bird, TE11 mode excitation of flanged circular coaxial waveguides with an extended center conductor, *IEEE Trans. Antennas Propag.*, Vol. 35, pp. 1358–1366, 1987.

15. T. S. Bird, G. L. James, and S. J. Skinner, Input mismatch of TE11 mode coaxial waveguide feeds, *IEEE Trans. Antennas Propag.*, Vol. 34, pp. 1030–1033, 1986.
16. G. L. James, Cross-polarization performance of flanged cylindrical and conical waveguides, *Proc. IREE (Aust.)*, Vol. 40, pp. 180–184, 1979.
17. B. MacA. Thomas, Design of corrugated conical horns, *IEEE Trans. Antennas Propag.*, Vol. 26, pp. 367–372, 1978.
18. P. J. B. Clarricoats and A. D. Olver, *Corrugated Horns and Microwave Antennas*, Peter Peregrinus Ltd., London, 1984.
19. G. L. James and B. MacA. Thomas, TE11 to HE11 cylindrical waveguide mode converters using ring-loaded slots, *IEEE Trans. Microwave Theory Tech.*, Vol. 30, pp. 278–285, 1982.
20. B. MacA. Thomas, G. L. James, and K. J. Greene, Design of wide-band corrugated conical horns for Cassegrain antennas, *IEEE Trans. Antennas Propag.*, Vol. 34, pp. 750–757, 1986.
21. G. L. James, Design of wide-band compact corrugated horns, *IEEE Trans. Antennas Propag.*, Vol. 32, pp. 1134–1138, 1984.
22. C. del Rio, R. Gonzalo, and M. Sorolla, High purity Gaussian beam excitation by optimal horn antenna, *Proc. ISAP*, pp. 1133–1136, 1996.
23. Z. Ying, A. A. Kishk, and P.-S. Kildal, Broadband compact horn feed for prime-focus reflectors, *Electron. Lett.*, Vol. 31, pp. 1114–1115, 1995.
24. H. Wolf, The scrimphorn, a new compact multimode horn for array application, *J. Int. Nice Antennes–JINA*, pp. 446–449, 1988.
25. M. Takabayashi, H. Deguchi, S. Makino, and T. Katagi, A design method of double bands dual mode conical horn antenna by using generalized transmission equation, *Antennas Propag. Soc. Int. Symp.,* Vol. 2, pp. 922–925, 1996.
26. T. Satoh, Dielectric-loaded horn antenna, *IEEE Trans. Antennas Propag.*, Vol. 20, pp. 199–201, 1972.
27. K. Raghavan, D. N. Paul, and P. B. Bains, Low cross-polarization radiator of circularly polarized radiation, U.S. Patent No. 4,972,199, 1990.
28. L. Staveley-Smith, W. E. Wilson, T. S. Bird, M. J. Disney, R. D. Ekers, K. C. Freeman, R. F. Haynes, M. W. Sinclair, R. A. Vaile, R. L. Webster, and A. E. Wright, The Parkes 21cm multibeam receiver, *Publ. Astron. Soc. Aust.*, Vol. 14, No. 3, pp. 243–248, 1996.
29. A. W. Rudge and N. A. Adatia, New class of primary-feed antennas for use with offset parabolic-reflector antennas, *Electron. Lett.*, Vol. 11, pp. 597–599, 1975.
30. A. Bhattacharyya and G. Goyette, A novel horn radiator with high aperture efficiency and low cross-polarization and applications in arrays and multibeam reflector antennas, *IEEE Trans. Antennas Propag.*, Vol. 52, pp. 2850–2859, 2004.
31. B. MacA. Thomas, Prime-focus one- and two-hybrid-mode feeds, *Electron. Lett.*, Vol. 6, pp. 460–461, 1970.
32. R. F. Thomas and D. A. Bathkar, A dual hybrid mode feed horn for DSN antenna performance enhancement, *JPL DSN Progress Report 42-22*, pp. 101–108, 1974.
33. G. G. Gentili, E. Martini, R. Nesti, and G. Pelosi, Performance analysis of dual profile corrugated circular waveguide horns for radioastronomy applications, *IEE Proc. Microwaves Antennas Propag.*, Vol. 148, pp. 119–122, 2001.
34. C. Granet and T. S. Bird, Optimization of corrugated horn radiation patterns via a spline-profile, in *ANTEM*, Montréal, Canada, 27–29 July 2002, pp. 307–310.
35. G. L. James, P. R. Clark, G. R. Graves, and T. S. Bird, The S/X-band feed for the Tasmanian earth resources satellite station (TERSS), *J. Electr. Electron. Eng. Aust.*, Vol. 15, pp. 307–314, 1995.
36. R. I. Henderson and P. J. Richards, Compact circularly-polarised coaxial feed, *IEE Int. Conf. Antennas Propag. (ICAP)*, pp. 327–330, 1995.

37. J. C. Chen, P. H. Stanton, and H. F. Reilly, X/X/Ka-band horn design, *IEEE AP-S Int. Symp. Dig.*, Vol. 3, pp. 2022–2025, 1996.
38. D. Green and D. Smith, Wide flare-angle horn antenna with means for radiating low levels of crosspolarization in two widely separated frequency bands, *IEEE AP-S Int. Symp. Dig.*, Vol. 3, pp. 2026–2029, 1996.
39. S. Amari and J. Bornemann, A study of the symmetry of field patterns of elliptic horns fed by elliptic apertures, *IEEE Antennas Propag. Symp.*, pp. 1992–1995, 1996.
40. M. E. J. Jeuken and L. F. G. Thurlings, The corrugated elliptical horn antenna, *IEEE Antennas Propag. Symp.*, pp. 9–12, 1975.
41. R. F. E Guy and R. W. Ashton, Cross-polar performance of an elliptical corrugated horn antenna, *Electron. Lett.*, Vol. 15, pp. 400–402, 1979.
42. Lier, A dielectric hybrid mode antenna feed: a simple alternative to the corrugated horn, *IEEE Trans. Antennas Propag.*, Vol. 34, pp. 21–29, 1986.
43. G. L. James, Propagation and radiation from partly filled elliptical waveguide, *IEE Proc.*, Vol. 136, Pt. H; pp. 195–201, 1989.
44. P. R. Clark and G. L. James, Ultra-wideband hybrid-mode feeds, *Electron. Lett.*, Vol. 31, pp. 1968–1969, 1995.
45. E. Lier, Y. Rahmat-Samii, and S. R. Rengarajan, Application of rectangular and elliptical dielcore feed horns to elliptical reflector antennas, *IEEE Trans. Antennas Propag.*, Vol. 39, pp. 1592–1597, 1991.
46. P. R. Clark and G. L. James, Analysis of hybrid-mode feed horns with simulated dielectric material, *IEEE AP-S Int. Symp. Dig.*, Vol. 2, pp. 772–775, 1994.
47. W. M. Truman and C. A. Balanis, Optimum design of horn feeds for reflector antennas, *IEEE Trans. Antennas Propag.*, Vol. 22, pp. 585–586, 1974.
48. T. S. Bird, Mode matching analysis of arrays of stepped rectangular horns and application to satellite design, *IEE Int. Conf. Antennas Propag. (ICAP)*, pp. 849–852, 1991.
49. A. W. Love, The diagonal horn antenna, *Microwave J.*, Vol. 5, pp. 117–122, March 1962.
50. R. C. Johnson and H. Jasik (Ed.), *Antenna Engineering Handbook*, McGraw-Hill, New York, 1984.
51. G. N. Tsandoulas and W. D. Fitzgerald, Aperture efficiency enhancement in dielectrically loaded horns, *IEEE Trans. Antennas Propag.*, Vol. 20, pp. 69–74, 1972.
52. R. E. Lawrie and L. Peters, Modifications of horn antennas for low sidelobe levels, *IEEE Trans. Antennas Propag.*, Vol. 14, pp. 605–610, 1966.
53. G. H. Bryant, Propagation in corrugated waveguides, *Proc. IEE*, Vol. 116, pp. 203–213, 1969.
54. R. Baldwin and P. A. McInnes, A rectangular corrugated feed horn, *IEEE Trans. Antennas Propag.*, Vol. 23, pp. 814–817, 1975.
55. C. Dragone, A rectangular horn of four corrugated plates, *IEEE Trans. Antennas Propag.*, Vol. 33, pp. 160–164, 1985.
56. L. C. da Silva and S. Ghosh, A comparative analysis between rectangular corrugated waveguides with plain and corner filled corrugations, *IEEE AP-S Int. Symp.*, pp. 976–979, 1990.
57. P.-S. Kildal, A small dipole fed resonant reflector antenna with high efficiency, low cross-polarization and low sidelobes, *IEEE Trans. Antennas Propag.*, Vol. 33, pp. 1386–1391, 1985.
58. G. T. Poulton and T. S. Bird, Improved rear-radiating cup feeds, in *IEEE Antennas Propagation Symposium*, Philadelphia, 9–13 June 1986, Vol. 1, pp. 79–82.
59. D. W. Griffin, Improvements in or relating to radiating waveguide feeders for radio-frequency electromagnetic waves, U.K. Patent 825,532, 1959.

60. R. Schwerdtfeger, A coaxial dual mode feed system, in *IEEE Antennas Propagation Symposium*, Seattle, 18–22 June 1979, Vol. 1, pp. 286–289.
61. T. S. Bird, A ring-slot flange for rear-radiating cup feeds, in *IEEE Antennas Propagation Symposium*, Blacksburg, VA, 1987, Vol. II, pp. 672–675.
62. P.-S. Kildal, The hat feed: a dual-mode rear radiating waveguide antenna having low cross-polarization, *IEEE Trans. Antennas Propag.*, Vol. 35, pp. 1010–1016, 1987.
63. P.-S. Kildal and J. Yang, FDTD optimization of the bandwidth of the hat feed for mm-wave reflector antennas, *IEEE AP-S Int. Symp.*, Vol. 3, pp. 1638–1641, 1997.
64. A. A. Kishk and L. Shafai, Optimization of microstrip feed geometry for prime focus reflector antennas, *IEEE Trans. Antennas Propag.*, Vol. 37, pp. 445–451, 1989.
65. A. K. Singh and G. Kumar, EMCP microstrip antennas as feed for satellite receiver, *IEEE AP-S Int. Symp.*, Vol. 2, pp. 1274–1277, 1996.
66. J. R. James and P. S. Hall, *Handbook of Microstrip Antennas*, Peter Peregrinus Ltd., London, UK, 1989.
67. R. A. Sainati, *CAD of Microstrip Antennas for Wireless Applications*, Artech House, Boston, 1996.
68. A. Kumar and H. D. Hristov, *Microwave Cavity Antennas*, Artech House, Norwood, MA, 1989.
69. A. W. Rudge, K. Milne, A. D. Olver, and P. Knight (Eds.), *The Handbook of Antenna Design*, Vols. 1 and 2, Peter Peregrinus Ltd., London, UK, 1986.
70. Y. T. Lo and S. W. Lee, *Antenna Handbook*, Van Nostrand Reinhold, New York, 1988.
71. W. A. Imbriale, Applications of the method of moments to thin-wire elements and arrays, in *Numerical and Asymptotic Techniques in Electromagnetics*, R. Mittra (Ed.), Springer-Verlag, New York, 1975.
72. D. G. Kiely, *Dielectric Aerials*, Methuen, London, 1953.
73. C. Salema, C. Fernandes, and R. K. Jha, *Solid Dielectric Horn Antennas*, Artech House, Norwood, MA, 1998.
74. K. A. S. Qassim and N. J. McEwan, Focal plane dielectric rod arrays with printed launchers, *IEE Int. Conf. Antennas Propag. (ICAP)*, Vol. 2, pp. 792–795, 1993.
75. L. R. Lewis, M. Fassett, and J. Hunt, A broadband stripline array element, *IEEE AP-S Int. Symp.*, pp. 335–337, 1974.
76. D. H. Schaubert, E. L. Kollberg, T. L. Korzeniowski, T. Thungren, J. F. Johansson, and K. S. Yngvesson, Endfire tapered slot antennas on dielectric substrates, *IEEE Trans. Antennas Propag.*, Vol. 33, pp. 1392–1400, 1985.
77. P. J. Gibson, The Vivaldi aerial, in *9th European Microwave Conference*, Brighton, UK, 1979, pp. 101–105.
78. E. Gazit, Improved design of Vivaldi antenna, *IEE Proc.*, Vol. 135, pp. 89–92, 1988.
79. J. D. S Langley, P. S. Hall, and P. Newham, Novel ultrawide-bandwidth Vivaldi antenna with low crosspolarisation, *Electron. Lett.*, Vol. 29, pp. 2004–2005, 1993.
80. V. H. Rumsey, On the design and performance of feeds for correcting spherical aberration, *IEEE Trans. Antennas Propag.*, Vol. 18, pp. 343–351, 1970.
81. A. W. Love, Spherical reflecting antennas with corrected line sources, *IRE Trans. Antennas Propag.*, Vol. 10, pp. 529–537, 1962.
82. A. W. Love and J. J. Gustincic, Line source feed for a spherical reflector, *IEEE Trans. Antennas Propag.*, Vol. 16, pp. 132–134, 1968.
83. A. W. Love, Scale model development of a high efficiency dual polarized line feed for the Arecibo spherical reflector, *IEEE Trans. Antennas Propag.*, Vol. 21, pp. 628–639, 1973.

84. T. Pratt and E. D. R. Shearman, Sectoral hoghorn: a new form of line feed for spherical reflector aerials, *Electron. Lett.*, Vol. 5, pp. 1–2, 1969.

85. T. S. Bird, Coaxial feed array for a short focal-length reflector, *IEEE AP-S Int. Symp. Dig.*, Vol. 3, pp. 1618–1621, 1997.

86. N. Amitay and M. J. Gans, Design of rectangular horn arrays with oversized aperture elements, *IEEE Trans. Antennas Propag.*, Vol. 29, pp. 871–884, 1981.

87. T. S. Bird, Analysis of mutual coupling in finite arrays of different-sized rectangular waveguides, *IEEE Trans. Antennas Propag.*, Vol. 38, pp. 166–172, 1990.

88. M. C. Bailey and C. W. Bostian, Mutual coupling in a finite array of circular apertures, *IEEE Trans. Antennas Propag.*, Vol. 22, pp. 178–184, 1974.

89. T. S. Bird, Mutual coupling in arrays of coaxial waveguides and horns, *IEEE Trans. Antennas Propag.*, Vol. 52, pp. 821–829, 2004.

90. D. H. Schaubert, J. A. Aas, M. E. Cooley, and N. E. Buris, Moment method analysis of infinite stripline-fed tapered slot antenna arrays with a ground plane, *IEEE Trans. Antennas Propag.*, Vol. 42, pp. 1161–1166, 1994.

91. W. A. Imbriale, *Large Antennas of the Deep Space Network*, Wiley-Interscience, Hoboken, NJ, 2003.

92. T. Kitsuregawa, *Advanced Technology in Satellite Communication Antennas*, Artech House, Norwood, MA, 1990.

93. P. F. Goldsmith, *Quasioptical Systems—Gaussian Beam Quasioptical Propagion and Applications*, IEEE Press, Piscataway, NJ, 1998.

94. G. L. James, *Geometrical Theory of Diffraction for Electromagnetic Waves*, 3rd ed., Peter Peregrinus Ltd., London, 1986.

95. J. Uher, J. Bornemann, and U. Rosenberg, *Waveguide Components for Antenna Feed Systems: Theory and CAD*, Artech House, Norwood, MA, 1993.

96. G. L. James, P. R. Clark, and K. J. Greene, Diplexing feed assemblies for application to dual-reflector antennas, *IEEE Trans. Antennas Propag.*, Vol. 51, pp. 1024–1029, 2003.

97. F. Alessandri, M. Mongiardo, and R. Sorrentino, Computer-aided design of beam forming networks for modern satellite antennas, *IEEE Trans. Microwave Theory Tech.*, Vol. 40, pp. 1117–1127, 1992.

98. T. S. Bird and C. Sroka, Design of the Ku-band antennas for the Galaxy HS601C satellites, *J. Electr. Electron. Eng. Aust.*, Vol. 12, pp. 267–273, 1992.

99. R. J. Mailloux, *Phased Array Antenna Handbook*, Artech House, Norwood, MA, 1993.

100. S. M. Sherman, *Monopulse Principles and Techniques*, Artech House, Norwood, MA, 1984.

101. K. M. Lee and R. S. Chu, Design and analysis of a multimode feed horn for a monopulse feed, *IEEE Trans. Antennas Propag.*, Vol. 36, pp. 171–181, 1988.

102. B. K. Watson, N. D. Dang, and S. Ghosh, A mode extraction network for RF sensing in satellite reflector antennas, *IEE Int. Conf. Antennas Propag. (ICAP)*, pp. 323–327, 1981.

103. P. J. B. Clarricoats and R. D. Elliot, Multimode corrugated waveguide feed for monopulse radar, *Proc. IEE*, Vol. 128, Pt. H, pp. 102–110, 1981.

104. H. Z. Zhang, C. Granet, and M. A. Sprey, A compact Ku-band monopulse horn, *Microwave Optical Tech. Lett.*, Vol. 34, pp. 9–13, 2002.

105. W. X. Zhang, *Engineering Electromagnetism: Functional Methods*, Ellis Horwood, New York, 1991.

106. P. P. Silvester and R. L. Ferrari, *Finite Elements for Electrical Engineers*, 3rd ed., Cambridge University Press, New York, 1996.

107. J. J. H. Wang, Generalized moment methods in electromagnetics, John Wiley & Sons, Hoboken, NJ, 1991.

108. M. B. Steer, J. W. Bandler, and C. M. Snowden, Computer-aided design of RF and microwave circuits and systems, *IEEE Trans. Microwave Theory Tech.*, Vol. 50, pp. 996–1005, 2002.

109. F. Arndt, J. Brandt, V. Catina, J. Ritter, I. Rullhusen, J. Dauelsberg, U. Hilgefort, and W. Wessel, Fast CAD and optimization of waveguide components and aperture antennas by hybrid MM/FE/MoM/FD methods—state-of-the-art and recent advances, *IEEE Trans. Microwave Theory Tech.*, Vol. 52, pp. 292–305, 2004.

110. S. M. Rao, D. R. Wiltron, and A. W. Glisson, Electromagnetic scattering by surfaces of arbitrary shape, *IEEE Trans. Antennas Propag.*, Vol. 30, pp. 409–418, 1982.

111. www.ansoft.com/.

112. www.cst.de/Content/Products/MWS/Overview.aspx.

CHAPTER 19

Near-Field Scanning Measurements: Theory and Practice

MICHAEL H. FRANCIS and RONALD C. WITTMANN

19.1 INTRODUCTION

As antennas increase in size or operating frequency, it becomes increasingly difficult to obtain adequate real estate to measure antenna properties in the far field. Typically, a distance of nD^2/λ is specified as the beginning of the far-field region, where n is often 2 for routine work, D is the diameter of the smallest sphere that encloses the antenna's radiating parts, and λ is the wavelength. At this distance and for a spherical wavefront, the phase differs by $\pi/8$ between the center and edge of the antenna. For precision measurements, n often must be much greater than 2 [1, Chap. 14]. Table 19.1 shows the far-field distances for some sample antenna diameters and operating frequencies. We see that even for antennas of modest size, the far-field distance can become prohibitively large when the frequency is high enough. Thus there are situations where measuring an antenna in the near field would be advantageous because large amounts of real estate are not required. There are other advantages as well. Since the near-field method is usually employed inside a chamber, it is not subject to the effects of the weather and also provides a more secure environment.

Determining the far-field pattern of an antenna from near-field measurements requires a mathematical transformation and correction for the characteristics of the measuring antenna (hereafter referred to as the probe). The first efficient near-field-to-far-field transformation for the planar geometry was developed by Kerns and Dayhoff [2] at the National Bureau of Standards (now the National Institute of Standards and Technology). Subsequently Kerns [3, 4] introduced probe correction. (Probe correction was developed earlier for a two-dimensional case (scanning on a circle) by Brown and Jull [5].) Within a few years the transformations were formulated for the spherical geometry by Jensen [6, 7] and Larsen [8] at the Technical University of Denmark. Wacker [9] of the National Bureau of Standards also made important contributions to the early development of spherical near-field scanning. The cylindrical transformation was developed, using reciprocity, by Leach and Paris [10–11] at the Georgia Institute of Technology. Subsequently, Yaghjian [12] at the National Bureau of Standards elaborated the cylindrical

Modern Antenna Handbook. Edited by Constantine A. Balanis

TABLE 19.1 Sample Far-Field Distances

Diameter (m)	Frequency (GHz)	Far-Field Distance (m)
1	3	20
3	3	180
10	3	2000
1	30	200
3	30	1800
10	30	20000
1	100	667
3	100	6000
10	100	66700

transformation based on a scattering-matrix approach. A detailed history of the development of near-field scanning methods can be found in Gillespie [13], especially the papers by Baird et al. [14], Hansen and Jensen [15], and Joy [16]. An extensive bibliography is given in these historical papers and also in Yaghjian [17]. A good, complementary reference that also treats all three geometries is Appel-Hansen et al. [18].

In principle, it is possible to transform to the far-field from measurements made on an arbitrary near-field surface. However, the transformation can be done efficiently for only three scanning geometries: planar (Section 19.2.3), spherical (Section 19.2.4), and Cylindrical (Section 19.2.5). These transformations require complex data (amplitude and phase).

19.2 NEAR-FIELD MEASUREMENT THEORY

19.2.1 Overview

Near-field measurement methods require the development of formulas that describe the coupling of two antennas in close proximity. Avoiding the polarization issue for now, we outline the process using the simpler example of a scalar field $u(\mathbf{r})$. (Near-field scanning measurements have also been developed for applications in acoustics [19–22].)

First, we expand the field of the test antenna in a complete set of modes $u_n(\mathbf{r})$:

$$u(\mathbf{r}) = a_0 \sum_{n=1}^{N} t_n u_n(\mathbf{r}) \tag{19.1}$$

where a_0 is the excitation. (In cases where modes are not discrete, the sum is replaced by an integral.) It must be possible to represent $u(\mathbf{r})$ to acceptable accuracy by use of a finite summation. The expansion (19.1) must be valid throughout the region of interest, including the region where measurements are made.

Practical probes do not (or only approximately) measure the field at a point in space. This fact is especially important near a source where fields may vary significantly over the volume of the probe. To account for probe effects, it is necessary to catalog the response of the probe to each mode in the expansion (19.1):

$$w_n'(\mathbf{r}', \mathbf{R}') = [\mathcal{P} u_n]\left(\mathbf{r}', \mathbf{R}'\right) \tag{19.2}$$

where the arguments must specify probe position (location and orientation). In this case, the vector $\mathbf{r}'$ gives the location and the matrix $\mathbf{R}'$ (Section 19.2.2.2.) determines the orientation. The probe is represented as a linear operator $\mathcal{P}$. (The probe can be modeled effectively as a differential operator [23], [24, Chap. 8]). As discussed in subsequent sections, w_n' can be calculated from the known receiving function of the probe.

The response of the probe to the field in Eq. (19.1) can be approximated as a superposition of the responses of the probe to the contributing modes:

$$\frac{w'(\mathbf{r}, \mathbf{R})}{a_0} = \sum_{n=1}^{N} w_n'(\mathbf{r}, \mathbf{R}) t_n \tag{19.3}$$

We assume that the presence of the probe does not affect the test-antenna field. Discrepancies due to probe–test antenna interaction are called multiple-reflection or mutual-coupling errors. (Multiple reflections are accounted for in developments that use a full scattering-matrix theory [12, 25, 26]; however, practical implementations (so far) retain only the direct interaction term discussed here.)

To determine the coefficients t_n we can, in principle, make independent measurements at the probe positions (locations and orientations) $(\mathbf{r}_i, \mathbf{R}_i)$, $1 \leq i \leq P$, $P \geq N$, and find the least-squares solution of the system of linear equations generated from the *transmission equation* (19.3). This would require $O(N^3)$ operations using, for example, Gaussian elimination. For many practical antenna problems $10^4 < N < 10^6$ and computational efficiency is a major concern. For this reason, near-field measurements are commonly made so that data points lie on planar, spherical, or cylindrical surfaces. The separation of coordinates, orthogonalities, and simple rotational and translational properties, which characterize the modes in these geometries, allow for significant reduction in processing time. Computational complexity is $O(N \log N)$ for the planar and cylindrical geometries, and $O(N^{3/2})$ for the spherical geometry. (Recent developments [27] suggest the possibility of a practical $O(N \log N)$ algorithm for spherical scanning.)

As separation increases, the coupling between test antenna and probe reduces to a simple asymptotic form. The resulting simpler theory (Friis transmission equation) and data acquisition requirements are the primary motivations for making *far-field measurements*. For electrically large antennas, however, obtaining adequate separation may not be practical. Additionally, the ability to use smaller separation distances means that many antenna measurements can be made within a controlled laboratory environment. This fact may result in improved security and accuracy.

Theory presented here represents a mathematical ideal. An effective, practical implementation requires the spirit of intelligent compromise that embodies the "art" of measurement.

19.2.2 General Considerations

19.2.2.1 Maxwell's Equations and Plane Waves We consider time-harmonic fields with frequency $f = \omega/2\pi$, wavenumber $k = 2\pi/\lambda$, and wavelength λ. (Time domain near-field measurements are discussed in [28].) The time-dependent factor $\exp(-i\omega t)$ is suppressed. In free space, the complex electromagnetic fields obey the equations

$$\nabla \cdot \mathbf{E}(\mathbf{r}) = \nabla \cdot \mathbf{H}(\mathbf{r}) = 0 \tag{19.4}$$

$$\frac{1}{k}\nabla \times \mathbf{E}(\mathbf{r}) = iZ_0\mathbf{H}(\mathbf{r})$$
$$\frac{1}{k}\nabla \times \mathbf{H}(\mathbf{r}) = \frac{1}{iZ_0}\mathbf{E}(\mathbf{r}) \tag{19.5}$$

The impedance of free space is $Z_0 = \sqrt{\mu_0/\varepsilon_0} \approx 377\ \Omega$. In particular, *the electric and magnetic fields each satisfy the Helmholtz equation*:

$$(\nabla^2 + k^2)\mathbf{E}(\mathbf{r}) = \mathbf{0} \tag{19.6}$$
$$(\nabla^2 + k^2)\mathbf{H}(\mathbf{r}) = \mathbf{0}$$

Perhaps the simplest solution of Eqs. (19.4)–(19.6) is the plane wave,

$$\mathbf{E}_{\text{pw}}(\mathbf{r}) = \frac{1}{2\pi}\mathbf{A}(\hat{\mathbf{k}})\exp(i\mathbf{k}\cdot\mathbf{r}) \tag{19.7}$$
$$\mathbf{H}_{\text{pw}}(\mathbf{r}) = \frac{1}{iZ_0}i\hat{\mathbf{k}} \times \mathbf{E}_{\text{pw}}(\mathbf{r})$$

where

$$\mathbf{k}\cdot\mathbf{A}(\hat{\mathbf{k}}) = 0 \tag{19.8}$$
$$\mathbf{k}\cdot\mathbf{k} = k^2 \tag{19.9}$$
$$\hat{\mathbf{k}} = \mathbf{k}/k \tag{19.10}$$

When $\mathbf{k}$ is real, $\mathbf{E}_{\text{pw}}$ is organized into planes of constant phase that are perpendicular to $\mathbf{k}$. These phase fronts propagate in the $\mathbf{k}$ direction at the speed of light $c = \lambda f \approx 3 \times 10^8$ m/s. We define the unit vector corresponding to $\mathbf{a}$ as $\hat{\mathbf{a}} = \mathbf{a}/\sqrt{\mathbf{a}\cdot\mathbf{a}}$ so that $\hat{\mathbf{a}}\cdot\hat{\mathbf{a}} = \mathbf{1}$; however, $\hat{\mathbf{a}}\cdot\hat{\mathbf{a}}^* \neq 1$ when $\mathbf{a}$ is complex.

19.2.2.2 Coordinate Systems The laboratory system is specified by the set $(\mathbf{O}, \hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}})$, where $\mathbf{O}$ is the origin and the unit vectors $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ give the directions of the x, y, and z axes in a right-handed Cartesian coordinate system. We also embed a local Cartesian coordinate system $(\mathbf{O}_0, \hat{\mathbf{x}}_0, \hat{\mathbf{y}}_0, \hat{\mathbf{z}}_0)$ in an antenna. The location of the antenna is given by the coordinates of the point $\mathbf{O}_0$ in the laboratory coordinate system. The orientation of the antenna is defined by expressing $\hat{\mathbf{x}}_0$, $\hat{\mathbf{y}}_0$, and $\hat{\mathbf{z}}_0$ in terms of $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$; for example, $\hat{\mathbf{x}}_0 = \mathbf{R}\hat{\mathbf{x}}$, $\hat{\mathbf{y}}_0 = \mathbf{R}\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}_0 = \mathbf{R}\hat{\mathbf{z}}$, where $\mathbf{R}$ is the 3×3 matrix (Cartesian basis) that induces a specified rotation [29, Eq. (C.45)].

The *minimum sphere* is the smallest sphere, centered on $\mathbf{O}_0$, that encloses the antenna. The mathematical description of an antenna is most efficient if $\mathbf{O}_0$ is chosen so that the minimum sphere is as small as possible. Additionally, we assume that $\hat{\mathbf{z}}_0$ *is directed away from the antenna*, usually along an axis of mechanical symmetry. This axis often corresponds closely to the axis of maximum gain or effective area. See Figures 19.1 and 19.18.

The term *reference position* is used to describe the state $(\mathbf{O}_0, \hat{\mathbf{x}}_0, \hat{\mathbf{y}}_0, \hat{\mathbf{z}}_0) = (\mathbf{O}, \hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}})$ for the test antenna and $(\mathbf{O}'_0, \hat{\mathbf{x}}'_0, \hat{\mathbf{y}}'_0, \hat{\mathbf{z}}'_0) = (z'\hat{\mathbf{z}}, -\hat{\mathbf{x}}, \hat{\mathbf{y}}, -\hat{\mathbf{z}})$ for the probe. When the test antenna and probe are in their reference positions, they face each other along the laboratory z axis (Figure 19.1). For any valid z', there will be a plane $z = z_0$ such that the test antenna is contained in the region $z < z_0$ and the probe is contained in the region $z > z_0$. It is convenient to begin in the reference positions when describing the translations and rotations that effectively move the probe over the measurement surface. (In some near-field scanning implementations, the test antenna rotates in the laboratory coordinate system. See Figure 19.14, for example.)

19.2.2.3 Waveguide Feeds We assume that an antenna is coupled to a generator or to a load through a single-moded, lossless waveguide feed (which may be coaxial). Phase is measured relative to a reference plane that is usually located at the point (waveguide port) where the antenna connects to the RF system. See Refs. 25 and 30 for a complete discussion of waveguide circuit theory.

The power accepted by a transmitting antenna is

$$P_0 = K|a_0|^2\left(1 - |\Gamma|^2\right) \tag{19.11}$$

where a_0 is the (complex) amplitude of the incident waveguide mode and Γ is the reflection coefficient looking into the antenna through its feed.

The constant of proportionality K is arbitrary since it depends on the normalization chosen for the waveguide modes. Once K is specified, $|a_0|$ can be uniquely determined by a measurement of power. (The phase of a_0 is often not important.) For convenience we use

$$K = \frac{1}{2k^2 Z_0} \tag{19.12}$$

Thus a_0 has the dimension V/m.

19.2.2.4 Transmitting and Receiving Functions Far from an antenna that is transmitting, the radiated electromagnetic fields can be characterized simply:

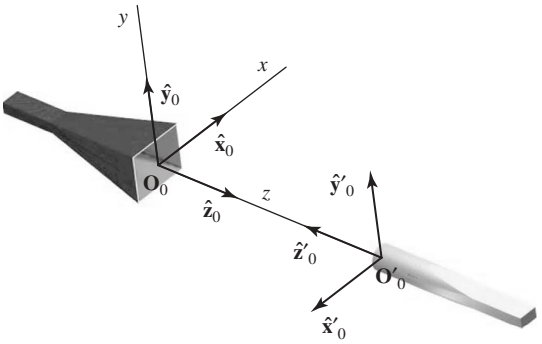


Figure 19.1 A test antenna and a probe in their reference positions in the laboratory coordinate system; that is, $(\mathbf{O}_0, \hat{\mathbf{x}}_0, \hat{\mathbf{y}}_0, \hat{\mathbf{z}}_0) = (\mathbf{O}, \hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}})$ and $\left(\mathbf{O}'_0, \hat{\mathbf{x}}'_0, \hat{\mathbf{y}}'_0, \hat{\mathbf{z}}'_0\right) = \left(z'\hat{\mathbf{z}}, -\hat{\mathbf{x}}, \hat{\mathbf{y}}, -\hat{\mathbf{z}}\right)$.

$$\begin{aligned}\mathbf{E}(\mathbf{r}) &\underset{r\to\infty}{\sim} a_0\mathbf{t}(\hat{\mathbf{r}})\frac{\exp(ikr)}{ikr}, \qquad \hat{\mathbf{r}}\cdot\mathbf{t}(\hat{\mathbf{r}}) = 0\\ \mathbf{H}(\mathbf{r}) &\underset{r\to\infty}{\sim} \frac{1}{iZ_0} i\hat{\mathbf{r}}\times\mathbf{E}(\mathbf{r})\end{aligned} \tag{19.13}$$

The *transmitting function* $\mathbf{t}(\hat{\mathbf{r}})$ depends only on direction, which is specified by the unit vector $\hat{\mathbf{r}}$.

Next, consider a plane wave (Eq. (19.7))

$$\mathbf{E}_{\text{pw}}(\mathbf{r}) = \frac{1}{2\pi}\mathbf{A}(\hat{\mathbf{k}})\exp(i\mathbf{k}\cdot\mathbf{r})$$

that is incident on an antenna that is receiving. The *receiving function* $\mathbf{s}(\hat{\mathbf{k}})$ is defined so that the received signal is

$$w = (1 - \Gamma\Gamma_\ell)b_0 = \mathbf{s}(\hat{\mathbf{k}})\cdot\mathbf{A}(\hat{\mathbf{k}}) \tag{19.14}$$

Here, b_0 is the (complex) amplitude of the emergent waveguide mode and Γ_ℓ is the reflection coefficient of the load; that is, w is the amplitude of the emergent waveguide mode that would be present were the load nonreflecting.

The forms of the functions $\mathbf{t}$ and $\mathbf{s}$ depend on the position (location and orientation) of the antenna in the laboratory coordinate system. For the special case $(\mathbf{O}_0, \hat{\mathbf{x}}_0, \hat{\mathbf{y}}_0, \hat{\mathbf{z}}_0) = (\mathbf{O}, \hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}})$ in which the antenna is located at the origin and is pointed in the $\hat{\mathbf{z}}$ direction, we add a subscript 0, so that

$$\mathbf{t}(\hat{\mathbf{k}}) \to \mathbf{t}_0(\hat{\mathbf{k}}), \quad \mathbf{s}(\hat{\mathbf{k}}) \to \mathbf{s}_0(\hat{\mathbf{k}}) \tag{19.15}$$

For the special case $(\mathbf{O}_0, \hat{\mathbf{x}}_0, \hat{\mathbf{y}}_0, \hat{\mathbf{z}}_0) = (\mathbf{O}, -\hat{\mathbf{x}}, \hat{\mathbf{y}}, -\hat{\mathbf{z}})$ in which the antenna is located at the origin and is pointed in the $-\hat{\mathbf{z}}$ direction, we add a subscript π, so that

$$\mathbf{t}(\hat{\mathbf{k}}) \to \mathbf{t}_\pi(\hat{\mathbf{k}}), \quad \mathbf{s}(\hat{\mathbf{k}}) \to \mathbf{s}_\pi(\hat{\mathbf{k}}) \tag{19.16}$$

The functions $\mathbf{t}_\pi(\hat{\mathbf{r}})$ and $\mathbf{s}_\pi(\hat{\mathbf{r}})$ can be obtained by rotating $\mathbf{t}_0(\hat{\mathbf{r}})$ and $\mathbf{s}_0(\hat{\mathbf{r}})$ by 180° about the laboratory y axis.

In accordance with our conventions, we may take

$$\mathbf{s}(\hat{\mathbf{k}}) = \mathbf{t}(-\hat{\mathbf{k}}) \tag{19.17}$$

when an antenna is *reciprocal*. See Kerns [25, Chap. II, Appendix A] for a detailed discussion of reciprocity.

19.2.2.5 Basic Antenna Parameters Following IEEE Standard 145-1993 [31], the *gain* $G(\hat{\mathbf{r}})$ and *effective area* $\sigma(\hat{\mathbf{r}})$ of an antenna are

$$G(\hat{\mathbf{r}}) = \frac{4\pi}{1 - |\Gamma|^2}\,\|\mathbf{t}_0(\hat{\mathbf{r}})\|^2 \tag{19.18}$$

$$\sigma(\hat{\mathbf{r}}) = \frac{\lambda^2}{1 - |\Gamma|^2} \|\mathbf{s}_0(\hat{\mathbf{r}})\|^2 \tag{19.19}$$

When an antenna is reciprocal (Eq. (19.17)),

$$\sigma(\hat{\mathbf{r}}) = \frac{\lambda^2}{4\pi} G(-\hat{\mathbf{r}}) \tag{19.20}$$

We use the term *pattern* to refer to a transmitting or receiving "function" of unspecified normalization. Any pattern can be normalized to satisfy Eq. (19.18) or (19.19) if Γ is known and if the gain or effective area is available in some direction.

The *directivity* $D(\hat{\mathbf{r}})$ is given by

$$D(\hat{\mathbf{r}}) = G(\hat{\mathbf{r}}) \frac{P_0}{P_T} = \frac{4\pi}{\int_0^{2\pi} \int_0^{\pi} \|\mathbf{t}_0(\hat{\mathbf{r}})\|^2 \sin\theta \, d\theta \, d\varphi} \|\mathbf{t}_0(\hat{\mathbf{r}})\|^2 \tag{19.21}$$

where P_T is the total radiated power. Due to ohmic losses, $P_T < P_0$ and so $G(\hat{\mathbf{r}}) < D(\hat{\mathbf{r}})$. Equation (19.21) holds regardless of the normalization used for the pattern. Calculation of directivity in planar or cylindrical scanning is subject to truncation error because the measurement surfaces never completely enclose the test antenna. Spherical scanning, however, provides the simple result (see Eq. (19.60))

$$\int_0^{2\pi} \int_0^{\pi} \|\mathbf{t}_0(\hat{\mathbf{r}})\|^2 \sin\theta \, d\theta \, d\varphi = \sum_{nm} \left(\left| t_{nm}^1 \right|^2 + \left| t_{nm}^2 \right|^2 \right) \tag{19.22}$$

In addition to the above parameters, the *equivalent isotropic radiated power* (EIRP) [31]

$$EIRP(\hat{\mathbf{r}}) = G(\hat{\mathbf{r}}) P_0 \tag{19.23}$$

is often of interest, particularly for communication satellite applications. Near-field measurement of EIRP may require a specialized technique since the antenna's waveguide port is not accessible in some systems [32].

19.2.2.6 Miscellaneous Conventions We assume the "standard" branch when specifying square roots; that is, if

$$a + ib = \sqrt{x + iy}$$

then $a \geq 0$, $b \geq 0$ when $y \geq 0$, and $b < 0$ when $y < 0$. In particular, *the square root of a nonzero real number is either positive real or positive imaginary*.

Quantities associated with a probe are marked with primes (single or double) when such distinctions are desirable.

In this development, the probe always receives and the test antenna always transmits. Transmission equations for transmitting probes and receiving test antennas may be found by straightforward generalization of our arguments. For the convenience of the reader, however, Appendix A lists transmission equations for transmitting probes and receiving test antennas.

For the most part, theories of near-field planar, spherical, and cylindrical scanning have developed independently. We have attempted to bring some uniformity to the notation. Here, for example, (1) modal functions are dimensionless, (2) pattern functions (and the scattering matrix, in general) are dimensionless, and (3) the modal coefficients carry the dimensions of the electric field (V/m).

19.2.3 Planar Scanning

19.2.3.1 Plane Wave Representation Scalar solutions of Eq. (19.6), separated in Cartesian coordinates (Figure 19.2), are given by plane waves

$$\begin{aligned} u_{\alpha\beta}(\mathbf{r}) &= \frac{1}{2\pi}\exp(i\mathbf{k}\cdot\mathbf{r}) \\ &= \frac{1}{2\pi}\exp(i\alpha x + i\beta y)\exp(i\gamma z) \end{aligned} \tag{19.24}$$

where

$$\mathbf{k}\cdot\mathbf{k} = k^2$$

In Eq. (19.24) we restrict $\mathbf{k}$ so that

$$\begin{gathered} k_x = \alpha, \quad k_y = \beta, \quad k_z = \gamma \\ -\infty < \alpha, \beta < \infty \\ \gamma = \sqrt{k^2 - \alpha^2 - \beta^2} \end{gathered} \tag{19.25}$$

In formulas where α, β, or γ appear explicitly, it is to be understood that $\mathbf{k} = \alpha\hat{\mathbf{x}} + \beta\hat{\mathbf{y}} + \gamma\hat{\mathbf{z}}$.

When γ is real (corresponding to $k^2 > \alpha^2 + \beta^2$), $u_{\alpha\beta}(\mathbf{r})$ describes a plane wave that propagates into the half-space $z > z_0$. When γ is imaginary (corresponding to $k^2 < \alpha^2 + \beta^2$), $u_{\alpha\beta}(\mathbf{r})$ is an "evanescent" wave that decays exponentially as z increases. When

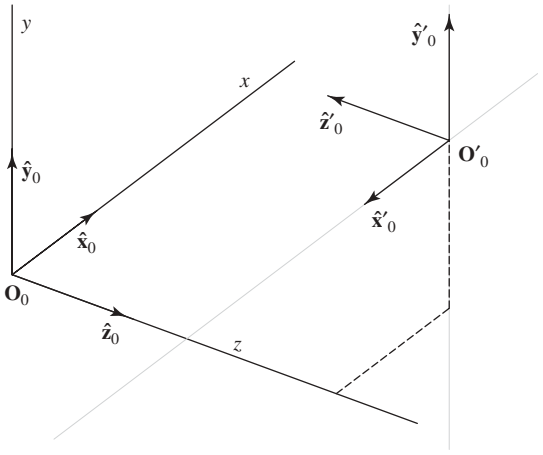


Figure 19.2 Planar scanning geometry. The scan plane is $z = d$. The test antenna is in its reference position (see Figure 19.1). The probe is shown at the position (x, y, z) with $x = d/2$, $y = d/2$, and $z = d$. Furthermore, $\hat{\mathbf{x}}'_0 = -\hat{\mathbf{x}}$, $\hat{\mathbf{y}}'_0 = \hat{\mathbf{y}}$, and $\hat{\mathbf{z}}'_0 = -\hat{\mathbf{z}}$.

sources are confined to the region $z < z_0$, the functions $u_{\alpha\beta}(\mathbf{r})$ form a complete basis for representing fields in the free-space region $z > z_0$.

Vector solutions of Eqs. (19.4) and (19.6) follow directly from the scalar solutions (Eq. (19.24)):

$$\begin{aligned} \mathbf{m}_{\alpha\beta}(\mathbf{r}) &= \frac{1}{k}\nabla \times \mathbf{n}_{\alpha\beta}(\mathbf{r}) = u_{\alpha\beta}(\mathbf{r})\mathbf{e}^1(\hat{\mathbf{k}}) \\ \mathbf{n}_{\alpha\beta}(\mathbf{r}) &= \frac{1}{k}\nabla \times \mathbf{m}_{\alpha\beta}(\mathbf{r}) = u_{\alpha\beta}(\mathbf{r})\mathbf{e}^2(\hat{\mathbf{k}}) \end{aligned} \tag{19.26}$$

$$\hat{\mathbf{k}} \cdot \mathbf{m}_{\alpha\beta}(\mathbf{r}) = \hat{\mathbf{k}} \cdot \mathbf{n}_{\alpha\beta}(\mathbf{r}) = 0 \tag{19.27}$$

Here,

$$\begin{aligned} \mathbf{e}^1(\hat{\mathbf{r}}) &= \begin{cases} \dfrac{\hat{\mathbf{z}} \times \hat{\mathbf{r}}}{\sqrt{1 - (\hat{\mathbf{z}} \cdot \hat{\mathbf{r}})^2}}, & \hat{\mathbf{r}} \neq \pm\hat{\mathbf{z}} \\ \pm\hat{\mathbf{y}}, & \hat{\mathbf{r}} = \pm\hat{\mathbf{z}} \end{cases} \\ \mathbf{e}^2(\hat{\mathbf{r}}) &= i\hat{\mathbf{r}} \times \mathbf{e}^1(\hat{\mathbf{r}}) \end{aligned} \tag{19.28}$$

Equation (19.27) is a direct restatement of Eq. (19.4).

The electromagnetic fields of a test antenna (in the reference position) can be expanded as

$$\begin{aligned} \mathbf{E}(\mathbf{r}) &= a_0 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[t^1_{\alpha\beta}\mathbf{m}_{\alpha\beta}(\mathbf{r}) + t^2_{\alpha\beta}\mathbf{n}_{\alpha\beta}(\mathbf{r})\right] \frac{d\alpha\, d\beta}{\gamma k} \\ iZ_0\mathbf{H}(\mathbf{r}) &= a_0 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[t^2_{\alpha\beta}\mathbf{m}_{\alpha\beta}(\mathbf{r}) + t^1_{\alpha\beta}\mathbf{n}_{\alpha\beta}(\mathbf{r})\right] \frac{d\alpha\, d\beta}{\gamma k}, \quad z > z_0 \end{aligned} \tag{19.29}$$

When γ is real,

$$\frac{d\alpha\, d\beta}{\gamma k} \to \sin\theta\, d\theta\, d\varphi \tag{19.30}$$

so that $d\alpha\, d\beta/\gamma k$ is a differential solid angle. For $\alpha^2 + \beta^2 > k^2$, the integrand in Eq. (19.29) is exponentially attenuated with increasing $|\gamma|$. The resulting bandlimit has important consequences in the practical implementation of the theory.

We write

$$\mathbf{t}_0(\hat{\mathbf{k}}) = t^1_{\alpha\beta}\mathbf{e}^1(\hat{\mathbf{k}}) + t^2_{\alpha\beta}\mathbf{e}^2(\hat{\mathbf{k}}) \tag{19.31}$$

and, following Kerns [25] and Yaghjian [33, Eq. 8],

$$\mathbf{E}(\mathbf{r}) = \frac{a_0}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbf{t}_0\left(\hat{\mathbf{k}}\right) \exp(i\mathbf{k} \cdot \mathbf{r}) \frac{d\alpha\, d\beta}{\gamma k}, \quad z > z_0 \tag{19.32}$$

Equation (19.32) is a superposition of vector plane waves of the form in Eq. (19.7) in which $\mathbf{k}$ is restricted according to Eq. (19.25), so that $k_z = \gamma$ is positive real or positive imaginary.

A stationary-phase evaluation of Eq. (19.32) suggests the asymptotic result

$$\mathbf{E}(\mathbf{r}) \underset{r\to\infty}{\sim} a_0 \mathbf{t}_0(\hat{\mathbf{r}}) \frac{\exp(ikr)}{ikr} \tag{19.33}$$

that confirms $\mathbf{t}_0(\hat{\mathbf{r}})$ as the transmitting function. Equation (19.33) can be justified rigorously when the source is finite [34]. (A finite source can be enclosed within a sphere of finite radius.)

The transverse field is

$$\mathbf{E}_t(\mathbf{r}) = \mathbf{E}(\mathbf{r}) - \hat{\mathbf{z}}\hat{\mathbf{z}} \cdot \mathbf{E}(\mathbf{r}) = \frac{a_0}{2\pi} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbf{t}_t(\hat{\mathbf{k}}) \exp(i\mathbf{k} \cdot \mathbf{r}) \frac{d\alpha\, d\beta}{\gamma k} \tag{19.34}$$

This is a two-dimensional Fourier transform that can be inverted to give

$$\mathbf{t}_t(\hat{\mathbf{k}}) = \frac{\gamma k}{2\pi} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\mathbf{E}_t(\mathbf{r})}{a_0} \exp(-i\mathbf{k} \cdot \mathbf{r}) dx\, dy, \quad z = d > z_0 \tag{19.35}$$

Because $\hat{\mathbf{k}} \cdot \mathbf{t}_0(\hat{\mathbf{k}}) = 0$,

$$\mathbf{t}_0(\hat{\mathbf{k}}) = \frac{1}{\gamma} \mathbf{k} \times [\mathbf{t}_t(\hat{\mathbf{k}}) \times \hat{\mathbf{z}}] \tag{19.36}$$

Thus the transmitting function is determined by the transverse electric field on any plane $z = d > z_0$.

19.2.3.2 Probe Response We begin with the test antenna and probe in their reference positions. In planar scanning, the response of the probe is recorded as the probe is translated (no rotation allowed) over the measurement plane $z = z' = d$. See Figure 19.1.

In the laboratory coordinate system, position (location and orientation) of the probe is described by the parameter set (x, y, z) with the probe reference position corresponding to $(0, 0, d)$. The position of a similarly translated test antenna is also described by these parameters with the test antenna reference position corresponding to $(0, 0, 0)$.

Consider the output when the test antenna is at its reference position and the probe is at (x', y', z'). The same output will be observed when the probe is at $(0, 0, 0)$ and the test antenna is at $(-x', -y', -z')$. This fact may be expressed mathematically as

$$w'(\mathbf{r}') = [\mathcal{P} \cdot \mathbf{E}](x', y', z') = [\mathcal{P} \cdot \tilde{\mathbf{E}}](0, 0, 0) \tag{19.37}$$

Here, $\mathcal{P}$ is a linear operator that represents the probe, $\mathbf{E}(\mathbf{r})$ is the field of the test antenna (at its reference position), and $\tilde{\mathbf{E}}(\mathbf{r})$ is the field of the translated test antenna.

Now, let the "test antenna" radiate the plane wave

$$\mathbf{E}_{\alpha\beta}(\mathbf{r}) = \frac{1}{2\pi} \mathbf{t}_0(\hat{\mathbf{k}}) \exp(i\mathbf{k} \cdot \mathbf{r})$$

so that

$$\tilde{\mathbf{E}}_{\alpha\beta}(\mathbf{r}) = \frac{1}{2\pi} \mathbf{t}_0(\hat{\mathbf{k}}) \exp[i\mathbf{k} \cdot (\mathbf{r} + \mathbf{r}')] = \mathbf{E}_{\alpha\beta}(\mathbf{r}) \exp(i\mathbf{k} \cdot \mathbf{r}') \tag{19.38}$$

From Eqs. (19.37) and (19.38)

$$w'_{\alpha\beta}(\mathbf{r}') = \exp(i\mathbf{k}\cdot\mathbf{r}')\left[\boldsymbol{\mathcal{P}}\cdot\mathbf{E}_{\alpha\beta}\right](0,0,0) \tag{19.39}$$

or

$$w'_{\alpha\beta}(\mathbf{r}') = \mathbf{s}'_{\pi}(\hat{\mathbf{k}})\cdot\mathbf{t}_0(\hat{\mathbf{k}})\exp(i\mathbf{k}\cdot\mathbf{r}') \tag{19.40}$$

where we have applied Eqs. (19.14) and (19.16). Equation (19.40) expresses the "obvious" fact that the output of a probe can vary only by a phase factor as it is translated from location to location in the presence of an incident (propagating) plane wave field. In the notation of Section 19.2.2.4,

$$\mathbf{s}'_{\pi}(\hat{\mathbf{k}}) = \mathbf{s}'_{\pi}(\hat{\mathbf{k}})\exp(i\mathbf{k}\cdot\mathbf{r}') \tag{19.41}$$

where $\mathbf{s}'_{\pi}(\hat{\mathbf{k}})$ is the receiving function of the translated probe at the location $\mathbf{r}'$.

19.2.3.3 Kerns Transmission Equation In Eq. (19.32) we have a modal expansion corresponding to Eq. (19.1). In Eq. (19.40) we have the response of the probe to each mode corresponding to Eq. (19.2). Finally, corresponding to Eq. (19.3), we have the *Kerns transmission formula* [25]

$$\frac{w'(\mathbf{r})}{a_0} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathbf{s}'_{\pi}(\hat{\mathbf{k}})\cdot\mathbf{t}_0(\hat{\mathbf{k}})\exp(i\mathbf{k}\cdot\mathbf{r})\frac{d\alpha\,d\beta}{\gamma k} \tag{19.42}$$

Equation (19.42) is valid when multiple reflections can be ignored and when the probe and test antenna can be physically separated by a plane perpendicular to $\hat{\mathbf{z}}$.

The Fourier transforms in Eq. (19.42) may be inverted to yield

$$\begin{aligned} T'_{\alpha\beta} &= \mathbf{s}'_{\pi}(\hat{\mathbf{k}})\cdot\mathbf{t}_0(\hat{\mathbf{k}}) = s^{1\prime}_{\alpha\beta}t^{1}_{\alpha\beta} - s^{2\prime}_{\alpha\beta}t^{2}_{\alpha\beta} \\ &= \frac{\gamma k}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{w'(\mathbf{r})}{a_0}\exp(-i\mathbf{k}\cdot\mathbf{r})dx\,dy, \quad z = d \end{aligned} \tag{19.43}$$

with Eq. (19.31) and

$$\mathbf{s}'_{\pi}(\hat{\mathbf{k}}) = s^{1\prime}_{\alpha\beta}\mathbf{e}^{1}(\hat{\mathbf{k}}) + s^{2\prime}_{\alpha\beta}\mathbf{e}^{2}(\hat{\mathbf{k}}) \tag{19.44}$$

Since $T'_{\alpha\beta}$ can be computed from data $w'(\mathbf{r})/a_0$ measured on the plane $z = d$, Eq. (19.43) provides one equation for the two unknown components of $\mathbf{t}_0(\hat{\mathbf{k}})$. Data from a second probe are needed to provide sufficient information to solve for $\mathbf{t}_0(\hat{\mathbf{k}})$. With two probes, we have the linear system

$$\begin{pmatrix} T'_{\alpha\beta} \\ \\ T''_{\alpha\beta} \end{pmatrix} = \begin{pmatrix} s^{1\prime}_{\alpha\beta} & -s^{2\prime}_{\alpha\beta} \\ \\ s^{1\prime\prime}_{\alpha\beta} & -s^{2\prime\prime}_{\alpha\beta} \end{pmatrix}\begin{pmatrix} t^{1}_{\alpha\beta} \\ \\ t^{2}_{\alpha\beta} \end{pmatrix} \tag{19.45}$$

Here, quantities associated with the second probe are marked with a double prime. Thus

$$\begin{pmatrix} t^{1}_{\alpha\beta} \\ \\ t^{2}_{\alpha\beta} \end{pmatrix} = \frac{-i}{\hat{\mathbf{k}}\cdot\left(\mathbf{s}'_{\pi}\times\mathbf{s}''_{\pi}\right)}\begin{pmatrix} -s^{2\prime\prime}_{\alpha\beta} & s^{2\prime}_{\alpha\beta} \\ \\ -s^{1\prime\prime}_{\alpha\beta} & s^{1\prime}_{\alpha\beta} \end{pmatrix}\begin{pmatrix} T'_{\alpha\beta} \\ \\ T''_{\alpha\beta} \end{pmatrix} \tag{19.46}$$

Obviously, probes should be selected so that $\mathbf{s}'_\pi\left(\hat{\mathbf{k}}\right) \times \mathbf{s}''_\pi\left(\hat{\mathbf{k}}\right) \neq \mathbf{0}$. If the first probe is linearly polarized, a "second" probe may be realized by rotating the first probe by 90° about its z axis.

In the most general setting, the Kerns transmission formula may be written

$$\frac{w'}{a_0} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbf{s}'(\hat{\mathbf{k}}) \cdot \mathbf{t}(\hat{\mathbf{k}}) \frac{d\alpha\, d\beta}{\gamma k} \tag{19.47}$$

This describes the coupling between two arbitrarily positioned antennas, provided they can be separated by a planar surface (which is used to define the z direction). From this point of view, both the spherical and cylindrical transmission formulas may be considered special cases.

19.2.4 Spherical Scanning

19.2.4.1 Spherical Wave Representation Scalar solutions of Eq. (19.6), separated in spherical coordinates (Figure 19.3), are given by spherical waves

$$\begin{aligned} u_{nm}(\mathbf{r}) &= f_n(kr) Y_{nm}(\hat{\mathbf{r}}) \\ f_n(x) &= i^n h_n^{(1)}(x) \end{aligned} \tag{19.48}$$

(See Ref. 35, for example). Here, the $h_n^{(1)}$ are spherical Hankel functions of the first kind and the Y_{nm} are (scalar) spherical harmonics. The multipole index is a nonnegative integer, $n = 0, 1, \ldots$. For a given n, the azimuthal (or component) index is an integer m

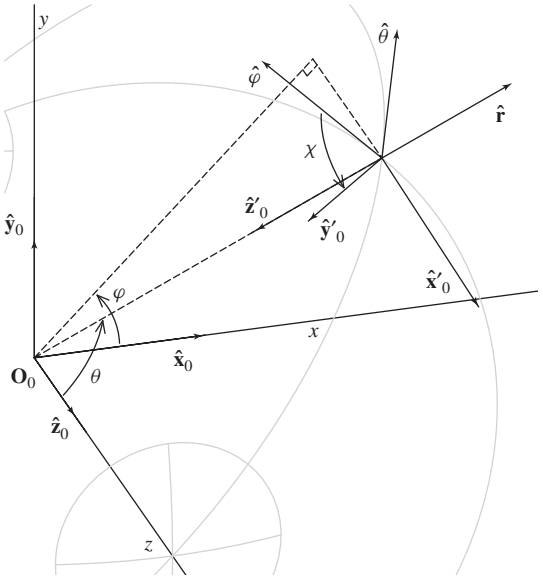


Figure 19.3 Spherical scanning geometry. The scan sphere is $r = d$. The test antenna is in its reference position (see Figure 19.1). The probe is shown at the position $(r, \varphi, \theta, \chi)$ with $r = d$, $\varphi = \pi/4$, and $\theta = \pi/3$. Furthermore, $\hat{\mathbf{y}}'_0 = \hat{\boldsymbol{\varphi}}$ when $\chi = 0$, $\hat{\mathbf{y}}'_0 = \hat{\boldsymbol{\theta}}$ when $\chi = -\pi/2$, and $\mathbf{z}'_0 = -\hat{\mathbf{r}}$.

in the range $-n \le m \le n$. (The nth multipole has $2n+1$ components.) Scalar spherical waves satisfy the radiation condition

$$u_{nm}(\mathbf{r}) \underset{r\to\infty}{\sim} Y_{nm}(\hat{\mathbf{r}}) \frac{\exp(ikr)}{ikr} \tag{19.49}$$

and the functions $u_{nm}(\mathbf{r})$ form a complete basis for representing fields in free space outside the minimum sphere $r = r_0$ that contains the sources. The plane wave representation

$$\begin{aligned} u_{nm}(\mathbf{r}) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} Y_{nm}\left(\hat{\mathbf{k}}\right) u_{\alpha\beta}(\mathbf{r}) \frac{d\alpha\, d\beta}{\gamma k} \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} Y_{nm}\left(\hat{\mathbf{k}}\right) \exp(i\mathbf{k}\cdot\mathbf{r}) \frac{d\alpha\, d\beta}{\gamma k}, \quad z > 0 \end{aligned} \tag{19.50}$$

can be established rigorously. See Eq. (19.24).

Vector solutions of Eqs. (19.4) and (19.6) follow directly from the scalar solutions (Eq. (19.48)):

$$\begin{aligned} \mathbf{m}_{nm}(\mathbf{r}) &= \frac{1}{k}\nabla \times \mathbf{n}_{nm}(\mathbf{r}) = \frac{1}{\sqrt{n(n+1)}}\mathbf{L}u_{nm}(\mathbf{r}) \\ \mathbf{n}_{nm}(\mathbf{r}) &= \frac{1}{k}\nabla \times \mathbf{m}_{nm}(\mathbf{r}) \end{aligned} \tag{19.51}$$

where the differential operator $\mathbf{L}$ is

$$\mathbf{L} = \frac{1}{i}\mathbf{r}\times\nabla = \frac{i}{\sin\theta}\hat{\boldsymbol{\theta}}\partial_\varphi - i\hat{\boldsymbol{\varphi}}\partial_\theta = \frac{i}{\sin\theta}\partial_\varphi\hat{\boldsymbol{\theta}} - \frac{i}{\sin\theta}\partial_\theta \sin\theta\hat{\boldsymbol{\varphi}} \tag{19.52}$$

That $\mathbf{m}_{nm}$ and $\mathbf{n}_{nm}$ are solutions of Eq. (19.6) is a consequence of the commutation relationships $\nabla^2\mathbf{L} = \mathbf{L}\nabla^2$ and $\nabla^2\nabla = \nabla\nabla^2$. Written out more fully,

$$\begin{aligned} \mathbf{m}_{nm}(\mathbf{r}) &= f_n(kr)\mathbf{X}^1_{nm}(\hat{\mathbf{r}}) \\ \mathbf{n}_{nm}(\mathbf{r}) &= g_n(kr)\mathbf{X}^2_{nm}(\hat{\mathbf{r}}) - \frac{\sqrt{n(n+1)}}{ikr} f_n(kr)\mathbf{X}^0_{nm}(\hat{\mathbf{r}}) \end{aligned} \tag{19.53}$$

with

$$\begin{aligned} f_n(x) &= i^n h_n(x) \underset{x\to\infty}{\sim} \frac{\exp(ix)}{ix} \\ g_n(x) &= \left[\frac{1}{ix}\frac{d}{dx}(xf_n)\right](x) \underset{x\to\infty}{\sim} \frac{\exp(ix)}{ix} \end{aligned} \tag{19.54}$$

The (vector) spherical harmonics [36, Chap. 9]

$$\begin{aligned} \mathbf{X}^0_{nm}(\hat{\mathbf{r}}) &= \hat{\mathbf{r}}Y_{nm}(\hat{\mathbf{r}}) \\ \mathbf{X}^1_{nm}(\hat{\mathbf{r}}) &= \frac{1}{\sqrt{n(n+1)}}\mathbf{L}Y_{nm}(\hat{\mathbf{r}}) = -\left(\hat{\boldsymbol{\theta}}\frac{m}{\sin\theta} + i\hat{\boldsymbol{\varphi}}\partial_\theta\right)\frac{Y_{nm}(\hat{\mathbf{r}})}{\sqrt{n(n+1)}} \\ \mathbf{X}^2_{nm}(\hat{\mathbf{r}}) &= i\hat{\mathbf{r}}\times\mathbf{X}^1_{nm}(\hat{\mathbf{r}}) \end{aligned} \tag{19.55}$$

form an orthonormal set

$$\int_0^{2\pi}\int_0^{\pi} \mathbf{X}_{\nu\mu}^{a*}(\hat{\mathbf{r}})\cdot\mathbf{X}_{nm}^{b}(\hat{\mathbf{r}})\sin\theta\, d\theta\, d\varphi = \delta_{ab}\delta_{\nu n}\delta_{\mu m} \tag{19.56}$$

We have the asymptotic formulas

$$\begin{aligned}\mathbf{m}_{nm}(\mathbf{r}) &\underset{r\to\infty}{\sim} \mathbf{X}_{nm}^{1}(\hat{\mathbf{r}})\frac{\exp(ikr)}{ikr}\\ \mathbf{n}_{nm}(\mathbf{r}) &\underset{r\to\infty}{\sim} \mathbf{X}_{nm}^{2}(\hat{\mathbf{r}})\frac{\exp(ikr)}{ikr}\end{aligned} \tag{19.57}$$

and the corresponding plane wave representations

$$\begin{aligned}\mathbf{m}_{nm}(\mathbf{r}) &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathbf{X}_{nm}^{1}(\hat{\mathbf{k}})\exp(i\mathbf{k}\cdot\mathbf{r})\frac{d\alpha\, d\beta}{\gamma k}\\ \mathbf{n}_{nm}(\mathbf{r}) &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathbf{X}_{nm}^{2}(\hat{\mathbf{k}})\exp(i\mathbf{k}\cdot\mathbf{r})\frac{d\alpha\, d\beta}{\gamma k},\quad z>0\end{aligned} \tag{19.58}$$

The electromagnetic fields of a test antenna (in the reference position) can be written as the superposition

$$\begin{aligned}\mathbf{E}(\mathbf{r}) &= a_0\sum_{n=1}^{\infty}\sum_{m=-n}^{n}\left[t_{nm}^{1}\mathbf{m}_{nm}(\mathbf{r})+t_{nm}^{2}\mathbf{n}_{nm}(\mathbf{r})\right]\\ iZ_0\mathbf{H}(\mathbf{r}) &= a_0\sum_{n=1}^{\infty}\sum_{m=-n}^{n}\left[t_{nm}^{2}\mathbf{m}_{nm}(\mathbf{r})+t_{nm}^{1}\mathbf{n}_{nm}(\mathbf{r})\right],\quad r>r_0\end{aligned} \tag{19.59}$$

There is no monopole term ($n=0$) in spherical wave expansions of the electromagnetic fields in free space. Using Eq. (19.57), we obtain the asymptotic formula

$$\mathbf{E}(\mathbf{r}) \underset{r\to\infty}{\sim} a_0\mathbf{t}_0(\hat{\mathbf{r}})\frac{\exp(ikr)}{ikr}$$

where the transmitting function is

$$\mathbf{t}_0(\hat{\mathbf{r}}) = \sum_{n=1}^{\infty}\sum_{m=-n}^{n}\left[t_{nm}^{1}\mathbf{X}_{nm}^{1}(\hat{\mathbf{r}})+t_{nm}^{2}\mathbf{X}_{nm}^{2}(\hat{\mathbf{r}})\right] \tag{19.60}$$

From Eq. (19.59),

$$\begin{aligned}\mathbf{E}_t(\mathbf{r}) &= \mathbf{E}(\mathbf{r})-\hat{\mathbf{r}}\hat{\mathbf{r}}\cdot\mathbf{E}(\mathbf{r})\\ &= a_0\sum_{n=1}^{\infty}\sum_{m=-n}^{n}\left[f_n(kr)\,t_{nm}^{1}\mathbf{X}_{nm}^{1}(\mathbf{r})+g_n(kr)\,t_{nm}^{2}\mathbf{X}_{nm}^{2}(\mathbf{r})\right]\end{aligned} \tag{19.61}$$

An application of the orthogonalities in Eq. (19.56) yields

$$t_{nm}^{1} = \frac{1}{f_n(kr)} \int_0^{2\pi} \int_0^{\pi} \mathbf{X}_{nm}^{1*}(\hat{\mathbf{r}}) \cdot \frac{\mathbf{E}(\mathbf{r})}{a_0} \sin\theta \, d\theta \, d\varphi$$
$$t_{nm}^{2} = \frac{1}{g_n(kr)} \int_0^{2\pi} \int_0^{\pi} \mathbf{X}_{nm}^{2*}(\hat{\mathbf{r}}) \cdot \frac{\mathbf{E}(\mathbf{r})}{a_0} \sin\theta \, d\theta \, d\varphi, \quad r = d > r_0 \tag{19.62}$$

Thus the transmitting pattern is determined by the tangential electric field on any sphere $r = d > r_0$.

19.2.4.2 Probe Response We begin with the test antenna and probe in their reference positions. In spherical scanning, the response of the probe is recorded as the probe is rotated over a measurement sphere of radius $r = r' = d$. First the probe is rotated by angle $\chi = \chi'$ about the laboratory z axis. Then it is rotated by angle $\theta = \theta'$ about the laboratory y axis. Finally, the probe is again rotated about the laboratory z axis by an angle $\varphi = \varphi'$. The members of the triplet (φ, θ, χ) are collectively known as *Euler angles*.† (The positive sense of these angles is determined, in the usual manner, with the "right-hand rule.") The angles φ' and θ' fix the location of the probe on the measurement sphere. The angle χ' specifies a rotation about the probe axis. When $\chi' = 0$, $\hat{\mathbf{y}}_0' = \hat{\boldsymbol{\varphi}}$, and when $\chi' = -\pi/2$, $\hat{\mathbf{y}}_0' = \hat{\boldsymbol{\theta}}$. See Figure 19.3.

In the laboratory coordinate system, position (location and orientation) of the probe is described by the parameter set $(r, \varphi, \theta, \chi)$ with the probe reference position corresponding to $(d, 0, 0, 0)$. The position of a similarly rotated test antenna is also described by these parameters with the test antenna reference position corresponding to $(0, 0, 0, 0)$.

Consider the output when the test antenna is at its reference position and the probe is at $(r', \varphi', \theta', \chi')$. The same output will be observed when the probe is at its reference position and the test antenna is at $(0, -\chi', -\theta', -\varphi',)$. This fact may be expressed mathematically as

$$w'(r', \varphi', \theta', \chi') = [\mathcal{P} \cdot \mathbf{E}](r', \varphi', \theta', \chi',) = \left[\mathcal{P} \cdot \tilde{\mathbf{E}}\right](r', 0, 0, 0) \tag{19.63}$$

Here $\mathcal{P}$ is a linear operator that represents the probe, $\mathbf{E}(\mathbf{r})$ is the electric field of the test antenna (at its reference position), and $\tilde{\mathbf{E}}(\mathbf{r})$ is the field of the rotated test antenna.

Now, let the "test antenna" radiate the vector spherical wave $\mathbf{m}_{nm}(\mathbf{r})$ or $\mathbf{n}_{nm}(\mathbf{r})$. As it turns out, rotated n-poles can be reexpressed as superpositions of multipole components with the same multipole index n:

$$\tilde{\mathbf{m}}_{nm}(\mathbf{r}) = \sum_{\mu=-n}^{n} \mathbf{m}_{n\mu}(\mathbf{r}) D_{\mu m}^{n}(-\chi', -\theta', -\varphi')$$
$$\tilde{\mathbf{n}}_{nm}(\mathbf{r}) = \sum_{\mu=-n}^{n} \mathbf{n}_{n\mu}(\mathbf{r}) D_{\mu m}^{n}(-\chi', -\theta', -\varphi') \tag{19.64}$$

†The reader should be aware that several definitions of Euler angles are commonly used.

We apply the Euler angles as rotations about axes fixed in the laboratory coordinate system. An equivalent result may be obtained by rotating about axes fixed in the body being moved. Begin with the laboratory and body coordinate systems coincident: first rotate by angle φ about the body z axis; then rotate by angle θ about the body y axis; finally, rotate by angle χ about the body z axis.

We discuss the rotation functions $D^n_{\mu m}$ later. From Eqs. (19.63) and (19.64)

$$w^{1'}_{nm}\left(r', \varphi', \theta', \chi'\right) = \left[\mathcal{P} \cdot \tilde{\mathbf{m}}_{nm}\right]\left(r', 0, 0, 0\right) = \sum_{\mu=-n}^{n} \sigma^{1'}_{n\mu}(r') D^n_{\mu m}(-\chi', -\theta', -\varphi')$$

$$w^{2'}_{nm}\left(r', \varphi', \theta', \chi'\right) = \left[\mathcal{P} \cdot \tilde{\mathbf{n}}_{nm}\right]\left(r', 0, 0, 0\right) = \sum_{\mu=-n}^{n} \sigma^{2'}_{n\mu}(r') D^n_{\mu m}(-\chi', -\theta', -\varphi') \tag{19.65}$$

The translated probe coefficients are

$$\begin{aligned} \sigma^{1'}_{n\mu}(r) &= \left[\mathcal{P} \cdot \mathbf{m}_{n\mu}\right](r, 0, 0, 0) \\ \sigma^{2'}_{n\mu}(r) &= \left[\mathcal{P} \cdot \mathbf{n}_{n\mu}\right](r, 0, 0, 0) \end{aligned} \tag{19.66}$$

Thus the response of the probe to the modes $\mathbf{m}_{nm}(\mathbf{r})$ or $\mathbf{n}_{nm}(\mathbf{r})$ can be summarized in terms of the $2n + 1$ numbers $\sigma^{1'}_{n\mu}$ or $\sigma^{2'}_{n\mu}$ that are the responses of the probe, in its reference position, to the n-pole component functions.

Appendix C.2 discusses the evaluation of the translated probe coefficients (Eq. (19.66)), which depend on the known receiving function of the probe.

19.2.4.3 Jensen Transmission Equation In Eq. (19.59) we have a modal expansion corresponding to Eq. (19.1). In Eq. (19.65) we have the response of the probe to each mode corresponding to Eq. (19.2). Finally, corresponding to Eq. (19.3), we have the *Jensen transmission formula* [6]

$$\frac{w'(r, \varphi, \theta, \chi)}{a_0} = \sum_{n=1}^{\infty} \sum_{m=-n}^{n} \sum_{\mu=-n}^{n} \left[\sigma^{1'}_{n\mu}(r) t^1_{nm} + \sigma^{2'}_{n\mu}(r) t^2_{nm}\right] D^n_{\mu m}(-\chi, -\theta, -\varphi) \tag{19.67}$$

A full discussion of the rotation functions[†]

$$D^n_{\mu m}(-\chi, -\theta, -\varphi) = \exp(i\mu\chi) d^n_{\mu m}(-\theta) \exp(im\varphi) \tag{19.68}$$

is beyond the scope of this chapter. The orthogonalities of the $D^n_{\mu m}$ can be used to solve for the unknown $t^{1,2}_{nm}$. In particular, were the translated probe coefficients $\sigma^{1'}_{n\mu}$ and $\sigma^{2'}_{n\mu}$ significant for $|\mu| \leq N_p$, then measurements would be required at $2N_P + 1$ values of χ for each (φ, θ). In order to require only two spin values per direction and to retain the advantages of orthogonality, standard practice employs special probes [26] designed so that $\sigma^{1'}_{n\mu} = \sigma^{2'}_{n\mu} = 0$ unless $\mu = \pm 1$.

Consider such a $\mu = \pm 1$ *probe* and define the *measurement vector*

$$\mathbf{w}'(\mathbf{r}) = w'_\theta(r, \varphi, \theta)\hat{\boldsymbol{\theta}} + w'_\varphi(r, \varphi, \theta)\hat{\boldsymbol{\varphi}} \tag{19.69}$$

[†]We define the rotation functions according to the convention of Rose [37] and Tinkham [38]. To use the convention of Edmonds [39], substitute $D^n_{\mu m}(\varphi, \theta, \chi)$ for $D^n_{\mu m}(-\chi, -\theta, -\varphi)$. To use the convention of Hansen [26], substitute $D^n_{\mu m}(\chi, \theta, \varphi)$ for $D^n_{\mu m}(-\chi, -\theta, -\varphi)$.

$$\begin{aligned} w'_{\varphi}(r,\varphi,\theta) &= w'(r,\varphi,\theta,\chi=0) \\ w'_{\theta}(r,\varphi,\theta) &= w'(r,\varphi,\theta,\chi=-\pi/2) \end{aligned} \tag{19.70}$$

With

$$d^{n}_{\pm 1,m}(-\theta)\exp(im\varphi) = \sqrt{\frac{4\pi}{2n+1}}\left(\hat{\boldsymbol{\theta}} \mp i\hat{\boldsymbol{\varphi}}\right)\cdot \mathbf{X}^{1}_{nm}(\hat{\mathbf{r}}) \tag{19.71}$$

the Jensen transmission formula Eq. (19.67) can be written [23, Eq. (27)]

$$\frac{\mathbf{w}'(\mathbf{r})}{a_0} = \sum_{n=1}^{\infty}\sum_{m=-n}^{n}\left[T^{1\prime}_{nm}(r)\,\mathbf{X}^{1}_{nm}(\hat{\mathbf{r}}) + T^{2\prime}_{nm}(r)\,\mathbf{X}^{2}_{nm}(\hat{\mathbf{r}})\right] \tag{19.72}$$

where in matrix form

$$\begin{pmatrix} T^{1\prime}_{nm} \\ T^{2\prime}_{nm} \end{pmatrix} = -i\sqrt{\frac{4\pi}{2n+1}}\begin{pmatrix} \sigma^{1\prime}_{n1} - \sigma^{1\prime}_{n,-1} & \sigma^{2\prime}_{n1} - \sigma^{2\prime}_{n,-1} \\ \sigma^{1\prime}_{n1} + \sigma^{1\prime}_{n,-1} & \sigma^{2\prime}_{n1} + \sigma^{2\prime}_{n,-1} \end{pmatrix}\begin{pmatrix} t^{1}_{nm} \\ t^{2}_{nm} \end{pmatrix} \tag{19.73}$$

Equations (19.67) and (19.72) are valid when multiple reflections can be ignored and when the probe and test antenna can be physically separated by a plane perpendicular to $\hat{\mathbf{r}}$.

Application of the orthogonalities in Eq. (19.56) gives

$$\begin{aligned} T^{1\prime}_{nm} &= \int_0^{2\pi}\int_0^{\pi} \mathbf{X}^{1*}_{nm}(\hat{\mathbf{r}})\cdot\frac{\mathbf{w}'(\mathbf{r})}{a_0}\sin\theta\,d\theta\,d\varphi \\ T^{2\prime}_{nm} &= \int_0^{2\pi}\int_0^{\pi} \mathbf{X}^{2*}_{nm}(\hat{\mathbf{r}})\cdot\frac{\mathbf{w}'(\mathbf{r})}{a_0}\sin\theta\,d\theta\,d\varphi, \quad r=d \end{aligned} \tag{19.74}$$

Finally, Eq. (19.73) provides two equations that can be solved for t^{1}_{nm} and t^{2}_{nm}:

$$\begin{pmatrix} t^{1}_{nm} \\ t^{2}_{nm} \end{pmatrix} = \frac{i\sqrt{\dfrac{2n+1}{16\pi}}}{\sigma^{1\prime}_{n1}\sigma^{2\prime}_{n,-1} - \sigma^{1\prime}_{n,-1}\sigma^{2\prime}_{n1}}\begin{pmatrix} \sigma^{2\prime}_{n1} + \sigma^{2\prime}_{n,-1} & -\sigma^{2\prime}_{n1} + \sigma^{2\prime}_{n,-1} \\ -\sigma^{1\prime}_{n1} - \sigma^{1\prime}_{n,-1} & \sigma^{1\prime}_{n1} - \sigma^{1\prime}_{n,-1} \end{pmatrix}\begin{pmatrix} T^{1\prime}_{nm} \\ T^{2\prime}_{nm} \end{pmatrix} \tag{19.75}$$

Thus the transmitting function is determined by data $\mathbf{w}'(\mathbf{r})/a_0$ measured on the sphere $r = d$. Unlike planar and cylindrical scanning, where the second probe may be distinct, the present formulation of spherical scanning assumes that one probe is used in two orientations. If this probe is circularly polarized, then Eqs. (19.73) will not be independent; however, if a complete set of measurement data is obtained for both a right- and a left-circularly polarized probe, then the resulting set of four equations can be solved for t^{1}_{nm} and t^{2}_{nm} in a least-squares sense.

An example $\mu = \pm 1$ probe is the conceptual *ideal probe* that directly measures a component of the transverse electric field; however, physical $\mu = \pm 1$ probes are not difficult to construct (see Appendix A.2). In fact, any practical probe will approach a $\mu = \pm 1$ probe as the separation between test antenna and probe increases.

Formula (19.71) can be derived from first principles by forcing Eqs. (19.61) and (19.67) to agree when the probe is ideal.

19.2.5 Cylindrical Scanning

19.2.5.1 Cylindrical Wave Representation Scalar solutions of Eq. (19.6), separated in cylindrical coordinates (Figure 19.4), are given by the cylindrical waves

$$u_{n\beta}(\mathbf{r}) = \frac{1}{\sqrt{2\pi}} F_n(\kappa\rho)\, \Psi_n(\hat{\boldsymbol{\rho}}) \exp(i\beta y) \tag{19.76}$$

$$F_n(x) = i^n \sqrt{\frac{\pi}{2}} H_n^{(1)}(x)$$

$$\Psi_n(\hat{\boldsymbol{\rho}}) = \frac{1}{\sqrt{2\pi}} \left(\frac{z + ix}{\rho}\right)^n = \frac{\exp(in\Phi)}{\sqrt{2\pi}} \tag{19.77}$$

Here, the $H_n^{(1)}$ are cylindrical Hankel functions of the first kind [40, Chap. 9] and

$$\begin{aligned} \rho &= \sqrt{r^2 - y^2} \\ \kappa &= \sqrt{k^2 - \beta^2} \\ \hat{\boldsymbol{\rho}} &= \frac{z\hat{\mathbf{z}} + x\hat{\mathbf{x}}}{\rho} \end{aligned} \tag{19.78}$$

In order to keep the reference positions of the test antenna and the probe on the z axis, we take the y axis as the axis of cylinders of constant ρ. Cylindrical waves propagate or decay away from the y axis when κ is real ($k > \beta$) or imaginary ($k < \beta$), respectively. If sources are confined to the region $\rho < \rho_0$, then the functions $u_{n\beta}$ form a complete basis for representing fields in free space $\rho > \rho_0$. The plane wave decomposition

$$\begin{aligned} u_{n\beta}(\mathbf{r}) &= \int_{-\infty}^{\infty} \Psi_n(\hat{\boldsymbol{\kappa}}) u_{\alpha\beta}(\mathbf{r}) \frac{d\alpha}{\gamma} \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \Psi_n(\hat{\boldsymbol{\kappa}}) \exp(i\mathbf{k}\cdot\mathbf{r}) \frac{d\alpha}{\gamma}, \quad z > 0 \end{aligned} \tag{19.79}$$

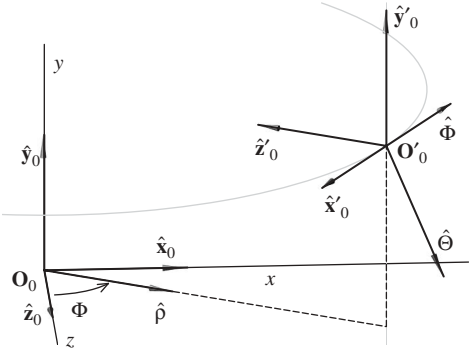


Figure 19.4 Cylindrical scanning geometry. The scan cylinder is $\rho = d$. The test antenna is in its reference position (see Figure 19.1). The probe is shown at the position (ρ, Φ, y) with $\rho = d$, $\Phi = \pi/3$, and $y = d/2$. Furthermore, $\hat{\mathbf{x}}'_0 = -\hat{\boldsymbol{\Phi}}$, $\hat{\mathbf{y}}'_0 = \hat{\mathbf{y}}$, $\hat{\mathbf{z}}'_0 = -\hat{\boldsymbol{\rho}}$, and $\hat{\boldsymbol{\Theta}} = \hat{\boldsymbol{\Phi}} \times \hat{\mathbf{r}}$.

where

$$\hat{\boldsymbol{\kappa}} = \frac{\gamma \hat{\mathbf{z}} + \alpha \hat{\mathbf{x}}}{\kappa} \tag{19.80}$$

can be established rigorously [41, Eqs. (3.753) and (3.754)]. See Eq. (19.24).

Vector solutions of Eqs. (19.4) and (19.6) follow directly from the scalar solutions of Eq. (19.76):

$$\begin{aligned} \mathbf{m}_{n\beta}(\mathbf{r}) &= \frac{1}{k}\nabla \times \mathbf{n}_{n\beta}(\mathbf{r}) = \frac{1}{i\kappa}\hat{\mathbf{y}} \times \nabla u_{n\beta} \\ &= \frac{1}{\sqrt{2\pi}}\left[-\frac{nF_n(\kappa\rho)}{\kappa\rho}\hat{\boldsymbol{\rho}} + G_n(\kappa\rho)\hat{\boldsymbol{\Phi}}\right]\Psi_n(\hat{\boldsymbol{\rho}})\exp(i\beta y) \end{aligned} \tag{19.81}$$

$$\begin{aligned} \mathbf{n}_{n\beta}(\mathbf{r}) &= \frac{1}{k}\nabla \times \mathbf{m}_{n\beta}(\mathbf{r}) = -\frac{\beta}{ik}\hat{\mathbf{y}} \times \mathbf{m}_{n\beta}(\mathbf{r}) - \frac{\kappa}{ik}u_{n\beta}(\mathbf{r})\hat{\mathbf{y}} \\ &= \frac{1}{\sqrt{2\pi}}\left[\frac{\beta}{ik}G_n(\kappa\rho)\hat{\boldsymbol{\rho}} + \frac{\beta}{ik}\frac{nF_n(\kappa\rho)}{\kappa\rho}\hat{\boldsymbol{\Phi}} - \frac{\kappa}{ik}F_n(\kappa\rho)\hat{\mathbf{y}}\right]\Psi_n(\hat{\boldsymbol{\rho}})\exp(i\beta y) \end{aligned} \tag{19.82}$$

with

$$\begin{aligned} F_n(x) &= i^n\sqrt{\frac{\pi}{2}}H_n^{(1)}(x) \underset{x\to\infty}{\sim} \frac{\exp(ix)}{\sqrt{ix}} \\ G_n(x) &= \frac{1}{i}\frac{d}{dx}F_n(x) \underset{x\to\infty}{\sim} \frac{\exp(ix)}{\sqrt{ix}} \end{aligned} \tag{19.83}$$

From Eqs. (19.79), (19.81), and (19.82), we have the plane wave decompositions

$$\begin{aligned} \mathbf{m}_{n\beta}(\mathbf{r}) &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\boldsymbol{\Psi}_n^1(\hat{\mathbf{k}})\exp(i\mathbf{k}\cdot\mathbf{r})\frac{d\alpha}{\gamma} \\ \mathbf{n}_{n\beta}(\mathbf{r}) &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\boldsymbol{\Psi}_n^2(\hat{\mathbf{k}})\exp(i\mathbf{k}\cdot\mathbf{r})\frac{d\alpha}{\gamma}, \quad z > 0 \end{aligned} \tag{19.84}$$

The angular functions are

$$\begin{aligned} \boldsymbol{\Psi}_n^1(\hat{\mathbf{r}}) &= \hat{\mathbf{y}} \times \hat{\boldsymbol{\rho}}\Psi_n(\hat{\boldsymbol{\rho}}) = \Psi_n(\hat{\boldsymbol{\rho}})\hat{\boldsymbol{\Phi}} \\ \boldsymbol{\Psi}_n^2(\hat{\mathbf{r}}) &= i\hat{\mathbf{r}} \times \boldsymbol{\Psi}_n^1(\hat{\boldsymbol{\rho}}) = -i\Psi_n(\hat{\boldsymbol{\rho}})\hat{\boldsymbol{\Theta}} \end{aligned} \tag{19.85}$$

The electromagnetic fields of a test antenna (in the reference position) can be written as the superposition

$$\begin{aligned} \mathbf{E}(\mathbf{r}) &= a_0\sum_{n=-\infty}^{\infty}\int_{-\infty}^{\infty}\left[t_{n\beta}^1\mathbf{m}_{n\beta}(\mathbf{r}) + t_{n\beta}^2\mathbf{n}_{n\beta}(\mathbf{r})\right]\frac{d\beta}{k} \\ iZ_0\mathbf{H}(\mathbf{r}) &= a_0\sum_{n=-\infty}^{\infty}\int_{-\infty}^{\infty}\left[t_{n\beta}^2\mathbf{m}_{n\beta}(\mathbf{r}) + t_{n\beta}^1\mathbf{n}_{n\beta}(\mathbf{r})\right]\frac{d\beta}{k}, \quad \rho > \rho_0 \end{aligned} \tag{19.86}$$

A substitution of Eq. (19.84) yields the plane wave expansion

$$\mathbf{E}(\mathbf{r}) = \frac{a_0}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathbf{t}_0(\hat{\mathbf{k}})\exp(i\mathbf{k}\cdot\mathbf{r})\frac{d\alpha\, d\beta}{\gamma k}, \quad z > \rho_0$$

where the transmitting function is represented by the Fourier series

$$\mathbf{t}_0(\hat{\mathbf{r}}) = \sum_{n=-\infty}^{\infty}\left[t^1_{n,ky/r}\boldsymbol{\Psi}^1_n(\hat{\mathbf{r}}) + t^2_{n,ky/r}\boldsymbol{\Psi}^2_n(\hat{\mathbf{r}})\right] \tag{19.87}$$

By inverting the Fourier transform and applying the orthogonality

$$\int_0^{2\pi}\Psi^*_\nu(\hat{\boldsymbol{\rho}})\Psi_n(\hat{\boldsymbol{\rho}})d\Phi = \delta_{\nu n} \tag{19.88}$$

we can solve Eq. (19.86) for the cylindrical-mode coefficients

$$\begin{aligned} t^1_{n\beta} &= \frac{1}{G_n(\kappa\rho)}\frac{k}{2\pi}\int_0^{2\pi}\int_{-\infty}^{\infty}\left(\hat{\boldsymbol{\Phi}} + \frac{n\beta}{\kappa^2\rho}\hat{\mathbf{y}}\right)\cdot\frac{\mathbf{E}(\mathbf{r})}{a_0}\exp(-in\Phi - i\beta y)\, dy\, d\Phi \\ t^2_{n\beta} &= \frac{1}{F_n(\kappa\rho)}\frac{k}{2\pi}\int_0^{2\pi}\int_{-\infty}^{\infty}\frac{k}{i\kappa}\hat{\mathbf{y}}\cdot\frac{\mathbf{E}(\mathbf{r})}{a_0}\exp(-in\Phi - i\beta y)\, dy\, d\Phi, \quad \rho = d > \rho_0 \end{aligned} \tag{19.89}$$

Thus the transmitting function is determined by the tangential electric field on any cylinder $\rho = d > \rho_0$.

19.2.5.2 Probe Response We begin with the test antenna and probe in their reference positions. In cylindrical scanning, the response of the probe is recorded as the probe is translated and rotated over a measurement cylinder of radius $\rho = \rho' = d$. The probe is translated in the $\hat{\mathbf{y}}$ direction by a distance $y = y'$. Then the probe is rotated about the y axis by an angle $\Phi = \Phi'$. (The positive sense of Φ is determined, in the usual manner, with the "right-hand rule.") See Figure 19.4.

In the laboratory coordinate system, position (location and orientation) of the probe is described by the parameter set (ρ, Φ, y) with the probe reference position corresponding to $(d, 0, 0)$. The position of a similarly rotated and translated test antenna is also described by these parameters with the test antenna reference position corresponding to $(0, 0, 0)$.

Consider the output with the test antenna at its reference position and the probe at (ρ', Φ', y'). The same output will be observed with the probe at its reference position and the test antenna at $(0, -\Phi', -y')$. This fact may be expressed mathematically as

$$w'(\mathbf{r}') = [\boldsymbol{\mathcal{P}}\cdot\mathbf{E}](\rho', \Phi', y') = \left[\boldsymbol{\mathcal{P}}\cdot\tilde{\mathbf{E}}\right](\rho', 0, 0) \tag{19.90}$$

Here $\boldsymbol{\mathcal{P}}$ is a linear operator that represents the probe, $\mathbf{E}(\mathbf{r})$ is the field of the test antenna (at its reference position), and $\tilde{\mathbf{E}}(\mathbf{r})$ is the field of the rotated and translated test antenna.

Now, let the "test antenna" radiate the vector cylindrical wave $\mathbf{m}_{n\beta}$ or $\mathbf{n}_{n\beta}$. Then,

$$\begin{aligned} \tilde{\mathbf{m}}_{n\beta}(\mathbf{r}) &= \mathbf{m}_{n\beta}(\rho, \Phi + \Phi', y + y') = \mathbf{m}_{n\beta}(\mathbf{r})\exp(in\Phi' + i\beta y') \\ \tilde{\mathbf{n}}_{n\beta}(\mathbf{r}) &= \mathbf{n}_{n\beta}(\rho, \Phi + \Phi', y + y') = \mathbf{n}_{n\beta}(\mathbf{r})\exp(in\Phi' + i\beta y') \end{aligned} \tag{19.91}$$

From Eqs. (19.90) and (19.91),

$$\begin{aligned} w_{n\beta}^{1'}(\mathbf{r}') &= \left[\boldsymbol{\mathcal{P}} \cdot \tilde{\mathbf{m}}_{n\beta}\right](\rho', 0, 0) = \sigma_{n\beta}^{1'}(\rho') \exp(in\Phi' + i\beta y') \\ w_{n\beta}^{2'}(\mathbf{r}') &= \left[\boldsymbol{\mathcal{P}} \cdot \tilde{\mathbf{n}}_{n\beta}\right](\rho', 0, 0) = \sigma_{n\beta}^{2'}(\rho') \exp(in\Phi' + i\beta y') \end{aligned} \tag{19.92}$$

The translated probe coefficients are

$$\begin{aligned} \sigma_{n\beta}^{1'}(\rho) &= \left[\boldsymbol{\mathcal{P}} \cdot \mathbf{m}_{n\beta}\right](\rho, 0, 0) \\ \sigma_{n\beta}^{2'}(\rho) &= \left[\boldsymbol{\mathcal{P}} \cdot \mathbf{n}_{n\beta}\right](\rho, 0, 0) \end{aligned} \tag{19.93}$$

Thus $w_{n\beta}^{1'}(\mathbf{r}')$ and $w_{n\beta}^{2'}(\mathbf{r}')$ depend on the transverse location of the probe through a simple phase factor.

Appendix C.3 discusses the evaluation of the translated probe coefficients in Eq. (19.93), which depend on the known receiving function of the probe.

19.2.5.3 Transmission Equation In Eq. (19.86) we have a modal expansion corresponding to Eq. (19.1). In Eq. (19.92) we have the response of the probe to each mode corresponding to Eq. (19.2). Finally, corresponding to Eq. (19.3), we have the transmission formula [10–12]

$$\frac{w'(\mathbf{r})}{a_0} = \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\sigma_{n\beta}^{1'}(\rho) t_{n\beta}^{1} + \sigma_{n\beta}^{2'}(\rho) t_{n\beta}^{2}\right] \exp(in\Phi + i\beta y) \frac{d\beta}{k} \tag{19.94}$$

(Brown and Jull [5] earlier developed a two-dimensional form of this equation.) Equation (19.94) is valid when multiple reflections can be ignored and when the probe and test antenna can be physically separated by a plane perpendicular to $\hat{\boldsymbol{\rho}}$.

Inversion of the Fourier transform and use of the orthogonality in Eq. (19.88) gives

$$\begin{aligned} T_{n\beta}'(\rho) &= \sigma_{n\beta}^{1'} t_{n\beta}^{1} + \sigma_{n\beta}^{2'} t_{n\beta}^{2} \\ &= \frac{k}{4\pi^2} \int_0^{2\pi} \int_{-\infty}^{\infty} \frac{w'(\mathbf{r})}{a_0} \exp(-in\Phi - i\beta y)\, dy\, d\Phi, \quad \rho = d \end{aligned} \tag{19.95}$$

Since $T_{n\beta}'$ can be computed from data $w'(\mathbf{r})/a_0$ measured on the cylinder $\rho = d$, Eq. (19.95) provides one equation for the two unknown components of $\mathbf{t}_0(\hat{\mathbf{k}})$. Data from a second probe are needed to provide sufficient information to solve for $\mathbf{t}_0(\hat{\mathbf{k}})$. With two probes, we have the linear system

$$\begin{pmatrix} T_{n\beta}' \\ T_{n\beta}'' \end{pmatrix} = \begin{pmatrix} \sigma_{n\beta}^{1'} & \sigma_{n\beta}^{2'} \\ \sigma_{n\beta}^{1''} & \sigma_{n\beta}^{2''} \end{pmatrix} \begin{pmatrix} t_{n\beta}^{1} \\ t_{n\beta}^{1} \end{pmatrix} \tag{19.96}$$

Here, quantities associated with the second probe are marked with a double prime. Thus

$$\begin{pmatrix} t_{n\beta}^{1} \\ t_{n\beta}^{1} \end{pmatrix} = \frac{1}{\sigma_{n\beta}^{1'}\sigma_{n\beta}^{2''} - \sigma_{n\beta}^{1''}\sigma_{n\beta}^{2'}} \begin{pmatrix} \sigma_{n\beta}^{2''} & -\sigma_{n\beta}^{2'} \\ -\sigma_{n\beta}^{1''} & \sigma_{n\beta}^{1'} \end{pmatrix} \begin{pmatrix} T_{n\beta}' \\ T_{n\beta}'' \end{pmatrix} \tag{19.97}$$

If the first probe is linearly polarized, then a "second" probe may be realized by rotating the first probe by 90° about its z axis.

19.3 NEAR-FIELD MEASUREMENT PRACTICE

19.3.1 General Considerations for Near-Field Scanning

19.3.1.1 Overview Here (Section 19.3.1) we discuss practical issues that are substantially the same for all implementations of near-field scanning methods. Later sections deal with geometry-specific considerations: planar scanning (Section 19.3.2), spherical scanning (Section 19.3.3), and cylindrical scanning (Section 19.3.4).

19.3.1.2 Implementation

RF Equipment Transformation from near field to far field requires complex data (both amplitude and phase information). Figure 19.5 shows a typical antenna measurement system. Key components include a phase/amplitude receiver and a phase-stable signal source, such as that provided by a synthesizer. Because cables may be long, remote mixing is often used to reduce attenuation.

Alignment Equipment Requirements and techniques vary from geometry to geometry, but the basic alignment tools are the same. These include laser straightness interferometers, electronic levels, and theodolites (with autocollimation capability). A laser tracker can greatly reduce the time required by and improve the accuracy of some procedures.

Probe Characterization To evaluate the far-field properties of the test antenna from near-field measurements, it is necessary to know the far-field properties of a reference antenna (probe). These properties (gain, polarization, and pattern) are usually measured with traditional far-field techniques, which are not in the scope of this chapter. See Refs. 42–44.

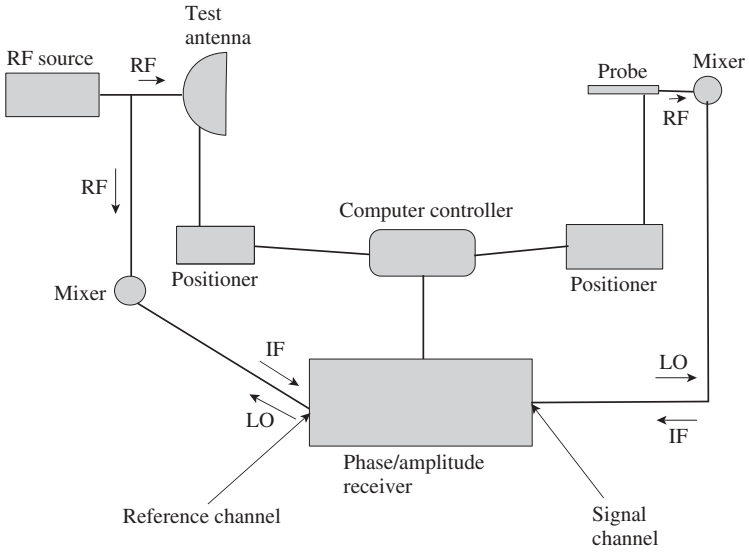


Figure 19.5 Typical components of a near-field measurement system.

Absorber and Absorber Placement The use of a good quality microwave absorber helps to reduce the effects of stray signals. As discussed previously, the planar technique is best used for moderately to highly directive antennas. Since the antenna under test (AUT) is fixed in the laboratory, it is usually sufficient to place absorber only in front of the test antenna. The mounting structure behind the probe is especially critical. More absorber coverage is needed for nondirective and/or rotating test antennas. For many cylindrical and spherical scanning implementations, a full anechoic chamber is required. Typically, a pyramidal absorber is used. The optimal size of the pyramids depends on the operational frequency range.

Microwave absorber (even that treated with fire retardant salts) is flammable [45]. Once ignited, absorber burns hot and produces many toxic gases, including cyanide, carbon monoxide, carbon dioxide, and hydrochloric acid. Potential ignition sources should be treated with care. These include high power RF radiation (beyond the rating of the absorber), soldering and welding irons, and high intensity lights.

Cross-Talk and Leakage Cross-talk is the unwanted coupling between two ports of a measurement system or subsystem. Examples include signals traveling from the reference channel to the measurement channel of the receiver; signals traveling from the x polarization port to the y polarization port of a dual-port probe; and signals traveling from the LO (local oscillator) port to the RF port of a remote mixer. A good receiver will have 80 dB or better of isolation between the reference and measurement channels. Cross-talk problems are often due to poor isolation in switches, mixers, isolators, or dual-port probes. The direct way to test for cross-talk is to insert a signal into one port and measure the output at the other port.

Leakage is unwanted coupling due to imperfections in the cables, connectors, and joints used to link the components of a measurement system. One technique of checking for leakage is to make three near-field scans: a reference scan is taken with the AUT transmitting and the probe receiving the signal; a second scan is taken with the transmit signal terminated at the antenna and a third is taken with the receive signal terminated. The second scan detects leakage in the circuit preceding the transmitting antenna The third scan detects leakage in the circuit following the receiving antenna. (See Figure 19.6.)

The bias is the average leakage signal over the scan area. A nonzero bias can lead to a spurious peak in the z direction.

A cross-talk/leakage contribution that is 20 dB lower than a signal of interest results in a measurement uncertainty of about 1 dB.

Cable Flexing Ideally, the stressing of cables, connectors, and rotary joints should have no effect on a measurement. Practically, of course, there will always be some undesirable variation in amplitude and (especially) phase. The severity of the problem depends on the quality and condition of the components. If the errors are repeatable, one can use the reflectometer setup illustrated in Figure 19.7 to estimate the effect of cable flexing and to correct for it. In this procedure, the probe (or test antenna) is replaced with a reflectometer. The reflected signal is measured with respect to some reference point on the scan surface, where for convenience the amplitude and phase are set to 1.0 and 0°, respectively. The measured signal at other points on the surface will be given by $1 + 2\varepsilon$ for the amplitude and $2\delta\phi$ for the phase, where ε is the amplitude error and $\delta\phi$ is the phase error of the one-way path. An alternative method of measuring these errors is to use the three-cable method of Hess [46]. This method involves the use of three cables (with unknown properties) in pairs to determine the effect of cable flexing.

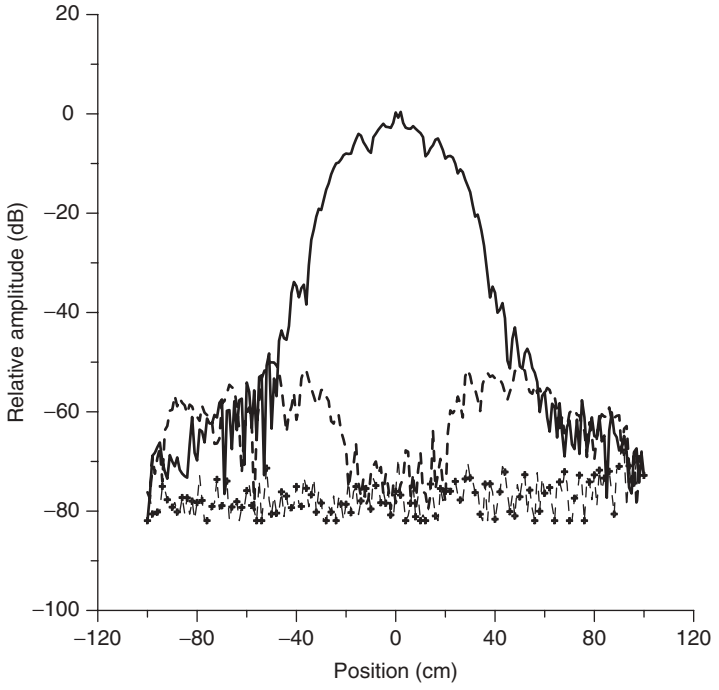


Figure 19.6 Sample leakage test results. The solid line represents the measured near field of the test antenna. The dashed line is a test result showing significant leakage problems. The $-+-+$ line is a test result showing more acceptable leakage levels.

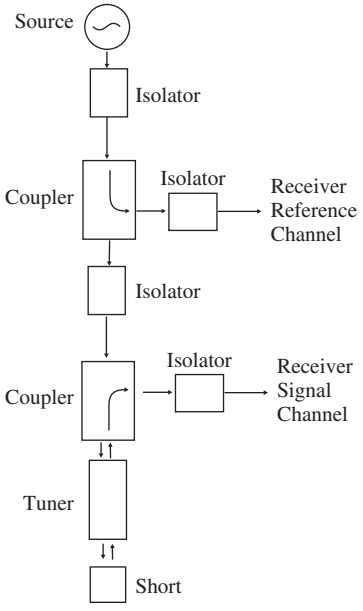


Figure 19.7 Reflectometer setup for estimating errors due to cable flexing.

Noise and Dynamic Range Parseval's theorem states that the root sum square (RSS) noise in the near field is equal to the RSS noise in the far field. The far-field peak is greater than the near-field peak by approximately the gain of the antenna. This implies that for a 30-dB gain antenna, if the noise level were at −40 dB relative to the near-field peak, we would expect it to be at about −70 dB relative to the far-field peak. A noise level that is 20 dB lower than a signal of interest results in a measurement uncertainty of about 1 dB.

Dynamic range is affected, not only by the noise level, but also by receiver saturation at high signal levels.

Nonlinearity Many receivers that have been manufactured, since about 1990, and used for antenna measurements are fairly linear in their optimal operating ranges. However, other receivers, especially those manufactured earlier, may exhibit significant nonlinearity. The effects of nonlinearity can be mitigated through comparison with a calibrated rotary vane or step attenuator. The system linearity can be evaluated by inserting a calibrated rotary vane attenuator. The setting of the attenuator is varied and the receiver output compared to the calibrated value of the attenuator. One obtains results similar to those in Table 19.2, where measurements are typically made relative to some reference point in the near field. (Table 19.2 is not intended to represent any particular receiving system.) In Table 19.2, column 2 represents the best value of the attenuation for the attenuator setting in column 1. Column 3 is the receiver reading for that attenuation. For example, for an attenuator setting of 30 dB, the calibrated attenuation is 30.02 dB and the receiver reading is −29.96 dB. That is, the receiver reading is 0.06 dB high since it should be −30.02 dB. This information can be used to generate a correction to the receiver output to obtain a calibrated value for each near-field data point.

Many modern devices, however, have better linearity than the rotary vane attenuator. In this case, it makes no sense to construct and use a table to correct the receiver output;

TABLE 19.2 Attenuator Calibration

Attenuator Setting (dB)	Calibrated Attenuator Value (dB)	Receiver Reading (dB)
0	0	0
2	2.00	− 2.00
4	4.00	− 4.00
6	6.01	− 6.01
8	8.01	− 8.01
10	10.01	−10.01
12	12.00	−12.00
14	14.00	−14.00
16	15.99	−15.99
18	17.99	−17.99
20	19.99	−19.99
25	24.98	−24.99
30	30.02	−29.96
35	35.02	−35.08
40	40.03	−40.12

however, this linearity test is still useful to detect abnormalities due to, for example, a faulty or improperly biased mixer.

Normalization There are several methods of determining the normalization ratios (b_0'/a_0 and b_0''/a_0). (1) Measure the insertion loss (a measurement of the power with the probe antenna and test antenna inserted in the circuit relative to the power with the generator and load connected directly together without the antennas) at a near-field reference point, as illustrated in Figure 19.8. The insertion loss measurement requires multiple connections and disconnections, so a connector of high quality is required. (2) In the gain substitution method, a known reference antenna is scanned to calibrate the measurement system. In this method we do not need to know the probe gain. (3) If the ohmic losses are negligible or can be calculated, then the gain can be determined from the directivity, Eq. (19.21). (4) The normalization can be accomplished using absolute power measurements.

To normalize the two polarization components relative to each other, it is necessary to refer them to a common signal. This can be done by measuring the signal at some near-field reference point of each component relative to the through connection (with the generator and load hooked directly together). Both the amplitude and phase must be determined in order to perform the probe correction accurately.

19.3.2 Planar Scanning Measurements

19.3.2.1 Overview of Planar Scanning In the planar scanning geometry, the test antenna is normally fixed and the probe antenna moves over the surface of a plane in front of the AUT. The probe and scanner act as an $x-y$ recorder. Because of truncation effects, the planar near-field scanning geometry is most useful for measuring moderately to highly directive antennas. For antennas with broad beams, the use of the planar near-field method is not advised since much of the energy will not be intercepted by a finite measurement

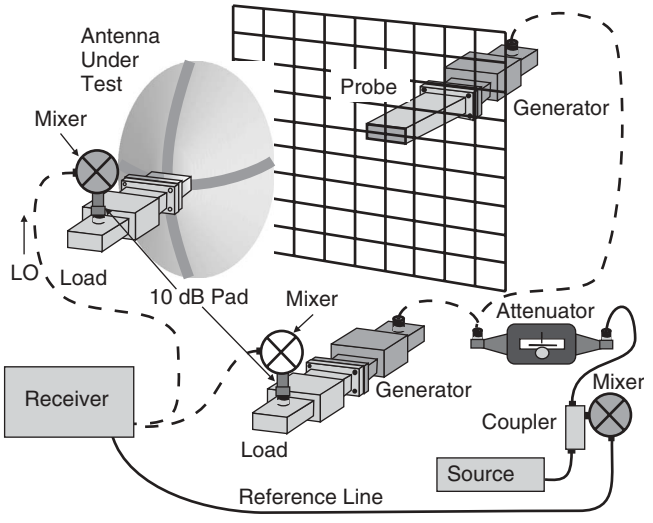


Figure 19.8 Illustration of the insertion loss measurement. The attenuator setting is usually higher for the generator-to-load connection to avoid nonlinearity effects and damage to the mixer.

region. Data are usually acquired "on the fly" in the scanning direction (e.g., the y direction). The probe is then stepped in the orthogonal direction (e.g., the x direction) and scanning reinitiated in the scanning direction.

19.3.2.2 Implementation

Alignment A properly aligned planar scanning measurement setup satisfies the following requirements. (1) The x and y axes are perpendicular. Furthermore, the plane over which the probe moves must be flat to within a small fraction of a wavelength (typically 0.01λ–0.02λ). (2) The probe z axis is perpendicular to the scan plane. The probe y axis must be parallel to the laboratory y axis. (3) The AUT coordinate system must be parallel to the laboratory coordinate system. (Translational displacement from the reference position produces only a phase factor in the far field, which is usually unimportant.)

The flatness required of the scan plane depends on the accuracy requirements and the nature of the error [47, Eqs. (53)–(61)]. For example, periodic position errors produce larger errors in the pattern, but these errors are concentrated in specific directions. The scan plane can be efficiently aligned or characterized by use of a laser tracker. Often, a mirror is positioned parallel to the scanner as a reference in the alignment of the probe and AUT.

The alignment of the probe may be accomplished by aligning a theodolite to the reference mirror mentioned above and using the autocollimation feature of the theodolite. Mirrors used for alignment should be of high quality with good optical surfaces and parallel front and back surfaces.

The AUT is generally aligned relative to an aligned probe.

Sampling Criteria Sampling considerations require that two data points be acquired for the shortest period in the measured data. Provided the measurements are acquired in a region where the evanescent (nonpropagating modes) are unimportant, increments of less than $\lambda/2$ in both x and y will generally satisfy the sampling criteria. However, there are error mechanisms that have periods that are less than a wavelength. In these cases, a sample spacing of less than $\lambda/2$ is necessary to prevent such errors from being aliased into the pattern.

Failure to satisfy the sampling criteria will lead to aliasing. One way to test for aliasing is to acquire data at intervals of much less than $\lambda/2$ (e.g., 0.1λ) and to check whether the pattern changes when fewer points (e.g., every other point) are used. Figure 19.9 shows how the root mean square error changes with spacing interval for a sample slotted waveguide array. Since this array had low sidelobes and measurements were required down to -60 dB, it was necessary to use spacing intervals of 0.4λ or less. The smaller spacing size is required because of the high frequency errors (see Figure 19.10) that occur due to probe–AUT multiple reflections. These errors are associated with the regular spacing between the slotted elements.

Scan Area Truncation The theory assumes an infinite scan plane and this is obviously not practical. Acquiring data over a finite area has two effects. First, the pattern beyond $\pm\theta_v$ is invalid [47], where

$$\theta_v = \arctan\left(\frac{L-a}{2d}\right) \tag{19.98}$$

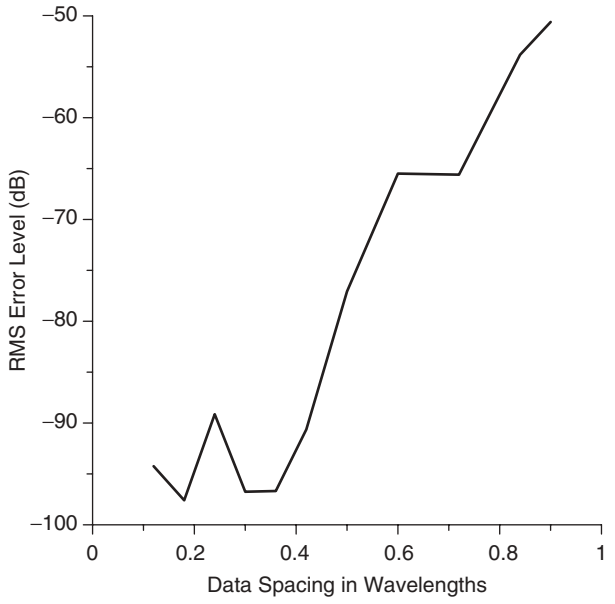


Figure 19.9 Example of the RMS error as a function of spacing size. Larger spacings lead to higher errors due to aliasing effects.

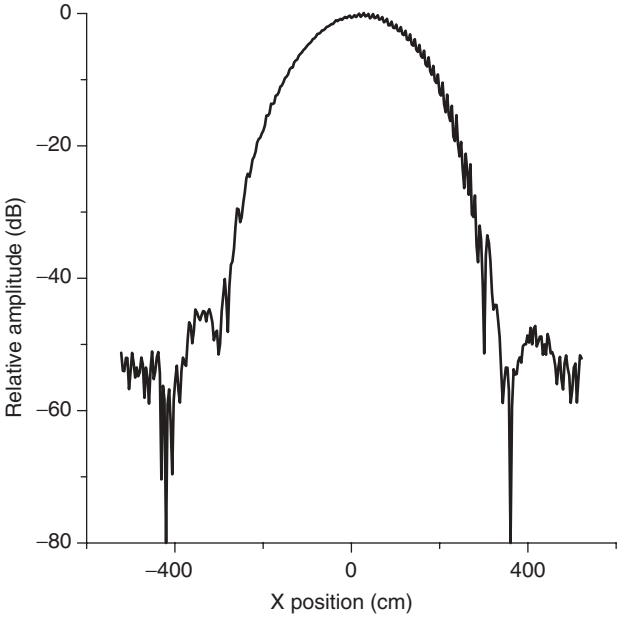


Figure 19.10 Near-field cut showing the ripple due to multiple reflections between the probe and the slots of the Ultralow Sidelobe Array (ULSA). The period of these ripples requires spacing sizes of less than about 0.38λ to avoid unacceptable aliasing errors.

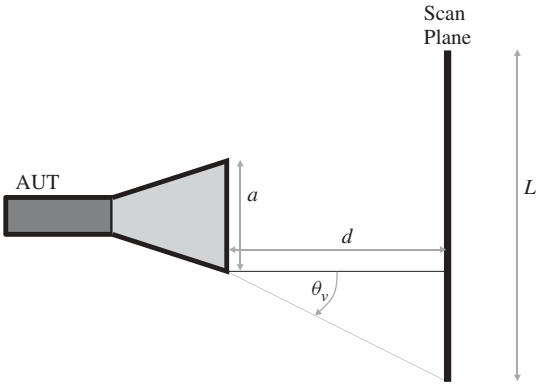


Figure 19.11 Illustration of the planar truncation geometry showing the angle of validity θ_v.

L is the scan length, a is the AUT diameter, and d is the separation distance between the AUT aperture and the scan plane (see Figure 19.11). Second, there are errors in the valid region due to ringing caused by the discontinuity at the edge of the scan plane. Yaghjian [48, p. 29] showed that this truncation error can be estimated from a knowledge of the measured data on the perimeter of the scan area. To determine whether the scan area is large enough, we set the data in the outer perimeter of the scan area to zero and observe how much the computed far field changes. As a rule of thumb, the scan area should be large enough so that measurements, at the edges, are at least 30 dB below (and preferably 40 dB or more below) the near-field peak.

Positioning Tolerances and Position Correction The fast Fourier transform, which provides an efficient near-field-to-far-field transform for the planar geometry, requires measurement points to be located on the plane $z = d$ in a regular grid, equispaced in x and equispaced in y. Errors can occur in the z coordinate (perpendicular to the scan plane) due to imperfections in the scanner (see Figure 19.12). Errors can occur in the step coordinate due to inertia in the scanner and errors in the y rails, which limit the step resolution (see Figure 19.13). Lastly, errors can occur in the scan coordinate due to errors in the x rails, the position encoders, the trigger timing accuracy, and the time to acquire a measurement. Newell [47, Eqs. (53)–(61)] and Yaghjian [48, Eq. (61)] may be consulted to estimate the errors in the far-field pattern due to position errors. Alternatively, one can use a laser tracker to determine the actual positions where the data are acquired and use a position-correction method such as that developed by Wittmann et al. [49] to perform an efficient near-field-to-far-field transform.

In addition, other sampling grids have been proposed [50–53]. The method of Wittmann et al. [49] can be used to process data acquired on a "nonstandard" grid.

Choosing a Separation Distance Several factors must be taken into account when choosing a probe–test antenna separation distance. First, we must ensure that evanescent (nonpropagating) modes do not contribute significantly to the measured near-field data (since it is extremely difficult to determine the response of the probe to these modes). A minimum separation of 3λ is often sufficient. The second consideration is that, as the separation distance increases, the valid region of the pattern decreases for a given scan area (see Eq. (19.98)). Thus we must choose a separation distance that allows adequate

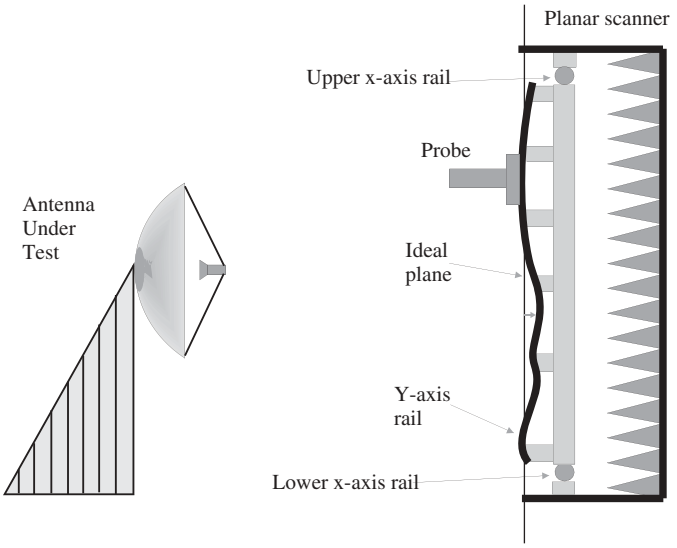


Figure 19.12 Illustration of z position error.

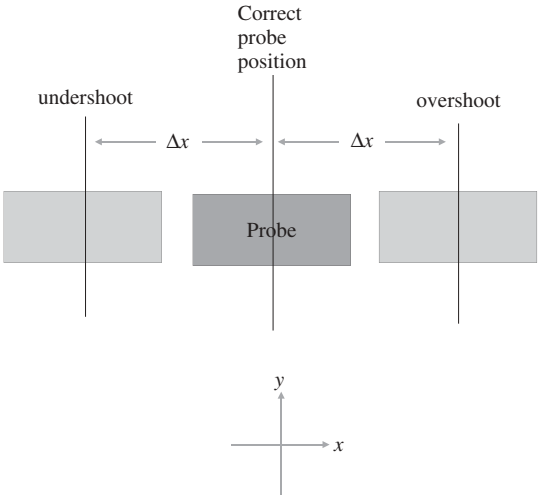


Figure 19.13 Illustration of the step position error. For the NIST planar scanner the step axis corresponds to the horizontal (x) axis.

pattern coverage. The final consideration is that of probe–AUT multiple reflections. When the signal travels from the transmitting antenna to the receiving antenna, not all the signal is received. Some of this signal is scattered from the receiving antenna back to the transmitting antenna and re-reflected to the receiving antenna. These reflections are not accounted for in the practical implementation of the theory. Multiple reflection errors generally increase with decreasing separation distance. It is preferable to choose a separation distance where the amplitude of the variation is smaller and where the gradient

is small. The multiple reflection effect can be estimated by taking measurements on two planes separated by $\lambda/4$, calculating the far field for each measurement, and computing the difference, direction by direction.

In choosing a separation distance, the need of smaller separation distances to increase pattern coverage (and improve signal to noise due to limited power) must be counterbalanced with the need to increase separation distance to decrease the effects of multiple reflections and evanescent modes.

Choosing a Probe As shown in Section 19.2.3.3, probe correction for planar near-field measurements is performed direction by direction after the transformation to the far field. One consequence is that if the probe has a null in its pattern, the uncertainty in the probe correction for that direction will be large. In planar scanning the probe should have good sensitivity in any direction for which we need an accurate determination of the AUT pattern.

Experience shows that the closer the probe and test antenna are in gain, the worse the multiple reflections will be. Thus the gain of the probe should be considerably less than that of the test antenna. However, if the gain is too small, then the insertion loss will be large and cannot be measured accurately. These factors lead us to generally choose a probe that has a gain that is approximately 20–30 dB less than that of the test antenna.

Finally, the return loss of the probe should be at least 10 dB (preferably 20 dB or more) to reduce the impedance mismatch. Corrections can be made for the impedance mismatch, but larger mismatch corrections are usually accompanied by larger measurement uncertainties.

Probe Correction The theory of probe correction (Section 19.2.3.3) requires the complex far-field patterns of two probes. The amplitude and phase of the second probe pattern must be known relative to the first probe pattern. If the second probe is the first probe rotated by 90° (as might be the case if one uses a linearly polarized probe to measure first one polarization and then the other), then the second probe pattern is the first probe pattern rotated by 90°. (For circularly polarized probe, the orthogonal polarization cannot be obtained by rotating the probe.) If the probe is a single antenna with more than one port (for different polarizations), then we must know the relative phase and amplitude between the ports. The amplitude difference can be calculated from the difference in gain between the ports. The phase difference must be measured (in the far field) by switching between the ports [43].

19.3.3 Spherical Scanning Measurements

19.3.3.1 Overview of Spherical Scanning A basic spherical scanning geometry is shown in Figure 19.3. There are a number of practical implementations of spherical scanning. (See Figures 19.14 and 19.15 for examples.) In all of them, the probe *effectively moves over the surface of a sphere*. We focus on the implementation of Figure 19.14 in which the test antenna is mounted on a roll-over-azimuth positioner. In this case, in contrast to the planar scanning case, it is the probe, rather than the AUT, that is fixed. The test antenna rotates in the two angular coordinates (see Figure 19.14). The spin angle χ is controlled at the probe, which is located on the laboratory z axis.

For AUTs with broad beams, the spherical near-field method is preferred because radiation in all directions is intercepted by the probe. Either θ' or φ' can be used as

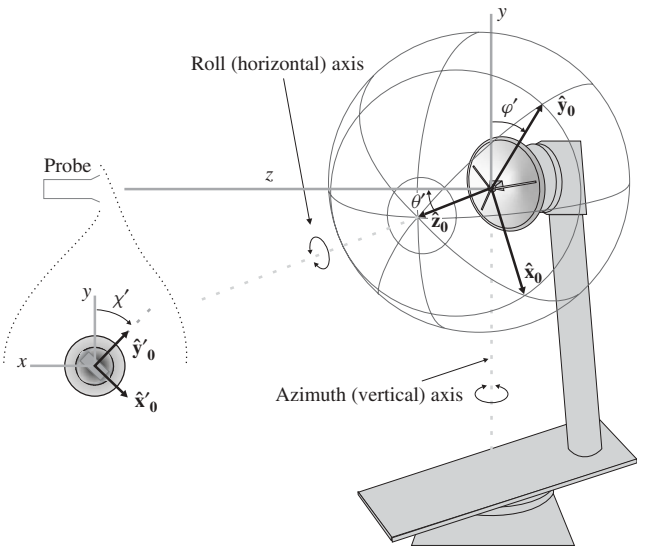


Figure 19.14 Spherical scanning with a roll-over-azimuth positioner. The inset shows the probe as viewed from the test antenna.

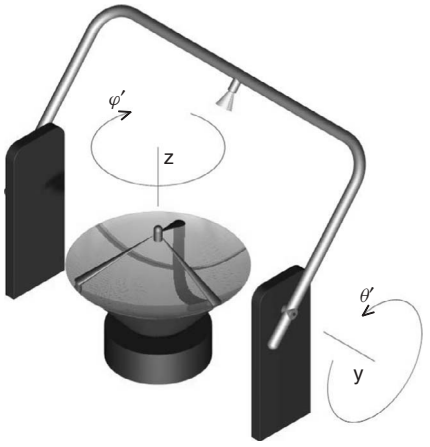


Figure 19.15 Spherical scanning with the probe on an arch and the test antenna on an azimuth rotator.

the scanning direction. The test antenna is then stepped in the orthogonal direction and scanning reinitiated in the scanning direction.

19.3.3.2 Implementation

Alignment A properly aligned roll-over-azimuth measurement setup satisfies the following requirements. (1) The azimuth (vertical) axis must be coincident with the laboratory y axis. The roll (horizontal) axis must be coincident with the laboratory z axis

when $\theta = 0$. (2) The AUT coordinate system must be coincident with the laboratory coordinate system when $\theta = 0$ and $\varphi = 0$. (3) The probe z axis must be coincident with the laboratory z axis. The probe y axis must be parallel to the laboratory y axis when $\chi = 0$.

Actual alignment procedures will vary from situation to situation. Because adjustments tend to be mutually dependent, it is necessary to iterate any procedure until further improvement is not practical. Perfect alignment is prevented, of course, by mechanical imperfections. For example, due to a lack of stiffness, it is generally not possible to maintain alignment of the z_a axis as a function of the roll angle φ. The alignment for spherical scanning is considerably more complex and time consuming than the alignment for planar or cylindrical scanning.

Sampling Criteria We must obtain enough data to determine the contributions of all relevant spherical modes. The magnitudes of the modal coefficients are observed to drop sharply for $n > N$, where $N \sim kr_0$ and r_0 is the radius of the minimum sphere. (The minimum sphere is the smallest sphere, centered on the origin of the laboratory coordinate system, that encloses the radiating parts of the AUT.) Sampling considerations then imply that

$$\Delta\theta, \Delta\varphi \leq \frac{2\pi}{2N+1} \tag{19.99}$$

To provide a margin of safety, N is often chosen so that $N = kr_0 + 10$. As in the planar case, undersampling leads to aliasing. A similar test can be performed for aliasing; that is, we obtain data with spacing about one-tenth that required by Eq. (19.99) and compare the results using all of the data with those obtained using every other point, then every third point, and so forth. In spherical scanning, oversampling can be used to reduce the effects of room reflections.

Scan Area Truncation In theory, data must be acquired over a complete sphere. Sometimes there are blockages (e.g., from mounting structures) so that obtaining data over a complete sphere becomes impossible. Acquiring data over a partial sphere has two effects. First, the pattern beyond $\pm\theta_v$ is invalid [26, p. 233], where

$$\theta_v = \theta_m - \arcsin\left(\frac{r_0}{d}\right) \tag{19.100}$$

where θ_m is the angle to which we obtain measured near-field data (see Figure 19.16), and d is the radius of the measurement sphere. Second, as in the planar case, there are errors in the valid region due to ringing. To determine whether the scan area is large enough, we set the data in the outer perimeter of the scan area to zero and observe how much the computed far field changes. As a rule of thumb, the scan area should be large enough so that measurements, at the edges, are at least 30 dB below (and preferably 40 dB or more below) the near-field peak. Some mitigation of truncation effects can be accomplished with a procedure such as that outlined in Wittmann et al. [54].

Positioning Tolerances and Position Correction The standard algorithm, which provides an efficient near-field-to-far-field transform for the spherical geometry, assumes that measurement points are located on the sphere $r = d$ in a regular grid, equispaced in θ and equispaced in φ. The position error must be small compared to the maximum

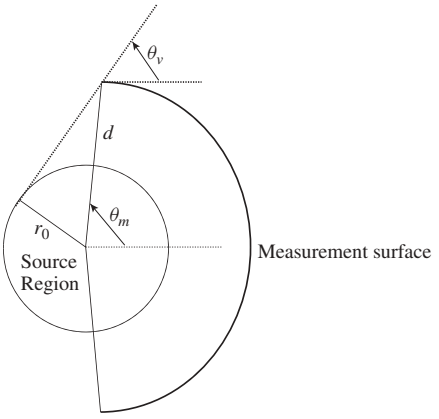


Figure 19.16 The spherical truncation geometry illustrating the angle of validity θ_v.

sample size allowed by Eq. (19.99). If one is able to determine the actual positions where the data are acquired, a position-correction method, such as that developed by Wittmann et al. [55], will provide an efficient near-field-to-far-field transform.

In addition, other sampling schemes have been proposed [56]. The method of Wittmann et al. [55] can be used to process data acquired on a "nonstandard" grid.

Choosing a Separation Distance In choosing the separation distance between the AUT and the probe (the measurement radius), we must consider the effects of probe–AUT multiple reflections and the effects of noise. An increased separation will reduce the effects of multiple reflections. However, increasing the separation tends to decrease the signal level so that the relative noise level is higher. Unlike the situation for planar scanning, an increase in the separation distance (measurement radius) decreases the effect of truncation for spherical scanning, Eq. (19.100). In spherical scanning, the separation between probe and AUT can never be less than the radius of the minimum sphere.

Choosing a Probe As shown in Section 19.2.4.3, the spherical transform relies on the probe having some special properties. The probe must have only spherical modes corresponding to $\mu = \pm 1$. That is, the probe response must have the form $A\cos(\chi + \chi_0)$, where χ is the rotation angle about the z axis of the probe. Also, the spherical transform theory assumes that the second probe (used for measuring the orthogonal polarization) is the same as the first probe but rotated 90°. Since a circularly polarized antenna rotated by 90° has the same polarization but a different tilt angle, this means we cannot use a circularly polarized probe in spherical scanning (but see the discussion after Eq. (19.75)).

As in the planar case, the gain of the probe should be considerably less than that of the test antenna. However, if the gain is too small, then the insertion loss will be large and cannot be measured accurately. These factors lead us to generally choose a probe that has a gain that is approximately 20–30 dB less than the test antenna.

Finally, the return loss of the probe should be at least 10 dB (preferably 20 dB or more) to reduce the impedance mismatch. Corrections can be made for the impedance mismatch, but larger mismatch corrections are usually accompanied by larger measurement uncertainties.

Probe Correction The theory of probe correction (Section 19.2.4.3) requires the complex far-field pattern of a single probe that is essentially linearly polarized. This probe is used in two χ orientations.

19.3.4 Cylindrical Scanning Measurements

19.3.4.1 Overview of Cylindrical Scanning

In common implementations of cylindrical scanning, both the probe and the AUT move. The AUT rotates in the angular coordinate Φ, while the probe is usually translated in the y (vertical) direction. The net result is equivalent to a probe moving over the surface of a cylinder about the test antenna (see Figure 19.4). The cylindrical scanning method is particularly useful for fan-beam antennas, which have a broad beam in one plane but a narrow beam in the orthogonal plane. For antennas with broad beams in both planes, the use of the spherical near-field method is preferred because radiation in all directions will be intercepted by the probe. Either Φ or y can be used as the scanning direction; that is, either the AUT is stepped in Φ and the probe scanned in y, or the probe is stepped in y and the AUT scanned in Φ.

19.3.4.2 Implementation

Alignment Cylindrical scanning is typically accomplished by use of an azimuth rotator for the Φ rotation and a vertical transport to move the probe in the laboratory y direction. The rotator is aligned so that its axis is coincident with the laboratory y axis. The probe is aligned so that $\hat{\mathbf{y}}'_0 = \hat{\mathbf{y}}$ and $\hat{\mathbf{z}}'_0 = -\hat{\mathbf{z}}$ at $\Phi = 0$ (see Figure 19.4).

Sampling Criteria For the angular coordinate the sampling requirement for cylindrical scanning is similar to that for the spherical scanning case (see Eq. (19.99)). For the translational coordinate the sampling requirement is similar to that for the planar scanning case (see Section 19.3.2.2). That is,

$$\begin{aligned} \Delta\Phi &\le \frac{2\pi}{2N+1} \\ \Delta y &\le \frac{\lambda}{2} \end{aligned} \tag{19.101}$$

where $N \sim k\rho_0$ and ρ_0 is the radius of the minimum cylinder. (The minimum cylinder is the smallest cylinder, centered on the y axis, that encloses the radiating parts of the AUT.)

Scan Area Truncation The theory requires data over an infinite cylinder. Acquiring data over a finite cylinder has two effects. First, the pattern beyond the cutoff angle is invalid. For the angular coordinate, the azimuthal angle of validity A_v is similar to the spherical case (Figure 19.16) and is given by

$$A_v = \Phi_m - \arcsin\left(\frac{\rho_0}{d}\right) \tag{19.102}$$

where Φ_m is the truncation angle for the near-field scan and d is the radius of the measurement cylinder. The elevation angle of validity E_v is similar to the planar angle of validity (Figure 19.11) and is given by

$$E_v = \arctan\left(\frac{L_y - a_y}{2d}\right) \tag{19.103}$$

where L_y is the scan length in the y direction, and a_y is the antenna extent in the y direction. Second, as in the planar case, there are errors in the valid region due to ringing. To determine whether the scan area is large enough, we set the data in the outer perimeter of the scan area to zero and observe how much the computed far field changes. As a rule of thumb, the scan area should be large enough so that measurements, at the edges, are at least 30 dB below (and preferably 40 dB or more below) the near-field peak.

Positioning Tolerances and Position Correction The standard algorithm, which provides an efficient near-field-to-far-field transform for the cylindrical geometry, requires measurement points to be located on the cylinder $\rho = d$ in a regular grid, equispaced in y and equispaced in Φ. The position error must be small compared to the maximum sample spacing allowed by Eq. (19.101). A probe-position correction algorithm for cylindrical scanning, such as those developed for spherical and planar scanning, has not been implemented to date, although such a scheme is certainly possible.

In addition, other sampling grids have been proposed [57].

Choosing a Separation Distance In choosing the separation distance between the test antenna and the probe (the measurement radius), we must consider the effects of probe–AUT multiple reflections, truncation, and noise. An increased separation will reduce the effects of multiple reflections. However, increasing the separation distance decreases the signal level, so that the relative noise level is higher. Increasing the separation distance has mixed results on the truncation in cylindrical scanning. It decreases the effects of truncation for the Φ direction (increases the angle of validity) but increases the truncation effect in the y direction (reduces the angle of validity). For cylindrical scanning, the separation between the probe and AUT cannot be less than the radius of the minimum cylinder.

Choosing a Probe As in the planar case, the gain of the probe should be considerably less than that of the test antenna. However, if the gain is too small, then the insertion loss will be large and cannot be measured accurately. These factors lead us to generally choose a probe that has a gain that is approximately 20–30 dB less than the test antenna.

The return loss of the probe should be at least 10 dB (preferably 20 dB or more) to reduce the impedance mismatch. Corrections can be made for the impedance mismatch, but larger mismatch corrections are usually accompanied by larger measurement uncertainties.

Probe Correction The theory of probe correction (Section 19.2.5.3) requires the complex far-field patterns of two probes. The amplitude and phase of the second probe pattern must be known relative to the first probe pattern. If the second probe is the first probe rotated by 90° (as might be the case if one uses a linearly polarized probe to measure first one polarization and then the other), then the second probe pattern is the first probe pattern rotated by 90°. (For a circularly polarized probe, the orthogonal polarization cannot be obtained by rotating the probe.) If the probe is a single antenna with more than one port (for different polarizations), then we must determine the relative phase and amplitude between the ports. The amplitude difference can be calculated from the difference in gain between the ports. The phase difference must be measured (in the far field) by switching between the ports [43].

19.3.5 Sources of Measurement Uncertainty

No measurement is truly complete without an uncertainty analysis that provides an estimate of how large the measurement error might be. A detailed analysis for estimating the errors in planar scanning has been presented by Newell [47]. For spherical scanning, an analysis of error sources has been done by Hansen [26, Chap. 6]. The sources of uncertainty are similar in all three scanning geometries. Tables 19.3 and 19.4 show the sources of uncertainty for test antenna gain and pattern, respectively. These tables are similar to the 18-term uncertainty table found in Newell [47].

Measurements of the AUT gain and the relative pattern share many sources of uncertainty. There are exceptions, however. The "probe relative pattern" uncertainty does not affect the resulting gain uncertainty of the AUT unless the AUT's main beam is steered off axis. Also, when the second probe is the same as the first probe but rotated 90°, the AUT relative pattern uncertainty is not affected by uncertainty in the "probe gain" or in the "impedance mismatch."

Near-field scanning has been used successfully to measure low sidelobe arrays [58, 59]. One example, the pattern of the ULSA for 3.0 GHz, is shown in Figure 19.17. The uncertainties for the measurement for this antenna are discussed in detail in Francis et al. [58]. The uncertainties relative to peak for directions other than the main beam and the grating lobe (at about 46°) are shown in Table 19.5. These values correspond to a coverage factor of 2. (Were we to assume a normal distribution, a coverage factor of 2 would correspond to a 95% (2σ) confidence level.) If an uncertainty source in Table 19.4 is not listed in Table 19.5 that is because its contribution is −100 dB or less. Table 19.6 shows the corresponding uncertainties versus sidelobe level.

TABLE 19.3 Sources of Uncertainty for Gain Measurement

Probe gain	Probe–AUT multiple reflections
Probe alignment	System nonlinearity
Probe relative pattern	System dynamic range
AUT alignment	Flexing cables
Normalization constant	Temperature (drift) effects
Impedance mismatch	Room scattering
Aliasing	Leakage and cross-talk
Measurement area truncation	Random errors
Probe position errors	

TABLE 19.4 Sources of Uncertainty for Relative Pattern Measurement

Probe polarization ratio	Probe position errors
Probe gain	Probe–AUT multiple reflections
Probe alignment	System nonlinearity
Probe relative pattern	System dynamic range
AUT alignment	Flexing cables
Normalization constant	Temperature (drift) effects
Impedance mismatch	Room scattering
Aliasing	Leakage and cross-talk
Measurement area truncation	Random errors

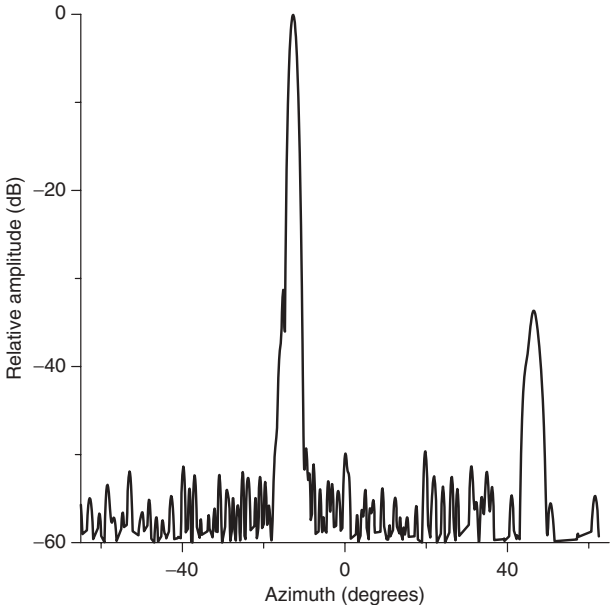


Figure 19.17 Azimuth cut of the far-field pattern of the ULSA as determined on the NIST planar near-field range. This pattern generally agrees with the ULSA pattern measured on a far-field range to within the uncertainties of measuremment. Note the grating lobe at 46°.

TABLE 19.5 Relative Pattern Uncertainties for the ULSA

Uncertainty Source	Uncertainty Level (dB relative to peak)
Probe–AUT multiple reflections	−65
Room scattering	−70
Cable flexing	−70
Position errors	−75
Random errors	−75
Leakage and cros-stalk	−75
Truncation	−90
Aliasing	−90
Root sum square	*−62*

TABLE 19.6 Relative Pattern Uncertainties Versus Sidelobe Level

Sidelobe Level (dB relative to peak)	Uncertainty (dB)
−30	±0.25
−45	+1.2, −1.3
−55	+3.2, −5.0
−60	+5.0, −13.3

In the main beam region and the direction of the grating lobe, the probe–AUT multiple reflections are greater, and separate uncertainty tables are required for these directions.

19.4 SUMMARY

There are three geometries in which near-field scanning can be performed efficiently: planar, spherical, and cylindrical. As compared to far-field measurements, near-field scanning measurements use less real estate and provide a more controlled environment.

ACKNOWLEDGMENTS

The authors thank Jeff Guerrieri and Katie MacReynolds of the National Institute of Standards and Technology for their assistance in creating the figures and in acquiring near-field data. We also thank numerous reviewers for critical comments. In particular, we acknowledge detailed and lengthy discussions with Dr. Doren Hess of MI Technologies.

APPENDIX A: TRANSMITTING PROBE, RECEIVING TEST ANTENNA

Transmission formulas for cases when the probe transmits and the test antenna receives can be developed via slight generalization of the foregoing arguments. We list some key formulas.

Planar scanning: Corresponding to Eq. (19.42),

$$\frac{w(\mathbf{r})}{a_0'} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbf{t}_\pi'(-\hat{\mathbf{k}}) \cdot \mathbf{s}_0(-\hat{\mathbf{k}}) \exp(i\mathbf{k}\cdot\mathbf{r}) \frac{d\alpha\, d\beta}{\gamma k} \tag{19.104}$$

Spherical scanning: Corresponding to Eqs. (19.67), (19.60), and (19.104),

$$\frac{w(r,\varphi,\theta,\chi)}{a_0'} = \sum_{n=1}^{\infty}\sum_{m=-n}^{n}\sum_{\mu=-n}^{n} \left[\tau_{n\mu}^{1\prime}(r)s_{nm}^1 + \tau_{n\mu}^{2\prime}(r)s_{nm}^2\right] D_{\mu m}^n(-\chi, -\theta, -\varphi) \tag{19.105}$$

$$\mathbf{s}_0(\hat{\mathbf{r}}) = \sum_{n=1}^{\infty}\sum_{m=-n}^{n} \left[s_{nm}^1 \mathbf{X}_{nm}^1(\hat{\mathbf{r}}) + s_{nm}^2 \mathbf{X}_{nm}^2(\hat{\mathbf{r}})\right] \tag{19.106}$$

$$\begin{aligned} \tau_{nm}^{1\prime}(r) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbf{t}_\pi'\left(-\hat{\mathbf{k}}\right) \cdot \mathbf{X}_{nm}^1(-\hat{\mathbf{k}}) \exp(i\gamma r) \frac{d\alpha\, d\beta}{\gamma k} \\ \tau_{nm}^{2\prime}(r) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbf{t}_\pi'(-\hat{\mathbf{k}}) \cdot \mathbf{X}_{nm}^2(-\hat{\mathbf{k}}) \exp(i\gamma r) \frac{d\alpha\, d\beta}{\gamma k} \end{aligned} \tag{19.107}$$

Cylindrical scanning: Corresponding to Eqs. (19.94), (19.87), and (19.132),

$$\frac{w(\mathbf{r})}{a_0'} = \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\tau_{n,-\beta}^{1\prime}(\rho)\, s_{n,-\beta}^1 + \tau_{n,-\beta}^{2\prime}(\rho)\, s_{n,-\beta}^2\right] \exp(in\Phi + i\beta y) \frac{d\beta}{k} \tag{19.108}$$

$$\mathbf{s}_0(\hat{\mathbf{r}}) = \sum_{n=-\infty}^{\infty} \left[s_{n,ky/r}^1 \boldsymbol{\Psi}_n^1(\hat{\mathbf{r}}) + s_{n,ky/r}^2 \boldsymbol{\Psi}_n^2(\hat{\mathbf{r}})\right] \tag{19.109}$$

$$\tau^{1'}_{n,-\beta}(\rho) = \int_{-\infty}^{\infty} \mathbf{t}'_{\pi}(-\hat{\mathbf{k}}) \cdot \boldsymbol{\Psi}^{1}_{n}(-\hat{\mathbf{k}}) \exp(i\gamma\rho) \frac{d\alpha}{\gamma}$$
$$\tau^{2'}_{n,-\beta}(\rho) = \int_{-\infty}^{\infty} \mathbf{t}'_{\pi}(-\hat{\mathbf{k}}) \cdot \boldsymbol{\Psi}^{2}_{n}(-\hat{\mathbf{k}}) \exp(i\gamma\rho) \frac{d\alpha}{\gamma} \tag{19.110}$$

In all geometries

$$\frac{w}{a'_0} = \frac{w'}{a_0} \tag{19.111}$$

when both the probe and test antenna are reciprocal.

APPENDIX B: SPECIAL PROBES

The use of $\mu = \pm 1$ probes leads to significant simplifications in the practical implementation of spherical near-field scanning techniques. These probes are often constructed by attaching a section of circular waveguide to a rectangular waveguide feed via a "smooth" transition. The guides are joined so they have a common axis and the transition region is designed so that the probe has reflection symmetry through two perpendicular planes. Figure 19.18 shows a $\mu = \pm 1$ probe intended for applications near 16 GHz.

The first few circular waveguide modes and their cutoff frequencies f_c are given in Table 19.7 [60, Sec. 2.3], where c is the speed of light and d is the inside diameter of the guide. Because of symmetry, the rectangular guide will not respond to $\mathrm{TE}_{\mu v}$ or

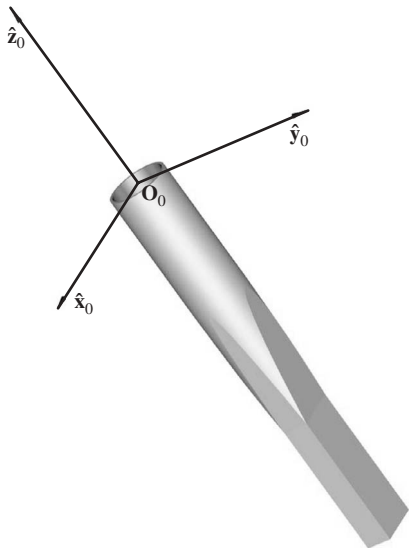


Figure 19.18 A Ku-band probe for spherical scanning measurements constructed from a section of WR-62 waveguide, followed by a symmetric transition to circular guide, followed by a section of circular waveguide. The length of each section is 5 cm. The inside diameter of the circular guide is 2 cm. The waveguide flange is not shown. The fiducial coordinate system has been chosen so that the origin is at the center of the circular aperture, the z axis is parallel to the probe axis, and the y axis is perpendicular to the broad side of the rectangular guide.

TABLE 19.7 Circular Waveguide Modes

Mode	f_c
TE_{11}	$1.841\dfrac{c}{\pi d}$
TM_{01}	$2.405\dfrac{c}{\pi d}$
TE_{21}	$3.054\dfrac{c}{\pi d}$
TE_{01}, TM_{11}	$3.832\dfrac{c}{\pi d}$
TE_{31}	$4.201\dfrac{c}{\pi d}$

$TM_{\mu\nu}$ modes when μ is even. To obtain the required $\mu = \pm 1$ dependence, the circular waveguide must filter out only the TE_{31} (and higher order) modes. When $d = 2.2$ cm, the TE_{31} mode is exponentially attenuated below 18.2 GHz and the probe will be useful over the full Ku band (12.4–18 GHz).

APPENDIX C: TRANSLATED PROBE COEFFICIENTS

C.1 Planar Geometry

The Kerns transmission formula (Eq. (19.42)) can be written

$$\frac{w'(\mathbf{r})}{a_0} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\sigma_{\alpha\beta}^{1\prime}(z)t_{\alpha\beta}^{1} + \sigma_{\alpha\beta}^{2\prime}(z)t_{\alpha\beta}^{2}\right]\exp(i\alpha x + i\beta y)\frac{d\alpha\, d\beta}{\gamma k} \tag{19.112}$$

where the translated probe coefficients are

$$\begin{aligned}\sigma_{\alpha\beta}^{1\prime}(z) &= \left[\mathcal{P}\cdot\mathbf{m}_{\alpha\beta}\right](0,0,z) = s_{\alpha\beta}^{1\prime}\exp(i\gamma z)\\ \sigma_{\alpha\beta}^{2\prime}(z) &= \left[\mathcal{P}\cdot\mathbf{n}_{\alpha\beta}\right](0,0,z) = -s_{\alpha\beta}^{2\prime}\exp(i\gamma z)\end{aligned} \tag{19.113}$$

This form is directly comparable to Eqs. (19.67) and (19.94). Evaluation of the translated probe coefficients is considerably more complicated in the spherical and cylindrical geometries.

C.2 Spherical Geometry

The translated probe coefficients are defined in Eq. (19.66). Using the Kerns transmission formula (19.42) and plane wave representations for the spherical waves (19.58),

$$\begin{aligned}\sigma_{nm}^{1\prime}(r) &= \left[\mathcal{P}\cdot\mathbf{m}_{n\mu}\right](r,0,0,0) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathbf{s}_{\pi}'(\hat{\mathbf{k}})\cdot\mathbf{X}_{nm}^{1}(\hat{\mathbf{k}})\exp(i\gamma r)\frac{d\alpha\, d\beta}{\gamma k}\\ \sigma_{nm}^{2\prime}(r) &= \left[\mathcal{P}\cdot\mathbf{n}_{n\mu}\right](r,0,0,0) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathbf{s}_{\pi}'(\hat{\mathbf{k}})\cdot\mathbf{X}_{nm}^{2}(\hat{\mathbf{k}})\exp(i\gamma r)\frac{d\alpha\, d\beta}{\gamma k}\end{aligned} \tag{19.114}$$

Introducing the spherical-harmonic expansion

$$\mathbf{s}'_{\pi}(\hat{\mathbf{r}}) = \sum_{\nu=1}^{\infty} \sum_{\mu=-\nu}^{\nu} \left[s^{1\prime}_{\nu\mu} \mathbf{X}^1_{\nu\mu}(\hat{\mathbf{r}}) + s^{2\prime}_{\nu\mu} \mathbf{X}^2_{\nu\mu}(\hat{\mathbf{r}}) \right] \tag{19.115}$$

we have

$$\begin{aligned} \sigma^{1\prime}_{nm} &= \sum_{\nu=1}^{\infty} \left[B\,(\nu, -m|nm)\, s^{1\prime}_{\nu,-m} + C\,(\nu, -m|nm)\, s^{2\prime}_{\nu,-m} \right] \\ \sigma^{2\prime}_{nm} &= -\sum_{\nu=1}^{\infty} \left[C\,(\nu, -m|nm)\, s^{1\prime}_{\nu,-m} + B\,(\nu, -m|nm)\, s^{2\prime}_{\nu,-m} \right] \end{aligned} \tag{19.116}$$

with

$$\begin{aligned} B\,(\nu, -m|nm) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{X}^1_{\nu,-m}(\hat{\mathbf{k}}) \cdot \mathbf{X}^1_{nm}(\hat{\mathbf{k}}) \exp(i\gamma r) \frac{d\alpha\, d\beta}{\gamma k} \\ C\,(\nu, -m|nm) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{X}^2_{\nu,-m}(\hat{\mathbf{k}}) \cdot \mathbf{X}^1_{nm}(\hat{\mathbf{k}}) \exp(i\gamma r) \frac{d\alpha\, d\beta}{\gamma k} \end{aligned} \tag{19.117}$$

Products of vector spherical harmonics can be expanded in scalar spherical harmonics [24, Chap. 6]:

$$\begin{aligned} \mathbf{X}^1_{\nu,-m}(\hat{\mathbf{k}}) \cdot \mathbf{X}^1_{nm}(\hat{\mathbf{k}}) &= \sum_{\ell=|\nu-n|}^{\nu+n} b\,(\ell 0|\nu, -m, nm)\, Y_{\ell 0}\left(\hat{\mathbf{k}}\right) \\ \mathbf{X}^2_{\nu,-m}(\hat{\mathbf{k}}) \cdot \mathbf{X}^1_{nm}(\hat{\mathbf{k}}) &= \sum_{\ell=|\nu-n|}^{\nu+n} c\,(\ell 0|\nu, -m, nm)\, Y_{\ell 0}\left(\hat{\mathbf{k}}\right) \end{aligned} \tag{19.118}$$

so that

$$\begin{aligned} B\,(\nu, -m|nm) &= \sum_{\ell=|\nu-n|}^{\nu+n} \sqrt{\pi\,(2\ell+1)}\, f_{\ell}\,(kr)\, b\,(\ell 0|\nu, -m, nm) \\ C\,(\nu, -m|nm) &= \sum_{\ell=|\nu-n|}^{\nu+n} \sqrt{\pi\,(2\ell+1)}\, f_{\ell}\,(kr)\, c\,(\ell 0|\nu, -m, nm) \end{aligned} \tag{19.119}$$

where we have used Eqs. (19.48), and (19.50) and the fact that

$$Y_{nm}\,(\hat{\mathbf{z}}) = \sqrt{\frac{2n+1}{4\pi}} \delta_{m0} \tag{19.120}$$

The coefficients in Eq. (19.118) are [24, see Chap. 7]

$$b\left(\ell 0|\nu,-m,nm\right)=\sqrt{\frac{\left(2\ell+1\right)\left(2\nu+1\right)\left(2n+1\right)}{16\pi\,\nu\left(\nu+1\right)n\left(n+1\right)}}\left[\ell\left(\ell+1\right)-\nu\left(\nu+1\right)-n\left(n+1\right)\right]$$
$$\times\begin{pmatrix}\ell & \nu & n\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}\ell & \nu & n\\ 0 & -m & m\end{pmatrix} \tag{19.121}$$

$$c\left(\ell 0|\nu,-m,nm\right)=-\sqrt{\frac{\left(2\ell+1\right)\left(2\nu+1\right)\left(2n+1\right)}{16\pi\,\nu\left(\nu+1\right)n\left(n+1\right)}}\sqrt{\left[\ell^2-\left(\nu-n\right)^2\right]\left[\left(\nu+n+1\right)^2-\ell^2\right]}$$
$$\times\begin{pmatrix}\ell-1 & \nu & n\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}\ell & \nu & n\\ 0 & -m & m\end{pmatrix} \tag{19.122}$$

The 3-j symbols (which are scalars, not matrices) in these formulas may be computed recursively.

For a $\mu=\pm1$ probe, we require the special cases

$$b\left(\ell 0|\nu,\pm1,n,\mp1\right)=\sqrt{\frac{\left(2\ell+1\right)\left(2\nu+1\right)\left(2n+1\right)}{4\pi}}$$
$$\times\frac{\left[\ell\left(\ell+1\right)-\nu\left(\nu+1\right)-n\left(n+1\right)\right]^2}{4\nu\left(\nu+1\right)n\left(n+1\right)}\begin{pmatrix}\ell & \nu & n\\ 0 & 0 & 0\end{pmatrix}^2 \tag{19.123}$$

$$c\left(\ell 0|\nu,\pm1,n,\mp1\right)=\pm\sqrt{\frac{\left(2\ell+1\right)\left(2\nu+1\right)\left(2n+1\right)}{4\pi}}$$
$$\times\frac{\left[\ell^2-\left(\nu-n\right)^2\right]\left[\left(\nu+n+1\right)^2-\ell^2\right]}{4\nu\left(\nu+1\right)n\left(n+1\right)}\begin{pmatrix}\ell-1 & \nu & n\\ 0 & 0 & 0\end{pmatrix}^2 \tag{19.124}$$

The 3-j symbols in Eqs. (19.123) and (19.124) may be evaluated from the formulas

$$\begin{pmatrix}\ell & \nu & n\\ 0 & 0 & 0\end{pmatrix}=0 \tag{19.125}$$

when $\ell+\nu+n$ is odd, $\ell<|\nu-n|$, or $\ell>\nu+n$,

$$\sum_{\ell}\left(2\ell+1\right)\begin{pmatrix}\ell & \nu & n\\ 0 & 0 & 0\end{pmatrix}^2=1 \tag{19.126}$$

$$\xi_\ell\begin{pmatrix}\ell-1 & \nu & n\\ 0 & 0 & 0\end{pmatrix}^2=\xi_{\ell+1}\begin{pmatrix}\ell+1 & \nu & n\\ 0 & 0 & 0\end{pmatrix}^2 \tag{19.127}$$

$$\xi_\ell=\left[\ell^2-\left(\nu-n\right)^2\right]\left[\left(\nu+n+1\right)^2-\ell^2\right] \tag{19.128}$$

Asymptotic formulas are available for $\sigma_{nm}^{1\prime}$ and $\sigma_{nm}^{2\prime}$. Beginning with Eq. (19.114), a stationary-phase evaluation yields

$$\sigma_{nm}^{1\prime}(r) \underset{r\to\infty}{\sim} \mathbf{s}_{\pi}'(\hat{\mathbf{z}}) \cdot \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbf{X}_{nm}^{1}(\hat{\mathbf{k}}) \exp(i\gamma r)\frac{d\alpha\, d\beta}{\gamma k}$$

$$\underset{r\to\infty}{\sim} 2\pi f_n(kr)\mathbf{s}_{\pi}'(\hat{\mathbf{z}}) \cdot \mathbf{X}_{nm}^{1}(\hat{\mathbf{z}}) \tag{19.129}$$

$$\sigma_{nm}^{2\prime}(r) \underset{r\to\infty}{\sim} \mathbf{s}_{\pi}'(\hat{\mathbf{z}}) \cdot \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbf{X}_{nm}^{2}(\hat{\mathbf{k}}) \exp(i\gamma r)\frac{d\alpha\, d\beta}{\gamma k}$$

$$\underset{r\to\infty}{\sim} 2\pi g_n(kr)\, \mathbf{s}_{\pi}'(\hat{\mathbf{z}}) \cdot \mathbf{X}_{nm}^{2}(\hat{\mathbf{z}}) \tag{19.130}$$

where

$$\mathbf{X}_{nm}^{1}(\hat{\mathbf{z}}) = i\hat{\mathbf{z}} \times \mathbf{X}_{nm}^{2}(\hat{\mathbf{z}}) = \sqrt{\frac{2n+1}{16\pi}}\,(\hat{\mathbf{x}} \pm i\hat{\mathbf{y}})\,\delta_{m,\pm 1} \tag{19.131}$$

Formulas (19.129) and (19.130) provide a convenient numerical check in the evaluation of the translated probe coefficients. They also suggest, even for intermediate separations when the test antenna is in the "far field" of the probe, that practical probes are approximately $\mu = \pm 1$ and that on-axis properties of $\mathbf{s}_{\pi}'$ are most important.

C.3 Cylindrical Geometry

The translated probe coefficients are defined in Eq. (19.93). Using the Kerns transmission formula (19.42) and plane wave representations for the cylindrical waves (19.84),

$$\begin{aligned}
\sigma_{n\beta}^{1\prime}(\rho) &= \left[\boldsymbol{\mathcal{P}} \cdot \mathbf{m}_{n\beta}\right](\rho, 0, 0) = \int_{-\infty}^{\infty} \mathbf{s}_{\pi}'(\hat{\mathbf{k}}) \cdot \boldsymbol{\Psi}_{n}^{1}(\hat{\mathbf{k}}) \exp(i\gamma\rho)\frac{d\alpha}{\gamma} \\
\sigma_{n\beta}^{2\prime}(\rho) &= \left[\boldsymbol{\mathcal{P}} \cdot \mathbf{n}_{n\beta}\right](\rho, 0, 0) = \int_{-\infty}^{\infty} \mathbf{s}_{\pi}'(\hat{\mathbf{k}}) \cdot \boldsymbol{\Psi}_{n}^{2}(\hat{\mathbf{k}}) \exp(i\gamma\rho)\frac{d\alpha}{\gamma}
\end{aligned} \tag{19.132}$$

With the Fourier series representation

$$\mathbf{s}_{\pi}'(\hat{\mathbf{r}}) = \sum_{\nu=-\infty}^{\infty} \left[s_{\nu,ky/r}^{1\prime}\boldsymbol{\Psi}_{\nu}^{1}(\hat{\mathbf{r}}) + s_{\nu,ky/r}^{2\prime}\boldsymbol{\Psi}_{\nu}^{2}(\hat{\mathbf{r}})\right] \tag{19.133}$$

it follows that

$$\begin{aligned}
\sigma_{n\beta}^{1\prime}(\rho) &= \sum_{\nu=-\infty}^{\infty}\int_{-\infty}^{\infty} s_{\nu\beta}^{1\prime}\boldsymbol{\Psi}_{\nu}^{1}(\hat{\mathbf{k}}) \cdot \boldsymbol{\Psi}_{n}^{1}(\hat{\mathbf{k}}) \exp(i\gamma\rho)\frac{d\alpha}{\gamma} \\
&= \sqrt{2\pi}\sum_{\nu=-\infty}^{\infty} s_{\nu\beta}^{1\prime}\frac{1}{2\pi}\int_{-\infty}^{\infty} \Psi_{\nu+n}(\hat{\boldsymbol{\kappa}}) \exp(i\gamma\rho)\frac{d\alpha}{\gamma} \\
&= \sqrt{2\pi}\sum_{\nu=-\infty}^{\infty} s_{\nu\beta}^{1\prime}u_{\nu+n,\beta}(\rho\hat{\mathbf{z}})
\end{aligned}$$

$$= \frac{1}{\sqrt{2\pi}} \sum_{\nu=-\infty}^{\infty} s_{\nu\beta}^{1'} F_{\nu+n}(\kappa\rho) \tag{19.134}$$

where we have used Eqs. (19.76) and (19.79). Similarly,

$$\sigma_{n\beta}^{2'}(\rho) = -\frac{1}{\sqrt{2\pi}} \sum_{\nu=-\infty}^{\infty} s_{\nu\beta}^{2'} F_{\nu+n}(\kappa\rho) \tag{19.135}$$

Asymptotic formulas are available for $\sigma_{n\beta}^{1'}$ and $\sigma_{n\beta}^{2'}$ [12, 61]: Beginning with Eq. (19.134), a stationary-phase evaluation yields

$$\begin{aligned}
\sigma_{n\beta}^{1'}(\rho) &= \sum_{\nu=-\infty}^{\infty} \int_{-\infty}^{\infty} s_{\nu\beta}^{1'} \Psi_\nu(\hat{\boldsymbol{\kappa}}) \Psi_n(\hat{\boldsymbol{\kappa}}) \exp(i\gamma\rho) \frac{d\alpha}{\gamma} \\
&\underset{\rho\to\infty}{\sim} \sqrt{2\pi} \sum_{\nu=-\infty}^{\infty} s_{\nu\beta}^{1'} \frac{1}{2\pi} \int_{-\infty}^{\infty} \Psi_n(\hat{\boldsymbol{\kappa}}) \exp(i\gamma\rho) \frac{d\alpha}{\gamma} \\
&\underset{\rho\to\infty}{\sim} \sqrt{2\pi} \sum_{\nu=-\infty}^{\infty} s_{\nu\beta}^{1'} u_{n\beta}(\rho\hat{\mathbf{z}}) \\
&\underset{\rho\to\infty}{\sim} \frac{1}{\sqrt{2\pi}} F_n(\kappa\rho) \sum_{\nu=-\infty}^{\infty} s_{\nu\beta}^{1'}
\end{aligned} \tag{19.136}$$

and similarly,

$$\sigma_{n\beta}^{2'}(\rho) \underset{\rho\to\infty}{\sim} -\frac{1}{\sqrt{2\pi}} F_n(\kappa\rho) \sum_{\nu=-\infty}^{\infty} s_{\nu\beta}^{2'} \tag{19.137}$$

But

$$\mathbf{s}_\pi'\left(\frac{\kappa}{k}\hat{\mathbf{z}} + \frac{\beta}{k}\hat{\mathbf{y}}\right) = \frac{1}{\sqrt{2\pi}} \sum_{\nu=-\infty}^{\infty} \left(s_{\nu\beta}^{1'}\hat{\boldsymbol{\Phi}} - i s_{\nu\beta}^{2'}\hat{\boldsymbol{\Theta}}\right) \tag{19.138}$$

so that

$$\begin{aligned}
\sigma_{n\beta}^{1'}(\rho) &\underset{\rho\to\infty}{\sim} F_n(\kappa\rho)\hat{\boldsymbol{\Phi}} \cdot \mathbf{s}_\pi'\left(\frac{\kappa}{k}\hat{\mathbf{z}} + \frac{\beta}{k}\hat{\mathbf{y}}\right) \\
\sigma_{n\beta}^{2'}(\rho) &\underset{\rho\to\infty}{\sim} -i F_n(\kappa\rho)\hat{\boldsymbol{\Theta}} \cdot \mathbf{s}_\pi'\left(\frac{\kappa}{k}\hat{\mathbf{z}} + \frac{\beta}{k}\hat{\mathbf{y}}\right)
\end{aligned} \tag{19.139}$$

Formulas (19.139) provide a convenient numerical check in the evaluation of the translated probe coefficients. They also suggest, even for intermediate separations when the test antenna is in the "far field" of the probe, that the $\Phi = 0$ cut of $\mathbf{s}_\pi'$ is most important.

REFERENCES

1. J. S. Hollis, T. J. Lyon, and L. Clayton, *Microwave Antenna Measurements*, 3rd ed., Scientific-Atlanta, Inc., 1985.

2. D. M. Kerns and E. S. Dayhoff, Theory of diffraction in microwave interferometry, *J. Res. Natl. Bur. Stand.*, Vol. 64B, pp. 1–13, January–March 1960.
3. D. M. Kerns, Analytical techniques for the correction of near-field antenna measurements made with an arbitrary but known antenna, presented at the URSI-IRE Meeting, Washington, DC, 29 April–2 May, 1963.
4. D. M. Kerns, Correction of near-field antenna measurements made with an arbitrary but known antenna, *Electron. Lett.*, Vol. 6, pp. 346–347, 28 May, 1970.
5. J. Brown and E. B. Jull, The prediction of aerial radiation patterns from near-field measurements, *Proc. Inst. Elec. Eng.*, pp. 635–644, November 1961.
6. F. Jensen, Electromagnetic near-field far-field correlations, Ph.D. dissertation, Technical University of Denmark, Lyngby, 1970.
7. F. Jensen, On the probe compensation for near-field measurements on a sphere, *Arch. Elek. Übertagung.*, Vol. 29, pp. 306–308, July/August 1975.
8. F. Holm Larsen, Probe correction of spherical near-field measurements, *Electron. Lett.*, Vol. 13, pp. 393–395, July 1977.
9. P. F. Wacker, Non-planar near-field measurements: spherical scanning, National Bureau of Standards (U.S.), NBSIR 75-809, June 1975.
10. W. M. Leach, Jr. Probe compensated near-field measurements on a cylinder, Ph.D. dissertation, Georgia Institute of Technology, Atlanta, GA, August 1972.
11. W. M. Leach, Jr. and D. T. Paris, Probe compensated near-field measurements on a cylinder, *IEEE Trans. Antennas Propag.*, Vol. 21, pp. 435–445, July 1973.
12. A. D. Yaghjian, Near-field antenna measurements on a cylindrical surface: A source scattering-matrix approach, National Bureau of Standards Tech. Note 696, September 1977.
13. E. S. Gillespie (Ed.), Special Issue on Near-Field Scanning Techniques, *IEEE Trans. Antennas Propag.*, Vol. 36, pp. 725–901, June 1988.
14. R. C. Baird, A. C. Newell, and C. F. Stubenrauch, A brief history of near-field measurements at the National Bureau of Standards, *IEEE Trans. Antennas Propag.*, Vol. 36, pp. 727–733, June 1988.
15. J. E. Hansen and F. Jensen, Spherical near-field scanning at the Technical University of Denmark, *IEEE Trans. Antennas Propag.*, Vol. 36, pp. 734–739, June 1988.
16. E. B. Joy, A brief history of the development of the near-field measurement technique at the Georgia Institute of Technology, *IEEE Trans. Antennas Propag.*, Vol. 36, pp. 740–745, June 1988.
17. A. D. Yaghjian, An overview of near-field antenna measurements, *IEEE Trans. Antennas Propag.*, Vol. 34, pp. 30–45, January 1986.
18. J. Appel-Hansen, E. S. Gillespie, T. G. Hickman, and J. D. Dyson, Antenna measurements, in *The Handbook of Antenna Design*, A. W. Rudge, K. Milne, A. D. Olver, and P. Knight (Eds.), Peter Peregrinus Ltd., London, 1982, Chap. 8.
19. D. M. Kerns, Scattering matrix description and nearfield measurements of electroacoustic transducers, *J. Acoust. Soc. Am.*, Vol. 57, pp. 497–507, February 1975.
20. A. D. Yaghjian, Simplified approach to probe-corrected spherical near-field scanning, *Electron. Lett.*, Vol. 20, pp. 195–196, 1 March 1984.
21. R. C. Wittmann, Probe-corrected spherical near-field scanning in acoustics, *IEEE Trans. Instrum. Meas.*, Vol. 41, pp. 17–21, February 1992.
22. T. B. Hansen, Probe-corrected near-field measurements on a truncated cylinder, *J. Acoust. Soc. Am.*, Vol. 119, pp. 792–807, February 2006.
23. A. D. Yaghjian and R. C. Wittmann, The receiving antenna as a linear differential operator: application to spherical near-field measurements, *IEEE Trans. Antenna Propag.*, Vol. 33, pp. 1175–1185, November 1985.

24. R. C. Wittmann and C. F. Stubenrauch, Spherical near-field scanning: Experimental and theoretical studies, National Institute of Standards and Technology, NISTIR 3955, July 1990.
25. D. M. Kerns, Plane-wave scattering matrix theory of antennas and antenna–antenna interactions, National Bureau of Standards (U.S.) Monograph 162, 1981.
26. J. E. Hansen (Ed.), *Spherical Near-Field Antenna Measurements*, Peter Peregrinus Ltd., London, 1988.
27. V. Rokhlin, and M. Tygert, Fast algorithms for spherical harmonic expansions, *SIAM J. Sci. Comp.*, Vol. 27, pp. 1903–1928, 2006.
28. T. B. Hansen and A. D. Yaghjian, *Plane-Wave Theory of Time Domain Fields*, IEEE Press, Piscataway, NJ, 1999.
29. A. Messiah, *Quantum Mechanics*, Vol. 2, North-Holland Publishing Company, Amsterdam, 1963.
30. D. M. Kerns and R. W. Beatty, *Basic Theory of Waveguide Junctions and Introductory Microwave Network Analysis*, Pergamon, Oxford, UK, 1967.
31. *IEEE Standard Definitions of Terms for Antennas*, IEEE Std 145- 1993.
32. A. C. Newell, R. D. Ward, and E. J. McFarlane, Gain and power parameter measurements using planar near-field techniques, *IEEE Trans. Antennas Propag.*, Vol. 36, pp. 792–803, June 1988.
33. A. D. Yaghjian, Efficient computation of antenna coupling and fields within the near-field region, *IEEE Trans. Antennas Propag.*, Vol. 30, pp. 113–128, January 1982.
34. C. Müller, *Foundations of the Mathematical Theory of Electromagnetic Waves*, Springer-Verlag, Berlin, 1969.
35. R. C. Wittmann, Spherical wave operators and the translation formulas, *IEEE Trans. Antennas Propag.*, Vol. 36, pp. 1078–1087, August 1988.
36. J. D. Jackson, *Classical Electrodynamics*, 3rd ed., John Wiley & Sons, Hoboken NJ, 1999.
37. M. E. Rose, *Elementary Theory of Angular Momentum*, John Wiley & Sons, Hoboken, NJ, 1957.
38. M. Tinkham, *Group Theory and Quantum Mechanics*, McGraw-Hill, New York, 1964.
39. A. R. Edmonds, *Angular Momentum in Quantum Mechanics*, 3rd ed., Princeton University Press, Princeton, NJ, 1974.
40. M. Abramowitz and I. E. Stegun (Eds.), *Handbook of Mathematical Functions*, National Bureau of Standards (U.S.) Applied Mathematics Series Vol. 55, 1972.
41. I.S. Gradshteyn, and I. M. Ryzhik. *Table of Integrals, Series, and Products*, Academic Press, San Diego, 1980.
42. A. G. Repjar, A. C. Newell, and M. H. Francis, Accurate determination of planar near-field correction parameters for linearly polarized probes, *IEEE Trans. Antennas Propag.*, Vol. 36, pp. 855–868, June 1988.
43. A. C. Newell, M. H. Francis, and D. P. Kremer, The determination of near-field correction parameters for circularly polarized probes, *Proc. Antenna Measurement Tech. Assoc.*, Vol. 6, pp. 3A3-2–3A3-29, 1984.
44. A. C. Newell, D. P. Kremer, and J. R. Guerrieri, Improvements in polarization measurements of circularly polarized antennas, *Proc. Antenna Measurement Tech. Assoc.*, Vol. 11, pp. 1-30–1-35, October 1989.
45. D. P. Kremer, W. J. Parker, A. C. Newell, and F. Mayo-Wells, Tests of the fire performance of microwave absorber, *Proc. Antenna Measurement Tech. Assoc.*, Vol. 12, pp. 3–19, 1990.
46. D. W. Hess, Principle of the three-cable method for compensation of cable variations, *Proc. Antenna Measurement Tech. Assoc.*, Vol. 14, pp. 10-26–10-32, 1992.
47. A. C. Newell, Error analysis techniques for planar near-field measurements, *IEEE Trans. Antennas Propag.*, Vol. 36, pp. 754–768, June 1988.

48. A. D. Yaghjian, Upper-bound errors in far-field antenna parameters determined from planar near-field measurements, National Bureau of Standards Tech. Note 667, October 1975.

49. R. C. Wittmann, B. K. Alpert, and M. H. Francis, Near-field antenna measurements using nonideal measurements locations, *IEEE Trans. Antennas Propag.*, Vol. 46, pp. 716–722, May 1998.

50. Y. Rahmat-Samii, V. Galindo-Israel, and R. Mittra, A plane-polar approach for far-field construction from near-field measurements, *IEEE Trans. Antennas Propag.*, Vol. 28, pp. 216–230, March 1980.

51. R. G. Yaccarino, Y. Rahmat-Samii, and L. I. Williams, The bi-polar planar near-field measurement technique—Part II: Near-field to far-field transformation and holographic imaging methods, *IEEE Trans. Antennas Propag.*, Vol. 42, pp. 196–204, February 1994.

52. L. I. Williams, Y. Rahmat-Samii, and R. G. Yaccarino, The bi-polar planar near-field measurement technique—Part I: Implementation and measurement comparisons, *IEEE Trans. Antennas Propag.*, Vol. 42, pp. 184–195, February 1994.

53. O. M. Bucci, F. D'Agostino, C. Gennarelli, G. Riccio, and C. Savarese, Probe compensated far-field reconstruction by near-field planar spiral scanning, *Proc. IEE Microwave Antennas Propag.*, Vol. 149, pp. 119–123, April 2002.

54. R. C. Wittmann, C. F. Stubenrauch, and M. H. Francis, Spherical scanning measurements using truncated data sets, *Proc. Antenna Measurement Techn. Assoc.*, Vol. 24, pp. 279–283, November 2002.

55. R. C. Wittmann, B. K. Alpert, and M. H. Francis, Near-field, spherical scanning antenna measurements with non-ideal probe locations, *IEEE Trans. Antennas Propag.*, Vol. 52, pp. 2184–2186, August 2004.

56. O. M. Bucci, F. D'Agostino, C. Gennarelli, G. Riccio, and C. Savarese, Field recovery over a sphere from a minimum number of data over a spiral, in *ANTEM 2000*, Winnipeg, Canada, July 2000, pp. 187–190.

57. O. M. Bucci, C. Gennarelli, G. Riccio, and C. Savarese, Probe compensated NF-FF transformation with helicoidal scanning, *J. Electromagn. Waves Appl.*, Vol. 14, pp. 531–549, 2000.

58. M. H. Francis, A. C. Newell, K. R. Grimm, J. Hoffman, and H. Schrank, Planar near-field measurement of low sidelobe antennas, *J. Res. Natl. Inst. Stand. Tech.*, Vol. 99, pp. 143–167, March–April 1994.

59. M. H. Francis, A. C. Newell, K. R. Grimm, J. Hoffman, and H. Schrank, Comparison of ultralow-sidelobe antenna far-field patterns using the planar-near-field method and the far-field method, *IEEE Antennas Propag. Mag.*, Vol. 37, pp. 7–15, December 1995.

60. N. Marcuvitz, *Waveguide Handbook*, McGraw-Hill, New York, 1951.

61. G. V. Borgiotti, Integral equation formulation for probe corrected far-field reconstruction from measurements on a cylinder, *IEEE Trans. Antennas Propag.*, Vol. 26, pp. 572–578, July 1978.

CHAPTER 20

Antenna Measurements

CONSTANTINE A. BALANIS and CRAIG R. BIRTCHER

20.1 INTRODUCTION

Some antennas, because of their complex structural configurations or excitation methods, cannot be investigated analytically. Although the number of radiators that fall into this category has diminished because of advanced analytical methods (such as the method of moments, finite-difference time-domain, finite-element method, etc.) that have developed over the past few years, there are still a fair number that have not been examined analytically. Measurements are a practical method for characterizing the performance of such antennas. In addition, experimental results are often needed to validate theoretical data or to verify antenna construction.

Experimental investigations suffer from a number of drawbacks such as:

1. For pattern measurements, the distance to the far-field region ($r > 2D^2/\lambda$) may be too long even for outdoor ranges. It also becomes difficult to keep unwanted reflections from the ground and the surrounding objects below acceptable levels.
2. In many cases, it may be impractical to move the antenna from the operating environment to the measuring site.
3. For some antennas, such as phased arrays, the time required to measure the necessary characteristics may be enormous.
4. Outside measuring systems provide an uncontrolled environment, and they do not possess an all-weather capability.
5. Enclosed measuring systems usually cannot accommodate large antenna systems (such as ships, aircraft, or large spacecraft).
6. Measurement techniques, in general, are expensive.

The parameters that often best describe an antenna system's performance are the pattern (amplitude and phase), gain, directivity, efficiency, impedance, current distribution, and polarization. Each of these topics is addressed briefly in this chapter, which has been largely adapted from Chapter 17 of *Antenna Theory: Analysis and Design* [1]. A more extensive and exhaustive treatment of these and other topics can be found in the *IEEE*

Modern Antenna Handbook. Edited by Constantine A. Balanis

Standard Test Procedures for Antennas [2], in a summarized journal paper [3], and in a book on microwave antenna measurements [4]. Much of the material in this chapter is drawn from these four sources. The authors recommend that the IEEE publication on test procedures for antennas becomes part of the library of every practicing antenna and microwave engineer.

There are a variety of antenna ranges, instrumentation, and techniques with which the parameters that characterize an antenna can be measured. These include outdoor ranges, rectangular and tapered anechoic chambers [5], compact and extrapolation ranges [6], and near-field probing techniques [6–9]. Some of these are more appropriate than the others for measuring certain classes of antennas or at particular operating frequencies. Although the topic of near-field measurements is included in this chapter, it is a subject of considerable complexity and is treated separately in Chapter 19 of this handbook.

20.2 ANTENNA RANGES

20.2.1 Fundamental Considerations

It is usually most convenient to perform antenna measurements with the test antenna in its receiving mode. If the test antenna is reciprocal, the receiving mode characteristics (gain, radiation pattern, etc.) are identical to those transmitted by the antenna. The ideal condition for measuring far-field radiation characteristics then is the illumination of the test antenna by plane waves: those uniform in amplitude and phase. Although this ideal condition is not achievable, it can be approximated by separating the test antenna from the illumination source by a large distance on an outdoor range. As this separation distance increases, the curvature of the spherical phasefront produced by the source antenna becomes more planar over the test antenna aperture. If the separation distance is equal to the inner boundary of the far-field region, $2D^2/\lambda$, where D is the largest dimension of the antenna or scatterer, then the maximum phase error of the incident field from an ideal plane wave is about 22.5°, as shown in Figure 20.1. In addition to phasefront curvature due to finite separation distances, reflections from the ground and

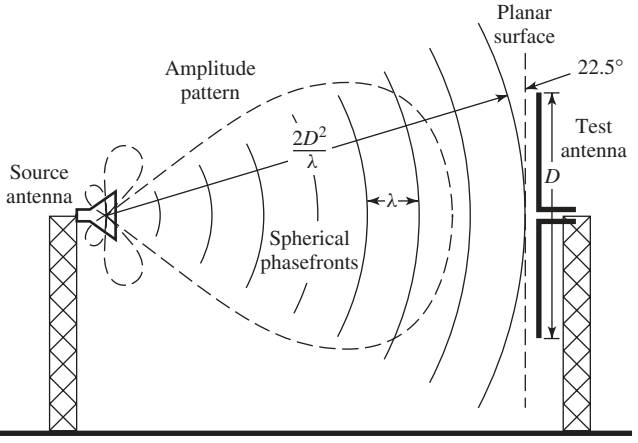


Figure 20.1 The phase error across the aperture of a test antenna in the far field when illuminated by a spherical wave.

nearby objects are possible sources of degradation of the test antenna illumination. Thus optimizing the planarity of the wavefronts illuminating the test antenna, while reducing or controlling environmental reflections, are the primary objectives of facilities designed to measure the far-field characteristics of antennas.

Measurements of test antennas are performed on antenna *ranges*. Antenna test facilities are categorized as either *outdoor* or *indoor* ranges, and there are limitations associated with both of them. Outdoor ranges are not protected from environmental conditions. Clearly, any kind of precipitation will have a negative impact on radiation measurements, but wind can also have the deleterious effect of causing undesired movements of the test and/or source antennas. Radiated phase is particularly susceptible to wind-induced error. Furthermore, when the measurements involve proprietary or classified information, achieving a satisfactory level of security may not be possible on an outdoor range. Another complication for outdoor ranges is their stray radiated emissions, which must be sufficiently low in power to preclude interfering with neighboring communications or navigation systems. This requirement conflicts with the increased radiated signal power necessary to compensate for the $1/r^2$ space attenuation that the signal accrues in propagating across the large separation distance between the test and auxiliary antennas. On very large outdoor ranges, the inconvenience of traveling between the transmitting and receiving antenna locations can be nontrivial.

Indoor facilities are primarily limited by space restrictions. Because some of the antenna characteristics are measured in the receiving mode and require that the far-field criteria be satisfied, the ideal field incident upon the test antenna should be a uniform plane wave. To meet this specification, a large space is usually required. The finite space enclosed by an indoor facility limits the size and/or frequency range of the test antenna that can be measured. On balance, however, indoor ranges are generally more convenient than, and are usually favored over, outdoor ranges.

20.2.2 Reflection Ranges

Antenna measurement ranges can also be categorized as either *reflection* or *free-space* ranges. The reflection ranges, if judiciously designed [10], can create a constructive interference in the region of the test antenna, which is referred to as the "quiet zone." This is accomplished by designing the ranges so that specular reflections from the ground, as shown in Figure 20.2, combine constructively with the direct illumination.

Usually it is desirable for the illuminating field to have a small and symmetric amplitude taper. This can be achieved by adjusting the transmitting antenna height while maintaining constant that of the receiving antenna. These ranges are of the outdoor type, where the ground is the reflecting surface. Although they are usually employed in the ultra-high frequency (UHF) region for measuring the patterns of moderately broad beam antennas, they can also be used for systems operating at higher frequencies.

20.2.3 Free-Space Ranges

Free-space ranges are designed to suppress the contributions from the surrounding environment and include *elevated ranges, slant ranges* [11], *anechoic chambers*, and *compact ranges* [6].

20.2.3.1 Elevated Ranges Elevated ranges are usually designed to operate mostly over smooth terrains. The antennas are mounted on towers or roofs of adjacent buildings.

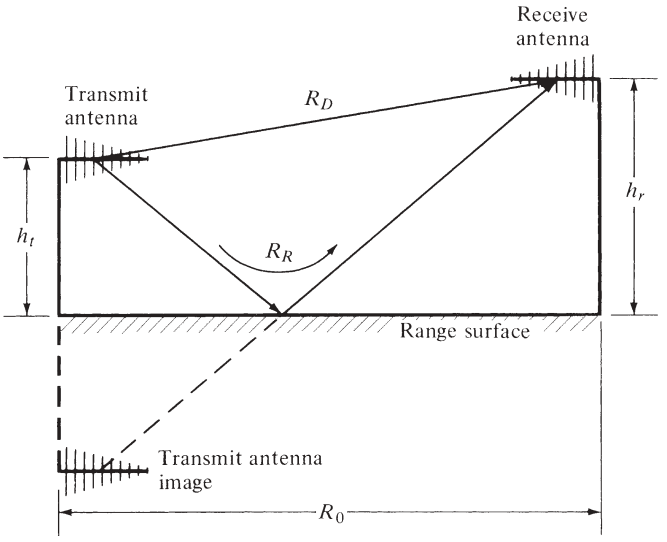


Figure 20.2 Geometrical arrangement for reflection range. (From Ref. 10, © 1973 IEEE.)

These ranges are used to test physically large antennas. A geometrical configuration is shown in Figure 20.3a. The contributions from the surrounding environment are usually reduced or eliminated [7] by (1) carefully selecting the directivity and sidelobe level of the source antenna; (2) clearing the line-of-sight between the antennas; (3) redirecting or absorbing any energy that is reflected from the range surface and/or from any obstacles that cannot be removed; and (4) utilizing special signal-processing techniques such as modulation tagging of the desired signal, or by using short pulses combined with range gating.

In some applications, such as between adjacent mountains or hilltops, the ground terrain may be irregular. For these cases, it is more difficult to locate the specular reflection points (points that reflect energy toward the test antenna). To take into account the irregular surface, scaled drawings of the vertical profile of the range are usually constructed from data obtained from the U.S. Geological Survey. The maps show ground contours [12], and they give sufficient details that can be used to locate the specular reflection points, determine the level of energy reflected toward the test antenna, and make corrections if it is excessive.

20.2.3.2 Slant Ranges Slant ranges [11] are designed so that the test antenna, along with its positioner, are mounted at a fixed height on a nonconducting tower while the source (transmitting) antenna is placed near the ground, as shown in Figure 20.3b. The source antenna is positioned so that the pattern maximum, of its free-space radiation, is oriented toward the center of the test antenna. The first null is usually directed toward the ground specular reflection point to suppress reflected signals. Slant ranges, in general, are more compact than elevated ranges in that they require less real estate.

20.2.3.3 Anechoic Chambers To provide a controlled environment, all-weather capability, security, and to minimize electromagnetic interference, indoor anechoic chambers have been developed as an alternative to outdoor testing. By this method, the

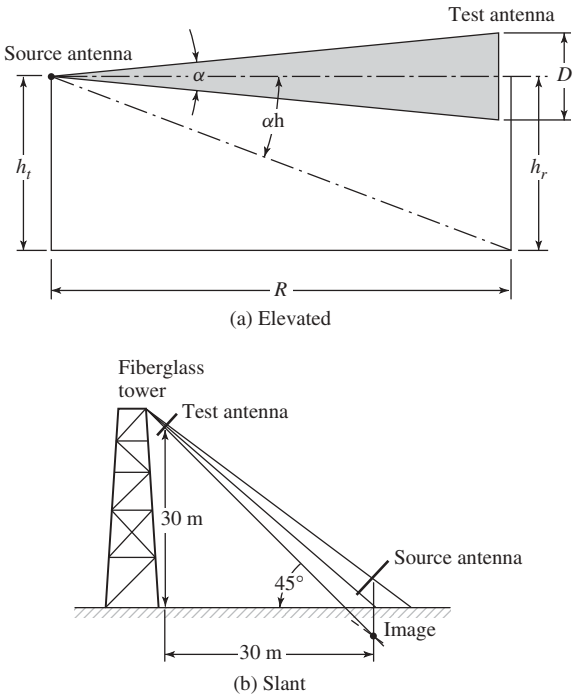


Figure 20.3 Geometries of (a) elevated and (b) slant ranges. (Part (a) from Ref. 2; part (b) from Ref. 11, © 1966 IEEE.)

testing is performed inside a chamber having walls that are covered with radiofrequency (RF) absorbers. The availability of commercial high quality RF absorbing material, with improved electrical characteristics, has provided the impetus for the development and proliferation of anechoic chambers. Anechoic chambers are mostly utilized in the microwave region, but materials have been developed [13] which provide a normal-incidence reflection coefficient of −40 dB at frequencies as low as 100 MHz. In general, as the operating frequency is lowered, the thickness of RF absorbing material must be increased to maintain a given level of reflectivity performance. An RF absorber that meets the minimum electrical requirements at the lower frequencies usually possesses improved performance at higher frequencies.

Presently there are two basic types of anechoic chamber designs: the *rectangular* and the *tapered chamber*. The design of each is based on geometrical optics techniques, and each attempts to reduce or to minimize specular reflections. The geometrical configuration of each, with specular reflection points depicted, is shown in Figure 20.4.

The rectangular chamber [14] is usually designed to simulate free-space conditions and maximize the volume of the quiet zone. The design takes into account the pattern and location of the source and the frequency of operation, and it assumes that the receiving antenna at the test point is isotropic. Reflected energy is minimized by the use of high quality RF absorbers. Despite the use of RF absorbing material, significant specular reflections can occur, especially at large angles of incidence.

Tapered anechoic chambers [15] take the form of a pyramidal horn. They begin with a tapered chamber that leads to a rectangular configuration at the test region, as shown in

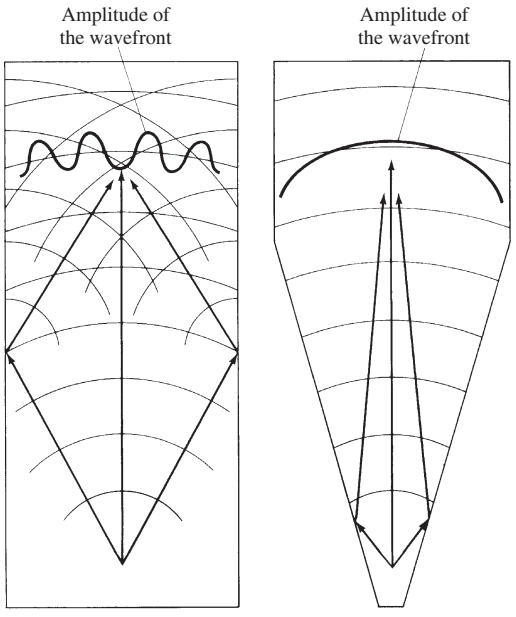


Figure 20.4 (a) Rectangular and (b) tapered anechoic chambers and the corresponding sidewall specular reflections. (From Ref. 3, © 1978 IEEE.)

Figure 20.4b. At the lower end of the frequency band for which the chamber is designed, the source is usually placed near the apex so that the reflections from the sidewalls, which contribute to the illuminating fields in the region of the test antenna, occur near the source antenna. For such paths, the phase difference between the direct radiation and that reflected from the walls near the source can be made very small by properly locating the source antenna near the apex. Thus the direct and reflected rays near the test antenna region add vectorially and provide a relatively smooth amplitude illumination taper. This can be illustrated by ray-tracing techniques.

As the frequency of operation increases, it becomes increasingly difficult to place the source sufficiently close to the apex such that the phase difference between the direct and specularly reflected rays can be maintained below an acceptable level. For such applications, reflections from the walls of the chamber are suppressed by using high gain source antennas whose radiation toward the walls is minimal. In addition, the source is moved away from the apex, and it is placed closer to the end of the tapering section so as to simulate a rectangular chamber.

20.2.4 Compact Ranges

Microwave antenna measurements require that the radiator under test be illuminated by a uniform plane wave. This is usually achieved only in the far-field region, which in many cases dictates very large distances, as illustrated in Figure 20.1. The requirement of an ideal plane wave illumination can be nearly achieved by utilizing a compact range.

A *c*ompact *a*ntenna *t*est *r*ange (CATR) is a collimating device that generates nearly planar wavefronts in a very short distance (typically 10–20 meters) compared to the

$2D^2/\lambda$ (minimum) distance required to produce the same size test region using the standard system configuration of testing shown in Figure 20.1. Some attempts have been made to use dielectric lenses as collimators [16], but generally the name compact antenna test range refers to one or more curved metallic reflectors that perform the collimating function. Compact antenna test ranges are essentially very large reflector antennas designed to optimize the planar characteristics of the fields in the near field of the aperture. Compact range configurations are often designated according to their analogous reflector antenna configurations: *parabolic, Cassegrain, Gregorian*, and so forth.

One compact range configuration is that shown in Figure 20.5, where a source antenna is used as an offset feed that illuminates a paraboloidal reflector, which converts the impinging spherical waves into plane waves [6]. Geometrical optics (GO) is used in Figure 20.5 to illustrate general CATR operation. The rays from a feed antenna can, over the main beam, be viewed as emanating from a point at its phase center. When the phase center of the feed is located at the prime focus of a parabolic reflector, all rays that are reflected by the reflector and arrive at a plane transverse to the axis of the parabola have traveled an equal distance. Therefore the field at the aperture of the reflector has a uniform phase (e.g., that of a plane wave).

In addition to GO, analysis and design of CATRs have been performed with a number of other analytical methods. Compact range test zone fields have been predicted by the method of moments (MoM), but at high frequencies, the large electrical size of the CATR system makes the use of MoM, finite-difference time-domain (FDTD), and finite-element method (FEM) impractical. High frequency techniques, however, are well suited for compact range analysis because the fields of interest are near the specular reflection direction, and the reflector is electrically large. The geometrical theory of diffraction (GTD) is, in principle, an appropriate technique, but it is difficult to implement for serrated-edge reflectors due to the large number of diffracting edges. To date, physical optics (PO) is probably the most practical and efficient method of predicting the performance of CATRs [17, 18].

The major drawbacks of compact ranges are aperture blockage, direct radiation from the source to the test antenna, diffractions from the edges of the reflector and feed

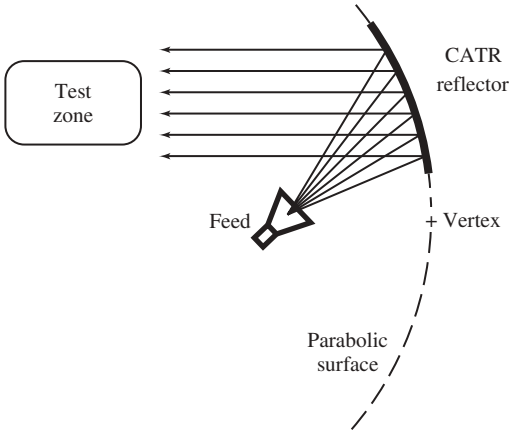


Figure 20.5 A compact antenna test range (CATR) synthesizes planar phasefronts by collimating spherical waves with a section of paraboloidal reflector.

support, depolarization coupling between the two antennas, and wall reflections. The use of an offset feed eliminates aperture blockage and reduces diffractions. Direct radiation and diffractions can be reduced further if a reflector with a long focal length is chosen. With such a reflector, the feed can then be mounted below the test antenna and the depolarization effects associated with curved surfaces are reduced. Undesirable radiation toward the test antenna can also be minimized by the use of high quality absorbing material. These and other concerns are discussed briefly.

20.2.4.1 CATR Designs Four reflector configurations that have been commercially developed are briefly discussed: the *single paraboloid*, the *dual parabolic-cylinder*, the *dual shaped-reflector*, and the *single parabolic-cylinder systems*. The first three configurations are relatively common fully collimating compact ranges; the fourth is a hybrid approach that combines aspects of compact range technology with near-field/far-field (NF/FF) techniques.

The single paraboloidal reflector CATR design is illustrated in Figure 20.5. As with all compact range designs, the feed antenna is offset by some angle from the propagation direction of the collimated energy. An installation of this type of CATR system is depicted in Figure 20.6. This is done to eliminate blockage and to reduce scattering of the collimated fields by the feed. To achieve this offset, the reflector is a sector of a paraboloid that does not include the vertex. This design is referred to as a *virtual vertex* compact range. With only one reflector, the paraboloidal CATR has a minimum number of surfaces and edges that can be sources of quiet-zone ripple. Feed spillover into the quiet zone is also low with this design since the feed antenna is pointed almost directly away from the test zone. On the other hand, it is more difficult and costly to produce a high precision surface that is curved in two planes (three dimensionally) compared to producing a reflector that is curved in only one plane (two dimensionally). In addition, it has been reported that the single paraboloidal reflector design depolarizes the incident fields to a greater degree than other CATR designs. This is due to the relatively low *f/d*

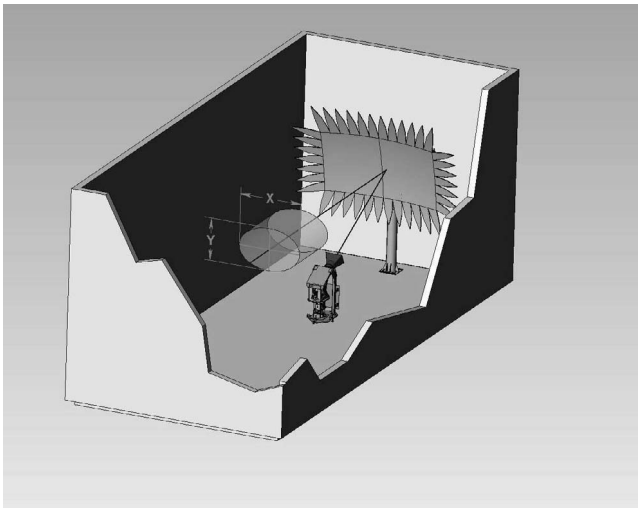


Figure 20.6 An artist's rendering of an anechoic chamber equipped with a single paraboloidal reflector CATR. (Courtesy of MI Technologies.)

ratio needed to simultaneously maintain the feed antenna between the test zone and the reflector while keeping the test zone as close as possible to the reflector aperture [19]. A photograph of a single paraboloidal CATR reflector, manufactured by MI Technologies [20], is shown in Figure 20.7.

The dual parabolic-cylinder reflector concept is illustrated in Figure 20.8, and a sketch of an installation of such a system is shown in Figure 20.9. A photograph of a dual parabolic-cylinder CATR system manufactured by March Microwave System B.V. is shown in Figure 20.10. It consists of two parabolic cylinders arranged so that one is curved in one plane (vertical or horizontal) while the other is curved in the orthogonal plane. The spherical phasefronts radiated by the feed antenna are collimated first in the horizontal or vertical plane by the first reflector, then are collimated in the orthogonal plane by the second reflector [21]. Because the boresight of the feed antenna is directed at almost 90° to the plane wave propagation direction, direct illumination of the test zone

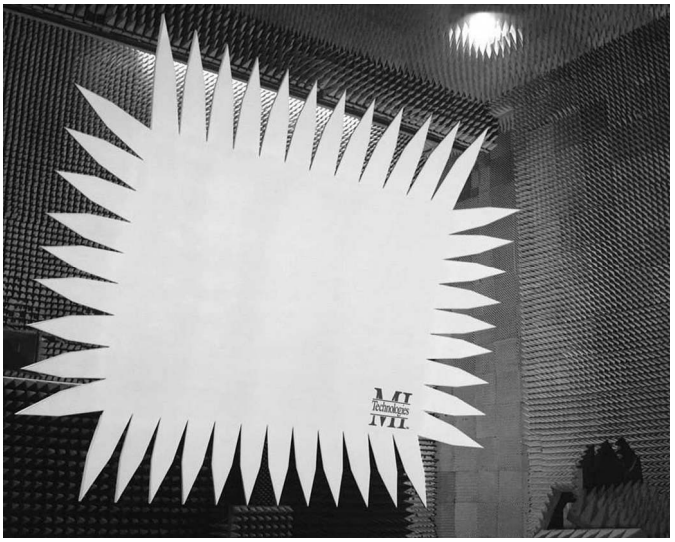


Figure 20.7 A photograph of a single paraboloidal CATR reflector. (Courtesy of MI Technologies.)

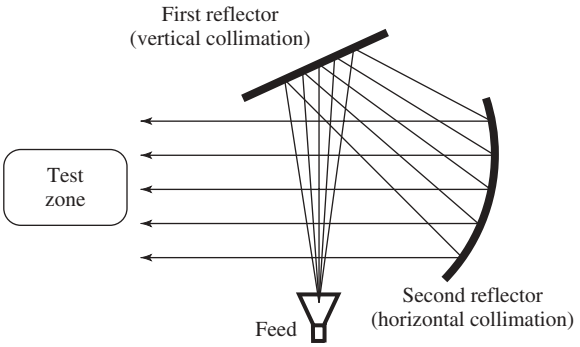


Figure 20.8 A dual parabolic-cylinder compact range collimates the fields in one plane with the first reflector and then collimates the fields in the orthogonal plane with the second reflector.

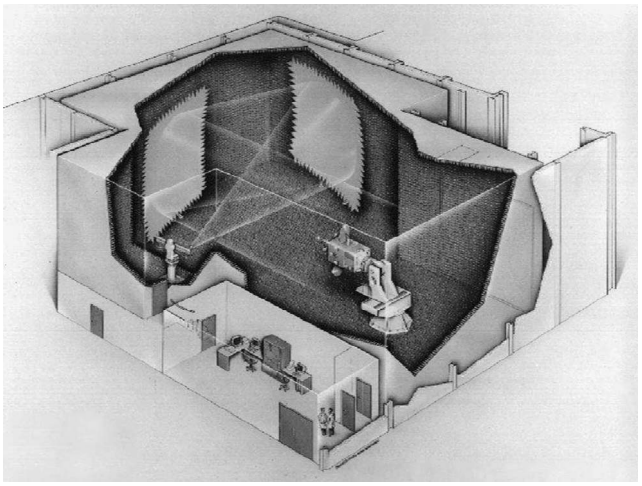


Figure 20.9 A dual parabolic-cylinder CATR installation. (Courtesy of March Microwave Systems B.V., Nuenen, The Netherlands.)

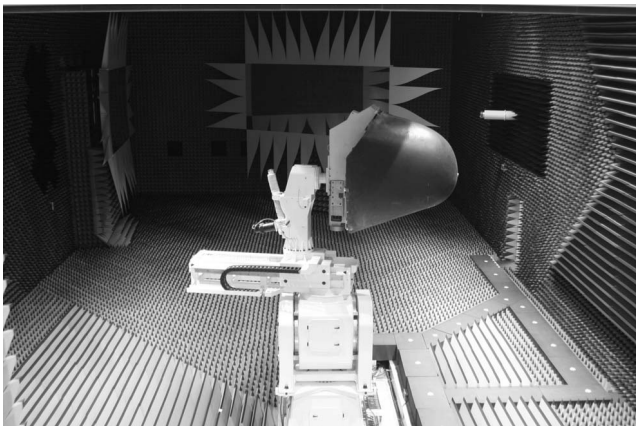


Figure 20.10 A photograph of a dual parabolic-cylinder CATR system. (Courtesy of SELEX Sensors and Airborne Systems, Edinburgh.)

by the feed can be relatively high. In practice, the effects of quiet-zone contamination from feed spillover are virtually eliminated through the use of range gating. Relatively low cross polarization is produced with this design because the doubly folded optics results in a long focal length main reflector.

The dual shaped-reflector CATR, shown schematically in Figure 20.11, is similar in design to a Cassegrain antenna, but the reflector surfaces are altered from the classical parabolic/hyperbolic shapes. An iterative design process is used to determine the shapes of the subreflector and main reflector needed to yield the desired quiet-zone performance. The shape of the subreflector maps the high gain feed pattern into a nearly optimum illumination of the main reflector. An almost uniform energy density illuminates the

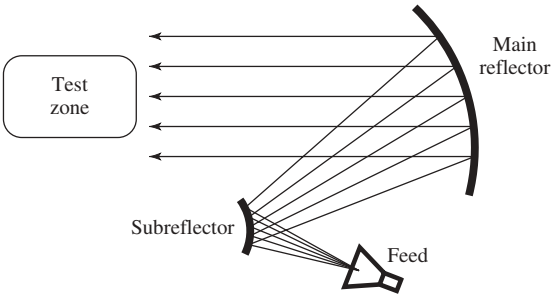


Figure 20.11 Dual shaped-reflector compact range analogous to a Cassegrain antenna system.

central part of the main reflector while the amplitude tapers toward the reflector edges. This design results in a very high illumination efficiency (the power of the collimated quiet-zone fields relative to the system input power) [23]. Two of the consequences of this high illumination efficiency are (1) the reduction of spillover into the chamber reduces range clutter, and (2) the increased RF power delivered to the target increases system sensitivity. A photograph of a Harris Corporation [24] dual shaped-reflector system is shown in Figure 20.12.

The single parabolic-cylinder reflector system is essentially half of the dual parabolic-cylinder CATR of Figures 20.9 and 20.10. The reflector has a parabolic curvature in the vertical plane and is flat in the horizontal plane. This semicompact antenna test range collimates the fields only in the vertical plane, producing a quiet zone that consists of cylindrical waves, as shown in Figure 20.13 [25–27]. Such a compact range configuration is utilized in the ElectroMagnetic Anechoic Chamber (EMAC) at Arizona State University (ASU) (Figure 20.14) [26, 27].

This *single-plane collimating range* (SPCR) approach results in a number of advantages and compromises compared to conventional CATR systems and near-field/far-field

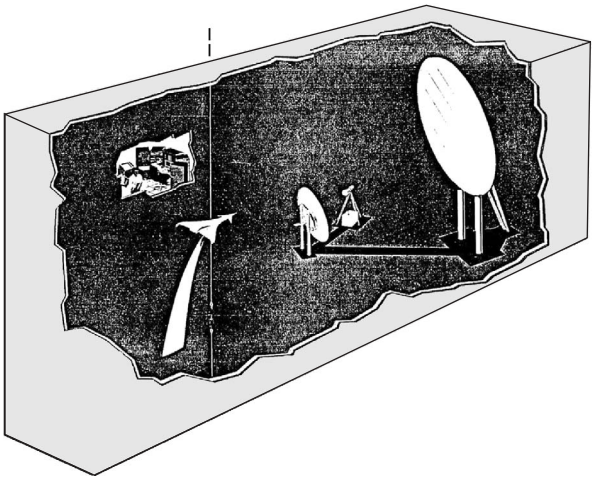


Figure 20.12 A dual shaped-reflector compact range system manufactured by Harris Corporation [22].

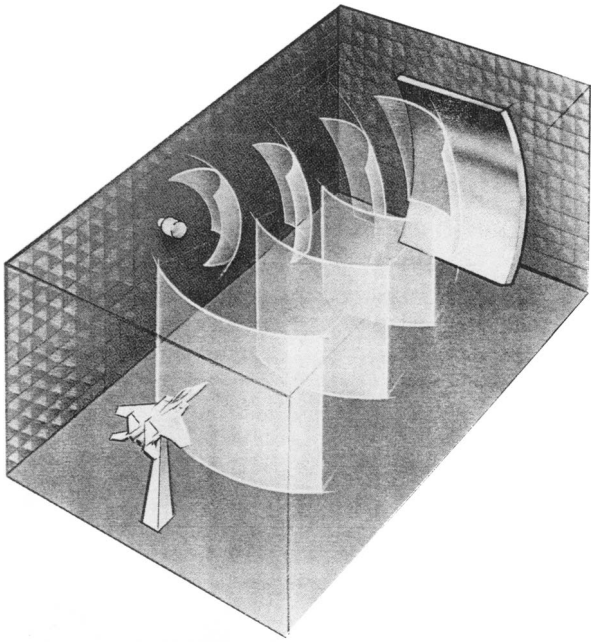


Figure 20.13 ASU single-plane collimating range (SPCR) produces a cylindrical wave in the quiet zone. (Artist rendering by Michael Hagelberg.)

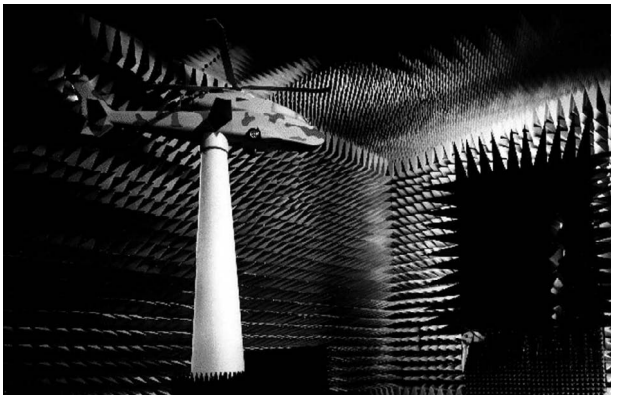


Figure 20.14 A photograph of a scale model helicopter in the quiet zone of ASU's single-plane collimating range.

(NF/FF) systems. For antennas that are small compared to the curvature of the cylindrical phasefront, far-field radiation patterns can be measured directly. Because of the folded optics, the radius of the cylindrical phasefront produced by the SPCR is larger than the radius of the spherical phasefront obtainable by separating the source antenna from the test antenna in a direct illumination configuration within the same anechoic chamber. Thus with the SPCR it is possible to measure directly the far-field patterns of larger

antennas compared to those directly measurable on an indoor far-field range. When the size of the antenna is large relative to the curvature of the cylindrical phasefront, a NF/FF transformation is used to obtain the far-field pattern. However, because the fields are collimated in the vertical plane, only a one-dimensional transformation is required. This greatly simplifies the transformation algorithm. Most importantly, there is a one-to-one correlation between a single azimuthal pattern cut measured in the near-field and the corresponding predicted far-field pattern. The data acquisition time is identical to that of conventional CATRs, and the NF/FF calculation time is nearly negligible. Another advantage of the SPCR is the size of the quiet zone. In the vertical plane, the quiet-zone dimension compared to the SPCR reflector is similar to that of conventional CATRs (about 50–60%). However, in the horizontal plane, the quiet zone is nearly 100% of the horizontal dimension of the reflector. For a given size anechoic chamber and reflector, targets having much larger horizontal dimensions (e.g., yaw patterns of aircraft) can be measured using the SPCR than is possible using a conventional CATR. The SPCR system is relatively inexpensive; the manufacturer estimates that its cost is about 60% of conventional CATR systems.

An example of an antenna radiation pattern measured under this cylindrical wave illumination is shown in Figure 20.15. In Figure 20.15a the uncorrected near-field azimuthal pattern of a 1.8-m dish antenna, measured using ASU's SPCR, is compared with its far-field pattern, measured on an outdoor antenna range at 10.7 GHz. After the NF/FF transformation has been applied to the near-field measurement, the far-field pattern of the antenna agrees extremely well with its outdoor measurement, as seen in Figure 20.15b.

In addition to the added complexity of NF/FF transformation considerations, this cylindrical wave approach has other disadvantages compared to conventional CATR designs. Because the quiet-zone fields are expanding cylindrically as they propagate along the axis of the range, a large portion of the anechoic chamber is directly illuminated. This should be carefully considered in the design of the sidewalls of the anechoic chamber to control range clutter. Also, some measurement sensitivity is sacrificed for the same reason.

20.2.4.2 CATR Performance A perfect plane wave would be produced by a CATR if the reflector has an ideal parabolic curvature, is infinite in size, and is fed by a point source located at its focus. Of course, CATR reflectors are of finite size, and their surfaces have imperfections; thus the test zone fields they produce can only approximate plane waves. Although there are different configurations of CATR, their test zone fields have some common characteristics. The usable portion of the test zone consists of nearly planar wavefronts and is referred to as the *quiet zone*. Outside the quiet zone, the amplitude of the fields decreases rapidly as a function of distance transverse to the range axis. The size of the quiet zone is typically about 50–60% of the dimensions of the main reflector. Although the electromagnetic field in the quiet zone is often a very good approximation, it is not a "perfect" plane wave. The imperfections of the fields in the quiet zone relative to an ideal plane wave are usually represented as phase errors, and ripple and taper amplitude components. These discrepancies from an ideal plane wave, which occur over a specified test zone dimension, are the primary figures of merit of CATRs. For most applications, phase deviations of less than 10°, peak-to-peak amplitude ripples of less than 1 dB, and amplitude tapers of less than 1 dB are considered adequate. More stringent quiet-zone specifications may be required to measure, within acceptable error levels, low sidelobe antennas and low-observable scatterers. The sources of quiet-zone taper and ripple are well known, but their minimization has been a source of much debate.

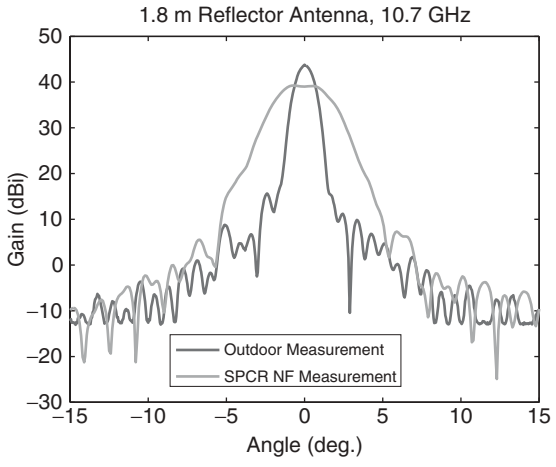
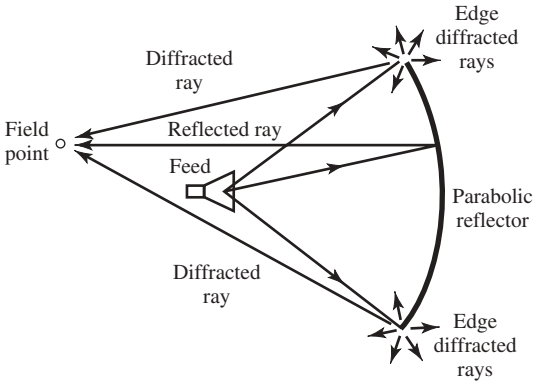


Figure 20.15 Comparisons of amplitude radiation patterns of a 1.8-meter dish antenna at 10.7 GHz, measured on an outdoor far-field range and measured at ASU's EMAC facility: (a) before applying the far-field transformation and (b) after the NF/FF transformation.

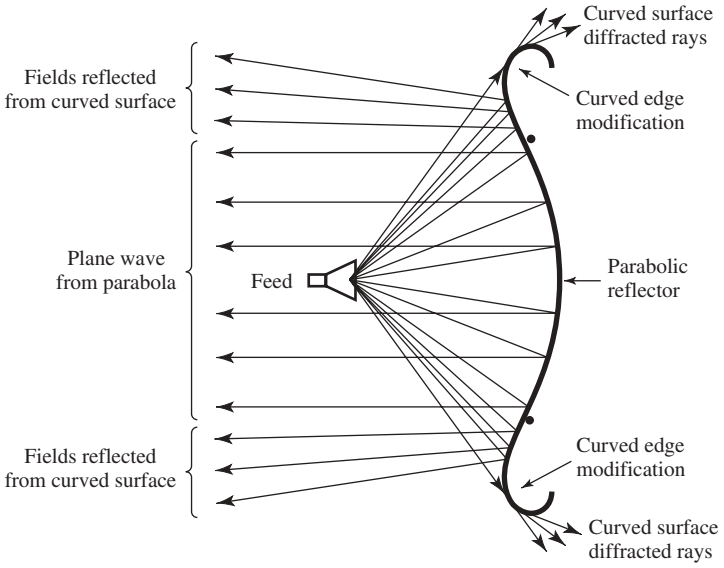
Amplitude taper across the quiet zone can be attributed to two sources: the feed pattern and space attenuation. That portion of the radiation pattern of the feed antenna that illuminates the CATR reflector is directly mirrored into the quiet zone. For example, if the 3-dB beamwidth of the feed is equal to about 60% of the angle formed by lines from the reflector edges to the focal point, then the feed will contribute 3 dB of quiet-zone amplitude taper. In general, as the directivity of the feed antenna increases, quiet-zone amplitude taper increases. Usually, low gain feed antennas are designed to add less than a few tenths of a decibel of amplitude taper. The other contributor of amplitude taper, the $1/r^2$ space attenuation, occurs because of the spherical spreading of the uncollimated radiation from the feed. Although the total path from the feed to the quiet zone is a constant, the distance from the feed to the reflector varies. These differences in the

propagation distances from the feed to various points across the reflector surface cause amplitude taper in the quiet zone due to space attenuation. This taper is asymmetric in the plane of the feed offset.

Amplitude and phase ripple are primarily caused by diffractions from the edges of the reflector. The diffracted fields spread in all directions, which, along with the specular reflected signal, form constructive and destructive interference patterns in the quiet zone, as shown in Figure 20.16a. Considerable research has been done on reflector edge terminations in an effort to minimize quiet-zone ripple. Reflector edge treatments are the physical analogs of windowing functions used in Fourier transforms. Edge treatments reduce the discontinuity of the reflector/free-space boundary, caused by the finite size of

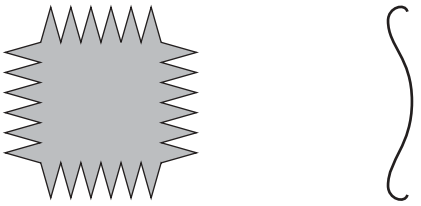


(a) Knife-edge



(b) Rolled-edge

Figure 20.16 Amplitude and phase ripple in the quiet-zone fields produced by a compact antenna test range caused by the phasor sum of the reflected and diffracted rays from the reflector. (From Ref. 30, © 1987 IEEE.)



(a) Front view of a serrated-edge CATR reflector.

(b) Side view of a rolled-edge CATR reflector.

Figure 20.17 Two common CATR reflector edge treatments that are used to reduce the diffracted fields in the quiet zone.

the reflector, by providing a gradually tapered transition. Common reflector edge treatments include serrations and rolled edges, as shown in Figure 20.17. The serrated edge of a reflector tapers the amplitude of the reflected fields near the edge. An alternate interpretation of the effects of serrations is based on edge diffraction. Serrations produce many low-amplitude diffractions as opposed to, for example, the large-amplitude diffractions that would be generated by the four straight edges and corners of a rectangular knife-edged reflector. These small diffractions are quasirandomized in location and direction; hence there is a propensity for them to cancel each other out when added together in the quiet zone. Although most serrated-edge CATRs have triangular serrations, curving the edges of each serration can result in improved performance at high frequencies [28].

A number of blended, rolled-edge treatments have been suggested as alternatives to serrations and have been implemented to gradually redirect energy away from the quiet zone, as shown in Figure 20.16b [29–31]. In these designs, the concave parabolic surface of the reflector is blended into a convex surface that wraps around the edges of the reflector and terminates behind it. The predicted quiet-zone fields produced by a knife-edged reflector compared to those produced by a rolled-edged reflector are shown in Figure 20.18 and demonstrate the effectiveness of this edge treatment.

Another method of reducing quiet-zone ripple is to taper the illumination amplitude near the reflector edges. This can be accomplished with a high gain feed or the feed can consist of an array of small elements designed so that a null in the feed pattern occurs at the reflector edges [32–35]. Finally, the surface currents on the reflector can be terminated gradually at the edges by tapering the conductivity and/or the impedance of the reflector via the application of lossy material.

The frequency of operation of a CATR is determined by the size of the reflector and its surface accuracy. The low frequency limit is usually encountered when the reflector is about 25–30 wavelengths in diameter [36]. Quiet-zone ripple becomes large at the low frequency limit. At high frequencies, reflector surface imperfections contribute to the quiet-zone ripple. A rule of thumb used in the design of CATRs is that the surface must deviate less than about 0.007λ from that of a true paraboloid [37]. Since the effects of reflector surface imperfections are additive, dual reflector systems must maintain twice the surface precision of a single reflector system to operate at the same frequency. Many CATR systems can operate from 1 to 100 GHz.

Quiet-zone probing is the most direct method of measuring the deviations of the fields produced in a measurement range from those of an ideal plane wave. To probe the quiet

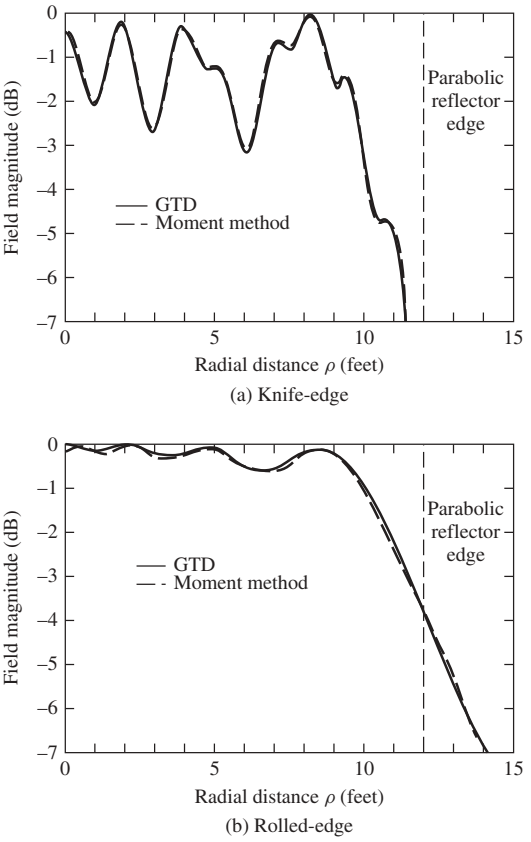


Figure 20.18 Predicted quiet-zone field amplitude versus transverse distance for knife-edge and rolled-edge reflectors. (From Ref. 30, © 1987 IEEE.)

zone, a small antenna or open-ended waveguide is moved in a direction perpendicular to the propagating fields with the aid of a linear scanner. The received amplitude and phase are recorded as a function of transverse position, usually along horizontal and vertical lines that cross at the center of the quiet zone. This is done for both horizontal and vertical polarizations, and at many different frequencies.

Examples of quiet-zone field probe results are shown in Figures 20.19–20.23. In these figures, the amplitude and phase variations of the quiet-zone fields produced by ASU's SPCR (Figures 20.13 and 20.14) are plotted for the case of vertical polarization and at a frequency of 7.0 GHz [27, 38]. The solid curve of Figure 20.19 is the measured amplitude as a function of vertical distance through the quiet zone. To resolve this measured scan into taper and ripple components, a least-squares fit curve (the dashed line of Figure 20.19) was calculated. The least-squares curve represents the amplitude taper of the fields. The value of the taper is defined as the difference between the maximum and minimum values that occur within a specified region. In this case, the amplitude taper over a 1.20-meter quiet zone (-60 to $+60$ cm) is 0.45 dB.

Normalizing the measured amplitude by the least-squares taper results in a plot, shown in Figure 20.20, from which the amplitude ripple can be found. Again, the value of the

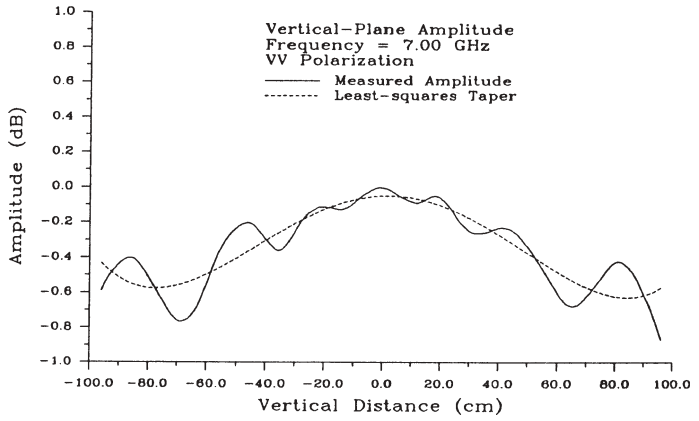


Figure 20.19 The measured amplitude of the quiet-zone field in the vertical plane produced by ASU's SPCR at 7.0 GHz (solid line), and a calculated least-squares fit curve (dashed line) that corresponds to the amplitude taper.

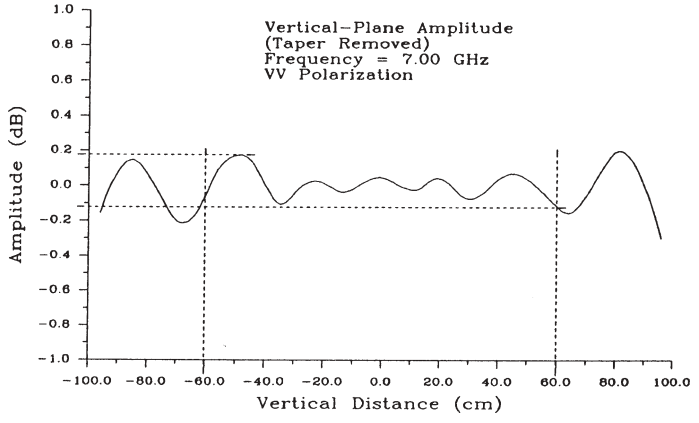


Figure 20.20 Normalizing the measured amplitude by the least-squares fit curve allows the amplitude ripple to be determined.

amplitude ripple is defined as the difference between the maximum and minimum values that occur within the specified zone. For a 1.20-meter zone in the vertical plane, the amplitude ripple at 7.0 GHz and vertical polarization is 0.3 dB.

The measured phase in the vertical plane is shown in Figure 20.21. Because the fields are collimated in the vertical plane, the phase is essentially flat, with a peak-to-peak variation over the 1.20-meter quiet zone of 2.0°.

Since the SPCR produces cylindrical waves, the measured phase in the horizontal plane of the quiet zone has a circular curvature, as shown in Figure 20.22. In this case, the figure of merit is the deviation from an ideal cylinder. Normalizing the measured horizontal-plane phase to an ideal cylinder results in the plot of Figure 20.23. It can now be seen that the horizontal-plane phase variation (relative to a cylinder) over the entire 1.92-meter length of the linear scanner is less than 2.5°.

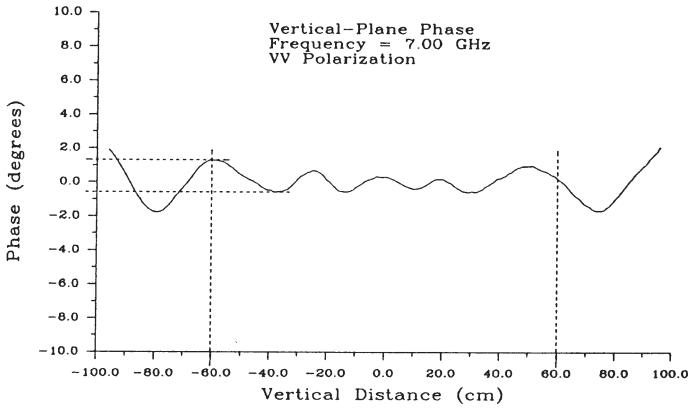


Figure 20.21 The measured phase produced by the SPCR in the vertical plane at 7.0 GHz.

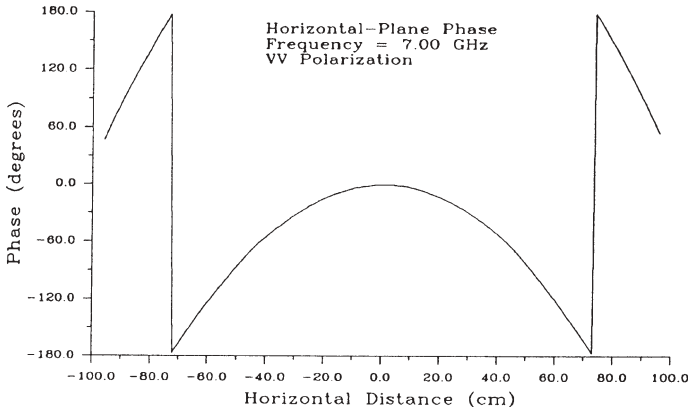


Figure 20.22 The measured phase produced by the SPCR in the horizontal plane at 7.0 GHz. The phase exhibits the curvature of a cylindrical wave.

In summary, compact antenna test ranges enable the measurement of full-sized antennas in very short distances, usually within the controlled environment of an anechoic chamber. A compact antenna test range can be used to accomplish any type of antenna testing (including radiation patterns, gain, and efficiency) that can be performed on an outdoor facility.

20.2.5 Near-Field/Far-Field Methods

The distribution of the fields radiated by an antenna does not assume the same amplitude (or phase) pattern at all radial distances; it evolves with increasing distance from the antenna. Three regions surrounding an antenna are recognized and are classified according to the development of the field distribution [1], as shown in Figure 20.24. In the space immediately surrounding an antenna (closer than approximately $0.62\sqrt{D^3/\lambda}$, where D is the largest dimension of the antenna), there is a significant electric field component that does not propagate radially and the energy is predominantly reactive—hence the

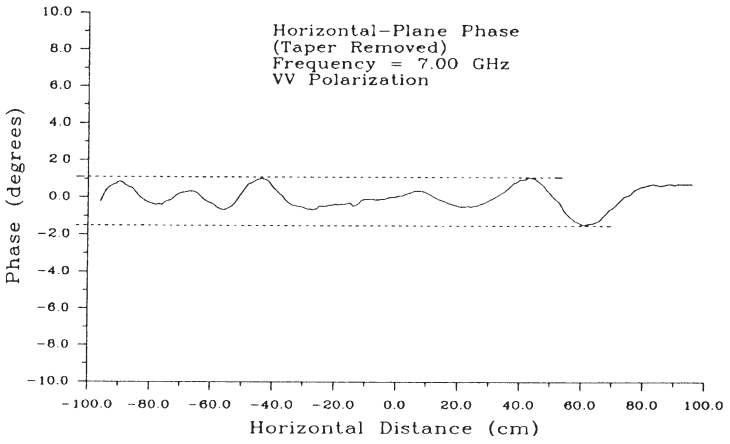


Figure 20.23 After normalizing the measured phase to an ideal cylinder, the phase variations in the horizontal plane are seen to be less than 2.5° over the 1.92-meter length of the available linear scanner.

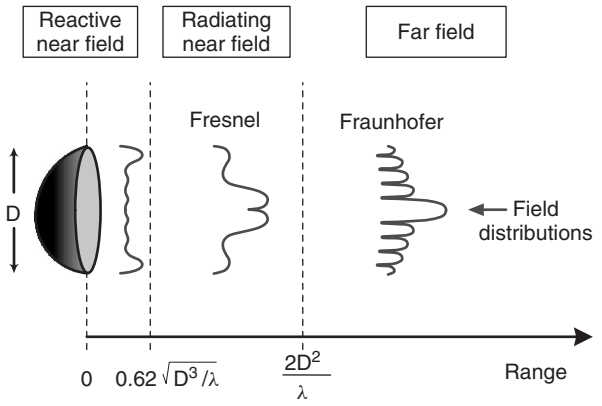
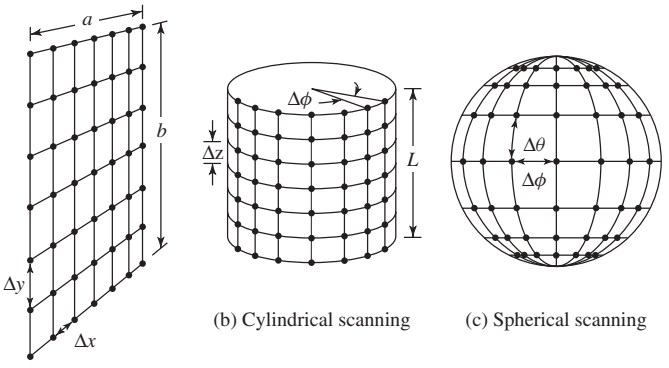


Figure 20.24 The evolution as a function of radial distance of the amplitude field distributions emanating from an antenna.

term "reactive near field." From approximately $0.62\sqrt{D^3/\lambda}$ to $2D^2/\lambda$, the propagation vectors gradually become, as the distance increases, radial. However, the contributions to the total radiated field at any observation point in this region do not have mutually identical phases. The main lobe of the radiation pattern in this *radiating near-field* region is frequently broadened and bifurcated. At ranges beyond approximately $2D^2/\lambda$, the differences in phase among the field components contributed from all parts of the antenna are negligible, and the radiation pattern approaches the same shape that would basically be observed at an infinite distance from the antenna.

The dimensions of a conventional test range can be reduced by making measurements in the radiating near field, and then using analytical methods to transform the measured data to the far-field radiation characteristics [6–9, 30]. These are referred to as *near-field*



(a) Planar scanning (b) Cylindrical scanning (c) Spherical scanning

Figure 20.25 The three near-field scanning surfaces that permit convenient data acquisition (planar, cylindrical, and spherical).

to far-field (NF/FF) methods. The advantages of these methods are offset by their more complex measurement systems, more extensive calibration procedures, and sophisticated data processing requirements. Nonetheless, the availability of "turnkey" NF/FF systems can make them an attractive alternative to other antenna measuring systems.

In this method, the amplitude and phase of the test antenna's electric field is sampled at predetermined positions over a geometric surface in the near field. From the near-field data, an angular spectrum of basis functions is generated. The basis functions are either planar, cylindrical, or spherical waves depending on the measurement scan geometry. A "probe correction" is applied to the angular spectrum to account for the directive and polarization effects of the measuring probe. The far-field radiation patterns of the test antenna can then be computed using the vector wave functions that are orthogonal to the scanning surface. Three coordinate systems (*planar, cylindrical*, and *spherical*) satisfy the geometrical requirements for near-field to far-field transformations and are conducive to convenient data acquisition. Those three near-field scanning surfaces, and the maximum data acquisition spacings that will ensure the accurate construction of far-field patterns according to sampling theory, are depicted in Figure 20.25.

The efficacy of (NF/FF) measurements is illustrated in Figure 20.26. The sum and difference patterns of a four-foot reflector test antenna, transformed from near-field data measured with a planar system, are compared with those patterns measured on far-field ranges. The agreement between the three patterns in both figures is excellent.

Near-field techniques are discussed in much greater detail in Chapter 19 of this handbook.

20.3 MICROWAVE ABSORBING MATERIAL

In this context, the terms *microwave absorbing material* or *absorbers* refer to materials that are used to cover a surface in order to reduce the reflection of electromagnetic waves that would have otherwise occurred at that surface. These surfaces are usually the walls, floor, and ceiling of an anechoic chamber, as depicted in Figure 20.27, but it is often desirable to minimize the reflections from other objects as well.

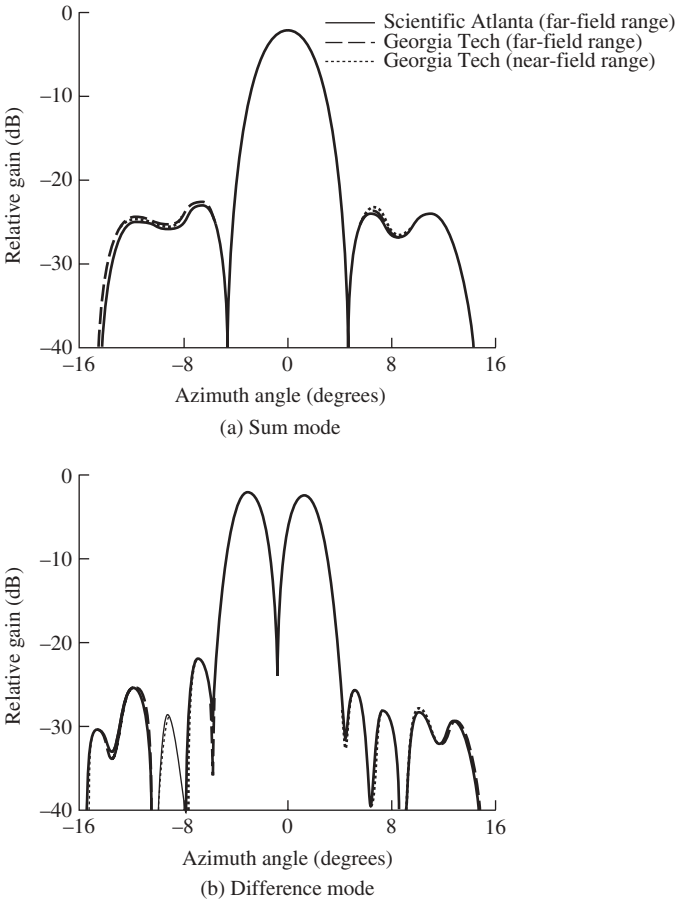


Figure 20.26 Measured and computed sum and difference mode principal-plane far-field patterns for a four-foot parabolic reflector. (From Ref. 9, © 1978 IEEE.)

On indoor antenna ranges, absorbers are ubiquitous and crucial. Some knowledge of absorbers (the various types, shapes and sizes, and their performance) is of immense value to the antenna measurement engineer/technician, even on ranges other than anechoic chambers. Microwave absorbing material can also be used to modify the radiation patterns of antennas, accentuate their front/back ratios, improve the isolation between antennas, and suppress resonant fields inside conducting enclosures, among other applications.

Microwave absorbing materials can basically be categorized into five classifications based on their construction and absorbing mechanisms:

1. *Carbon-loaded absorbers* are electrically lossy. They are generally ultra-broadband. The parameters that most strongly impact their performance include their thickness, shape, and carbon loading density.
2. *Sintered ferrite material*, usually in tile form, absorbs electromagnetic energy via magnetic loss. It exhibits favorable performance at lower frequencies (10–1000 MHz) and is compact in size compared to carbon-loaded absorber.

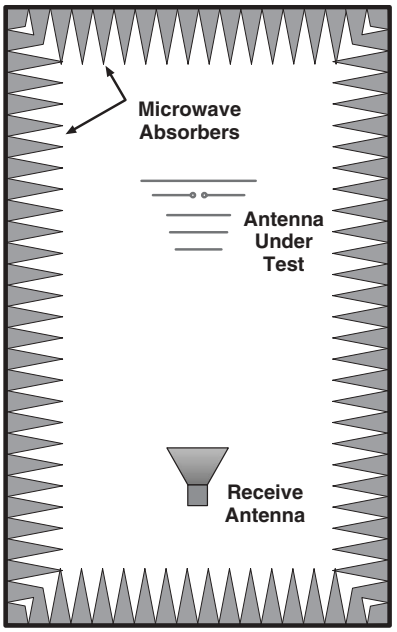


Figure 20.27 Shaped, carbon-loaded foam microwave absorbers line the surfaces of anechoic chambers to approximate the conditions of free space suitable for antenna and radar cross-section measurements.

3. By combining ferrite tiles with carbon-loaded absorbers, *hybrid absorbers* are capable of functioning over broader frequency ranges.
4. *Ferrite-loaded elastomeric absorbers* are thin sheets of rubber-like material that can be conformally applied to curved structures.
5. *Salisbury screens* achieve absorption via phase cancellation between the incident wavefront and its reflection off a conductive backplane that is located a quarter-wavelength behind its 377-Ω/sq impedance front surface. Absorbers of this class are thin and narrowband.

Carbon-loaded absorbers are by far the most prevalent and most effective type of absorber for frequencies above 1 GHz. The most common carrier material for the carbon is polyurethane foam. It has mechanical and electrical properties that are appropriate for this application. Expanded polystyrene is also used as a substrate for carbon-based absorbers. Both materials are relatively lightweight, can be cut or formed into various shapes, and have low dielectric constants. A key parameter affecting the frequency of operation of carbon-foam absorbers is thickness. An absorber that is three or more free-space wavelengths thick may achieve a reflectivity level of -55 dB at normal incidence. An absorber that is one wavelength thick will typically have a reflectivity of -35 dB or higher. Thus the lowest expected frequency of operation will play a large part in choosing the thickness of material used to cover the surfaces of an anechoic chamber, for example. This also explains the preference for ferrite tiles over carbon-foam absorber for frequencies much below 500 MHz. Although carbon-foam absorbers of up to 144 inches thick are commercially available, they are expensive and they occupy a lot of space.

Two other parameters that impact the performance of carbon-foam absorbers can be viewed as interrelated: the doping density of the carbon and the shape of the absorber. To minimize the reflectivity of an absorber, the impedance of the illuminated side must approach that of free space. One method to achieve a graduation of doping density is to laminate together relatively thin sheets of carbon-loaded foam such that the most lightly doped layer is on the illuminated side, and subsequent layers are increasingly doped with a higher percentage of carbon, terminating in the most heavily doped layer that will be placed against the otherwise reflective surface. Such laminates have limited performance due to the discontinuities at the layer interfaces. A much better way to effectively achieve the desired gradient between free space and fully doped absorber is to cut or form shapes on the illumination side of uniformly doped material. The shape that provides the best normal-incidence performance is the pyramid, as shown in Figure 20.28. The fact that the urethane foam substrate material is simultaneously tapered is not inconsequential to the excellent performance achieved with the pyramidal shape. Since the relative permittivity of the substrate material is not equal to unity, an abrupt discontinuity (a flat surface) between it and free space would have a negative impact on reflectivity.

While the normal-incidence performance of pyramidal absorbers is excellent, their reflectivity increases as a function of incidence angle [40]. Furthermore, as the incidence angle approaches grazing, the reflectivity is higher for incidence parallel to the rows of pyramids than at other angles. Additionally, there is an angle of incidence that is normal to the flat sides of the pyramids, which results in a relative peak in the backscatter. This can be detrimental to radar cross-section measurement ranges for which backscatter from the anechoic chamber itself must be minimized. One shape that provides moderate performance over a broad angular region of incidence, particularly at millimeter-wave frequencies, is referred to as "convolute." Convolute (egg crate-shaped) absorbers, shown in Figure 20.29, have continuously curved surfaces. By eliminating any flat surfaces, convolute absorbers reduce incidence angle-specific scattering.

The incidence angles of the most significant stray illumination on the interior surfaces of an anechoic chamber can be predicted according to geometrical optics or ray tracing. Most of the energy incident on the endwalls, for example, will be nearly normal to those walls; thus the use of pyramidal absorbers are most appropriate there. Near

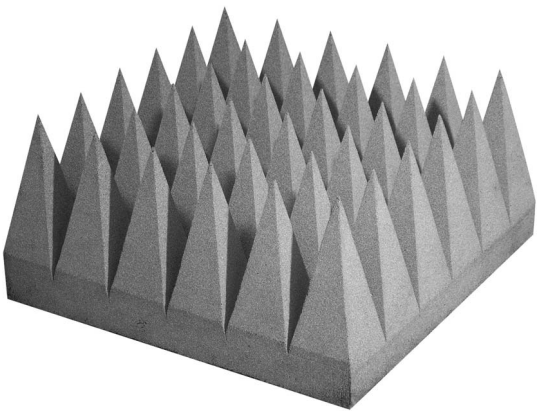


Figure 20.28 Pyramidal carbon-foam absorbers generally have the lowest normal-incidence reflectivity and operate over a very wide bandwidth.

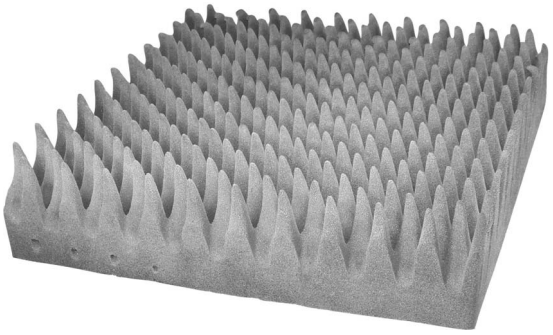


Figure 20.29 Convolute absorber provides moderate reflectivity over a wide range of incidence angles at up to millimeter-wave frequencies.

the middle of the chamber, most of the illumination on the sidewalls, floor, and ceiling will be at near-grazing angles to those surfaces. For this reason, particularly for radar cross-section measurement applications in which minimizing backscatter from the chamber is paramount, wedge absorbers are frequently used. A piece of wedge absorber is shown in Figure 20.30. When the lengths of the wedges are oriented parallel to the longitudinal axis of the chamber, the majority of the energy that is not absorbed by the wedge material scatters bistatically in directions approximately according to Snell's law, rather than backscatter to be received by the radar. That bistatically scattered energy is subsequently terminated in the absorber on the endwalls.

There are many other possible shapes and configurations of carbon-loaded absorbers [41] including pyramids of honeycomb material for high power application and for ventilation, closed-cell foams and vinyl covered materials for all-weather applications, walkway material, and exotic shapes to enhance performance.

Additional details about ferrite tiles are of greater use to the designers of anechoic and semianechoic chambers than to the antenna measurement engineer; and Salisbury screens,

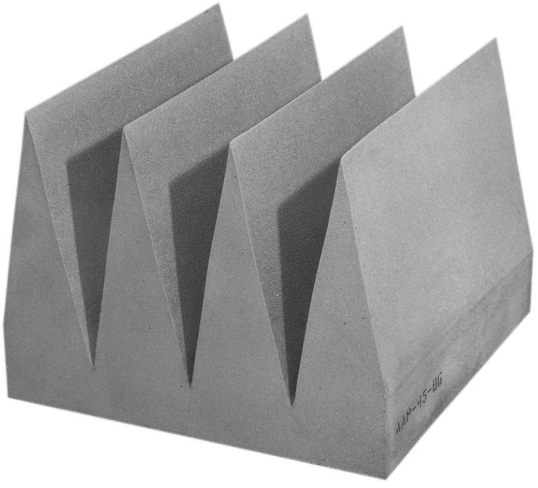


Figure 20.30 Wedge absorber exhibits reduced backscatter at near-grazing incidence angles.

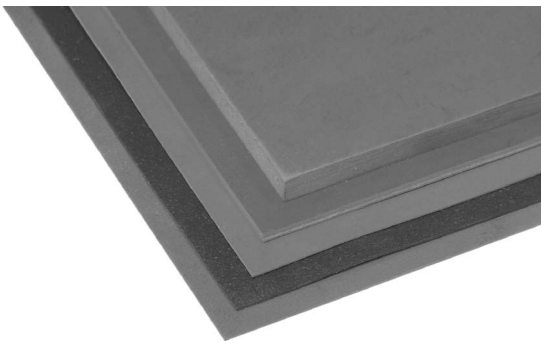


Figure 20.31 An assortment of ferrite-loaded elastomeric absorber material.

having been supplanted by superior materials for most applications, are infrequently encountered. Ferrite-loaded elastomeric absorbers (pictured in Figure 20.31), on the other hand, have a broad range of applications. These are highly engineered absorbers that are loaded with very specific quantities and combinations of a wide array of powdered ferrites, iron, and other compounds. A review of these materials can be found in Ref. 42. The substrate materials for the powdered ferrites can be a variety of polymers including silicone, nitrile, urethane, neoprene, and natural rubber.

This type of absorber can be further classified according to its designed purpose. In one design, sometimes called magRAM (a contraction of *magnetic* and *radar absorbing material*), the mix of ferrites and thickness of the substrate are designed to optimize the suppression of specular reflections from metallic surfaces [43, 44]. In addition to the matrix thickness, the performance is dependent on the ferrite/rubber ratio and the material properties. Since the permeability and permittivity of these materials are higher than unity, they are very thin (a few millimeters) at microwave frequencies. Such absorbers exhibit resonant behavior, with a relative null in reflectivity at a particular frequency [45]. Two or more layers having different resonant frequencies can be laminated together to increase their operational bandwidth. One commercially available dual-band absorber, for example, has a nominal reflectivity of better than -17 dB from approximately 7.2 to 18 GHz.

The other type of ferrite-loaded elastomeric absorber is sometimes referred to as surface wave absorbing material (SWAM). By suppressing surface currents before they encounter a discontinuity (such as the edge of a ground plane), SWAM reduces the fields that would otherwise be reradiated.

20.4 INSTRUMENTATION

The instrumentation required to accomplish a measuring task depends largely on the functional requirements of the design (Figure 20.32) An antenna-range instrumentation must be designed to operate over a wide range of frequencies, and it usually can be classified into five categories [7]:

1. Source antenna and transmitting system
2. Receiving system

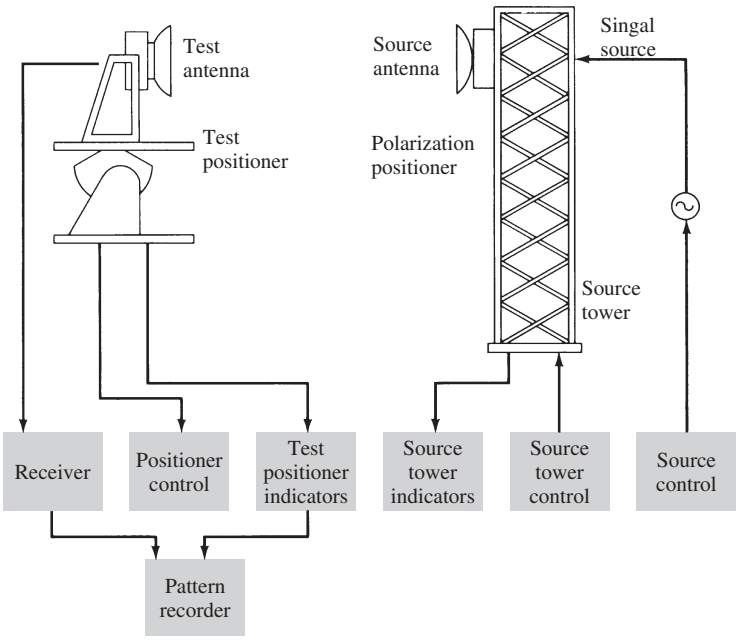


Figure 20.32 Instrumentation for typical antenna-range measuring system. (From Ref. 2, © 1979 IEEE.)

3. Positioning system
4. Recording system
5. Data processing system

A block diagram of a system that possesses these capabilities is shown in Figure 20.33.

The source antennas are usually log-periodic antennas for frequencies below 1 GHz, families of parabolas with broadband feeds for frequencies above 400 MHz, and even large horn antennas. The dual-ridged horn shown in Figure 20.34 operates from 500 MHz to 6 GHz. The system must be capable of controlling the polarization. Continuous rotation of the polarization can be accomplished by mounting a linearly polarized source antenna on a polarization positioner. Antennas with circular polarization can also be designed, such as crossed log-periodic arrays, which are often used in measurements. Some broadband antennas are not appropriate as feed antennas for CATRs, however, due to the displacement of their phase centers as a function of frequency.

The transmitting RF source must be selected [2] so that it has frequency control, frequency stability, spectral purity, controllable power level, and possibly modulation. The receiving system could be as simple as a bolometer detector, followed possibly by an amplifier, and a recorder. More elaborate and expensive receiving systems that provide greater sensitivity, precision, and dynamic range can be designed. One such system is a heterodyne receiving system [2], which uses double conversion and phase locking, which can be used for amplitude measurements. A dual-channel heterodyne system design is also available [2], and it can be used for phase measurements.

The vector network analyzer (VNA) has become an attractive alternative to the piecemeal component instrumentation described above. Hewlett-Packard's 8510 C is pictured

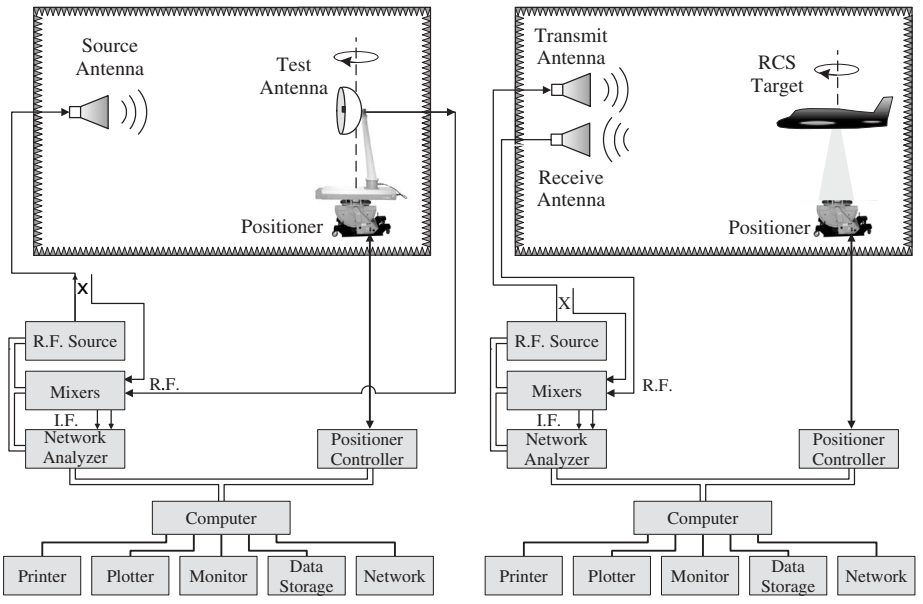


Figure 20.33 Block diagrams of typical network analyzer-based instrumentations, configured for (a) antenna pattern measurements and (b) radar cross-section measurements.

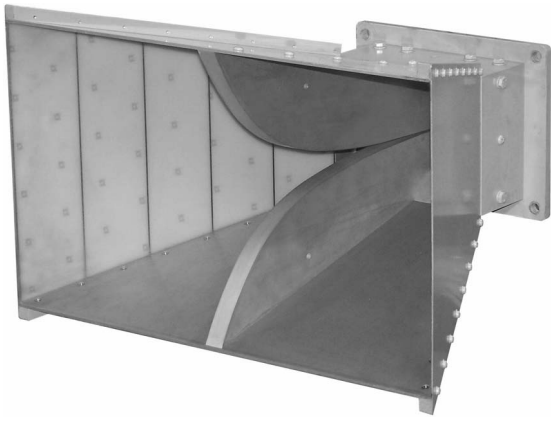


Figure 20.34 This dual-ridged horn, which operates from 500 MHz to 6 GHz, is suitable as a broadband source antenna.

in Figure 20.35. Since an antenna pattern measurement can be conceptually viewed as the S_{21} between the input connectors of the source and test antennas, the VNA is well suited for this application. VNAs exhibit excellent phase stability, spectral purity, sensitivity, and dynamic range (well in excess of 100 dB). Furthermore, some of today's VNAs can operate in both pulsed-RF and swept frequency modes.

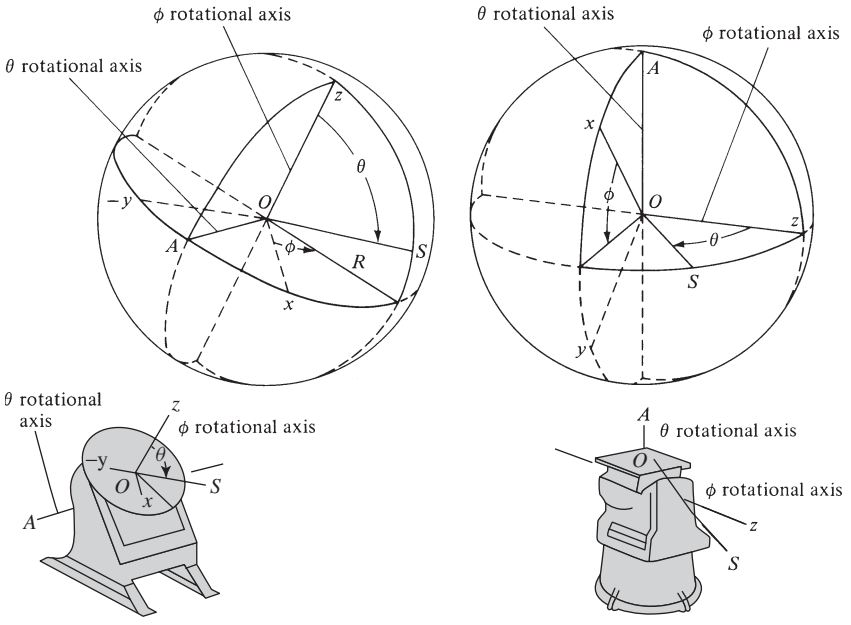
To achieve the desired plane cuts, the mounting structures of the system must have the capability to rotate in various planes. This can be accomplished by utilizing

Figure 20.35 The HP8510 vector network analyzer has become a fairly common antenna-range instrumentation.

rotational mounts (pedestals), two of which are shown in Figure 20.36. Tower-model elevation-over-azimuth pedestals are also available [7]. A *model tower*, shown in Figure 20.37, is a commonly used positioning system for antenna pattern measurements. The test antenna is mounted on the roll axis at the top of the tower, which places it in the quiet zone of the measurement range. A linear offset axis is located beneath the tower, and it is used to place the phase center of the antenna above the rotation center of the azimuthal table to which it is bolted. The elevation axis at the bottom is useful for tilting the top of the tower down to ground level for ease of test antenna mounting. The test antenna can be positioned to any spherical orientation with this combination of roll and azimuthal axes.

Modern antenna measurement facilities utilize a computer and software to coordinate the actions of the positioning and measuring systems, to store and process the acquired data, and to display or plot the resulting measurements. A communication bus connects the computer to the positioner controller and instruments. The IEEE-488 GPIB (general purpose instrument bus) is frequently used.

There are a number of commercially available computer software packages that are specifically designed for antenna measurement ranges. Such software packages provide an immediate measurement automation and data acquisition solution, usually with a lot of features and convenient graphical user interface (GUI). One such software package, *ARCS*, is available from March Microwave Systems B.V., the leading manufacturer of dual parabolic-cylinder compact range systems. ARCS includes fully automated acquisition software for radiation and RCS patterns, 2D and 3D visualization codes, and a suite of postmeasurement processing tools.



(a) Azimuth–over–elevation positioner (b) Elevation–over–azimuth positioner

Figure 20.36 Azimuth-over-elevation and elevation-over-azimuth rotational mounts. (From Ref. 2, © 1979 IEEE.)

Figure 20.37 The front and side views of a model tower positioning system. A roll axis is located at the top of the tower, which is mounted on a fore-aft linear offset axis. This in turn is mounted on the azimuthal table of an AZ/EL positioner.

It is not unheard of for an antenna measurement facility to trade time for money by writing its own custom data acquisition software. A third software option that lies between these two extremes is to use a generalized instrument control and data acquisition software like LabVIEW^{TM} [46] as a shortcut to writing an antenna measurement-specific package.

20.5 RADIATION PATTERNS

The radiation patterns (amplitude and phase), polarization, and gain of an antenna, which are used to characterize its radiation capabilities, are measured on the surface of a constant radius sphere. Any position on the sphere is identified using the standard spherical coordinate system of Figure 20.38. Since the radial distance is maintained fixed, only the two angular coordinates (θ, ϕ) are needed for positional identification. A representation of the radiation characteristics of the radiator, as a function of θ and ϕ for a constant radial distance and frequency, is defined as the *pattern* of the antenna.

In general, the pattern of an antenna is three dimensional. Because it is impractical to measure a three-dimensional pattern, a number of two-dimensional patterns are measured. They are used to construct a three-dimensional pattern. The number of two-dimensional patterns needed to construct faithfully a three-dimensional graph is determined by the functional requirements of the description and the available time and funds. The minimum number of two-dimensional patterns is two, and they are usually chosen to represent the orthogonal principal E- and H-plane patterns.

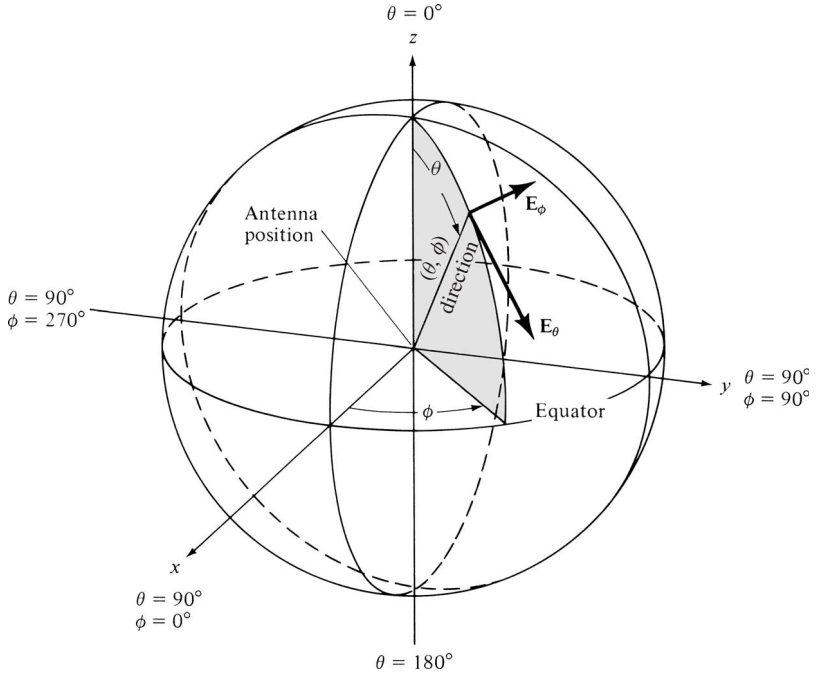


Figure 20.38 Spherical coordinate system geometry. (From Ref. 2, © 1979 IEEE.)

A two-dimensional pattern is also referred to as a *pattern cut*, and it is obtained by fixing one of the angles (θ or ϕ) while varying the other. For example, by referring to Figure 20.38, pattern cuts can be obtained by fixing ϕ_j $(0 \leq \phi_j \leq 2\pi)$ and varying θ $(0 \leq \theta \leq \pi)$. These are referred to as elevation patterns. Similarly, θ_i can be maintained fixed $(0 \leq \theta_i \leq \pi)$ while ϕ is varied $(0 \leq \phi \leq 2\pi)$. These are designated as azimuthal patterns.

The patterns of an antenna can be measured in the transmitting or receiving mode. The mode is dictated by the application. However, if the radiator is reciprocal, as is the case for most practical antennas, then either the transmitting or receiving mode can be utilized. For such cases, the receiving mode is selected. Unless otherwise specified, it is assumed here that the measurements are performed in the receiving mode.

20.5.1 Amplitude Pattern

The total amplitude pattern of an antenna is described by the vector sum of the two orthogonally polarized radiated field components. The pattern on a conventional antenna range can be measured using the system of Figure 20.32 with an appropriate detector. The receiver may be a simple bolometer (followed possibly by an amplifier), a double-conversion phase-locking heterodyne system [7, Figure 14], or any other design.

In many applications, the movement of the antenna to the antenna range can significantly alter the operational environment. Therefore, in some cases, antenna pattern measurements must be made *in situ* to preserve the environmental performance characteristics. A typical system arrangement that can be used to accomplish this is shown in Figure 20.39. The source is mounted on an airborne vehicle, which is maneuvered through space around the test antenna and in its far field, to produce a plane wave and to provide the desired pattern cuts. The tracking device provides to the recording equipment the angular position data of the source relative to a reference direction. The measurements can be conducted either by a point-by-point or by a continuous method. Usually the continuous technique is preferred.

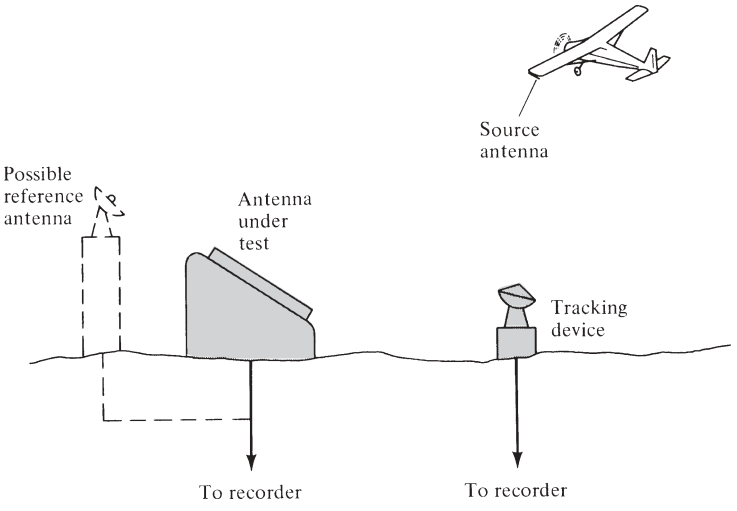


Figure 20.39 System arrangement for *in situ* antenna pattern measurements. (From Ref. 2, © 1979 IEEE.)

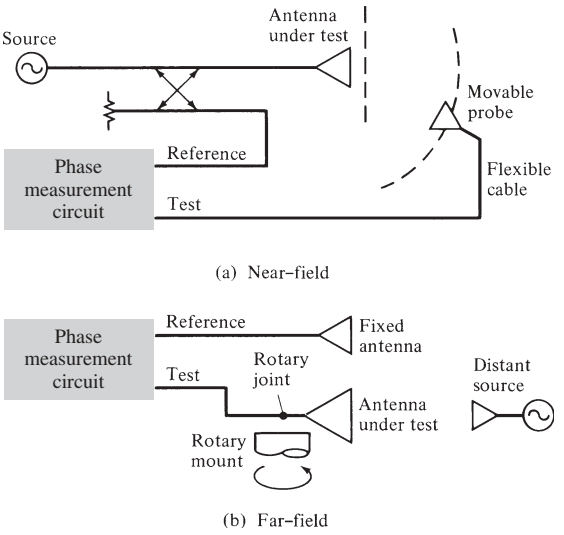


Figure 20.40 Near-field and far-field phase pattern measuring systems. (From Ref. 2, © 1979 IEEE.)

20.5.2 Phase Measurements

The phase pattern of the field, in the direction of the unit vector $\hat{\mathbf{u}}$, is given by the $\psi(\theta, \phi)$ phase function see Eq. (13-63), Chapter 13 of [1]. For linear polarization, $\hat{\mathbf{u}}$ is real, and it may represent $\hat{\mathbf{a}}_\theta$ or $\hat{\mathbf{a}}_\phi$ in the direction of θ or ϕ.

The phase of an antenna is periodic, and it is defined in multiples of 360°. In addition, the phase is a relative quantity, and a reference must be provided during measurements for comparison.

Two basic system techniques that can be used to measure phase patterns at short and long distances from the antenna are shown in Figure 20.40. For the design of Figure 20.40a, a reference signal is coupled from the transmission line, and it is used to compare, in an appropriate network, the phase of the received signal. For large distances, this method does not permit a direct comparison between the reference and the received signal. In these cases, the arrangement of Figure 20.40b can be used in which the signal from the source antenna is received simultaneously by a fixed antenna and the antenna under test. The phase pattern is recorded as the antenna under test is rotated while the fixed antenna serves as a reference. The phase measuring circuit may be the dual-channel heterodyne system [7, Figure 15].

20.6 GAIN MEASUREMENTS

The most important figure of merit that describes the performance of a radiator is the gain. There are various techniques and antenna ranges that are used to measure the gain. The choice of either depends largely on the frequency of operation.

Usually free-space ranges are used to measure the gain above 1 GHz. In addition, microwave techniques, which utilize waveguide components, can be used. At lower frequencies, it is more difficult to simulate free-space conditions because of the longer

wavelengths. Therefore between 0.1 and 1 GHz, ground-reflection ranges are utilized. Scale models can also be used in this frequency range. However, since the conductivity and loss factors of the structures cannot be scaled conveniently, the efficiency of the full-scale model must be found by other methods to determine the gain of the antenna. This is accomplished by multiplying the directivity by the efficiency to result in the gain. Below 0.1 GHz, directive antennas are physically large and the ground effects become increasingly pronounced. Usually the gain at these frequencies is measured *in situ*. Antenna gains are not usually measured at frequencies below 1 MHz. Instead, measurements are conducted on the field strength of the ground wave radiated by the antenna.

Usually there are two basic methods that can be used to measure the gain of an electromagnetic radiator: *absolute-gain* and *gain-transfer* (or *gain-comparison*) measurements. The absolute-gain method is used to calibrate antennas that can then be used as standards for gain measurements, and it requires no a priori knowledge of the gains of the antennas. Gain-transfer methods must be used in conjunction with standard gain antennas to determine the absolute gain of the antenna under test.

The two antennas that are most widely used and universally accepted as gain standards are the resonant $\lambda/2$ dipole (with a gain of about 2.1 dB) and the pyramidal horn antenna (with a gain ranging from 12 to 25 dB). Both antennas possess linear polarizations. The dipole, in free space, exhibits a high degree of polarization purity. However, because of its broad pattern, its polarization may be suspect in other than reflection-free environments. Pyramidal horns usually possess, in free space, slightly elliptical polarization (axial ratio of about 40 to infinite dB). However, because of their very directive patterns, they are less affected by the surrounding environment.

20.6.1 Absolute-Gain Measurements

There are a number of techniques that can be employed to make absolute-gain measurements. Very brief reviews of each are included here. More details can be found in Refs. 3 and 4. All of these methods are based on the Friis transmission formula, which assumes that the measuring system employs, each time, two antennas. The antennas are separated by a distance R, and it must satisfy the far-field criterion of each antenna (see Figure 2.31 of [1]).

20.6.1.1 Two-Antenna Method

In a logarithmic decibel form we can write

$$(G_{0t})_{\text{dB}} + (G_{0r})_{\text{dB}} = 20\log_{10}\left(\frac{4\pi R}{\lambda}\right) + 10\log_{10}\left(\frac{P_r}{P_t}\right) \tag{20.1}$$

where

$(G_{0t})_{\text{dB}}$ = gain of the transmitting antenna (dB)
$(G_{0r})_{\text{dB}}$ = gain of the receiving antenna (dB)
P_r = received power (W)
P_t = transmitted power (W)
R = antenna separation (m)
λ = operating wavelength (m)

If the transmitting and receiving antennas are identical ($G_{0t} = G_{0r}$), Eq. (20.1) reduces to

$$(G_{0t})_{\text{dB}} = (G_{0r})_{\text{dB}} = \frac{1}{2}\left[20\log_{10}\left(\frac{4\pi R}{\lambda}\right) + 10\log_{10}\left(\frac{P_r}{P_t}\right)\right] \tag{20.2}$$

By measuring R, λ, and the ratio of P_r/P_t, the gain of the antenna can be found. At a given frequency, this can be accomplished using the system of Figure 20.41a. The system is simple and the procedure straightforward. For continuous multifrequency measurements, such as for broadband antennas, the swept-frequency instrumentation of Figure 20.41b can be utilized.

20.6.1.2 Three-Antenna Method If the two antennas in the measuring system are not identical, three antennas (*a, b, c*) must be employed and three measurements must be made (using all combinations of the three) to determine the gain of each of the three. Three equations (one for each combination) can be written, and each takes the form of Eq. (20.1). Thus *(a–b combination)*

$$(G_a)_{\text{dB}} + (G_b)_{\text{dB}} = 20\log_{10}\left(\frac{4\pi R}{\lambda}\right) + 10\log_{10}\left(\frac{P_{rb}}{P_{ta}}\right) \tag{20.3a}$$

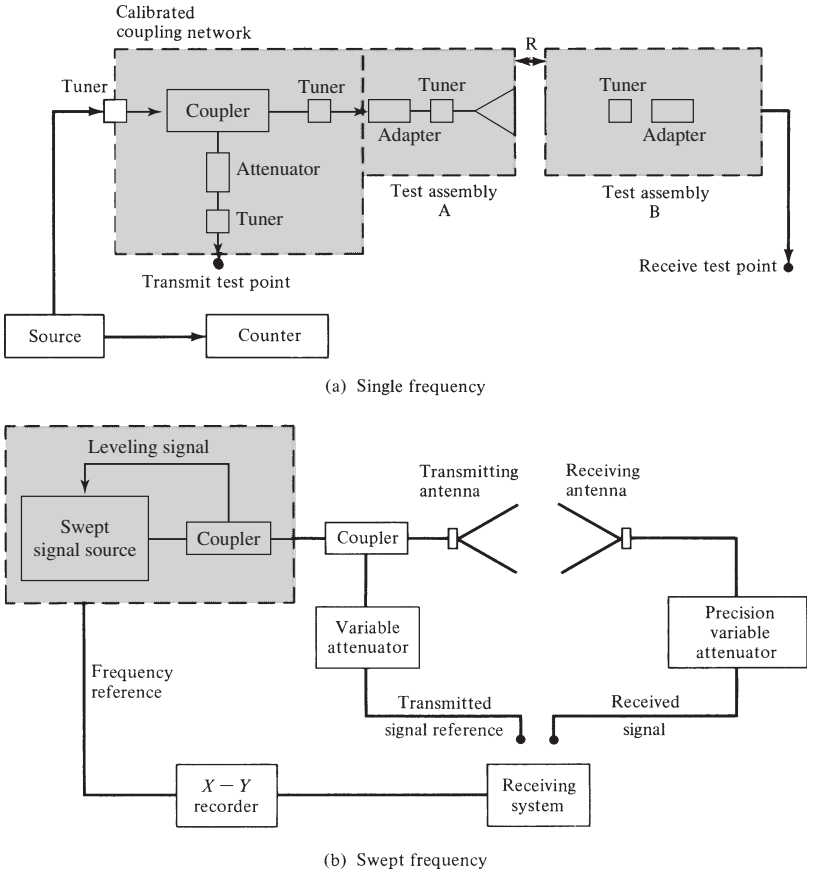


Figure 20.41 Typical two- and three-antenna measuring systems for single- and swept-frequency measurements. (From Ref. 4.)

(a–c combination)

$$(G_a)_{\text{dB}} + (G_c)_{\text{dB}} = 20\log_{10}\left(\frac{4\pi R}{\lambda}\right) + 10\log_{10}\left(\frac{P_{rc}}{P_{ta}}\right) \tag{20.3b}$$

(b–c combination)

$$(G_b)_{\text{dB}} + (G_c)_{\text{dB}} = 20\log_{10}\left(\frac{4\pi R}{\lambda}\right) + 10\log_{10}\left(\frac{P_{rc}}{P_{tb}}\right) \tag{20.3c}$$

From these three equations, the gains $(G_a)_{\text{dB}}$, $(G_b)_{\text{dB}}$, and $(G_c)_{\text{dB}}$ can be determined provided R, λ, and the ratios of P_{rb}/P_{ta}, P_{rc}/P_{ta}, and P_{rc}/P_{tb} are measured.

The two- and three-antenna methods are both subject to errors. Care must be utilized so (1) the system is frequency stable; (2) the antennas meet the far-field criteria; (3) the antennas are aligned for boresight radiation; (4) all the components are impedance and polarization matched; and (5) there is a minimum of proximity effects and multipath interference.

Impedance and polarization errors can be accounted for by measuring the appropriate complex reflection coefficients and polarizations and then correcting accordingly the measured power ratios. The details for these corrections can be found in Refs. 2 and 3. There are no rigorous methods to account for proximity effects and multipath interference. These, however, can be minimized by maintaining the antenna separation by at least a distance of $2D^2/\lambda$, as is required by the far-field criteria, and by utilizing RF absorbers to reduce unwanted reflections. The interference pattern that is created by the multiple reflections from the antennas themselves, especially at small separations, is more difficult to remove. It usually manifests itself as a cyclic variation in the measured antenna gain as a function of separation.

20.6.1.3 Extrapolation Method The extrapolation method is an absolute-gain method, which can be used with the three-antenna method, and it was developed [16] to rigorously account for possible errors due to proximity, multipath, and nonidentical antennas. If none of the antennas used in the measurements are circularly polarized, the method yields the gains and polarizations of all three antennas. If only one antenna is circularly polarized, this method yields only the gain and polarization of the circularly polarized antenna. The method fails if two or more antennas are circularly polarized.

The method requires both amplitude and phase measurements when the gain and the polarization of the antennas are to be determined. For the determination of gains, amplitude measurements are sufficient. The details of this method can be found in Refs. 3 and 47.

20.6.1.4 Ground-Reflection Range Method A method that can be used to measure the gain of moderately broadbeam antennas, usually for frequencies below 1 GHz, has been reported [48]. The method takes into account the specular reflections from the ground (using the system geometry of Figure 20.2), and it can be used with some restrictions and modifications with the two- or three-antenna methods. As described here, the method is applicable to linear antennas that couple only the electric field. Modifications must be made for loop radiators. Using this method, it is recommended that the linear vertical radiators be placed in a horizontal position when measurements are made. This

is desired because the reflection coefficient of the earth, as a function of incidence angle, varies very rapidly for vertically polarized waves. Smoother variations are exhibited for horizontally polarized fields. Circularly and elliptically polarized antennas are excluded, because the earth exhibits different reflective properties for vertical and horizontal fields.

To make measurements using this technique, the system geometry of Figure 20.2 is utilized. Usually it is desirable that the height of the receiving antenna h_r be much smaller than the range R_0 ($h_r \ll R_0$). Also the height of the transmitting antenna is adjusted so that the field of the receiving antenna occurs at the first maximum nearest to the ground. Doing this, each of the gain equations of the two- or three-antenna methods take the form of

$$(G_a)_{\text{dB}} + (G_b)_{\text{dB}} = 20\log_{10}\left(\frac{4\pi R_D}{\lambda}\right) + 10\log_{10}\left(\frac{P_r}{P_t}\right) - 20\log_{10}\left(\sqrt{D_A D_B} + \frac{rR_D}{R_R}\right) \tag{20.4}$$

where D_A and D_B are the directivities (relative to their respective maximum values) along R_D, and they can be determined from amplitude patterns measured prior to the gain measurements. R_D, R_R, λ, and P_r/P_t are also measured. The only quantity that needs to be determined is the factor r, which is a function of the radiation patterns of the antennas, the frequency of operation, and the electrical and geometrical properties of the antenna range.

The factor r can be found by first repeating the above measurements but with the transmitting antenna height adjusted so that the field at the receiving antenna is minimized. The quantities measured with this geometry are designated by the same letters as before but with a prime ($'$) to distinguish them from those of the previous measurement.

By measuring or determining the parameters

1. R_R, R_D, P_r, D_A, and D_B at a height of the transmitting antenna such that the receiving antenna is at the first maximum of the pattern
2. R'_R, R'_D, P'_r, D'_A, and D'_B at a height of the transmitting antenna such that the receiving antenna is at a field minimum

it can be shown [48] that r can be determined from

$$r = \left(\frac{R_R R'_R}{R_D R'_D}\right)\left[\frac{\sqrt{(P_r/P'_r)(D'_A D'_B)}R_D - \sqrt{D_A D_B}R'_D}{\sqrt{(P_r/P'_r)}R_R + R'_R}\right] \tag{20.5}$$

Now all parameters included in Eq. (20.4) can be either measured or computed from measurements. The free-space range system of Figure 20.41a can be used to perform these measurements.

20.6.2 Gain-Transfer (Gain-Comparison) Measurements

The method most commonly used to measure the gain of an antenna is the gain-transfer method. This technique utilizes a gain standard (with a known gain) to determine absolute gains. Initially relative gain measurements are performed, which, when compared with the known gain of the standard antenna, yield absolute values. The method can be used with free-space and reflection ranges and for *in situ* measurements.

The procedure requires two sets of measurements. In one set, using the test antenna as the receiving antenna, the received power (P_T) into a matched load is recorded. In the other set, the test antenna is replaced by the standard gain antenna and the received power (P_S) into a matched load is recorded. In both sets, the geometrical arrangement is maintained intact (other than replacing the receiving antennas), and the input power is maintained the same.

Writing two equations of the form of (20.1) or (20.4), for free-space or reflection ranges, it can be shown that they reduce to [7]

$$(G_T)_{\text{dB}} = (G_S)_{\text{dB}} + 10\log_{10}\left(\frac{P_T}{P_S}\right) \tag{20.6}$$

where $(G_T)_{\text{dB}}$ and $(G_S)_{\text{dB}}$ are the gains (in dB) of the test and standard gain antennas.

System disturbance during replacement of the receiving antennas can be minimized by mounting the two receiving antennas back-to-back on either side of the axis of an azimuth positioner and connecting both of them to the load through a common switch. One antenna can replace the other by a simple, but very precise, 180° rotation of the positioner. Connection to the load can be interchanged by proper movement of the switch.

If the test antenna is not too dissimilar from the standard gain antenna, this method is less affected by proximity effects and multipath interference. Impedance and polarization mismatches can be corrected by making proper complex reflection coefficient and polarization measurements [8].

If the test antenna is circularly or elliptically polarized, gain measurements using the gain-transfer method can be accomplished by at least two different methods. One way would be to design a standard gain antenna that possesses circular or elliptical polarization. This approach would be attractive in mass productions of power-gain measurements of circularly or elliptically polarized antennas.

The other approach would be to measure the gain with two orthogonal linearly polarized standard gain antennas. Since circularly and elliptically polarized waves can be decomposed to linear (vertical and horizontal) components, the total power of the wave can be separated into two orthogonal linearly polarized components. Thus the total gain of the circularly or elliptically polarized test antenna can be written

$$(G_T)_{\text{dB}} = 10\log_{10}(G_{TV} + G_{TH}) \tag{20.7}$$

G_{TV} and G_{TH} are, respectively, the partial power gains with respect to vertical-linear and horizontal-linear polarizations.

G_{TV} is obtained, using Eq. (20.6), by performing a gain-transfer measurement with the standard gain antenna possessing vertical polarization. The measurements are repeated with the standard gain antenna oriented for horizontal polarization. This allows the determination of G_{TH}. Usually a single linearly polarized standard gain antenna (a linear $\lambda/2$ resonant dipole or a pyramidal horn) can be used, by rotating it by 90°, to provide both vertical and horizontal polarizations. This approach is very convenient, especially if the antenna possesses good polarization purity in the two orthogonal planes.

The techniques outlined above yield good results provided the transmitting and standard gain antennas exhibit good linear polarization purity. Errors will be introduced if either one of them possesses a polarization with a finite axial ratio. In addition, these techniques are accurate if the tests can be performed in a free-space, a ground-reflection, or an extrapolation range. These requirements impose a low frequency limit of 50 MHz.

Below 50 MHz, the ground has a large effect on the radiation characteristics of the antenna, and it must be taken into account. It usually requires that the measurements are performed on full-scale models and *in situ*. Techniques that can be used to measure the gain of large HF antennas have been devised [49–51].

20.7 DIRECTIVITY MEASUREMENTS

If the directivity of the antenna cannot be found using solely analytical techniques, it can be computed using measurements of its radiation pattern. One of the methods is based on the approximate expressions of Kraus or Tai and Pereira. [see (2.26) and (2.30a) of Ref. 1]. The computations can be performed very efficiently and economically with modern computational facilities and numerical techniques.

The simplest, but least accurate, method requires that the following procedure is adopted:

1. Measure the two principal E- and H-plane patterns of the test antenna.
2. Determine the half-power beamwidths (in degrees) of the E- and H-plane patterns.
3. Compute the directivity using (2.26) or (2.30a) of [1].

The method is usually employed to obtain rough estimates of directivity. It is more accurate when the pattern exhibits only one major lobe, and its minor lobes are negligible.

The other method requires that the directivity be computed using (2.35) of [1], where P_{rad} is evaluated numerically. The $F(\theta_i, \phi_j)$ function of (2.43) of [1] represents the radiation intensity or radiation pattern, and it will be obtained by measurements. $U_{\max}$ represents the maximum radiation intensity of $F(\theta, \phi)$ in all space, as obtained by the measurements.

The radiation pattern is measured by sampling the field over a sphere of radius r. The pattern is measured in two-dimensional plane cuts with ϕ_j constant ($0 \le \phi_j \le 2\pi$) and θ variable ($0 \le \theta \le \pi$), as shown in Figure 20.42, or with θ_i fixed ($0 \le \theta_i \le \pi$) and ϕ variable ($0 \le \phi \le 2\pi$). The first are referred to as elevation or great-circle cuts, whereas the second represent azimuthal or conical cuts. Either measuring method can be used.

The spacing between measuring points is largely controlled by the directive properties of the antenna and the desired accuracy. The method is most accurate for broadbeam antennas. However, with the computer facilities and the numerical methods now available, this method is very attractive even for highly directional antennas. To maintain a given accuracy, the number of sampling points must increase as the pattern becomes more directional.

The above discussion assumes that all the radiated power is contained in a single polarization, and the measuring probe possesses that polarization. If the antenna is polarized such that the field is represented by both θ and ϕ components, the *partial directivities* $D_\theta(\theta, \phi)$ and $D_\phi(\theta, \phi)$ must each be found. This is accomplished from pattern measurements with the probe positioned, respectively, to sample the θ and ϕ components. The *total directivity* is then given by

$$D_0 = D_\theta + D_\phi \tag{20.8}$$

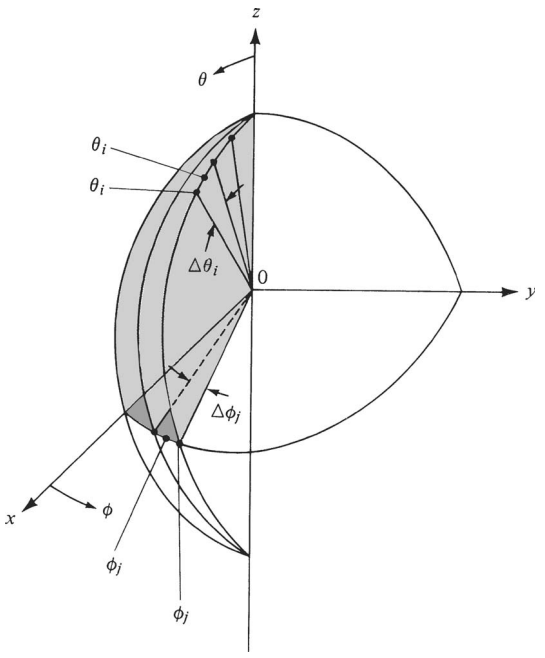


Figure 20.42 Digitization scheme of pattern in spherical coordinates.

where

$$D_\theta = \frac{4\pi U_\theta}{(P_{\text{rad}})_\theta + (P_{\text{rad}})_\phi} \tag{20.8a}$$

$$D_\phi = \frac{4\pi U_\phi}{(P_{\text{rad}})_\theta + (P_{\text{rad}})_\phi} \tag{20.8b}$$

$U_\theta, (P_{\text{rad}})_\theta$ and $U_\phi, (P_{\text{rad}})_\phi$ represent the radiation intensity and radiated power as contained in the two orthogonal θ and ϕ field components, respectively.

The same technique can be used to measure the field intensity and to compute the directivity of any antenna that possesses two orthogonal polarizations. Many antennas have only one polarization (θ or ϕ). This is usually accomplished by design and/or proper selection of the coordinate system. In this case, the desired polarization is defined as the *primary polarization*. Ideally, the other polarization should be zero. However, in practice, it is nonvanishing, but it is very small. Usually it is referred to as the *cross polarization*, and for good designs it is usually below -40 dB.

The directivity of circularly or elliptically polarized antennas can also be measured. Precautions must be taken [2] as to which component represents the primary polarization and which the cross-polarization contribution.

20.8 RADIATION EFFICIENCY MEASUREMENTS

The radiation efficiency is defined as the ratio of the total power radiated by the antenna to the total power accepted by the antenna at its input terminals during radiation. System

factors, such as impedance and/or polarization mismatches, do not contribute to the radiation efficiency because it is an inherent property of the antenna.

A common method of measuring the efficiency of an antenna involves the physical realization of the above definition. The radiated power can be determined by measuring the power density radiated by the antenna for all directions over a spherical surface and then integrating the measured samples. The input power can be obtained by either utilizing a power meter in place of the test antenna, or through a knowledge of the output power of the signal source and the S_{21} values of the devices and transmission lines between the source and the antenna, and taking into account any impedance mismatches. However, due to the relationship between gain and input power, the efficiency reduces to [52]

$$e_{cd} = \frac{1}{2\pi} \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} G_T(\theta, \phi) \sin(\theta) d\theta d\phi \tag{20.9}$$

This *pattern integration* method is time consuming, but it is the most rigorous of the efficiency measurement techniques. To avoid an accumulation of errors, however, extra care should be taken at each step of the measurement.

The radiation efficiency can also be defined, using the direction of maximum radiation as reference, as

$$e_{cd} = \text{radiation efficiency} = \frac{\text{gain}}{\text{directivity}} \tag{20.10}$$

where the definitions of directivity and gain imply that they are measured or computed in the direction of maximum radiation. Using techniques that were outlined in Sections 20.4 and 20.5 for the measurements of the gain and directivity, the radiation efficiency can then be computed using Eq. (20.10).

If the antenna is very small and simple, it can be represented as a series network. For antennas that can be represented by such a series network, the radiation efficiency can also be defined by $e_{cd} = R_r/(R_r + R_L)$, where R_r is the radiation resistance and R_L is the loss resistance of the antenna. This definition enables the measurement of efficiency using the *Wheeler cap* method [53, 54]. For these antennas, the real part of the input impedance is equal to the total antenna resistance, which consists of the radiation resistance and the loss resistance. Therefore the input resistance can be measured directly using a vector network analyzer. According to Wheeler, the radiation resistance can be canceled out by enclosing the antenna in a conducting shell. When so enclosed, the input resistance of the antenna then consists of only the loss resistance. The efficiency can thus be found by making two input resistance measurements: with the antenna radiating in free space, and when it is enclosed in a conducting shell.

Because the loss resistance of antennas coated with lossy dielectrics or antennas over lossy ground cannot be represented in series with the radiation resistance, this method cannot be used to determine their radiation efficiency.

20.9 IMPEDANCE MEASUREMENTS

Associated with an antenna, there are two types of impedances: a *self* and a *mutual* impedance. When the antenna is radiating into an unbounded medium and there is no coupling between it and other antennas or surrounding obstacles, the self-impedance is also the driving-point impedance of the antenna. If there is coupling between the antenna

under test and other sources or obstacles, the driving-point impedance is a function of its self-impedance and the mutual impedances between it and the other sources or obstacles. In practice, the driving-point impedance is usually referred to as the input impedance.

To attain maximum power transfer between a source or a source–transmission line and an antenna (or between an antenna and a receiver or receiver transmission line), a conjugate match is usually desired. In some applications, this may not be the most ideal match. For example, in some receiving systems, minimum noise is attained if the antenna impedance is lower than the load impedance. However, in some transmitting systems, maximum power transfer is attained if the antenna impedance is greater than the load impedance. If conjugate matching does not exist, the power lost can be computed [2] using

$$\frac{P_{\text{lost}}}{P_{\text{available}}} = \left| \frac{Z_{\text{ant}} - Z_{cct}^*}{Z_{\text{ant}} + Z_{cct}} \right|^2 \tag{20.11}$$

where

Z_{ant} = input impedance of the antenna
Z_{cct} = input impedance of the circuits that are connected to the antenna at its input terminals

When a transmission line is associated with the system, as is usually the case, the matching can be performed at either end of the line. In practice, however, the matching is performed near the antenna terminals, because it usually minimizes line losses and voltage peaks in the line and maximizes the useful bandwidth of the system.

In a mismatched system, the degree of mismatch determines the amount of incident or available power that is reflected at the input antenna terminals into the line. The degree of mismatch is a function of the antenna input impedance and the characteristic impedance of the line. These are related to the input reflection coefficient and the input VSWR at the antenna input terminals by the standard transmission-line relationships of

$$\frac{P_{\text{refl}}}{P_{\text{inc}}} = |\Gamma|^2 = \frac{|Z_{\text{ant}} - Z_c|^2}{|Z_{\text{ant}} + Z_c|^2} = \left| \frac{\text{VSWR} - 1}{\text{VSWR} + 1} \right|^2 \tag{20.12}$$

where

$\Gamma = |\Gamma| e^{j\gamma}$ = voltage reflection coefficient at the antenna input terminals
VSWR = voltage standing wave ratio at the antenna input terminals
Z_c = characteristic impedance of the transmission line

Equation (20.12) shows a direct relationship between the antenna input impedance (Z_{ant}) and the VSWR. In fact, if Z_{ant} is known, the VSWR can be computed using Eq. (20.15).

Once the reflection coefficient is completely described by its magnitude and phase, it can be used to determine the antenna impedance by

$$Z_{\text{ant}} = Z_c \left[\frac{1 + \Gamma}{1 - \Gamma} \right] = Z_c \left[\frac{1 + |\Gamma| e^{j\gamma}}{1 - |\Gamma| e^{j\gamma}} \right] \tag{20.13}$$

The reflection coefficient is also equal to S_{11}, and the mutual coupling between two antennas in proximity to one another is S_{21}. The instrument best suited for measuring S-parameters is the network analyzer. The convenience, versatility, and precision of a network analyzer makes it extremely attractive as the instrumentation for an antenna measurement range.

It is beyond the scope of this chapter to completely describe network analyzer operation. However, a simplified overview can be summarized as follows:

1. The start and stop frequencies, number of frequency points, source power, and other measurement parameters are specified.
2. The VNA is calibrated by measuring its response to a number of *calibration standards* (typically an open circuit, short circuit, and an impedance-matched termination). The calibration corrects for any residual frequency response, power loss, impedance mismatch, and so on in the instrument and cables leading up the calibration plane to which the test antenna(s) will be connected.
3. The antenna(s) are then connected to the test cables, and the S-parameters are measured.

The input impedance is generally a function of frequency, geometry, method of excitation, and proximity to its surrounding objects. Because of its strong dependence on the environment, it should usually be measured *in situ* or in an anechoic chamber. A simple test for environmental influence on the input impedance is to repeat the measurement after changing the orientation of the antenna. If the measured impedance changes, then scattering from the surrounding objects is disturbing the measurement.

Mutual impedances, which take into account interaction effects, are usually best described and measured by the cross-coupling coefficients S_{mn} of the device's (antenna's) scattering matrix. The coefficients of the scattering matrix can then be related to the coefficients of the impedance matrix [55].

20.10 CURRENT MEASUREMENTS

The current distribution along an antenna is another very important antenna parameter. A complete description of its amplitude and phase permits calculation of the radiation pattern.

There are a number of techniques that can be used to measure the current distribution [56–59]. One of the simplest methods requires that a small sampling probe, usually a small loop, be placed near the radiator. On the sampling probe, a current is induced which is proportional to the current of the test antenna.

The indicating meter can be connected to the loop in many different ways [56]. If the wavelength is very long, the meter can be consolidated into one unit with the measuring loop. At smaller wavelengths, the meter can be connected to a crystal rectifier. In order not to disturb the field distribution near the radiator, the rectifier is attached to the meter using long leads. To reduce the interaction between the measuring instrumentation and the test antenna and to minimize induced currents on the leads, the wires are wound on a dielectric support rod to form a helical choke. Usually the diameter of each turn, and

spacing between them, is about $\lambda/50$. The dielectric rod can also be used as a support for the loop. To prevent a dc short circuit on the crystal rectifier, a bypass capacitor is placed along the circumference of the loop.

There are many other methods, some of them more elaborate and accurate, and the interested reader can refer to the literature [56–59].

20.11 POLARIZATION MEASUREMENTS

The polarization of a wave is defined as *the curve traced by the instantaneous electric field, at a given frequency, in a plane perpendicular to the direction of wave travel*. The far-field polarization of an antenna is usually measured at distances where the field radiated by the antenna forms, in a small region, a plane wave that propagates in the outward radial direction.

In a similar manner, the polarization of the antenna is defined as *the curve traced by the instantaneous electric field radiated by the antenna in a plane perpendicular to the radial direction*, as shown in Figure 20.43a. The locus is usually an ellipse. In a spherical coordinate system, which is usually adopted in antennas, the polarization ellipse is formed by the orthogonal electric field components of E_θ and E_ϕ. The sense of rotation, also referred to as the sense of polarization, is defined by the sense of rotation of the wave as it is observed along the direction of propagation (see Figure 20.43b).

The general polarization of an antenna is characterized by the axial ratio (AR), the sense of rotation (CW or CCW, RH or LH), and the tilt angle τ. The tilt angle is used to identify the spatial orientation of the ellipse, and it is usually measured clockwise from the reference direction. This is demonstrated in Figure 20.43a, where τ is measured clockwise with respect to $\hat{\mathbf{a}}_\theta$, for a wave traveling in the outward radial direction.

Care must be exercised in the characterization of the polarization of a receiving antenna. If the tilt angle of an incident wave that is polarization matched to the receiving antenna is τ_m, it is related to the tilt angle τ_t of a wave transmitted by the same antenna by

$$\tau_t = 180^\circ - \tau_m \tag{20.14}$$

if a single coordinate system and one direction of view are used to characterize the polarization. If the receiving antenna has a polarization that is different from that of the incident wave, the polarization loss factor (PLF) can be used to account for the polarization mismatch losses.

The polarization of a wave and/or an antenna can best be displayed and visualized on the surface of a Poincaré sphere [60]. Each polarization occupies a unique point on the sphere, as shown in Figure 20.44. If one of the two points on the Poincaré sphere is used to define the polarization of the incident wave and the other the polarization of the receiving antenna, the angular separation can be used to determine the polarization losses. The procedure requires that the complex polarization ratios of each are determined, and they are used to compute the polarization efficiency in a number of different ways. The details of this procedure are well documented, and they can be found in Refs. 2 and 3.

Practically, it is very difficult to design radiators that maintain the same polarization state in all parts of their pattern. A complete description requires a number of measurements in all parts of the pattern. The number of measurements is determined by the required degree of polarization description.

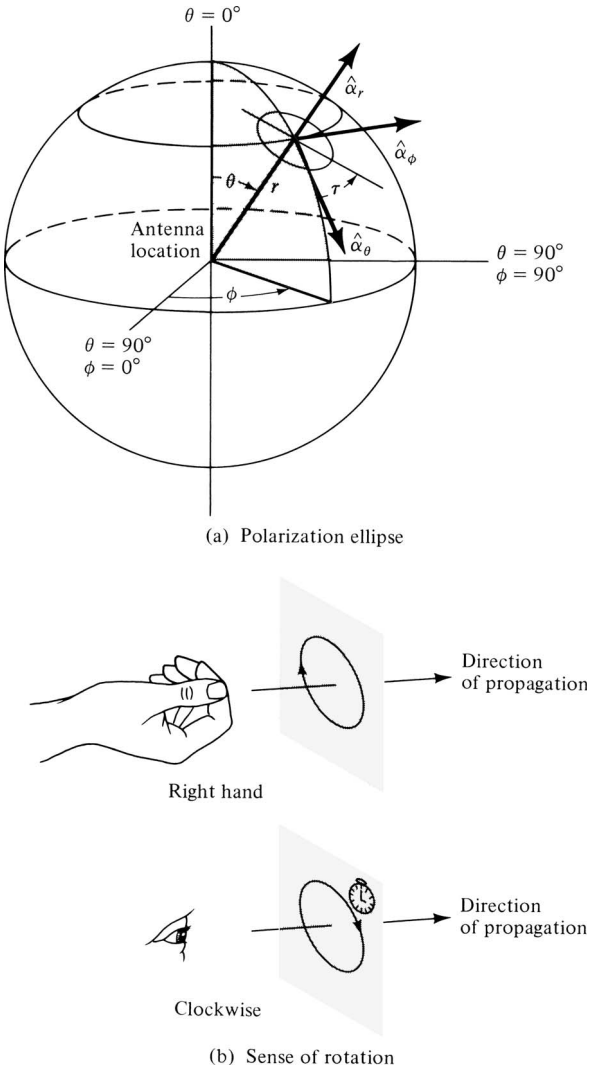


Figure 20.43 Polarization ellipse and sense of rotation for antenna coordinate system. (From Ref. 2, © 1979 IEEE.)

There are a number of techniques that can be used to measure the polarization state of a radiator [2, 3], and they can be classified into three main categories:

1. Those that yield partial polarization information. They do not yield a unique point on the Poincaré sphere.
2. Those that yield complete polarization information but require a polarization standard for comparison. They are referred to as *comparison methods*.
3. Those that yield complete polarization information and require no a priori polarization knowledge or no polarization standard. They are designated as *absolute methods*.

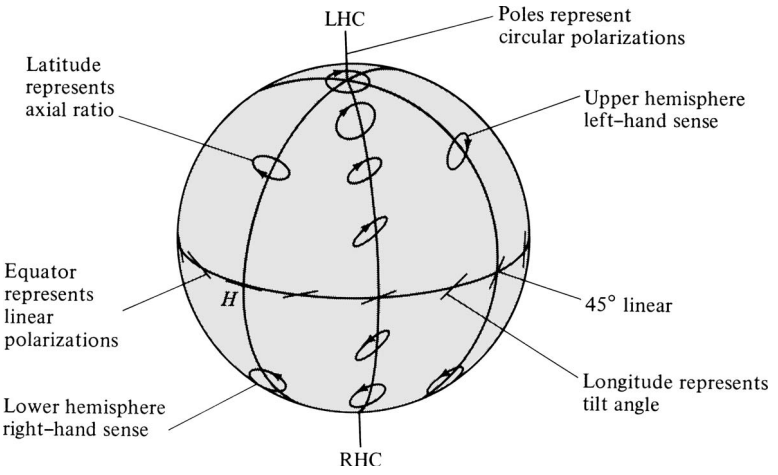


Figure 20.44 Polarization representation on the Poincaré sphere. (From Ref. 3, © 1978 IEEE.)

The method selected depends on such factors as the type of antenna, the required accuracy, and the time and funds available. A complete description requires not only the polarization ellipse (axial ratio and tilt angle), but also its sense of rotation (CW or CCW, RH or LH).

In this chapter, a method is discussed that can be used to determine the polarization ellipse (axial ratio and tilt angle) of an antenna but not its sense of rotation. This technique is referred to as the *polarization-pattern method*. The sense of polarization or rotation can be found by performing auxiliary measurements or by using other methods [7].

To perform the measurements, the antenna under test can be used either in the transmitting or in the receiving mode. Usually the transmitting mode is adopted. The method requires that a linearly polarized antenna, usually a dipole, be used to probe the polarization in the plane that contains the direction of the desired polarization. The arrangement is shown in Figure 20.45a. The dipole is rotated in the plane of the polarization, which is taken to be normal to the direction of the incident field, and the output voltage of the probe is recorded.

If the test antenna is linearly polarized, the output voltage response will be proportional to $\sin\psi$ (which is the far-zone field pattern of an infinitesimal dipole). The pattern forms a figure-eight, as shown in Figure 20.45b, where ψ is the rotation angle of the probe relative to a reference direction. For an elliptically polarized test antenna, the nulls of the figure-eight are filled and a dumbbell polarization curve (usually referred to as a *polarization pattern*) is generated, as shown in Figure 20.45b. The dashed curve represents the polarization ellipse.

The polarization ellipse is tangent to the polarization pattern, and it can be used to determine the axial ratio and the tilt angle of the test antenna. The polarization pattern will be a circle, as shown in Figure 20.45b, if the test antenna is circularly polarized. Ideally, this process must be repeated at every point of the antenna pattern. Usually it is performed at a number of points that describe sufficiently well the polarization of the antenna at the major and the minor lobes.

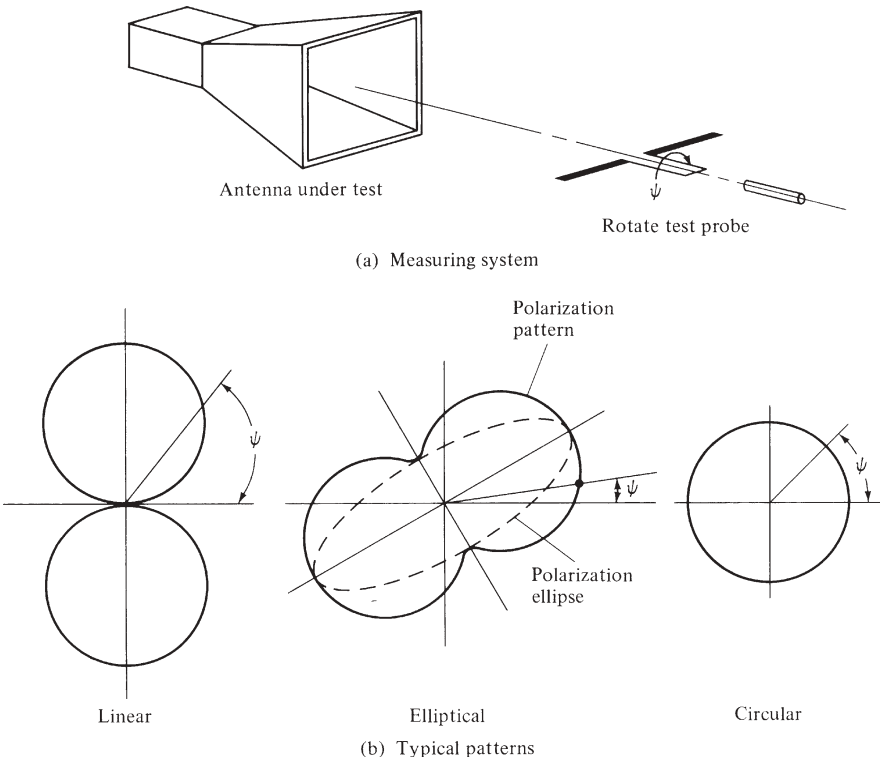


Figure 20.45 Polarization measuring system and typical patterns.

In some cases the polarization needs to be known over an entire plane. The axial ratio part of the polarization state can be measured using the arrangement of Figure 20.45a, where the test probe antenna usually is used as a source while the polarization pattern of the test antenna is being recorded while the test antenna is rotated over the desired plane. This arrangement does not yield the tilt angle or sense of rotation of the polarization state. In order to obtain the desired polarization pattern, the rate of rotation of the linear probe antenna (usually a dipole) is much greater than the rotation rate of the positioner over which the test antenna is mounted and rotated to allow, ideally, the probe antenna to measure the polarization response of the test antenna at that direction before moving to another angle. When this is performed over an entire plane, a typical pattern recorded in decibels is shown in Figure 20.46 [61], and it is referred to as the *axial ratio pattern*. It is apparent that the axial ratio pattern can be inscribed by inner and outer envelopes. At any given angle, the ratio of the outer and inner envelope responses represents the axial ratio. If the pattern is recorded in decibels, the axial ratio is the difference between the outer and inner envelopes (in dB); 0-dB difference represents circular polarization (axial ratio of unity). Therefore the polarization pattern of Figure 20.46 indicates that the test antenna it represents is nearly circularly polarized (within 1 dB; axial ratio less than 1.122) at and near $\theta = 0^\circ$ and deviates from that almost monotonically at greater angles (typically by about 7 dB maximum; maximum axial ratio of about 2.24).

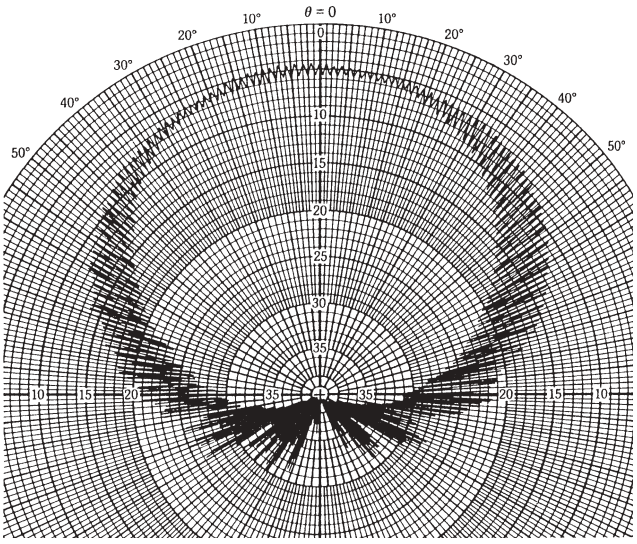


Figure 20.46 Pattern of a circularly polarized test antenna taken with a rotating, linearly polarized, source antenna. (From Ref. 61.)

The sense of rotation can be determined by performing auxiliary measurements. One method requires that the responses of two circularly polarized antennas, one responsive to CW and the other to CCW rotation, be compared [56]. The rotation sense of the test antenna corresponds to the sense of polarization of the antenna that produced the more intense response.

Another method would be to use a dual-polarized probe antenna, such as a dual-polarized horn, and to record simultaneously the amplitude polarization pattern and the relative phase between the two orthogonal polarizations. This is referred to as the *phase–amplitude* method, and it can be accomplished using the instrumentation of Figure 20.47. Double-conversion phase-locked receivers can be used to perform the amplitude and phase comparison measurements.

Another absolute polarization method, which can be used to completely describe the polarization of a test antenna, is referred to as the *three-antenna* method [2, 3]. As its name implies, it requires three antennas, two of which must not be circularly polarized.

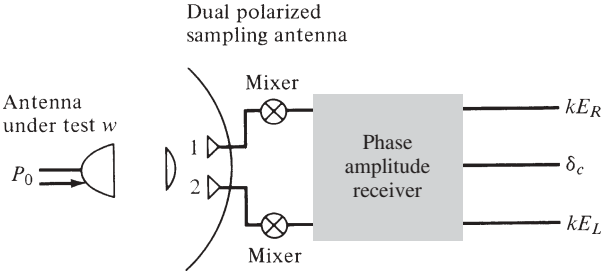


Figure 20.47 System configuration for measurements of polarization amplitude and phase. (From Ref. 3, © 1978 IEEE.)

There are a number of transfer methods [2, 3], but they require calibration standards for complete description of the polarization state.

20.12 SCALE MODEL MEASUREMENTS

In many applications (such as with antennas on ships, aircraft, large spacecraft, etc.), the antenna and its supporting structure are so immense in weight and/or size that they cannot be moved or accommodated by the facilities of an antenna measuring range. In addition, a movement of the structure to an antenna range can eliminate or introduce environmental effects. To satisfy these system requirements, *in situ* measurements are usually executed.

A technique that can be used to perform antenna measurements associated with large structures is *geometrical scale modeling*. Geometrical modeling is employed to (1) physically accommodate, within small ranges or enclosures, measurements that can be related to large structures; (2) provide experimental control over the measurements; and (3) minimize costs associated with large structures and corresponding experimental parametric studies.

Geometrical scale modeling by a factor of n (n smaller or greater than unity) requires the scaling indicated in Table 20.1. The primed parameters represent the scaled model while the unprimed represent the full-scale model. For a geometrical scale model, all the linear dimensions of the antenna and its associated structure are divided by n, whereas the operating frequency and the conductivity of the antenna material and its structure are multiplied by n. In practice, the scaling factor n is usually chosen greater than unity.

Ideal scale modeling for antenna measurements requires exact replicas, both physically and electrically, of the full-scale structures. In practice, however, this is closely approximated. The most difficult scaling is that of the conductivity. If the full-scale model possesses excellent conductors, even better conductors will be required in the scaled models. At microwave and millimeter-wave frequencies this can be accomplished by utilizing clean polished surfaces, free of films and other residues.

Geometrical scaling is often used for pattern measurements. However, it can also be employed to measure gain, directivity, radiation efficiency, input and mutual impedances, and so forth. For gain measurements, the inability to properly scale the conductivity can be overcome by measuring the directivity and the antenna efficiency and multiplying the two to determine the gain. Scalings that permit additional parameter changes are available [62]. The changes must satisfy the *theorem of similitude*.

TABLE 20.1 Geometrical Scale Model

Scaled Parameters	Unchanged Parameters
Length: $l' = l/n$	Permittivity: $\epsilon' = \epsilon$
Time: $t' = t/n$	Permeability: $\mu' = \mu$
Wavelength: $\lambda' = \lambda/n$	Velocity: $\upsilon' = \upsilon$
Capacitance $C' = C/n$	Impedance: $Z' = Z$
Inductance: $L' = L/n$	Antenna gain: ${G_0}' = G_0$
Echo area: ${A_e}' = A_e/n^2$	
Frequency: $f' = nf$	
Conductivity: $\sigma' = n\sigma$	

20.13 TROUBLESHOOTING

After an antenna is designed using either analytical methods or computer software, it is desirable to verify the design and its performance via measurements. The analytical or CAD geometry of the antenna is reproduced in the physical world for operation at a particular range of frequencies, and selected parameters (e.g., gain, input impedance, E- and H-plane radiation patterns) are chosen to be measured for comparison with their theoretically predicted counterparts. Occasionally, the comparisons between the measured and predicted values exhibit discrepancies that are greater than acceptable. For this discussion, it is assumed that the predictions are ideal and free of error. In which case, it is up to the measurement engineer/technician to determine the sources of the discrepancies, and to correct them if possible. Sometimes the "error" is unique to the set of circumstances involved in a particular measurement and cannot be anticipated here. There are, however, a number of common sources for these discrepancies, which are discussed in this section.

20.13.1 Dimensional Discrepancies

One of the most likely sources of discrepancy between measured and predicted antenna performance is a lack of congruency between the CAD model of the antenna and its physical realization. Clearly, measurements of an antenna cannot be expected to agree well with predictions of an antenna having a different geometry.

The actively radiating element(s) is critical, and errors in its dimensions can cause a frequency shift in its resonance and changes in input impedance and can alter the gain characteristics of the antenna. The structures in close proximity to the active element, including cavities, substrate heights and parasitic elements, can also be very sensitive to small dimensional variations. Also be alert to the separation distances between these structures and the active element. Higher-order effects can be linked to mechanical discrepancies that are not immediately obvious, resulting in performance discrepancies at certain frequencies and observation angles. For example, ground plane edges that are imperfectly straight and/or that have angular orientations different from the design (usually orthogonal to the antenna's principal planes) can redirect scattered fields away from the plane of the measurement.

The electrical properties of materials (substrates, superstrates, loading ferrites and dielectrics, etc.) should be identical to those in the CAD model. For cases in which materials with the designed properties cannot be obtained, it may be necessary to repeat the predictions using the properties of the materials actually used in the physical antenna.

Finally, as the measurement frequency increases, the absolute tolerances that must be held on the dimensions and material parameters of the physical antenna decrease. While good agreement might be obtainable with an antenna that is constructed to within 1.0 mm of the designed dimensions at a frequency of 200 MHz, for example, to achieve a similar level of agreement at 20 GHz, the dimensions would need to be within 0.01 mm of those specified.

20.13.2 Angular Misalignments

Angular misalignments are another possible source of error (as it is implicitly defined in this section, e.g., the discrepancy between measurements and predictions) that might

be overlooked by the newly practicing measurement engineer, particularly with respect to gain and pattern measurements. It is most convenient to reference alignments to the vertical, enabling the use of spirit levels to verify that antenna surfaces and edges are orthogonal to the vertical, and carpenter's protractors to measure intended angles relative to the vertical. Alignments that can usually be measured using these tools include:

- Polarization angles (for linearly polarized antennas) of the source, gain standard, and test antennas
- Elevation angles of the source, gain standard, and test antennas
- Rotation axes of the positioning system

The source and test antennas should be located in the same horizontal plane (at the same height above a level floor). In a compact range environment, this would usually be located at the vertical center of the main reflector.

The gains of calibration standards are defined at their boresight directions. After ensuring the other alignments, if the gain standard is mounted on the positioning system, the azimuthal table of the positioner might be employed to find the peak gain and hence its boresight direction. Another strategy to align a horn antenna (e.g., a standard gain horn—SGH) in azimuth is to temporarily affix a flat mirror across its aperture. Assuming that the boresight direction of the SGH is perpendicular to its aperture, the boresight direction will be pointed directly at the location of a laser pointer when the beam of the laser is reflected back to its point of origin by the mirror. These methods can also be used to align the test antenna in azimuth.

A common practice is to utilize a support structure between the positioning system and the test antenna to raise the antenna up to the measurement plane and to isolate it from the positioner. This support structure might be an expanded polystyrene column. If the base of the column is centered on the azimuthal rotation axis but the top of the column is not, then the top of the column (along with the test antenna) will move in a circle while an azimuthal pattern is measured. Furthermore, if the antenna mounted at the top of such a column is aligned to the vertical at one azimuthal orientation (e.g., at the beginning of an azimuthal pattern), its orientation relative to the vertical will continuously change as a function of azimuthal angle. To detect such a condition, place a stationary reference pointer close to the side of the support column, near its top. With the azimuthal axis of the positioning system rotating continuously, observe the separation distance between the pointer and the column. If this distance is constant (within some small tolerance), the column is properly aligned, and its top is nearly centered on the rotation axis. If the separation between the pointer and the column varies with azimuthal angle, realign the column to correct this condition.

For copolarized measurements, these alignments enhance measurement accuracy and help to ensure that measured pattern cuts are in the intended planes about the antenna. The cross-polarized fields, on the other hand, are extremely sensitive to these alignments. Very careful attention to all alignments are imperative in order to accurately measure the cross-polarized component patterns of most linearly polarized antennas.

20.13.3 RF Cables

While it is nearly universal that an RF cable connects a test antenna to the receiving instrumentation during a measurement, it is rare that such a cable is included in the

analytical model. Therefore its presence represents a common source of possible measurement error. How much of an error can vary widely depending on the impact that its presence has on the value to be measured. For many classes of test antennas, the cable is practically invisible. For example, an antenna mounted on one side of a ground plane usually has its input connector located on the opposite side. If the ground plane is large, the cable has essentially no impact on the measurements, especially in the half-space of the antenna element side of the ground plane. Similarly, for antennas mounted on scale models, the cable can be routed through the inside of the model to exit a long distance away from the active element, and at a point not in the plane of the measurement. Also, it is usually possible to route an RF cable connected to a highly directive test antenna in such a way as to minimize the impact of its presence.

The presence of the RF cable becomes a serious impediment to measurement accuracy when the electrical size of the antenna is small, when the directivity of the antenna is low, or when the antenna system is balanced and a balun is not integral to its design. For some types of instrumentation and test antennas, the ideal solution is to eliminate the RF cable. An example might involve building a small, battery-powered CW oscillator into the body of a cell phone for measuring the radiation patterns of cell phone antennas. This strategy is not an option for network analyzer-based instrumentations.

Sometimes the only option is to attempt to minimize the effect of the cable. Among the methods to do so include the following:

- Avoid routing the length of the cable parallel with the receiving polarization. Currents induced on the cable reradiate, and the cable behaves like an electrically long monopole. If the reradiated fields are cross polarized to those of the intended measurement, their received amplitude is minimized.
- Avoid routing the cable through the plane of the measurement.
- Use a small-diameter cable in the vicinity of the antenna in an attempt to minimize scattering and/or shadowing effects.
- Use ferrite beads on the cable to choke off stray currents.
- Apply microwave absorber on the outside of the cable. There are ferrite-loaded elastomeric tubes intended for this application. A possible contraindication to this strategy occurs when the increased diameter of the cable plus absorber would increase shadowing effects.
- There are instances in which it is advantageous to route the cable along the rotation axis through a hollow support structure. Although the gain and frequency response errors still exist, the contribution to the radiation pattern from the cable can be made constant.

There are other possible measurement issues associated with RF cables. One of these occurs when the measurement frequency exceeds the operating frequency limit of the cable. In general, cable losses (at a given frequency) decrease as the overall diameter of the cable increases. Cable losses are generally undesirable, and it is reasonable to select the largest cable that is consistent with any mechanical constraints of the antenna/support system (the ability of the cable to withstand small-radius bends, to fit through an aperture, to be lightweight enough to be supported by the antenna/support structure, etc.). However, the single-mode operating frequency of the cable decreases with increasing diameter. When the measurement frequency exceeds the cable's operational frequency limit (the

cutoff frequency), the cable can support modes other than the TEM mode. When this occurs, the additional transmission modes can interfere destructively with the desired TEM mode, resulting in dramatic reductions in the transmitted power at some frequencies.

Another possible cable-related issue involves RF leaks. If the shielding of the cable, due to either wear, damage, or by design, is not significantly greater than the lowest amplitude that is measurable, fields passing through the outer conductor of the cable can cause measurement errors. This is not a common occurrence. However, if a leaky cable or connector is suspected, replace the antenna under test with a matched load and observe the received power. If the received power is above the noise floor of the instrumentation, the cable likely needs to be replaced. Keep in mind that semirigid cables (those with solid copper outer conductors) inherently have a higher degree of shielding than flexible cables (those having an outer conductor comprised of many small metal strands braided together). That statement is not meant to imply that flexible cables are deficient in shielding. Most double-shielded flexible cables (having outer conductors comprised of both braided wires and thin metal foil) have excellent shielding performance that exceeds the requirements of nearly any antenna measurement application.

In general, keep in mind that coaxial cables and their connectors can wear out and fail. Clean the mating surfaces of connectors frequently. Dirty connectors accelerate wear and decrease performance. If the power transmitted through a cable fluctuates abruptly when it or its point of connection is being flexed, it probably needs to be repaired or replaced.

20.13.4 Environmental Reflections

Implicit in the measurement of antenna gain or radiation patterns is that a test signal only propagates from one antenna in a straight line directly to a receiving antenna. However, energy from the transmitting antenna propagates in varying degrees outward in *all* directions, and the receiving antenna can receive energy from *any* direction. Any nearby object is a potential source of reflection that can cause measurement error by introducing an undesired signal path in addition to the path directly between the two antennas. Since microwave absorbers are finitely effective, even an absorber-covered wall inside an anechoic chamber can be such a reflecting object.

The following is a hypothetical example to illustrate how an environmental reflection can cause a measurement error. Consider the azimuthal radiation pattern measurement of a +20-dBi horn antenna that is being operated in the receiving mode. The auxiliary antenna is the transmitting antenna, and to simplify the discussion it is postulated to be an isotropic radiator. The pattern measurement begins at $\theta = 0^\circ$, with the boresight direction of the horn pointed toward the auxiliary antenna. Measurements of the horn's main beam are comparatively immune from environmental reflections because the relatively high gain of its main lobe biases the reception in favor of energy coming from the direction of the auxiliary antenna (the desired signal path direction). The horn rotates in azimuth as the pattern measurement progresses until the horn is oriented at $\theta = 37^\circ$, where a null in its radiation pattern is pointed toward the auxiliary antenna. The true value of this null is −40 dBi. However, at this orientation the boresight direction of the horn is pointed at the specular reflection point on the anechoic chamber wall, between it and the auxiliary antenna. At this angle and frequency, the absorber on the wall has a reflectivity of −35 dB, and the propagation distance from the auxiliary antenna to the wall and then to the horn is such that the reflected signal is in phase with the direct signal from the auxiliary

antenna to the horn. The reduction in power of the reflected signal caused by its longer propagation distance and the associated additional $1/r^2$ space attenuation is neglected for this example. Because the reflected signal is received through the main beam of the horn, the resulting erroneously measured radiation pattern exhibits a small -14.99-dBi lobe at $\theta = 37^\circ$ instead of the -40-dBi null.

Although the consequences to the measured pattern were modest, the previous example is representative of an error due to an environmental reflection. The impact of an environmental reflection can, of course, be more severe than was illustrated above. When such an error is suspected, due to a discrepancy between measurement and prediction or an asymmetry in the measured pattern of a symmetrical antenna, two courses of action are indicated: first, locate the source of the reflection; second, reduce or eliminate the reflection (or its effects).

Identifying the source of a reflection may not always be trivial. As in the example, the orientation of the test antenna can be an important clue to its location. If a time-domain response of the measurement environment is available, it can be a valuable tool in finding the reflection by indicating the scattered signal path length relative to the desired signal path length. The location of a reflection might be inferred by deliberately enhancing the reflectivity at the suspected location and observing the resulting effects on the measurement. If the intentional disturbance to the radiation pattern is similar in nature and occurs at the same observation angle as the one suspected, then it is probable that the location has been identified. In more difficult cases, the offending signal might be reflected from more than one surface before it is receivable by the test antenna.

Having identified the point of reflection, it may simply be a matter of employing some local treatment by, for example, applying absorbing material to its surface to reduce its reflectivity. If the surface is already covered by absorber, it might be possible to reduce the reflectivity further by replacing the existing absorber with some that is more appropriate for the application or frequency in question. Alternatively, the reflectivity of the absorber-covered surface might be modified by applying pieces of additional absorber, between the pyramids of carbon-foam material, for example. Finally, it might be possible to modify the geometry of the measurement configuration either to scatter the energy away from the AUT (antenna under test) or to orient a relative null in the pattern of the auxiliary antenna toward the point of reflection.

REFERENCES

1. C. A. Balanis, *Antenna Theory Analysis and Design*, 3rd ed., John Wiley & Sons, Hoboken, NJ, 2005, Chap. 17.
2. *IEEE Standard Test Procedures for Antennas*, IEEE Std 149-1979, published by IEEE, Inc., 1979, distributed by Wiley-Interscience.
3. W. H. Kummer and E. S. Gillespie, Antenna measurements—1978, *Proc. IEEE*, Vol. 66, No. 4, pp. 483–507, April 1978.
4. J. S. Hollis, T. J. Lyon, and L. Clayton, Jr., *Microwave Antenna Measurements*, Scientific-Atlanta, Atlanta, GA, July 1970.
5. E. F. Buckley, *Modern Microwave Absorbers and Applications*, Emerson & Cuming, Canton, MA.
6. R. C. Johnson, H. A. Ecker, and J. S. Hollis, Determination of far-field antenna patterns from near-field measurements, *Proc. IEEE*, Vol. 61, No. 12, pp. 1668–1694, December 1973.

7. J. Brown and E. V. Jull, The prediction of aerial patterns from near-field measurements, *IEE (London)*, Paper No. 3469E, pp. 635–644, November 1961.
8. D. T. Paris, W. M. Leach, Jr., and E. B. Joy, Basic theory of probe-compensated near-field measurements, *IEEE Trans. Antennas Propag.*, Vol. 26, No. 3, pp. 373–379, May 1978.
9. E. B. Joy, W. M. Leach, Jr., G. P. Rodrigue, and D. T. Paris, Applications of probe-compensated near-field measurements, *IEEE Trans. Antennas Propag.*, Vol. 26, No. 3, pp. 379–389, May 1978.
10. L. H. Hemming and R. A. Heaton, Antenna gain calibration on a ground reflection range, *IEEE Trans. Antennas Propag.*, Vol. 21, No. 4, pp. 532–537, July 1973.
11. P. W. Arnold, The "slant" antenna range, *IEEE Trans. Antennas Propag.*, Vol. 14, No. 5, pp. 658–659, September 1966.
12. A. W. Moeller, The effect of ground reflections on antenna test range measurements, *Microwave J.*, Vol. 9, pp. 47–54, March 1966.
13. W. H. Emerson, Electromagnetic wave absorbers and anechoic chambers through the years, *IEEE Trans. Antennas Propag.*, Vol. 21, No. 4, pp. 484–490, July 1973.
14. M. R. Gillette and P. R. Wu, RF anechoic chamber design using ray tracing, *1977 Int. IEEE/AP-S Symp. Dig.*, pp. 246–252, June 1977.
15. W. H. Emerson and H. B. Sefton, An improved design for indoor ranges, *Proc. IEEE*, Vol. 53, pp. 1079–1081, August 1965.
16. J. R. Mentzer, The use of dielectric lenses in reflection measurements, *Proc. IRE*, Vol. 41, pp. 252–256, February 1953.
17. P. A. Beeckman, Prediction of the Fresnel region field of a compact antenna test range with serrated edges, *Proc. IEE*, Vol. 133, Pt. H, No. 2, pp. 108–114, April 1986.
18. H. F. Schluper, Compact antenna test range analysis using physical optics, in *AMTA Proceedings*, Seattle, WA, October 1987, pp. 309–312.
19. H. F. Schluper and V. J. Vokurka, Troubleshooting limitations in indoor RCS measurements, *Microwaves & RF*, pp. 154–163, May 1987.
20. http://www.mi-technologies.com.
21. V. J. Vokurka, Seeing double improves indoor range, *Microwaves & RF*, pp. 71–76, 94, February 1985.
22. http://web.mit.edu/istgroup/ist/research/microwaveY2004.html.
23. T. Harrison, A new approach to radar cross-section compact range, *Microwave J.*, pp. 137–145, June 1986.
24. http://www.harris.com.
25. K. W. Lam and V. J. Vokurka, Hybrid near-field/far-field antenna measurement techniques, in *AMTA Proceedings*, Boulder, CO, October 1991, pp. 9-29–9-34.
26. C. R. Birtcher, C. A. Balanis, and V. J. Vokurka, RCS measurements, transformations, and comparisons under cylindrical and plane wave illumination, *IEEE Trans. Antennas Propag.*, Vol. 42, No. 3, pp. 329–334, March 1994.
27. C. R. Birtcher, C. A. Balanis, and V. J. Vokurka, Quiet zone scan of the single-plane collimating range, in *AMTA Proceedings*, Boulder, CO, October 1991, pp. 4-37–4-42.
28. H. F. Schluper, Verification method for the serration design of CATR reflectors, in *AMTA Proceedings*, Monterey, CA, October 1989, pp. 10-9–10-14.
29. W. D. Burnside, M. C. Gilreath, and B. Kent, A rolled edge modification of compact range reflectors, presented at *AMTA Conference*, San Diego, CA, October 1984.
30. W. D. Burnside, M. C. Gilreath, B. M. Kent, and G. L. Clerici, Curved edge modification of compact range reflector, *IEEE Trans. Antennas Propag.*, Vol. 35, No. 2, pp. 176–182, February 1987.

31. M. R. Hurst and P. E. Reed, Hybrid compact radar range reflector, in *AMTA Proceedings*, Monterey, CA, October 1989, pp. 8-9–8-13.
32. J. P. McKay and Y. Rahmat-Samii, Multi-ring planar array feeds for reducing diffraction effects in the compact range, in *AMTA Proceedings*, Columbus, OH, October 1992, pp. 7-3–7-8.
33. J. P. McKay, Y. Rahmat-Samii, and F. M. Espiau, Implementation considerations for a compact range array feed, in *AMTA Proceedings*, Columbus, OH, October 1992, pp. 4-21–4-26.
34. J. P. McKay and Y. Rahmat-Samii, A compact range array feed: tolerances and error analysis, *1993 Int. IEEE/AP-S Symp. Dig.*, Vol. 3, pp. 1800–1803, June 1993.
35. J. P. McKay, Y. Rahmat-Samii, T. J. De Vicente, and L. U. Brown, An X-band array for feeding a compact range reflector, in *AMTA Proceedings*, Dallas, TX, October 1993, pp. 141–146.
36. H. F. Schluper, J. Van Damme, and V. J. Vokurka, Optimized collimators—theoretical performance limits, in *AMTA Proceedings*, Seattle, WA, October 1987, p. 313.
37. J. D. Huff, J. H. Cook, Jr., and B. W. Smith, Recent developments in large compact range design, in *AMTA Proceedings*, Columbus, OH, October 1992, pp. 5-39–5-44.
38. C. R. Birtcher, The electromagnetic anechoic chamber at Arizona State University: characterization and measurements, MSEE thesis, Department of Electrical Engineering, Arizona State University, 1992.
39. A. D. Yaghjian, An overview of near-field antenna measurements, *IEEE Trans. Antennas Propag.*, Vol. 34, pp. 30–45, January 1986.
40. S. A. Brumley, Evaluation of microwave anechoic chamber absorbing materials, MSEE thesis, Department of Electrical Engineering, Arizona State University, 1988.
41. S. Cui and D. S. Weile, Robust design of absorbers using genetic algorithms and the finite element-boundary integral method, *IEEE Trans. Antennas Propag.*, Vol. 51, No. 12, pp. 3249–3258, December 2003.
42. V. M. Petrov and V. V. Gagulin, Microwave absorbing materials, *Inorganic Materials*, Vol. 37, No. 2, pp. 93–98, 2001. (Translated from *Neorganicheskie Materialy*.)
43. H. M. Musal, Jr., and H. T. Hahn, Thin-layer electromagnetic absorber design, *IEEE Trans. Mag.*, Vol. 25, No. 5, pp. 3851–3853, September 1989.
44. S. S. Kim, S. B. Jo, K. I. Gueon, K. K. Choi, J. M. Kim, and K. S. Churn, Complex permeability and permittivity and microwave absorption of ferrite–rubber composite in X-band frequencies, *IEEE Trans. Mag.*, Vol. 27, No. 6, pp. 5462–5464, November 1991.
45. J. Y. Shin and J. H. Oh, The microwave absorbing phenomena of ferrite microwave absorbers, *IEEE Trans. Mag.*, Vol. 29, No. 6, pp. 3437–3439, November 1993.
46. http://www.ni.com/labview.
47. A. C. Newell, R. C. Baird, and P. F. Wacker, Accurate measurement of antenna gain and polarization at reduced distances by an extrapolation technique, *IEEE Trans. Antennas Propag.*, Vol. 21, No. 4, pp. 418–431, July 1973.
48. L. H. Hemming and R. A. Heaton, Antenna gain calibration on a ground reflection range, *IEEE Trans. Antennas Propag.*, Vol. 21, No. 4, pp. 532–537, July 1973.
49. R. G. FitzGerrell, Gain measurements of vertically polarized antennas over imperfect ground, *IEEE Trans. Antennas Propag.*, Vol. 15, No. 2, pp. 211–216, March 1967.
50. R. G. FitzGerrell, The gain of a horizontal half-wave dipole over ground, *IEEE Trans. Antennas Propag.*, Vol. 15, No. 4, pp. 569–571, July 1967.
51. R. G. FitzGerrell, Limitations on vertically polarized ground-based antennas as gain standards, *IEEE Trans. Antennas Propag.*, Vol. 23, No. 2, pp. 284–286, March 1975.
52. Antenna calculations, in *ETS-Lindgren Antenna Catalog*, ETS-Lindgren, Cedar Park, TX, 2002, p. 71.
53. H. A. Wheeler, The radiansphere around a small antenna, *Proc. IRE*, pp. 1325–1331, August 1959.

54. E. H. Newman, P. Bohley, and C. H. Walter, Two methods for the measurement of antenna efficiency, *IEEE Trans. Antennas Propag.*, Vol. 23, No. 4, pp. 457–461, July 1975.

55. R. E. Collin, *Foundations for Microwave Engineering*, McGraw-Hill, New York, 1992, pp. 248–257.

56. J. D. Kraus, *Antennas*, McGraw-Hill, New York, 1988.

57. G. Barzilai, Experimental determination of the distribution of current and charge along cylindrical antennas, *Proc. IRE (Waves and Electrons Sec.)*, pp. 825–829, July 1949.

58. T. Morita, Current distributions on transmitting and receiving antennas, *Proc. IRE*, pp. 898–904, August 1950.

59. A. F. Rashid, Quasi-near-zone field of a monopole antenna and the current distribution of an antenna on a finite conductive earth, *IEEE Trans. Antennas Propag.*, Vol. 18, No. 1, pp. 22–28, January 1970.

60. H. G. Booker, V. H. Rumsey, G. A. Deschamps, M. I. Kales, and J. I. Bonhert, Techniques for handling elliptically polarized waves with special reference to antennas, *Proc. IRE*, Vol. 39, pp. 533–552, May 1951.

61. E. S. Gillespie, Measurement of antenna radiation characteristics on far-field ranges, in *Antenna Handbook*, Y. T. Lo and S. W. Lee (Eds.), Van Nostrand Reinhold, New York, 1988, Chap. 32, pp. 32-1–32-91.

62. G. Sinclair, Theory of models of electromagnetic systems, *Proc. IRE*, Vol. 36, pp. 1364–1370, November 1948.

CHAPTER 21

Antenna Scattering and Design Considerations

OREN B. KESLER, DOUGLAS PASQUAN, and LARRY PELLETT

21.1 INTRODUCTION

An aircraft's radar signature can be limited by the sum of its apertures/antennas, but avionics requirements still demand traditional antenna performance. This contradiction of having a platform that is designed to have low radar cross section (RCS) or signature with minimal radio frequency (RF) reflections while simultaneously having avionics systems that require efficient RF radiation through antennas/radomes has profound implications on the design of antennas and platforms. Ultimately, the only way to achieve desired performance is by designing the antenna as a balanced system. Trade-offs between the platform signature and electronics subsystems must be considered together while understanding the system requirements. Typical requirements for low signature antennas include:

- Avionics/electronics system (gain, pattern, impedance, coupling, polarization, average power, peak power, bandwidth, sidelobes, grating lobes)
- Aerodynamic/structural performance (e.g., static loads, fatigue, flutter, weight, shear, impact strength, thermal)
- RCS (different modes)
- Materials (multiple layers, μ, ε, impedance mismatch)
- RM&S (reparability, maintainability, supportability, interchangeability)
- Manufacturability (cost, tolerance, special tooling)

This chapter is primarily about antenna scattering or the radar cross section of antennas and design approaches. Those interested in the general topic of radar cross section including many of the electromagnetic modeling concepts should Consult *Radar Cross Section* [1] and *Radar Cross Section Handbook* [2]. They Provide a much more in-depth background on the general topic of radar cross section.

Modern Antenna Handbook. Edited by Constantine A. Balanis

21.2 CONCEPT OF RADAR CROSS SECTION

By definition, RCS (σ) is a far-field parameter relating the scattered electric field to the incident electric field and it can be expressed as:

$$\sigma = \lim_{R \to \infty} 4\pi R^2 \quad \frac{|\mathbf{E}_s^2|}{|\mathbf{E}_i^2|} \tag{21.1}$$

where

$\mathbf{E}_s$ = electric field scattered or reflected by the target in the direction of the observer
$\mathbf{E}_i$ = electric field incident on the target
R = range from the observer to the scattering object
σ = radar cross section

Conceptually RCS is the "effective size" of the reflection of a target. The RCS of a target is a function of the polarization of the incident wave, the angle of incidence, the angle of observation, the geometry of the target, the electrical properties of the target, and the frequency of operation. RCS is expressed in various ways. For a three-dimensional object, the most common way of expressing RCS (m^2) is using meters squared or for normalized values, decibels per square meter (dBsm). Sometimes RCS is normalized and unitless by expressing scattering as the ratio RCS/λ^2 in dB. Representative values of some typical objects are shown in Table 21.1 [3].

Equation (21.1) is consistent with the physical model shown in Figure 21.1 in relating the received power to the reflected power. RCS or σ can be considered equivalent to the projected area of a sphere ($\sigma = \pi a^2$, a being the sphere's radius) that intercepts an amount of incident power and reflects a portion of that power toward an observer. RCS changes with various parameters. This conceptual model has an angular and electromagnetic dependence associated with the target. If we carry the concept of the sphere further, the "equivalent sphere" for the target changes with aspect angle, frequency, and polarization.

TABLE 21.1 RCS of Some Typical Objects

Object	RCS (m^2)	RCS (dBsm)
Pickup truck	200	23
Automobile	100	20
Jumbo jet liner	100	20
Cabin cruiser boat	10	10
Four-passenger jet	2	3
Adult male	1	0
Conventional missile	0.5	−3
Bird	0.01	−20
Insect	0.00001	−50

Source: From Ref. 3.

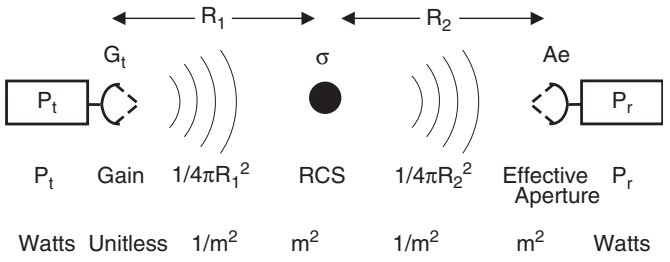


Figure 21.1 Radar-target detection model illustrated using a linear flow diagram relating power transmitted (P_t) to power received (P_r) via a scatterer, σ.

As a physical starting point, modeling the radar's system performance along with the power flow is a valuable method to provide insight into an antenna's RCS. Figure 21.1 models the power flow from a radar to a target/scatterer and finally to a receiver that detects the power. This is the general radar range formulation. As ilustrated in Figure 21.1, the radar is bistatic meaning that the transmitter and receiver are in different locations. This model provides physical insight into radar detection with respect to scattering.

The radar range equation is derived using the model in Figure 21.1. Solving for the power received (P_r) we obtain

$$P_r = P_t \cdot G_t \cdot \frac{1}{4\pi R_1^2} \cdot \sigma \cdot \frac{1}{4\pi R_2^2} \cdot A_e \cdot p_i \tag{21.2}$$

where

P_t = power transmitted
G_t = gain of transmit antenna
R_1 = range from transmit antenna to scatterer
σ = radar cross section (RCS) of a target or scatterer
R_2 = range from scatterer to receiver
A_e = effective area of receive antenna
P_r = power received at the terminal
λ = wavelength
p_i = polarization factor

The effective aperture (A_e) of an antenna (Eq. (21.3)) is the concept of how much power is extracted over an equivalent area. If the receive antenna is efficient, we can substitute Eq. (21.3) for A_e. If the transmit and receive sites are the same (i.e., monostatic case => $R_1 = R_2 = R$ and $G_1 = G_2 = G$) and if polarization is aligned or copolarized ($p_i = 1$), then we generate the familiar Eq. (21.4), the radar range equation. This version of the RCS equation is simplified such that any effects of pulse width and receiver bandwidth are not included (Eq. (21.4) is for continuous wave illumination), but even so it is a very accurate model for the physics of target detection. Equation (21.4) can easily

be transformed into a form that includes pulse train/receiver bandwidth effects related to a signal-to-noise (S/N) detection model of pulse radars or other radar systems.

$$A_e = \frac{\lambda^2}{4\pi} G \tag{21.3}$$

$$P_r = \sigma \frac{G^2 \lambda^2}{(4\pi)^3 R^4} P_t \tag{21.4}$$

Examining the radar range equation and the model in Figure 21.1, we see that σ, or RCS, represents the physical area that intercepts the incoming power density and then reradiates a power. It relates the transmitted power to the power reflected (radiated) in the direction toward the receiver. This is equivalent to the concept definition described in Eq. (21.1). Another way of looking at the radar range equations is that both σ and A_e (effective aperture) represent the interaction with the power density (watts/meter^2) incident at that location. The radar range equation can be seen as two serial Friis transmission equations, in which σ contains the power extracted from the incident field, P_r for the first of the Friis equations, and then becomes the source (or P_t, power transmitted) for the second Friis transmission equation. This is shown in Figure 21.1. By using the definition for A_e shown in Eq. (21.3), and using σ at the target for the power density intercepted and power reradiated, it is easy to recast Eq. (21.2) into the radar range equation:

$$\underbrace{P_r = P_t \cdot G_t \cdot \frac{1}{4\pi R^2}}_{\text{Friis Equation 1}} \cdot \overbrace{\underbrace{A_{e\sigma} \cdot G_\sigma}_{\sigma}}^{} \cdot \underbrace{\frac{1}{4\pi R^2} \cdot A_e}_{\text{Friis Equation 2}} \tag{21.5}$$

RCS or σ is range independent and is a function of the target (shape, composition, etc.). Implied in this definition of σ is that the target is in the far field of the transmitter and receiver and wavefronts at the target as well as the receiver are planar. Furthermore, in keeping with comparing antennas and σ aligned with the concept of antennas, σ can be considered an area intercepting an amount of power density that is reradiated.

Two characteristics suppressed in the radar range equation (Eq. (21.4)) should be further understood before continuing; those are polarization and differences between bistatic/monostatic RCS.

The polarization response of a target's RCS can be quite different depending on the target's geometry and the radar's transmit/receive polarization. The enhancement of RCS with respect to polarization at certain frequencies is so pronounced that radar designers make use of these characteristics to provide maximum system performance. Because of these characteristics, radar designs can make use of vertical polarization, horizontal polarization, circular polarization, and so on depending on the target's polarization characteristics.

Polarization is typically expressed as a scattering matrix that relates the scattered E-field RCS with respect to the incident polarization. A general form of the polarization scattering matrix (SM) used in conjunction with Eq. (21.1) is given as

$$SM = \begin{bmatrix} \sqrt{\sigma_{11}} e^{j\phi_{11}} & \sqrt{\sigma_{12}} e^{j\phi_{12}} \\ \sqrt{\sigma_{21}} e^{j\phi_{21}} & \sqrt{\sigma_{22}} e^{j\phi_{22}} \end{bmatrix} \tag{21.6}$$

where the RCS or σ is for various combinations of transmit (subscript 1) and receive (subscript 2) polarizations and $e^{j\phi}$ establishes the phase relationship of the reflected field to the incident field. Equation (21.6) can be formulated using any two orthogonal polarizations. Typically, horizontal and vertical polarizations are chosen, which results in the scattering matrix being recast as

$$SM = \begin{bmatrix} \sqrt{\sigma_{HH}}e^{j\phi_{HH}} & \sqrt{\sigma_{HV}}e^{j\phi_{HV}} \\ \sqrt{\sigma_{VH}}e^{j\phi_{VH}} & \sqrt{\sigma_{VV}}e^{j\phi_{VV}} \end{bmatrix} \tag{21.7}$$

Using this notation, σ_{VV} is the vertical RCS component for a scatterer that was illuminated by a vertical E-field and σ_{VH} and σ_{HV} are the cross-polarized RCS contributions. If the system is linear and reciprocal, σ_{VH} equals σ_{HV}. Figure 21.2 is an example of the scattering matrix description of scattering for a vertically polarized incident field. As shown in Figure 21.2, a vertical E-field illuminates and scatters from a vertical thin flat plate generating the displayed scattering matrix. The incident condition has the energy normal to the edge of the thin flat plate such that for this example the angular dependence has been suppressed. The scattering matrix is shown for this single incident angle. The scattering matrix indicates that the scattered RCS for this case will have a magnitude of $\sigma = L^2/\pi$ (assumption is $L \gg \lambda$) and will be collinear with the incident or vertical E-field. The horizontal RCS will be negligible.

Polarization effects will seem to be somewhat minimized in the remainder of this chapter but for completeness the polarization effects must be understood. Antennas can set up numerous modes that both cancel and provide enhanced antenna performance depending on polarization. A good example is the conformal annular slot antenna mounted on the upper fuselage of an aircraft. The slots on each side of the antenna are phased to provide a null on boresight (if on top of an aircraft, the boresight direction would be zenith) and summed to provide vertical polarization on the horizon. Figure 21.3 illustrates an example of a conformal annular cavity backed slot antenna fed to generate a null on zenith and enhanced radiation (gain) toward the horizon. If one were to look at the antenna performance, one might be inclined to consider only the σ_{VV} components of its scattering matrix since that is where the antenna has gain. This would be a mistake since the electromagnetic components that make up the RCS may not combine in the same way as the electromagnetic components combine as antenna energy.

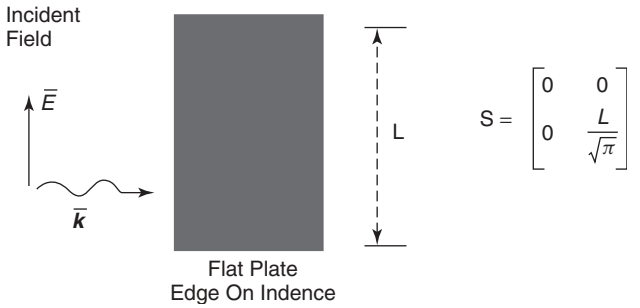


Figure 21.2 Example of scattering matrix for a very thin flat plate having a width $= L$. The incident E-field is collinear with the plate edge. Illumination direction is normal to the plate edge. The resulting scattered field or RCS is collinear with the incident E-field. No scattered energy is observed when illuminated by a horizontal E-field. **k** is the propagation vector. (From Ref. 4.)

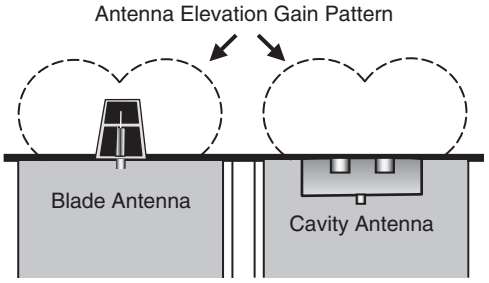


Figure 21.3 Different antennas with similar gain and patterns may have different RCS due to the summation of the antenna modes and structural/resonant mode. The structural mode of the sleeve monopole blade antenna in this figure would have significantly higher RCS.

One other definition that needs to be established is the difference between monostatic RCS and bistatic RCS. Monostatic RCS is the RCS measured when the radar transmitter and receiver are in the same direction. Bistatic RCS is the RCS when the radar transmitter and the receiver are in different directions. Monostatic radar systems are prevalent and as a result when the single term "RCS" is discussed, it most often refers to monostatic RCS.

Another characteristic of the monostatic radar system when compared to a bistatic radar system is the monostatic azimuth RCS pattern has a lobe structure that has twice the frequency (in angle space) as the bistatic case; in other words, the monostatic's RCS spatial frequency is twice that of a bistatic RCS pattern for a given scatterer or an antenna aperture. This is caused by the transmit and receive antennas moving in angle simultaneously and creating an effective two times angle change for any given angle change with respect to the target. Figure 21.4 illustrates the effect with the antenna shown having a bistatic RCS pattern twice as wide (angle extent) as the monostatic RCS pattern.

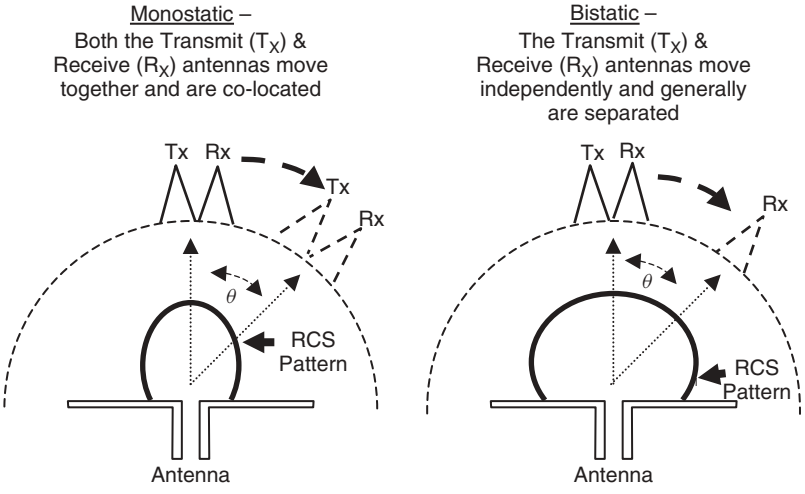


Figure 21.4 Monostatic RCS refers to having the transmitter and receiver in the same direction and typically if it is not specified, RCS refers to monostatic.

21.3 RADAR CROSS SECTION OF ANTENNAS: MULTIPLE MODES

The function of an antenna is to provide a method of transforming radiated electromagnetic energy into currents/voltage to be used by circuits, or reciprocally, take power generated by the circuit and through currents/voltage on the antenna transform that power into radiated electromagnetic waves. The antenna is a transducer converting energy from one form to another. This coupling is based on the resonant modes (current/voltage) on the antenna structure, in which substantial gain in the coupling is possible based on antennas with excellent coupling qualities. In addition, the antenna also contains structures that are not involved with the coupling of power. These structures are used to support or protect the antenna. Other structures such as fasteners may be in the vicinity of the radiating structure and will also modify and influence the antenna performance by interacting with the electromagnetic energy.

To gain insight into antenna scattering, it is convenient to break up the scattering mechanism into different components. Earliest work in this area was by Green [5] in which antenna scattering was broken up into two components: residual scattering and antenna mode scattering. It is important to realize that these terms are not standard. In some texts, the antenna mode is considered to be only the energy that is delivered to the antenna terminals while any residual component is what remains. This would allow for an antenna mode to be zero if it is correctly matched and if all energy delivered to the antenna terminals was absorbed. The current that is on the antenna and reradiates along with any other structural scattering is combined as the residual scattering mode. The concept of breaking up antenna RCS modes into an antenna mode that can be zero with the correct match allows the problem to be worked such that if the residual scattering mode can also be designed as zero, you have a low RCS antenna. This residual remainder would include the reradiated energy from the currents on the antenna as well as the surrounding structural components. This antenna/residual mode concept is particularly useful when working with arrays. This chapter evolves the concept of breaking antenna scattering into components along a slightly different model.

Here we intend to break up antenna scattering into three components:

- *Structural mode scattering* is a function of the structure and is not an effect of the antenna being resonant or operating as an antenna. Structural scattering is caused by discontinuities, impedance mismatch, material boundaries, and so on. Of particular importance is that radiating elements outside their band of operation (when they are no longer resonant or antenna elements) are considered structural scattering mechanisms.
- *Antenna resonant mode scattering* is the scattering caused by the resonating currents on the antenna. Currents on the antenna will deliver energy to the terminals but these same currents may reradiate regardless of the antenna terminal load. Since this mode is directly a function of the currents on the antenna, the scattering from the antenna resonant mode would be proportional to the gain and pattern of the antenna.
- *Antenna terminal mode scattering* is a function of how the antenna is terminated or loaded. As an antenna couples energy from free space to its terminal, characteristics at that terminal or its load will determine how much energy is reradiated as antenna terminal mode RCS. In concept, if the terminal is shorted, all of the energy received by the antenna would be reradiated.

(If one were to equate these concepts to Green's development [5], one would sum the *structural mode scattering* plus the *antenna resonant mode scattering* to equal the *residual mode scattering*. Many texts use residual mode scattering for all scattering except the antenna terminal mode scattering.)

Tight coupling between RCS and antenna performance forces tough decisions during the design of the antenna. This definition also allows one to partition the problem into three very different categories and to systematically attack antenna scattering reductions.

Studying the three modes separately would seem to be an oversimplification of the problem, but by separating the three modes and studying the behavior of these modes we obtain insight into the antenna scattering problem. RCS when converted to m^2 (instead of its usual form of dBsm) is a linear system and is a sum of its individual responses. The total integration of all scattering modes on the antenna will be dominated by the scattering mechanism that is the largest. Based on the relationship between the RCS of antennas and their structures, it is apparent that one must study the antenna support structure as well as the antenna launch structure. The antenna terminal and resonant modes will be investigated first since they are unique to an antenna. Figure 21.3 illustrates how two antennas with very similar gain patterns would have very different RCS values. The blade antenna having a large amount of metal protruding from the surface and making a right angle with the ground plane will have an RCS much greater than that of the cavity antenna due to structural scattering.

21.4 ANTENNA TERMINAL MODE SCATTERING AND RESONANT MODE SCATTERING FOR A SMALL SINGLE WIRE ANTENNA ELEMENT

We have already described how the radar range equation can be viewed as a back-to-back Friis transmission equation (Eq. (21.5)) in which σ represents the amount of energy intercepted and then reradiated. In a similar way, the RCS of the antenna in relationship to gain is found the same way. Using the receive port on the radar range equation for this illustration, we find that the amount of power intercepted by the receive antenna is A_e, the effective aperture. Now applying the same principle as the definition of RCS, the RCS of an antenna should be given by the $A_{e\sigma}$, the power density intercepted by the antenna, multiplied by the gain G_σ of the antenna. Equation (21.8) models this physical situation for an antenna (at σ consistent with Eq. (21.5)) and is the logical derivative of the serial Friis equations.

$$\sigma = A_{e\sigma} \cdot G_\sigma \tag{21.8}$$

An alternate way of determining this equation is to use the fundamental equation for radar cross section, Eq. (21.4). Hansen [6] and Montgomery et al. [7] derive the same equation for a simplified antenna having its RCS due to the collection of the energy by the aperture and then the reradiation of that energy with the antenna shorted. If we have an antenna with an effective area A_e that is immersed in a field with a power density of S_i, then the power received by the antenna as a scatterer is

$$P_r = S_i \cdot A_e \tag{21.9}$$

If the antenna is set up to reradiate its received energy, and as a first approximation it is assumed that the antenna is an isotropic radiator, the omnidirectional energy will

be equally distributed over the entire surface of a sphere about the antenna, represented by $1/4\pi R^2$. To make this equation a bit more general, the concept of gain (G) is used to indicate how much energy is radiated in any given direction when compared to an isotropic radiator. With the physics modeled as such, the reflected power density S_r will be

$$S_r = P_r \cdot G \cdot \frac{1}{4\pi R^2} = \frac{S_i A_e}{4\pi R^2} G \tag{21.10}$$

and using the definition of RCS with power densities replacing electric fields

$$\sigma = \lim_{R\to\infty} (4\pi R^2) \frac{|E_s^2|}{|E_i^2|} = \lim_{R\to\infty} (4\pi R^2) \frac{|S_s|}{|S_i|} = \lim_{R\to\infty} (4\pi R^2) \frac{S_i A_e G}{4\pi R^2 S_i} = A_e \cdot G \tag{21.11}$$

As shown, Eqs. (21.8) and (21.11) are the same using both of these techniques to derive the relationship of RCS with the gain of an antenna. Now using the definition of A_e and G with Eq. (21.11) gives us the typical equation for antenna resonant mode scattering:

$$\sigma = \frac{G^2 \lambda^2}{4\pi} \tag{21.12}$$

Equation (21.12) is the typical equation for antenna resonant mode RCS. It still is not complete in that accommodations for the terminal impedance or load are not factored in Eq. (21.12) but for practical application, Eq. (21.12) is a good benchmark on what an individual radiating antenna will have as its RCS based on antenna factors (remember that polarization effects have been simplified, this is for a monostatic condition, and no accommodations for load impedance or antenna mismatch have been modeled). From Eq. (21.12) we see that the RCS is related to the gain of the antenna and has the same pattern as the gain squared of the antenna. Intuitively, this is sometimes called the "cat's eye" effect and it can be somewhat interpreted that for an antenna to receive energy that antenna will scatter energy and the higher the efficiency in receiving energy (higher gain) the higher the scattering or RCS. In effect, the currents that resonate on the antenna to deliver energy to a circuit will also be the currents that reradiate and cause the antenna resonant mode RCS. (See Figure 21.5.)

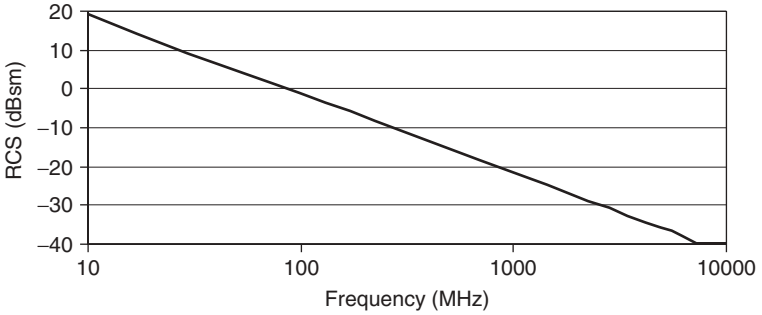


Figure 21.5 Based on the derivation of Eq. (21.12), for a 0-dBi antenna gain, a good approximation for the antenna resonant mode RCS is plotted versus frequency. This represents the contribution from one of the multiple antenna RCS modes.

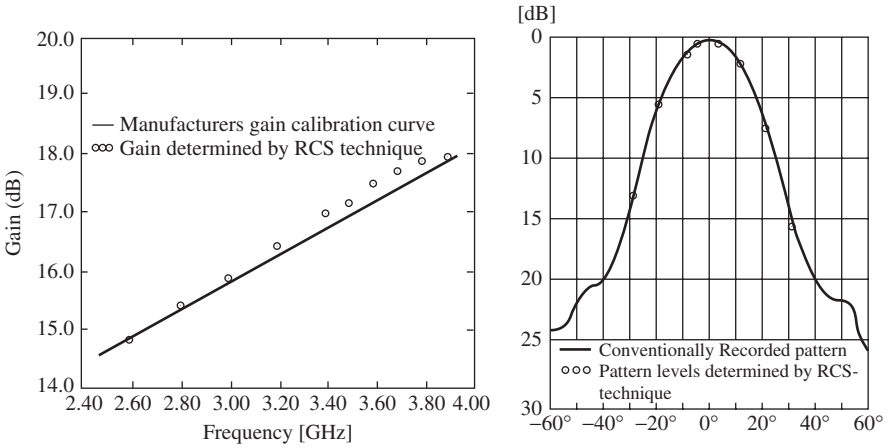


Figure 21.6 Pyramidal standard gain horn antenna gain and pattern computed from measuring the RCS and calculating the resultant gain and pattern. (From Ref. 8.)

Another interesting form that Eq. (21.12) can take is to find the RCS in terms of the effective aperture, A_e. Using Eq. (21.3) we can recast the equation to the form shown in Eq. (21.13). What is significant is that the form of Eq. (21.13) is similar to the normal RCS return of a metallic flat plate with the area of a metallic flat plate, A, replaced with the antenna effective area, A_e. Intuitively, this makes sense especially for large aperture antennas or arrays in which the antenna occupies an area and the RCS is proportional to the area squared.

$$\sigma = 4\pi A_e^2/\lambda^2 \tag{21.13}$$

Figure 21.6 is an example of measuring an antenna's RCS and determining its gain using these concepts. In this example, a pyramidal standard gain horn antenna had its RCS measured in an anechoic chamber. From these RCS measurements, the antenna's gain and pattern were generated utilizing the relationship between RCS and antenna gain/pattern.

21.5 IMPEDANCE EFFECTS ON ANTENNA SCATTERING

Antenna gain is the metric that defines the coupling between the electromagnetic fields and circuits. For an elemental antenna, it is convenient to model the antenna and its load impedance by a lumped Thevenin equivalent source. The Thevenin source is V_A, and the equivalent impedance $Z_A = R_A + jX_A$, where R_A is the antenna radiation resistance and X_A is related to the energy stored in the near field [5].

Figure 21.7 illustrates a model for this condition and Z_L is the load seen by the antenna. From this simple model it can be shown that, as a maximum, only 50% of the power captured by the antenna can be delivered to the load, Z_L, and that occurs for the ideal case of $Z_L = Z_A^*$, the complex conjugate. If $Z_L \neq Z_A^*$, the reradiated power is increased until the antenna is shorted, at which point all of the captured power is reradiated. For the general case, the remaining power that is not dissipated in Z_L must be dissipated in R_A and is the antenna resonant mode contribution [9].

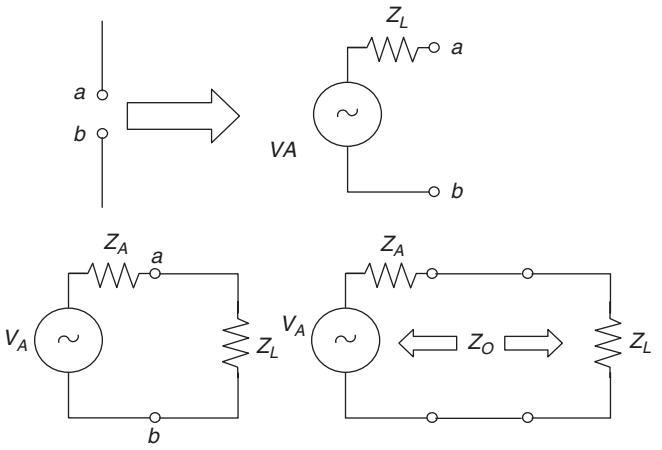


Figure 21.7 A dipole is modeled as a Thevenin equivalent circuit having a driving source of V_A, an equivalent impedance of Z_A, and a load impedance of Z_L with a transmission line impedance of Z_0.

Deriving the general case of antenna scattering with different impedances has been done in a number of ways. Hansen [6] used an approach favored by Collin [10] and Montgomery et al. [7] that incorporates many of the previous concepts put forward. The derivation makes use of the idea that the scattered electric field, $\mathbf{E}_S$, is equal to the total electric field, $\mathbf{E}_T$, minus the incident electric field, $\mathbf{E}_I$.

$$\mathbf{E}_S = \mathbf{E}_T - \mathbf{E}_I \tag{21.14}$$

For this concept, the antenna is considered to have a single port, M (for simplification, more ports could be included), and the electric field is assumed to be expanded in a series of spherical waves, N modes. The complex applied and reflected electric field waves in the feed network are represented by a and b, respectively, and the incoming and outgoing spherical waves are represented by c_n and d_n. The scattering matrix for this geometry, $\mathbf{S}$, can be written

$$\begin{vmatrix} b \\ d_1 \\ \vdots \\ d_N \end{vmatrix} = \begin{vmatrix} S_{00} & S_{01} & \cdots & S_{0N} \\ S_{10} & \cdot & \cdot & \vdots \\ \cdots & \cdot & \cdot & \vdots \\ S_{N0} & \cdot & \cdots & S_{NN} \end{vmatrix} \begin{vmatrix} a \\ c_1 \\ \vdots \\ c_N \end{vmatrix} \tag{21.15}$$

For this derivation in which we initially are assuming the antenna to be radiating only, all of the incoming spherical waves, c_n, are set to 0 and the scattering matrix reduces to $b = S_{00} \cdot a$ with S_{00} as the antenna reflection coefficient with a wave incident from the source. To obtain the total field, the antenna is removed and $d_n = c_n$; that is, the incoming and outgoing spherical waves are equivalent and this results in the identity matrix. When the antenna is driven by the incident fields, the mode in the feed reflects at the load impedance and $a = \Gamma b$, where Γ is the reflection coefficient. We now have all of the components to solve for the scattering electric field by subtracting the incident field from the total field. The scattered field is obtained by the summation of

all d_n and using the open condition to separate components into a scattering mode and antenna mode.

$$d_n^s = d_n^{sc} + \frac{S_{n0}(1+\Gamma)}{(1-\Gamma S_{00})(1+S_{00})} \sum_1^N S_{0m} c_m \tag{21.16}$$

where

$d_n^s = E_s$, total scattered field
$d_n^{sc} = E_{sc}$, field when the antenna is shorted

This ultimately becomes

$$E_{\text{scat}} = E_{\text{short}} - E_{\text{ant}} \frac{I_0 Z_L}{I_a (Z_a + Z_L)} \tag{21.17}$$

where

I_0 = current on a short circuit antenna from incident external fields
I_a = antenna driving current
Z = impedance consistent with the definitions shown in Figure 21.7

Finally, for a small elemental antenna, Eq. (21.17) can be simplified to

$$\sigma = \frac{G^2 \lambda^2}{4\pi} |\Gamma - 1|^2 \tag{21.18}$$

If we look at the conditions of an open circuit, matched load, and shorted feed for a single small, very thin, wire antenna, we obtain for the scattering normal to the wire

$$\sigma = \begin{cases} 0 \text{ (or } -\infty \text{ dBsm)}, & \text{open circuit feed} => \Gamma = 1 \\ \dfrac{G^2 \lambda^2}{4\pi}, & \text{matched load} => \Gamma = 0 \\ \dfrac{G^2 \lambda^2}{4\pi} \times 4, & \text{shorted feed} => \Gamma = -1 \end{cases}$$

If the antenna's terminal is open for this small elemental antenna, the antenna no longer operates as an antenna. With the open terminal, the two halves of the thin dipole would behave as two individual structural scatterers (they still could be two antennas, operating at twice the initial frequency with their terminal shorted). This is an example of an antenna becoming only a structural scatterer outside its band of operation. With a matched load, we get the equation developed earlier in this chapter, which is equivalent to the RCS equal to the antenna resonant mode, Eq. (21.12). With the feed shorted, all of the energy that would have been delivered to the load would be reradiated along with the antenna resonant mode component. Summing both the components

TABLE 21.2 Comparison of Small Antenna Scattering from Antennas with Various Impedance Conditions Across the Terminals[a]

	λ/10 Dipole			λ/2 Dipole		
RCS (dBsm)	Open ($\Gamma = 1$)	Load (matched)	Short ($\Gamma = -1$)	Open ($\Gamma = 1$)	Load (matched)	Short ($\Gamma = -1$)
Simplified formula (dBsm)	$-\infty$	−17.9	−11.9	$-\infty$	−17.1	−11.1
MOM code (dBsm)	−75	−17.9	−11.9	−35.7	−17.2	−11.2

[a]Comparison is between Eq. (21.18) and a method of moments analysis.

Source: From Ref. 11.

from the antenna resonant and antenna terminal modes increases the RCS by a factor of 4 (or 6 dB).

Example 1 Equation (21.18) can be used to find the RCS for different small, thin wire antennas with varying loads. In this example, calculations using Eq. (21.18) are compared to a method of moments solution (Table 21.2) for a short dipole ($\frac{1}{10}$ wavelength) and a half-wave dipole having three different load conditions—matched, shorted, and open.

Gain of λ/10 dipole (tuned) = 1.8 dBi
Gain of λ/2 dipole = 2.2 dBi → (1.643 linear)
Matched load for a λ/10 antenna is $R_L = 1.5\ \Omega + j1000\ \Omega$ (inductive)
Matched load for a λ/2 antenna is $R_L = 72\ \Omega + j0\ \Omega$
Frequency for this example is 1 GHz

For the half-wave dipole with a matched load, we obtain

$$\sigma = \frac{G^2\lambda^2}{4\pi}|\Gamma - 1|^2 \rightarrow G = (1.643)^2, \quad \Gamma = 0 \text{ (matched)}, \quad \lambda = 0.3 \text{ meter}$$

$$\rightarrow [(1.643)^2 \times (1)^2 \times 0.3^2]/4\pi$$

$$\rightarrow 0.01933 \text{ m}^2$$

$$\sigma = 10 \log_{10}(0.01933) = -17.1 \text{ dBsm}$$

In this manner, all of the different scattering conditions can be calculated using Eq. (21.18) for different load conditions. The method of moments comparison shows good agreement and both of these antennas exhibit the characteristic that their signature increases 6 dB when the antenna match is replaced by a short. For the two antennas in this example, when they are terminated in an open, they have very low scattering since the antennas no longer resonate or operate like antennas at the analyzed frequency.

Figure 21.8 pulls many of these concepts together using a time-domain scattering measurement of a standard gain horn antenna. The time-domain measurements clearly separate out the different scattering modes. Also, the two conditions measured include the antenna terminated with a matched load and with a short. As shown, the terminal mode is reduced by 20 dB with the matched load.

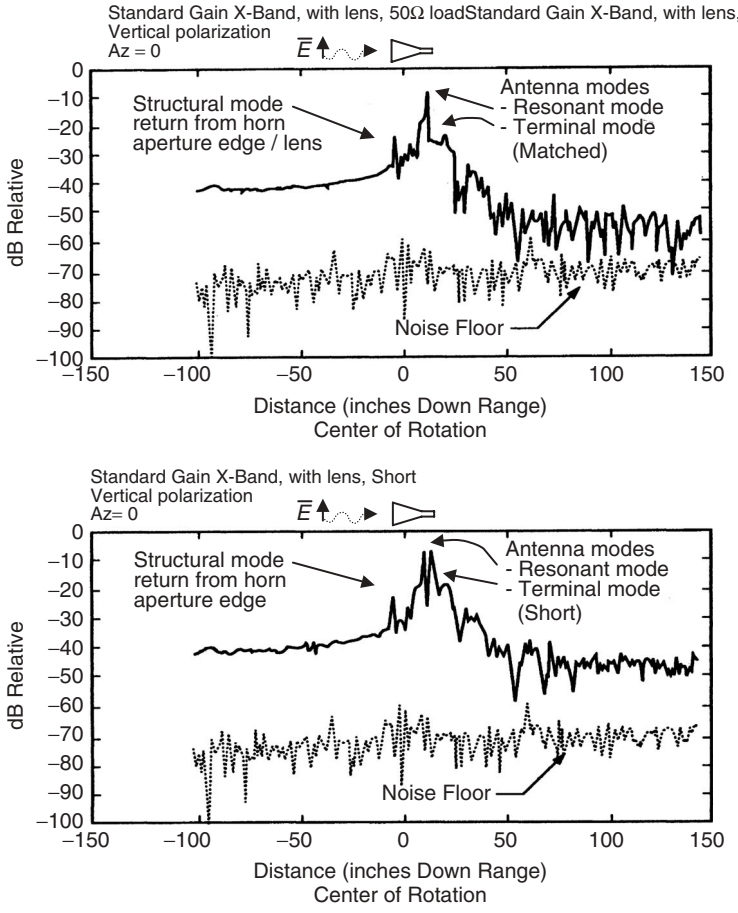


Figure 21.8 Band-limited impulse scattering response as a function of range for a standard gain horn. Frequency band is 6.27–12.32 GHz for vertical E-field polarization, with azimuth looking right into the horn throat. The graphic at the top of each graph indicates the geometry. As shown with the antenna shorted, the antenna mode consists of two almost equal amplitude returns—one from the antenna terminal mode reflecting 50% of the energy and one from the antenna resonant mode reflecting 50% of the energy. (From Ref. 12.)

Even though the equations developed in this section were derived for simple, small element antennas to relate gain to RCS, they are very powerful in the development of antennas and their understanding. Equation (21.18) can be used to understand the boundaries of an RCS versus antenna design and the system impacts. In practical designs, these equations are too simple to fully model antennas when the complexity of the antennas increases. Effects such as RCS impacts from cross-polarization effects coupling other antenna modes into RCS returns, interaction of arrays, impedance mismatches in load impedance down the transmission line, and in some cases continuous current aperture antennas require full electromagnetic modeling of the radiating structure to fully optimize the antenna/RCS trade.

21.6 UNIFORM TOTAL ANTENNA RADIATION AND SCATTERING REPRESENTATION

Most antenna radiation and scattering representations are formulated separately. In the following we give a uniform representation where both radiation and scattering are given together as a single formulation as shown in Figure 21.9. The formulation clearly shows the relationship between the antenna match, s_{11}, the receive and transmit patterns, the antenna termination load reflection coefficient, Γ, and the terminal scattering mode and the residual mode (structural scattering mode plus antenna resonant mode scattering).

Define the Green's function $G(R)$ in Eq. (21.19) and the far-field scattered field in Eq. (21.20). The far-field scattered field on the left side of the equation is independent of R due to the limit process. The far field is transverse to the radial vector and depends on the angles and the incident polarization.

$$G(R) \equiv \frac{e^{-jkR}}{4\pi R} \tag{21.19}$$

$$\begin{bmatrix} E_H^{sf} \\ E_V^{sf} \end{bmatrix} \equiv \lim_{R\to\infty} \frac{1}{\sqrt{4\pi}G(R)} \begin{bmatrix} E_H^s(R,\theta,\phi) \\ E_V^s(R,\theta,\phi) \end{bmatrix} \tag{21.20}$$

With these definitions, the linear relationship between the incident and scattered fields is given in Eq. (21.21). The notation used is as follows:

s_{11} = antenna voltage reflection coefficient
g_H, g_V = antenna H and V polarization gains
λ = wavelength,
E_H^i, E_V^i = incident plane waves at the origin for H and V polarizations
a_1, b_1 = incident and reflected voltage traveling waves for the antenna port
ϕ_{H1}, ϕ_{V1} = H and V antenna pattern phases

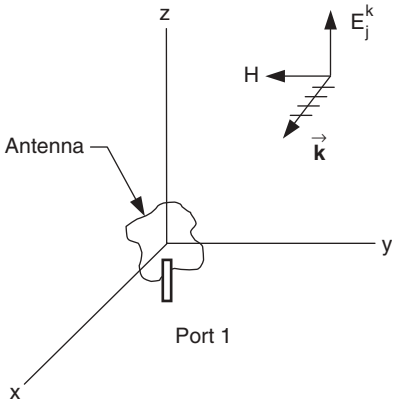


Figure 21.9 The geometry and conditions used in the derivation of the complete scattering and radiation conditions.

$\sigma^s_{HH}, \sigma^s_{HV}, \sigma^s_{VH}, \sigma^s_{VV}$ = residual mode scattering radar cross sections
$\phi^s_{HH}, \phi^s_{HV}, \phi^s_{VH}, \phi^s_{VV}$ = residual mode scattering voltage phases

$$\begin{bmatrix} b_1 \\ E^{sf}_H \\ E^{sf}_V \end{bmatrix} = \begin{bmatrix} s_{11} & \lambda\sqrt{\frac{g_H}{4\pi}}e^{j\phi_{H1}} & \lambda\sqrt{\frac{g_V}{4\pi}}e^{j\phi_{V1}} \\ \sqrt{g_H}e^{j\phi_{H1}} & \sqrt{\sigma^s_{HH}}e^{j\phi_{HH}} & \sqrt{\sigma^s_{HV}}e^{j\phi_{HV}} \\ \sqrt{g_V}e^{j\phi_{V1}} & \sqrt{\sigma^s_{VH}}e^{j\phi_{HV}} & \sqrt{\sigma^s_{VV}}e^{j\phi_{VV}} \end{bmatrix} \begin{bmatrix} a_1 \\ E^i_H \\ E^i_V \end{bmatrix} \tag{21.21}$$

The matrix in Eq. (21.21) is not symmetric even though the scattering is assumed reciprocal. This is because the dependent and independent variables are of different units and were selected to represent the antenna radiation and scattering. The formulation can easily be generalized to include nonreciprocal materials by distinguishing transmit and receive antenna gains and phases. The formulation is for monostatic scattering but is easily generalized to bistatic scattering.

The matrix has all elements of the antenna and scattering. The matrix is partitioned in Eq. (21.22) to show these components.

$$M_3 \equiv \left[\begin{array}{c|cc} s_{11} & \lambda\sqrt{\frac{g_H}{4\pi}}e^{j\phi_H} & \lambda\sqrt{\frac{g_V}{4\pi}}e^{j\phi_V} \\ \hline \sqrt{g_H}e^{j\phi_H} & \sqrt{\sigma^s_{HH}}e^{j\phi_{HH}} & \sqrt{\sigma^s_{HV}}e^{j\phi_{HV}} \\ \sqrt{g_V}e^{j\phi_V} & \sqrt{\sigma^s_{VH}}e^{j\phi_{VH}} & \sqrt{\sigma^s_{VV}}e^{j\phi_{VV}} \end{array}\right] \tag{21.22}$$

The upper left element is the antenna reflection coefficient. The lower left is the vector of transmit voltage patterns for the H and V polarizations. The upper right element is the vector of receive pattern length (square root of the effective aperture area). The bottom right matrix represents the residual mode (structural mode and antenna resonant mode) scattering.

When the antenna is terminated with a load, its linear relationship is given in terms of the load reflection coefficient in Eq. (21.23).

$$a_1 = \Gamma b_1 \tag{21.23}$$

$$\begin{bmatrix} E^{sf}_H \\ E^{sf}_V \end{bmatrix} = \left\{ \Gamma(1-\Gamma s_{11})^{-1} \begin{bmatrix} \sqrt{g_H}e^{j\phi_{H1}} \\ \sqrt{g_V}e^{j\phi_{V1}} \end{bmatrix} \begin{bmatrix} \lambda\sqrt{\frac{g_H}{4\pi}}e^{j\phi_{H1}} \\ \lambda\sqrt{\frac{g_V}{4\pi}}e^{j\phi_{V1}} \end{bmatrix}^T + \begin{bmatrix} \sqrt{\sigma^s_{HH}}e^{j\phi_{HH}} & \sqrt{\sigma^s_{HV}}e^{j\phi_{HV}} \\ \sqrt{\sigma^s_{VH}}e^{j\phi_{VH}} & \sqrt{\sigma^s_{VV}}e^{j\phi_{VV}} \end{bmatrix} \right\} \begin{bmatrix} E^i_H \\ E^i_V \end{bmatrix} \tag{21.24}$$

$$M_2 \equiv \left\{ \Gamma(1-\Gamma s_{11})^{-1} \begin{bmatrix} \sqrt{g_H} e^{j\phi_{H1}} \\ \sqrt{g_V} e^{j\phi_{V1}} \end{bmatrix} \begin{bmatrix} \lambda\sqrt{\dfrac{g_H}{4\pi}} e^{j\phi_{H1}} \\ \lambda\sqrt{\dfrac{g_V}{4\pi}} e^{j\phi_{V1}} \end{bmatrix}^T + \begin{bmatrix} \sqrt{\sigma^s_{HH}} e^{j\phi_{HH}} & \sqrt{\sigma^s_{HV}} e^{j\phi_{HV}} \\ \sqrt{\sigma^s_{VH}} e^{j\phi_{VH}} & \sqrt{\sigma^s_{VV}} e^{j\phi_{VV}} \end{bmatrix} \right\} \tag{21.25}$$

The antenna RCS is then given using the linear relation in Eq. (21.24).

$$\begin{bmatrix} \sigma_{HH} & \sigma_{HV} \\ \sigma_{VH} & \sigma_{VV} \end{bmatrix} = \begin{bmatrix} \dfrac{|E^{sf}_{HH}|^2}{|E^i_H|^2} & \dfrac{|E^{sf}_{HV}|^2}{|E^i_V|^2} \\ \dfrac{|E^{sf}_{VH}|^2}{|E^i_H|^2} & \dfrac{|E^{sf}_{VV}|^2}{|E^i_V|^2} \end{bmatrix} \tag{21.26}$$

Each RCS element has the following form: a square of the magnitude of the sum of the antenna terminal mode and the residual complex components. Equation (21.24) also shows that the antenna gain patterns and the antenna match can be measured on an RCS range capable of measuring phase. The antenna patterns and match can be determined by measuring the antenna scattering when the antenna is terminated with a matched load, short, and an open and then using Eq. (21.24) to solve for the desired antenna parameters. Appel-Hansen used a simplified version of this concept to generate the RCS of a horn using antenna measurements as shown in Figure 21.6.

The formulation has been kept simple to illustrate the factors. The representation can readily be generalized to include multiple antenna ports, multiple incidence angles, and multiple scattering angles. The representation is basically of the same form as Eqs. (21.21) and (21.22). The partitioned square s_{11} matrix is the scattering matrix for the ports. The other partitioned matrices are not necessarily square. The top right and bottom left matrices are, respectively, the port received patterns and the transmit patterns. The remaining partitioned matrix relates the incident waves to the received waves when all of the ports are matched. Since the incident angles and received angles may differ, the multistatic problem is represented. A difference is that the load and antennas, and the other matrices do not commute, resulting in a slightly different form for the representation corresponding to M_2.

Another representation of antenna scattering results from Eq. (21.24). As derived from the antenna/scattering matrix, a term for the terminal mode, Γ, is shown and a term, A, for the residual (antenna structural mode + antenna resonant mode) can be introduced. Equation (21.18) can be rewritten

$$\sigma = \frac{G^2\lambda^2}{4\pi}|\Gamma - A|^2 \tag{21.27a}$$

where Γ and A are complex numbers. Since $|\Gamma| \leq 1$, if $|A| > 1$ then $|\Gamma - A|$ will never be a small number and the scattering from the antenna will remain large. For a half-wave dipole, $A \sim 1$ and substituting in the impedance concepts from Section 21.5, we obtain

$$\sigma = \frac{G^2\lambda^2}{4\pi}\left|\frac{Z_L - Z_A}{Z_L + Z_A} - 1\right|^2 \tag{21.27b}$$

The uniform antenna and scattering representation in Eqs. (21.21) through (21.26) clearly illuminate the previous simpler Eq. (21.18). The simpler representation of Green does not consider the interaction between the antenna and load match. Assuming a perfectly matched antenna ($s_{11} = 0$) and equal transmit and receive antennas, the residual RCS can be reconciled in the uniform and the simpler representations. If the antenna is not matched then the "A" term in the simpler representation must incorporate parts of the antenna and load match and the residual components.

Example 2 Equation (21.27a) can be used to develop a closed form solution for a resonant antenna with its terminal shorted. In practice, this can be a special antenna known as "chaff." Chaff is an antenna that is released into the air and is used to confuse radar systems by extracting energy from a radar pulse and then, since the antenna is shorted at the terminal, all of the energy is reradiated, increasing the clutter and helping to hide objects. For chaff, the terminal is shorted such that $Z_L = 0$ and for this example the chaff is assumed to be one-half wavelength and $A \sim 1$. Equation (21.27a) becomes

$$\sigma = \frac{\lambda^2}{4\pi} \cdot (1.643)^2 \cdot 4$$

$$\sigma = \frac{\lambda^2}{\pi} \cdot (1.643)^2$$

$$\sigma = 0.8593\ \lambda^2 \quad \text{or} \quad \sigma/\lambda^2 \sim 0.86$$

Calculate the scattering from half-wavelength chaff at 4.02 GHz. $Z_L = 0$

$$\lambda = \frac{3 \times 10^8}{4.02 \times 10^9} = 0.07463 \text{ m}$$

such that

$$\sigma = 0.8593\ (0.07463)^2 = 4.786 \times 10^{-3}\ \text{m}^2$$

$$\sigma = 10 \log_{10}(4.786 \times 10^{-3}) = 10\ (-2.32)$$

$$\sigma = -23.2 \text{ dBsm}$$

Figure 21.10 is the measurement of a shorted half-wavelength dipole that matches the example for a shorted dipole antenna that could be used as chaff around 4 GHz.

Another example of antenna scattering is shown in Figure 21.11. Figure 21.11 shows scattering measurements of a horn antenna that illustrate the complexity of antenna scattering beyond a small dipole. These examples continue to show the scattering reductions that are achieved by matching the antenna (antenna terminal mode = 0) and the effects of a short circuit terminal such that all of the received energy is reradiated.

The "*A*" term from Eq. (21.27a) varies with angle and its magnitude can be compared to the terminal/resonant mode with respect to angle. On axis (incident angle 0°), where the antenna has substantial gain, the terminal mode and resonant mode dominate. As one moves in aspect, the gain of the antenna drops while the structural aspects of the antenna become much more visible (side of the horn). The structural mode does not change with terminal impedance and, from Figure 21.11, it is apparent that beyond ±15°, the structural mode component of this antenna's RCS is dominant. This also reinforces the

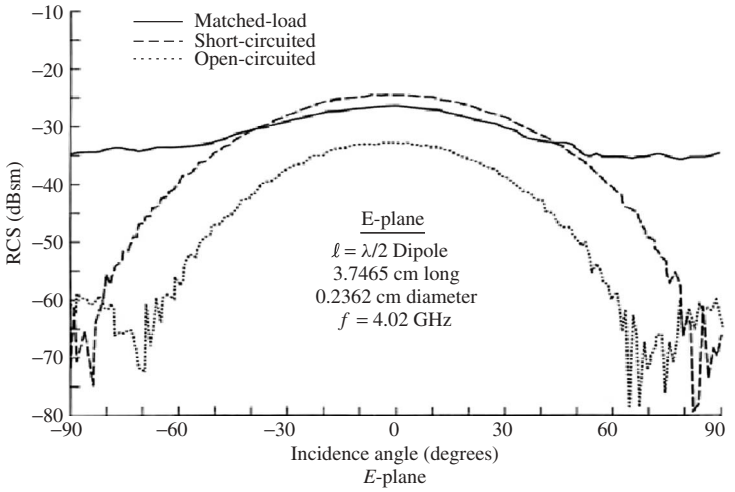


Figure 21.10 Measured scattering from a half-wave dipole that matches the conditions of Example 2. For the shorted condition, the scattering has a ∼−23- to 24-dBsm radar cross section. (From Ref. 13.)

concept that both the terminal and resonant modes are proportional to gain. As one moves off axis, the horn antenna's gain pattern decreases and the magnitudes of the terminal and resonant modes similarly decrease. The A term can have significant variations depending on the structural scattering of the antenna and is a function of all the parameters that can vary with respect to RCS. In this example, the variation with incident angle for A is clear.

Another characteristic exhibited by the horn example and contrasted with Example 1—the small, very thin, dipole antenna—is that the antennas behave differently with their terminals open. For the small, very thin dipole, leaving the terminal open resulted in an antenna with very low scattering since the resultant antenna had very little structural scattering and very small antenna gain (antenna would operate as two antennas at twice the frequency but at the frequency analyzed it does not resonate (low gain)). The horn measurement from Figure 21.11 illustrates that leaving the terminals open for the horn does not necessarily result in a low scattering antenna. In the case of the horn, even with the terminal open, captures significant energy and reradiates it as scattering. The behavior of the small, very thin dipole with the open terminal creating a minimum scattering condition is rare and should be considered the exception.

21.7 STRUCTURAL ANTENNA CONTRIBUTIONS

With much of an antenna's RCS return caused by structural or non-antenna-related mechanisms (especially antenna elements outside their operational band), many techniques can be applied to understand and reduce the antenna's structural RCS. As described in earlier sections, structural RCS is not proportional to the antenna gain so the antenna's structural RCS can be modified significantly with little or no impact on antenna performance. Figure 21.12 illustrates some of the many ways in which structural scattering

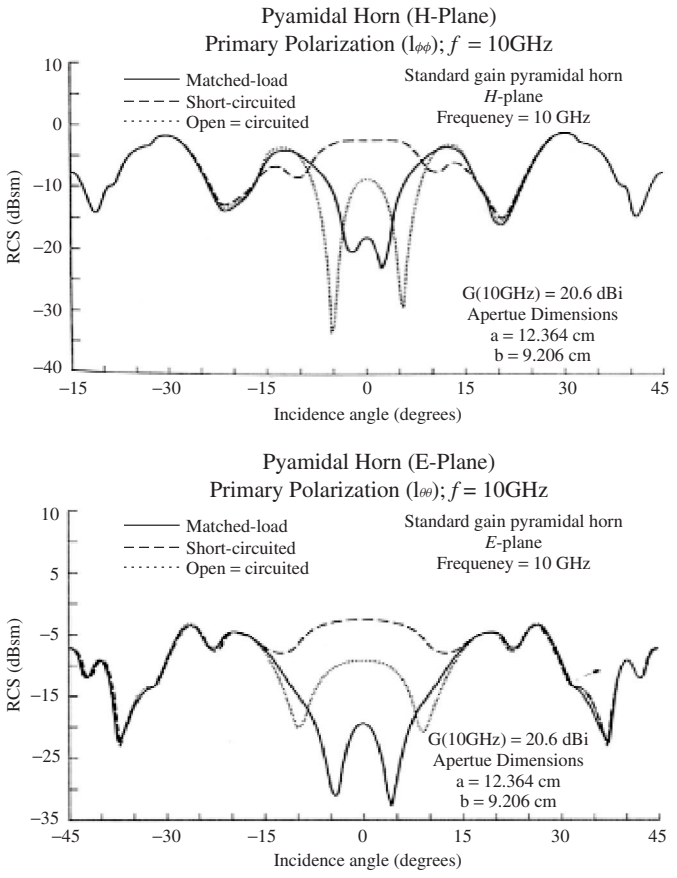


Figure 21.11 RCS azimuth response of a pyramidal horn with the horn terminated in a matched load, short circuit, and open circuit. The antenna related modes (terminal and resonant) dominate on boresight (0° incident angle) and the structural mode scattering dominates beyond ±15° incident angle.

RCS can be generated. From a systems input, we will consider the antenna to contain not only the radiation structure but also any of the adjacent structure that would be required for strength, installation, or attachment. Typically, one would want to reduce structural scattering to a minimum to achieve low RCS antenna designs.

As with most scattering systems, the primary scattering control mechanisms include shaping, minimizing discontinuities, using materials to absorb energy, and energy cancellation. Figure 21.13 illustrates the primary methods of structural scattering reduction or control, and we have conceptually seen how structural scattering can be reduced with minimal loss of antenna performance by substituting the sleeve monopole blade antenna with a conformal annular slot antenna in Figure 21.3.

21.7.1 Shaping

Shaping makes use of the concept that there may be volume or sectors in space in which you would want to control your RCS at the expense of other sectors. For example, on a

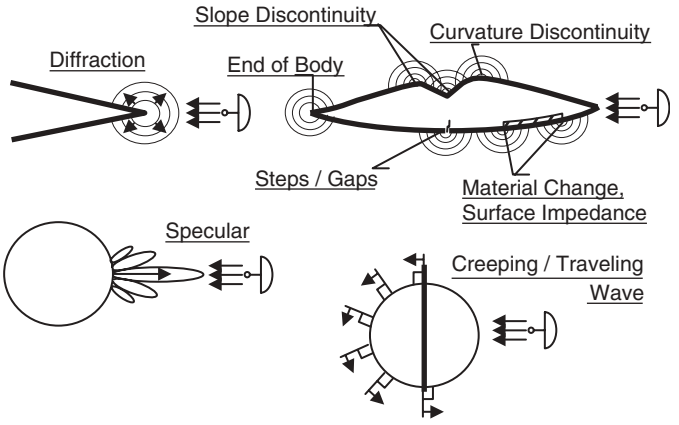


Figure 21.12 Typical scattering mechanisms. Antenna RCS structural mode control requires controlling these types of scattering phenomena. (From Refs. 1 and 2.)

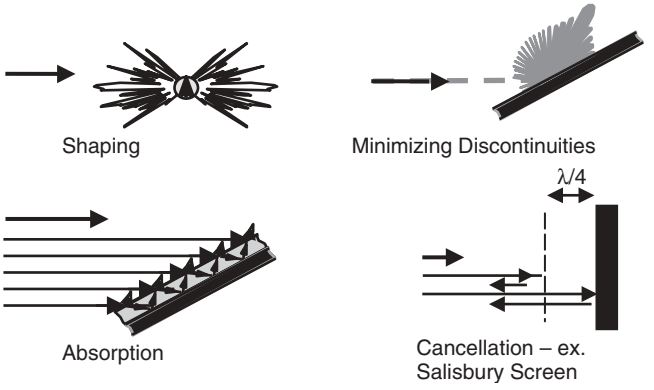


Figure 21.13 Primary scattering control mechanisms for any scatterer. (From Ref. 1.)

dish antenna you may wish to shape the feed struts to have a reduced scattering shape for antenna performance. An antenna performance (instead of RCS) reason to shape the struts may be to reduce retroreflections back into the dish and thus reduce antenna sidelobe structure and antenna "noise."

Since shaping does not include energy absorption, the total energy scattered for a given intercept shape will be conserved. In contrast, the integral over 4π steradians of the monostatic RCS is not conserved. Figure 21.14 shows a number of shapes and their RCS values (physical optics approximation) when illuminated from the same direction. In the case of the ogive, sphere, and flat plate, the RCS toward the observer can vary greatly given an equivalent intercept area and the example illustrates how shaping can be used to control the RCS.

Example 3 Using four of the shapes in Figure 21.14 (sphere, flat plate, diagonal flat plate, ogive) each having an equivalent incident intercept area (measured at 0°, normal incidence, direction $\vec{\mathbf{k}}$ as shown in Figure 21.14), we are able to see the wide variation

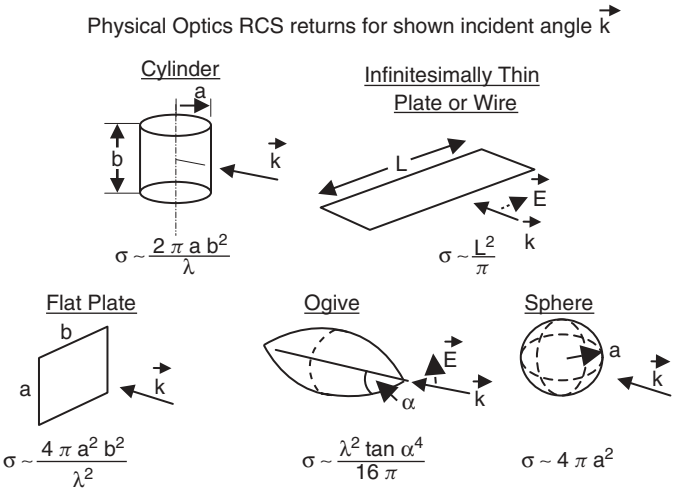


Figure 21.14 Closed form physical optics RCS approximations demonstrating the range of monostatic RCS returns for given shapes illuminated at "normal" incidence (shown by $\vec{\mathbf{k}}$). For a given physical intercept area, the RCS can vary considerably due to shaping. (From Ref. 3.)

in RCS due to shaping. The intercept area for each shape will be 0.7854 m^2 and the frequency is 3 GHz.

$$\lambda = \frac{300 \times 10^6}{3000 \times 10^6} = 0.1 \text{ m}$$

Sphere: A sphere with a projected intercept area of 0.7854 m^2

$$\rightarrow \pi a^2 = 0.7854 \rightarrow a = 0.5 \text{ m}$$
$$\rightarrow \sigma = \pi a^2 = \pi(0.5)^2 = 0.7854 \rightarrow 10 \log_{10}(0.7854) = -1.05 \text{ dBsm}$$

Flat Plate: A square flat plate (edges $a = b$) with a projected intercept area of 0.7854 m^2

$$\rightarrow 0.7854 \text{ m}^2 = \text{a}^2; \quad a = 0.886 \text{ m}$$
$$\rightarrow \sigma = \frac{4\pi a^2 b^2}{\lambda^2} = \frac{4\pi a^4}{\lambda^2} = \frac{4\pi(0.886)^4}{0.1^2} = 774.36 \text{ m}^2 = 10\log_{10}(774.36)$$
$$= 28.9 \text{ dBsm}$$

Diagonal Flat Plate: Has the same projected area as the flat plate.

Ogive: Dimensions, diameter = 1 m, length = 3.7 m and an included half-angle, $\alpha = 30.1^\circ$

$$\rightarrow \sigma \text{(from Ref. 4)} \sim -39 \text{ dBsm}$$

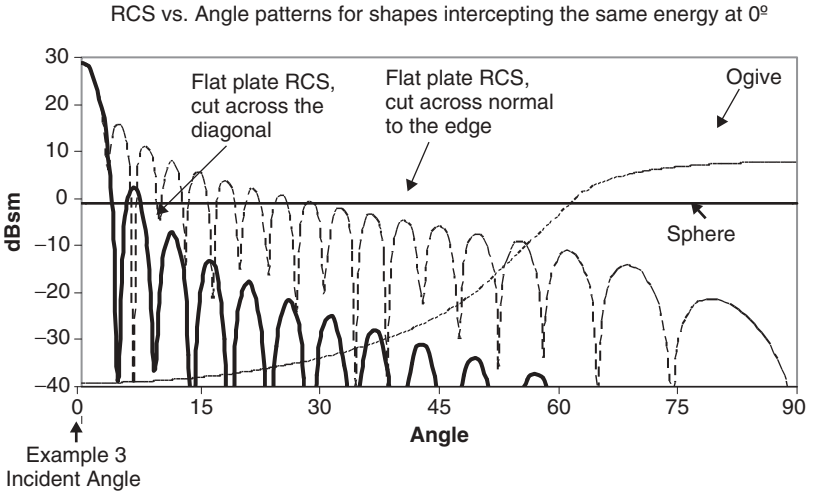


Figure 21.15 RCS returns for a number of shapes having equivalent intercept areas. (From Refs. 3 and 14.)

Each shape at 0° incident angle intercepts the same power (watts/m^2) with an intercept area defined as $=> 0.7854$ m^2. From Figure 21.15 and looking at incident angle 0°, we see that the plates have the highest return of 28.9 dBsm, the sphere has a return of -1 dBsm, and the ogive has a return of -39 dBsm. This example illustrates how shaping causes large variations in RCS. For this example, the data at 0° compares the RCS for shapes having the same intercept area but the RCS varies by almost 70 dB.

Figure 21.15 also provides an example of how shaping can impact RCS as a function of incident angle. Certainly when the incident angle changes, the four shapes in Example 3 each will have different intercept areas (except for the two flat plates) but even near normal, where their intercept areas are similar, each shape has very distinct angular scattering characteristics. The sphere is a constant at any angle. The two flat plates have identical specular (0° incident angle) RCS values of 28.9 dBsm but their angular behavior is quiet different. The flat plate edge-on has a sidelobe structure that rolls off with a $((\sin x)/x)^2$ behavior with the first sidelobe $\sim$13 dB down, the second $\sim$17 dB down, and so on. With the plate just rotated 45° such that the angular plane cuts across an apex instead of an edge, the sidelobe structure rolls off with a $((\sin x)/x)^4$ behavior with the first sidelobe at $\sim$26 dB down and the second sidelobe $\sim$34 dB down. The ogive with a pointed tip aligned at 0° has the lowest RCS on axis but as you move in angle, the ogive's signature increases rapidly. These four examples illustrate the large swings in RCS that are possible with shaping.

For antennas, shaping is a primary method for controlling RCS. For a typical airborne antenna in which we may wish to reduce the RCS in a sector, we may decide to shape the scattering mechanisms such that they have minimum scattering in a particular direction. If the radiating structures are also shaped and made conformal to the platform's surface, the antenna pattern may be impacted since the projected aperture in the RCS reduced direction could also lead to a reduced antenna aperture, thus reducing the antenna's gain. Figure 21.3 illustrates this concept. In Figure 21.3, we have designed the antenna radiating slots and attachment to be conformal to the surface of the platform. This minimizes

the structural scattering that would occur from an equivalent antenna blade that has an effective flat plate scatter as well as the corner reflector that is created by the antenna blade being at a right angle to the surface of the platform.

21.7.2 Minimizing Discontinuities

Once the antenna has been optimally shaped consistent with the vehicle signature requirements, any discontinuities need to be understood and minimized. Discontinuities in this context could be remnants of antenna resonating structure that are out of band and now are considered structural scattering elements. Figure 21.16 illustrates two common discontinuity geometries and approximate maximum signature returns from those discontinuities. Equations (21.27c) and (21.28) (see Figure 21.16) are for perfect electrical conductor (PEC) conditions and are equivalent to scattering from a flat plate.

$$\sigma \sim 4\pi(Lh)^2/\lambda^2 \tag{21.27c}$$

$$\sigma \sim 4\pi(LW)^2/\lambda^2 \tag{21.28}$$

The angular dependence for both nonlinear discontinuities and linear discontinuities is shown in Figure 21.17. The effective length L_e determines the maximum return when used with Eqs. (21.27c) and (21.28). The RCS return for nonlinear discontinuities will have a varying specular return based on the effective length, L_e, and that specular RCS return will occur over the region in which L_e has a surface normal. For a linear discontinuity, the RCS return falls off with a characteristic $((\sin x)/x)^2$ function and is similar to the roll-offs shown in Figure 21.17.

$$L_e = \sqrt{\lambda R/2} \tag{21.29}$$

$$f(\theta) = ((\sin N)/N)^2 \cos^2\theta \tag{21.30}$$

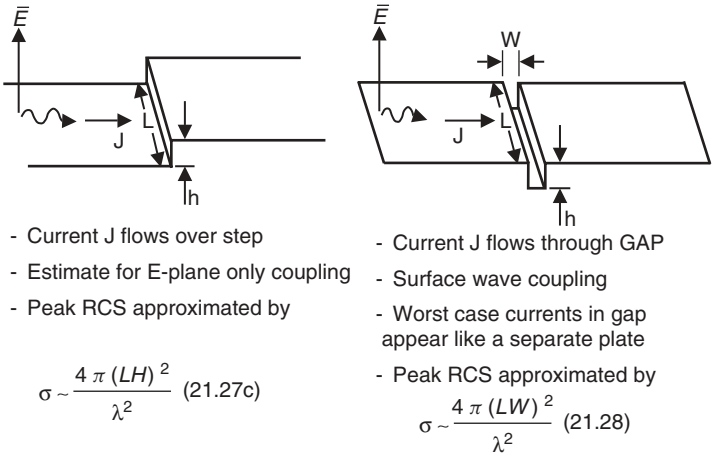


Figure 21.16 Physical optics approximation of gaps and step RCS returns. Assumption is $L \gg \lambda$. Scattering approximates a flat plate return, that is, $\sigma = 4\pi(\text{Area})^2/\lambda^2$. Many antenna elements out of band exhibit this type of structural scattering. (From Ref. 15.)

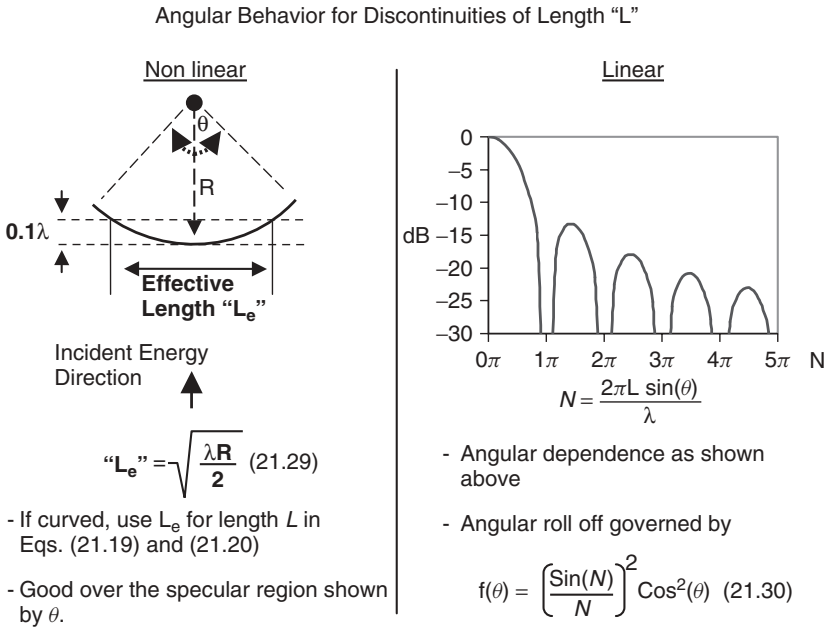


Figure 21.17 RCS versus angle behavior derived using a physical optics approximation for both a nonlinear (or curved) discontinuity and a linear discontinuity. The nonlinear or curved discontinuity will scatter with a magnitude similar to a linear scatter with an effective length of L_e in the direction perpendicular to the surface normal.

Certainly if all of the discontinuities can be faired in or covered over, one can minimize the contribution from these scattering sources. A number of the concepts illustrate the use of a PEC material covering or filling the gap. These are good techniques, but unfortunately, the nature of antennas is that not all of the slots can be covered or filled with PEC since the slot will no longer be functional as a radiator.

21.7.3 Absorbers

The nature of antennas is that you cannot fill all of the slots or gaps with a PEC to reduce discontinuities since many of these gaps are radiating elements within bands of operation. Often, absorbing materials are used to reduce the scattering from these discontinuities. An example of such a material is magnetic RAM used as a surface wave absorber. This technique is used to reduce the scattered energy from a discontinuity and, by reducing the scattering, the resultant RCS is reduced.

Figure 21.18 illustrates the geometry for scattering reduction of a radiation slot in a ground plane using RAM to attenuate energy before it scatters from the discontinuity. Using the geometry in Figure 21.18, Strattan [16] solved two transcendental equations with assumptions that $\varepsilon_1 > 15$ and $\mu_1 > 1$ with a resultant $\varepsilon_1\mu_1 > 20$ to generate the propagation constant in the RAM:

$$\beta = \frac{2\pi}{\lambda_0}\left[1 + \frac{\mu_1}{\varepsilon_1}\tan^2\left(\sqrt{\mu_1\varepsilon_1}\frac{2\pi}{\lambda_0}d\right)\right]^{0.5} \tag{21.31}$$

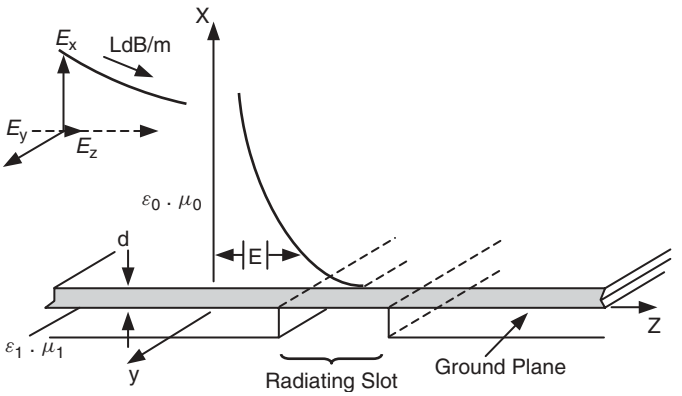


Figure 21.18 Geometry and interaction of an incident transverse magnetic (TM) field with radar absorbing material (RAM) over a ground plane and a radiating slot. (From Ref. 16.)

where

μ_1 = relative permeability
ε_1 = relative permittivity
d = thickness of the material
λ_0 = free-space wavelength at frequency of interest

and the lossy part of Eq. (21.31)

$$\alpha = -\text{Im}(\beta) \tag{21.32}$$

leading to a loss factor L of

$$L = 8.69\alpha \text{ dB/m} \tag{21.33}$$

For the RAM coating to operate properly, one would want the surface to have inductive impedance. Surface impedances that are inductive guide the Poynting vector into the RAM, allowing the material to attenuate the energy. The easiest way to create an inductive surface impedance is to ensure that the material is less than a quarter-wavelength in height with $\frac{1}{4}\lambda_1$ referenced as the wavelength within the material. If the surface is not inductive (it would be capacitive), then the energy is shed away from the surface. (see Figure 21.19.)

To meet the criteria of remaining inductive, the surface wave absorber often has a tapered thickness with the thinnest part of the RAM away from the scattering mechanism. The thinner section of the taper is designed to inductively couple the highest frequency surface energy into the RAM. The RAM tapers to thicker levels to further attenuate lower frequency RCS (having a larger wavelength).

Using lossy material to reduce structural scattering will impact antenna performance. Figure 21.18 illustrates a geometry in which a lossy RAM is used over a radiating slot. Deriving the impact of an arbitrary material in front of an antenna is developed using Figure 21.20.

Using the geometry in Figure 21.20, a set of Fresnel coefficients are generated along with Snell's law for complex media; and finally a transmission coefficient from Region

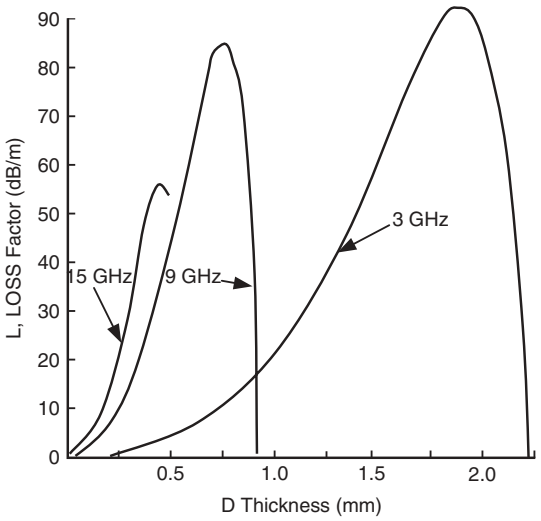


Figure 21.19 Surface wave attenuation curves for Rockwell NR-95. Attenuation levels are very large for thin layers of material since the geometry allows the surface material to extend from the scattering mechanism for some distance. (From Ref. 16.)

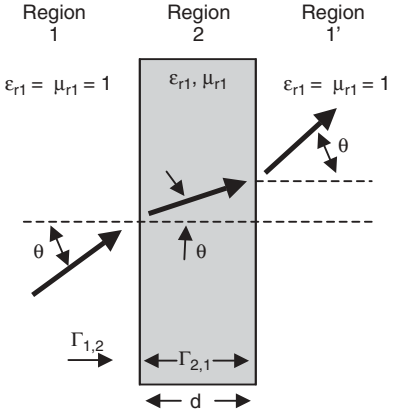


Figure 21.20 The geometry for an arbitrary material (complex μ,ε) over a radiating element.

1 through Region 2 back into Region 1′ can be generated for use with antennas in arbitrary matter. To generate the transmission coefficient, one must first generate the Fresnel coefficients for parallel and perpendicular polarizations incident on the geometry shown in Figure 21.20:

$$_{\parallel}\Gamma_{1,2} = \frac{\sqrt{\dfrac{\mu_{2r}}{\varepsilon_{2r}}}\cos\theta_2 - \sqrt{\dfrac{\mu_{1r}}{\varepsilon_{1r}}}\cos\theta_1}{\sqrt{\dfrac{\mu_{2r}}{\varepsilon_{2r}}}\cos\theta_2 - \sqrt{\dfrac{\mu_{1r}}{\varepsilon_{1r}}}\cos\theta_1} \tag{21.34}$$

$$\perp\Gamma_{1,2} = \frac{\sqrt{\frac{\mu_{2r}}{\varepsilon_{2r}}}\cos\theta_1 - \sqrt{\frac{\mu_{1r}}{\varepsilon_{1r}}}\cos\theta_2}{\sqrt{\frac{\mu_{2r}}{\varepsilon_{2r}}}\cos\theta_1 - \sqrt{\frac{\mu_{1r}}{\varepsilon_{1r}}}\cos\theta_2} \tag{21.35}$$

with Snell's law as

$$\sin\theta_2 = \frac{\sqrt{\mu_{1r}\varepsilon_{1r}}}{\sqrt{\mu_{2r}\varepsilon_{r2}}}\sin\theta_1 \tag{21.36}$$

And finally the transmission coefficient, T, can be generated when all of the internal reflections are accounted for and using the geometric series we obtain

$$\parallel/\perp T = \frac{(1 - \parallel/\perp\Gamma_{1,2}^2)e^{-jk_2 d\cos\theta_2}}{1 - \perp/\parallel\Gamma_{1,2}^2 e^{-j2k_2 d\cos\theta_2}} \tag{21.37}$$

Shown in Figure 21.21 is a typical example of the transmission loss that occur when using magnetic RAM over a slot or in front of a radiating structure.

There are many materials that can be used to control radiation properties of antennas for both antenna performance and RCS. Some (certainly there are many more materials than listed here) of the materials that may be used include:

- Alternate magnetic materials
- Dielectric RAM
- Volumetric materials
- Salisbury screens

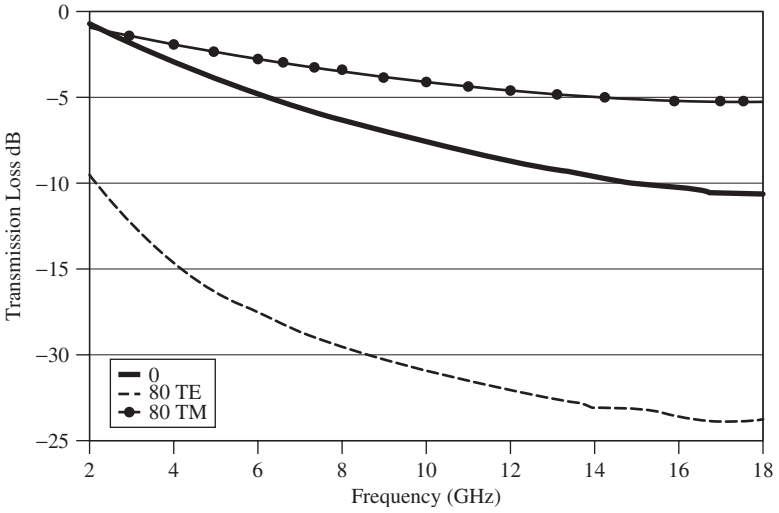


Figure 21.21 Typical loss experienced when radiating through a magnetic RAM. These curves were generated using a RAM available from Steward and using Eq. (21.37). The Steward material, a FeSi product from their November 1995 catalogue, is 40 mils thick and has permittivity and permeability as shown in Table 21.3.

- Jaumann absorbers
- Circuit analog absorbers

Each of these materials has its own unique characteristics and can be used to control fields around antennas. For this chapter, the case of magnetic RAM was developed as an example. To understand and model the effects of the many other scattering controlling materials, once again one is directed to Knott [1].

21.7.4 Cancellation

Cancellation techniques rely on the existence of multiple scattering mechanisms to be combined in such a way that the resultant scattering from the system is reduced over a select frequency band or angular aspect.

One of the simplifications used in deriving Eq. (21.18) for a simple radiation antenna element is that the terminal mode and the resonant mode were collocated. Since they are collocated, the scattering that occurs when the antenna is shorted ($\Gamma = -1$) is twice that of just the resonant mode. If an antenna can be constructed that has the antenna resonant mode separated from the antenna terminal mode, one can have these two scattering mechanisms interact to provide some reduction of the antenna's RCS. With a separation distance of d between the two scattering mechanisms, Eq. (21.18) can be rewritten

$$\sigma = \frac{G^2\lambda^2}{4\pi} \quad |\Gamma e^{-j2kd} - 1|^2 \tag{21.38}$$

with the phase factor $e^{-j2\beta d}$ being frequency dependent. In this situation, a short circuit at a quarter-wavelength distance can cause the summation of the antenna resonant mode and antenna terminal mode to provide some cancellation. One example of such an antenna may be a horn that transitions into its waveguide section gradually such that the antenna terminal position is set by the location of the coax to waveguide adapter and the antenna resonant mode would occur in the throat of the horn. This would be an example of cancellation occurring internal to the antenna with the resonant and terminal modes interacting.

Figure 21.22 illustrates the amplitude and phase deltas possible to achieve a certain level of cancellation. If over a set of frequencies or aspect angles the reflection from the terminal and the reflection from the antenna resonant modes are close enough in amplitude and close to 180° out of phase, then RCS reduction is possible.

TABLE 21.3 Steward FeSi RAM

Frequency	ε_{real}	ε_{imag}	μ_{real}	μ_{imag}
2 GHz	23.5	1.20	2.45	0.81
5 GHz	22.8	1.02	1.88	0.91
10 GHz	22.4	1.15	1.39	0.91
15 GHz	22.1	1.06	1.01	0.78

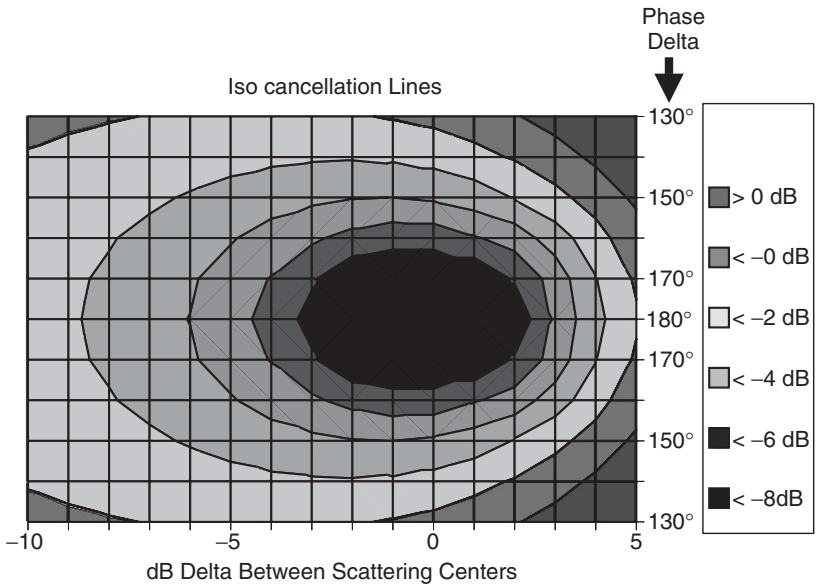


Figure 21.22 Isocancellation lines indicate the ability of two scattering centers to cancel each other. If the two signals are identical (0-dB delta or difference) and 180° apart, the cancellation is perfect.

Example 4 Given that an antenna has a resonant mode scattering RCS of +2 dBsm with phase of 35° and a structural mode scattering RCS of −3 dBsm with phase of 185° (terminal mode assumed zero), find the resultant new RCS when these two scattering contributions are combined.

The easiest way is to use Figure 21.22 and determine where the vector differences are located on the graph. Delta amplitude between the two scatterers is −5 dB (−3 dBsm −2 dBsm). Delta phase between the two scatterers is 150° (185° − 35°). From Figure 21.22, the delta −5 dB/150° point is found to be between the −4 and −6 dB cancellation lines. Applying the range of values from Figure 21.22 (−4 to −6 dB) to the original scatter, +2 dBsm, we find that the new RCS should be between −2 and −4 dBsm.

Mathematically the combined RCS can be calculated as follows:

$$\begin{aligned}
\textbf{Vector 1} \rightarrow \textbf{2 dB/35}^\circ \quad & \textbf{Vector 2} \rightarrow \textbf{−3 dB/185}^\circ \\
\text{Convert to linear} \rightarrow 10^{2/20}/35^\circ \quad & \rightarrow 10^{-3/20}/185^\circ \\
\text{Linear phasor} \rightarrow 1.26/35^\circ \quad & \rightarrow 0.708/185^\circ \\
\text{Real and imaginary} \rightarrow 1.26\cos(35^\circ) + j1.26\sin(35^\circ) & \\
\rightarrow 0.708\cos(185^\circ) + j0.708\sin(185^\circ) & \\
\text{Linear vector} \rightarrow 1.03 + j0.723 \quad & \rightarrow -0.705 - j0.062 \\
\text{Combine} \rightarrow 0.325 + j0.661 &
\end{aligned}$$

$$\text{Create phasor} \rightarrow \sqrt{[(0.325)^2 + (0.661)^2]}$$
$$\rightarrow \sqrt{[0.106 + 0.437]} = \sqrt{0.543} = 0.737$$
$$\text{Convert back to dBsm} \rightarrow 20\log_{10}(0.737)$$
$$\text{Resultant new RCS} \rightarrow -2.65 \text{ dBsm}$$

This compares favorably with using Figure 21.22, which indicated that the new RCS was between -4 dBsm $< \sigma < -2$ dBsm.

Another form of cancellation can occur external to the antenna as the complexity of the antenna increases. Equation (21.38) was qualified by being accurate for a simple element radiator, such as a dipole. As the complexity of an antenna increases, the summation of the antenna from a gain perspective versus an RCS perspective can be quite different due to the two times around path for RCS compared to the antenna geometry.

Figure 21.23 illustrates how the two-way path from multiple elements can be used for signature control. If the antenna is matched such that the antenna terminal mode is very small, then the primary scattering occurs from the antenna resonant mode for each slot (assuming that the structural mode is zero). In Figure 21.23 this is illustrated by two slot antennas. The two slot antennas are designed to sum to as large as possible on the horizon. This is achieved as shown in the graph of Figure 21.23, where the distance d between the slots along the direction of the gain pattern, is such that the slots sum on the horizon. The antenna pattern shows a peak around -50° from zenith. The drop in antenna gain near grazing incidence is caused by the slot element pattern falling off due to the aperture projected on the horizon shrinking. For the RCS geometry, the d has an effective length of $2d$ since RCS is a two-way or round-trip effect. This two-way effect can cause signature reduction if designed correctly. In some ways, this is similar to the phenomena we will discuss when we get to arrays.

Another method that is similar to cancellation was developed by Dr. Munk [17]. Dr. Munk determined that if designed correctly the structural scattering mechanisms and the antenna resonant mode can sum such that the RCS is reduced. Dr. Munk recognized that if one had an array of dipoles and they were illuminated from an external source, the resultant resonant mode scattering would occur both in the backscatter direction as well as in the forward scatter direction. Figure 21.24a illustrates this situation. If one then places that array in front of a ground plane, the forward scatter will be intercepted by

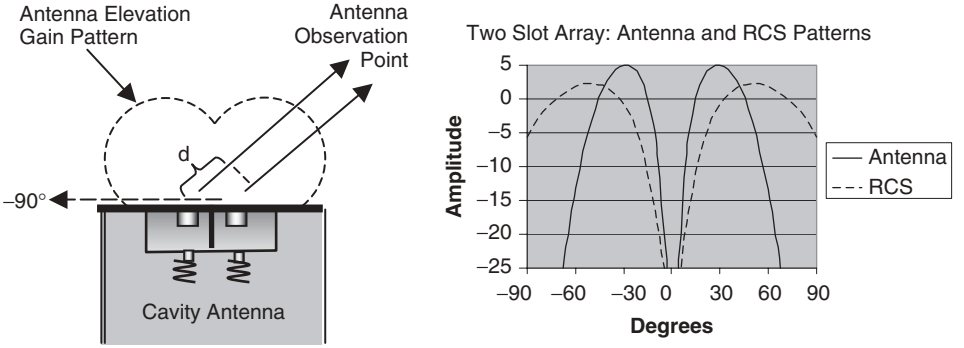


Figure 21.23 Cancellation from multiple elements can provide some RCS reduction.

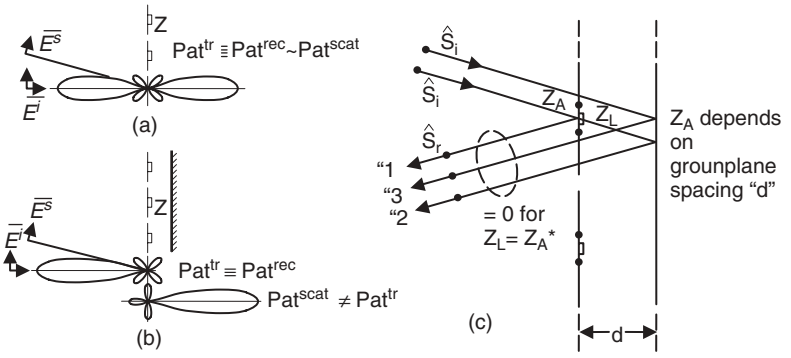


Figure 21.24 Array of dipoles in front of a ground plane allows the physics to have the antenna resonant mode and the structural mode to interact and with the correct impedance and spacing one can achieve low RCS. (From Ref. 17.)

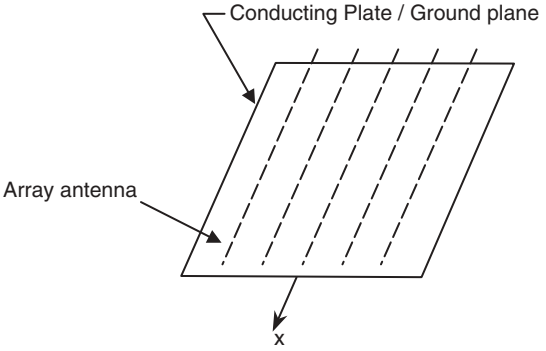


Figure 21.25 An array of dipoles in front of a conductive ground plane can be used to extend the concept in Figure 21.24 to a two-dimensional system.

the ground plane and be reflected back in the direction of the antenna mode scattering backscatter from the array of dipole antennas. Figure 21.24b,c illustrates that situation, and if the impedances and ground plane are designed correctly, one can reduce RCS.

Figure 21.25 illustrates how one would expand this technique to a two-dimensional array. To balance the scattering from the antenna and the ground plane, one would use Eq. (21.13) for the antenna's effective aperture and the equation for a flat plate shown in Figure 21.12 to correctly size the scattering from the ground plane:

$$\frac{4\pi A_e^2}{\lambda^2} = \frac{4\pi A^2}{\lambda^2} \tag{21.39}$$

where

A_e = effective area of receive antenna
λ = wavelength
A = area of the conduction plate = ab

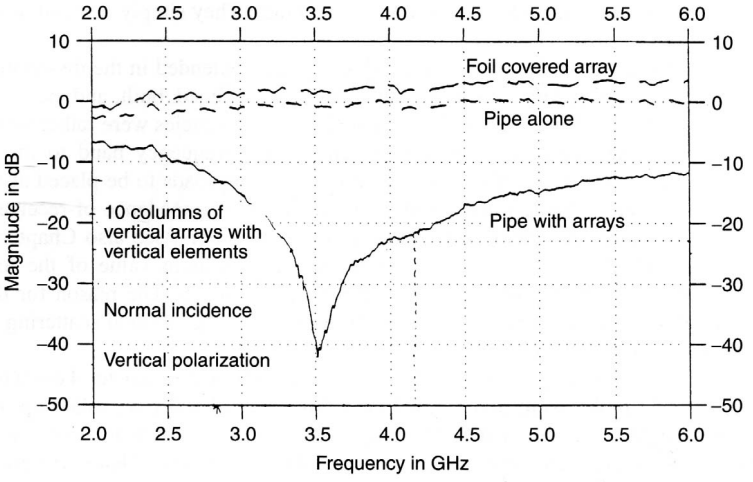


Figure 21.26 Measured backscatter from an array of dipoles in front of a circular ground plane. The three curves represent the circular ground plane (pipe), a foil covering the circular array of antennas and the circular gound plane, and the ground plane with a circular array or resistively loaded dipoles. (From Ref. 18.)

By balancing these two equations, we can satisfy the requirement to have the two scattering mechanisms with near equivalent RCS amplitudes. This concept can be extended to other geometries such as a circular ground plane with an array of resistively loaded dipoles in front of the circular ground plane. Figure 21.26 is a measurement of such a system containing an array of resistively loaded dipoles in front of a circular ground plane.

21.8 ARRAYS

Arrays have proved to be a solution for many antenna systems but they must be designed differently to control scattering. It is well known that a flat plate has a large reflected (also defined as forward scatter) scatter component and a relatively small backscatter (nonnormal incidence) component as shown in Figure 21.27 and in Figure 21.13. If possible, emulating a flat plate with a discrete set of element should have the same effect such that the antenna could have significant gain in a given direction along with a large forward scatter but in the backscatter (monostatic) direction, it would have a reduced backscatter.

To determine the spacing required to primarily have the energy reflect as forward scattered energy instead of backscattered energy, a small linear subset of a two-dimensional array is analyzed. Figure 21.28 models this geometry.

For simplicity, the elements that form the array in Figure 21.28 are assumed to have an isotropic or omniscattering pattern. This is a good assumption since most antenna elements in an array are electrically small and have a broad antenna pattern. If the scattering from the elemental antenna has significant directivity, then the results of this array analysis would need to be modified by the element scattering pattern.

Figure 21.27 A flat plate has a relatively low backscatter compared with its forward scattering. Arrays may have a large backscatter associated with "Bragg" lobes.

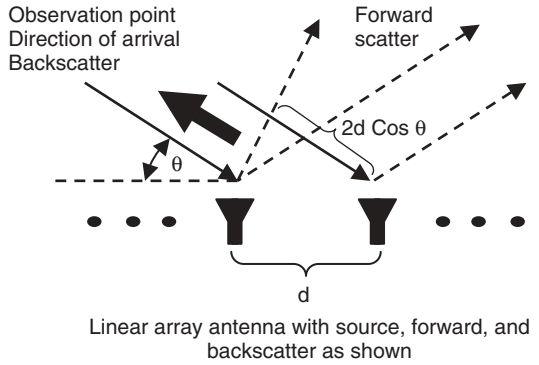


Figure 21.28 Scattering from an array of antennas. Forward scatter is determined by Snell's law and the summation of all the individual scatterers. Backscatter will be determined by how each element sums in the backscatter direction.

For the array to sum in the backscatter direction, all of the elements would need to add in phase in that direction. The term "in phase" for backscatter indicates that the elements are related by integer multiples of 360°. If this occurs, then a backscatter lobe will be generated. These lobes have been named "Bragg" lobes. From Figure 21.28 we see that the distance has a factor of 2 applied to account for the backscatter's two-way path and is adjusted for the angle of incidence (with respect to the plane of the array). From this, the Bragg lobe condition occurring for an array spacing of d is

$$d = \frac{n\lambda}{2\cos\theta} \tag{21.40}$$

where

d = array element-to-element spacing
λ = wavelength
θ = angle of incidence with respect to grazing
n = integer = 1, 2, 3 ...

To satisfy the maximum element-to-element spacing preventing Bragg lobes, we use $n = 1$ and the incident angle at grazing or $\theta = 0$.

$$d \leq \lambda/2 \tag{21.41}$$

This is a familiar expression in near-field sampling of antennas and is considered the Nyquist sample interval that is used to sample a continuous aperture to generate the aperture's far field. A more general form of the equation given an element-to-element spacing and a desire to work the problem with respect to which aspect incident angles (not just grazing) and frequency combinations will crate a Bragg lobe, we obtain

$$f_{\text{GHz}} = \frac{0.15}{d \cos\theta} \tag{21.42}$$

where f_{GHz} is frequency in gigahertz (GHz) and d is in meters.

In this manner, arrays can be used to approximate continuous surfaces and the artifacts of scattering from arrays, mainly Bragg lobes, can be accounted for and designed to control an array's backscatter.

21.9 RADOMES AND FREQUENCY SELECTIVE SURFACES

21.9.1 Radomes

One way to separate the structural mode from the antenna modes and attack them separately is by the use of a radome. Radomes have always been an important part of antenna design and integration by providing a protective surface around the antenna elements. The protective surface allows the antenna elements to have much lower environmental requirements, opening up the design space for the antenna elements. For example, a nose radome needs to withstand high speeds and loads, lightning strikes, bird strikes, rain erosion, and so on, but the antenna behind the radome can have many of these requirements reduced. As avionic systems have become more complex, these radomes have evolved from "dust" covers that just provide protection to a complex system whose function is more aptly described as a "prescription lens." For antenna scattering, radomes also provide an ability to separate out different scattering modes and thus allowing for improved signature design freedom for the antenna elements.

Figure 21.29 shows how a radome supports the RCS reduction of antenna systems. Figure 21.29a illustrates the radome's transmission/backscatter performance. This example is for a bandpass radome. For the bandpass radome, the out-of-band performance has low reflectivity (due to its shape) by behaving similar to a reflective surface shaped for low scattering. Over the "in-band" region, the radome becomes transmissive and its backscatter is considered to be somewhat "high" since with a transparent radome, the cavity behind the radome generates a large backscatter. Figure 21.29b shows how the antenna behaves conceptually. The antenna's RCS has a reduced backscatter in the operational band, but outside the antenna's band of operation, the antenna has a large backscatter. Together, the radome and antenna provide a reduced broadband backscatter system with the radome handling the backscatter both above and below band and the antenna handling the backscatter in-band. This is shown in Figure 21.29c. Since the antenna has no out-of-band backscatter requirements, there are more choices for the antenna and its design. The radome's ability to eliminate much of the structural scattering from the antenna design is similar to a traditional antenna's ability to cover much of the environmental requirements for the system without imposing those requirements on the antenna.

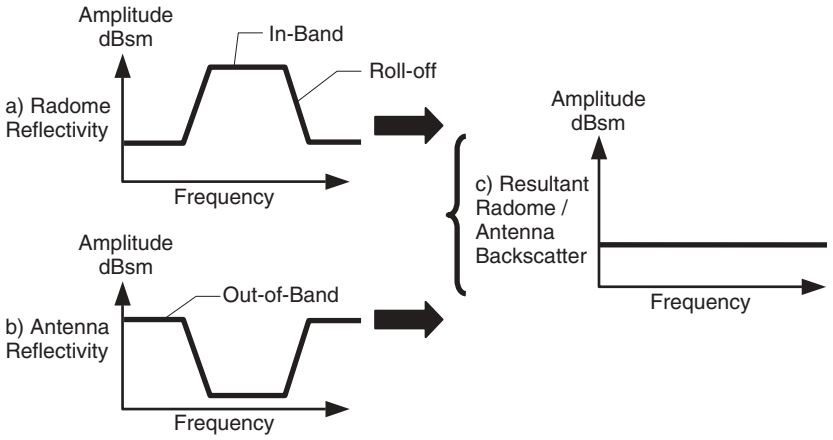


Figure 21.29 Use of radomes to control scattering by having the radome responsible for the out-or-band RCS and the antenna responsible for the in-band RCS. Together as a system, the system has low backscatter.

21.9.2 Frequency Selective Surfaces

Radomes that have passband characteristics usually are fabricated using frequency selective surface (FSS) materials. In reality, most materials/systems have transmission or reflection characteristics that change with frequency. The traditional radome behaves as an FSS with a very low loss passband at the operational band and higher reflections outside the passbands.

The RAM in Figure 21.20 illustrates performance that could be utilized as an FSS. In Figure 21.21, the transmission coefficient at low band has very little loss while at high frequencies and higher incident angles, the loss increases substantially. This RAM has characteristics of a lowpass FSS. A grid of wires, each with spacing less than 0.1λ, will reflect almost 100% of the energy incident on it while at higher frequencies, the grid will be fairly transmissive. The grid's behavior is characterized as a highpass FSS. Another lowpass system that could be used as an FSS is the skin depth of thin metallic materials. The skin depth δ is given as

$$\delta = \frac{1}{\sqrt{\pi f \mu \sigma}} \tag{21.43}$$

and

$$E(z) = E_0 e^{-z/\delta} \tag{21.44}$$

where

δ = skin depth, distance at which the field drops 64%
f = frequency
μ = magnetic permeability (free space = $4\pi \times 10^{-7}$ h/m)
σ = material conductivity
E_0 = field strength outside the metal
z = depth of penetration into the metal

Example 5 If it was desired to achieve a 12-dB reduction in one-way transmission at 10 GHz, what would be the required thickness of the copper to achieve this shielding? What would be the one-way transmission reduction at 1 GHz for this thickness of copper?

Skin depth is given as

$$\rightarrow \delta = 1/\sqrt{\pi(10 \times 10^9)(4\pi \times 10^{-7}\ \text{H/m})(5.8 \times 10^7\ \text{S/m})} = 6.6 \times 10^{-4}\ \text{mm}$$

To achieve 12 dB reduction,

$$\rightarrow -12\ \text{dB} => 10^{-12/20} = 0.25 \text{ of original field}$$

$$\rightarrow E(z) = 0.25E_0 = E_0 e^{-z/\delta} => -z/\delta = \ln(0.25) = -1.39$$

$$\rightarrow -z/\delta = -z/(6.6 \times 10^{-4}\ \text{mm}) = -1.39, \quad z \rightarrow = 9.2 \times 10^{-4}\ \text{mm of copper}$$

Now to determine the field transmission at 1 GHz, we use

$$\rightarrow \delta = 1/\sqrt{\pi(1 \times 10^9)(4\pi \times 10^{-7}\ \text{H/m})(5.8 \times 10^7\ \text{S/m})} = 21 \times 10^{-4}\ \text{mm}$$

$$\rightarrow -z/\delta = -9.2 \times 10^{-4}\ \text{mm of copper}/21 \times 10^{-4}\ \text{mm} = -0.438$$

$$\rightarrow -0.438 = \ln(\text{field strength}) \rightarrow e^{-0.438} = e^{\ln(\text{field strength})} \rightarrow 0.645 = \text{field strength}$$

$$\rightarrow \text{to obtain dB} \rightarrow 20\log(0.645) = -3.8\ \text{dB}$$

Thus a thin copper layer will operate as a lowpass FSS; and 9.2×10^{-4} mm of copper will reduce the one-way transmission at 10 GHz by 12 dB while reducing the one-way transmission at 1 GHz by only 3.8 dB. (See Figure 21.30.)

Bandpass or bandstop FSS performance can be achieved by etching periodic shapes into a continuous metallic layer. For cases in which the "element" is metallic and the remainder of the surface metal has been etched away, the performance is of a bandstop or reflective surface in band. For cases in which the "element" is the etched unit and the remainder of the surface is metallic, the system performs as a bandpass FSS. Practical implementation of an FSS typically utilizes a nonmetallic composite surface that is metallized and then the FSS elements are formed using chemical etch or other

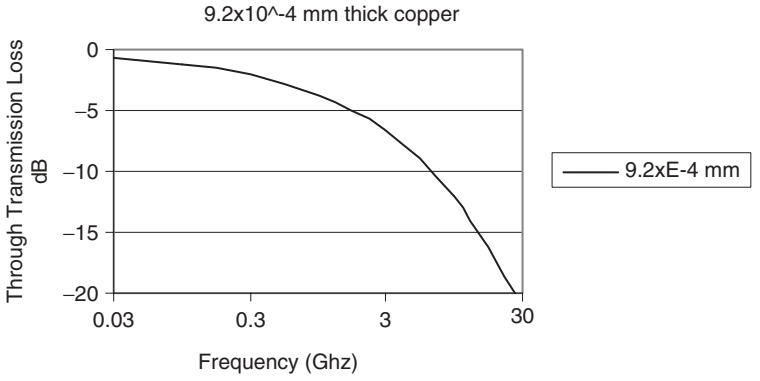


Figure 21.30 Analytically modeled thin layer of metal behaves as a lowpass frequency selective surface (FSS).

methods. Since the radome and elements are in close proximity to the antenna, the metal and dielectric are usually formed from highly conductive materials and low loss dielectrics.

Figure 21.31 illustrates some of these bandpass/bandstop FSS geometries. It is beyond the scope of this chapter but many of these shapes have been analyzed for their performance in systems and the reader is referred to Mittra, Chan, and Cwik [19] for the analytical underpinnings needed to support the design of bandpass/bandstop FSSs. An example of an FSS system and its performance is shown in Figure 21.32. Figure 21.32 shows the frequency response of the Jerusalem cross FSS that is operated as a bandstop element. The antenna element is on the order of 0.5λ in size at the design frequency center. Center design frequency is at 8 GHz and at that center frequency the FSS system has a reflection coefficient of 1 or is 100% reflective. Out of band, the FSS system becomes

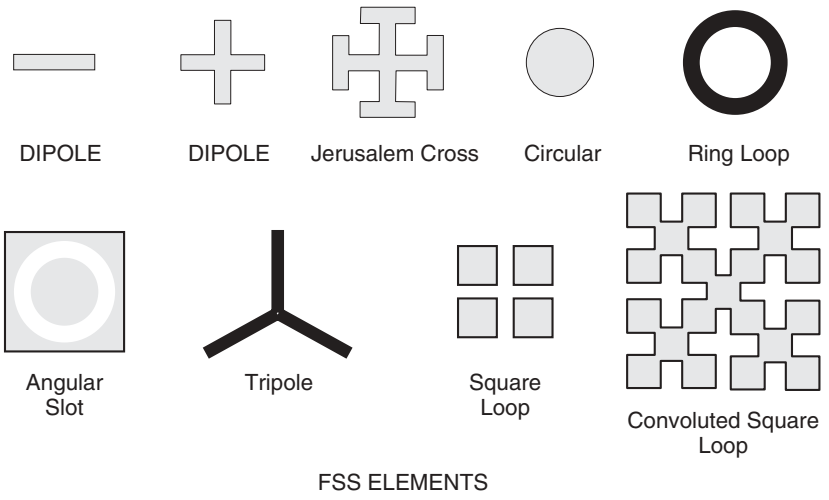


Figure 21.31 Bandpass FSS elements can come in many different shapes, each with their own frequency response.

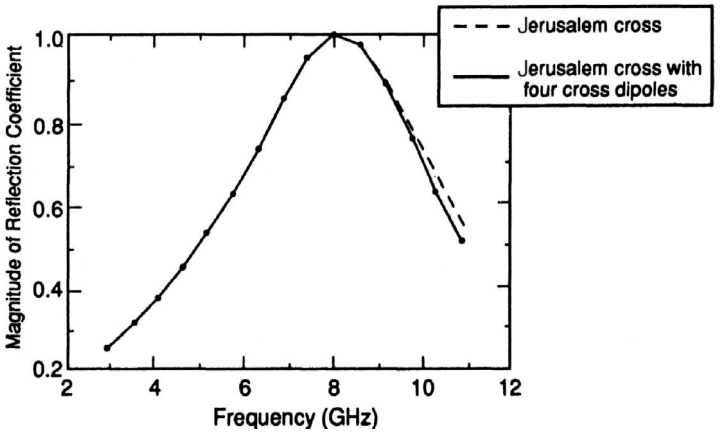


Figure 21.32 Frequency response of bandstop FSS centered at 8 GHz. (From Ref. 19.)

more reflective as one moves away from the design frequency. FSS layers can be used to provide additional performance in roll-off but usually with each layer, additional loss will be incurred since none of the layers is totally lossless.

Beyond the frequency response, one must also be concerned with other FSS characteristics such as the FSS's behavior with respect to incident angles and incident polarization. The transverse electric field performance, in particular, becomes poor as the incident angle moves toward grazing. Figure 21.33 shows frequency versus incident angle transmission performance for an FSS used in a radome wall. As shown, this is a bandpass system that has approximately 12% bandwidth measured with 1 dB, one way as the metric. The FSS supports transverse electric field transmission incident out to 65°.

In a radome there are many ways to utilize an FSS to achieve performance. Figure 21.34 illustrates ways that FSS radomes can be used to reduce backscatter. System trades for FSS radomes with different performance are discussed later.

One method (method a) of integrating an FSS with an antenna, shown in Figure 21.34, is to install the FSS just above the antenna at FSS Surface B in the cavity, away from the outer mold line surface. Since the FSS is not part of the exterior radome (no FSS in

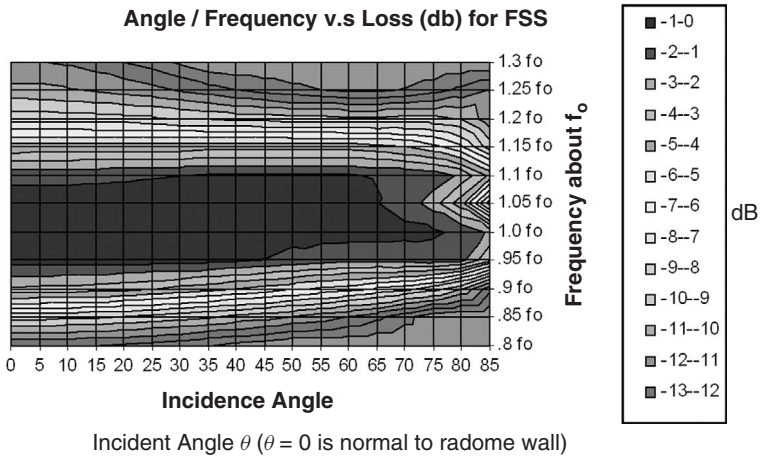


Figure 21.33 Bandpass FSS radome wall performance with respect to angle and frequency. Transmission is problematic for a transverse electric field beyond 65° from normal.

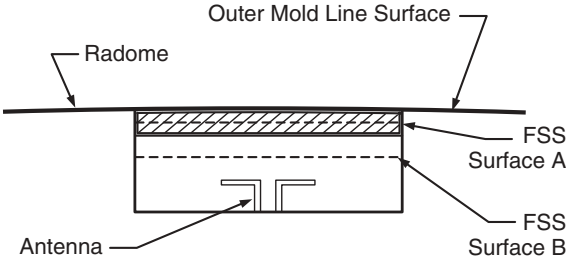


Figure 21.34 Radome/FSS integration concepts to support electrical, RCS, and environmental requirements. (From Ref. 20.)

FSS Surface A), its environmental requirements are reduced. The following trades are available:

- Better choice of materials since exterior environmental is not a concern
- Lowest cost solution since the surface can be contoured as desired (usually flat or at least curved in a single dimension because of material flexibility)
- Good radar antenna performance since there is maximum flexibility in material choices and coupling between antenna and FSS
- Simplified analysis for both RCS and RF due to the planar nature of the FSS and antenna
- Internal reflections limit scattering reduction due to cavity effects
- Avoids difficult environmental requirements (such as protection against lightning strikes)

A second method (method b) of integrating an FSS with an antenna, shown in Figure 21.34, is to install a highpass FSS just above the antenna at FSS Surface B in the cavity, away from the surface, and a lowpass FSS in FSS Surface A, the radome. This allows some flexibility in tailoring both FSSs to meet bandwidth requirements. Since the FSS is a part of both the radome (FSS in FSS Surface A) and FSS Surface B, this hybrid approach does have compromises. Advantages and disadvantages are:

- Some flexibility in materials but having an FSS in Surface A will require it to meet environmental requirements.
- Slight increase in weight.
- Increase in cost, considered moderate cost solution with FSS surface A needing to conform to the surface and reduced material choices.
- Some internal scattering eliminated (compared to method a) with the FSS in FSS Surface A.

A third method (method c) of integrating an FSS with an antenna, shown in Figure 21.34, is to install a bandpass FSS just above the antenna at FSS Surface A in the exterior radome. Since the FSS is a part of the radome (FSS in FSS Surface A), all RF, environmental, scattering, and structural requirements need to be satisfied by this system. Advantages and disadvantages are:

- Low to moderate weight (elimination of second radome reduces weight).
- Higher cost—complex electrical structure in an environmentally challenged location with reduced material selection drives cost.
- Reduced radar performance; since FSSs are tied to radome shape, there may be conflicts if radome shape has high incident angles to RF (antenna) energy.
- Higher analytical complexity for both RCS and antenna performance.
- Internal cavity scattering is minimized.

All three concepts—method a, method b, and method c—provide advantages and disadvantages. As in all designs, understanding the system requirements and what drives the extremes in the decision is critical.

21.10 SUMMARY

Finally, in putting it all together, a low observable/RF system has many competing variables. Environmental considerations include lightning, hail, static, rain erosion, opening and closing loads, static loads, fatigue loads, structural resonances, weight, impact strength and repair, humidity intrusion, and CTE. Operational considerations include produceability, maintainability, repairability, and interchangeability.

Antenna considerations include gain, pattern, boresight, image lobes, impedance, mutual coupling, transmit power, loss, and polarization. And finally, the scattering considerations are shown in Figure 21.35.

As a final checklist for low observable antenna designs, the following is a summary of concepts from this chapter:

- Know and understand the requirements and how those requirements map to the antenna system variables.
- Conformal is good; understand the as-is installed performance. Select antenna types with low backscatter potential.
- Decompose the antenna RCS problem into the different scattering modes:
  - Structural mode
  - Antenna resonant mode
  - Antenna terminal mode
- Apply scattering reduction techniques to the different modes that yield the best combination of backscatter and antenna performance.
- Using vector math, attack the largest scattering contributors first.
- Understand and control the four scattering mechanisms as applied to the antenna system.
- Radomes are our friends: use them to separate requirements from the antenna—RCS, environmental, and so on.

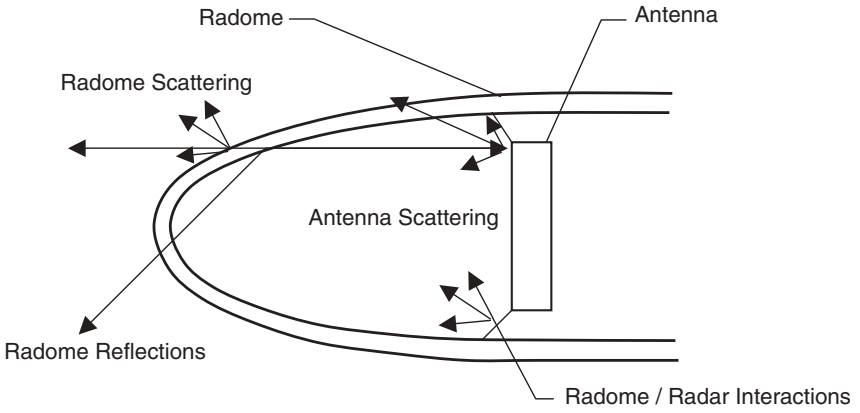


Figure 21.35 Scattering considerations when integrating a radome with an antenna.

ACKNOWLEDGMENT

We thank Dr. Brian Baertlein and Dr. Ed Joy for their review and guidance in developing this chapter.

REFERENCES

1. E. F. Knott, J. F. Shaeffer, and M. T. Tuley, *Radar Cross Section*, Scitech Publishing, Raleigh, NC, 2004.
2. G. T. Ruck, D. E. Barrick, W. D. Stuart, and C. K. Krichbaum, *Radar Cross Section Handbook*, Vols. I and II, Plenum Press, New York, 1970.
3. C. A. Balanis, *Antenna Theory: Analysis and Design*, 3rd ed., John Wiley & Sons., Hoboken, NJ, 2005, Chapter 2.
4. J. W. Crispin, Jr. and K. M. Siegel, *Methods of Radar Cross Section Analysis*, Academic Press, New York, 1968.
5. R. B. Green, The general theory of antenna scattering, Report 1223-17, Antenna Laboratory, Ohio State University, 30 November 1963.
6. R. C. Hansen, Relationships between antennas as scatterers and as radiators, *Proc. IEEE*, Vol. 77, pp. 659–662, May 1989.
7. C. G. Montgomery, R. H. Dicke, and E. M. Purcell, *Principles of Microwave Circuits*, Boston Technical Publishers, Boston, 1964 (also available in the original MIT Rad Lab Series).
8. J. Appel-Hansen, Accurate determination of gain and radiation patterns by radar cross section measurements, *IEEE Trans. Antenna Propag.*, Vol. 27, pp. 640–646, September 1979.
9. J. D. Kraus, *Antenna*, McGraw-Hill, New York, 1950.
10. R. E. Collin, *Field Theory of Guided Waves*, 2nd ed., IEEE Press, Piscataway, NJ, 1991.
11. D. Deschweinitz, D. Kitchen, and E. Ingrassia, Internal Document, July 1995.
12. J. F. Lindsey, Radar cross-section effects relating to a horn Antenna, *IEEE Trans. Antenna Propag.*, Vol. 37, pp. 257–260, February 1989.
13. C. A. Balanis, *Antenna Theory: Analysis and Design*, 3rd ed., John Wiley & Sons, Hoboken, NJ, 2005, pp. 98–105.
14. A. Brown, Stealth design fundamentals, *Lockheed Horizons*, Issue 31, pp. 6–12, August 1992.
15. C. Larson, C. Carter, and G. Niizawa, *Low Observable for Designers*, Lockheed Martin Technical Institute, Short Course Notes.
16. R. D. Strattan, A design procedure for surface wave absorbers (U), in *1980 Radar Camouflage Symposium*, March 1981, pp. 117–122.
17. B. A. Munk, *Finite Antenna Arrays and FSS*, John Wiley & Sons, Hoboken, NJ, 2003, Chap. 2.
18. J. I. Simon, Impedance properties of periodic linear arrays conformal to a dielectric-clad infinite PEC cylinder, Ph.D. dissertation, Ohio State University, Department of Electrical Engineering, Columbus, OH, 1989.
19. R. Mittra, C. H. Chan, and T. Cwik, Techniques for analyzing frequency selective surfaces—a review, *Proc IEEE*, Vol. 76, pp. 1593–1621, December 1988.
20. B. McGarraugh and E. Ingrassia, Interim study on aperture trades, Lockheed Internal Document, August 1996.

PART V

ANTENNA APPLICATIONS

CHAPTER 22

Integrated Antennas for Wireless Personal Communications

YAHYA RAHMAT-SAMII, JERZY GUTERMAN, A. A. MOREIRA, and C. PEIXEIRO

22.1 INTRODUCTION

During the last decade, the popularity of personal wireless communication devices has grown tremendously, and according to market evaluations, this trend will continue. The explosion of cellular telephony undoubtedly places handsets among the most numerous of released mobile units (Figure 22.1). As the wireless local area network (WLAN) and personal area network (PAN) are becoming a standard, the vast majority of new laptop computers and personal digital assistants (PDAs) are being equipped with radio interfaces. Moreover, a single terminal often integrates several wireless systems, for example, novel mobile phones ("smart phones" or "blackberries") come with built-in WLAN and Bluetooth cards, while some laptops integrate cellular network radios giving access to the Internet in areas not covered by WLANs. In addition, some terminals are equipped with digital television and global positioning system (GPS) receivers.

Antennas play a paramount role in the optimal design of a personal wireless communication terminal unit. A properly designed and integrated radiator not only enables a reliable connectivity, independent of terminal orientation, but also helps prevent excessive battery energy consumption due to an improvement in efficiency. Additionally, the antenna design (see Figure 22.2) takes into account the housing effects within a small, densely packaged terminal. Furthermore, as the unit usually operates in close vicinity to the user, the mutual interaction between the antenna and human body has to be understood and accounted for in the antenna design and integration stages. The antenna implementation issues are schematically represented in Figure 22.2.

In general, antennas for mobile units are required to be small, lightweight, and low profile. The volume constraints, often combined with large bandwidth and high efficiency specifications, make the antenna design a challenging task. Moreover, multiple wireless systems often use a common antenna element, hence multiband designs are necessary. An overview chart of common wireless frequency bands, with corresponding bandwidths, is presented in Figure 22.3.

Modern Antenna Handbook. Edited by Constantine A. Balanis

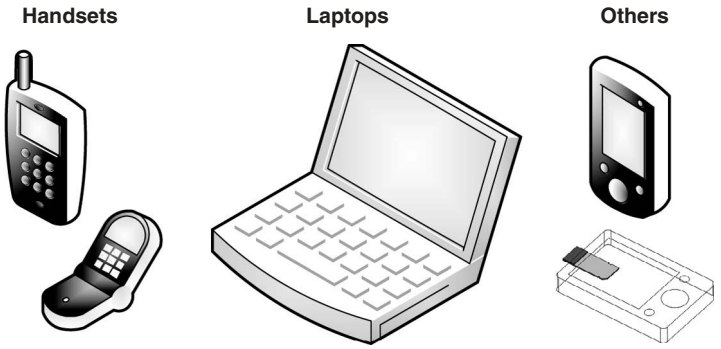


Figure 22.1 Types of mobile wireless unit terminals addressed in this chapter.

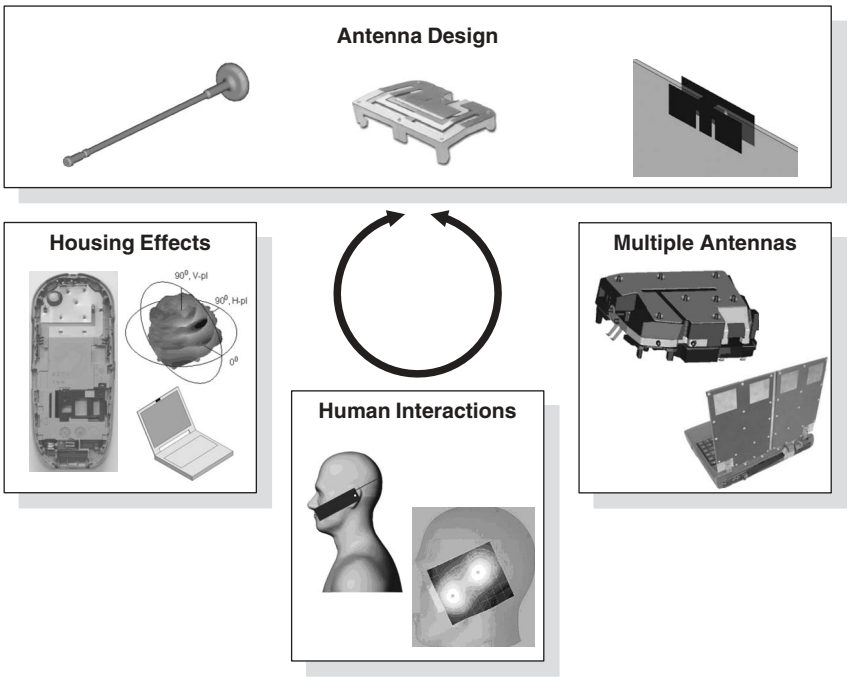


Figure 22.2 Integrated antenna design and related issues. The radiator performance depends on all environmental factors, which should also be considered in the antenna design stage.

Many of antennas for mobile terminals are required to have horizontal plane radiation patterns that are close to omnidirectional, as this favors maintaining a reliable wireless connection independent of terminal orientation. Due to the omnidirectional nature, the far-field radiation pattern of an integrated antenna can be significantly affected by other nearby terminal components as well as by the presence of the operator. Therefore the antenna housing effects and interaction with the human have to be carefully studied in order to select the best antenna design and radiator location within the terminal.

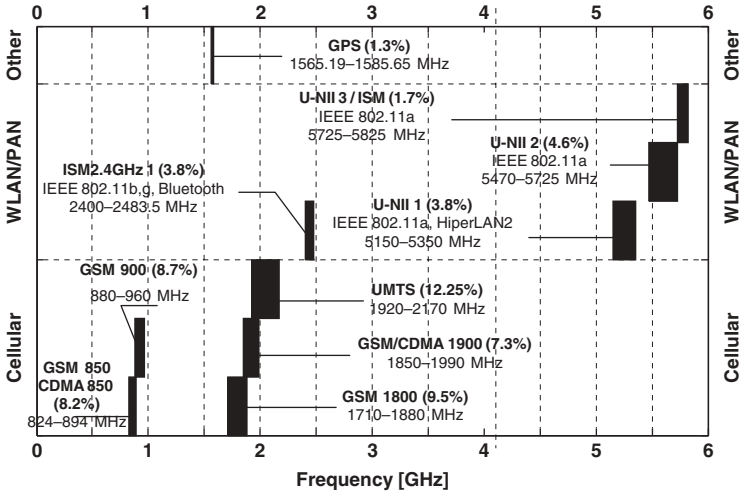


Figure 22.3 Overview of wireless bands widely used in wireless personal communications. The chart presents the general bands allocation, which may vary for different geographical regions.

The antenna position within the terminal also affects the amount and distribution of radio frequency (RF) power absorbed by the user's body and nearby objects. These effects are most pronounced in the handset terminals, where in normal operating conditions, the radiating elements are in close vicinity to human tissues and wearable fabrics. Moreover, the maximum radiated power by the terminal antenna is significantly higher for cellular than for WLAN units. This is because the distance to a cellular base station can be much larger than to a WLAN access point.

The global pursuit of wireless system reliability and the growing demand for high spectrum efficiency are witnessed through the recent achievements of diversity, smart antennas, and multiple-input multiple-output (MIMO) techniques. As a result, integrated multiple element antenna arrangements are becoming a trend in mobile terminal development. The design challenges in these arrangements are related to terminal size constraints since small spacing relative to the wavelength between elements causes high mutual coupling, degrades the performance, and lowers the total efficiency of multiple antenna arrangements.

The aim of this chapter is a systematic presentation of basic design concepts of integrated antennas for wireless terminals. The material presented is divided into four main sections:

- Fundamentals (Section 22.2)
- Integrated handset antennas (Section 22.3)
- Integrated antennas for laptops (Section 22.4)
- Integrated antennas for other mobile devices (Section 22.5)

For each mobile terminal type (Figure 22.1), the antenna design procedure is described, dealing with the issues depicted in Figure 22.2. After presenting the most common

antenna configurations and design methodologies, the housing effects and human interactions are described. Finally, for each mobile unit type, the integration of multiple antennas is discussed and illustrated with representative examples.

22.2 FUNDAMENTALS

This section contains a brief review of important design parameter definitions for integrated antennas. An outline of the characteristics used for an individual antenna element is followed by the list of measurements used in evaluating multiantenna arrangements.

22.2.1 Individual Antenna Design Parameters

22.2.1.1 Antenna Impedance, Reflection Coefficient, and Operation Band

The antenna input impedance is defined as the ratio of the voltage $V(0)$ to the current $I(0)$ at the antenna feed point:

$$Z_{\text{in}} = \frac{V(0)}{I(0)} = \frac{V_0^+ + V_0^-}{V_0^+ - V_0^-} Z_0 \tag{22.1}$$

where Z_0 denotes the characteristic impedance of the feeding transmission line, and V_0^+, V_0^- are the amplitudes of incident and reflected voltage waves at the antenna feed point, respectively.

The S_{11} parameter is the antenna voltage reflection coefficient, which is defined as the amplitude of the reflected voltage, V_0^-, normalized to the incident voltage amplitude at the feed point

$$S_{11} = \frac{V_0^-}{V_0^+} = \frac{Z_{\text{in}} - Z_0}{Z_{\text{in}} + Z_0} \tag{22.2}$$

In wireless applications, the antenna operation frequency band is often defined as the impedance match band, where $|S_{11}|$ is below a certain level (typically -6 or -10 dB, which corresponds to 25.1% or 10% reflected power, respectively).

22.2.1.2 Directivity and Gain

The directivity of an antenna in a given direction (θ, ϕ), $D(\theta, \phi)$, is the ratio of the radiated power density $S(\theta, \phi)$ to the radiated power density averaged in all directions, or

$$D(\theta, \phi) = \frac{4\pi S(\theta, \phi)}{\int_0^{2\pi} \int_0^{\pi} S(\theta, \phi) \sin\theta d\theta d\phi} \tag{22.3}$$

The gain of an antenna in a given direction (θ, φ), $G(\theta, \varphi)$, takes into account the radiation efficiency η_{cd} of the antenna as well as the directivity. It is defined as

$$G(\theta, \phi) = \frac{4\pi r^2 S(\theta, \phi)}{P_{in}} = \eta_{cd} D(\theta, \phi) \tag{22.4}$$

where P_{in} denotes power accepted (input) by antenna.

If the direction (θ, ϕ) is not specified, D and G are usually taken in the direction of maximum radiation.

22.2.1.3 Absorbed Power and Specific Absorption Rate For a handset antenna located in the vicinity of the human biological tissue, some portion of the antenna delivered power is absorbed in the tissue. The total power absorbed in the lossy tissues can be defined as

$$P_{\text{abs}} = \frac{1}{2}\iiint_V \sigma |\mathbf{E}|^2 dV \tag{22.5}$$

where σ is conductivity of the tissue, $\mathbf{E}$ is the electric field intensity, and V is the volume of the biological tissue. The radiated power to the far-field region can be obtained from

$$P_{\text{rad}} = \frac{1}{2}\text{Re}\oiint_S (\mathbf{E} \times \mathbf{H}^*) \cdot \hat{\mathbf{n}}\, dS \tag{22.6}$$

where $\mathbf{E}$ and $\mathbf{H}$ are electric and magnetic field intensities on the surface S completely enclosing the antenna and the biological tissue, and $\hat{\mathbf{n}}$ is the outward unit vector normal to the surface. One may now define the efficiency of the antenna/tissue system as

$$\eta_a = \frac{P_{\text{rad}}}{P_{\text{rad}} + P_{\text{abs}}} = \frac{P_{\text{rad}}}{P_{\text{del}}} \tag{22.7}$$

where P_{del} is delivered by antenna. The total antenna efficiency η_a is the product of the reflection efficiency $(1 - |\boldsymbol{S}_{11}|^2)$, conduction efficiency, dielectric efficiency, and antenna–tissue efficiency.

The specific absorption rate (SAR) is an important parameter when discussing the health risk associated with electromagnetic power absorption. SAR is defined as

$$\text{SAR} = \frac{\sigma}{2\rho}|\mathbf{E}|^2 \tag{22.8}$$

where ρ is the tissue density. The ANSI/IEEE standard C95.3–2002 RF Safety Guideline [1] suggests that the 1-g averaged peak SAR should not exceed 1.6 W/kg and the whole-body averaged peak SAR should be less than 0.08 W/kg. These guidelines are applicable to uncontrolled situations and therefore must be satisfied for personal handsets. The maximum value of 2 W/kg defined for 10-g averaged peak SAR is more commonly used in Europe [2, 3].

22.2.2 Multiple Antenna Design Parameters

22.2.2.1 Mutual Coupling In N-element antenna arrangements, the mutual coupling S_{ij} between ith and jth antenna elements is found by driving port j with an incident wave of amplitude a_j, and measuring the reflected wave amplitude, b_i, coming out of port i:

$$S_{ij} = \left.\frac{b_i}{a_j}\right|_{a_k=0 \text{ for } k\neq j} \tag{22.9}$$

The incident waves on all other k ports except for the jth port are set to zero, which means that all ports should be terminated in matched loads to avoid reflections.

For the N-antenna arrangement, the individual elements' reflection coefficients and mutual coupling parameters constitute the scattering matrix $[S]$:

$$\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1N} \\ S_{21} & \cdot & \cdots & S_{2N} \\ \vdots & \cdot & \ddots & \vdots \\ S_{N1} & \cdot & \cdots & S_{NN} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix} \tag{22.10}$$

22.2.2.2 Total Active Reflection Coefficient The total active reflection coefficient (TARC) is a new parameter used to evaluate the performance of a multiport antenna in a single measure. TARC is calculated as the square root of the incident power provided by all excitations minus the radiated power and then dividing by the incident power [4]. For lossless antennas it can be calculated as

$$\text{TARC} = \frac{\sqrt{\sum_{i=1}^{N} |b_i|^2}}{\sqrt{\sum_{i=1}^{N} |a_i|^2}} \tag{22.11}$$

TARC can be used as an alternative criterion to define the multiport antenna operating band, as described in Section 22.4.4.

22.3 INTEGRATED HANDSET ANTENNAS AND HUMAN INTERACTIONS

In the last fifteen years, mobile communications have experienced a tremendous development. The demand for innovative systems and services has created new and challenging handset requirements. More and more the handset has to be small, robust, aesthetically appealing, multiband, and with long-life (battery). Moreover, in the very competitive global market, low price can provide a definitive positive differentiation for success. As the antenna is an important component of the handset, new and more challenging requirements have to be faced by its designers.

New types of antennas have been developed and well known antenna types have been adapted to the new handset applications. Initially, external antennas such as whip and stubby antennas were used almost exclusively [5] (see Section 22.3.1). However, today internal antennas are mainly used. Many types of handset internal antennas have been created, but the planar inverted-F antenna (PIFA) remains one of the most popular [5–7].

With the demand for multistandard handsets and for higher and higher transmission rates, the integration of multiple antennas has become a need. Different antennas can be used for different systems as, for instance, a PIFA for mobile communications and an inverted-F antenna (IFA) for GPS [5] (see Section 22.3.2). However, more than one similar or different antenna can be used for the same system for diversity or MIMO enhancement [8, 9] (see Section 22.3.3.3).

The design of one or more antenna elements to be integrated into a small handset has to be carried out taking into account the housing effects of specific handset components. Due to the very compact structure of the handset, components such as the battery, the display, the speaker, and many others are in close vicinity to the antenna system and may have nonnegligible effects on its performance. Some of these components, as well

as the small ground plane and the plastic (or metal) cover, have to be included in the final design model.

Another very important issue is the interaction of the antenna with the user. This interaction has to be considered from two points of view. One is the impact of the user on the antenna performance. The user represents a large dielectric and lossy body that affects the antenna parameters (impedance, radiation pattern, and efficiency) [10–12]. Although different users have different characteristics and hold the handset in different positions, there are some basic guidelines that can be followed in the design procedure to minimize the user effects [10]. The other point of view is the impact of the handset electromagnetic (EM) radiation on the user. The absorption of EM energy by the human body is quantified in the SAR. The SAR of handsets is a very hot topic as the potential health risks have become a point of very intense and critical public discussions [13]. Handset manufacturers conduct SAR measurements to comply with the maximum SAR accepted levels [1, 2, 14]. Therefore the SAR of the handset has to be taken into account in the design procedure [15].

The structure and topics of this section are presented in the flowchart shown in Figure 22.4.

22.3.1 Overview of Handset External Antennas

In spite of the huge improvements in electronics that now enable small and feature-rich units, the physical constraints on the size of the antennas remain wavelength dependent and cannot be overlooked. In the first hand-held communication radios, a grounded (by the user) quarter-wavelength whip or helical antenna was virtually the only solution due to the long wavelengths used. These solutions were adopted in the beginning of modern mobile communication systems and can still be found in a few present-day hand-held phones. The major difference between the older and newer portable radios is due to the wavelength: in modern day handsets, the wavelength is much shorter, allowing for smaller and more complex structures. In many instances, the antennas are partially external to the main body of the handset and are typically referred to as external antennas. In some phones, like those of the "clamshell" type, the available space is minimal and unavoidably some sort of external antenna is likely to be the best solution. Additionally,

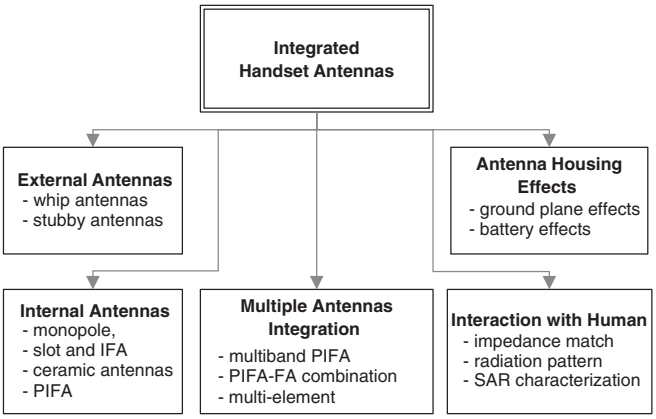


Figure 22.4 Flowchart of topics presented in Section 22.3.

when noncellular systems, like GPS, are to be included, a combination of external and internal antennas is often implemented.

In general, antennas for handsets have been required to be small, lightweight, and low profile and to have an omnidirectional radiation pattern in the horizontal plane. The most prevalent types of these external antennas still on the market today are the whip (monopoles) and stubby antennas. Figure 22.5a shows a whip antenna. At the top, a plastic cap is present so users can pull the whip out. The bottom part is connected to the feed point on the printed wiring board (PWB). Conventional whip antennas are quarter-wavelength ($\frac{1}{4}\lambda$) monopoles. Using the ground plane to create an image, the monopole unit behaves similarly to a dipole antenna. In practical configuration, a gain of 2.15 dB can be achieved. At the 800-MHz band, the radiation pattern is the typical donut shape, while at the 1900-MHz band, the butterfly-shape pattern is usually seen. The electrical length of the monopole determines the input impedance and the resonant frequency; however, the position of the monopole antenna in the handset does not significantly affect the impedance and the resonant frequency.

Whip antennas are used in the U.S. code division multiple access (CDMA) market because there are fewer base stations in a CDMA network, so a handset with only a PIFA hardly meets the performance requirements from mobile carriers. Also, when a handset with only a PIFA is put against the user's head, the gain can drop up to 3 dB. A PIFA–whip combination (Figure 22.5b) helps to overcome this drawback. Research on the coupling between the whip and the PIFA have been reported [16, 17].

External antennas are usually designed to be retractable so that when they are extended, better antenna performance against the human head, hand, and body can be achieved. On the one hand, this is an advantage; but on the other hand, movable parts are a drawback because the antenna can easily be damaged by the user. For the sake of ruggedness, retractable antennas tend to be avoided in handsets nowadays.

Stubby antennas are another example of external antennas. These antennas are either of the helix or of the meander type (shown in Figure 22.6). The helix is typically made of a metal wire, while the meander is basically made of a flex with printed wire and wrapped around. The meander is relatively easy to fabricate and more adequate for mass production.

Stubby antennas have less dependency on the size of the ground plane than internal antennas. Although more robust than the whip antenna, the stubby antenna still protrudes from the handset case and therefore is more prone to damage than internal antennas.

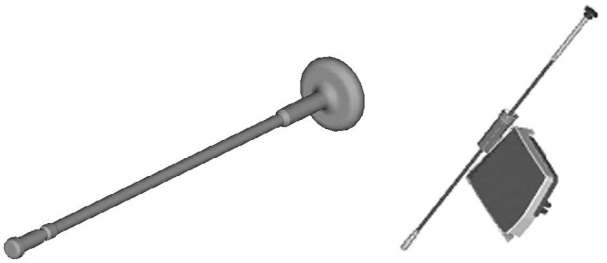


Figure 22.5 (a) Plastic-coated nickel titanic whip antenna with a plastic cap for extension or retraction; (b) PIFA–whip combination. (From Ref. 5, with permission.)

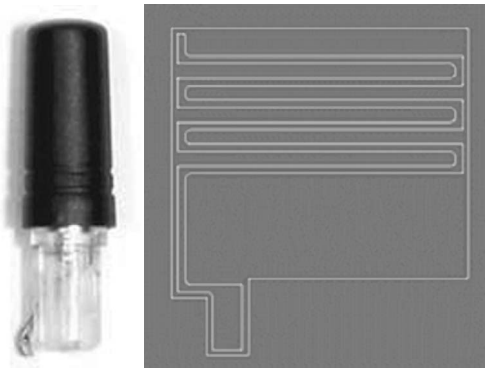


Figure 22.6 A stubby antenna and the meander. (From Ref. 5, with permission.)

Radiation losses due to the presence of the user have been compared for handsets with internal and external antennas, and it has been shown that handsets with built-in antennas are much less sensitive to how the phone is held than handsets with external antennas [18].

22.3.2 Overview of Handset Internal Antennas

Internal antennas have been adopted by most handset manufacturers, mainly because of mechanical and aesthetic reasons. Furthermore, these designs are inherently less prone to damage. Internal antenna solutions are usually based on planar antenna configurations. These are mostly PIFAs (planar inverted-F antennas) and folded monopoles [19, 20]. A schematic view of these two types of antennas is shown in Figure 22.7.

The main difference between the antenna types is that the ground plane directly underneath the monopole is cut out, while for the PIFA, the ground plane underneath the patch is required. It is generally accepted that the ground plane in the PIFA structure helps block the backward RF radiation and lower the SAR values [10].

Special designs of folded planar monopole antennas have been reported [21] with a broadband behavior potentially covering the bandwidth needs of Global System for Mobile Communication (GSM), Digital Communication System (DCS), Personal

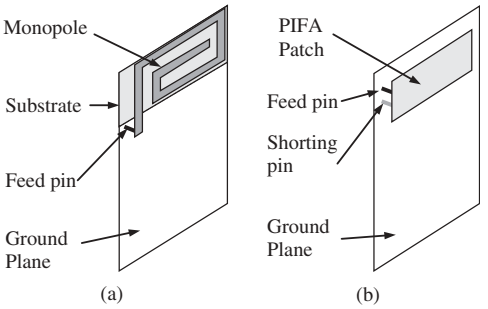


Figure 22.7 Configuration of two internal antennas: (a) planar monopole antenna and (b) PIFA. (From Ref. 5, with permission.)

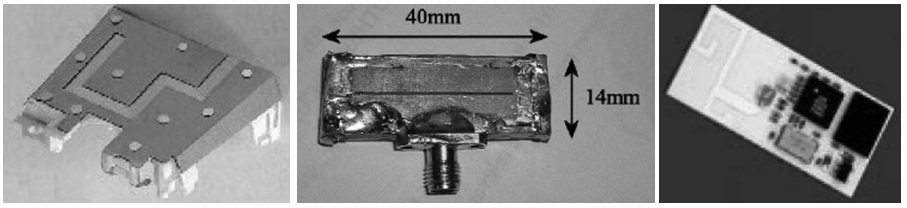


Figure 22.8 Internal antennas for mobile terminals: (a) dual-band GSM patch, (b) slot antenna, and (c) low temperature cofired ceramic (LTCC) antenna for Bluetooth. (From Ref. 22, with permission © 2005 IEEE.)

Communication System (PCS), Universal Mobile Communication System (UMTS), and Industrial Science Medical (ISM) bands.

In a recent survey of internal antennas used in GSM phones from 1998 to 2005 [7], different handset topologies, referred to as "clamshell" and "candy bar," as well as different antenna placements were examined. Of those types, the PIFAs appeared in several configurations such as PIFA combined with a switched matching circuit, stacked PIFAs, PIFAs with grounded parasitics, and PIFAs with special high resistivity materials. Different types of internal antennas were listed, such as PIFAs, ungrounded internal antennas, and slot antennas.

Relevant mechanical issues of internal antennas include the antenna volume (and its correlation with antenna performance and frequency of operation) and the production technology used to fabricate the antennas.

Antennas have evolved along with phone types in recent years. Common handset forms like clamshell, sliders, and dual-hinges include more frequency bands and larger screens. Moreover, antennas are going from single-band to dual-band, triple-band, or even quad-band combinations of 2G and 3G solutions.

Due to commercial reasons, terminal aesthetics are of great importance. This goes in favor of internally mounted antennas, such as those displayed in Figure 22.8 [22]. As stated before, the antenna performance is affected by the volume constraints, the close proximity to electronic components and the handset case and battery [23]. Another important feature affecting the antenna performance is the interaction with the user.

Since terminals are now multistandard, they must be multiband, implying the use of either multiple antennas or multiband antennas. Multiple antennas are also necessary in handsets that incorporate noncellular services such as GPS in the U.S. market (PIFA–IFA combinations are a possible total internal antenna solution [24]; a prototype is shown in Figure 22.9).

Future mobile terminals will require diversity schemes (this is already in use in Japanese and Korean systems). System aspects, such as the decision algorithms, affect the overall performance and have to be taken into account in the antenna design procedure. For handsets, due to the lack of space, the antennas must be electrically small [25] and inverted-F antennas appear to be a good candidate [26].

22.3.3 Integration of Multiple Antennas

With the huge increase of mobile communication and WLAN systems and services, the use of more than one antenna in a handset is often required. Some solutions have already

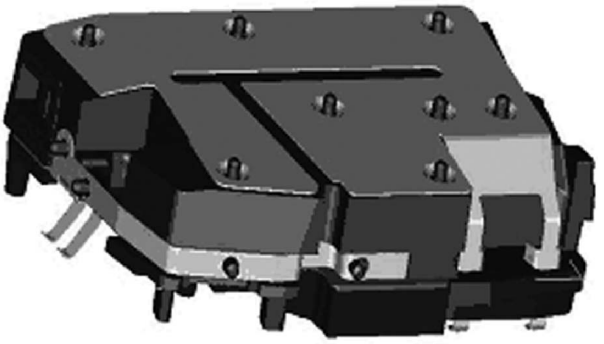


Figure 22.9 PIFA–IFA combination.

been reported [20–28] and advanced research activities in the field are being carried out worldwide. As the volume available in a small handset for the antenna system is very limited (typically less than 10–12 cm^3), integration of more than one antenna element is usually quite difficult. When two or more antennas are integrated in such a small volume, the separation between them is necessarily small and mutual coupling can be quite strong [29–32]. Although some techniques have been proposed to decrease or compensate for the mutual coupling [33–36], it still has some unwanted effects such as a decrease in efficiency, a change in the radiation pattern, and an increase in the correlation.

In order to fulfill the needs of various systems and standards, different antennas can be implemented for different standards: for instance, an inverted-F antenna (IFA) for GPS and a planar inverted-F antenna (PIFA) for mobile cellular communications [5]. However, several different or similar antennas can be used for the same standards [27, 37]. In this case, the antenna system provides diversity or MIMO enhancement.

In this section, a multiband PIFA design is introduced first. Then the integration of multiple antennas into a single handset is investigated. A combination of a cellular PIFA with a GPS IFA is described in Section 22.3.3.2 [5]. Section 22.3.3.3 contains a detailed description of a two wideband PIFA configuration [37].

22.3.3.1 Multiband PIFA Design A PIFA is a compact and low profile microstrip patch loaded with a short circuit. Its properties are well documented and have provided generalized use in small handset applications [6, 38]. Many variations have been introduced in the initial PIFA configuration. By cutting slots on the patch or by adding parasitic elements, dual-band or multiband behavior can be obtained [20, 39–42].

Figure 22.10 shows a triple-band PIFA integrated into a small 100×40 mm^2 handset [41]. The main requirement of the antenna is the ability to operate in GSM 1800 (1710–1880 MHz), UMTS (1900–2170 MHz), and IEEE 802.11a (5150–5350 MHz) bands. Due to the proximity of the two lower bands (20 MHz), the patch and the outer U slot are designed to form a wide band that will cover the two standards. The inner U slot is designed to operate in the IEEE 802.11a band. The required bandwidths are $\Delta f_1 = 170$ MHz (9.4%), $\Delta f_2 = 270$ MHz (13.3%), and $\Delta f_3 = 200$ MHz (3.8%), or, if the two lower bands are combined, $\Delta f_{12} = 460$ MHz (23.7%). The patch is printed on a 1.575-mm thick dielectric substrate ($\varepsilon_r = 2.2$) over a 10-mm air gap, which separates the

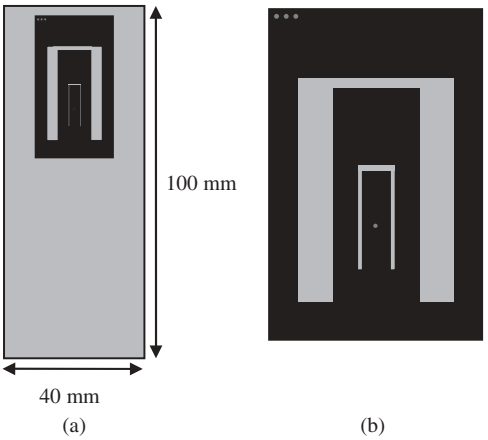


Figure 22.10 Triple-band PIFA configuration: (a) PIFA in the handset and (b) details of the patch and U slots.

substrate from the ground plane. The antenna is fed by a coaxial probe inside the inner U slot. Instead of the common short-circuit plate, three shorting pins are inserted near the upper corner of the patch [41]. A -6-dB S_{11} reference level (VSWR ≤ 3) is used to define impedance bandwidth. For small handsets, where very tight volume constraints exist, this -6-dB reference level is usually accepted.

A prototype has been fabricated (Figure 22.28) for proof of concept. The S_{11} of the antenna prototype has been measured with a vector network analyzer. The corresponding amplitude is shown in Figure 22.11. It is seen that $|S_{11}| \leq -6$ dB covers the specified frequency bands (1710–2170 + 5150–5350 MHz). The current distribution on the patch and ground plane is shown in Figure 22.12.

These results provide physical insights into the PIFA behavior. As expected, different current distribution features are obtained at different frequencies. At the GSM 1800 central frequency (1795 MHz), the more intense currents flow around the patch borders.

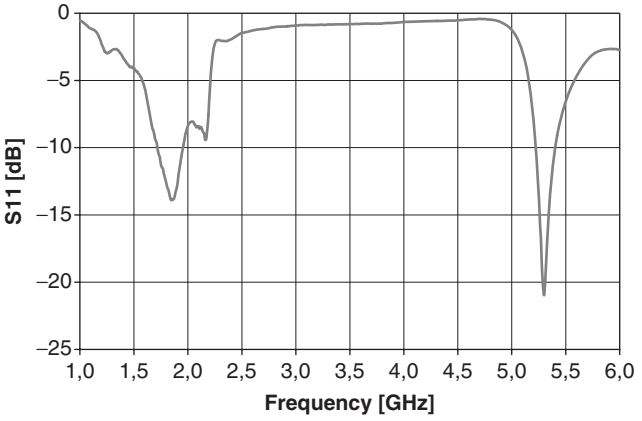


Figure 22.11 Triple-band PIFA experimental S_{11}.

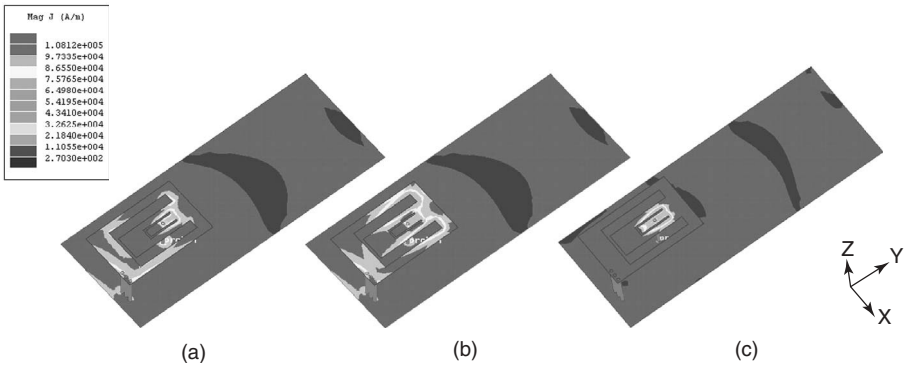


Figure 22.12 Simulation results for the current distribution on the PIFA (and ground plane): (a) 1795 MHz, (b) 2035 MHz, and (c) 5250 MHz.

TABLE 22.1 Summary of Triple-Band PIFA Radiation Pattern Experimental Results

Frequency (MHz)	Gain (dBi)	Half-Power Beamwidth (degrees)	
		E-Plane	*H*-Plane
1795	1.1	112	>180
2035	1.2	171	>180
5250	2.5	—	—

However, at the UMTS central frequency (2035 MHz), the currents are more important around the external U slot. Finally, at the IEEE 802.11a central frequency (5250 MHz), the current is very intense mainly around the inner U slot. The radiation pattern and gain of the triple-band PIFA have been measured in a far-field anechoic chamber. The results are shown in Figure 22.13.

A PIFA-like radiation pattern is obtained at 1795 and 2035 MHz. The very different radiation pattern obtained at 5250 MHz is due mainly to the currents flowing around the inner U slot and on the narrow patch that is created by the slot. Additionally, the cross-polarization level is quite high. This is usually not a problem in mobile communication applications, as usage is mainly in urban scenarios, where multiple reflections, scattering, and diffraction dramatically change the polarization of the electromagnetic waves. A summary of the radiation pattern characteristics is listed in Table 22.1.

22.3.3.2 PIFA–IFA Combination This section presents a combination of two different antennas to be used in a handset for different standards. An IFA is used for GPS (1565.19–1585.65 MHz), while a PIFA is used for PCS (824–894 MHz) and DCS (1850–1990 MHz) [5]. The corresponding configuration is shown in Figure 22.14.

The choice of two different antennas avoids the use of a triplexer or an RF switch and the inevitable associated insertion loss. This is an important feature for GPS, where the received power is only a few decibels above the thermal noise floor.

The advantage of the IFA combined with a PIFA over other GPS antenna solutions is that the IFA shares the PIFA dielectric substrate and ground plane. By doing so, a very effective and small ($26 \times 34 \times 10$ mm^3) triband antenna module is obtained. It is

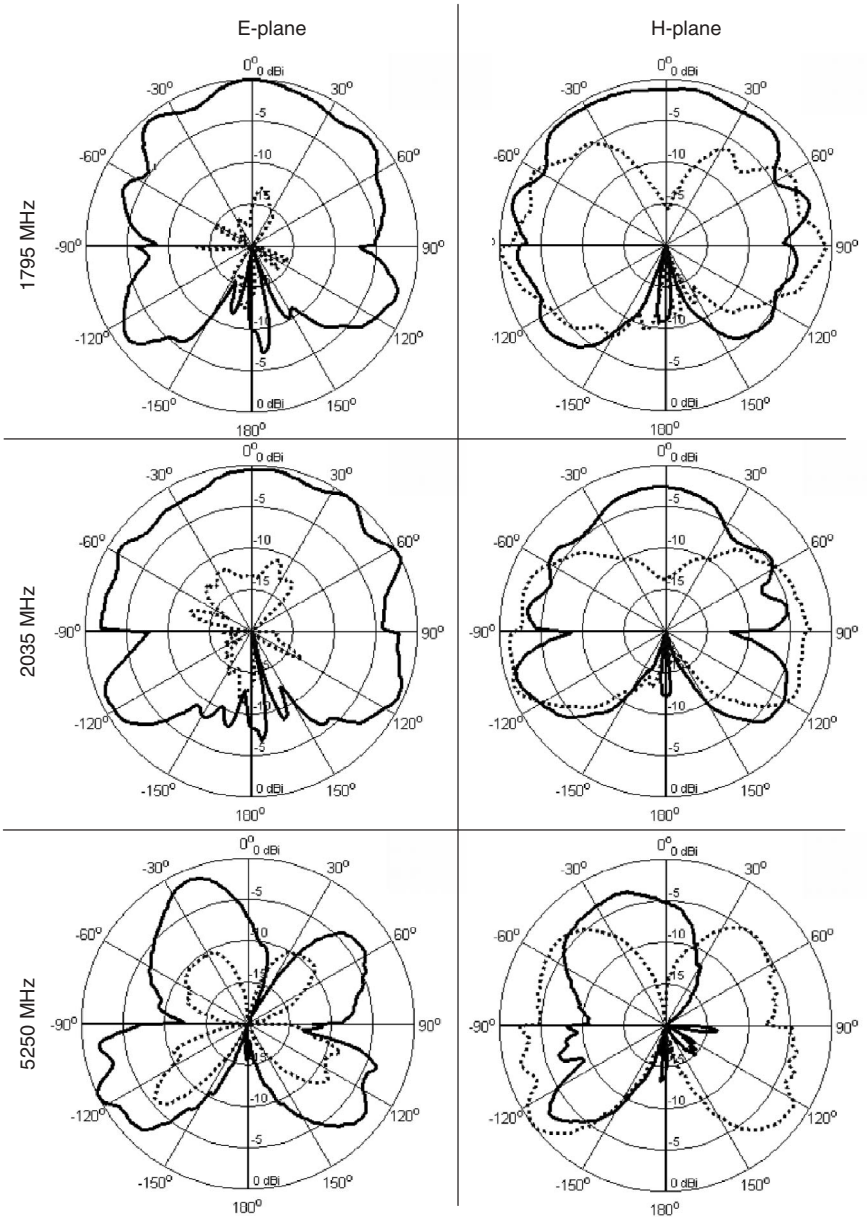


Figure 22.13 Radiation pattern (gain scale) of the triple-band PIFA: —— copolar, . . . cross-polar.

also worthy to note that the IFA produces a mainly linear polarization but right-handed circular polarization (RHCP) may be obtained in some directions if appropriate design optimization is carried out [5].

Along with its small volume, the antenna structure has also been optimized to improve the isolation between the IFA and the PIFA. Comparison of the simulation and

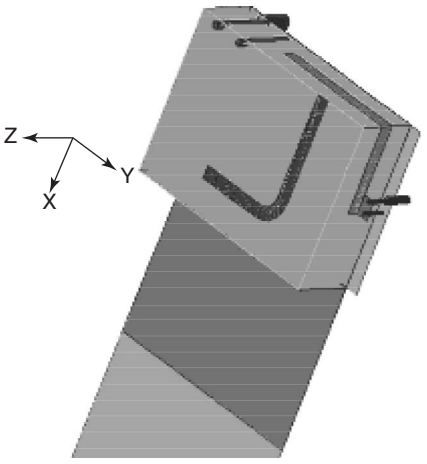


Figure 22.14 PIFA–IFA configuration.

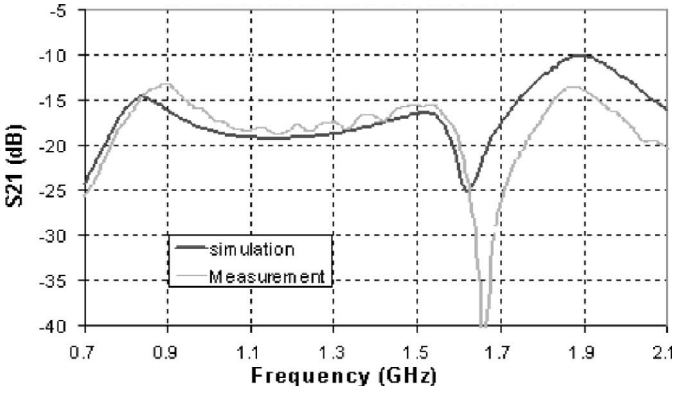


Figure 22.15 Isolation between the PIFA and the IFA.

experimental isolation results (S_{21}) between the IFA and the PIFA are shown in Figure 22.15. It is observed that S_{21} is always below -13 dB, which is remarkable for antennas so close to one another.

The radiation patterns of the IFA and PIFA are included in Section 22.3.5. A very detailed description of this IFA–PIFA combination can be found in Ref. 5.

22.3.3.3 Multielement Antenna Configurations It has been shown that multiple-input multiple-output (MIMO) wireless systems can provide increased capacity in rich multipath environments [43]. In mobile communication base stations, many antennas can be used. However, in the mobile terminals, only a few are affordable due to space limitations. In small handsets the space available for antenna integration is very scarce. With even two antennas, which is the case considered in this section, a remarkable increase in capacity can be obtained [9].

An antenna system comprised of two PIFAs integrated into a small handset is shown in Figure 22.16a. Along with the system, the antenna location in the typical spherical

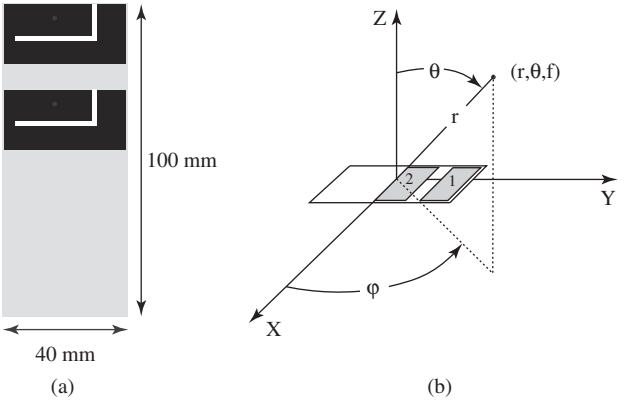


Figure 22.16 Two-PIFA configuration in a handset: (a) antenna geometry and (b) PIFAs in the spherical coordinate system.

system of coordinates is shown in Figure 22.16b. It has been developed to be used in future multistandard handsets covering 2G (GSM 1800) and 3G (UMTS) mobile communication and WLAN (IEEE 802.11b) systems. To provide the required wideband (36.9%), a multiresonance structure with an L-shaped slot is used. Moreover, to enlarge the bandwidth an air gap is used between the ground plane and the PIFA substrate [37].

Two almost similar PIFAs have been printed on top of each other in the upper part of the ground plane. To allow room for other components, especially the battery, the PIFAs have to be arranged in a very compact configuration, resulting in a distance between their adjacent edges of only 10 mm. The external dimensions of the patch and the feed point location have been optimized to provide the GSM 1800 ($f_1 = 1795$ MHz) and UMTS ($f_2 = 2035$ MHz) resonances. The L-shaped slot is intended to introduce the IEEE 802.11b resonance ($f_3 = 2442$ MHz). A radiation pattern with low directivity is also required. As usual, in mobile communication system applications, no polarization restrictions have been considered.

The simulated current distribution at the center frequency of the GSM 1800 band is shown in Figure 22.17a. A very similar result is obtained for the center frequency of the UMTS band. Figure 22.17b shows the current distribution at the center frequency of the IEEE 802.11b band. The two figures show quite different current features. While at 1795 MHz (and 2035 MHz) the currents are significant over almost the whole surface of the PIFAs, at 2442 MHz, the current is concentrated around the slots. Very weak currents are observed for both cases at the lower half of the ground plane. This is an important characteristic to minimize the effects of the user's hand.

Again for proof of concept, an antenna prototype has been fabricated and tested. Using a vector network analyzer the S parameters of the antenna prototype have been measured. The corresponding amplitude results are shown in Figure 22.18.

In this case $S_{11} \leq -6$ dB in the frequency range 1430–2760 MHz. Similarly, $S_{22} \leq -6$ dB in the frequency range 1700–2660 MHz. The highest mutual coupling (S_{12}) is -5.78 dB and is obtained in the low limit of the frequency range of interest (1710 MHz).

Assuming a propagation scenario where there is an incident field with uniform random distribution, the envelope correlation can be obtained from the S parameters of the

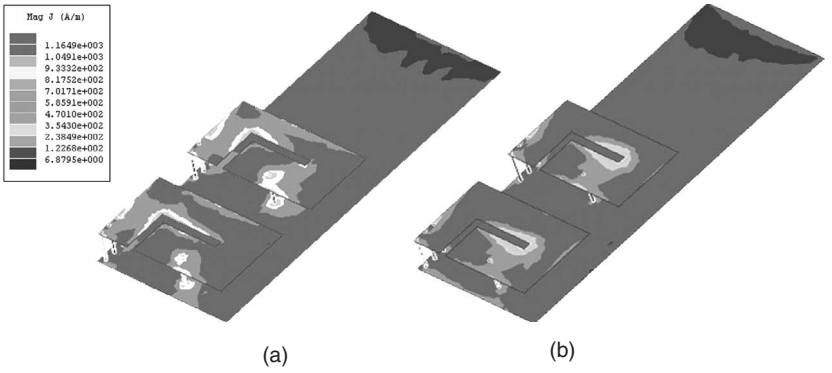


Figure 22.17 Current distribution on the PIFAs (and ground plane): (a) 1795 MHz and (b) 2442 MHz.

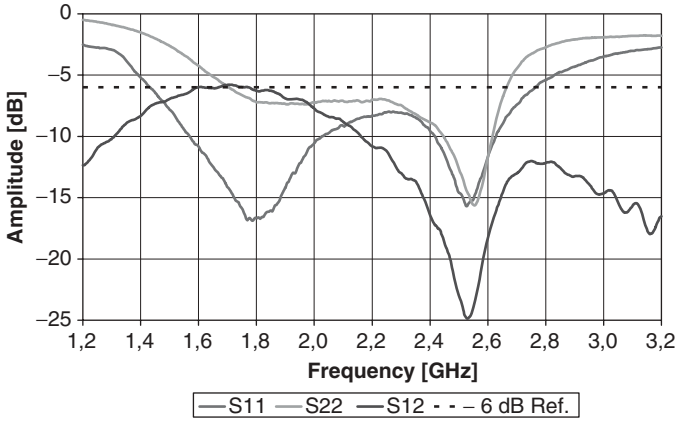


Figure 22.18 Experimental S parameter amplitude results of the two PIFAs.

antenna system as [44, 45]

$$\rho_e = \frac{|S_{11}^* S_{12} + S_{12}^* S_{22}|^2}{(1 - |S_{11}|^2 - |S_{12}|^2)(1 - |S_{22}|^2 - |S_{12}|^2)} \tag{22.12}$$

Using the above equation and the experimental S-parameter results (amplitude and phase), the experimental correlation can be calculated. Figure 22.19 shows the comparison of experimental and simulation correlation results. Except below the low frequency limit of interest (1710 MHz), the agreement is good. The experimental results confirm a low correlation (less than 0.3).

The far-field radiation pattern of the PIFA elements has also been measured in an anechoic chamber. The E-plane ($\phi = \pi/2$) and H-plane ($\phi = 0$) cuts have been measured with one element active and the other loaded with 50 Ω. Results for the center frequency of the IEEE 802.11b frequency band (2442 MHz) are shown in Figure 22.20. As expected, the radiation patterns of the two PIFAs are similar.

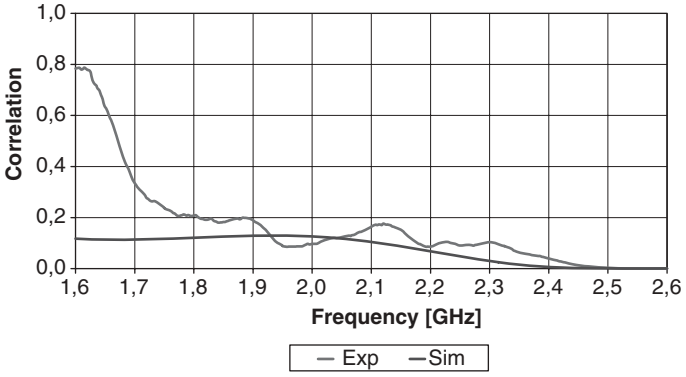


Figure 22.19 Comparison of simulation and experimental correlation results.

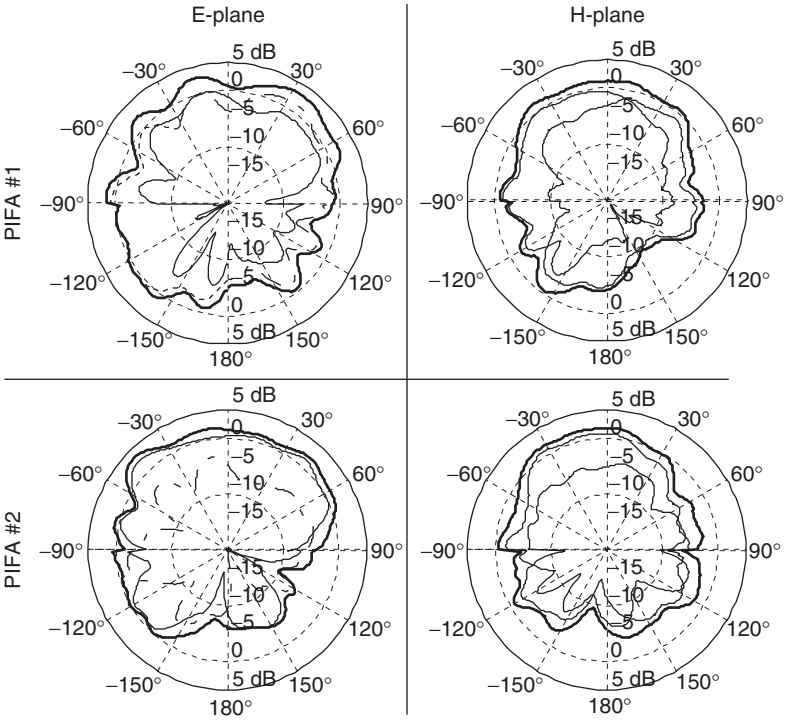


Figure 22.20 Far-field radiation patterns of PIFAs: total gain (bold solid line), copolar gain (thin solid line), and cross-polar gain (thin dashed line).

As required, very broad radiation patterns are obtained. The maximum gain is around 0 dBi. As expected, the cross-polarization level is quite high. However, as stated before, cross polarization is usually not a problem in mobile communication and WLAN systems. In fact, for most of the propagation scenarios severe multiple reflections and scattering are present and create cross polarization.

22.3.4 Antenna Housing Effects

It is well known that as long as the ground plane is large in comparison to the wavelength, the electrical performance of a printed antenna is independent of the ground plane size [46]. In fact, the ground plane can be considered infinite and image theory can be applied. However, for GSM 900, a typical 100×40 mm^2 handset is only about $(0.30 \times 0.12)\lambda_0^2$.

With such a small ground plane the antenna performance is highly dependent on the ground plane size and shape. Therefore the optimization of the antenna for a small handset has to be carried out by taking the antenna and the ground plane as a whole [47, 48]. In most cases the ground plane is predefined according to the handset size and shape and only (what was in the classical sense) the antenna can be optimized. However, the shape of the ground plane can be optimized using slots [49] if the electromagnetic compatibility with the other handset components allows for it.

The behavior of a printed antenna depends strongly on the nearby objects of the module where it is integrated. On a small handset, where size is critical, there are inevitably many other components close to the antenna. The battery, the speaker, the vibrator, and the camera are among the components that are usually integrated near the antenna. Most importantly, the battery must be analyzed as it shares the backside of the headset with the antenna. Additionally, its size and metallic structure make the battery the most influential component. Therefore an in-depth analysis of its effects is required.

In this section the effects of the ground plane size and of the battery proximity on the performance of a small wideband PIFA are analyzed.

22.3.4.1 Small Ground Plane Size Effects A small PIFA is used to evaluate the effects of the ground plane size [48]. Instead of a shorting plate, two shorting pins are used. A 10-mm air gap is used between the substrate and the ground plane to increase the bandwidth. The PIFA printed on a ground plane with variable length is shown in Figure 22.21. For lengths of the ground plane between 40 and 100 mm, the PIFA has

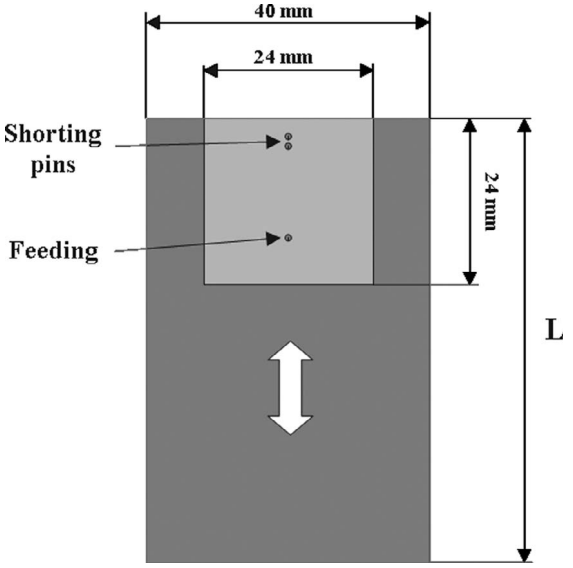


Figure 22.21 PIFA with variable ground plane length.

been matched to obtain a 2035-MHz resonance frequency (central frequency of UMTS band) and, at the same time, the largest possible bandwidth. Matching has been obtained by optimizing the position of the shorting pins and feed point without changing the size of the patch. The bandwidth and gain simulation results are shown in Figure 22.22.

In this case the bandwidth has been defined for a VSWR ≤ 2 ($S_{11} \leq -9.5$ dB) relative to 2035 MHz. From the results it can be seen that the widest bandwidth (620 MHz or 30.5%) is obtained for $L = 60$ mm, and the highest gain (2.78 dBi) is obtained for $L = 70$ mm.

An antenna prototype with the optimal 60-mm ground plane length has been fabricated and tested. Comparison of simulation and experimental input reflection coefficient (S_{11} amplitude) results is shown in Figure 22.23. Reasonable agreement is observed; the frequency shift is related to limitation of applied simulation method (infinite substrate has been assumed) and prototyping inaccuracy.

The far-field radiation pattern (at 2035 MHz) is shown in Figure 22.24. As expected, the radiation pattern exhibits very low directivity and resembles isotropic. This is an important characteristic for mobile communications applications. The maximum gain is about 1.8 dBi. The cross-polarization level is much higher than foreseen in the simulations. However, as stated before, cross polarization is usually not a problem in mobile communication systems. The main factor causing the discrepancy between simulation and

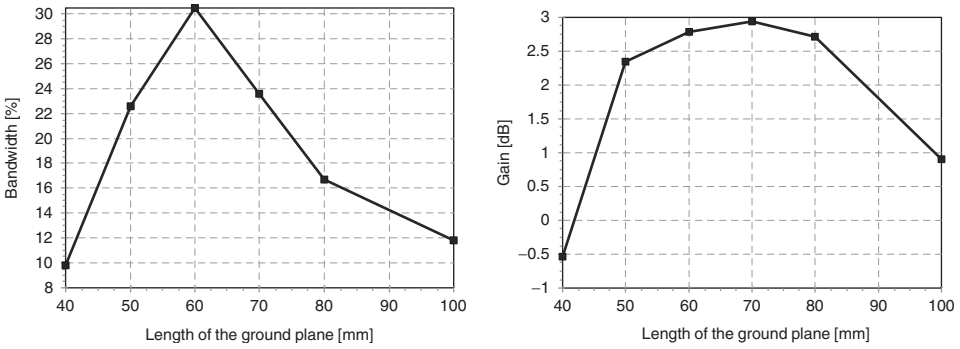


Figure 22.22 PIFA bandwidth and gain as a function of the ground plane length.

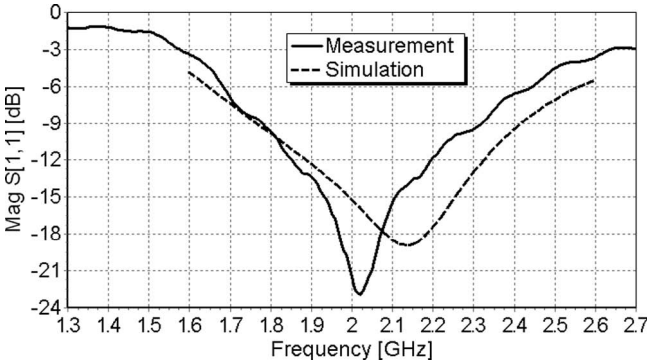


Figure 22.23 S_{11} of the optimized PIFA configuration.

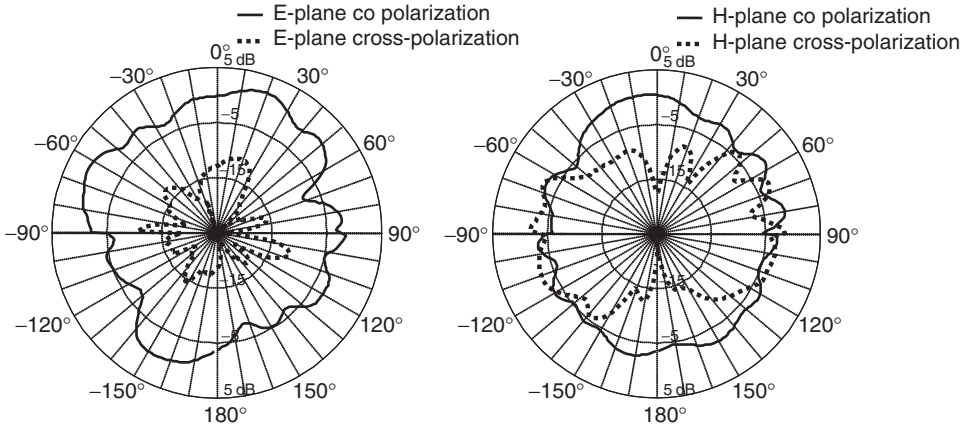


Figure 22.24 Measured radiation pattern (gain scale) of the optimized PIFA at 2035 MHz.

measurement results is radiation from the antenna setup feeding cable, not considered during the numerical studies.

22.3.4.2 Battery Effects In order to analyze the battery effects, a PIFA element described in the previous section is placed on the top edge of the ground plane (centered). This position is usually chosen to minimize the influence of the user (head and hand) on the antenna and to leave room for other devices, like the battery, for instance. In Refs. 50 and 51, the battery has been modeled as a metal box. The metal box has been implemented as a rectangular patch with a large number of shorting pins around the edges of the patch. The size of the rectangular patch has been chosen as 37×54 mm, which is the typical size of a battery used in most mobile phones. A 1.57-mm thick substrate is used on top of a 10-mm air gap. The antenna configuration, including battery, is shown in Figure 22.25. The simulations concerning battery effects on gain and bandwidth of the antenna have been made for different gaps between the PIFA and the battery model, that is, 2, 6, 10, 15, and 20 mm. The different gaps have been achieved by moving the battery model. The antenna was matched to obtain a 2035-MHz resonance frequency, which is the center frequency of the UMTS system, and the largest possible bandwidth. The matching has been done by moving the position of the shorting pins and feeding point. The patch size has remained unchanged. The simulation results for bandwidth and gain are shown in Figure 22.26.

In this case the (impedance) percentage bandwidth has been defined for a VSWR ≤ 2 ($S_{11} \leq -9.5$ dB) relative to 2035 MHz. From the results it can be seen that the widest bandwidth (360 MHz or 17.7%) and the highest gain (1.6 dBi) are obtained for the widest gap, that is, 20 mm.

For comparison purposes, the best bandwidth achieved for the same structure without the battery was 240 MHz (11.8% relative to 2035 MHz) and the corresponding gain was 0.9 dBi (please note that for a structure without battery this is not optimal ground plane size) [26]. This means that properly placing the battery in the handset can significantly increase the bandwidth and the gain of the antenna. In the present structure increases of 120 MHz (50%) in bandwidth and 0.7 dBi in gain have been obtained.

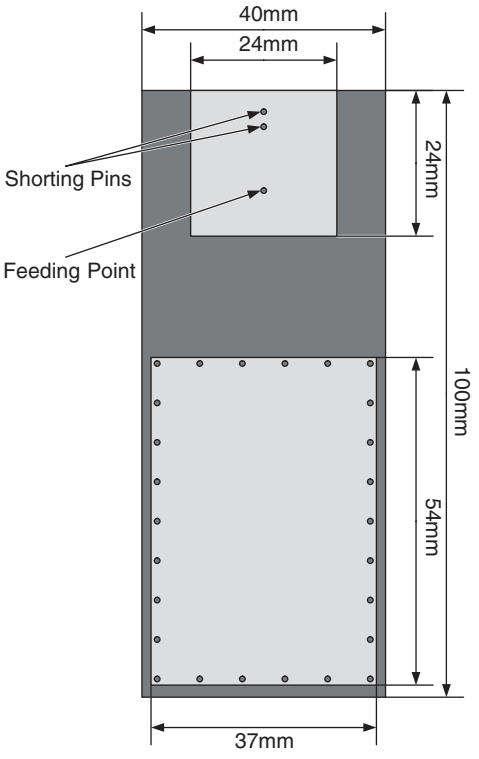


Figure 22.25 PIFA configuration with battery.

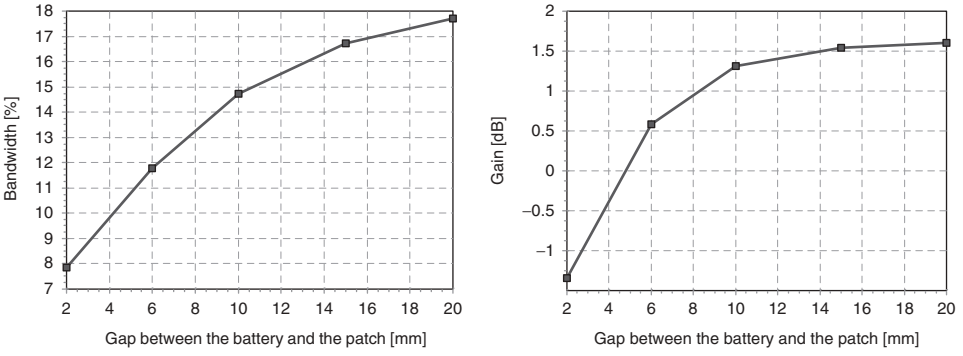


Figure 22.26 Bandwidth and gain as a function of the PIFA–battery gap.

To validate the model an antenna prototype has been fabricated and tested. Good agreement has been obtained between simulation and experimental input reflection coefficient results [50]. The far-field radiation pattern has been measured in an anechoic chamber. The results, at 2035 MHz, are shown on Figure 22.27. There is a good agreement between these experimental results and simulations [50]. The maximum experimental gain is 3.3 dBi, at 2035 MHz, for an angle of -29° in the E-plane. The cross-polarization level

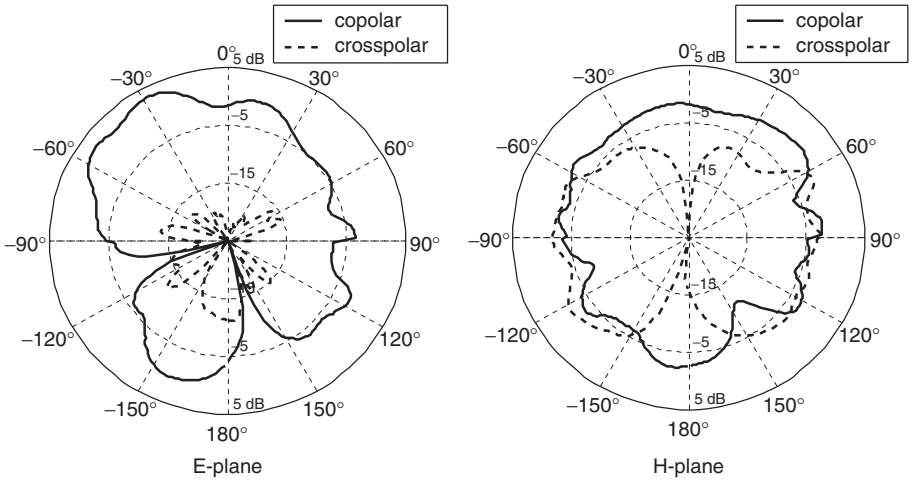


Figure 22.27 Experimental radiation pattern at 2035 MHz.

is higher than predicted by simulations. However, cross polarization is not critical in wireless communication systems. As mentioned before, the electromagnetic waves arrive at the mobile terminals after multiple reflection and scattering, which cause a significant change in the polarization. Moreover, the position in which the user holds the handset is not unique and affects the polarization mismatch. Analogously to Section 22.3.4.1 the discrepancy between simulation and measurement results is caused mainly by radiation from the antenna setup feeding cable. This factor disturbs the polarization measurement result and can lead to measured gain error (sign and value of gain shift depend on phases and amplitudes between antenna setup and cable contributions).

22.3.5 Human Interactions in Handset Antenna Design

The electromagnetic (EM) interaction between the human body and handset antennas has been an important topic in the last fifteen years. This topic is important from both the antenna performance and the human health points of view. On the one hand, the human body, specifically the head, affects antenna performance. Once the handset is put against the head, there is a degradation of the antenna impedance match and the radiation pattern and a decrease in radiation efficiency. On the other hand, human tissues are exposed to the EM radiation. This causes health concerns for users. The SAR (defined in Section 22.2) is used as an indication of how much EM radiation the human tissue absorbs. In the United States, the FCC sets the standard to be 1.6 mW/g (1.6 W/kg) averaged over 1 g of tissue [1]. In Europe, the standard is 2.0 mW/g (2.0 W/kg) averaged over 10 g of tissue [2].

The health requirements compete with the other classical specifications (bandwidth and efficiency), rendering the design of a handset antenna an even more challenging task. Ignoring the automatic power control of the handset, it could be said that an antenna with a good front-to-back ratio would reduce the head exposure to EM radiation. However, the front-to-back ratio and efficiency are defined in terms of the far field, and this makes it very difficult to relate the SAR distribution of an exposure to the structure of the incident EM fields. Therefore it is critical to understand the near-field distribution

and how it affects the SAR. In this section, the human effects on the antenna characteristics are discussed along with the evaluations of SAR for several configurations. Much research has been carried out in the development of computational techniques appropriate for antenna–human interaction modeling. Also, measurement techniques have been developed to validate the simulation results.

22.3.5.1 Effects on the Impedance Match The head absorbs certain amounts of the EM radiation and therefore can be considered as a load of the antenna. The ground plane can potentially act as an RF shield and maintain the impedance undisturbed. As an example, a handset triple-band PIFA has been used to characterize the effects of the antenna and head interactions. The antenna design has been described in Section 22.3.3.1. A photo of the prototype is shown in Figure 22.28.

The measurement system consists of a liquid tissue phantom head and hand, and a plastic holder to keep the PIFA in place against the side of the head. The head is a standard head used for SAR measurements (see Section 22.3.5.5). The PIFA has been placed in the standard "tilt" position, where the phone is tilted 15° away from the mouth. As there is no standard hand available, and no standard position for the hand, an improvised hand has been used. It consists of a rubber glove filled with the same liquid used in the head. The hand has been used in a realistic position. The experimental setup is shown in Figure 22.29.

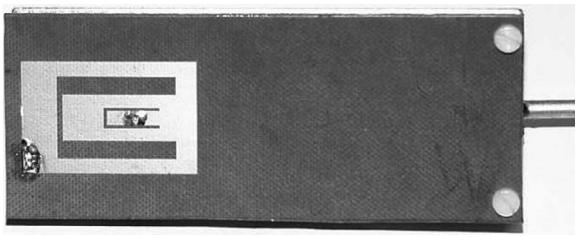


Figure 22.28 Photo of the handset triple-band PIFA prototype.

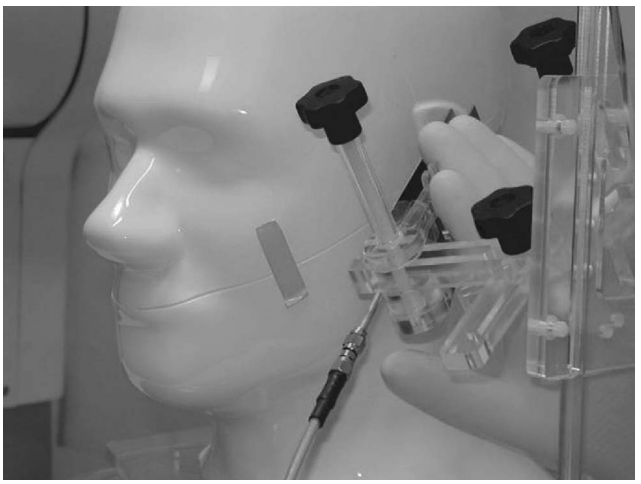


Figure 22.29 The experimental setup. (From Ref. 52, with permission.)

The input reflection coefficient of the PIFA has been measured in the following situations:

- In free space (freespace)
- In the plastic holder (holder)
- In the plastic holder near the head (holder + head)
- In the plastic holder near the head with the hand holding the handset away from the PIFA (holder + head + hand)(away)
- In the plastic holder near the head with the hand holding the handset and partially covering the PIFA (partial covering)
- In the plastic holder near the head with the hand holding the handset and completely covering the PIFA (complete covering)

The corresponding experimental results are shown in Figures 22.30 and 22.31. For each situation the label of the respective curve is shown in bold.

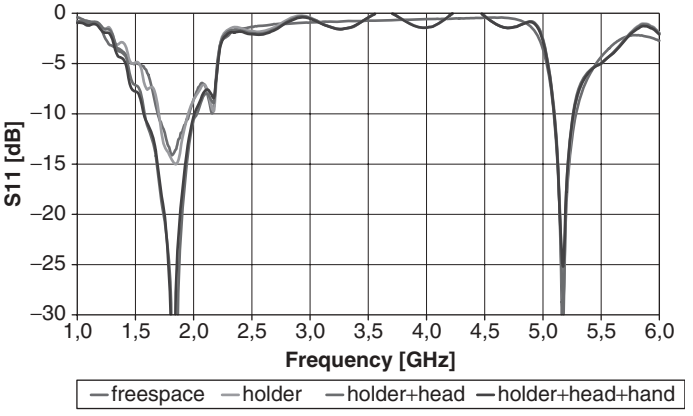


Figure 22.30 Effect of the plastic holder, head, and hand on S_{11}. (From Ref. 52, with permission.)

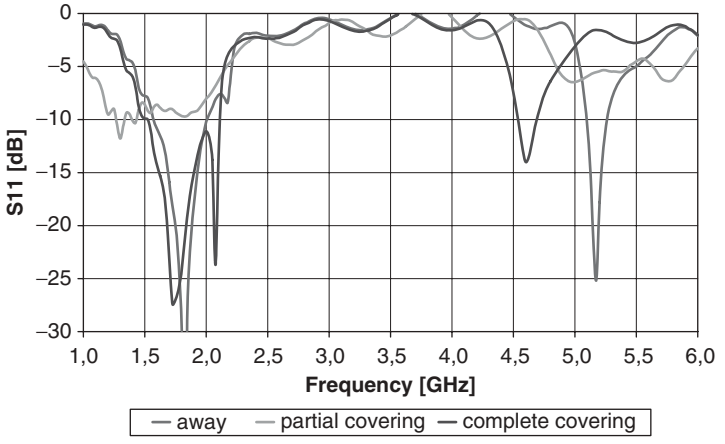


Figure 22.31 Effect of the position of the hand on $S_{11.}$ (From Ref. 52, with permission.)

As can be seen in Figure 22.30, the effects of the holder, the head, and the hand when holding the handset away from the PIFA are practically negligible. Only the matching in the lower bands (GSM1800 and UMTS) is affected. As the hand moves closer to the PIFA (Figure 22.31), the mismatch and detuning increase, reaching a maximum when the hand completely covers the PIFA. In a real handset, this proximity effect is less critical because the antenna is closed in a plastic case [52].

For yet another scenario, the return losses of both the PIFA and the IFA in a PIFA–IFA combination (see Section 22.3.3.2) are shown to demonstrate the head effect on the input impedance in Figures 22.32 and 22.33. At the 800-MHz band, the PIFA is slightly detuned and the bandwidth becomes narrower. At the 1900-MHz band and the GPS band, the impedances are almost unaffected by the phantom head.

Other antennas, such as the whip antenna or the stubby antenna, show even less effects because those antennas are external to the handset. Therefore the head will not block the antenna as much as it does in the PIFA–IFA combination. As mentioned before, the hand will also affect the impedance.

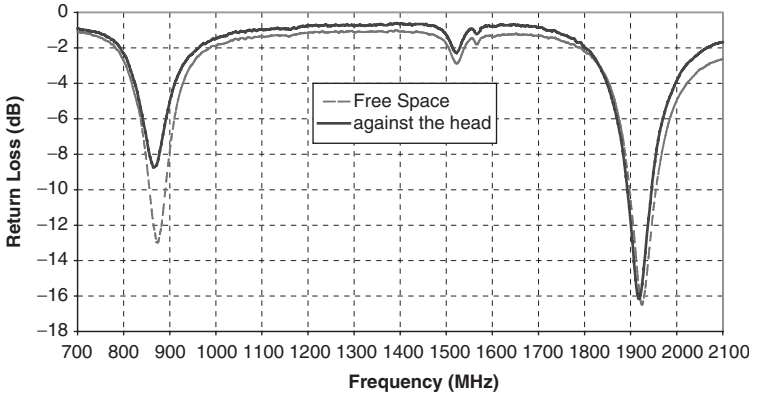


Figure 22.32 Head effect on the impedance match of the PIFA in the PIFA–IFA combination. (From Ref. 5, with permission.)

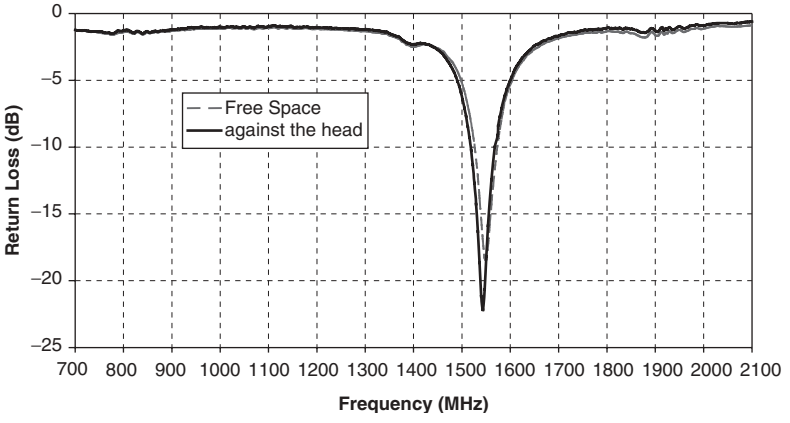


Figure 22.33 Head effect on the impedance match of the GPS IFA in the PIFA–IFA combination. (From Ref. 5, with permission.)

22.3.5.2 Head Effect on Radiation Patterns As shown in the previous section, the user's head does not significantly affect the input impedance of a handset antenna. However, as shown in Figure 22.34, the radiation pattern is greatly affected by the head. The back radiation (180°) has been strongly reduced, especially at the 800-MHz band. For the PIFA at 1900 MHz and the GPS IFA at 1575 MHz, the front radiation maintains free-space levels.

Comparison of the radiation patterns of an active whip and a parasitic whip in the presence of a head is shown in Figure 22.35. At 860 MHz, the parasitic whip shows the same behavior as the active whip, while at 1930 MHz, the pattern of the active whip still shows high back radiation toward the head. It can also be observed, by comparing Figures 22.34a and 22.35a, that, in the presence of the head, the gain of the whip, at the

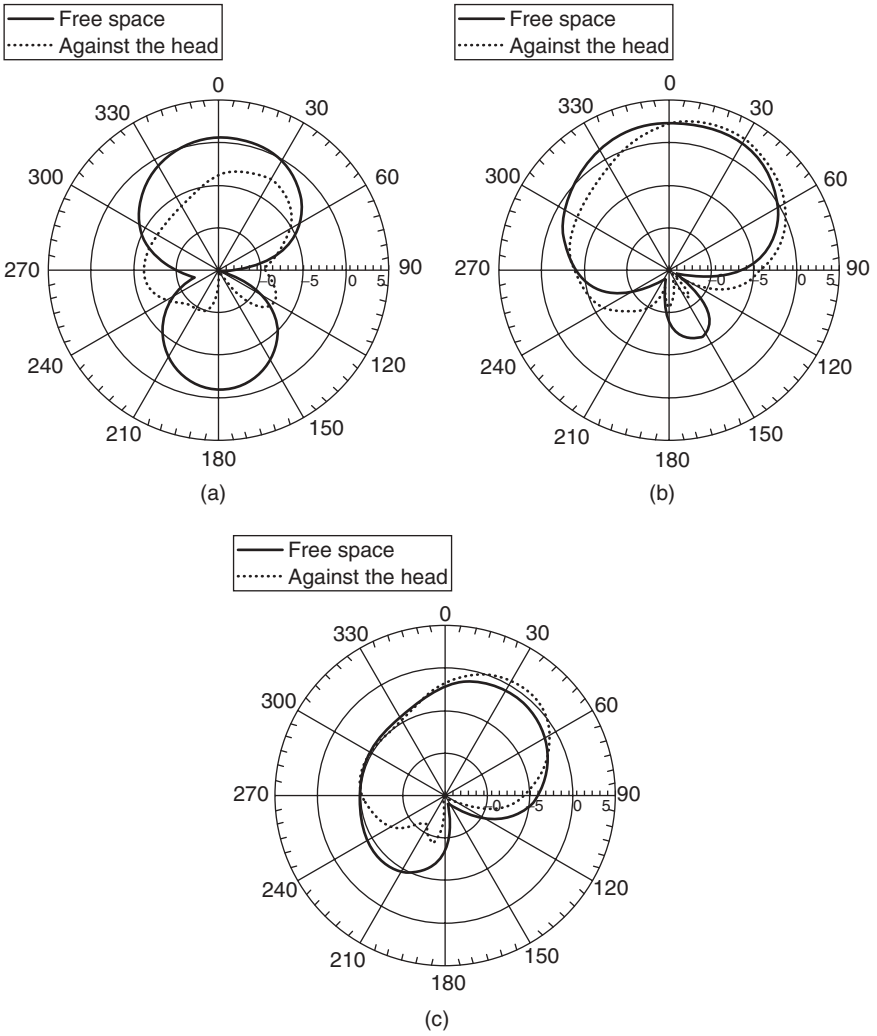


Figure 22.34 Head effect on the radiation patterns of the PIFA in a PIFA–IFA combination: (a) radiation pattern of the PIFA at 830 MHz, (b) radiation pattern of the PIFA at 1930 MHz, and (c) radiation pattern of the IFA at 1575 MHz. (From Ref. 5, with permission.)

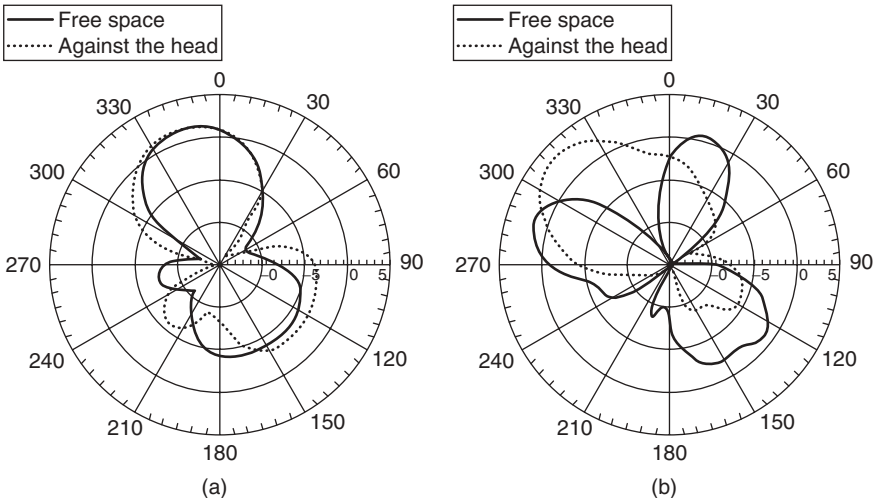


Figure 22.35 Head effect on the radiation patterns of whip antennas. Comparison of the active whip and the parasitic whip: (a) at 860 MHz and (b) at 1930 MHz. (From Ref. 5, with permission.)

800-MHz band, is about 3 dB above the gain of the PIFA. This is the major reason why the whip is still used as a handset antenna.

22.3.5.3 Head Effect on Antenna Efficiencies When the handset is held closely to the head, some of the radiated power will be absorbed by the head. It was shown that the efficiency of a PIFA could drop as much as 20–25% at the 800-MHz band, and 15–20% at the 1900-MHz band. The efficiency of an IFA dropped about 15–30% [10, 53].

Although the reduction of the antenna efficiency caused by the head cannot be avoided, some methods can be applied to minimize that unwanted effect [53–56].

22.3.5.4 SAR Characterization Clearly based on Eq. (22.8), SAR is a point quantity, and its value varies from one location to another. The accuracy and reliability of a given SAR value would depend on three parameters (tissue density, conductivity, and the electric field), but the most significant one of the three parameters is the induced electric field. In the antenna design, the target is to reduce the electric field in the head. The induced electric field is a complex function of frequency, handset size, the distance between the handset and the head, polarization, and the biological variables such as the tissue type. The typical ingredients of the head model are listed in Table 22.2 [53].

SAR can be obtained from detailed numerical computations [57]. Many different tools have been developed to simulate SAR. Finite-difference time-domain (FDTD) methods [58] are very suitable for SAR calculations because they allow antennas to be modeled in their true operating environment [10]. Depending on the human head model and the antenna model generated by the simulation tool, the calculated SAR values can be quite accurate. When it comes to a real handset, it is still difficult to simulate all of the features. So the numerical modeling of a handset for SAR can be costly, and it can take as long as several weeks due to the large amount of computations. Consequently, accurate SAR

TABLE 22.2 Typical Values of Tissue Parameters (Relative Permittivity ε_r, Effective Conductivity σ, and Density ρ)

	900 MHz		1800 MHz		
Tissue	ε_r	σ (S/m)	ε_r	σ (S/m)	ρ (kg/m^3)
Cartilage	42.65	0.782	40.21	1.287	1100
Muscle	55.95	0.969	54.44	1.389	1050
Eye	55.27	1.167	53.57	1.602	1020
Brain	45.8	0.766	43.54	1.153	1040
Dry skin	41.4	0.867	38.87	1.184	1090
Skull (bone)	16.62	0.242	15.56	0.432	1645

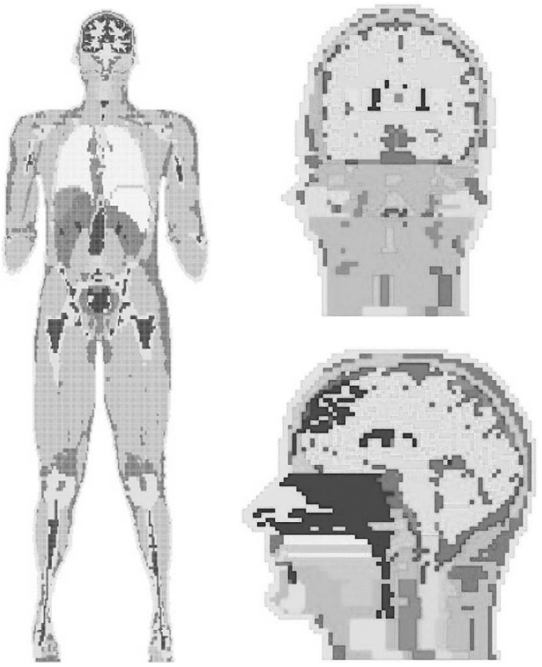


Figure 22.36 Representative human and head models.

values are still very difficult to achieve without measurement. In order to perform these simulations, representative human and head models (shown in Figure 22.36) are utilized.

22.3.5.5 SAR Measurement SAR distribution can be measured directly using body/head phantoms, robot arms, and associated test equipment. The Dissymmetric Assessment System (DASY) 4 robotic system as shown in Figure 22.37 is used to measure the actual SAR. By using conventional SAR measurements, one single test (one position and one frequency) takes about 15 minutes. To complete the head SAR measurement of one dual-band handset, it will take roughly one week. Given the large amount of reflection and scattering that a human body causes in the near field, it is

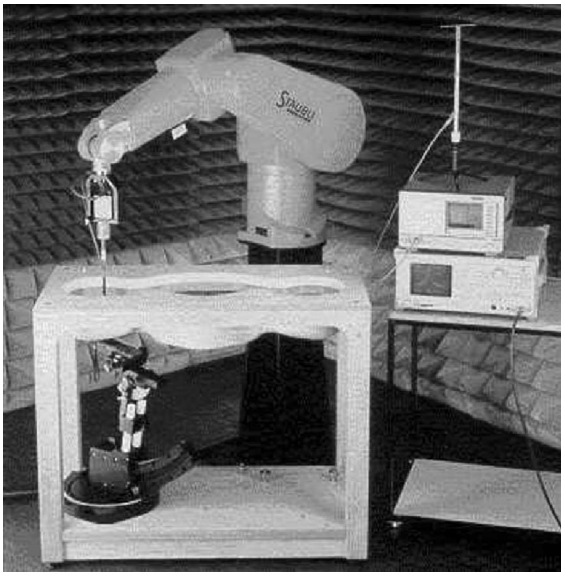


Figure 22.37 DASY 4 SAR measurement system. The electric probe controlled by the robot can reach into the fluid (head or body simulant) to measure the E-field inside the fluid. (From Ref. 5, with permission.)

TABLE 22.3 Ingredients of the Test Fluid for the FCC Head SAR Measurement

800 MHz Band		1900 MHz Band	
Ingredient	Head (% by weight)	Ingredient	Head (% by weight)
Deionized water	51.07	Deionized water	54.88
HEC	0.23	Butyl diglycol	44.91
Sugar	47.31	Salt	0.21
Preservative	0.24		
Salt	1.15		

typically necessary to measure the RF power deposited inside the tissue rather than the external incident EM fields. In the DASY (FCC compliant for conducting SAR tests), the SAR phantom is filled with the simulated human tissue fluid. Ingredients and the electric parameters of the fluid are listed in Tables 22.3 and 22.4. The robot-controlled electric probe is programmed to measure the electric field in volts per meter (V/m) inside the SAR fluid during the operation. The handset under test is required to be performed at the maximum output power, which is intended to represent the worst-case scenario.

In the head SAR measurement, the handset is usually measured at two positions, namely, the cheek position and the tilt position as shown in Figure 22.38. In the cheek position, the handset is pressed against the head with both the ear and the cheek touching the handset. In the tilt position the handset is rotated 15° from the cheek position. These two positions represent the two most typical user positions. Since the head can have an impact on the impedance of the antenna, the actual curve of the handset front cover

TABLE 22.4 Electric Parameters of the Head Tissue Simulant

	Head Tissue Simulant			
		Dielectric Parameters		
f (MHz)	Description	ε_r	σ (S/m)	Temperature (°C)
836.5	Recommended value	41.5	0.9	21
	+5% Window	39.4–43.6	0.86–0.95	N/A
1880	Recommended value	40	1.4	21
	+5% Window	38–42	1.33–1.47	N/A

TABLE 22.5 SAR Measurement Summary

		GSM 1800 (1750 MHz)		UMTS (1950 MHz)	
Antenna	Position	1-g Average SAR (W/kg)	10-g Average SAR (W/kg)	1-g Average SAR (W/kg)	10-g Average SAR (W/kg)
U-slotted PIFA	Cheek	0.715	0.463	0.584	0.366
U-slotted PIFA	Tilt	0.381	0.257	0.339	0.214

can be carefully designed to give the optimum SAR performance. However, antenna designers are usually not involved in the mechanical design of the handset. As a result, some other methods have to be applied for SAR reduction.

The measurement uncertainties are defined in the IEEE P1528 specification. Overall uncertainties must be below 30% for a 95% confidence level. An uncertainty in measurements of 30% may seem a little high, but it is small in decibels.

22.3.5.6 SAR Results for Representative Antennas The SAR measurement of some representative antennas has been reported in Ref. 66. The antennas were held in the standard "cheek" and "tilt" positions against the liquid tissue phantom head, filled with an appropriate liquid tissue simulant. A special isotropic field probe on the end of a robot arm was then used to measure the SAR within the volume of the head.

Figure 22.39 shows the SAR of the PIFA antenna being measured. The results are presented in Figures 22.40 and 22.41 [66].

The PIFA antenna exhibited relatively low SAR due to its directional radiation pattern and the shielding effect of its ground plane, as well as the larger distance between antenna and head that was required to accommodate the coaxial feed.

The results of the SAR measurements are summarized in Table 22.5. This provides information on the averages over 1 g and 10 g. It is the 10-g average that is used within the ICNIRP guidelines for exposure to electromagnetic fields [3], with the limit being set at 2 W/kg in the head and torso. All the measurements here fall well within this limit, although it should be noted that for the worst-case uncertainty an upper bound of 1.4 W/kg for the 10-g average is often used, and they all also meet this stricter requirement.

22.3.5.7 SAR and Efficiency versus Antenna Distance The preceding results for radiation efficiency and peak SAR in the head have all been provided for a given

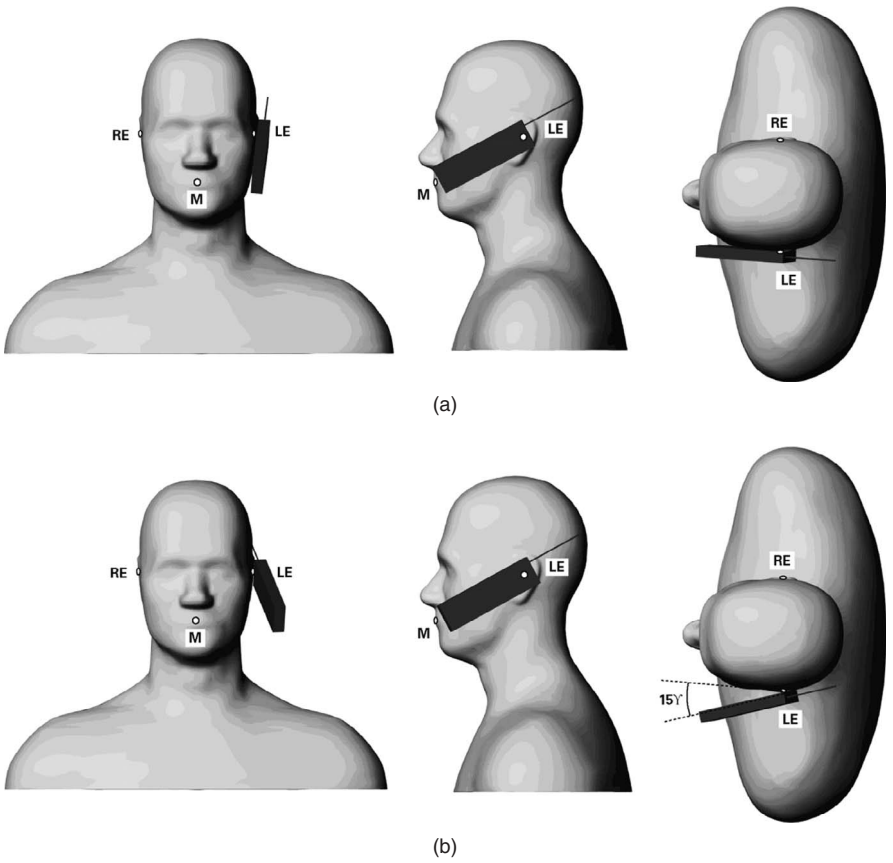


Figure 22.38 SAR test positions. (a) In the "touching cheek" position, the phone is angled from mouth to ear with the center of the phone's speaker aligned to the ear reference point (left ear (LE) or right ear (RE)). (b) In the "15° tilt" position, the phone remains on the mouth-to-ear angle (as for the "touching cheek" position), but the microphone end is rotated away from the head by 15°.

separation between the head and the handset. However, it is interesting to examine the effect of this distance on these parameters. Figure 22.42 presents the simulation results for the antenna efficiency and peak SAR (1-W delivered power) in the anatomical head model versus the distance, *b*, for the monopole and the back-mounted (internal) PIFA configurations with the handset upright. As might be expected, the radiation efficiency increases with distance, while the peak SAR decreases in a nearly exponential fashion. Similar results were also simulated for spherical head with different radii. Results are shown in Figure 22.43.

22.4 INTEGRATED ANTENNAS FOR LAPTOPS

Since the first portable personal computers were introduced back in the 1980s, their capabilities have grown tremendously. Today's laptops (also called notebook computers) are

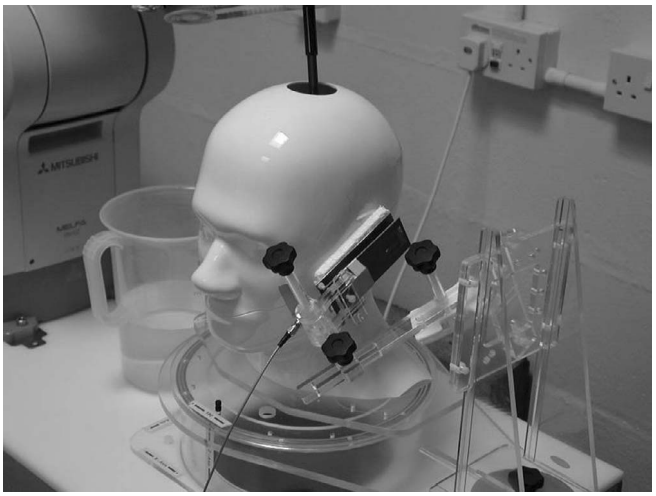


Figure 22.39 SAR measurement of U-slotted PIFA antenna.

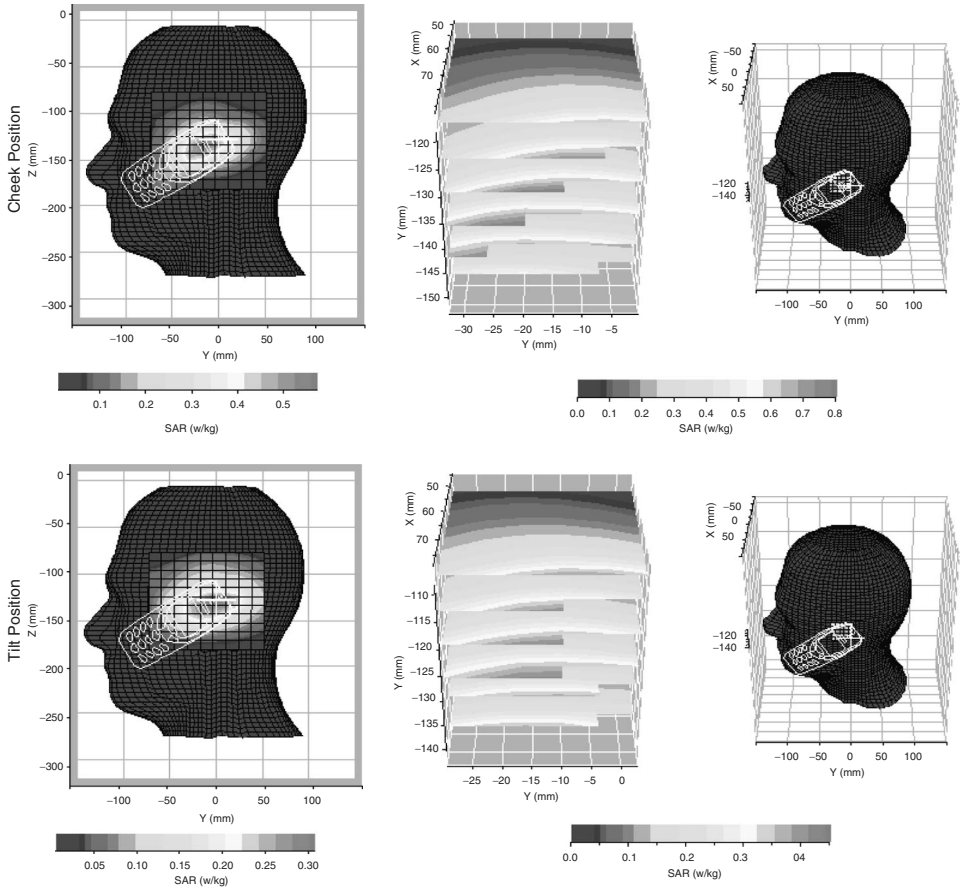


Figure 22.40 SAR from U-slotted PIFA at 1750 MHz.

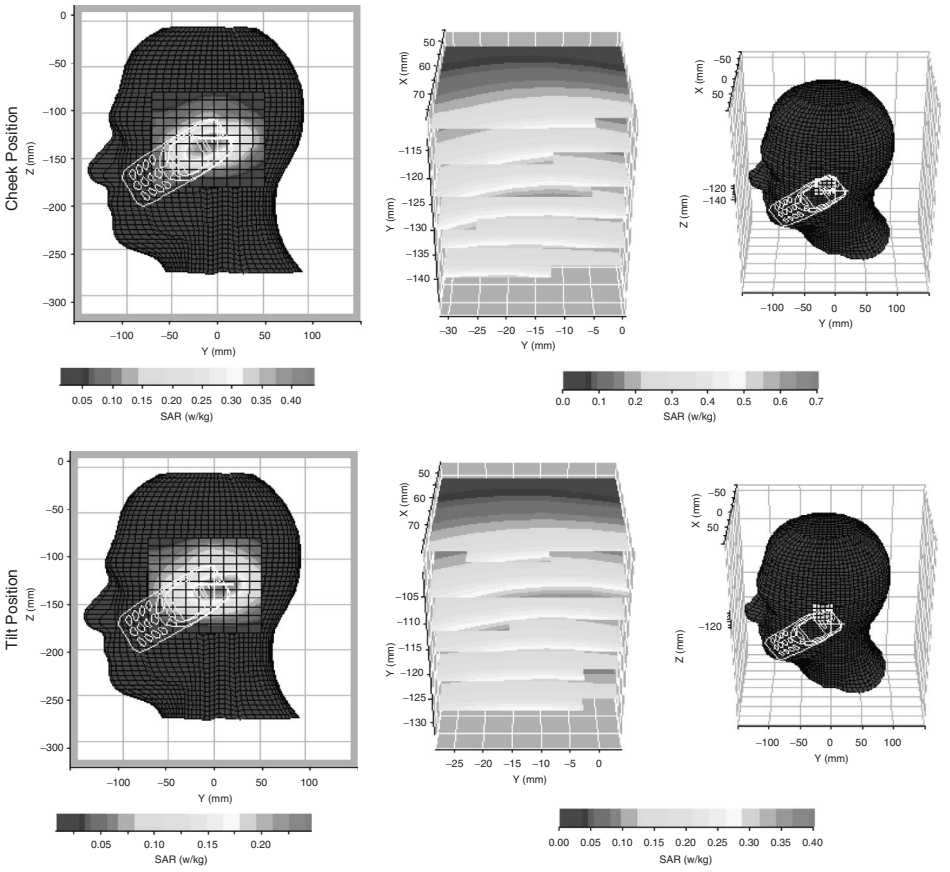


Figure 22.41 SAR from U-slotted PIFA at 1950 MHz.

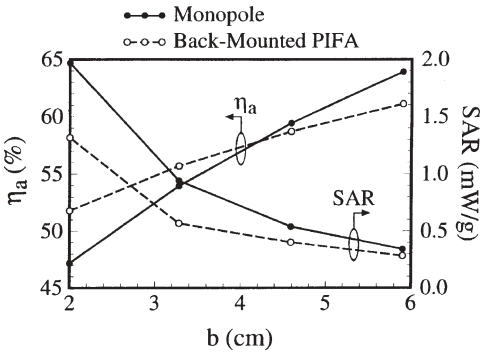


Figure 22.42 Simulated antenna efficiency and peak SAR (1-W delivered power) in the anatomical head model versus the distance b for the monopole and the back-mounted (internal) PIFA configurations with the handset upright.

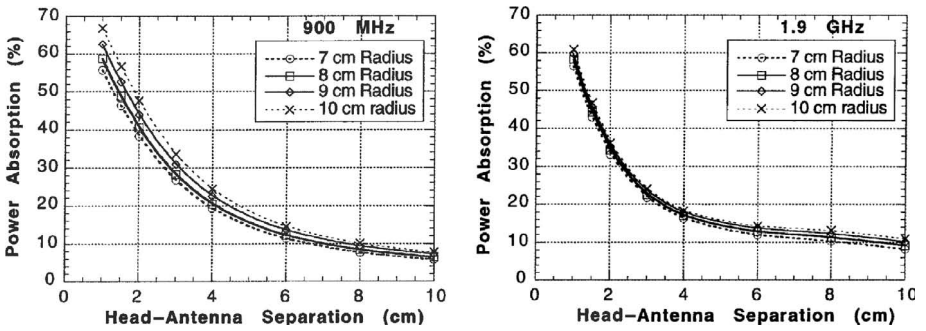


Figure 22.43 Simulated antenna efficiency versus the distance to the spherical head model for different head radii.

complex multifunctional devices offering high computational power, high storage capacity, and sophisticated audiovisual interfaces. Moreover, to fulfill needs of the "Information Society," the laptop's communication capabilities have steadily been expanding. Aiming at the portability, wireless interfaces are widely used to connect with peripheral devices [59] (wireless personal area networks, WPANs) and other computers [60] (wireless local area networks, WLANs). The integration of cellular network radios into some modern laptops gives users access to the Internet in areas not covered by WLANs. Ultimately, in the not too distant future, introduction of digital video broadcasting (DVB) television receivers built into portable computers is expected.

Laptop functionality expansion drives the evolution of integrated antennas. Moreover, the required laptop miniaturization and aesthetics impose additional design constraints for antenna engineers. Therefore a variety of new antennas dedicated to laptop computers has been investigated [61, 62]. Those designs have to meet certain specifications, which can be classified into system-related and integration-related requirements.

System-Related Laptop Antenna Design Requirements

- *Sufficient Impedance Match.* For laptop antennas operating in the transmit mode, this guarantees the reflection of RF power from the antenna port stays on an acceptably low level, which affects the system efficiency and therefore the battery lifetime. In the receive mode, good antenna impedance match improves the receiver sensitivity.
- *Multiband Operation.* Due to the high scale of integration, a single antenna element is often used to operate in more than one wireless system. It must provide a sufficient impedance match over several frequency bands. Some typical examples include dual-band 2.4 GHz/5.2 GHz [64, 79] and triple-band 2.4 GHz/5.2 GHz/5.7 GHz WLAN antennas [72, 82].
- *Omnidirectional Radiation Pattern.* This is the most adequate type of radiation pattern for laptop applications. It provides reliable wireless connectivity, independently of the terminal orientation. As laptop computers are often used in a horizontal position, the horizontal radiation pattern is most critical.
- *Antenna Polarization.* This is usually not a critical parameter for laptop applications, since laptops are used primarily in indoor environments, where there is intensive reflection and scattering [62].

Integration-Related Laptop Antenna Design Requirements

- *Miniaturization.* Although it seems that miniaturization is of much less importance than in the case of handsets, laptop antennas have to be integrated within very densely packed electronic devices, where there is little room for additional functions. The size, shape, and location of the antenna may be affected by other design constraints such as mechanical and industrial design [62]. The integration may be particularly challenging in new, very small ultraportable laptops and when multiple antenna arrangements are considered.
- *Aesthetics.* This is an important factor for consumers; therefore the antenna should not break the laptop sleek design lines. Generally, mechanically weak retractable antennas are substituted by internal antennas, invisible to the user.
- *Low Profile.* To minimize radiation from very high speed electronics, today's laptop computers are equipped with conducting covers or metallic shields just inside the plastic covers [62]. This condition, imposed by FCC emission requirements, significantly affects the antenna design. In particular, the metallic surroundings limit the thickness of the available space for an antenna and shields its radiation.

The design of a laptop integrated antenna, which meets all the above requirements, is a challenging task. The final antenna performance depends not only on the structure of the radiating element, but also on the antenna location within the laptop (see Section 22.4.3). The initial integration of wireless sub-systems into laptops used PC cards inserted into PC slots. This solution allows RF engineers to design an external component as a whole, without interfering with the already existing laptop's hardware. An inserted card contains the controller, the RF front-end, and an integrated antenna. The retractable or internal antenna is located in the extreme part of the card, protruding from the side of the portable computer (see Figure 22.44). With the use of wireless cards, users can easily add functionality to their laptops. Due to this convenience, external plug-in wireless interfaces are still very popular; some recent models are even housed in very compact universal service bus (USB) dongles (Figure 22.44). Antennas integrated in external interfaces, denoted here as external integrated laptop antennas, are described in Section 22.4.1.

As wireless technology becomes prevalent and less expensive, manufacturers are moving away from PC cards in favor of integrated implementations [62]. This trend goes with a general tendency of making portable computers more compact and sleek, as all the external and protruding components are being eliminated in order to facilitate the

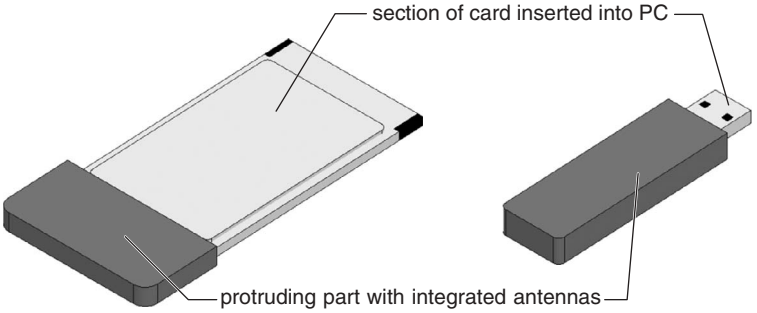


Figure 22.44 Wireless interfaces housed in PCMCIA card and USB dongle.

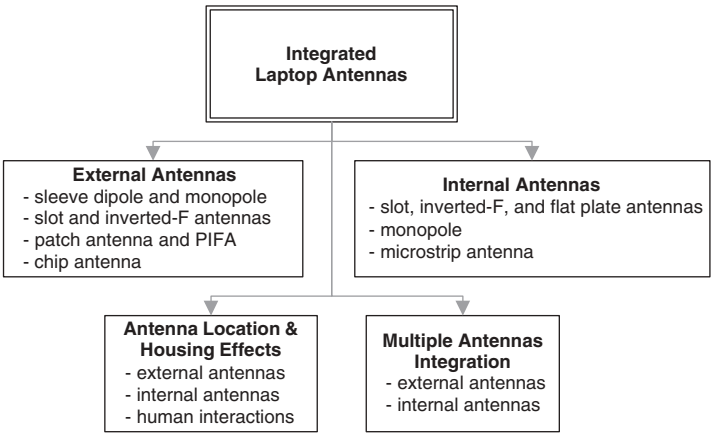


Figure 22.45 Flowchart of topics presented in Section 22.4.

laptop's use and reduce the risk of physical breakage. Integration of wireless communication devices into laptops requires a cooperation of antenna engineers with system designers. This is evident as the integrated antenna is an internal part of the laptop and has to fit inside a very limited volume, while the entire PC structure acts as a part of the antenna and contributes to the radiation performance. Therefore the internal laptop antenna location is a key issue and is discussed in Section 22.4.3.2. Typical built-in radiator structures are described in Section 22.4.2.

Today's and future laptop antenna systems are not limited to single element solutions. Two antenna diversity schemes are already used in vendor built-in and on-card products. In the near future, the achievement of multiple-input multiple-output (MIMO) techniques will be applied in laptops. In that case, the design of multiple antenna systems will be done under new constraints. The challenges and design methodologies of multiple antenna arrangements for laptops are discussed in Section 22.4.4. The flowchart of the topics presented in Section 22.4 is shown in Figure 22.45.

22.4.1 Overview of Laptop External Antennas

Antenna types integrated into laptop plug-in wireless interfaces are shown in Figure 22.46. In this chapter, the integration of antennas into personal computer memory card international associate (PCMCIA) cards is considered, although those antenna types are also used in smaller USB dongles.

22.4.1.1 Sleeve Dipole and Monopole Antenna The sleeve dipole (Figure 22.46a) has been used in the initial implementations of laptop wireless cards. It is a modification of the dipole antenna, where a center feed has been substituted by an end-feed method. The coaxial sleeve dipole cross section is shown in Figure 22.47a.

Its radiating structure is an asymmetric dipole made of conductors of different diameters and slightly different lengths. The thinner conductor, typically the extension of the coaxial feeding inner conductor, must have an appropriate length to achieve good antenna matching in the band of operation. The large diameter conductor sleeve must provide effective choking of the RF currents at its open end and at one-half of the radiating

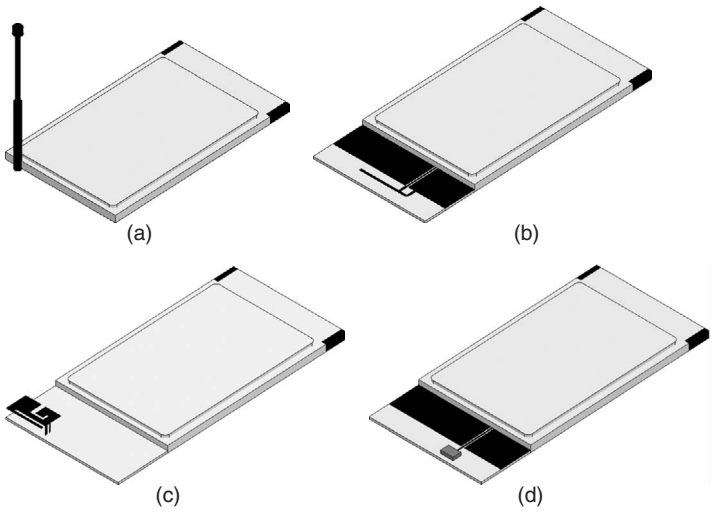


Figure 22.46 Integrated antennas into PCMCIA cards: (a) retractable sleeve dipole, (b) CPW-fed printed IFA [69], (c) triple-band PIFA [72], and (d) CPW-fed chip antenna [77].

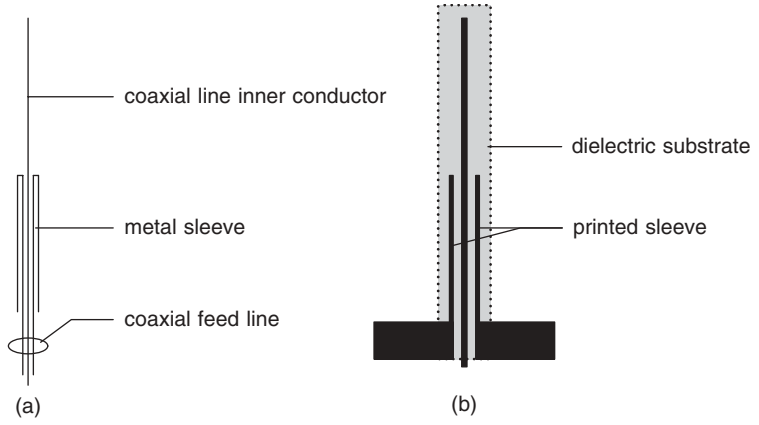


Figure 22.47 Sleeve dipole: (a) cross section of the coaxial version and (b) printed version.

dipole [19]. A sleeve dipole can be realized in a planar structure as a strip sleeve dipole (Figure 22.47b). Multiband strip sleeve dipoles have been reported in Refs. 63 and 64.

The sleeve dipole provides a narrower bandwidth than a simple dipole but is more robust for housing effects. However, during operation, its radiating segment has to be placed as far as possible from other laptop components. Therefore it has to be protruded from a card as a retractable or whip antenna (see Figure 22.46a). This antenna can also be mounted on the top of the laptop screen [64]. A sleeve dipole alone provides an omnidirectional radiation pattern adequate for mobile unit applications. The effects of antenna housing on radiation performance are discussed in Section 22.4.3.1.

A monopole antenna structure can be printed on a circuit board. Printed meander monopoles [20] integrated into a USB dongle [65] have been used as a laptop internal

antenna [66]. For PCMCIA card applications, a branched monopole antenna (also called F-shaped monopole antenna) [20] and a straight-F antenna [67] can also be applied.

22.4.1.2 Slot Antenna and Inverted-F Antenna (IFA) An IFA printed on a PCMCIA board is shown in Figure 22.46b. Its structure is shown in Figure 22.48b. The quarter-wavelength arm is parallel to the ground plane edge, which makes the IFA easy to integrate within a limited space. Essentially it is half of the traditional $\lambda/2$ slot antenna (Figure 22.48a) and their mechanisms of operation are analogous. By moving the feeding stub from the shorting stub to the open slot end, the IFA input impedance changes from very low to very high values. The feeding point is selected in order to obtain an impedance match to the 50-Ω line. The slot height is calculated according to the required antenna frequency band [68] and, for a 3.4% wide ISM 2.4-GHz band, ranges between 5 and 6 millimeters.

The IFA antenna can be printed on the protruding part of the PCMCIA board [69] and fed by a coplanar waveguide, as shown in Figure 22.46b. In multiple antenna arrangements (Section 22.4.4), where the miniaturization is even more important, the IFA arm can be bent or meandered [67]. The IFA element is also widely used as an internal antenna, integrated along the laptop's screen edge [62]. This application and future IFA developments are described in Section 22.4.2.1.

The slot antenna can be considered as a loaded version of the IFA, where the load is a quarter-wavelength stub. Since the quarter-wavelength stub itself is a narrowband system, the slot antenna has narrow bandwidth [62]. Classical slot antennas, due to bigger dimensions than IFAs, are implemented as laptop internal antennas [70]. In wireless plug-in interfaces, tapered meander slot antennas can be used [71].

22.4.1.3 Patch Antenna and Planar Inverted-F Antenna (PIFA) A PIFA is a modification of a half-wavelength long microstrip patch antenna. By using shorting walls or shorting pins, the antenna element is miniaturized. In a classical single-band PIFA, the patch element length is approximately a quarter wavelength. By cutting slots in the radiating patch, the current resonant path can be modified, thus allowing additional miniaturization. Moreover, a careful design of the patch shape and of the feed location and the use of shorting pins can result in the coexistence of several resonant paths, therefore producing a multiband operation. As an example of this, a triple-band, compact PIFA [72] integrated into a PCMCIA card is shown in Figure 22.46c.

A PIFA used as a microstrip component inherently requires a ground plane under the patch element. The antenna height above the ground plane determines the operation bandwidth, and in order to fulfill the specifications, it cannot be lower than a certain value. In the case of the triple-band PIFA example [72], the antenna height is 4 mm.

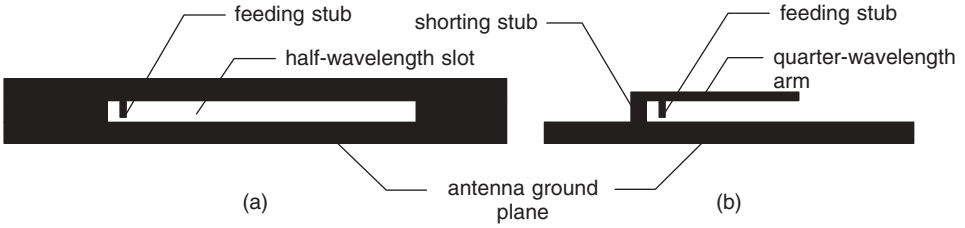


Figure 22.48 (a) Slot antenna and (b) inverted-F antenna structures.

Due to such a low profile structure, a PIFA is very robust against the effects of nearby components or antennas [72]. Consequently, it can easily be integrated into the protruded part of the PCMCIA card or inside the laptop case [73].

A simple microstrip patch antenna can be integrated into the backside of a laptop screen (see Section 22.4.2.3), but it is rather too big to be integrated inside a PCMCIA card. For those applications, double C-patch antennas have been proposed [74]. Recently, E-shaped patch antennas [75] have been used successfully in the two-element, wideband, and compact diversity arrangements for a PCMCIA platform [76].

22.4.1.4 Chip Antenna The chip antenna is a very compact surface mountable device (Figure 22.46d). It is comprised of a high permittivity dielectric body ($\varepsilon_r > 7$) with an embedded meandering metal line. In many practical implementations, low temperature cofired ceramic (LTCC) technology is used, meaning that conducting strips are printed on different ceramic layers and are connected by metal via posts forming a continuous three-dimensional (3D) path. The path shape depends on the application and required miniaturization and may take the form of a helix [20], meander [20], or spiral [77]. An example of rectangular 3D spiral chip antenna structure is shown in Figure 22.49.

There are two major types of ceramic chip antennas. One has a ground plane printed on the bottom of the ceramic base, and the other does not have a ground plane. The first one, more versatile for practical applications, has a narrower impedance bandwidth and lower radiation efficiency [20]. Therefore in the majority of today's designs the chip antenna is mounted on a circuit board portion without the underlying ground plane, as shown in Figure 22.49. The distance between the antenna elements and the ground plane affects the impedance bandwidth [78].

Thanks to the very small size and ability to be surface mounted, chip antennas are used in laptop wireless interfaces in both external plug-in devices and built-in antennas [78]. Due to the high permittivity of the ceramic substrates, the electromagnetic field is concentrated in this low loss part of the radiator. Therefore a chip antenna can achieve an acceptable radiation efficiency even when mounted on a high loss circuit board.

In this section, the most common antenna configurations used for laptop plug-in interfaces are described. In many practical applications the outlined structures are modified according to particular design constraints. The characteristics of an antenna integrated

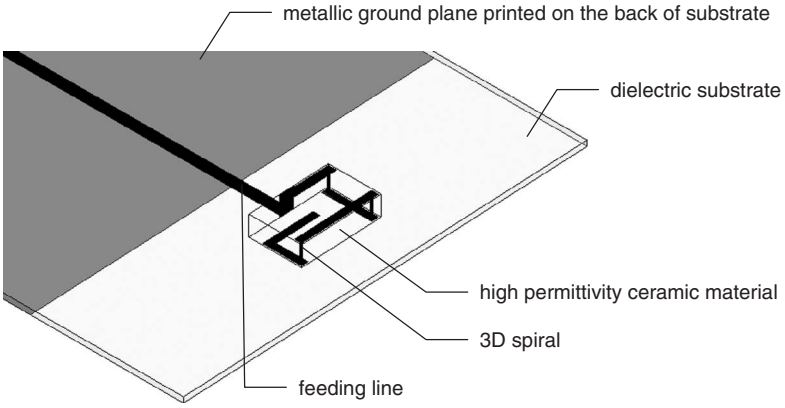


Figure 22.49 A 2.4/5.2-GHz dual-band chip antenna integrated on a dielectric board. (From Ref. 77.)

into a laptop system, in the case of both a plug-in card and an internal antenna, depend strongly on its location, distance to nearby components, and materials covering the device. The housing effects and antenna location are discussed in Section 22.4.3.

22.4.2 Overview of Laptop Internal Antennas

The majority of today's laptop computers are equipped with built-in wireless interfaces and do not require external WLAN/PAN plug-in cards. The user can rarely notice where the antenna element is located. Typically, the radiator is integrated within a thin space between the laptop's shielding layers and plastic case. The internal antenna size limitations are different from PCMCIA cards and USB dongle and depend strongly on location. Usually the most demanding design constraint is the built-in antenna thickness. For antennas integrated inside the screen rim, today's thin and large displays impose severe restriction on the antenna height.

22.4.2.1 Slot Antenna, Inverted-F Antenna (IFA), and Flat Plate Antenna

The structures of a slot antenna and an IFA have been described in Section 22.4.1.2. The IFA element is widely used as an internal antenna, integrated along the laptop screen edge [61, 62, 68, 79, 80, 83]. For this application, it can be cut from a thin metal sheet and fed by a miniaturized coaxial cable, as shown in Figure 22.50.

When mounted on a laptop screen edge, the IFA possesses good omnidirectional properties, as required for a mobile terminal application, and it also provides a fairly broad impedance bandwidth. Its properties, however, depend on the location as described in Section 22.4.3.2. Numerous modifications have been made to the IFA, leading to dual-band [79–81] and triple-band designs [82]. If the antenna height must be miniaturized, slots may be inserted in the ground plane as discussed previously [83].

The single-band, flat plate antenna [84] has a similar configuration to the IFA. It can also be cut from a thin metal sheet and fed by a miniaturized coaxial cable. The dual-band

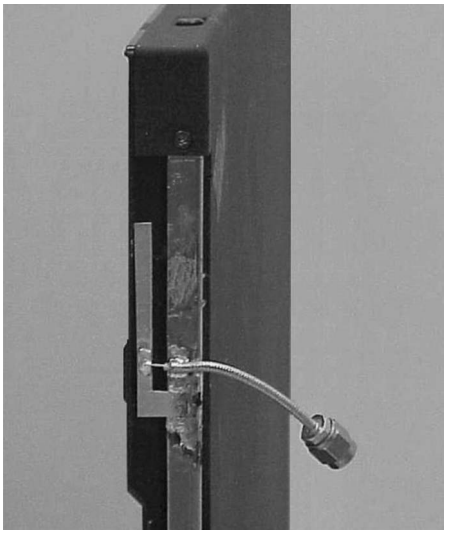


Figure 22.50 A 2.4-GHz inverted-F antenna integrated into a laptop prototype. (From Ref. 61, with permission © 2002 IEEE.)

flat plate antenna arm may have a more convoluted shape. Two examples of integration into a laptop screen rim have been described in Refs. 85 and 86.

The classical slot antenna occupies a larger volume than the IFA and usually possesses a narrower bandwidth. However, a slot structure has been proposed for dual-band laptop antennas integrated into a screen case [70].

22.4.2.2 Monopole Antenna A simple quarter-wavelength monopole is rarely used as an internal laptop antenna due to its relatively large size. However, several modifications have been proposed in order to use monopoles in laptop built-in applications [87]. A compact printed tab monopole, for instance, can achieve a broad 50% operation bandwidth and can be integrated on a dielectric board by being protruded over a laptop display panel as shown in Figure 22.51. For lower bandwidth requirements, it can be substituted by a single-band meander monopole or a dual-band branched monopole [66]. If a very low resonant frequency with relatively high bandwidth is required, as when used for digital television (DTV) reception, a U-shaped metal-plate monopole can be utilized [88].

22.4.2.3 Microstrip Antenna As mentioned in the introduction to Section 22.4, today's laptop computers are equipped with conducting covers or metallic shields just inside the plastic covers [62]. This condition, demanded by FCC emission requirements, has a significant effect on the antenna design. In particular, the metallic surround limits the thickness of the space available and shields the radiation of the antenna. Microstrip antennas seem to overcome the mechanical design constraints described above. They inherently need a ground plane and due to their conformal properties and low profile structure can easily be integrated between the metal shielding layers and the plastic cover.

A simple microstrip antenna can be integrated in the backside of a laptop screen within a thin volume between the TFT display panel and the plastic case (shown in Figures 22.52a and 22.53a). Unfortunately, when a simple patch antenna is mounted over a relatively large ground plane (constituted by the display panel metallic frame), the antenna predominantly radiates toward the back hemisphere of the display (front-to-back

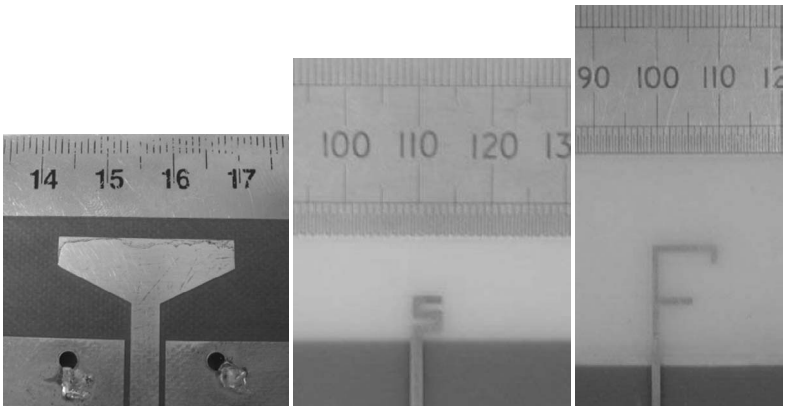


Figure 22.51 Printed monopole antennas: (a) wideband tab monopole, (b) miniature meander monopole, and (c) dual-band branched monopole. (Parts (b) and (c) from Ref. 66, with permission.)

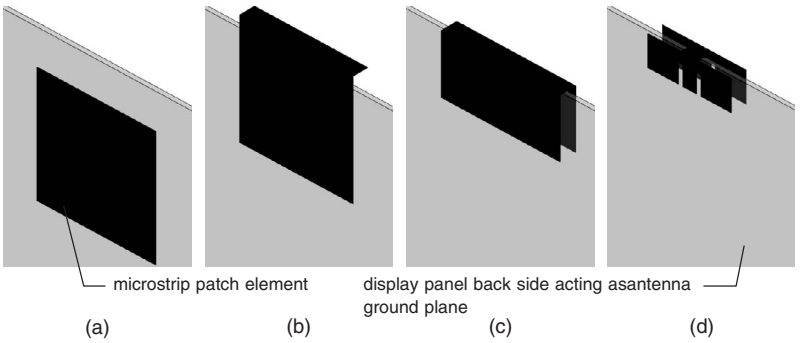


Figure 22.52 Evolution of microstrip patch antennas for laptops: (a) simple half-wavelength patch mounted on the laptop display backside, (b) L-bent protruded patch antenna (From Ref. 89), (c) omnidirectional wrapped patch antenna (From Ref. 90), and (d) dual-band back-to-back E-shaped antenna (From Ref. 91).

ratio above 15 dB). Its radiation pattern (Figure 22.55a) is far from omnidirectional, which is not adequate for mobile applications. In order to balance forward and backward radiation, the patch antenna can be protruded above the ground plane edge (Figure 22.52b). Moreover, by adding a bent section its bandwidth can be expanded [89]. This solution can reduce the front-to-back ratio to 2.7 dB. Further front-to-back ratio reduction, to values as low as 0.5 dB, can be achieved with the wrapped patch configuration [90] presented in Figure 22.52c. In this case the patch element conformally embraces the ground plane edge without needing the extended protrusion, allowing a more compact design. A miniaturized dual-band antenna with fairly omnidirectional total gain radiation patterns in both operating bands (Figure 22.55b,c) can be obtained when a wrapped rectangular patch is substituted by a back-to-back E-shaped patch [91]. This antenna structure and prototype are shown in Figures 22.52d and 22.53b, respectively, and its measured reflection coefficient in Figure 22.54.

For antenna integration inside the laptop keyboard, where size constraints are even more pronounced, a PIFA element can be used [73]. However, this antenna location does not provide an omnidirectional radiation pattern [62, 92], see Section 22.4.3.

22.4.3 Antenna Location and Housing Effects

In Sections 22.4.1 and 22.4.2, antenna elements used in laptop external and internal wireless interfaces have been described. Their characteristics, namely, radiation pattern and radiation efficiency, depend on the radiator location within the laptop [62, 92] and on the antenna position with respect to the dielectric and metallic surrounding structures [62, 93]. The overall built-in system performance is also affected by the distance between the RF front-end and the antenna, as a 0.5-m section of miniaturized coaxial cable operating at 5 GHz may introduce a 3-dB loss. Therefore the antenna location plays a critical role in the laptop wireless interface performance and has to be designed with special care. Additionally, for some antennas, the input matching may depend on the antenna location; therefore the element should be tuned after integration within the laptop. In this section, the characteristics of typical antenna locations for plug-in and built-in interfaces are discussed. Representative locations are depicted in Figure 22.56.

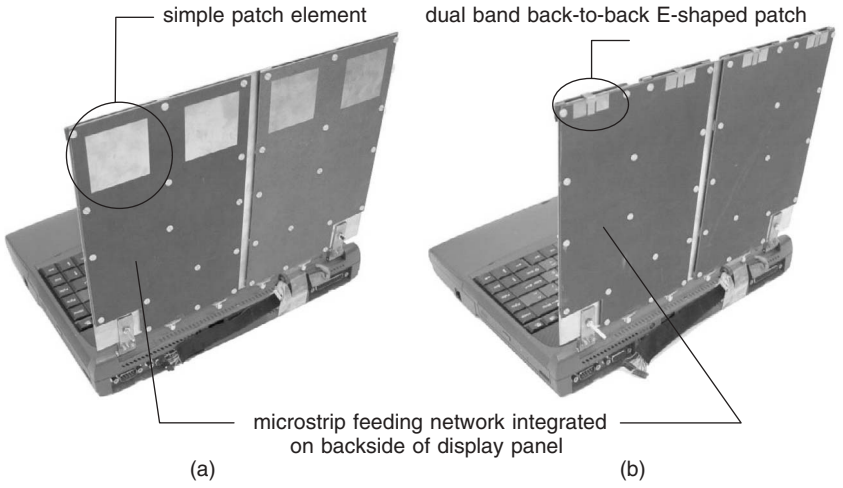


Figure 22.53 Microstrip antenna configurations integrated into a laptop prototype: (a) traditional 2.4-GHz rectangular patches and (b) quasi-omnidirectional dual-band 2.4/5.2-GHz back-to-back E-shaped patches. (From Ref. 96).

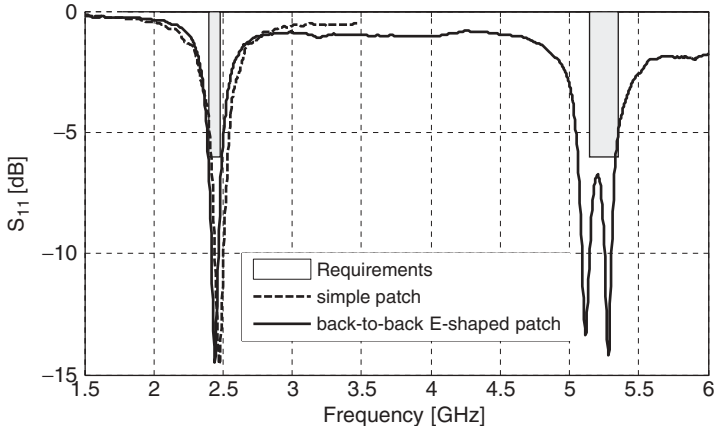


Figure 22.54 Measured reflection coefficient of a simple patch and a back-to-back E-shaped patch integrated into a laptop prototype.

22.4.3.1 External Antennas In the external plug-in wireless interface design, the choice of antenna location is very limited and does not allow laptop-independent performance. Usually the antenna is integrated in the most distant location from the laptop in order to reduce the effect of the laptop on the communication performance. The mechanisms that affect the on-card antenna performance are discussed next.

Effect of Nearby Components The antenna performance is particularly affected by the metal and lossy plastic laptop case. It is also dependent on the antenna separation from the laptop body. An experimental study has shown that by extending the default

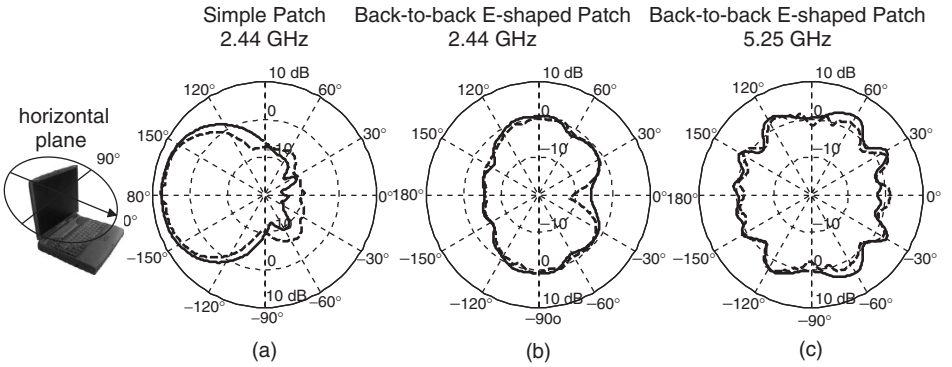


Figure 22.55 Horizontal plane (total gain) radiation patterns of a simple patch and a back-to-back E-shaped patch integrated into a laptop prototype (——— simulation, ------ measurement).

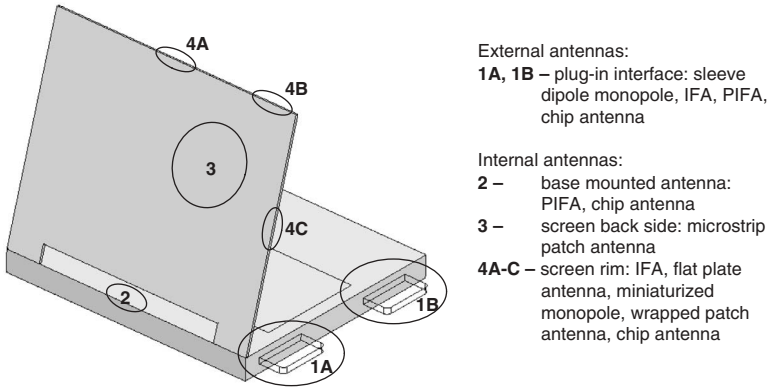


Figure 22.56 Common locations and corresponding applicable antenna types in a laptop (for simplicity, and due to physical symmetry, only left-side locations are shown).

PCMCIA card protrusion by 6 mm, the antenna sensitivity may be improved by almost 6 dB, which corresponds to over 60% range expansion [62]. The radiation efficiency of an antenna integrated inside the plug-in card also depends strongly on the dielectric losses inside the printed circuit board and device plastic case.

Effects of Laptop Base and Lid Parts These components, which in the microwave regime may be approximated by metal boxes [92], have dimensions of several wavelengths. On the one hand, they act as an expansion of the antenna ground plane, and on the other hand, they reflect and shield the antenna radiation. As a consequence, the far-field radiation pattern of an antenna integrated into a freestanding PCMCIA card is different from the one obtained when the card is inserted into the PC slot. Moreover, the pattern will be dependent on the PC slot location (see positions 1A and 1B in Figure 22.56). The variations in total gain radiation pattern when a retractable sleeve dipole and an integrated IFA are mounted in a PCMCIA card are shown in Figures 22.57 and 22.58, respectively. In those examples, the laptop opening angle is $\psi = 90^{\circ}$.

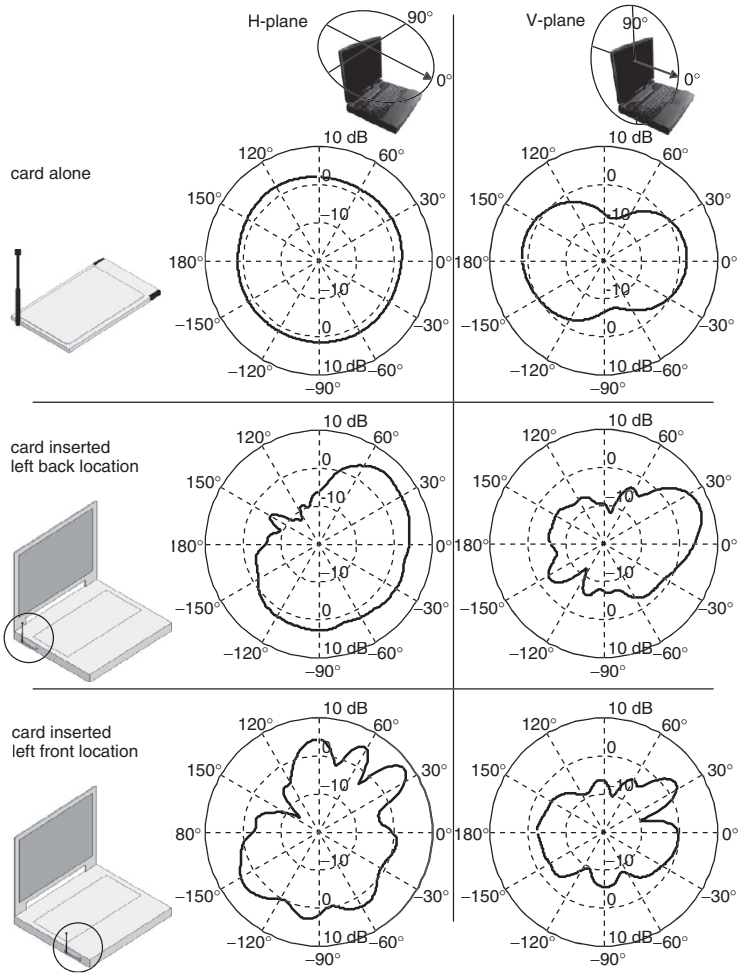


Figure 22.57 Sleeve dipole mounted on PCMCIA card: computed gain far-field radiation pattern at 2.44 GHz.

The sleeve dipole antenna possesses an omnidirectional radiation pattern when mounted vertically on a relatively small PCMCIA card (Figure 22.57, first row). When attached to the PC, the far-field radiation pattern is clearly modified and the horizontal plane differs significantly from omnidirectional. This is due to the laptop's screen, as its shielding effect is clearly visible for azimuth angles between 90° and 180°. The reflection from the screen contributes to an enhanced power radiation in the azimuth range 0°–90°. As the distance between the antenna and the screen changes (PCMCIA card is moved from location 1A to 1B), the interference pattern visible in this angular range also changes. This phenomenon can be explained intuitively with the aid of a virtual antenna array constituted by the sleeve dipole and its image representing the reflection from the screen. When the antennas' spacing increases (from below one-half wavelength in 1A to over two wavelengths in 1B), several lobes appear in the radiation pattern.

However, the vertical radiation pattern, of less importance for laptop applications, is also affected by the presence of the laptop structure.

The horizontal plane radiation pattern of an IFA element integrated into a freestanding PCMCIA card is less omnidirectional than the one obtained for the sleeve dipole (compare first rows of Figure 22.57 and 22.58).

The IFA element is, however, more robust to the influence of the laptop structure. When the antenna is mounted in the plane of the keyboard base (below the screen), the display panel shielding and reflection effects are not as pronounced as in the previous example. Other antenna types integrated into the card circuit board, like chip antennas, possess the same advantage. However, the effects of the laptop structure on the antenna performance also depend on the antenna dominant polarization and have to be evaluated separately for each antenna type.

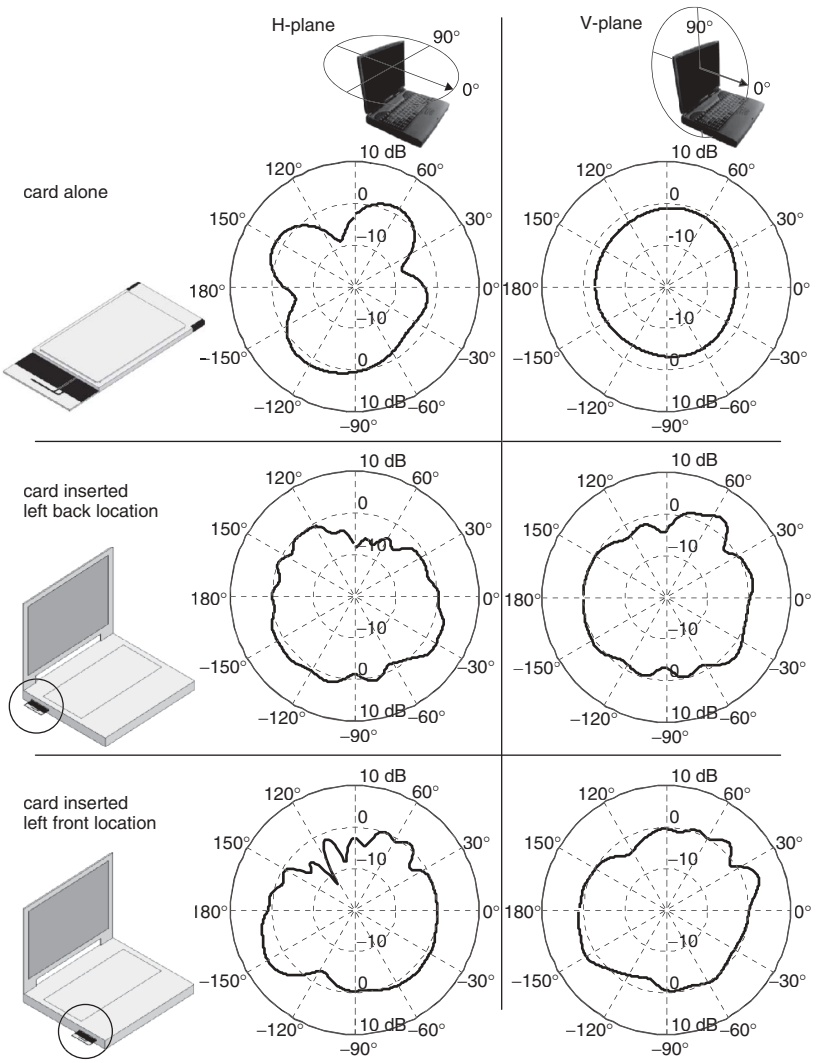


Figure 22.58 Printed IFA mounted on a PCMCIA card: computed gain at 2.44 GHz.

External Environment Influence The common disadvantage of on-card mounted (locations 1A, 1B) and keyboard base-mounted (location 2) antennas is the external environment influence, such as a metal desk and/or the user, on the antenna performance [94]. A metal desk may significantly shift the tuning of the antenna and create unwanted reflections that change the radiation pattern. The absorption of electromagnetic energy by the laptop user's hands and lap can have a dramatic effect on the antenna gain [62]. These effects are much less pronounced for antennas integrated inside the laptop screen (locations 3, 4A–C), which preselect those as the most beneficial antenna locations [94].

22.4.3.2 Internal Antennas The designer of an internal laptop wireless subsystem has a broader choice of antenna location than in the case of a PCMCIA card. Moreover, as in the design stage, the entire laptop structure is already known and the housing effects can be better predicted and taken into account. The possible antenna locations can be evaluated independently of the antenna element type with the aid of the electromagnetic visibility study (EVS) technique [92]. This method is based on illuminating the laptop structure with a plane wave incident from a variable angle and monitoring the excited surface current density at the points of interest. The best antenna locations are those where the surface current density is the highest and the least dependent on the illuminating wave incidence angle. The best results are obtained for location 4A (see Figure 22.56), next in ranking are locations 4B, 4C, and 3 [92, 94]. Locations 1A, 1B, and 2 have low EVS rankings.

The EVS results show that the locations on top of the display lead to an antenna performance almost similar to free space [94]. Therefore, for those positions, the influence of the laptop structure on the performance of an integrated antenna is minimized. It is impossible, however, to completely eliminate laptop structure effects. The mechanisms that affect the built-in antenna performance are discussed next.

Effects of Nearby Components These effects are critical for the overall antenna system performance and have to be considered at the antenna design stage. The metal laptop components near the antenna element play the role of a ground plane and affect the achievable bandwidth. The plastic cover layers modify the effective permittivity in the antenna neighborhood and therefore affect the antenna physical size. Moreover, the dielectric losses in the plastic cover degrade the radiation efficiency; when the laptop case consists of a very lossy material (e.g., carbon-fiber reinforced plastic—CFRP) it may be necessary to design a special "RF window" for the antenna [70].

Effects of Laptop Base and Lid The display panel and its frame basically act as a ground plane for antennas in locations 3 and 4A–C. For microstrip antennas integrated on the screen backside (location 3), the display size and the antenna location, with respect to its edges, determine the ratio between energy radiated in the back and in the front of the screen. However, for antennas integrated on the screen rim (IFA, flat plate antenna, modified monopole antenna, wrapped patch antenna) in locations 4A–C, the display size does not directly affect the radiation pattern. For those elements, the effect of the laptop base is much more pronounced.

The effect of the keyboard base on the antenna performance is demonstrated using the dual-band 2.4/5.2-GHz back-to-back E-shaped patch antenna mounted in location 4A. The reflection coefficients measured without the keyboard base and with the keyboard base for different laptop opening angles ψ are shown in Figure 22.59.

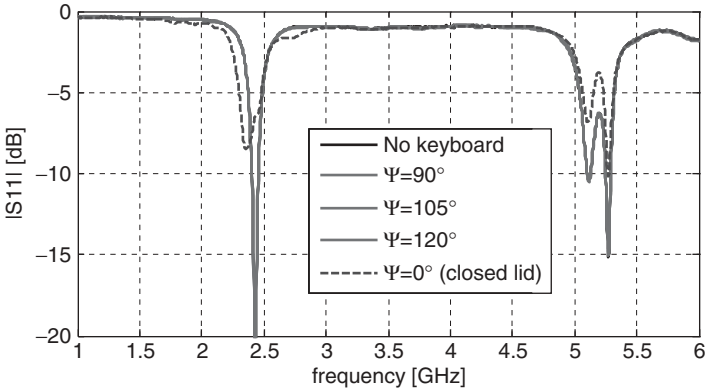


Figure 22.59 Keyboard effects on the measured input reflection coefficient of a top mounted back-to-back E-shaped patch antenna.

It is clearly seen that the presence of the keyboard base does not affect the antenna matching for normal operating angles, ψ (all solid curves overlap). This feature is very useful for the antenna designer: in the antenna matching procedure, the laptop base does not need to be modeled, which may drastically reduce the simulation complexity. When the laptop lid is closed, the antenna element is closer to the keyboard base and the input matching is slightly disturbed (dashed curve). This effect, however, may depend on the antenna type.

The effects of the laptop base on the radiation pattern are shown in Figure 22.60. The gain in the horizontal plane is not significantly affected by the keyboard, and the antenna maintains the required quasi-omnidirectional characteristics. The impact of the laptop keyboard base is clearly seen in the vertical plane radiation pattern. The reflection from the keyboard surface contributes to the enhanced radiation in the elevation range 0° to 90°. However, interference of direct and reflected waves in the keyboard results in strong maxima and deep nulls, hypothetically causing dead zones. In the elevation range, $-90^\circ$to 0°, the radiation is significantly reduced by the keyboard's blocking effect. However, there is still radiation originating from the waves diffracted from the keyboard's base edges. In the back screen hemisphere (90° to 270°), the radiation pattern is not severely affected by the keyboard. As a conclusion, it can be said that the keyboard changes the top mounted antenna radiation pattern; however, the effect on the horizontal plane is small.

Antenna Location Study According to the guidelines given above, the best integrated antenna performance can be achieved when the radiator is mounted inside the laptop's screen rim. Even though the central top location (4A) is the most beneficial, other design constraints may result in needing to use the other locations (e.g., 4B or 4C). In the design of multiple antenna arrangements, for instance, each radiator has to be integrated in distinct places. For each of the antenna locations described previously, the effects based on the laptop structure may be different and may have to be analyzed individually.

In the following section, the influence of antenna location on the radiation pattern of a 2.4-GHz IFA element and a 2.4/5.2-GHz back-to-back E-shaped patch antenna is presented. The horizontal and vertical plane radiation patterns are shown in Figures 22.61,

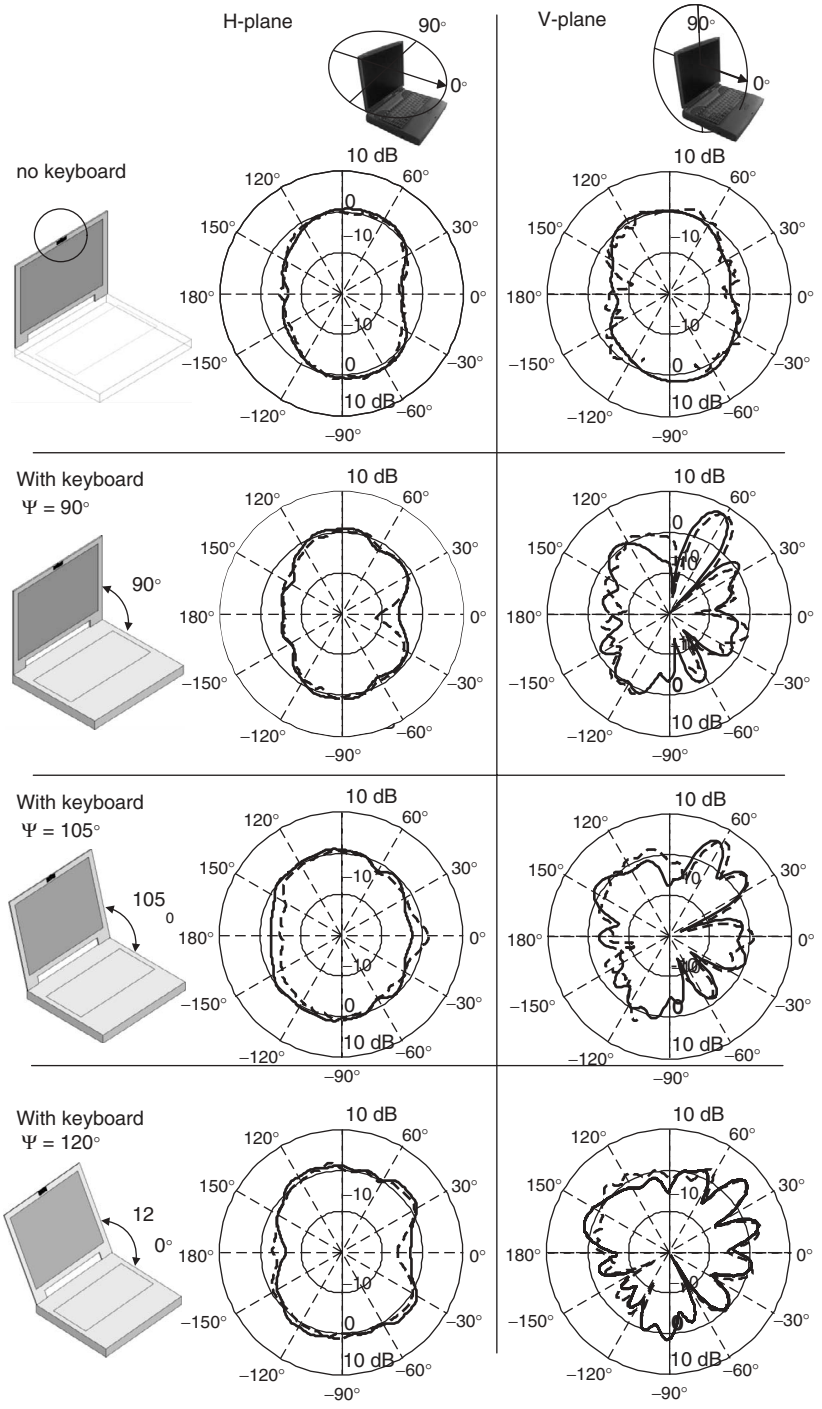


Figure 22.60 Keyboard effects on the radiation pattern of a top mounted back-to-back E-shaped patch antenna at 2.44 GHz (gain scale): ——— simulation, ------ measurement.

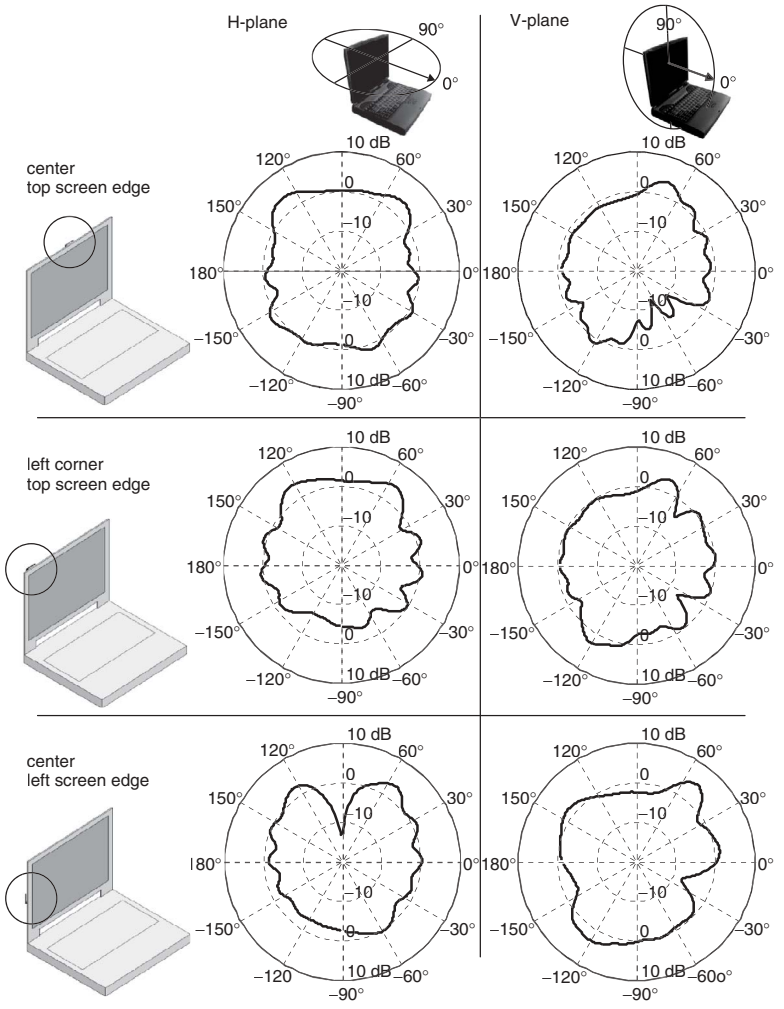


Figure 22.61 IFA integrated in the screen rim: computed gain at 2.44 GHz.

22.62 and 22.63. The first two rows of each figure correspond to antennas located on the top screen rim, in positions 4A and 4B. As mentioned previously, the antennas in these locations provide fairly omnidirectional radiation patterns. Moreover, the wrapped patch antenna is less sensitive to changing from center to corner location. The vertical plane radiation pattern is enhanced in the elevation range 0° to 90° (contribution of waves reflected from the keyboard base) and decreased in the elevation range -90° to 0° (laptop base blockage effect). The effects of the keyboard base are stronger when the antenna is located in the central position of the keyboard (position 4A) and weaker for the corner location (position 4B). For the 5.25-GHz wrapped patch element located in positions 4A and 4B, the effect of the keyboard structure is small because, in this frequency, the antenna radiates predominantly in the horizontal plane above the laptop base (Figure 22.63).

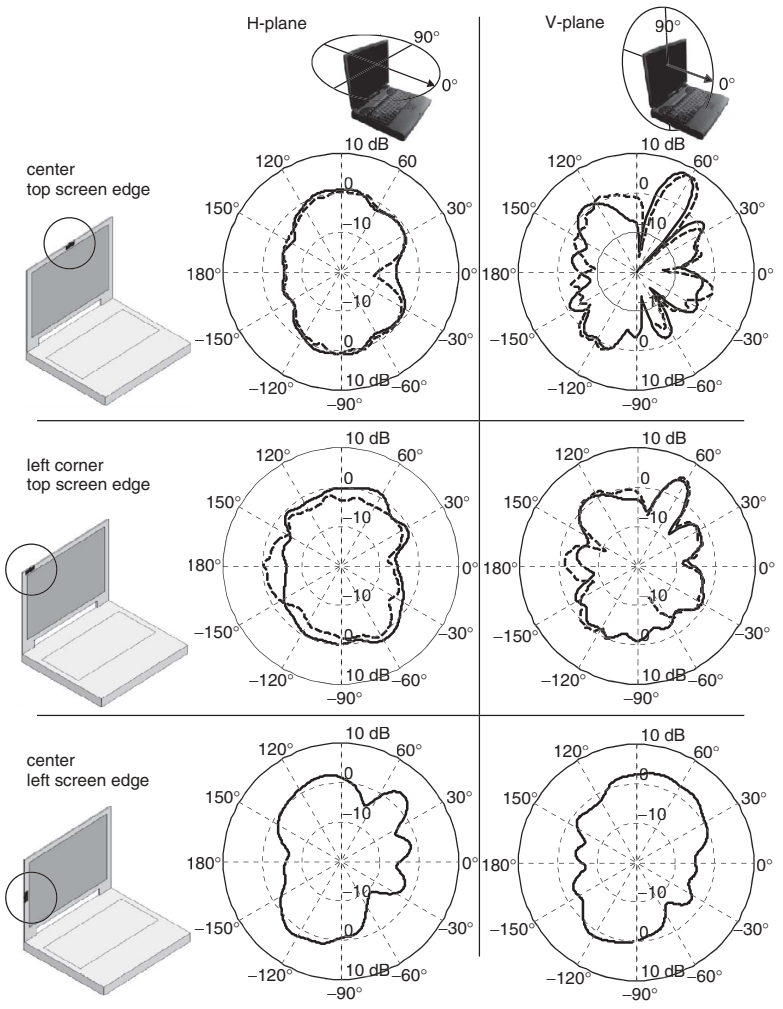


Figure 22.62 Back-to-back E-shaped patch antenna integrated in the screen rim: gain at 2.44 GHz (——— simulation, ------ measurement).

However, when the antenna is located on the side screen edge (position 4C), the dominant polarization changes by 90°, which can be used in polarization diversity schemes. If the keyboard structure was not present, the horizontal patterns for position 4C would strictly resemble the vertical pattern for position 4A (and analogously for the other planes). In the real scenario, the keyboard structure breaks the symmetry but the resemblance is still visible. As can be seen in Figure 22.63, the V-plane of position 4A is the H-plane of position 4C distorted by the keyboard, and the V-plane of position 4C is the distorted H-plane of position 4A.

The numerical results presented in this section have been obtained using 3 D full-wave simulation software. The laptop keyboard base has been modeled as a simple 225 × 260 × 30 mm^3 metal box. The screen has been substituted by a 1-mm thick metal plate. During the measurement process, the antennas integrated over a brass metal plate have been

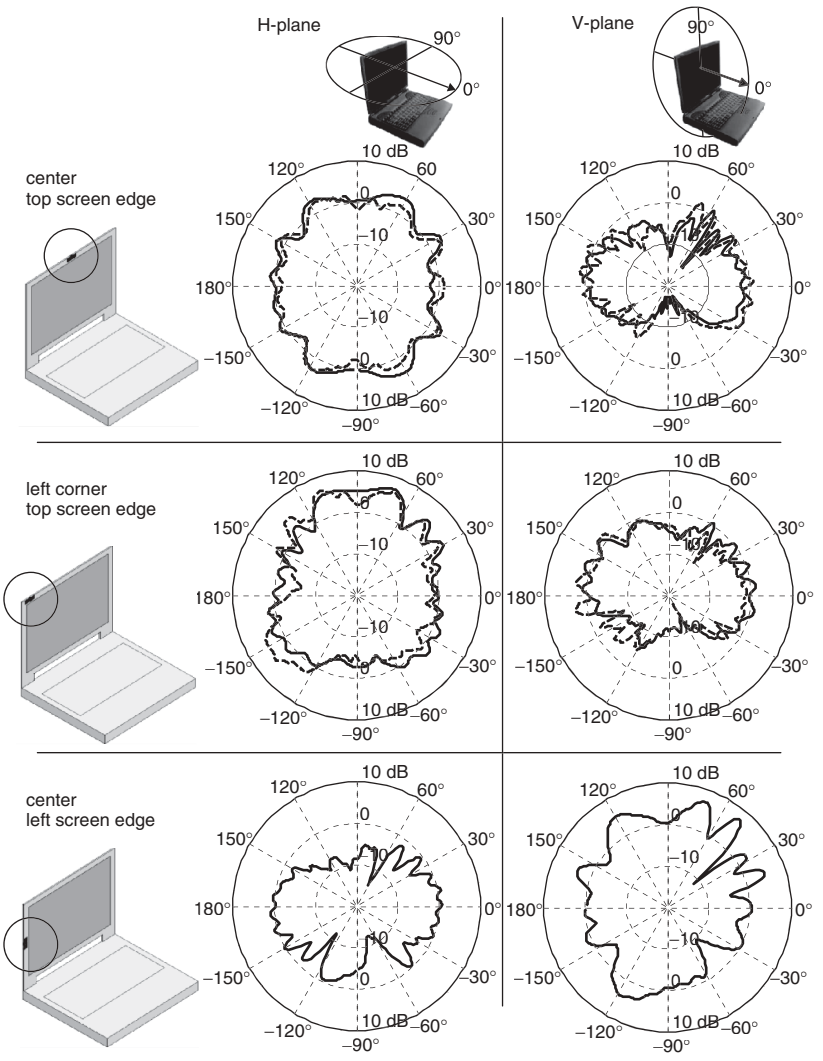


Figure 22.63 Back-to-back E-shaped patch antenna integrated in the screen rim: gain at 5.25 GHz (——— simulation, ------ measurement).

attached to a real Toshiba Satellite 4200 Series keyboard base, as shown in Figure 22.53. Radiation pattern measurements have been performed in the near-field spherical range at the University of California, Los Angeles. The very good agreement obtained between simulation and measurement results, confirms that the simplified keyboard model can be applied successfully to the radiation pattern calculation.

22.4.3.3 Human Interactions The electromagnetic interaction between the wireless interface and the user in laptops is of less importance than in handset terminals. In normal operation, the laptop's integrated antenna can be separated by several wavelengths from the user's body. Additionally, the maximum power radiated by a WLAN antenna is usually much lower than the maximum transmitted power by a cellular handset terminal.

Although not as prevalent in laptop structures, there is some disparity among the different antenna locations. The influence of the operator on the antenna performance is stronger when the PCMCIA wireless interface is inserted in location 1B (see Figure 22.56). The energy absorbed by the user's hand may deteriorate the radiation efficiency and for small separations even disturb the input matching [62, 95]. At 5 GHz, it was found that the user blocking effect may lower the radiation up to 20 dB [95]. For antennas located in the other positions, the effects are small.

The electromagnetic wave absorption by the operator depends on the antenna location and operator position (typing or not-typing). For the antenna in location 1B, the highest EM absorption in the typing and not-typing situations occurs in the hand and in the chest, respectively. The hands over the keyboard base have a significant blocking effect on the SAR of the head region. Thus, the peak SAR levels are low compared with the safety guidelines [95].

22.4.4 Integration of Multiple Antennas

The global pursuit of wireless system reliability and a growing demand for high spectrum efficiency have furthered recent achievements of diversity [96], smart antennas, and multiple-input multiple-output (MIMO) techniques [9]. Therefore integrated multielement antenna arrangements are highly beneficial in laptop applications and are expected to become a standard in the near future. The optimization of those arrangements incorporates aspects of antenna element design, integration, and multiantenna topology design. The multiantenna design paradigm differs for diversity, smart-antenna, and MIMO concepts [97]. This section deals with multiantenna arrangements for MIMO.

In the introduction to Section 22.4, the basic laptop internal antenna requirements have been outlined. In the design of a multielement antenna for a laptop, each antenna element should fulfill the requirements set forth, and, in particular, an omnidirectional radiation pattern for each element is beneficial for MIMO applications [97]. The integration of multiple antennas into limited spaces raises the additional problem of mutual interactions between antennas. Mutual coupling may have a significant effect in the MIMO channel capacity [98] and radiation efficiency [4]. In this section, examples of multielement antennas integrated into a plug-in PCMCIA interface and inside the laptop screen rim are presented. Topology effects on the antenna performance are also analyzed with special consideration of MIMO related requirements.

Two- and four-element compact MIMO antennas that are to be integrated into a PCMCIA card [72] are shown in Figure 22.64a,b. Each of the elements occupies a volume $17 \times 9 \times 4$ mm^3 and has three resonant frequencies at 2.45, 5.25, and 5.8 GHz, with bandwidths of 100, 200, and 150 MHz, respectively. The separation between the radiator centers in the two-element configuration is 27 mm ($<\lambda/4$). The four-element antenna is realized by adding two more elements on the opposite side of the ground plane. A PIFA element is relatively unaffected by another nearby PIFA element. Even with tight packaging, the antenna elements preserve a radiation pattern that is close to omnidirectional (for freestanding card operation).

The first row of the scattering matrix for the two-element and four-element compact PIFA arrangements are shown in Figure 22.65 (left column). In both cases, the input reflection coefficient (S_{11}) of the antenna elements is below -10 dB. However, due to the small element spacing and the narrow bandwidth a final tuning may be necessary after structure assembly. The mutual coupling in the two-element antenna is below

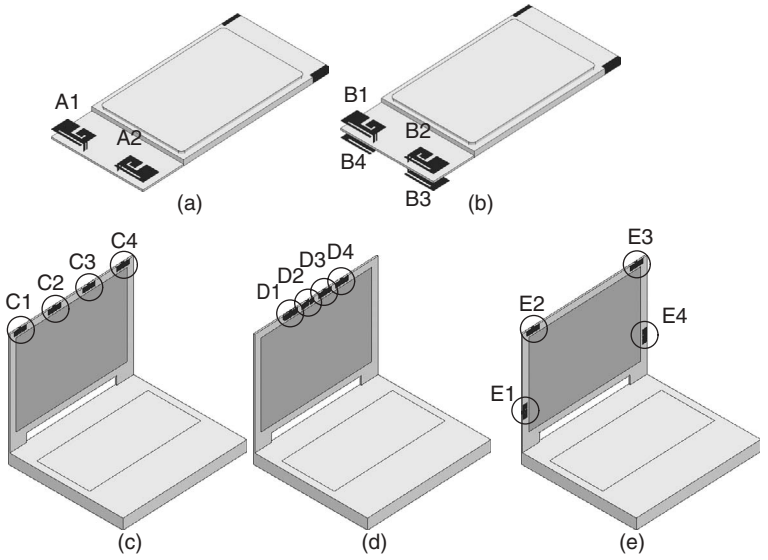


Figure 22.64 Multiple-antenna arrangement for MIMO enabled laptops: (a,b) two- and four-element compact triple-band PIFA arrangement for a PCMCIA card [72] and (c,d,e) four-element dual-band back-to-back E-shaped patch arrangements integrated into the screen rim.

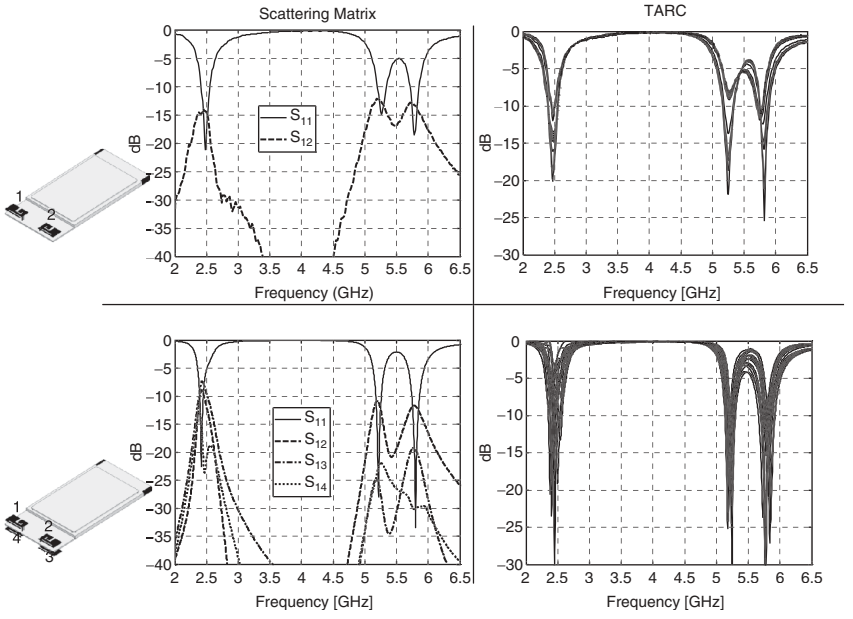


Figure 22.65 Measured scattering matrices and TARC of the compact triple-band PIFA arrangements for PCMCIA card. Each of 20 TARC curves is calculated for antennas excited at the same amplitude but with a random phase.

−14 dB within the 2.4-GHz ISM band, which is a very good result for such small antenna spacing. In the four-element antenna, the highest mutual coupling occurs between the top and bottom elements (B1–B4 and B2–B3) and reaches −7 dB. The increase of mutual coupling after adding two extra elements decreases the radiation efficiency, which can be observed with TARC analysis. In the right column of Figure 22.65, the results of TARC are shown; each curve corresponds to an excitation vector of unit magnitude and random phase. The random selection of the excitation phases simulates MIMO signaling and allows the estimation of array performance under system operation conditions.

It is apparent that the curves corresponding to different excitations are much more spread out for the four-element arrangement. It shows that the radiation efficiency is more dependent on the combinations of excitation signals for an antenna arrangement with high mutual coupling. In the worst case, at the 2.45-GHz band, the TARC of the four-element arrangement can reach −3 dB, which corresponds to a 50% radiation efficiency.

A similar analysis is conducted for the three configurations of four-element MIMO arrangements that are to be integrated into a laptop screen rim. Dual band, back-to-back E-shaped patch elements are used [91], as they provide fairly omnidirectional radiation patterns in both operating bands. The three antenna arrangements—linear with spacing 82 mm ($2\lambda/3$ at 2.45 GHz), linear with spacing 41 mm ($\lambda/3$ at 2.45 GHz), and maximum spacing with two elements on the side screen rims (spacing $>1\lambda$ at 2.45 GHz)—are shown in Figure 22.64c–e.

Selected elements of the scattering matrices and TARC are presented in Figure 22.66. The mutual coupling between antenna elements grows as the physical separation is decreased. The highest mutual coupling values occur within the 2.4-GHz ISM band for arrangement D. For the same arrangement, the mutual coupling is usually lower at the high frequency band, as the electrical separation between elements is larger with respect to the wavelength. As expected, the lowest mutual coupling values in the 2.4-GHz ISM band are obtained for the largest separation of elements (arrangement E). In this case, however, one phenomenon should be noted: mutual coupling between elements E2 and E3 is the highest in the high frequency band and approaches the value of mutual coupling for smaller spacing (compare with array C and D). This effect comes from the high surface currents excited along the edge of the ground plane at 5.25 GHz. It is worthwhile to highlight this effect to stress that, even for large antenna spacing, the laptop structure may cause unexpected interactions and to predict them correctly, all the antenna elements integrated into the laptop's real structure should be simulated.

TARC results for the screen integrated arrangements (right column of Figure 22.66) clearly show that the arrangements with the smallest spacing (D) may have the lowest efficiency. For this configuration, the TARC curves have the greatest spread, which indicates that the overall efficiency is more excitation dependent and, in the worst-case scenario, drops to 50% within the 2.4-GHz ISM band. For arrays C and E, the TARC curves have little spread and do not go above −6 dB for both operating frequency bands.

22.5 INTEGRATED ANTENNAS FOR OTHER MOBILE DEVICES

In Sections 22.3 and 22.4, antenna designs for handset and laptop applications have been addressed. Among all personal wireless communication devices available on the market, the majority of units can be classified into those two categories. Nevertheless, as the demand for wireless equipment grows and the price of miniature radio interfaces

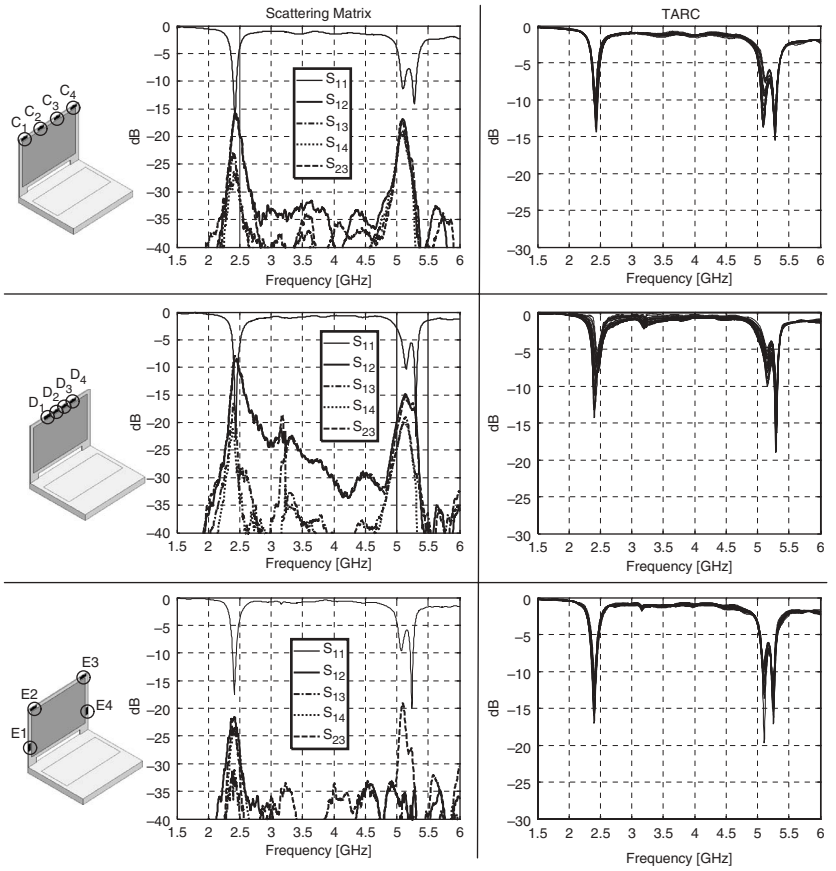


Figure 22.66 Scattering matrices and TARC of dual-band four-element arrangements integrated into the laptop screen rim. Each of 20 TARC curves is calculated for antennas excited at the same amplitude but with a random phase.

decreases, the number of different units undergoing an integration of wireless technology is increasing.

Due to high functionality, small size, and relatively long battery life (as compared to laptops), personal digital assistants (PDAs) are becoming very popular. Manufacturers usually equip the units with Bluetooth and WLAN connectivity. Thanks to their large screens, PDA terminals with integrated GPS receivers can work as portable navigation systems. Even recently, the integration of cellular radios into PDAs has become popular, leading to such designs as "smart phones" and "Blackberries®."

The design procedures for PDA antennas and handset antennas are very similar (Section 22.3), and the same antenna types can be used. As the PDA is usually a very densely packed device, the antenna size constraints are also severe. Analogous to handsets, integrated antennas for PDAs can be classified as external and internal. Typically, external radiators are stubby antennas, as shown in Figure 22.67. An internal PDA antenna may have an IFA (Sections 22.3.2 and 22.4.1.2) or a microstrip/PIFA structure (Sections 22.3.2, 22.3.3.1, and 22.4.1.3). The PDA case is often made of metal; therefore

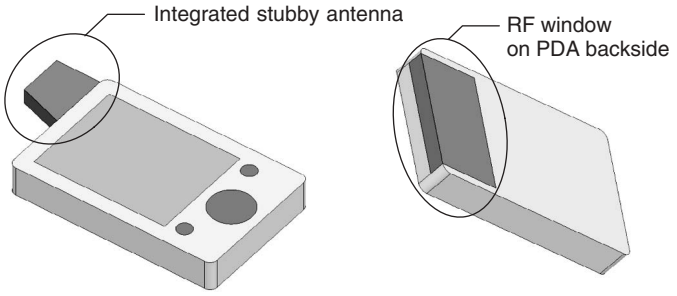


Figure 22.67 Integration of antennas into a PDA terminal.

the small "RF window" (Section 22.4.3.2) is usually placed in the PDA backside (see Figure 22.67).

Users of PDA devices that are not equipped with built-in wireless interfaces may use plug-in wireless cards inserted into a memory expansion slot (Figure 22.68). This solution, analogous to the PCMCIA/USB dongle for laptop computers, usually houses a miniature antenna in the part of the card that is protruding from the PDA. Due to extremely limited volume (to minimize the risk of physical breakage), the protruding part cannot be long, and very compact chip antennas (Section 22.4.1.4) and meander IFA antennas (Section 22.4.1.2) are often used.

A Bluetooth interface is commonly used to control and communicate with hands-free headsets. Those compact devices establish a short-range wireless connection with the cellular phone/PDA, allowing the user to lead a voice conversation. Chip antennas (Section 22.4.1.4) are commonly used in hands-free headsets.

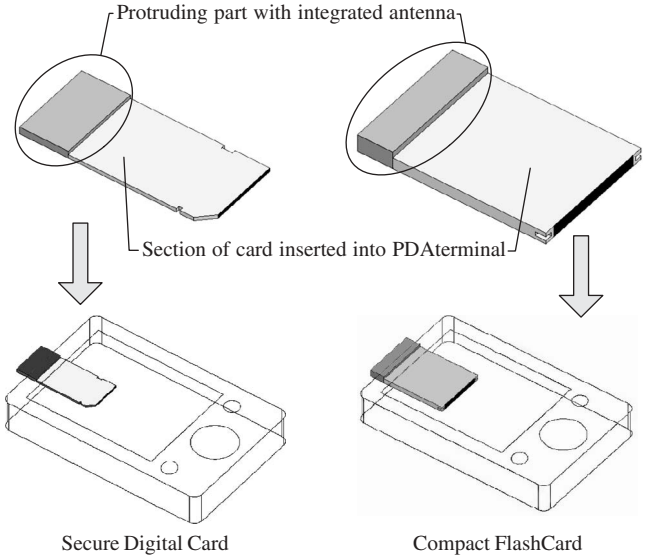


Figure 22.68 Wireless interfaces housed in miniature cards plugged into memory expansion slots of portable devices.

REFERENCES

1. IEEE C95.3-2002, IEEE recommended practice for measurements and computations of radio frequency electromagnetic fields with respect to human exposure to such fields, 100 kHz–300 GHz (Revision of IEEE C95.3-1991), 13 January 2003.
2. CENELEC, European Specification ES 59005, Considerations for the evaluation of human exposure to electromagnetic fields (EMFs) from mobile telecommunication equipment (MTE) in the frequency range from 30 MHz–6 GHz, Ref. No. ES 59005:1998 E, 1998.
3. ICNIRP Guidelines, Guidelines for limiting exposure to time-varying electric, magnetic, and electromagnetic fields (up to 300 GHz), International Commission of Non-Ionizing Radiation Protection, *Health Physics*, Vol. 74, No. 4, pp. 494–522, 1998.
4. M. Manteghi and Y. Rahmat-Samii, Multiport characteristics of a wide-band cavity backed annular patch antenna for multipolarization operations, *IEEE Trans. Antennas Propag.*, Vol. 53, No. 1, pp. 466–474, January 2005.
5. Y. Rahmat-Samii and Z. Li, Handset communication antennas, including human interactions, in *Wireless Networks: From the Physical Layer to Communication*, *Computing, Sensing and Control*, G. Franceschetti and S. Stornelli (Eds.), Academic Press, New York, 2006.
6. T. Taga, Analysis of planar inverted-F antennas and antenna design for portable radio equipment, in *Analysis, Design, and Measurement of Small and Low-Profile Antennas*, K. Hirasawa and M. Haneishi (Eds.), Artech House, Norwood, MA, 1992.
7. C. Rowell, A brief survey of internal antennas in GSM phones: 1998 to 2004, *Proc. UNSC/URS/ National Radio Science Meeting*, July 2005.
8. R. Vaughan and J. B. Andersen, *Channels, Propagation and Antennas for Mobile Communications*, IEE, London, 2003.
9. M. A. Jensen and J. W. Wallace, A review of antennas and propagation for MIMO wireless communications, *IEEE Trans. Antennas Propag.*, Vol. 52, No. 11, pp. 2810–2824, November 2004.
10. M. A. Jensen and Y. Rahmat-Samii, EM interaction of handset antennas and a human in personal communications, *Proc. IEEE*, Vol. 83, No. 1, pp. 7–17, January 1995.
11. D. Manteuffel, D. Heberling, and I. Wolf, EM user interaction—considerations for the design of mobile phones, *Proceedings of the International ITG-Conference on Antennas*, Berlin, Germany, September 2003.
12. C.-M. Su, C.-H. Wu, K.-L. Wong, S.-H. Yeh, and C.-L. Tang, User's hand effects on EMC internal GSM/DCS mobile phone antenna, *Proc. IEEE Int. Symp. Antennas Propag.*, pp. 2097–2100, June 2006.
13. J. C. Lin, A research program on the human impact of wireless telecommunications and other electromagnetic-field-emitting systems, *IEEE Antenna Propag. Mag.*, Vol. 46, No. 6, pp. 161–163 December 2004.
14. M. A. Stuchly and S. S. Stuchly, Experimental radio and microwave dosimetry, in *Handbook of Biological Effects of Electromagnetic Fields*, 2nd ed., C. Polk and E. Postow, (Eds.) CRC Press, Boca Raton, FL, pp. 295–336, 1996.
15. Federal Communications Commission, Order on Recon and FNPRM, 9 June 2005, FCC 05–122, WT Docket No. 01–309, http://www.fcc.gov/.
16. K. Ogawa, and T. Matsuyoshi, An analysis of the performance of a handset diversity antenna influenced by head, hand, and shoulder effects at 900 MHz. I. Effective gain characteristics, *IEEE Trans. Vehicular Technol.*, Vol. 50, No. 3, pp. 830–844 May 2001.
17. K. Ogawa, T. Matsuyoshi, and K. Monma, An analysis of the performance of a handset diversity antenna influenced by head, hand, and shoulder effects at 900 MHz. II. Correlation characteristics, *IEEE Trans. Vehicular Technol.*, Vol. 50, No. 3, pp. 845–853, May 2001.

18. P.-S. Kildal, and C. Carlsson, Comparison between head losses of 20 phones with external and built-in antennas measured in reverberation chamber, *Proc. IEEE Int. Symp. Antennas Propag.*, Vol. 1, pp. 436–439. June 2002.
19. K. Fujimoto and J. R. James (Eds.), *Mobile Antenna Systems Handbook*, 2 ed., Artech House, Norwood, MA, 2001.
20. K.-L. Wong, *Planar Antennas for Wireless Communications*, John Wiley & Sons, Hoboken, NJ, 2003.
21. S. Y. Lin, Multiband folded planar monopole antenna for mobile handset, *IEEE Trans. Antennas Propag.*, Vol. 52, No. 7, pp. 1790–1794, July 2004.
22. M. Martínez-Vázquez, ACE small terminal antennas activities: a review of the state of the art, *Proc. IEEE Int. Symp. Antennas Propag.*, Vol. 4A, pp. 93–96, July 2005.
23. Y. J. Cho, S. H. Hwang, and S.-O. Park, A dual-band internal antenna with a parasitic patch for mobile handsets and the consideration of the handset case and battery, *IEEE Antennas Wireless Propas. Lett.*, Vol. 4, No. 4, pp. 429–432, 2005.
24. Z. Li and Y. Rahmat-Samii, Optimization of PIFA–IFA combination in handset antenna designs, *IEEE Trans. Antennas Propag.*, Vol. 53, No. 5, pp. 1770–1778, May 2005.
25. P. N. Fletcher, M. Dean, and A. R. Nix, Mutual coupling in multi-element array antennas and its influence on {MIMO} channel capacity, *Electron. Lett.*, Vol. 39, No. 4, pp 342–344, 2003.
26. C. Waldschmidt, C. Kuhnert, M. Pauli and W. Wiesbeck, Integration of MIMO antenna arrays into hand-helds, in *IEE 5th International Conference on 3G Mobile Communication Technologies*, London, UK, 2004.
27. Z. Li and Y. Rahmat-Samii, Whip–PIFA combination for handset application: a hybrid circuit model and full-wave analysis, *Microwave Optical Technol. Lett.*, Vol. 44, No. 3, pp. 210–214, February 2005.
28. Z. N. Chen and M. Y. W. Chia, *Broadband Planar Antennas Design and Applications*, John Wiley & Sons, Chichester, UK, 2006.
29. D. M. Pozar, Input impedance and mutual coupling of rectangular microstrip antennas, *IEEE Trans. Antennas Propag.*, Vol. 30, No. 8, pp. 1191–1196, November 1982.
30. C. Waldschmidt and W. Wiesbeck, Quality measures and examples of arrays for MIMO in hand-held devices, *Proc. IEEE Int. Symp. Antennas Propag.*, Vol. 2, pp. 1293–1996, June 2004.
31. H. Carrasco, H. D. Hristov, R. Feick, and D. Cofré, Mutual coupling between planar inverted-F antennas, *Microwave Optical Technol. Lett.*, Vol. 42, No. 3, pp. 224–227, August 2004.
32. J. Thaysen, Mutual coupling between two identical planar inverted-F antennas, *Proc. IEEE Int. Symp. Antennas Propag.*, pp. 504–507, June 2002.
33. H. Steyskal and J. S. Herd, Mutual coupling compensation in small array antennas, *IEEE Trans. Antennas Propag.*, Vol. 38, No. 12, pp. 1971–1975, December 1990.
34. F. Yang and Y. Rahmat-Samii, Microstrip antennas integrated with electromagnetic band-gap (EBG) structures: a low mutual coupling design for array applications, *IEEE Trans. Antennas Propag.*, Vol. 51, No. 10, pp. 2936–2946, October 2003.
35. D. G. Fang, C. Z. Luan and Y. P. Xi, Mutual coupling in microstrip antenna arrays: evaluation, reduction, correction or compensation, in *Proceedings of the International Workshop on Antenna Technology: Small Antennas and Novel Metamaterials (IWAT)*, Singapore, March 2005, pp. 37–40.
36. A. Diallo, C. Luxey, P. Le Thuc, R. Star, and G. Kossiavas, Reduction of the mutual coupling between two planar inverted-F antennas working in close radiocommunication standards, *Proceedings of the 18th International Conference on Applied Electromagnetics and Communications (ICECom)*, Dubrovnik, Croatia, October 2005.

37. C. Peixeiro, Planar inverted-F antennas for MIMO applications in small multi-standard handsets, *Proceedings of the 13th International Conference on Telecommunications*, Madeira, Portugal, May 2006.

38. M. A. Jensen and Y. Rahmat-Samii, Performance analysis of antennas for hand-held transceivers using FDTD, *IEEE Trans. Antennas Propag.*, Vol. 42, No. 8, pp. 1106–1113, August 1994.

39. C. T. P. Pong, P. S. Hall, H. Ghafouri-Shoraz, and D. Wake, Triple band planar inverted-F antennas for handheld devices, *Electron. Lett.*, Vol. 36, No. 2, pp. 112–114, January 2000.

40. K.-L. Wong, *Compact and Broadband Microstrip Antennas*, John Wiley & Sons, Hoboken, NJ, 2002.

41. T. Gandara and C. Peixeiro, Compact triple-band double U-slotted planar inverted-F antenna, in *Proceedings of the IEEE 14th International Symposium on Personal, Indoor and Mobile Radio Communications*, Barcelona, Spain, September 2004.

42. T. Gandara, R. Urban, L. Fregoli, and C. Peixeiro, Planar inverted-F antennas for multi-standard handsets, in *Proceedings of the 18th International Conference on Applied Electromagnetics and Communications (ICECom)*, Dubrovnik, Croatia, October 2005.

43. G. J. Foschini and M. J. Gans, On limits of wireless communications in a fading environment when using multiple antennas, *Wireless Personal Commun.*, Vol. 6, No. 3, pp. 311–335, March 1998.

44. S. Blanch, J. Romeu, and J. Corbella, Exact representation of antenna system diversity performance from input parameter description, *Electron. Lett.*, Vol. 39, No. 9, pp. 705–707, May 2003.

45. A. Derneryd and G. Kristensson, Antenna signal correlation and its relation to the impedance matrix, *Electron. Lett.*, Vol. 40, No. 7, pp. 705–707, April 2004.

46. J. Huang, The finite ground plane effect on the microstrip antenna radiation pattern, *IEEE Trans. Antennas Propag.*, Vol. 31, No. 4, pp. 649–653, July 1983.

47. M.-C. Huynh and W. Stutzman, Ground plane effects on planar inverted-F antenna (PIFA) performance, *Proc. IEE Microwave Antennas Propag.*, Vol. 150, No. 4, pp. 209–213, August 2003.

48. R. Urban and C. Peixeiro, Ground plane size effects on a microstrip patch antenna for small handsets, in *Proceedings of the 15th International Conference on Microwaves, Radar and Wireless Communications (MIKON)*, Warsaw, Poland, May 2004.

49. P. Kabacik, A. Byndus, R. Hossa, and M. Bialkowski, An application of a narrow slot cut in the ground to improve multi-band operation of a small antenna, in *Proceedings of the 2003 IEEE Topical Conference on Wireless Communication Technology*, Honolulu, Hawaii, October 2003.

50. R. Urban and C. Peixeiro, Battery effects on a small handset planar inverted-F antenna, in *Proceedings of the Mobile & Wireless Communications Summit*, Lyon, France, June 2004.

51. M. Geissler, M. Gehrt, D. Heberling, P. Waldow, and I. Wolff, Investigation on radiation Q of integrated handset antennas, *Proc. IEEE Int. Symp. Antennas Propag.*, Vol. 3, pp. 54–57, July 2001.

52. ACE Deliverable 2.2D1, *Report on the State of the Art in Small Terminal Antennas: Technologies, Requirements and Standards*, M. Martinez Vázquez (Ed.), European Union project FP6-IST 508009, December 2004.

53. O. Kivekas, J. Ollikainen, T. Lehtiniemi, and P. Vainikainen, Bandwidth, SAR and efficiency of internal mobile phone antennas, *IEEE Trans. Antennas Propag.*, Vol. 46, pp. 71–86, February 2004.

54. L. C. Fung, S. W. Leung, and K. H. Chan, An investigation of the SAR reduction methods in mobile phone applications, in *IEEE International Symposium on Electromagnetic Compatibility*, Italy, 2002.

55. M. Sager, M. Forcucci, and T. kristensen, A novel technique to increase the realized efficiency of a mobile phone antenna placed beside a head-phantom, *Proc. IEEE Int. Symp. Antennas Propag.*, pp. 1013–1016, June 2003.
56. R. Y. Tay, Q. Balzano, and N. Kuster, Dipole configurations with strongly improved radiation efficiency for hand-held transceivers, *IEEE Trans. Antennas Propagat.*, Vol. 46, pp. 798–806, June 1998.
57. J. C. Lin, Specific absorption rates (SARs) induced in head tissues by microwave radiation from cell phones, *IEEE Antennas Propag. Mag.*, Vol. 42, pp. 138–139, October 2000.
58. A. Taflove, *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, Artech House, Norwood, MA, 1995.
59. J. Bray and C. Sturman, Bluetooth: *Connect Without Cables*, Prentice Hall, Englewood Cliffs, NJ, 2000.
60. J. Geier, Wireless Lans: *Implementing Interoperable Networks*, Macmillan Technical Publishing, Indianapolis, IN, 1999.
61. D. Liu, E. Flint, and B. Gaucher, Integrated laptop antennas—design and evaluation, in *Proceedings of the IEEE APS International Symposium and URSI Radio Science Meeting*, San Antonio, Texas, June 2002, Vol. 4, pp. 56–59.
62. D. Liu, B. P. Gaucher, E. B. Flint, T. W. Studwell, H. Usui, and T. J. Beukema, Developing integrated antenna subsystems for laptop computers, *IBM J. Res. Dev.*, Vol. 47, No. 2/3, pp. 355–367, March/May 2003.
63. T. L. Chen, Multi-band printed sleeve dipole antenna, *Electron. Lett.*, Vol. 39, No. 1, pp. 14–15, January 2003.
64. M. Ali, M. Okoniewski, M. A. Stuchly, and S. S. Stuchly, Dual-frequency strip-sleeve monopole for laptop computers, *IEEE Trans. Antennas Propag.*, Vol. 47, No. 1, pp. 317–323, February 1999.
65. C.-C. Lin, S.-W. Kuo, and H.-R. Chuang, A 2.4-GHz printed meander-line antenna for USB WLAN with Notebook-PC housing, *IEEE Microwave Wireless Components Lett.*, Vol. 15, No. 9, pp. 546–548, September 2005.
66. C. Peixeiro, C. A. Fernandes, A. A. Moreira, B. Kemp, M. Capstick, A. Burr, and H. Hofstetter, IST-2001-32125 FLOWS, Deliverable Number: D16, *Design and Performance of Antenna Prototypes*, November 2004.
67. H. Y. D. Yang, Miniaturized printed wire antenna for wireless communications, *IEEE Antennas Wireless Propag. Lett.*, Vol. 4, pp: 358–361, 2005.
68. D. Liu and B. P. Gaucher, The inverted-F antenna height effects on bandwidth, *Proc. IEEE Antennas Propag. Symp.*, Vol. 2A, pp. 367–370, July 2005.
69. C. Soras, M. Karaboikis, G. Tsachtsiris, and V. Makios, Analysis and design of an inverted-F antenna printed on a PCMCIA card for the 2.4 GHz ISM band, *IEEE Antenna Propag. Mag.*, Vol. 44, No. 1, pp. 37–44, February 2002.
70. R. Bancroft, Development and integration of commercially viable 802.11.a/b/g Hiperlan/WLAN antenna into laptop computers, *Proc. IEEE Int. Symp. Antennas Propag.*, Vol. 4A, pp. 231–234, July 2005.
71. C. M. Allen, A. Z. Elsherbeni, C. E. Smith, C.-W. P. Huang, and K.-F. Lee, Tapered meander slot antenna for dual band personal wireless communication systems, *Microwave and Optical Technol. Lett.*, Vol. 36, No. 5, pp. 381–385, March 2003.
72. M. Manteghi and Y. Rahmat-Samii, Novel compact tri-band two-element and four-element MIMO antenna designs, *Proc. IEEE Int. Symp. Antennas Propag.*, pp. 4443–4446, July 2006.
73. K. Ito and T. Hose, Study on the characteristics of planar inverted F antenna mounted in laptop computers for wireless LAN, *Proc. IEEE Int. Symp. Antennas Propag.*, Vol. 2, pp. 22–25, June 2003.

74. M. Sanad, A very small double C-patch antenna contained in a PCMCIA standard PC card, *Proc. IEEE Int. Symp. Antennas Propag.*, Vol. 1, pp. 117–120, April 1995.

75. F. Yang, X.-X. Zhang, X. Ye, and Y. Rahmat-Samii, Wide-band E-shaped patch antennas for wireless communications, *IEEE Trans. Antennas Propag.*, Vol. 49, No. 7, pp. 1094–1100, July 2001.

76. Y. Ge, K. P. Essele, and T. S. Bird, Compact diversity antenna for wireless devices, *Electron. Lett.*, Vol. 41, No. 2, pp. 52–53, January 2005.

77. C. L. Tang, 2.4/5.2 GHz dual-band chip antenna for WLAN application, *Proc. IEEE Int. Symp. Antennas Propag.*, Vol. 1A, pp. 454–457, June 2006.

78. Y. Dakeya, T. Suesada, K. Asakura, N. Nakajima, and H. Mandai, Chip multilayer antenna for 2.45 GHz-band application using LTCC technology, *IEEE MTT-S Int. Microwave Symp. Dig.*, Vol. 3, pp. 1693–1696, June 2000.

79. D. Liu, B. Gaucher, and T. Hildner, A dualband antenna for WLAN applications, in *Proceedings IEEE International Workshop on Antenna Technology*: *Small Antennas and Novel Metamaterials IWAT*, March 2005, pp. 201–204.

80. D. Liu and B. Gaucher, A branched inverted-F antenna for dual band WLAN applications, *Proc. IEEE Int. Symp. Antennas Propag.*, Vol. 3, pp. 2623–2626, June 2004.

81. J. Yeo, Y. J. Lee, and R. Mittra, A novel dual-band WLAN antenna for Notebook platforms, *Proc. IEEE Int. Symp. Antennas Propag.*, Vol. 3, No. 2, pp. 1439–1442, June 2004.

82. D. Liu and B. Gaucher, A triband antenna for WLAN applications, *Proc. IEEE Int. Symp. Antennas Propag.*, Vol. 2, pp. 18–21, June 2003.

83. Z. N. Chen, D. Liu, B. Gaucher, and T. Hildner, Reduction in antenna height by slotting ground plane, in *Proceedings of the IEEE International Workshop on Antenna Technology, Small Antennas, and Novel Metamaterials*, March 2006, pp. 25–28.

84. C. M. Su and K. L. Wong, Narrow flat-plate antenna for 2.4 GHz WLAN operation, *Electron. Lett.*, Vol. 39, pp. 344–345, February 2003.

85. K. Wang, L. Chou, and C. Su, Dual-band flat-plate antenna with a shorted parasitic element for laptop applications, *IEEE Trans. Antennas Propag.*, Vol. 53, No. 1, pp. 539–544, January 2005.

86. T. Ito, H. Moriyasu, and M. Matsui, A small antenna for laptop applications, in *Proceedings IEEE International Workshop on Antenna Technology, Small Antennas, and Novel Metamaterials*, March 2006, pp. 233–236.

87. J. M. Johnson and Y. Rahmat Samii, Wideband tab monopole antenna array for wireless adaptive and mobile information systems application, *Proc. IEEE Int. Symp. Antennas Propag.*, Vol. 1, pp. 718–721, July 1996.

88. C. M. Su, L. C. Chou, C. I. Lin, and K. L. Wong, Embedded DTV antenna for laptop application, *Proc. IEEE Int. Symp. Antennas Propag.*, Vol. 4B, pp. 68–71, July 2005.

89. J. Antoniuk, A. A. Moreira, and C. Peixeiro, L-bent omnidirectional patch antenna for wireless applications in laptop computers, *Proc. IEEE Int. Symp. Antennas Propag.*, Vol. 4A, pp. 355–358, July 2005.

90. J. Guterman, A. A. Moreira, and C. Peixeiro, Integration of omnidirectional wrapped microstrip antennas into laptops, *IEEE Antennas Wireless Propag. Lett.*, Vol. 5, pp. 141–144, April 2006.

91. J. Guterman, Y. Rahmat-Samii, A. A. Moreira, and C. Peixeiro, Quasi-omnidirectional dual-band back-to-back E-shaped patch antenna for laptop applications, *Electron. Lett.*, Vol. 42, No. 15, pp. 845–847, July 2006.

92. J. T. Bernhard, Analysis of integrated antenna positions on a laptop computer for mobile data communication, *Proc. IEEE Int. Symp. Antennas Propag.*, Vol. 4, pp. 2210–2213, July 1997.

93. F. M. Caimi and G. O'Neill, Antenna design for Notebook computers: pattern measurements and performance considerations, *Proc. IEEE Int. Symp. Antennas Propag.*, Vol. 4A, pp. 247–250, July 2005.

94. G. H. Huff, J. Feng, S. Zhang, G. Cung, and J. T. Bernhard, Directional reconfigurable antennas on laptop computers: simulation, measurement and evaluation of candidate integration positions, *IEEE Trans. Antennas Propag.*, Vol. 52, No. 12, pp. 3220–3227, December 2004.

95. J. Wang and O. Fujimura, EM interaction between a 5 GHz band antenna mounted PC and a realistic human body model, *IEICE Trans. Commun.*, Vol. E88-B, No. 6, pp. 2604–2608, June 2005.

96. R. G. Vaughan and J. B. Andersen, Antenna diversity in mobile communications, *IEEE Trans. Vehicular Technol.*, Vol. 36, No. 4, pp. 149–172, November 1987.

97. D. W. Browne, M. Manteghi, M. P. Fitz, and Y. Rahmat-Samii, Experiments with compact antenna arrays for MIMO radio communications, *IEEE Trans. Antennas Propag.*, Vol. 54, No. 11, pp. 3239–3250, November 2006.

98. J. W. Wallace and M. A. Jensen, Mutual coupling in MIMO wireless systems: a rigorous network theory analysis, *IEEE Trans. Wireless Commun.*, Vol. 3, No. 4, pp. 579–586, May 2004.

CHAPTER 23

Antennas for Mobile Communications

KYOHEI FUJIMOTO

23.1 INTRODUCTION

Remarkable progress in mobile communications systems has occurred in the last decade. Mobile phone systems have evolved from analog systems, called first generation (1G) systems, to digital systems, called second generation (2G) systems, and further to third generation (3G) systems, which are capable of multimedia transmission. Now the 3G systems are being advanced to fourth generation systems through 3.5G systems, which stand between 3G and 4G systems, in order to transition to 4G systems smoothly from 3G systems. In addition, various wireless mobile systems other than mobile phone systems have been deployed and service areas with rather narrow territories, ranging from very short distances to intermediate distances, whereas mobile phone systems provide nationwide services. These systems not only provide communication services but also perform control, data transmission, identification, and sensing, either through their network or with their own structure. Typical systems are UWB, RFID, Bluetooth, NFC (near-field communication) systems, WLAN (wireless local area network), and mobile WiMAX (worldwide interoperability for microwave access). Among them, some broadband systems feature capabilities of very high data-rate transmission, even while in motion at very high speed. Another significant possibility with these newly developed wireless systems is to achieve seamless communication links by connecting these systems to each other, including to mobile phone systems. Various antenna systems for use in these systems have been developed, and accordingly the technology has made progress along with the development of these systems.

A current trend in antenna design for mobile communication systems is to meet the requirements imposed for antennas used in 3G systems, advanced from 2G systems. Another trend is in antenna design for mobile wireless systems other than mobile phone systems.

The typical trend is personalization, which has been accelerated by the personal use of mobile phone systems to access personal information and data easily. This trend is observed in the use of wireless mobile systems. Antennas used for wireless mobile systems, including mobile phones, must be small in size, compact, and light in weight, and yet functional. In areas where 2G and 3G systems coexist, multiband antennas are

Modern Antenna Handbook. Edited by Constantine A. Balanis

required and, accordingly, small, compact, and lightweight multiband antennas have been used in both mobile terminals and base stations. Increased use of multimedia services has sped deployment of high speed, high data-rate transmission systems, for which advanced antenna systems like adaptive arrays and multiple-input multiple-output (MIMO) arrays have been developed.

Globalization has been promoted by use of mobile satellite systems; however, international roaming, made possible with some types of mobile phones, and connection of mobile systems to IP networks have led to even more global networks.

It is now understood that an antenna for mobile systems should not be treated independently as an isolated element, but instead should be designed by taking such parameters into consideration as propagation, system requirements, and environmental conditions [1], as Figure 23.1 illustrates.

The propagation problem should include parameters related not only to path loss, but also to signal transmission rate and bandwidth, delay spread, and Doppler shift in a Rayleigh fading environment. These parameters are particularly important in systems that employ digital modulation schemes, particularly ones with high data-rate transmission systems.

Antenna performance should meet the system specifications. In zone systems, radiation patterns are designed to match the zone patterns to serve impartially for mobile stations and, necessarily, to avoid interference through low sidelobes directed outside the zone, where the same frequency is used. Antennas should also function to reject interference and mitigate any bad effects of multipath fading.

Factors such as effective use of channels, type of information, and modulation scheme are significant in designing antennas for both mobile and base stations. Consideration of environmental conditions is also important in the design of antenna performance. In some ways, proximity effects may degrade antenna performance (e.g., materials near an antenna element may reduce antenna performance), but in other ways, they may enhance antenna performance by acting as parasitic parts of the radiator. The propagation path can be treated equivalently as a transmission circuit, and evaluation of its characteristics will assist in obtaining an optimum communication link, by which the best signal transmission

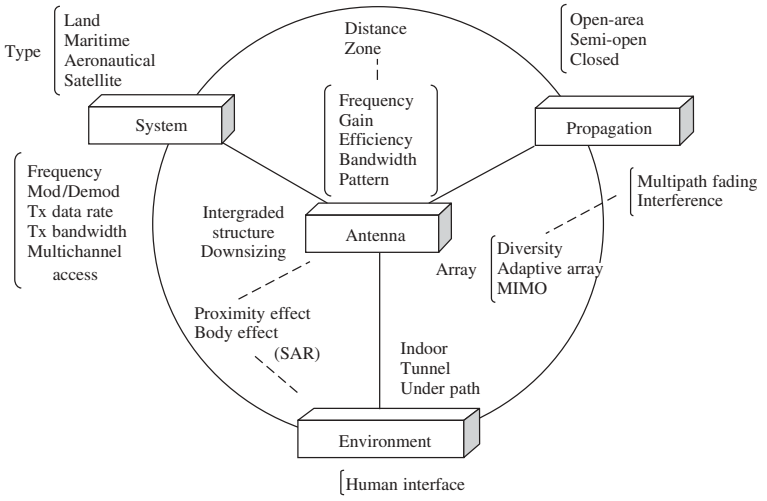


Figure 23.1 Antenna systems as an integral part of a mobile system.

TABLE 23.1 Antenna System as a Part of Mobile System

Requirement	Implication
Antenna as a system	Not as an isolated receive/transmit terminal
Designed to accommodate propagation effects	Diversity, rake reception, and introduction of adaptive antenna arrays
Transmission of high data-rate signals	Smart antennas, MIMO systems and SDMA
Compatible with environmental conditions	Pattern characteristics to match zone requirements and reduction of nearby obstacle effects
Downsizing	Small size, compact, light weight, built-in structure
Integration of antenna with platform and nearby materials	To include proximity effects in antenna design
Latest manufacturing technology	Exploitation of new composite materials, integrated electronic technology, and high density compact circuit structure
User-friendly and reliable performance	Minimum complexity in human interface; high reliability of mechanical design
EMC constraints	Reduction of EMI and SAR levels
Multimedia applications	Enhancement of antenna performance; wideband, multiband, MIMO system

is obtained. In mobile terminals, antennas should be designed so that SAR (specific absorption rate) values are as small as possible next to human brain tissue.

Table 23.1 summarizes the influence of system requirements on antenna design [1]. It must be stressed that the antenna should be considered as an integral part of the overall system.

23.2 BRIEF HISTORICAL REVIEW

Variation of related subjects and the evolution of antenna systems with time are exhibited in Table 23.2 in terms of frequency bands, systems, antenna types, propagation, and devices.

The first mobile communication was initiated in 1885 with the wireless telegraph between trains and stations, which was developed by Thomas Edison [2]. Trolley wires were used as the antenna, which coupled electrostatically with a metal sheet installed on the ceiling of the train, thus allowing the telegraph signal to be received. Edison also experimented with communication on a vehicle in 1901 [3, p. 133] using a thick cylindrical antenna placed on the roof of the vehicle. Practical mobile communication services started with the wireless telegraph on ships in 1889, developed by Guglielmo Marconi using long vertical wire antennas in various forms such as T, inverted-L, and umbrella shapes. It is interesting to note that portable equipment appeared in 1910 [3, p. 134].

Both World Wars I and II raised the need for advanced antennas and promoted the evolution of novel designs and technology [4]. Wire antennas were firmly established in

TABLE 23.2 Variation of Related Subjects and Evolution of Antenna Systems with Time

	1900–	1950–	1970–	1990–	2000–	2010–
Frequency	10 kHz<	30 MHz<	800 MHz<	1.9 GHz<	2 GHz<	3 GHz 5 GHz
System	Telegraph/telephone for train, ship, aircraft, police cars, portable receiving and transmitting	Vehicle and portable systems for business, navigation, taxis, pagers (tone), phone on vehicles, portable equipment	Mobile phones (analog), cordless phones, pagers (data), GPS	Mobile phones (digital), satcom, aircraft (voice and data), personal phones	Mobile phones (multi media), Wireless-Access (WLAN, WiMAX RFID, Bluetooth) ITS	BWA, NWA, UWB, Wearable com ITS, Adhoc NW
Antenna	Monopole/dipole, whip, Top-loaded monopole, Inverted L, loop	Blades, Ferrite coil antenna, Helical antenna	Corner reflector, PIFA, bifilar helix, MSA and arrays, Parallel-plate, printed antennas, LCX, diversity, Phased array	Meander line, Normal mode helix, Ceramic chip antennas, Adaptive array	Wideband, Multiband built-in antennas, Adaptive array, MIMO MP–TV reception	Small, compact, functional antennas, Implant antennas

Note: PIFA: Planar inverted-F antenna NWA: New wireless access MSA: Microstrip antenna BWA: Broadband wireless access LCX: Leaky coaxial cable UWB: Ultra wideband

the 1920s, while various advanced antenna designs and technologies still being used today became commonplace in the 1950s. In the 1960s a new antenna era emerged, triggered by revolutionary progress in semiconductors and integrated circuits, attributed initially to the defense industry but subsequently carried forward into the commercial sector. The demands of mobile communications inspired designers to create compact, lightweight, low cost, easy-to-manufacture radiating structures, compatible with the newly conceived integrated electronic packages. Creation of printed antenna technology, which lends itself to multifunction antenna devices [5], is most noteworthy. Some of the salient factors that have increasingly influenced antenna design in this era and continue to do so today are noted in Table 23.3, which clearly emphasizes that communications, and particularly mobile communication systems, are the most significant drivers of antenna technology at present.

Mobile communication systems have made rapid progress since mobile phone systems were introduced in 1979. The early analog systems (1G systems) conveyed mostly voice but demonstrated the benefits of communications on the move. It then became a good spur to advance other mobile systems, which dealt with both voice and data in digital form. The 2G systems have subsequently advanced to 3G systems, which have demonstrated transmission of images, both still and moving. Table 23.4 shows

TABLE 23.3 Salient Factors Related to Mobile Antenna Design

Factor	Trends
Spectral congestion and utilization	Improved performance; wide band, multiband, interference rejection, multibeam, and use of millimeter and submillimeter antennas
Explosive growth in mobile/personal communication systems	New small compact higher performing antennas for cellular and other mobile wireless systems
Increase in high rate information and data transmission	Wider bandwidth "smart" antennas and MIMO systems
Growth in mobile SATCOMS	Higher performance space-borne antennas offering multifunction operation, small, high-performance handset antennas
Link with IP and optical-fiber networks	Small high-performance antennas for mobile terminals
Intelligent traffic information, control, and management systems	Specifically designed antennas for both vehicles and roadside infrastructures
Development and application of new materials	Redesign of existing and creation of new functional, high-performance antennas, and application of metamaterials
Impact of computer modeling for analysis, design and measurement	Strengthens design methods to create higher performing antennas; development of EM simulation method to deal with small, complicated antenna structure
Public awareness of electromagnetic radiation and safety	Preference for lower transmitted powers antennas, reduction of EMI and SAR
Increase in functional antennas	Development of evaluation and simulation method
Development of wearable communication	Analysis and design of implant antennas

TABLE 23.4 Evolution of Mobile Systems with Time

Year					
1980–	1990–	2000–		2010–	
1G (analog) Voice	2G (digital) Voice and data	3G (multimedia) Multimedia	3.5G High rate data	3.9G Tx	4G High mobility, High speed Tx,
AMPS TACS, JTACS, NTT, NMT	GSM, PDC, PHS DECT IS-95, Cdma-one	IMT-2000, Cdma2000, WCDMA, Mobile SATCOM WLAN RFID	AdvIMT2000	LTE	IMT advanced, NWA, Ubiquitous NW
			EV-DO Rev A, Rev B HSDPA, BWA, WiMAX, UWB, NFC		
		MP-TV reception			

HSDPA: High speed down link packet access LTE: long term evolution EV-DO: (Evolution data-only) (Cdma 1× EVDO) BWA: Broadband wireless access NFC: Near field communications RFID: Radio frequency identification Tx: Transmission Rev A(B): EX-DO Rev A (B) NWA: New wireless access MP: Mobile phone UWB: Ultra wideband.

this evolution of mobile systems through time. The 2G systems showed the advantages of digital systems over analog systems with respect to increases in the available capacity and services provided. Digital systems have opened the way for global standardization of mobile communications. The typical European digital system, GSM (Global System for Mobile Communications), which initiated the operation of digital mobile systems first in the world, has paved the way for 3G systems, in which CDMA (code division multiple access) technology has been applied and more increased capacity and enhanced services, including a link to the Internet, have become available.

Now systems referred to as the IMT-2000 (International Mobile Telecommunication Systems)/UMTS (Universal Mobile Telecommunication Services) have made progress and are advancing toward 4G systems, which can deploy greatly enhanced services with higher quality and greater capacity for transmission.

Some recent mobile wireless data systems have become competitors to mobile phone systems in terms of high speed transmission with high mobility.

Antenna technology has made progress along with the advancement of these mobile systems.

Mobile phone systems, which began with automobile telephone systems, have now evolved into hand-held phone systems. Today automobiles with an installed mobile phone can hardly be seen, whereas the number of hand-held phones far surpasses the number of automobile phones.

There have been notable changes in antenna design for mobile phones in the recent decade. The monopole was used in both automobiles and portable equipment for a long time; however, use of planar antennas has prevailed in portable equipment as well as the monopoles. In current hand-held phones, planar antennas, typically PIFAs (planar inverted-F antennas), are employed. In some systems, particularly in GSM systems, variations of PIFAs have been used as built-in antennas. Presently, the built-in antenna has become a worldwide trend in mobile phones. Antennas in mobile terminals of 3G systems are various types of small, multiband, modified PIFAs. Further advancement in antenna systems is expected; for instance, we are likely to see mobile terminals with adaptive arrays (antennas having the capability of adjusting adaptively to environmental conditions so that variation of antenna performance due to environmental effects is minimized), as well as antennas with more sophisticated functions, yet with smaller and more-compact structures.

23.3 TRENDS

Typical trends in modern mobile systems are classified in five ways: personalization, globalization, increase of multimedia services, progress to multidimensional network, and software implementation. The trends are listed in Figure 23.2 [1], in which demands and antenna structure in conjunction with trends are described.

23.3.1 Personalization

Downsizing hardware in both mobile stations and base stations seems to have accelerated personalization. A typical example is observed in mobile phones. Mobile phones can be used for getting information of various contents, including games, movies, TV broadcasting, and music, which attracted many users and accelerated personalization for

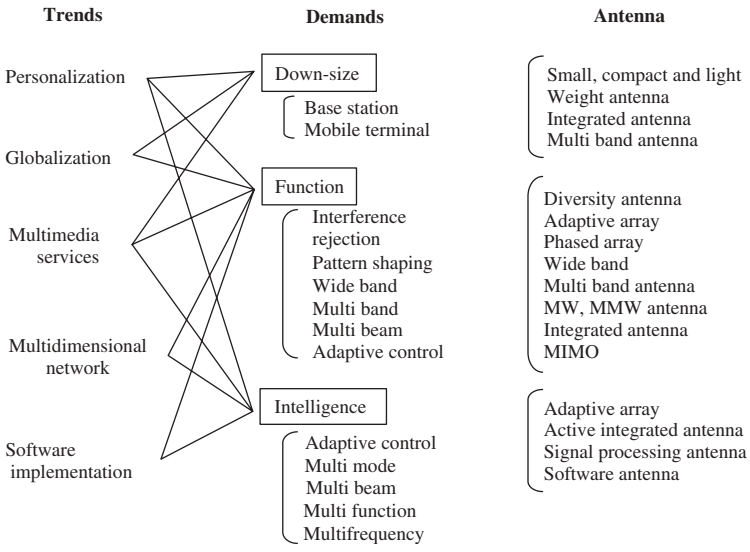


Figure 23.2 Typical trend of mobile communications and antennas.

mobile phone users. Mobile phones can access Internet facilities, so they can function as e-mail, trading, and ticketing terminals and game stations. Some types of mobile phones have a receiver of digital TV broadcasting installed. Hence mobile phones are no longer considered to be mere "telephones" but are information terminals, which deal with message exchange, transmitting/receiving moving images, entertainment, e-commerce, and so forth.

In another aspect, downsizing of mobile terminals has also given impetus to the personalization of mobile systems, because the smaller handsets are more convenient to carry and easier to operate. One of the biggest problems encountered is the realization of smaller antennas for downsized handsets without degrading system performance. Downsizing of base stations (BSs) has also required antennas to be small in size and light in weight as a consequence of the increase in the number of small base stations to serve increasing numbers of subscribers, and the need for more efficient usage of channels.

Recent development of wireless systems that are used over very short ranges, such as RFID (radiofrequency identification) and NFC (near-field communication) systems, has further accelerated personalization. Antennas used for these applications are usually small and compact, but not necessarily high gain.

In wireless broadband systems, various types of functional antennas like MIMO systems have been developed so that high speed and high data-rate transmissions are feasible.

23.3.2 Globalization

Globalization of mobile communications has included using satellites on low Earth orbit (LEO) or medium Earth orbit (MEO) as well as geostationary Earth orbit (GEO). However, global communication services depend not only on satellite systems but also on wired systems like IP-based networks (Internet Protocol networks), which have worldwide linkage and connections to mobile networks. In addition, there are also wireless systems in which mobile terminals can roam from one country to another country where

the same network services are available. Typical systems are GSM and some types of 3G systems, which have deployed their network in many countries. Dual- or triple-band antennas are mounted on mobile terminals for these systems.

23.3.3 Multimedia Services

Mobile systems are now capable of transmitting video images, both still and moving, as well as voice and data. Since the transmission rate has gradually been increased from the order of kilobits per second (kbps) to megabits per second (Mbps), various services, including videophone, movies, and large-scale packet data transmission, have become available. FOMA (the 3G system in Japan), for instance, which employs W-CDMA, can carry a maximum of 384 kbps of data presently; however, it is being raised to a maximum of 14.4 Mbps by introducing a HSDPA (high speed downlink packet access) system. This type of service is referred to as a 3.5G system, which stands for a system in transition from 3G to 4G. The data rate will be further raised to 100 Mbps, to provide further sophisticated multimedia services. In future wireless systems, the speed of data and information services is expected to be on the order of gigabytes per second (Gbps), even in high speed motion. For these systems, smart antennas such as adaptive arrays and MIMO (multiple-input multiple-output) systems will be applied.

23.3.4 Multidimensional Services and Network

There is a growing trend for certain systems to be integrated into other systems so that multidimensional services can be made available. A typical example is a combined communication system with broadcasting; TV broadcasting through a communication satellite, for example. An even more notable example is reception of terrestrial digital TV broadcasting by a mobile phone terminal. This type of service is referred to as "one-segment broadcasting," because one segment of the 13 OFDM (orthogonal frequency division multiplexing) segments is exclusively designed for broadcasting to mobile terminals, even when they are moving. Another example is the FMC (fixed mobile convergence) system, in which a mobile network is combined with a wired network in a home or office environment. Mobile terminals can be used exactly like fixed phones at home, while outside the home they function as a normal mobile phone. A mobile phone connects itself to either a fixed network or a mobile network, depending on the environment it finds, without conscious direction by the user.

In general, communication systems are now becoming integrated into multidimensional networks that embrace multi-informational media, multitransmission media, and multilayered networks. Information media are composed of both voice and nonvoice systems, including digital voice, sound, still and moving images, and computer data. Transmission media are both wire and wireless lines, including IP networks, radio, and optical networks. Together, they will constitute a seamless network regardless of the types of systems and communication ranges. Land, maritime, aeronautical, and satellite systems may have integrated structures to constitute complex multilayered networks, thus allowing global communications regardless of time and space. Wireless communication systems in ITSs (intelligent transportation systems) are now combined with wired systems using IP networks, which will expand the capabilities of information transmission and data exchange among transportation systems and roadside infrastructure.

Ubiquitous environments will be realized by means of wireless mobile systems, especially small mobile terminals, to which IPv6 (Internet Protocol version 6) networking will be employed.

The demand for more intelligent antennas that are integrated with adaptive control and software will continue unabated. Applications of higher frequencies such as in microwave, millimeter-wave, and even terahertz-wave regions, will be explored further to develop novel antenna systems.

23.3.5 Software Implementation

Implementation of software to interact with antenna systems will be increased in order to enhance antenna performance as well as system performance. An example is a mobile terminal containing software that enables the mobile terminal operating in one system to be switched to operate in another system, when a user desires to use a different system in a different environment. An operator who is using a GSM terminal can switch the terminal to receive the W-CDMA services when the operator enters a W-CDMA zone. A mobile system can correspond to several systems by means of software control, which reconfigures system structure as needed; for instance, a receiver can receive FM broadcasting or TV broadcasting, and can operate as a mobile phone terminal or an ITS terminal, without a change in hardware, by using software switching. Antennas in these systems are required to be wideband or multiband. Adaptive control would be desired for some cases where the system complexity increases. For some specific applications, antennas should have a reconfigurable structure, which is adaptively controlled by the software installed for that purpose.

23.4 TYPICAL ANTENNA ELEMENTS

Various antenna elements have been used for mobile communications. There is no single specific type of antenna element, but the designs differ depending on the practical mobile systems. Figure 23.3 illustrates the most typical antenna elements: typical elements having linear structure are the (a) dipole, (b) monopole, (c) loop, (d) inverted-L, (e) inverted-F, (f) normal mode helix, and (g) meander line; planar structures are the (h) microstrip antenna (MSA), (i) planar inverted-F antenna (PIFA), (j) parallel plate antenna, and (k) slot antenna; a bulk-structure antenna is depicted in (l), which is a small ceramic chip antenna.

In a bulk structure, a thin metallic element is either printed on or encapsulated in the high permeability ceramic material. The printed patterns, for example, are a type of normal mode helix, inverted-L, inverted-F, or other unspecified antenna type. Antennas made of material that is a synthetic resin combined with ceramic material have features of flexibility in manufacturing, and an antenna can easily conform to the place where the antenna must be mounted, even in narrow and uneven places. Typically, the material is less hard than ceramic and can be formed to fit many structures.

23.5 ANTENNAS FOR BASE STATIONS

The basic design concept of 2G base station antennas does not essentially change for 3G systems. However, 3G systems use different modulation schemes, by which the

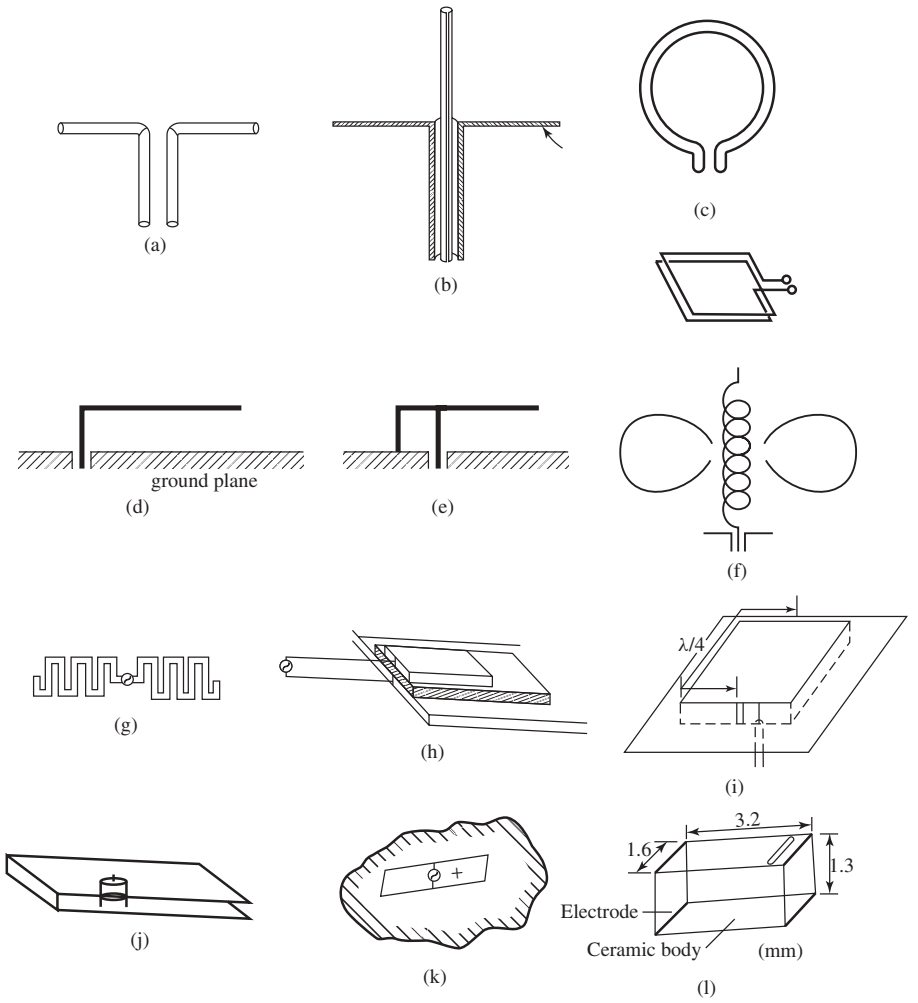


Figure 23.3 Typical antenna elements: (a) dipole, (b) monopole, (c), loop, (d) inverted-L antenna (ILA), (e) inverted-F antenna (IFA), (f) normal mode helical antenna (NMHA), (g) meander line antenna (MLA), (h) microstrip antenna (MSA),(i) planar inverted-F antenna (PIFA), (j) parallel plate, (k) slot, and (l) chip antenna.

transmission performance differs from that of 2G systems; thus there should be some differences in the antenna design.

Figure 23.4 indicates general items that should be taken into consideration when a base station antenna is designed. Figure 23.5 provides design issues in conjunction with requirements and antenna technologies. A major subject that affects antenna design in 3G systems is the pattern synthesis, which concerns beam shaping, multiband operation, downsizing, and sophistication of antenna systems.

Beam shaping includes beam tilting and low sidelobe beams in the vertical plane, along with uniform coverage patterns, sector beams, and multibeams in the horizontal plane.

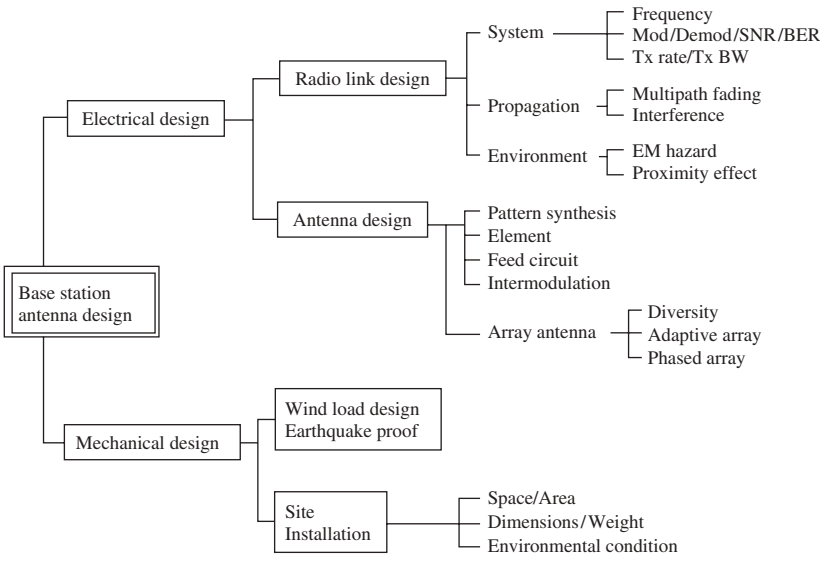


Figure 23.4 Key items in designing a base station antenna.

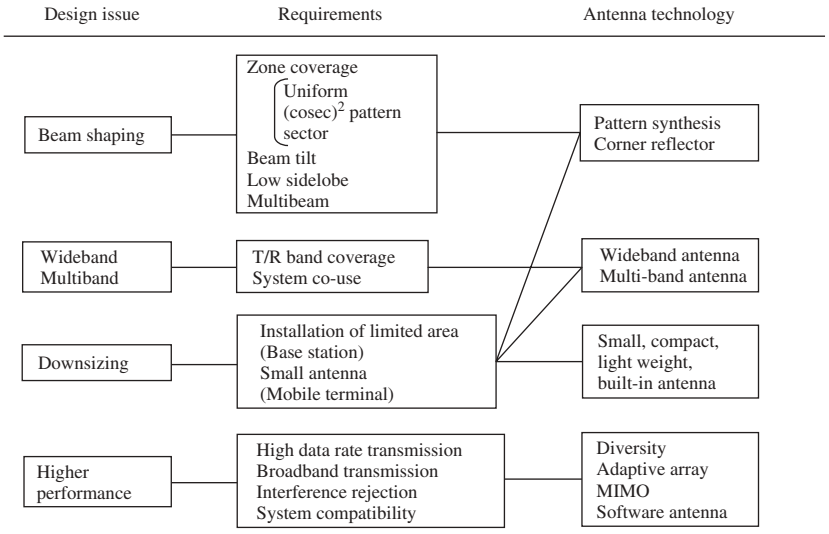


Figure 23.5 Base station antenna design issues, requirements, and antenna technology.

Multiband antennas are required for systems that provide services on both 2G and 3G systems.

Downsizing has become a serious concern in constructing base stations, since with the remarkable increase in the number of subscribers, cell sizes have become smaller to increase channel capacity, and accordingly the number of base stations has increased. Thus base stations often must be installed in limited spaces and places where heavy weight

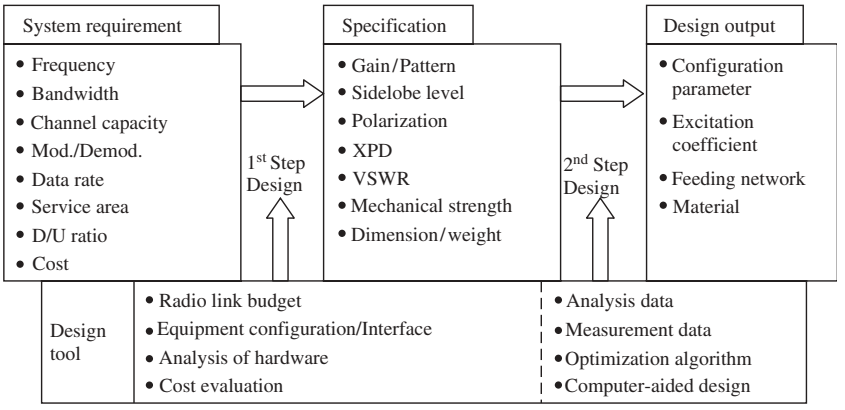


Figure 23.6 Design steps of base station antennas.

is not allowed. As a result, base-station antennas are often required to be downsized by substituting smaller, more compact, and lighter-weight structures. In order to reduce the number of antennas, because of the installation in confined spaces, multiband and multibeam antennas are being developed.

Figure 23.6 depicts the design steps of basestation antennas. Performance and cost sometimes are considered first, while determination of the electrical and mechanical parameters is taken in the second step [1].

There also have been urgent requirements for increasing coverage areas in 3G systems, so that service areas can include closed areas like rooms in buildings, underpasses, inside tunnels, subway stations, and so forth. In order to service such areas, small relay stations have been developed, and again, small, compact, and lightweight antennas have been used for these stations. Antennas used for these systems have planar structure.

In areas where 2G and 3G systems coexist, the number of antenna elements may be increased to serve three bands, for example, 0.8, 1.5, and 2 GHz, depending on national standards. However, since space to install antennas often is limited to narrow areas, the total number of antennas should be reduced. To meet this requirement, multiband antennas that are small in size have been developed. Multibeam antennas have also been developed for covering sector zone areas.

Now that the mobile phone system is evolving toward 4G systems, and other wireless mobile systems are also seeing rapid growth, feasibility studies for smart antenna systems such as MIMO and adaptive arrays are being done and an operational test is being performed in the Tokyo metropolitan area.

23.5.1 Propagation Problems

Propagation problems differ depending on environmental conditions. Propagation in mobile communications occurs within a diffraction region and differs from that in free space. Path loss in free-space propagation is simply proportional to the square of the distance, while that in mobile communications generally depends on the operating frequency, antenna height, and particular environmental conditions around the antenna as well as the propagation path.

Mobile propagation environments are very complicated but are roughly categorized into four kinds [6].

1. *Open Area:* There are few obstacles such as high trees or buildings in the propagation path. It can be said roughly that free spaces of about 300–400 m in length lie between the base and mobile stations. The propagation path is always on the line of sight (LoS).
2. *Suburban Area:* There are some obstacles around the mobile stations, but they are not dense. Roughly speaking, it is an area of trees and low houses.
3. *Urban Area:* There are many buildings or other high structures, and hence it is an area with high, close buildings, or a densely mixed area of buildings and high trees. No line of sight (NLoS) may exist in a typical propagation path.
4. *Closed Area:* Propagation is confined in limited areas like in buildings, tunnels, subway stations, and underpasses.

23.5.2 Propagation-Path Loss Characteristics

Path loss in the mobile propagation environment depends on the environmental conditions between base station and mobile station. In free space, it is proportional to d^{-2} (d is the distance between the base station and mobile station), whereas in mobile systems it should be $d^{-\alpha}$, where α varies from 3.5 to 7, depending on the propagation path. The path loss characteristics are major factors in designing a communication link, and antennas should be designed to meet the requirement associated with the link budget.

23.5.3 Base Station Antenna Techniques

23.5.3.1 Requirements for Base Station Antenna Systems Figure 23.7 illustrates typical technology necessary for designing base station antennas. In order for the base station to communicate with the mobile stations located in the service area, base station antennas must radiate uniformly inside the area. Moreover, antenna gain should be as high as possible. Since the pattern of the service area is specified, antenna gain

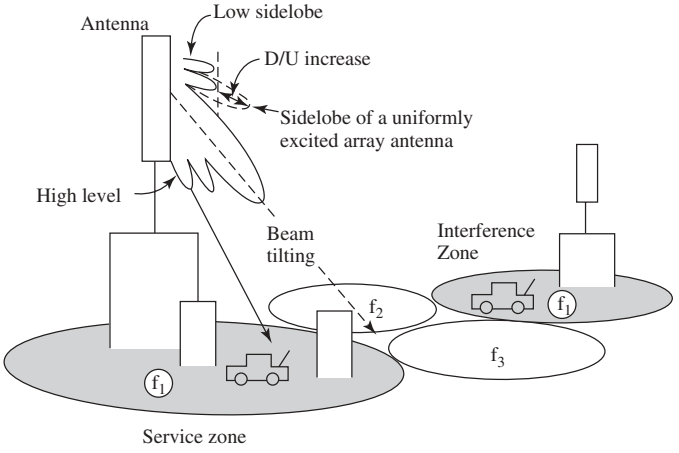


Figure 23.7 Effect of side reduction for frequency reuse.

cannot be increased by narrowing the beamwidth in the horizontal plane. Therefore it is customary that the antenna beam is narrowed in the vertical plane to increase gain; a vertical array of linear antennas is used for this purpose. Antennas with a gain in the range of 10–20 dBd are normally used for base station antennas in cellular systems.

Since a base station communicates with many mobile stations simultaneously, multiple channels must be handled. This requires wide-frequency characteristics and a function for branching and/or combining the channels. Hence base stations in the cellular system use one antenna for both transmitting and receiving. The required bandwidth of the antenna, for example, in the 800-MHz band Japanese PDC system, is more than 7% for the specified voltage standing wave ratio (VSWR) of less than 1.5. In addition, antennas should be designed for use with several systems (e.g., analog and digital systems). With that requirement, the antenna for the PDC in the 800-MHz band then needs frequency coverage between 810 and 960 MHz, a 17% relative bandwidth. Furthermore, since 3G systems have started services in areas where 2G systems remain in service, antennas for both 2G and 3G systems should be designed.

In 2G systems, the frequency-reuse concept is adopted in order to improve the channel capacity, and beam tilting is employed in base station antennas in order to avoid the co-channel interference between cells where the same frequency is used. In 3G systems, since the modulation scheme is CDMA, every cell uses the same frequency, and the beam-tilting concept is still applied to base station antennas. Interference suppression is still needed so as to increase the channel capacity. Shaped beam technology is still a major design subject.

Another important issue, which needs specific consideration in mobile communications, is to mitigate multipath fading, which deteriorates the signal quality in the narrowband modulation schemes. One way to overcome this problem is to apply diversity antenna systems.

Various antenna technologies for base stations other than diversity systems have been developed to enhance system performance. The adaptive array is one of them. Major subjects in the process of adaptive array development are communication frame format for beam forming, beam forming algorithm, array calibration, and simplification of hardware. The world's first application of adaptive arrays for mobile systems was to the PHS base station antennas in Japan in 1998. The PHS system is a likely system to apply adaptive arrays, because PHS employs a TDD (time division duplex) system and allows slow movement of mobile terminals. Thus the estimation of the propagation channel property for the downlink can be used for the uplink, which makes feasible the application of adaptive arrays to mobile systems. The iBurst system evolved from the PHS system, and it adopted adaptive arrays for the purpose of interference rejection and SDMA since 2004.

23.5.3.2 Types of Antenna Base station antenna configurations depend on the size and shape of the service area, the number of cells, and the number of channels. For limited service areas within a restricted angle in the horizontal plane, a corner reflector antenna is often used. When the service area is wide, as in a macrozone system, a linear array antenna, which has high directivity in the vertical plane, is used.

In the early stage of cellular system development, the base station antenna was mainly designed to achieve higher gain, and uniformly excited array antennas were usually used. However, the design concept of the base station antenna has been shifted from attainment of high gain to a greater ratio of desired-to-undesired signal strength (D/U), as cells have been divided into smaller subcells in order to increase the effectiveness of frequency

reuse. Main-beam tilting, either mechanically or electrically, has been adopted throughout the world and co-channel interference can be reduced by about 10 dB. The beam tilting was recognized to be essential for enhancing frequency reuse. Suppression of sidelobes adjacent to the main beam, achieved by synthesizing array antenna patterns appropriately, was also effective in decreasing the distance between cells.

As for diversity antennas, space diversity, in which two antennas are used with separation of 5–10 wavelengths, has commonly been used. Other diversity schemes, such as pattern diversity [7] and polarization diversity [7], have also been applied to base station antennas in commercial systems.

Polarization diversity has gained a new understanding: better performance compared to space diversity has been recognized in urban areas, particularly in dense areas like central Tokyo [8]. This is attributed to an increase in cross-polarization components in mobile phone propagation with the increase in mobile phone users. When a mobile phone is held in a normal talk position, the unit is slanted and produces both vertical and horizontal polarization components. Consequently, polarization diversity performs better than space diversity. Polarization diversity has been employed in CDMA-One systems and presently in IMT-2000 systems.

Figure 23.8 categorizes base station antennas from the viewpoints of types, functions, and antenna structures.

23.5.3.3 Design of Shaped-Beam Antennas

Shaped-Beam Antenna A base station antenna in a cellular system is required to illuminate the cell uniformly at as high a level as possible and radiate energy at as low a level as possible outside the cell, where the same frequency is reused. Beam shaping is required for forming (1) a sector beam in the horizontal plane and (2) a cosecant beam in the vertical plane. In the vertical plane, a tilted beam is also necessary in order to reduce interference.

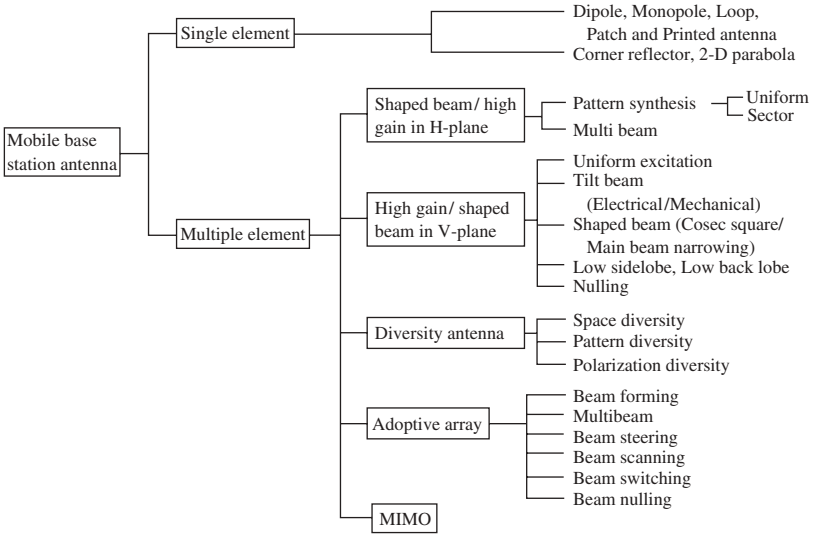


Figure 23.8 Classification of base station antennas.

In 2G mobile phone systems, a frequency reuse system is adopted in order to increase effectiveness of channel capacity. The zone system is designed by taking parameters such as D/R and C/I into consideration, where D/R stands for ratio of distance D between the centers of adjacent cells and cell radius R, and C/I stands for ratio of carrier level C and interference level I (Figure 23.9) at the edge of the cell and the worst value is given by including (C/I) ANT; power ratio on an antenna pattern at the desired wave direction and the interference wave direction, as follows [1]:

$$C/I = -\alpha \cdot \log\{R/(D-R)\} + (C/I)_{\text{ANT}} \tag{23.1}$$

where α denotes the propagation loss factor. Figure 23.10 expresses relations between D/R and C/I, and Figure 23.11 illustrates a typical radiation pattern of a base station antenna in the vertical plane and radiation levels at the edge of the cell and toward the

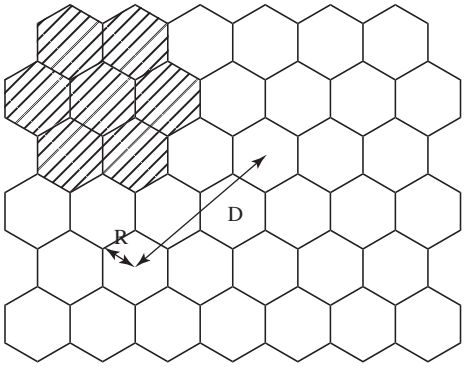


Figure 23.9 Cell configuration in cellular system.

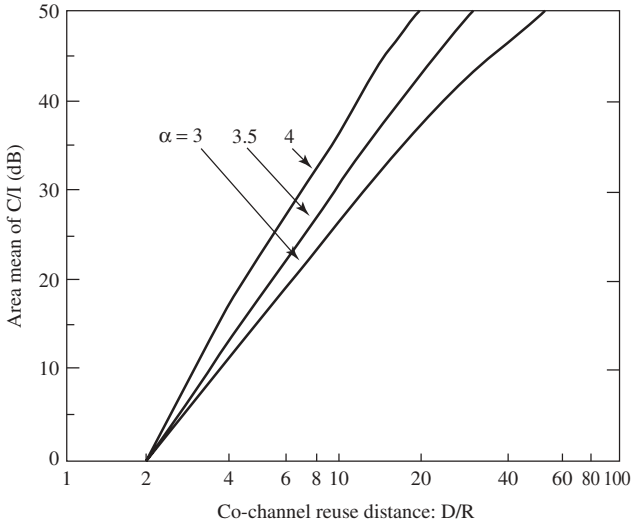


Figure 23.10 Relation between D/R and C/I.

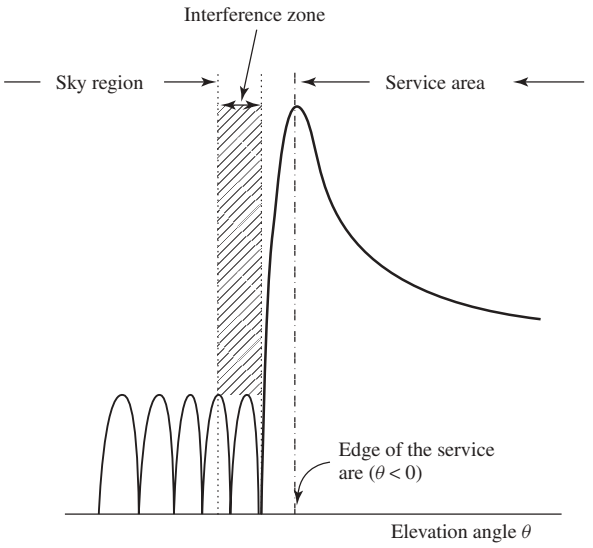


Figure 23.11 Typical radiation pattern of base station in the vertical plane.

interference direction. If the two directions are separated enough compared to 50% of the main beamwidth of the antenna, interference points exist in the sidelobe region of the radiation pattern, and the sidelobe level must be decreased to increase $(C/I)_{\mathrm{ANT}}$.

On the other hand, in W-CDMA systems the frequency reuse scheme is not employed, as the same frequencies are used in all cells. Hence allocation of base stations, and thus zone configuration, does not need to follow design concepts taken in 2G systems, but may depend on traffic distributions in the service area.

In order to further increase the effectiveness of frequency utilization, a system in which a cell is divided into sectors has been adopted in mobile phone systems. By using the sector zone system, the number of cells is increased effectively without an increase in the number of base stations. However, in the frequency reuse system, the effectiveness of frequency utilization cannot really be improved in proportion to the number of sectors, because it depends on the amount of frequency reuse, not on the number of sectors. On the contrary, in W-CDMA, the effectiveness increases as the number of sectors increases. In addition, with an increase in the number of sectors, beamwidths become narrower and hence antenna gain can be made higher.

Shaped Beam in the Horizontal Plane By sectoring a zone, the average number of mobile terminals in a sector will decrease and accordingly interference will be reduced. This will result in increasing channel capacity of the system.

Figure 23.12 shows the increase in channel capacity with the number of sectors [9]. Here the antenna pattern is expressed by

$$f(\theta) = (\cos\theta/2)^m \tag{23.2}$$

where

$$m = -3/[10\log(\cos\theta_b/4)],\ \theta_b = \text{beamwidth} \tag{23.3}$$

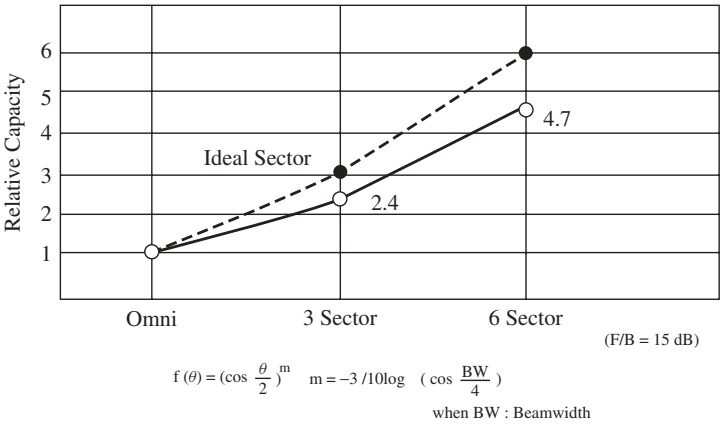


Figure 23.12 Channel capacity versus number of sectors.

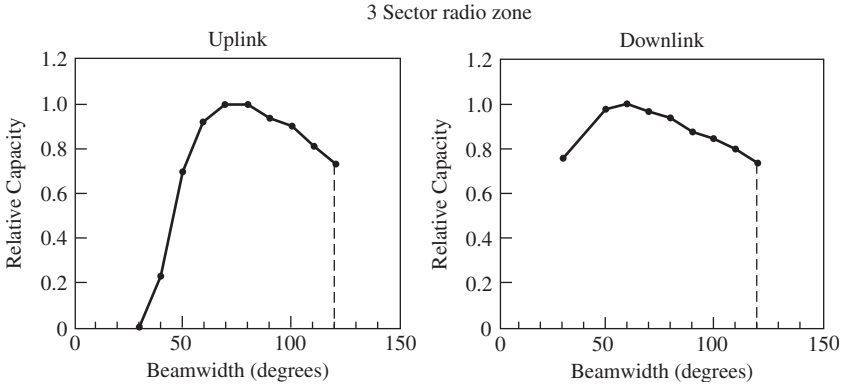


Figure 23.13 Channel capacity versus beamwidth in three-sector zone.

Variations of channel capacity with respect the beamwidth θ_b, calculated for cases where three- or six-sector zones are considered, are shown in Figures 23.13 and 23.14, respectively. In these figures, relative capacity in both uplink and downlink is shown [10]. Typical antennas having a sector beam are one with a corner reflector and a two-dimensional parabolic reflector antenna fed by two primary radiators [11].

The corner reflector antenna has an advantage in that the beamwidth can be adjusted by changing the aperture angle of the reflector. Figure 23.15 shows the fundamental configuration of the corner reflector antenna. The antenna element may be a vertically arrayed antenna. With vertically arrayed antennas, the corner reflector can achieve high directivity by narrowing the main beam in the vertical plane and can radiate a shaped radiation beam by controlling the excitation coefficient.

Figure 23.16 shows the relationship between the aperture angle of the corner reflector and a half-power beamwidth in the horizontal plane, as well as the relationship between the aperture angle and the directivity when the primary radiator is a half-wavelength

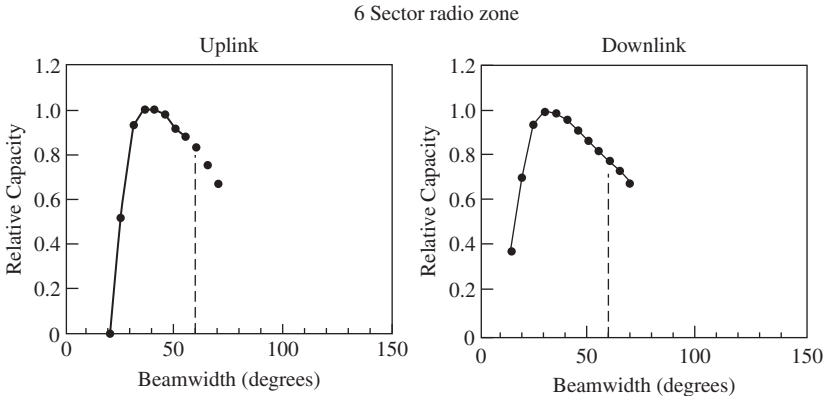


Figure 23.14 Channel capacity versus beamwidth in six-sector zone.

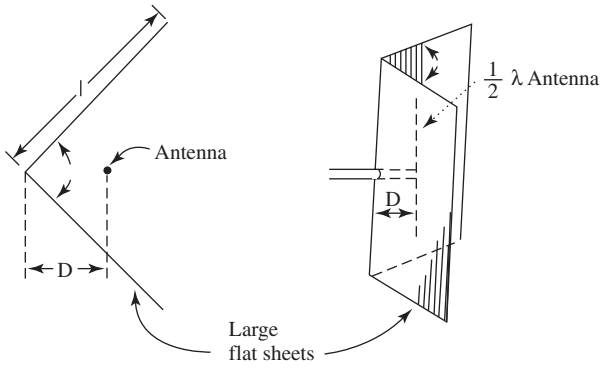


Figure 23.15 Fundamental configuration of a corner reflector antenna.

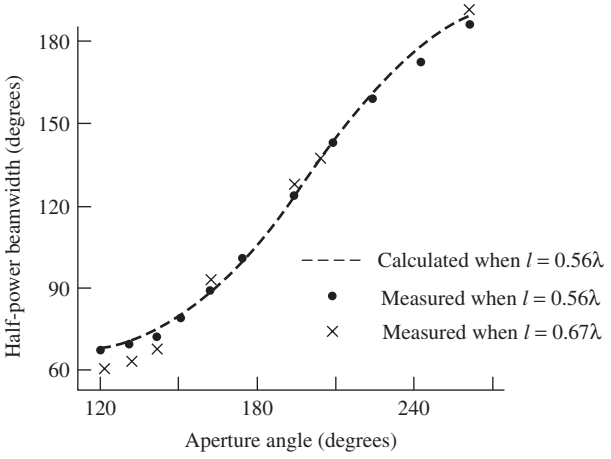


Figure 23.16 Aperture angle versus half-power beamwidth.

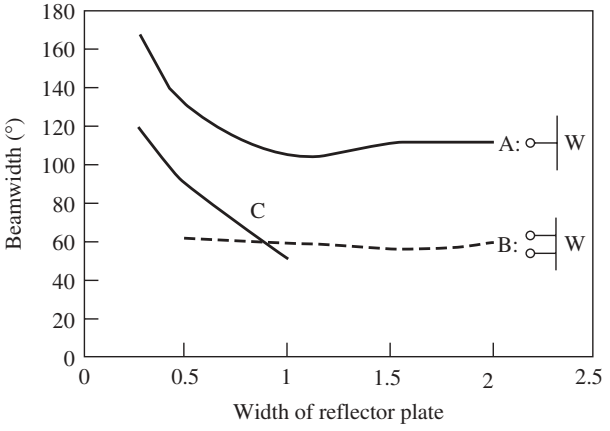


Figure 23.17 Beamwidth versus width of reflector plate: (A) with single dipole; (B) with two dipoles, using flat plate; and (C) using semi sphere reflector.

dipole. Sector beams with beamwidth from 60° to 180° can be obtained by setting the aperture angle from 60° to 270° [12].

When a flat plate is used as the reflector, the beamwidth becomes narrower as the reflector width becomes wider, as shown in Figure 23.17, in which the beamwidth is shown nearly constant for the reflector width greater than one wavelength. In the figure, A shows a case where a single dipole antenna is used and the beamwidth is about 120° with the reflector width of one wavelength (λ), while B shows when two dipole antennas with separation of 0.25λ are used and the beamwidth is about 60° with one wavelength reflector width. C depicts a case where a semicylinder is used as the reflector, by which the reflector can be made smaller as compared with a flat plate in order to achieve 120° beamwidth. For a 60° beamwidth, the cylinder diameter is 0.8λ, which is nearly the same as that of a flat reflector.

If a sector beam of more than 120° is desired, a corner reflector with an aperture angle of more than 180° is needed.

Shaped Beams in the Vertical Plane When a limited horizontal area is to be illuminated with equal received signal level from an antenna fixed at a certain height, as shown in Figure 23.18, an antenna with a cosecant-squared shaped-beam pattern in the vertical plane is used [1]. If the path loss is greater than that in free space (the path loss factor $\alpha \propto r^2$, the square of the propagation distance), as in the case of mobile communication systems, the p^{th} power order of a cosecant shaped-beam power pattern is necessary to achieve equal received signal level at all points in the area. However, in cellular systems, the significance of a shaped beam is recognized in suppression of the radiation toward the cell, where the same frequency is reused, rather than in uniform illumination of the self-zone. If frequencies are reused as in 2G systems, and the cells are closely spaced, a part of the main beam illuminates the reuse cell. Therefore it may be effective to tilt the main beam down to suppress interference, even if the received signal level within the self-zone weakens. By reducing the main beamwidth while maintaining the length of the antenna, it is possible to increase $(C/I)_{\text{ANT}}$. This concept can be applied to antennas in 3G systems.

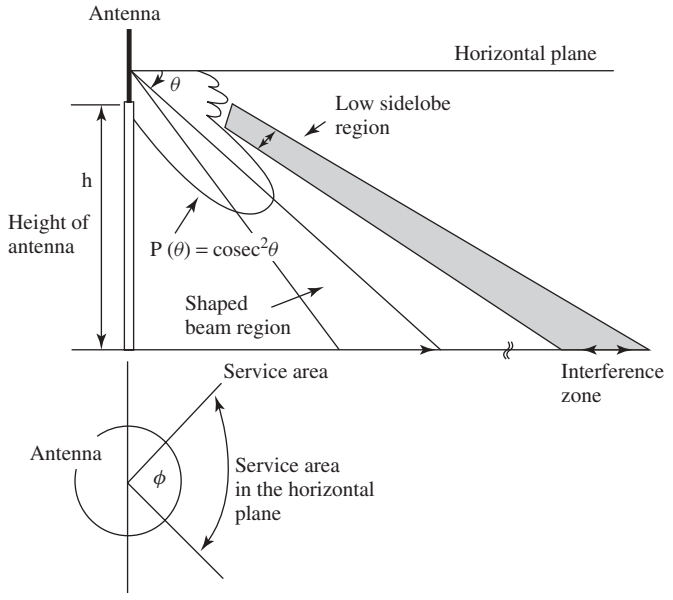


Figure 23.18 Shaped beam in horizontal plane.

The principal idea of the beam tilt-down technique is to tilt the main beam in order to suppress the level toward the reuse cell and to increase$(C/I)_{\mathrm{ANT}}$ in 2G systems. In this case, the carrier level also decreases in the zone edge. However, the interference level decreases more than the carrier level, so the total $(C/I)_{\mathrm{ANT}}$ increases. This is an advantage from the viewpoint of system design, and this technique is used in most cellular systems, including 3G systems, throughout the world.

In W-CDMA in Japan, array antennas are used, which consist of seven subarrays, each of which has four elements as shown in Figure 23.19. Four phase shifters are used to adjust phases in the array feed. The feed is composed of three parts; upper three, center, and lower three, as Figure 23.20 shows. The phase shifter consists of a triplate transmission line with meander line pattern and a movable dielectric sheet, which is placed over the triplate line to sandwich it, as Figure 23.21 shows. By moving and thus varying the area covering a part of the triplate line, phase velocity of the triplate line is varied and the phase of the array feeds can be controlled to obtain the designed pattern [13].

Shaped Beam with Locally Suppressed Sidelobe Level In order to reduce interference, sidelobe levels should be suppressed in some limited angles. Since the sidelobe suppression is performed in directions close to the main beam and the sidelobe level is closely related to the main beamwidth, it must be carried out very carefully. As shown in Figure 23.22, a main beam 30% narrower than that of a uniformly excited array was obtained by suppressing only several sidelobes near the main beam and setting the other sidelobes at a comparatively high level [14]. An antenna with this radiation pattern can increase the level at the zone edge by approximately 1.5 dB if the interference level is kept constant. This is a great advantage in the system design.

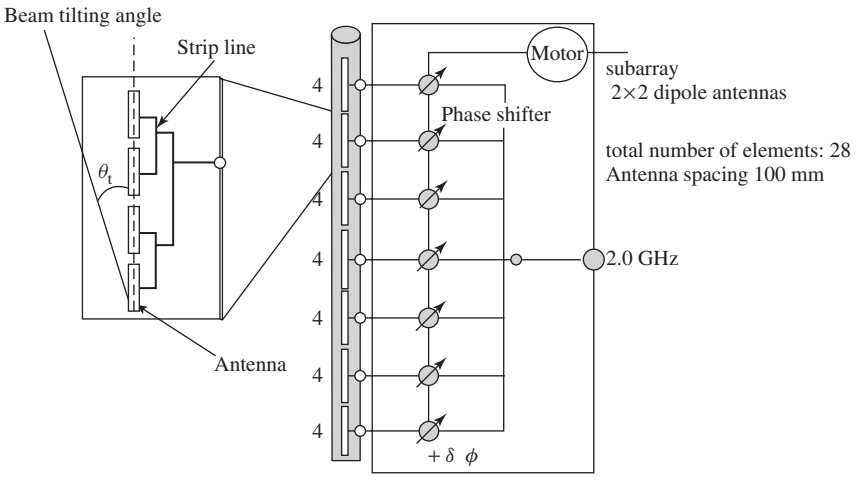


Figure 23.19 Base station antenna and beam tilting.

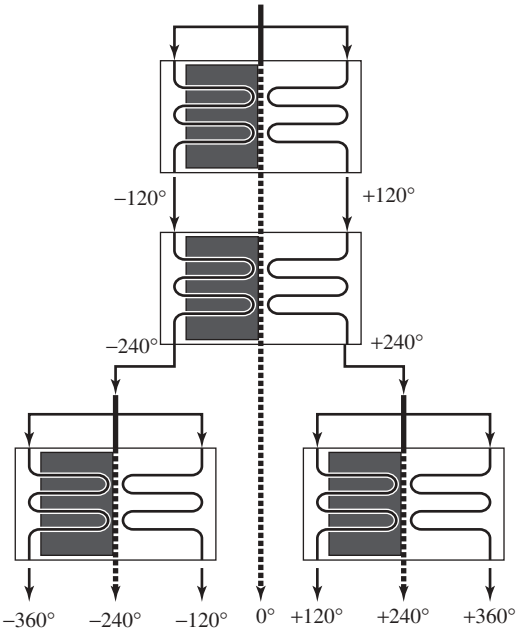


Figure 23.20 Feed network for seven-element array.

23.5.3.4 Wideband and Multiband Antennas In PDC systems in Japan, base stations employ wideband antennas, which operate for both transmitting and receiving bands, covering 800- and 900-MHz bands.

With the introduction of 3G systems, there are some areas where both 2G and 3G systems operate simultaneously and antennas for both systems are required. This requirement results in increased numbers of antennas; however, as the space to install base stations

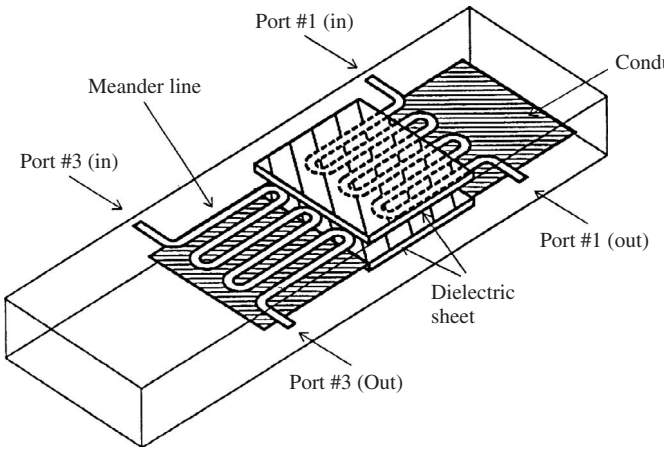


Figure 23.21 Configuration of variable phase shifter.

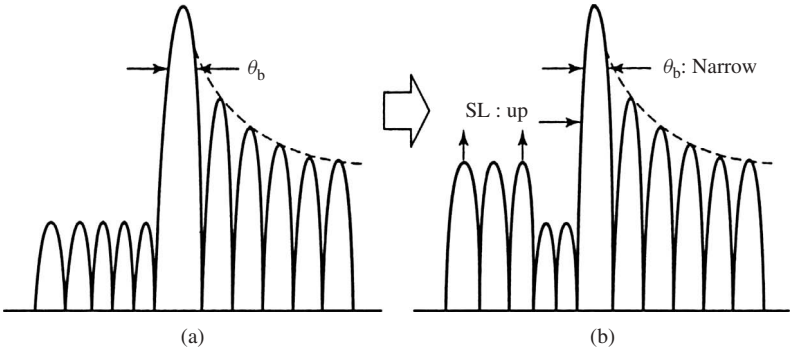


Figure 23.22 Narrow main beam designed by (a) suppressing sidelobe uniformly and (b) restricting several sidelobes.

is limited in confined areas, a decrease in the number of antennas is desired. Thus multiband antennas have been developed: two bands that cover 0.8- and 1.5-GHz bands in 2G systems, and three bands that cover 0.8-, 1.5-and 2-GHz bands in both 2G and 3G systems.

23.5.3.5 Downsizing of Base Stations and Antennas With the increasing number of subscribers, cell size is being made smaller and the number of base stations is increased. As a result, installation of base stations has been limited to some narrow areas and downsizing of base stations has been required. In accordance with this downsizing of base stations, antennas should also be downsized; antennas are designed to have smaller dimensions and the number of antennas is also reduced. Use of planar elements has prevailed in base stations as well as in mobile terminals. Use of multiband antennas is another promising way of downsizing.

23.5.3.6 Enhancement of Antenna Performance In order to mitigate Rayleigh fading effects, diversity and adaptive arrays have been used. The three major types of diversity schemes are space diversity, pattern diversity, and polarization diversity. The most popular system, space diversity, has been used in both base station and mobile terminals. The pattern diversity has been applied to create an omnidirectional pattern by using two 180° sector antennas. A pattern synthesized with locally suppressed sidelobe levels was first studied in the field of sector beam antennas. Polarization diversity has been employed in urban areas like Tokyo metropolitan areas, where both vertical and horizontal polarization components exist almost equally due to multipath propagation and use of mobile phone terminals. Mobile phones produce both polarizations because in the talk position, the mobile phone unit is slanted about 60° on average and the current flows on the unit, which align mainly in parallel with the unit, contribute to create both polarizations.

Another important antenna system is the adaptive array, which offers functions of multibeam forming, beam switching, beam steering, and null steering. The advantages of application of these functions have been verified by theory and also by operational tests. It was traditionally believed that implementation of adaptive arrays was very costly and complicated; however, they were introduced for PHS base stations in 1998 and their usefulness has been highly evaluated through their operation in Tokyo and some other cities in Asian countries. In 2002, an advanced adaptive array system was adopted by a PHS system, in which SDMA was introduced.

23.5.4 Practical Base Station Antennas

In Japan, mobile telephone services started in 1979. After the liberalization of mobile communication services in 1988 and terminal markets in 1994, the number of mobile phone subscribers increased rapidly. The digital system PDC (Personal Digital Communication) started its services in 1994. Four operators provide digital cellular services using frequency bands of 900 MHz and 1.5 GHz. In 2001, the third generation system IMT-2000 using 2 GHz started its services. The frequency bands being used for cellular mobile communications in Japan are shown in Figure 23.23. In Japan three operators have been assigned to use these frequency bands, which are shown in the figure with dark areas and hatched lines.

Base station antennas [1] are designed to operate in dual band and tripleband, employ electronic beam tilting and dual-beam techniques, and use diversity systems, not only space diversity schemes but also polarization and pattern diversity schemes. Small cells of radius less than 1 km have been used in highly dense areas such as Tokyo metropolitan areas. Large numbers of base station antennas were installed on civic buildings. Severe requirements, such as small installation space, low weight, and low wind load, are also allowed for in the design.

PHS (Personal Handy-phone System) services started in 1995. The cell size is designed to be very small with a radius of 100–500 m. Base station antennas are installed on various structures like telephone poles, on walls of buildings, and roofs of telephone booths. Low cost and very lightweight antennas have been used for base stations. PHS is the system that applied adaptive antennas first in the world.

In addition, PHS has been employed to operate a WLL (wireless local loop) system, with services commencing in 1997. Twelve antenna elements are used to compose an adaptive array based on the DBF (digital beam forming) technique, by which multiple

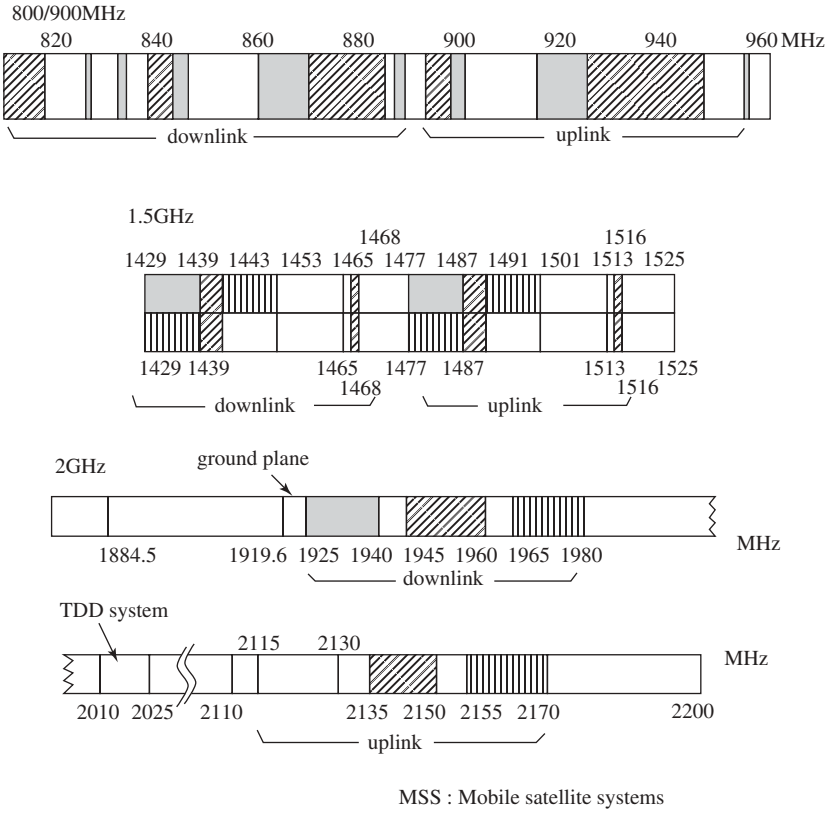


Figure 23.23 Frequency allocation for cellular systems in Japan.

beams are attained. The main beam and adaptive antenna performance contribute to increase the PHS channel capacity. In 2002, the iBurst system, which evolved from PHS, was introduced in some major cities in Australia and Africa. The system has employed adaptive arrays, which also use 12-element arrays, performing SDMA with three beams for spatial division and interference rejection. Table 23.5 shows typical mobile systems along with such parameters as data rate, coverage areas, and mobility in the present status.

23.5.4.1 Base Station Antennas for Cellular Systems The cell size has been made smaller in cellular mobile phone systems in order to increase the channel capacity in accordance with a rapid increase in the number of mobile phone users. Emphasis in designing antennas in 3G systems has been shifted from 2G systems to obtain lower sidelobes in the vertical plane and narrower sector beams in the horizontal plane in addition to beam tilting and multibeam capability. Downsizing is another significant design problem and thinning the antenna radome, in which an antenna array is housed, has made progress. In order to keep the number of antenna elements from increasing, multiband antennas, which cover both 2G and 3G frequency bands, and multibeam antennas, which illuminate both 60° and 120° sectors, have been developed.

TABLE 23.5 Typical Modern Mobile Systems

	System	Data Rate	Coverage	Mobility
2000	3G MP[a]	384 kbps∼2.4 Mbps	2 ∼10 km	High
	PHS	256 kbps	100 ∼200 m	
	WLAN (IEEE 802-11)	54 Mbps	∼100 m	Low
2006	3.5G MP	3.1∼14.4 Mbps	2 ∼10 km	High
	NG[b] PHS	20 Mbps	100 ∼200 m	
2010	iBurst (IEEE 802-20)	24.4 Mbps	∼50 km	
	WiMAX (IEEE 802.16e)	15 Mbps	∼50 km	
		75 Mbps		Low
	3.9G MP (LTE)[c]	100 Mbps	2 ∼10 km	High

[a]MP: Mobile phone
[b]NG: Next generation
[c]LTE: Long term evolution

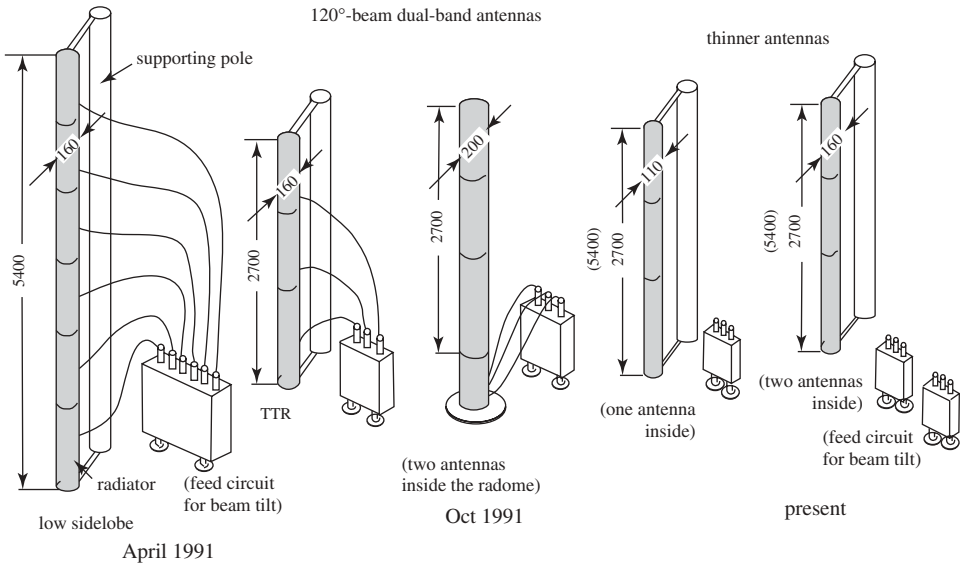


Figure 23.24 Base station antenna.

Figures 23.24 and 23.25 illustrate the history of development in base station antennas used in 2G and then 3G systems beginning in the late 1980s to the present. Figure 23.24 shows changes in antennas of dual band, having 120° sector beams. Figure 23.25 shows development of antennas for the 800-MHz band, having multibeams of 60° beamwidth.

23.5.4.2 Dual-Frequency Antennas The inside view of a practical base station antenna is shown in Figure 23.26. As dual-band array elements, the dual radiator configuration is employed. Printed dipole configurations are used as the 900- and 1500-MHz elements. The 1500-MHz element is placed in front of the 900-MHz element. Stubs are inserted between array elements in order to suppress mutual coupling due to the close spacing of the elements at 0.6 wavelength. A reflector placed behind these radiators achieves a 120° beam in the horizontal plane. The radome diameter is 100 mm.

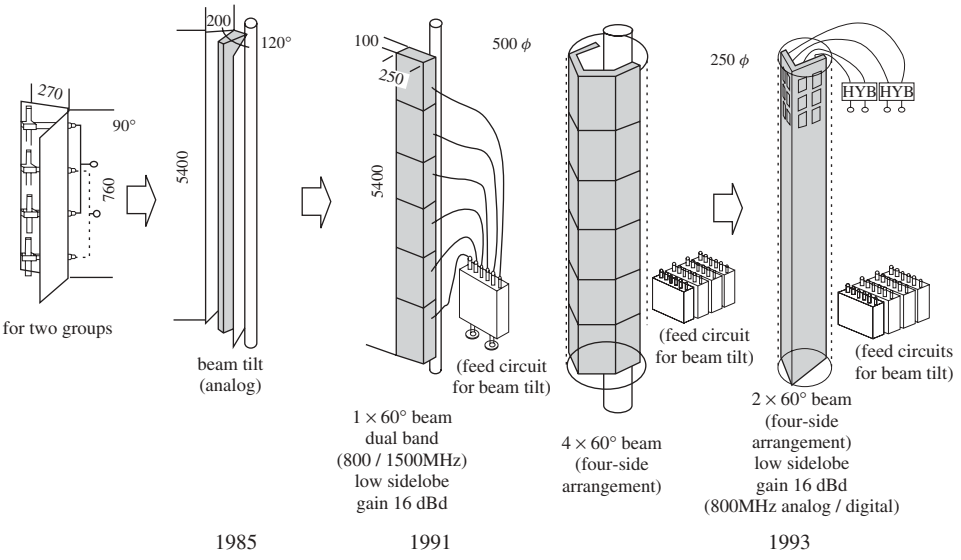


Figure 23.25 Base station antenna diversity effect versus correlation coefficient ρ_e.

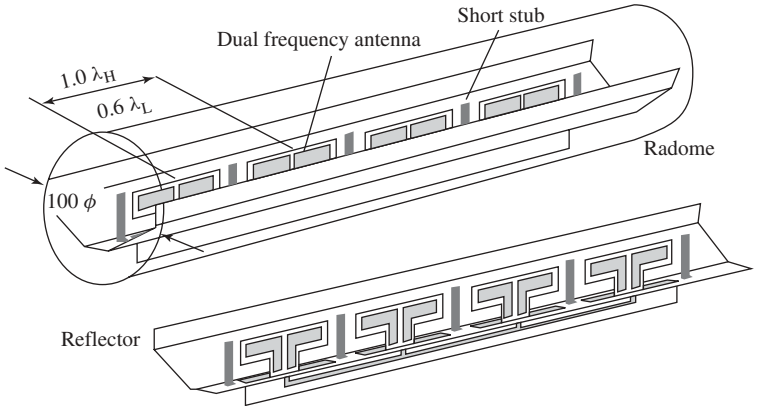


Figure 23.26 Internal view of a base station antenna.

The cross-sectional view of the antenna is shown in Figure 23.27. The feature of this configuration is the sidewalls attached to both sides of the reflector. Adding sidewalls is effective in achieving 120° beamwidths by a small reflector size. Figure 23.28 is the design chart for selecting radiator positions. By using this chart, a radiator position of 0.05 m is found for the 800-MHz band (810–960 MHz), and 0.07 m for the 1500-MHz band (1429–1501 MHz). Measured radiation patterns are shown in Figure 23.29a (800-MHz band) and Figure 23.29b (1500-MHz band). The bandwidths of the 800- and 1500-MHz bands, respectively, are 17% and 5%. At tilt angle of 6°, an antenna gain of 17.1 dBi is achieved in the 900-MHz band. Here, losses of the tilt panel of 1.2 dB and

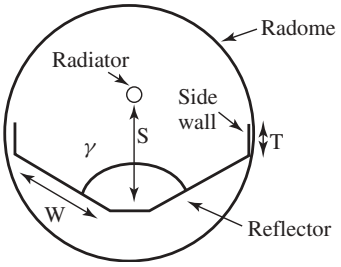


Figure 23.27 Cross-sectional view of a corner reflector antenna.

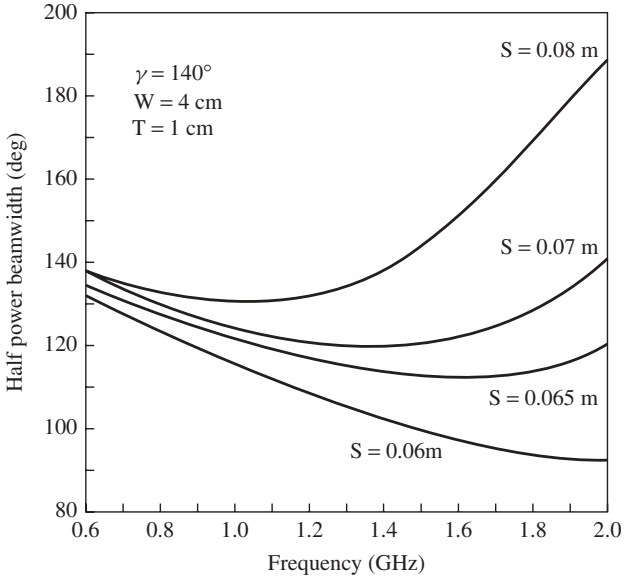


Figure 23.28 Design chart of corner reflector antenna; half-power beamwidth versus frequency with respect to the location of radiator.

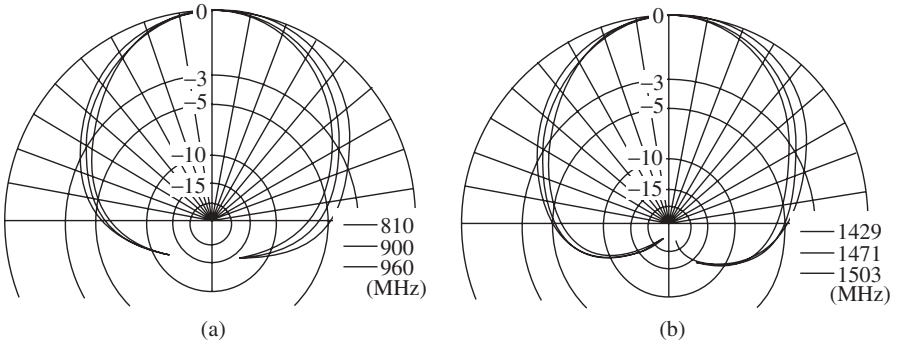


Figure 23.29 Radiation patterns in the horizontal plane: (a) 800-MHz band and (b) 1500-MHz band.

that of coaxial cables (length of 5 m) of 0.9 dB are included. At 1465 MHz, the antenna gain is 17.7 dBi, loss of the tilt panel is 1.8 dB, and loss of a coaxial cable is 1.0 dB.

23.5.4.3 Triple-Band Antenna IMT-2000 system uses the 2-GHz band and antennas additional to existing dual-band antennas or exchange of dual-band antennas have been required. Since there is little space to install additional antennas, multiband antennas have been developed: two types of dual-band antennas, which operate at 900 MHz/2 GHz and 1.5/2 GHz and one triple-band antenna [1], covering 0.9-, 1.5-, and 2-GHz bands. In IMT-2000 systems, new sector-beam antennas having 60° beamwidth are required in addition to 90° and 120° sector-beam antennas. By adding a parasitic element to dual-band antennas, triple-band antennas have been attained. Two horizontal parasitic elements are used for increasing the bandwidth at 900-MHz bands. Figure 23.30 illustrates the antenna configuration [16]. Current flows shown by arrows on the elements in Figure 23.31 express operation of three frequencies each [17]. Radiation patterns for each frequency are shown in Figure 23.32.

23.5.4.4 Dual-Beam Antenna In cellular systems, six-sectored cells are employed to increase the system capacity. Then a maximum of 18 antennas are needed when a diversity system is applied. Thus reduction in the number of antennas is desired to ease the antenna installation on the tower.

Figure 23.33 shows a 60° dual-beam antenna [18]. Two radiators are combined with a 90° hybrid circuit. By excitation of ports A and B, respectively, beams A and B are generated. By setting separation D of the radiator to be half a wavelength, the beamwidth of each beam becomes 60°, centering at $\theta_T = 30°$ directions. Two sidewalls are attached in order to refine beam shapes. The radiation pattern of this antenna is shown in Figure 23.34. The main parameter of designing a low sidelobe characteristic is length T of a sidewall. This result shows the effectiveness of the sidewall. By increasing the height, sidelobes near 70° can be reduced effectively. A parasitic element is placed in front of the radiator to achieve excellent 60° beamwidth and low sidelobe characteristics [19].

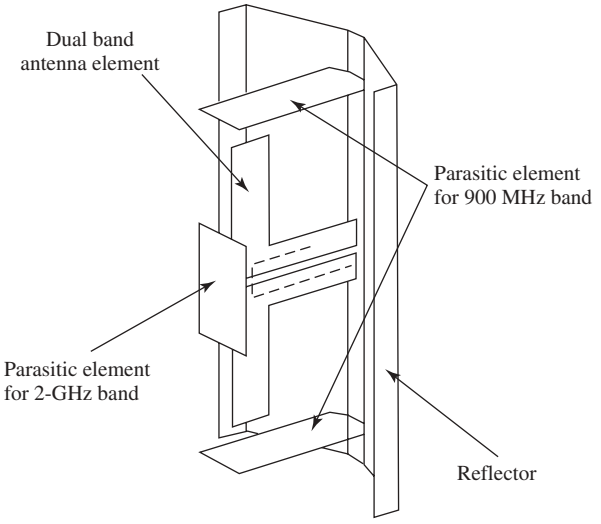


Figure 23.30 Configuration of a triple-band antenna.

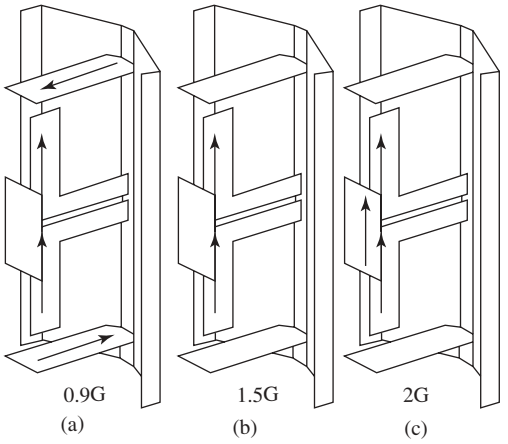


Figure 23.31 Current flow on the dipole element: (a) 900-MHz band, (b) 1.5-GHz band, and (c) 2-GHz band.

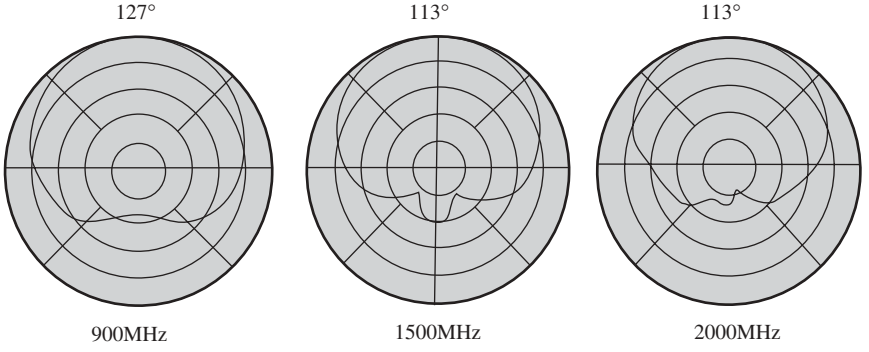


Figure 23.32 Radiation patterns at three frequencies.

23.5.5 Diversity Antenna

23.5.5.1 Space Diversity Shared use of an antenna element for reception and transmission is mandatory for reducing the number of antennas [1]. The approach to accommodate two sector elements inside a cylindrical radome is being applied to most base station antennas in order to reduce the apparent number of antennas for multi-sector cell sites. An example of this scheme is illustrated in Figure 23.35, which is a cross-sectional view of a three-sector antenna system consisting of two antennas, each of which covers a corresponding sector. In the 3G system in Japan, a six-sector zone system has been adopted in the 2-GHz band in addition to the three-sector zone system in 900- and 1500-MHz bands. Thus two types of antennas are required, one with 120°beamwidth (BW) for all three bands, and another with 60° BW for 2-GHz and 120°BW for both 900-MHz and 1.5-GHz bands. Three more 60° BW antennas at 2GHz are needed in order to cover the whole area. Two 60° BW antennas are placed side by side near the center of two 120° BW antennas for the lower frequency bands. The six-sector space diversity scheme is shown in Figure 23.36 [13].

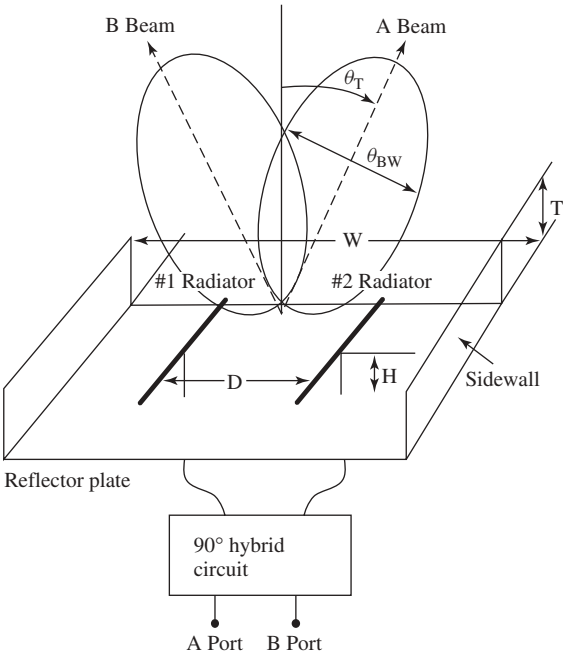


Figure 23.33 Configuration of dual-beam antenna.

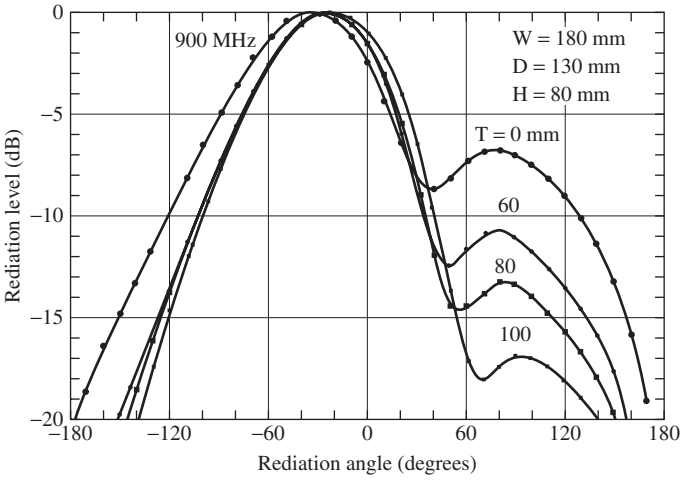


Figure 23.34 Radiation pattern with reduced sidelobes by means of sidewalls.

23.5.5.2 Polarization Diversity Polarization diversity, on the other hand, does not require two spatially separated antennas. Multiple dipole elements with orthogonal polarization can be alternately mounted on a piece of dielectric substrate in a vertical radome. Elements for the orthogonal polarization may be vertical/horizontal or $+45^\circ/-45^\circ$ crossed dipoles, depending on the particular design.

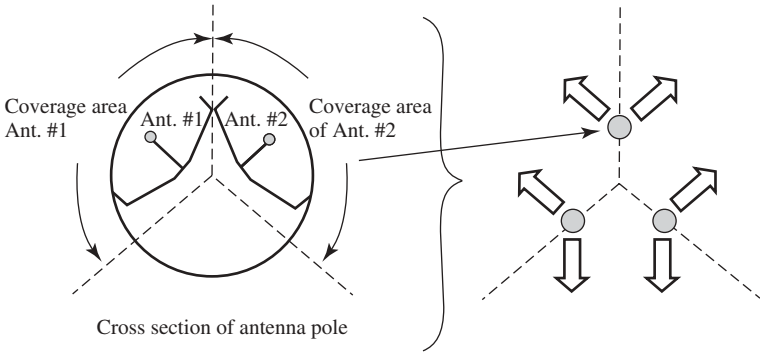


Figure 23.35 Arrangement of three-sector diversity antenna.

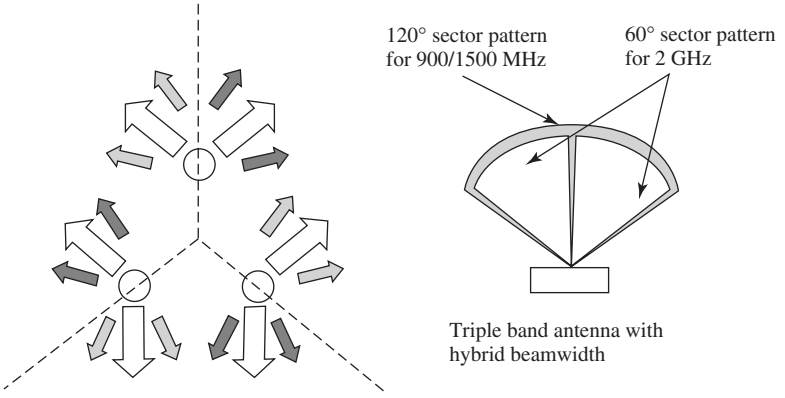


Figure 23.36 Horizontal pattern with triple-band antenna.

Polarization diversity is a well-known diversity technique; however, it has not been used in cellular phone systems. This was because it was not necessarily useful in the propagation environment where mobile terminals were mostly automobiles and vertical-polarization components dominated, as the mobile stations employed were a vertical trunk-lid element.

However, as the number of hand-held phone subscribers has increased, the propagation condition has changed. Horizontal-polarization components sometimes exceed vertical-polarization components, as users hold their phones in a tilted position, typically 60° from zenith in a talk position. The polarization diversity is then recognized as effective in the field where both vertical-and horizontal-polarization components exist.

Figure 23.37 shows a top view of a polarization diversity antenna having a compact structure, which should be compared with Figure 23.35. This is an antenna consisting of a single piece of vertical pole containing three sector transmitting and receiving arrays with diversity capability inside. Three vertical arrays A_1, A_2, and A_3 are bound together in the shape of a triangular pillar and are accommodated in a cylindrical radome. Array A_1, for example, contains a set of vertical and horizontal dipoles stacked alternately on a dielectric substrate to receive vertically polarized signal Rx-V_1 and horizontally polarized

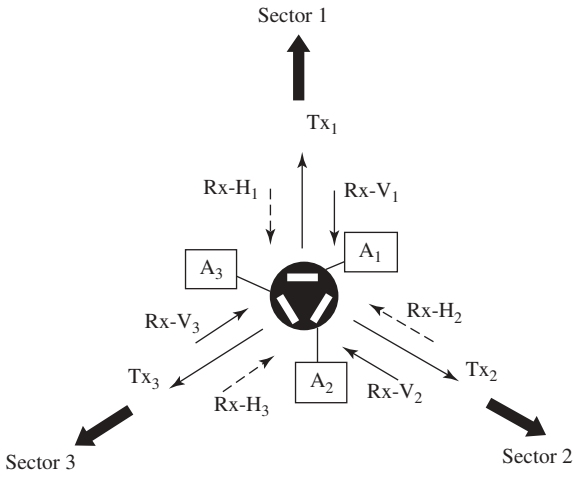


Figure 23.37 Three-sector polarization diversity system scheme.

signal Rx-H_1 for Sector 1. Either vertical or horizontal dipole elements of array A_1, or both, can be shared for transmission of the signal Tx1. The other two arrays, A_2 and A_3, have identical configuration to A_1.

By a number of tests [20–23] the effectiveness of polarization diversity was confirmed. Figure 23.38a shows the cumulative distribution of received level at a base station using space diversity, while the hand-held phone transmitted signals from a line-of-sight location in the base station coverage. The two solid lines on the left-hand side show the received levels at the two separate vertical ports and the dotted line shows the cumulative distribution of selection-combined diversity. Figure 23.38b is the case for line-of-sight location with polarization-diversity reception. The two solid lines show the distribution

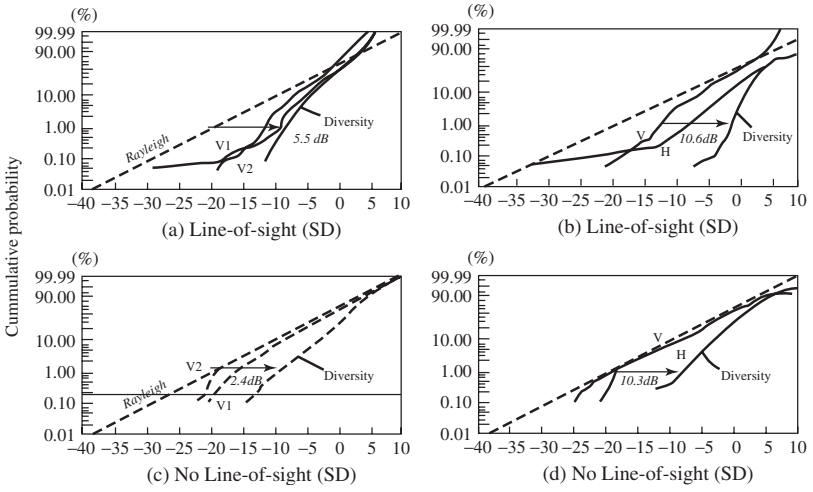


Figure 23.38 Cumulative distribution of received level at a base station: (a) and (b) for mobile phones being line of sight, and (c) and (d) for mobile phones being out of sight.

of output levels of vertical and horizontal ports and the dotted line again shows that of selection-combined diversity. As can be seen, the polarization-diversity gain is 12.0 dB at the 1% level of cumulative distribution, whereas that of space diversity is 6.2 dB, suggesting the effectiveness of polarization diversity. Similar results were obtained for the case where the mobile transmitted signals are from the out-of-sight area. The results are shown in Figure 23.38c, d. Although the effect is less significant than in the line-of-sight case, diversity gain is still greater for polarization diversity than for space diversity.

Polarization-diversity antennas feature a reduction in size and weight to cope with the stringent environmental requirements. Presently, polarization-diversity antennas have been employed in all CDMA-One base stations and some PDC base stations of KDDI of Japan [24, 25]. This has resulted in easy cell-site selection and reduction of construction cost because of the simple structure of the antennas in addition to superior performance.

Figure 23.39 shows an inner structure of this type of antenna and radome housing. Arrays for three sectors are accommodated in the shape of a triangular tube in a cylindrical radome of 23-cm diameter. In this example, three subarrays are stacked to achieve high gain. Down-tilt is performed by changing the phase angles of the top and bottom subarrays in the opposite sense relative to the central subarray. The tilt is given independently to each of the transmit/receive and vertical/horizontal polarization combinations. Each subarray consists of three vertically polarized dipoles and horizontally polarized dipoles stacked alternately [26].

Figure 23.40 shows a simplified structure of the dipoles consisting of the vertical array in vertical polarization [27]. A pair of dipoles is arrayed to make the beamwidth in the H-plane smaller, to match that of the E-plane. Similar dipoles are used for the horizontal-polarization array. The protruded portion of the parasitic element is effective in widening the frequency range over which the beamwidth in the magnetic plane is kept constant.

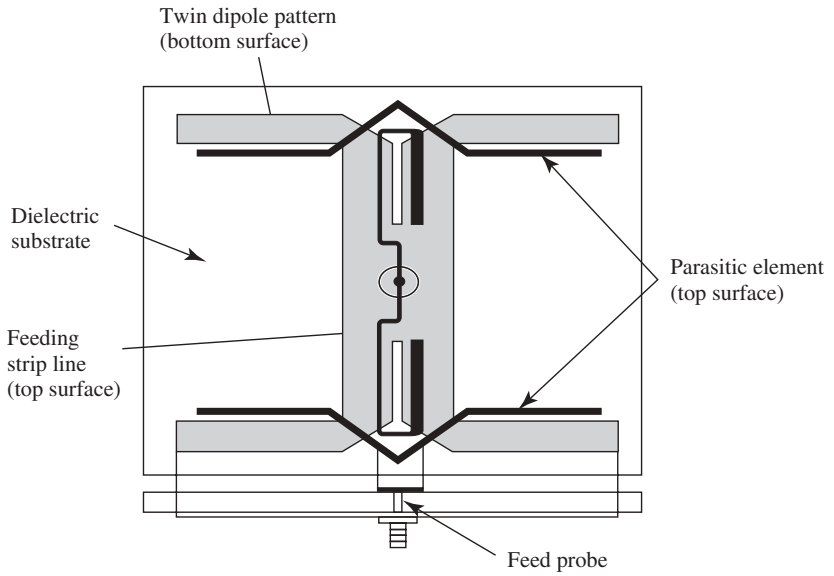


Figure 23.39 Printed twin dipole antenna.

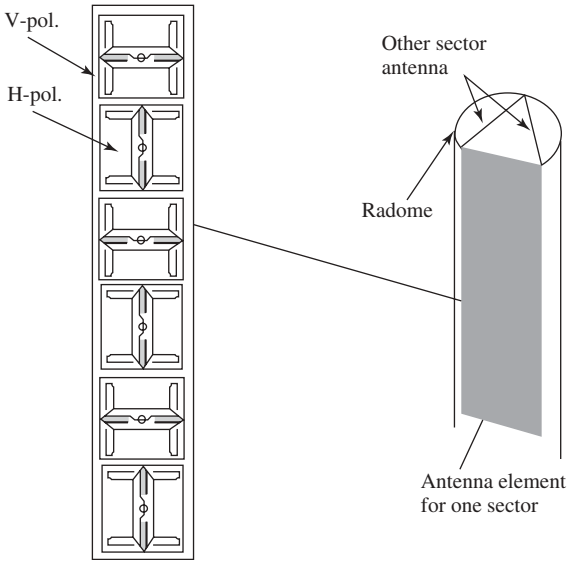


Figure 23.40 Radome housing of polarization diversity antenna.

One potential feature of this antenna is the capability of transmission diversity using different polarizations. Radiation patterns in the horizontal plane of this antenna for vertical and horizontal polarizations are shown in Figure 23.41. Great care has been taken to keep the beamwidth in both polarizations for transmit and receive bands almost identical. Cross-polarization characteristics are rather important for such antennas. Figure 23.42 shows those of the polarization-diversity antenna.

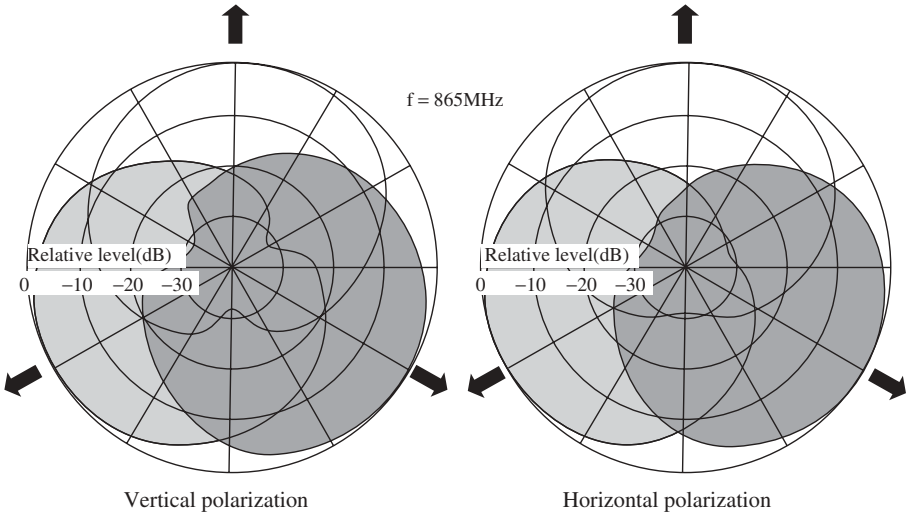


Figure 23.41 Horizontal plane patterns of three-sector antenna in V and H polarizations.

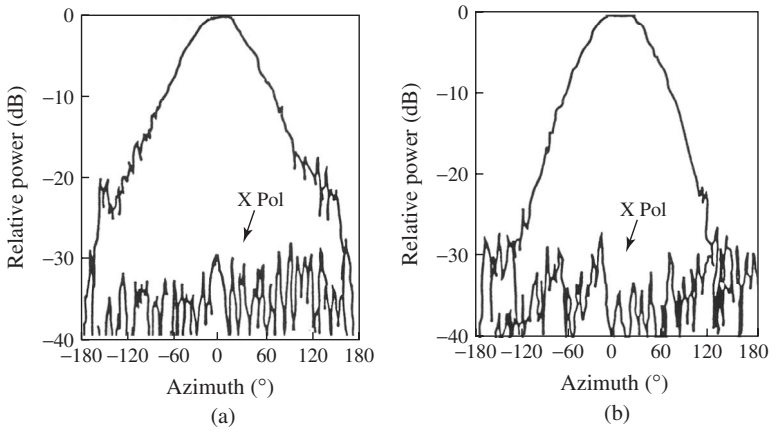


Figure 23.42 Cross-polarization characteristics of polarization-diversity antenna.

23.5.6 Antennas for Micro/Pico Cellular Systems

Cellular phone service [13, p. 984] has been extended to inside tunnels and subway stations, inside large buildings, in underground shopping malls, and so forth. In order to provide services for these areas, the relay station is installed between the outside base station and the areas to be covered. The relay station has a booster, which receives a downlink signal from the outside station, amplifies it, and reradiates it. As for the uplink signal, the operation is reversed in the same way. A flat antenna is used for the booster. For example, for tunnel systems, a low profile, two-element half-wave dipole antenna is used. The antenna configuration is shown in Figure 23.43. The antenna is installed on the wall of a tunnel with very low height. The radiation pattern is a figure eight in both E-and H-planes, having nulls in the direction normal to the antenna element.

In the tunnel system, an optical fiber cable is used to connect the outside station with the relay station, since the optical fiber has very low transmission loss and wide

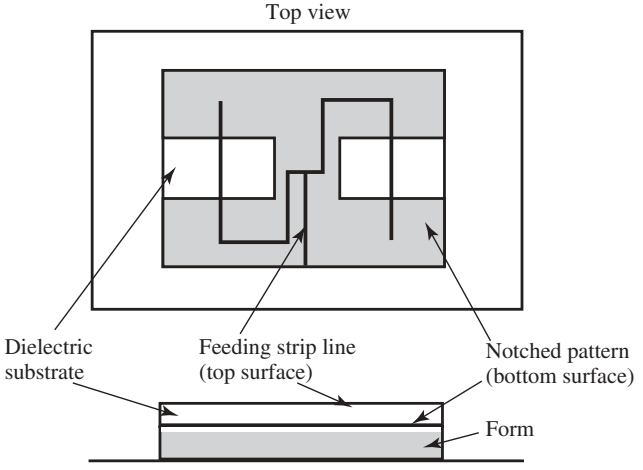


Figure 23.43 Planar notched structure two-element half-wave dipole antenna.

bandwidth. The optical signal is modulated with both 900-MHz and 1.5-GHz signals directly and conveyed by the optical fiber. In the downlink, from the outside station to the relay station, the received signal is reradiated by a low gain antenna to the relay station. The antennas are flat types and installed on the sidewall of a building or some other construction. The antennas are arranged one by one with separation of about a hundred meters. The reverse system is used for the uplink.

The subway link uses a similar system, by which mobile phone service can be made available in the subway stations and even on the trains.

23.5.7 Antennas for Personal Handy-phone System (PHS)

In the PHS system, a TDD (time division duplex) transmission scheme is employed. This system provides service at the 1900-MHz band. A block diagram of the transmission systems with transmission diversity is shown in Figure 23.44 [28]. The base station (BS) has two antennas and receivers. The portable station (PS) has only one antenna and receiver. In a TDD system, a single carrier frequency is used to provide two-way communication (upward channel—PS to BS; and downward channel—BS to PS). The BS is able to predict the received-signal strength at the PS because of the reciprocity between upward and downward channels. The BS receives an upward link signal from the PS using the diversity reception method and measures the received signal strength during a receiving period. In addition, it predicts which antenna gives the highest received signal strength at the PS. Then the selected antenna is used for transmission.

Diversity antenna setups are shown in Figure 23.45. The antenna installed orthogonal to a road has superior diversity gain. For antenna separation(s) larger than 50 cm, sufficient diversity gains are achieved.

The typical omnidirectional collinear antenna configuration is shown in Figure 23.46 [29]. The inner microstrip patch antennas are excited. The cylindrical parasitic elements are used as omnidirectional radiators. The element separation is 0.7 wavelength. Five radiation elements are employed for the upper and lower antennas, respectively. Antennas are covered with a radome whose diameter is 17 mm. Antenna beamwidth in the vertical plane is 16°. Antenna gain is 7.5 dBi.

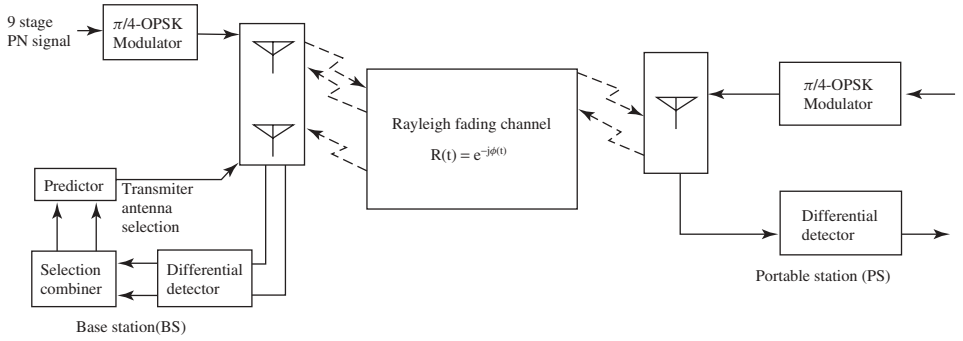


Figure 23.44 Block diagram of linear predictive transmitting diversity antenna system.

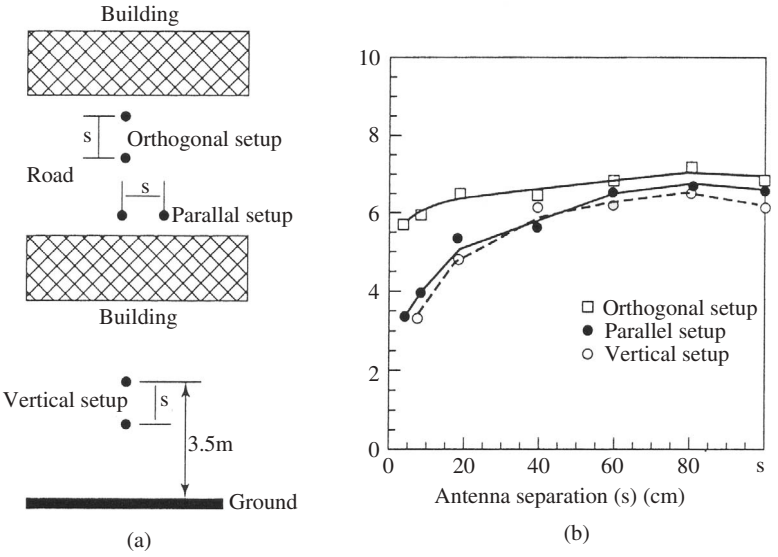


Figure 23.45 Diversity antenna: (a) antenna installation and (b) diversity gain.

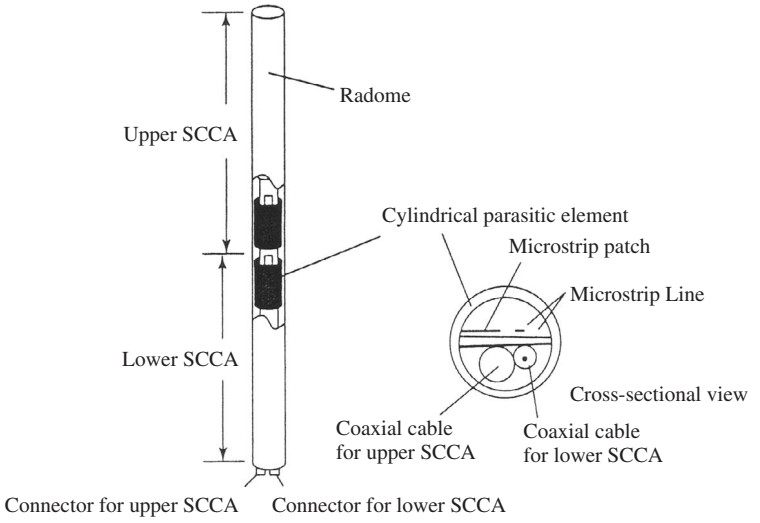


Figure 23.46 Configuration of typical omnidirectional collinear antenna.

23.5.8 Adaptive Array Antenna System

23.5.8.1 AAAS in PHS Among the endeavors to reduce the interference and improve the spectrum efficiency, as transmitting data and the number of users are increasing, study of the adaptive array antenna [13, p. 989–990] has been one of the most significant subjects to which antenna engineers devoted attention.

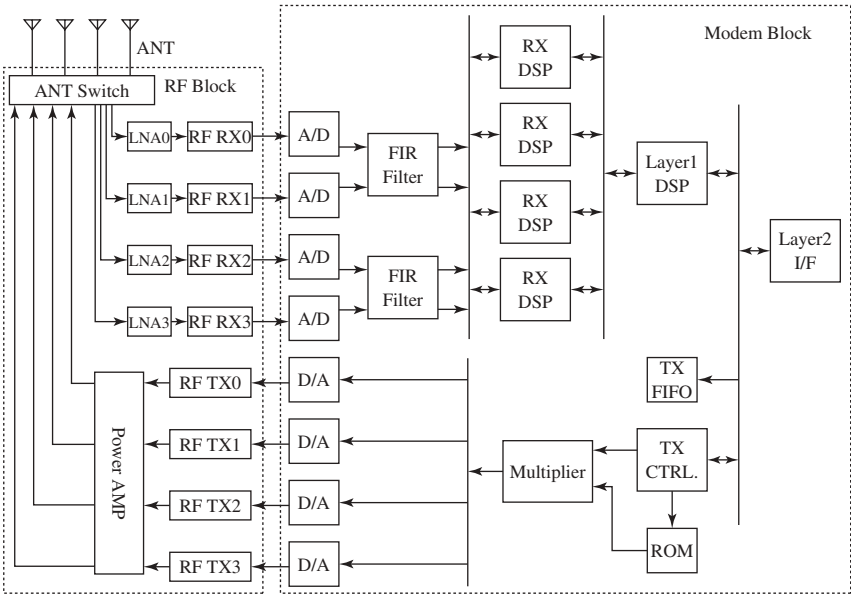


Figure 23.47 Block diagram of adaptive antenna system for base station.

The adaptive array system was first introduced in 1998. The block diagram of an adaptive base station antenna applied to PHS is shown in Figure 23.47 [30]. Four antennas are used for transmission and reception. Adequate weight is determined in the CPU module for the output signal of each antenna to achieve the best bit error rate (BER) value by using a constant modulus algorithm (CMA) concept. Antenna radiation pattern maximum is achieved in the desired signal direction, and pattern nulls are achieved in the undesired signal directions. Each antenna has gain of 10 dBi. Antenna spacing is 5 wavelengths. A radio unit is installed at the foot of the antenna.

23.5.8.2 AAAS in Wireless Local Loop (WLL) Systems An adaptive array antenna system (AAAS) applied to WLL systems [30] is called Super PHS–WLL, because the PHS standards are applied to the WLL, and wideband radios and spatial channel processing are employed to enhance system performance. By means of adaptive beam forming, the channel capacity is increased, the number of multipath signals is reduced, the coverage is expanded, and the flexible configuration of the coverage for each base station to match the local propagation environment is made feasible.

The system concept is briefly illustrated in Figure 23.48. The figure shows an example of a system configuration, which is composed of a BS (base station), single-and four-line SU (subscriber unit), and connection to the local exchange. The operating frequency of the system is the same as that of the PHS, 1880–1930 MHz, and the TDMA–SDMA system for the channel access and the duplex system, TDD (time division duplex), are employed.

The AAAS, which produces multiple beams by digital beam forming (DBF) technology, performs the function of SDMA. The beams are automatically directed to multiple SUs, wherever they are located. The received signal at the BS in the multipath environment is processed adaptively to enhance the processing gain, thus improving the quality

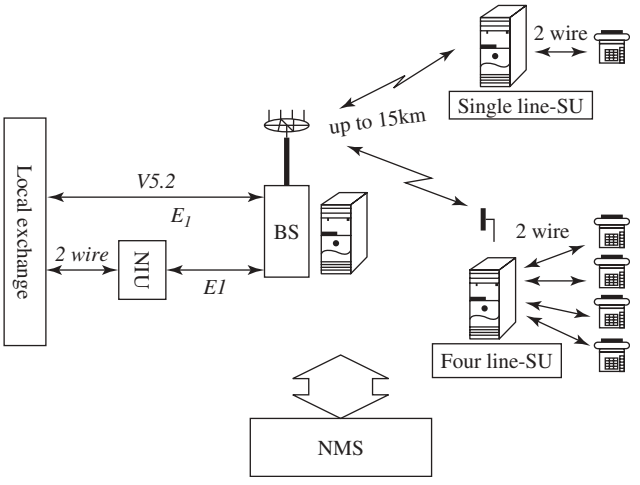


Figure 23.48 System concept of PHS WLL.

of the link. It also contributes to interference rejection by directing a null against it, and enhancing the reuse of the same frequency, thereby increasing the number of channels.

System performance that can dynamically adapt to changeable propagation environments is achieved by DBF technology. The DBF software processes, in real time, the phase and time differences among the incoming signals in the multipath environment so that the signal transmission can be directed to each of the desired SUs. By this means, a stable and robust link can be established.

Other factors that contribute to enhancing the channel capacity are the multiple-frequency operation by using multiple wideband radios in addition to the AAAS, and use of the PHS standard TDD, which has four time slots in one data frame that can be used for both transmitting and receiving on four channels simultaneously on one frequency.

A base station of the super PHS–WLL system, which is in operation, is composed of 12 antennas and 12 radios and uses 16 radiofrequencies for the 16-way multiplex operation. By combining the spatial channel processing with the AAAS, a spatial channel efficiency of about 2.5 is achieved. Consequently, the voice-channel capacity achieved is up to 155; that is, 4(time slot) $\times$ 16(frequency) $\times$ 2.5 (spatial channel efficiency) -5 (control channel). It can serve 2730 subscribers as the total traffic capacity and is capable of covering a range up to about 15 km.

23.5.8.3 AAAS in the iBurst Systems The iBurst system is one of the latest broadband wireless access systems, which features efficient frequency utilization achieved by means of an adaptive modulation scheme and application of an AAAS. The iBurst system was first introduced in Australia and began commercial service in 2000. It is presently deployed in some countries in southern Africa and Asia, while many other countries such as the United States, Canada, and Japan are closely investigating introduction of the system. Kyocera Corporation is playing a major role in the development and commercialization of the system. The system has been standardized by the ANSI as one of the high capacity (HC) SDMA systems and also will be taken up by the IEEE 802.20 Committee to be included in the standard as a wideband mobile broadband system (WMBS).

The iBurst system technology is based on PHS (Personal Handy-phone System), which employs a TDD/TDMA system. The TDD system is advantageous for realizing the AAAS, as the channel estimation, which is necessary for adaptive control of the antenna pattern, can easily be made, since the channel uses the same frequency for both transmitting and receiving. In addition, the system employs very advanced technologies such as an adaptive modulation scheme and advanced SDMA system. The block diagram of the system is shown in Figure 29.49. Figure 23.50 shows a base station installed on the top of a building, where a 12-element array and base station units, including T/R circuits, are shown. The small dimensions of the base station can be observed by comparing the man standing beside the base station units. Details of the antenna element are illustrated in Figure 23.51. The maximum total transmission rate is 32 Mbps, which is achieved by dividing a 5-MHz band into 8 carriers with 625-kHz separation, to which the assigned data rate is 1061 kbps for the downlink and 346 kbps for the uplink, respectively, and by applying 3-spatial multiplexing. The 32 Mbps is numerically evaluated as $[(1061 + 346) \times 8(\text{carriers}) \times 3(\text{SDMA}) - 1061(\text{control channel}) = 32.707$.

The AAAS adopts the MMSE algorithm, by which a high signal-to-interference + noise ratio (SINR) can be achieved. The basic functions of the AAAS are to form beam maximums toward desired signals, while directing nulls to the interference signals, and to perform spatial multiplexing, SDMA. A base station can serve 21 mobile terminals simultaneously, and mobile terminals can operate individually even when they are located as close as a few centimeters to each other.

The SDMA performance was verified by an experiment in which three terminals were located close to each other. Figure 23.52 shows the concept of the experiment. The average SINR at a mobile terminal was evaluated with and without the AAAS. The result is shown in Figure 23.53, which illustrates SINR versus separation of mobile terminals. It can be observed by the figure that SINR suddenly degrades without the AAA scheme as the separation of mobile terminals increases, whereas it does not change significantly with the AAA scheme.

The throughputs for both downlink and uplink were evaluated with the same model. Figures 23.54 and 23.55 illustrate throughputs obtained for the downlink and the uplink, respectively. The results imply that the SDMA system works effectively so that each terminal, although located very closely to others, has nearly the same transmission rate

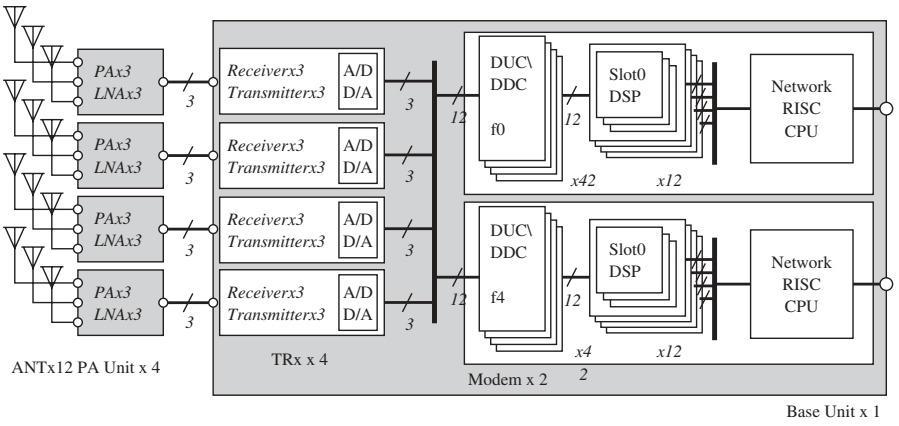


Figure 23.49 Block diagram of iBurst base station unit.

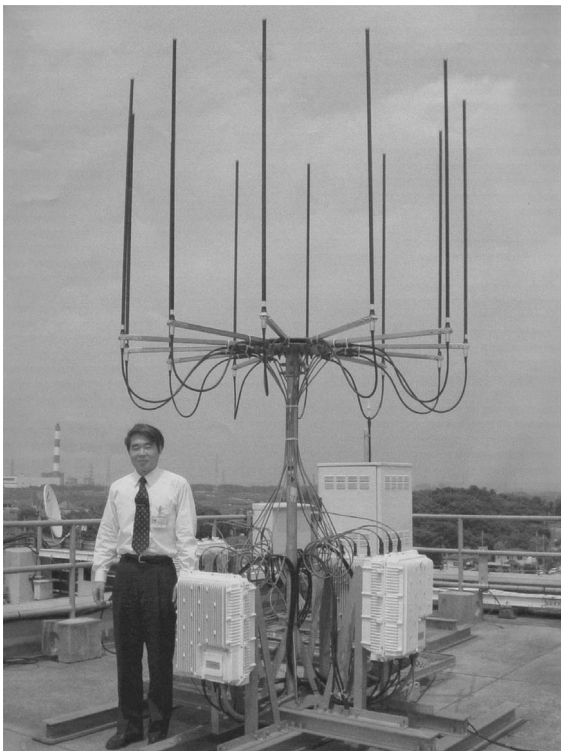


Figure 23.50 Practical antenna and base station units.

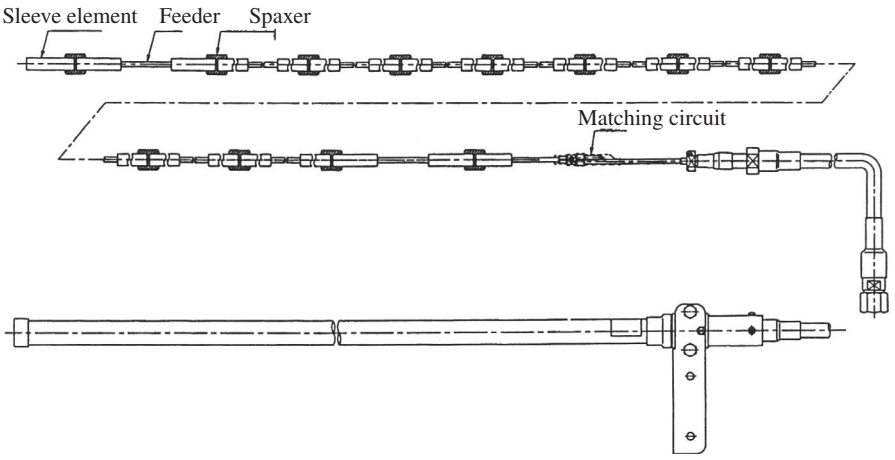


Figure 23.51 Antenna configuration.

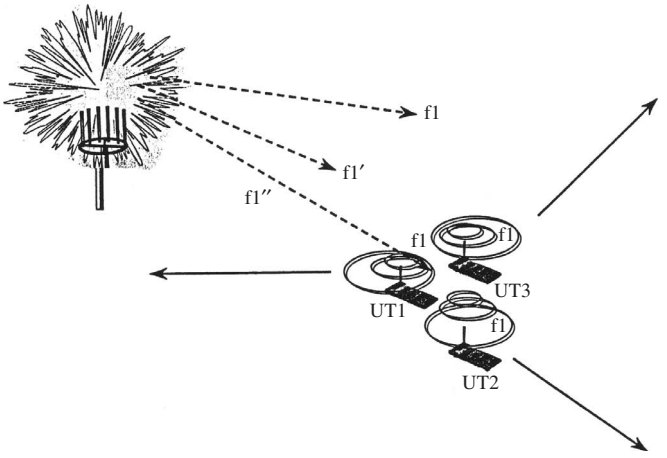


Figure 23.52 SDMA performance test with three mobile terminals.

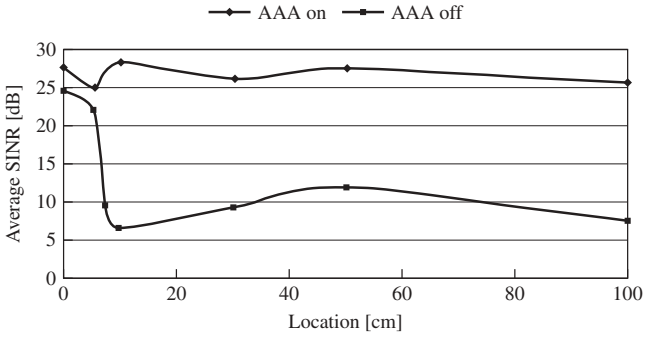


Figure 23.53 SINR versus mobile terminal location.

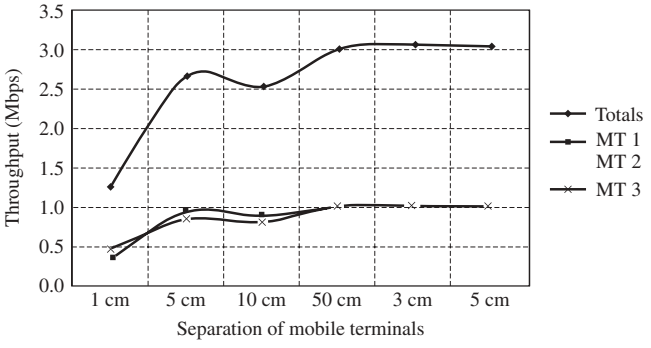


Figure 23.54 Throughput versus separation of mobile terminals (downlink).

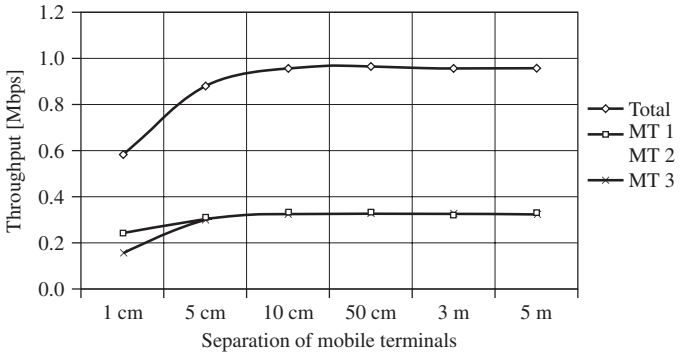


Figure 23.55 Throughput versus separation of mobile terminals (uplink).

as normal, even when the separation was as close as 1 cm, although the data rate was lower when compared with that for the separation of 5 cm or more.

The AAA performance was evaluated in the field, including a multipath environment. Two mobile terminals were used in the experiment and the cumulative distribution function (CDF) was evaluated with respect to SINR and throughput at a base station for a case where one mobile terminal acted as an interferer. The experiments were performed in both urban and suburban areas, and the speed of the mobile terminal was varied between 20 and 40 km/h in each case. Figure 23.56a provides results obtained in urban and suburban areas, respectively, with mobile speeds taken as the parameters. The superlative operation of using the SDMA system can be observed by these results. The CDF values evaluated with respect to SINR taken in suburban and urban areas, respectively, are shown in Figure 23.56b, c. The CDF versus throughput is shown in suburban (Figure 23.56d) and urban (Figure 23.56e) areas. In Figure 23.56, the dotted lines are for the case where the AAAS was not used. Again, the advantages of applying the AAAS can be observed.

23.6 ANTENNAS FOR SMALL MOBILE TERMINALS

23.6.1 Design of Small Mobile Terminal Antennas

In designing antennas for small mobile terminals [1], typically handsets, factors that should be taken into consideration are as follows:

- Small size
- Light weight
- Compact structure
- Low profile or flush mount
- Robustness
- Flexibility
- Low cost

For specific applications, there may be other important requirements for the design. Parameters pertaining to not only antennas but also communication systems, propagation

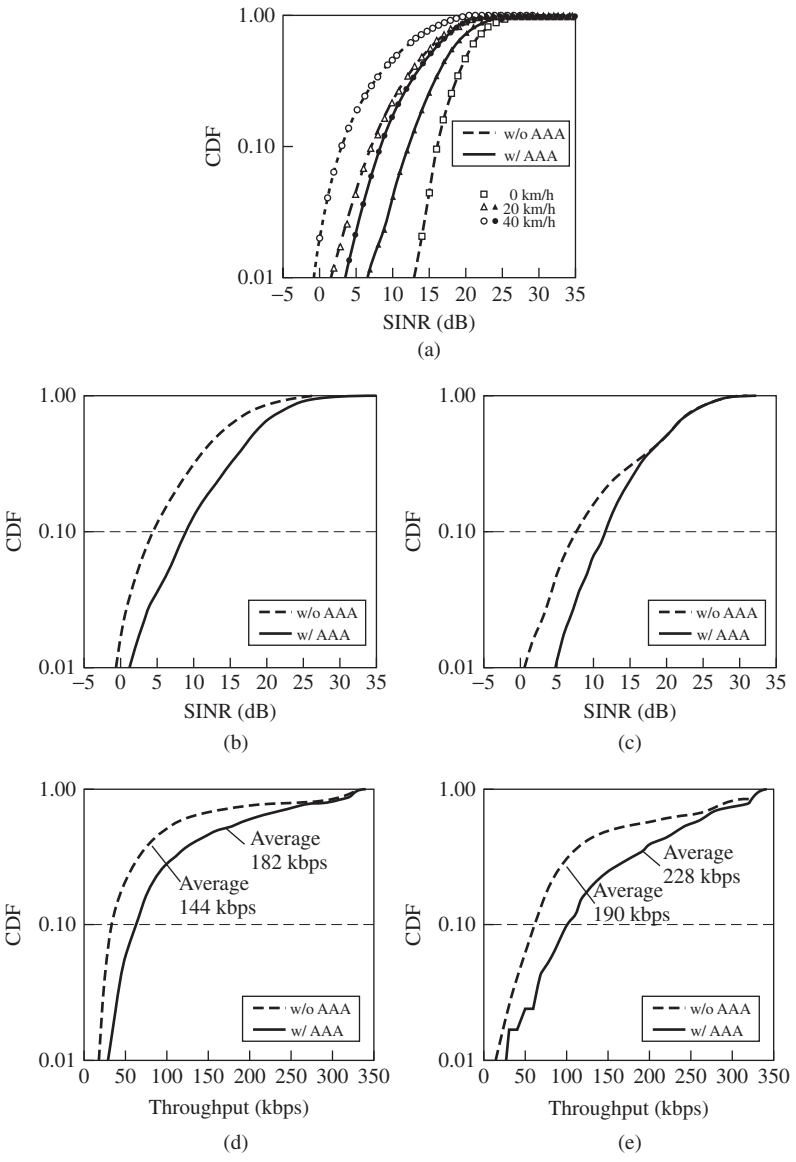


Figure 23.56 (a) CDF variation versus SINR for several speeds; (b) CDF versus SINR in suburban areas; (c) CDF versus SINR in urban areas; (d) CDF versus throughput in suburban areas; and (e) CDF versus throughput in urban areas.

problems, and environmental conditions are summarized in Figure 23.57. In fact, all of these parameters, more or less, should be taken into consideration when handset antennas are designed. When systems require particular specifications, antenna design should be made specifically to satisfy the requirements. Examples of such requirements are multiband operation and built-in structure. In GSM systems, various types of built-in multiband antennas have been used in small mobile terminals. In addition, after introduction of 3G

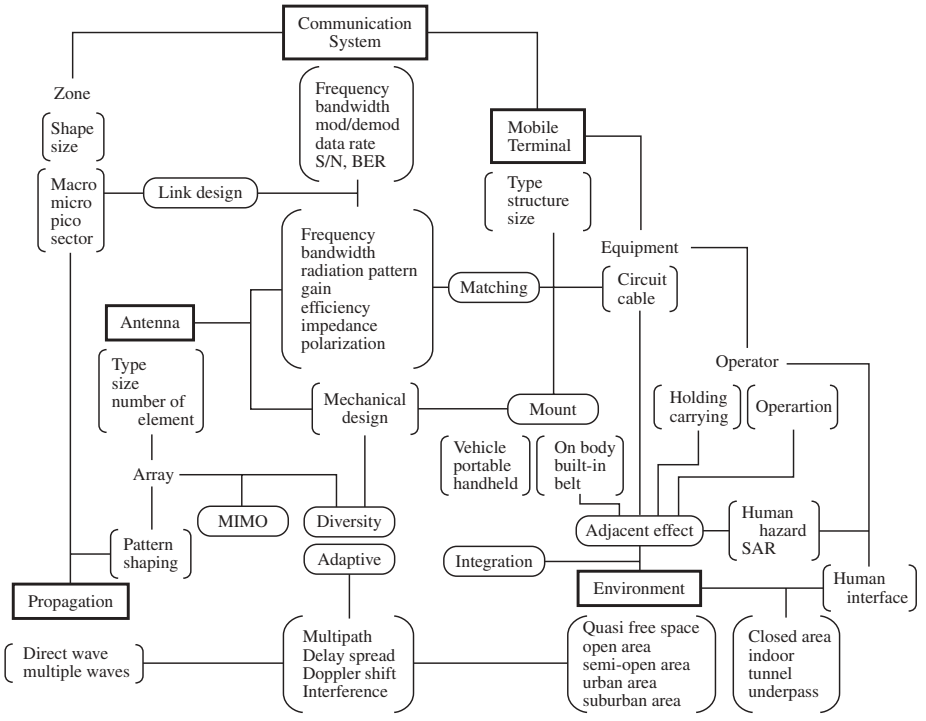


Figure 23.57 Design parameters necessary for mobile terminal antennas.

systems, there have been areas where both 2G and 3G systems operating with different frequency bands are in service and hence multiband antennas have been installed in mobile terminals operating in these areas. In addition, there have been various small mobile terminals to which nonmobile phone systems such as Bluetooth, RFID, NFC (near-field communication), and mobile WiMAX are added and function not only for the purpose of communication but also for control, data transmission, and identification. Accordingly, antennas that operate at frequencies corresponding to such systems should also be mounted. Propagation problems are presently diversified to include complicated environments such as inside buildings and subway stations and in underpasses. In addition, the problem now is to include higher frequency regions such as microwaves and millimeter waves.

23.6.2 Design Concept of Antenna Systems for Small Mobile Terminals

The design concept of antennas for mobile terminals [31] changed as mobile systems made progress and the analysis of antenna systems advanced.

In the early days of mobile communications, antennas used in portable equipment were simply monopoles of about a quarter-wavelength. Since the unit body was made of metal, and hence was assumed to be a conductive ground plane (GPL), the antenna system was considered to be equivalent to a half-wave dipole by taking the image of the quarter-wavelength monopole into consideration. Some time later, it was noticed that this was a misconception and the unit body should be treated as a part of the antenna element,

because the unit body could no longer be assumed to be a GPL, because currents, which contribute to radiation, flow on the surface of the unit body. Design of antennas for portable equipment based on this concept was first introduced in 1968 [32]. About ten years later, analysis of the current distributions on the unit body (rectangular conducting body) was made, and design parameters using the length and dimensions of the unit body were shown [33]. By this analysis, the optimum length of a monopole mounted on a rectangular conducting body was discussed; a box having a length of a half-wavelength, either a $\frac{5}{8}\lambda$ or $\frac{3}{8}\lambda$ monopole, is considered to be optimum in terms of input impedance and reduced current flows on the unit body. These reduced current flows on the unit body are meaningful, because the degradation of antenna performance due to proximity effects is reduced; for example, influence of the human head and hand, when the unit is in the talk position, should be small. This analysis established the design concept of antennas for mobile phones presently used worldwide [31]. Although current flows on the unit body may alter antenna performance and degrade it, those flows can be utilized in a positive way as a part of the antenna element to enhance antenna performance, since the?currents can contribute to radiation with the unit body acting as a radiator. Metal materials used for the unit body have gradually been replaced by plastic materials as mobile terminals have become smaller and lighter. Ground surfaces of the printed circuit board (PCB), on the rear side of the PCB, have effectively been taken as a GPL, which substitutes for the unit body.

In order to reduce the current flow on the GPL and thus reduce degradation of the antenna performance, a type of antenna having a balanced structure and fed by a balanced line was introduced [34]. An example is a folded dipole, in which current flow on the GPL can be reduced markedly. Reduction of the proximity effects, typically the effects of the human hand and head, has been verified by reduced distortions in radiation patterns for a balanced-antenna case.

As the number of subscribers increased, communication zones became smaller in order to increase the number of available channels; for example, a zone is less than 1 km in diameter, and zones are sectored into three or six. Accordingly, antenna design has changed to allow some degradation of the antenna performance, if system performance is not seriously deteriorated. Thus a GPL is purposely included in the antenna design as a part of the radiator, permitting some degradation of antenna performance due to the effect of the human hand and head.

The design concept of mobile terminal antennas has changed to develop small, built-in, and multiband antennas, because there are usually several systems in service in the same area. A single mobile phone terminal should correspond to them; for example, 2G systems in the 800-MHz and 1.5-GHz bands, and 1.8-GHz and 2-GHz bands for 3G and 3.5G systems. If a mobile phone terminal installs an additional system in the mobile phone like Bluetooth so that the mobile phone can be used for both systems, an antenna in addition to the mobile phone antenna should be mounted.

With wireless systems other than mobile phones, such as WiMAX, prevailing in numerous areas and providing service for broadband transmission, new antenna systems have been developed for these situations, by which high data-rate transmission (over 50 Mbps) can be supported. MIMO (multiple-input multiple-output) systems are used for such broadband systems; broadband data can be either transmitted or received without enlarging spectral bandwidth. In this system, data are divided into several channels, put

into several separate antennas through several channel transmitters, and then transmitted. The transmitted data are received by several corresponding antennas and are processed to combine the divided data and to recover the original data at the receiver.

Use of an internal antenna, replacing the monopole that has long been used in mobile terminals, is a trend in today's mobile terminals. Possible antenna characteristics, desired for the latest mobile terminals, are:

a) Having a magnetic current as a radiating source
b) Being a balanced type
c) Having two or more modes (e.g., two polarization components)

It should be noted that design of antennas for small mobile terminals essentially is based on the technologies common to the design and/or development of small antennas. The purpose of using magnetic current is to create a built-in small and low profile or flush-mounted antenna. A magnetic current source, for example, a loop, slot, or parallel plate, can produce a magnetic field vertical to the GPL, when the source is placed parallel to the GPL. With this structure, the GPL image of the source current will create an additional field in front of the GPL so that the field strength will be doubled. Consequently, antenna gain is doubled, compared with an antenna system in free space, yet the antenna is small and has a low profile. The PIFA was invented based on this concept. By means of a balanced mode, the currents on the GPL, which are generated by the antenna element, can be reduced, and the operator's effect on the antenna performance can be mitigated. In practice, a dipole, a loop, a normal mode helical dipole, a meander line dipole, and any other type of antenna having a balanced structure can be used for reducing current flows on the GPL and avoiding the body effect. Using two modes in an antenna system is the significant concept to create a small antenna. It has been shown that having two modes, such as a dipole and a loop, can contribute to enhancing antenna performance so that the antenna size is made smaller. An example is a pager antenna [35].

Now by combining any one of characteristics (a), (b), and (c) with the conventional design concept, an advanced antenna system can be developed. Introducing balanced modes into an antenna indicates that the GPL is not purposely utilized for enhancing antenna performance, but rather the GPL is separated from the antenna system.

The design concept is illustrated in Figure 23.58: stage I, use of a quarter-wavelength monopole and its image; stage II, combined design of a monopole and the equipment case as a radiator; the stage III, utilization of GPL to enhance performance of a built-in small antenna, and stage IV, separation of GPL from the antenna system.

The next stage will be design of an antenna such that the device, function, or software is integrated into an antenna system, so that sophisticated functions such as adaptive control and signal processing are achieved. Applying the concept of the integrated antenna system (IAS) is a promising way to realize such antennas. Unification of a device or RF circuitry into an antenna structure will bring about a new antenna having enhanced performance or improved characteristics. Integration of function or software into the antenna structure may also create a new antenna system that performs with intelligence.

Antenna systems now being developed should realize a system in which antenna, wireless transmission, signal processing, and network technology are unified to constitute an advanced sophisticated system.

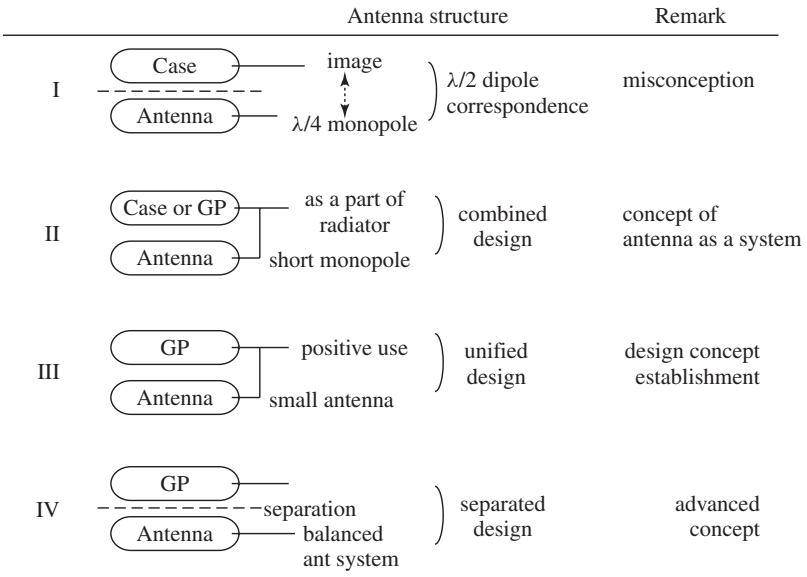


Figure 23.58 Process of design concept of mobile terminal antenna.

23.6.3 Typical Antenna Elements Used for Mobile Terminals

Typical antenna elements used for small mobile terminals are MP (monopole), NMHA (normal mode helical antenna), PIFA (planar inverted-F antenna), ceramic chip, meander line, and MSA (microstrip antenna). A majority of mobile terminals such as PDC handsets employ an MP element and a built-in PIFA as a pair of elements for a diversity antenna (Figure 23.59). The NMHA element is also used in small mobile terminals, where the NMHA is placed on the top of an MP element and operates when the MP is retracted. The NMHA element is used as the main antenna of PHS mobile terminals, since the operating frequency is 1.5 GHz, allowing the antenna length to be short enough to be fixed on the unit body. Present trends are to employ a built-in antenna. The most popular built-in antenna is a type of PIFA, but with various modifications from the original structure. Many GSM handsets have used such modified types of built-in, multiband PIFAs. Another antenna element has a meander line structure, which is useful in smaller-sized antennas. The smallest antenna presently used is a type of ceramic-chip antenna, having a size of several cubic millimeters; however, it is necessary to use some appropriate size of ground plane, on which the chip antenna is placed.

23.6.3.1 Monopole and Dipole A monopole is the most simple, thin, lightweight, low cost antenna element useful for small mobile terminals. It has long been used as the typical antenna element for small mobile terminals. The dipole antenna also has features of being simple, thin, lightweight, and low cost; however, it wasn't used until 3G systems, because the dipole length is too long to mount on the mobile terminals for frequency bands below 1 GHz.

Antenna performance depends not only on the antenna element but also on dimensions of the ground plane on which the antenna element is mounted. Figure 23.60 illustrates the gain of a MP with respect to the length of the unit, which is assumed to be a rectangular

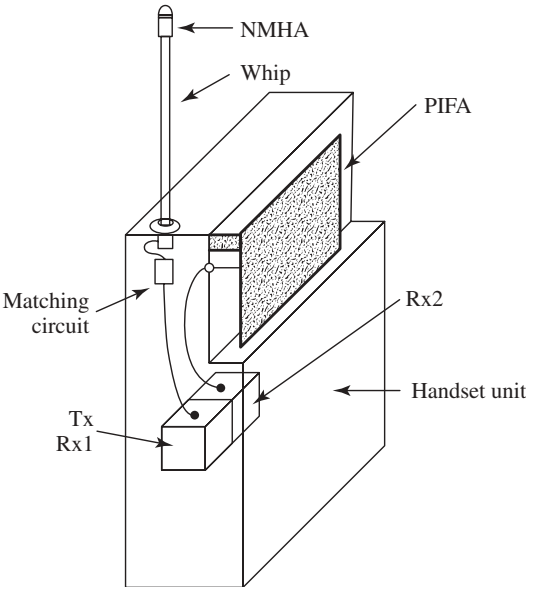


Figure 23.59 Antenna system in a PDC mobile terminal.

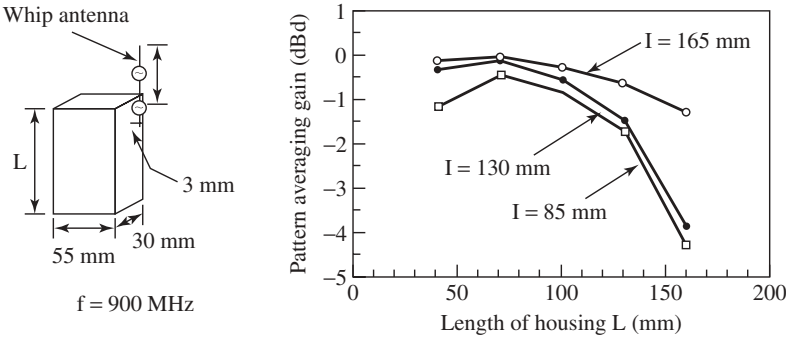


Figure 23.60 Gain of a monopole with respect to the length.

conducting box, simulating a handset unit [35]. The dimensions of the box are shown in the figure. Radiation currents flow heavily on the handset unit as well as the antenna element when the length of the unit is about $\frac{1}{4}\lambda$; however, the handset-unit current is reduced remarkably when the monopole length is about $\frac{1}{2}\lambda$ [1, p. 349]. Nevertheless, that is still not advantageous, because the input impedance is too high to match the load impedance. When the element length is either $\frac{3}{8}\lambda$ or $\frac{5}{8}\lambda$, handset current flows are relatively small and yet the input impedance characteristics are adequate to match the load impedance [36]. This concept has been applied to the design of PDC handset antennas.

23.6.3.2 Normal Mode Helical Antenna (NMHA) In order to make the antenna small, a traveling-wave structure can be utilized and the NMHA is a typical example.

The NMHA is essentially a helical antenna having a thin structure and thus radiating in the directions normal to the helical axis. The radiation patterns are the same as those of a short dipole (Figure 23.61a). The dimensions of the helix are usually much smaller than the operating wavelength, the length of an individual turn being a small fraction of a wavelength λ and the axial length being also much less than a $\frac{1}{4}\lambda$. With this structure, a NMHA is equivalently expressed by an array of small loops and short monopoles (Figure 23.61b) [37]. Inductive impedance is increased by the array of loops and compensates the capacitive impedance of short monopoles. As a consequence, the helical structure can be arranged to have a self-resonance property, although the antenna length (in the axial direction) is considerably shorter than that of a conventional resonant monopole or dipole antenna. Thus the antenna efficiency can be improved and extension of the antenna length may lead to increased bandwidth. The NMHA has been used in mobile phones worldwide.

Design parameters for a NMHA model depicted in Figure 23.62 [38] are provided by Figure 23.63. The relationship between the helix pitch p and the diameter D_h are given Figure 23.63, where the parameter used in Figure 23.63a is the number of turns MT, while that in Figure 23.63b is the radiation resistance R at resonance.

23.6.3.3 Meander Line Antenna (MLA) The meander line antenna also has a traveling-wave structure, by which the antenna length can be reduced. It has a periodical array structure of alternating square patterns as shown in Figure 23.64 [39]. With this pattern, the extended wire length can be made much longer than the initial antenna (dipole) length so that self-resonance can be attained. The resonance frequency is then lower and radiation resistance is higher than that of a dipole with the same length. This in turn implies that the antenna is effectively made small.

Impedance characteristics of a MLA being placed on the ground plane, as shown in Figure 23.64, are shown in Figure 23.65, where Figure 23.65a depicts input resistance R_1

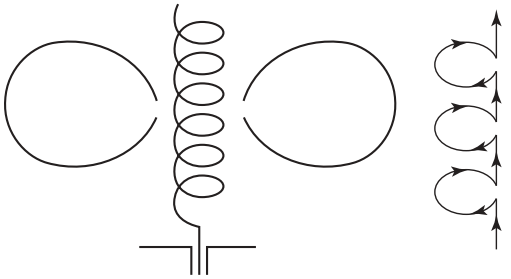


Figure 23.61 Normal mode helical antenna (NMHA) and its equivalent expression.

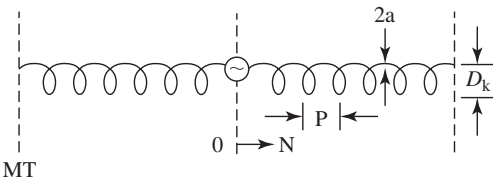


Figure 23.62 ?NMHA dipole model.

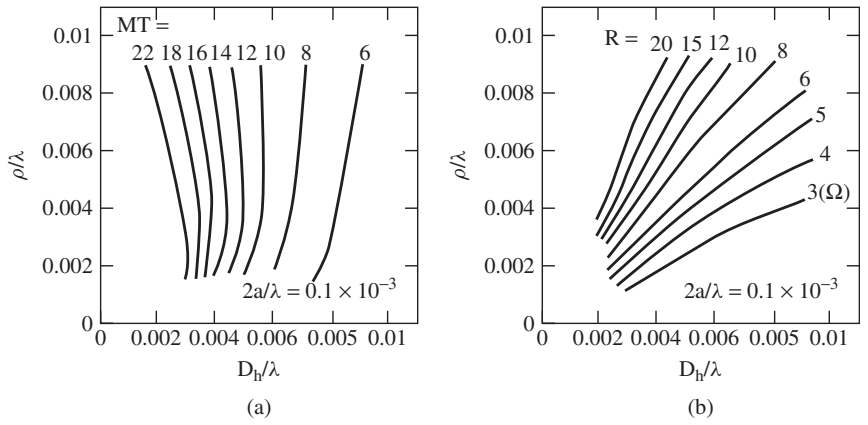


Figure 23.63 Design chart (P-versus D_b) of NMHA: (a) with parameter being the number of turns MT and (b) with parameter being radiation resistance R at resonance.

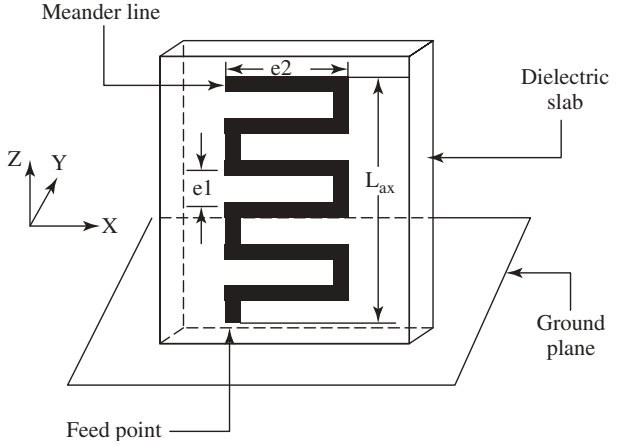


Figure 23.64 Meander line antenna placed on the finite ground plane.

and Figure 23.65b depicts input reactance X_1, with respect to the frequency, by taking the length L_{ax} as the parameter. Figure 23.65c depicts the first resonance frequency f_0 (in GHz) versus antenna length L_{ax}. The empirical expression is derived as

$$f_0 = 22.69L_{ax} - 0.7392 \tag{23.4}$$

when e1 and e2 are 3 mm, the substrate thickness t_s and width W_s are 3.17 mm and 11 mm, respectively, and the size of the ground plane is 59 × 29 mm.

A very small MLA having dimensions comparable to 0.1λ has been developed [39, 40]. The antenna structure is shown in Figure 23.66; radiation resistance R_{in} along with loss resistance R_l and radiation efficiency η with respect to the antenna length L are shown in Figure 23.67. In the figure, N stands for the number of repetitions of the rectangular wire patterns. Figure 23.68 illustrates a folded structure of MLA

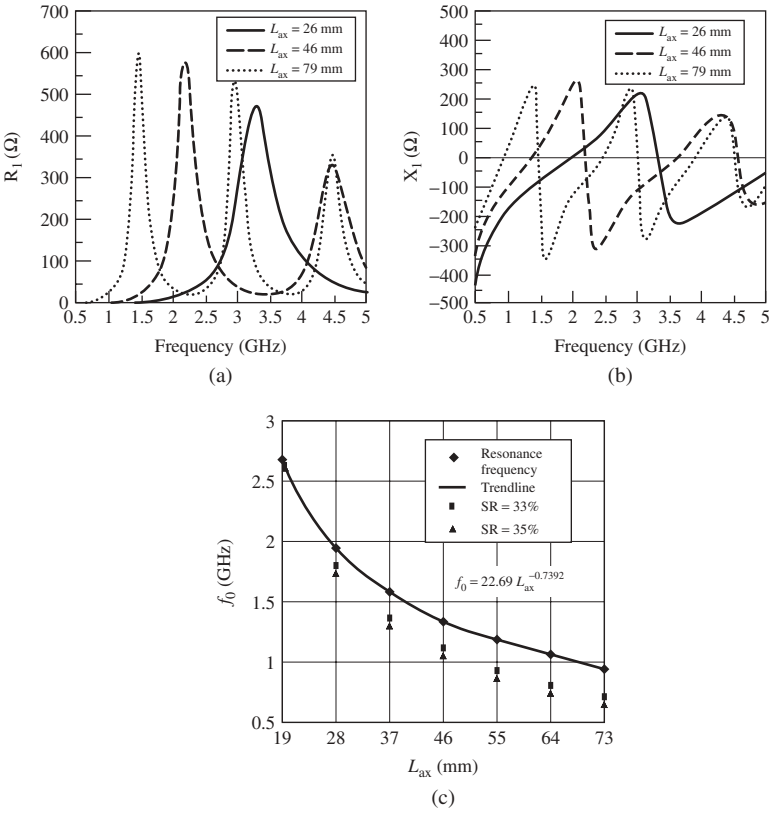


Figure 23.65 Characteristics of MLA shown in Figure 23.64: (a) input resistance, (b) input reactance, and (c) the first resonance frequency.

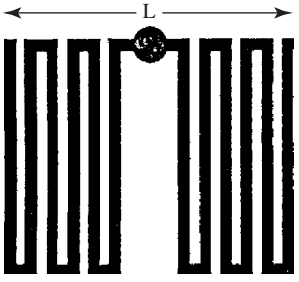


Figure 23.66 A very small MLA model.

shown in Figure 23.66 in order to obtain higher resonance resistance and thus higher radiation efficiency. Figure 23.69 gives radiation resistance R_{in} and loss resistance R_l and radiation efficiency with respect to the antenna length L. Figure 23.70 shows the relationships among the design parameters; antenna length L, width W, and number of turns N, for both original and folded types of small MLAs.

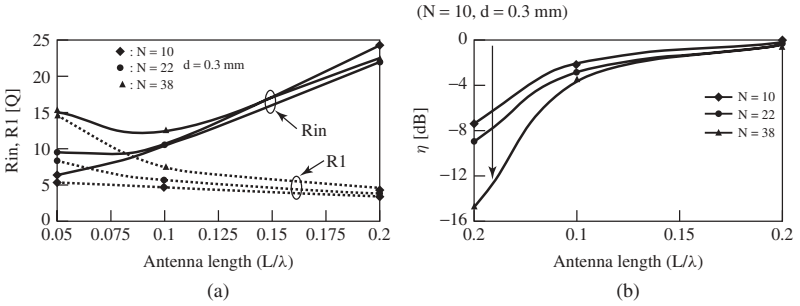


Figure 23.67 Characteristics of very small MLA: (a) input and loss resistances with respect to antenna length and (b) radiation efficiency with respect to antenna length.

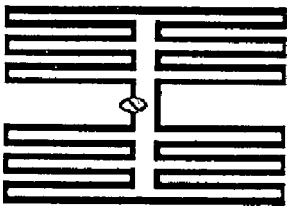


Figure 23.68 Folded type of very small MLA model.

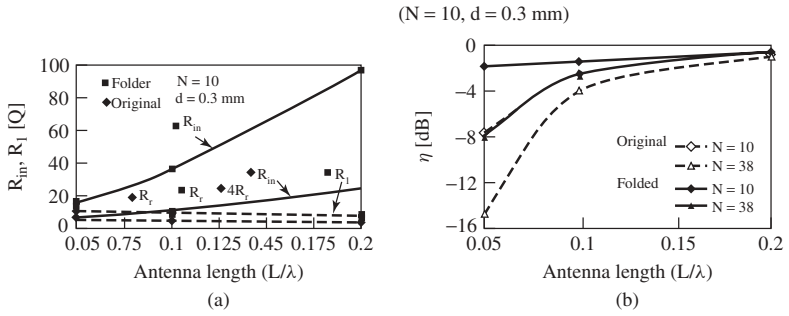


Figure 23.69 Characteristics of folded type MLA: (a) input resistance and loss resistance and (b) radiation efficiency.

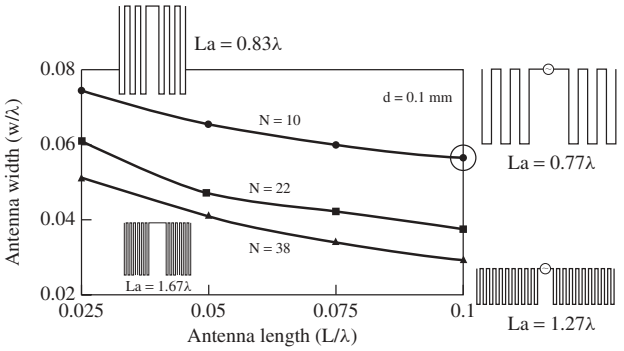


Figure 23.70 Design parameters (antenna length, width, and number of pattern repetition) for the resonance.

23.6.3.4 Inverted-F Antenna (IFA) The IFA can be treated equivalently as an antenna composed of two parts: an ILA with its image, and 2 two-wire shorted transmission lines. The two-wire transmission lines compensate capacitive impedance of the ILA and at the same time step up the input impedance so that matching to the 50-ohm load is made feasible. The input impedance of an IFA is expressed by

$$Z_{\text{in}} = 4Z_{\text{a}} \cap 2Z_b \tag{23.5}$$

where Z_a denotes the input impedance of the ILA with its image, Z_b is the impedance of the shorted two-wire transmission line, and $\cap$ expresses parallel combination of these two impedances [41].

23.6.3.5 Planar Inverted-F Antenna (PIFA) The PIFA (Figure 23.71) seems to be an antenna modified from an IFA by replacing the linear horizontal element of the IFA with a planar element. That is misleading, however. The radiation principle of the PIFA differs from that of the IFA. The source of radiation of a PIFA is the magnetic current on the peripheral aperture of the planar element, whereas that of the IFA is the electric currents on the linear F-shaped vertical and horizontal elements. In addition, the length of the peripheral aperture of the PIFA is about a half-wavelength, while that of the IFA is about a quarter-wavelength.

Study of the PIFA structure initially starts with a half-wave slot placed on the side of a rectangular conducting body as shown in Figure 23.72a. The slot is fed at a point near a shorted end in order to attain adequate impedance to match the load impedance. The idea was born in the process of developing handset antennas. The basic concept was to develop a small, low profile antenna, suitable for a handset antenna. In order to have a low profile or flush-mounted structure, use of a magnetic current as a source of radiation

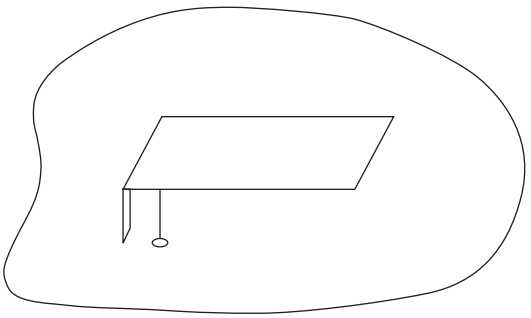


Figure 23.71 PIFA model.

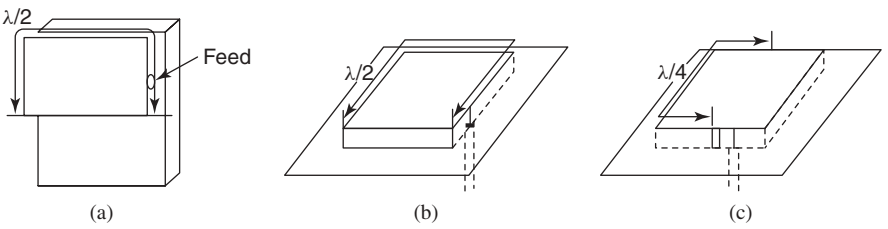


Figure 23.72 Process of PIFA development.

was considered. The slot antenna satisfies this requirement. Meanwhile, the microstrip antenna (MSA) had been known as a useful planar antenna, having a small and low profile structure. Then an attempt was made to place that slot antenna on a ground plane to form a planar structure like an MSA (Figure 23.72b) and the performance was confirmed to be satisfactory for handset use. Finally, the antenna configuration was arranged to have a structure as illustrated in Figure 23.72c, in which the shorted part is reduced to make the antenna size smaller. In a PIFA, its peripheral aperture of a half-wavelength contributes to radiation, whereas in the MSA a patch-end aperture of about a quarter-wavelength contributes to the radiation.

In fact, an attempt was made initially to increase the bandwidth of a linear IFA by replacing the linear horizontal element with a planar element; however, it was not successful, as the bandwidth increase obtained was only 1–2%, unless the height was otherwise increased.

An example of the input impedance of a similar antenna model is shown in Figure 23.73a [42]. The equivalent circuit of this antenna model is shown in Figure 23.73b. It is interesting to note that in the equivalent circuit there is a reactance jX. The bandwidth enhancement of an antenna system, being composed of a PIFA element placed on the handset unit, may be attributed to this reactance. This reactance has constant impedance over a much wider bandwidth than that of the antenna system.

In practice, there are many variations in a PIFA to be found. In fact, some can hardly be identified as a PIFA, because the modified antennas have entirely different appearances from the original structure.

23.6.3.6 Small Chip Antenna The small chip antenna (Figure 23.74) is a small bulk antenna, which is composed of an antenna element encapsulated by either ceramic material or a composite of ceramic and resin material having a high relative permeability. The antenna element is usually printed on the material substrate and encapsulated by material with relative permeability of 5 or higher. As a typical antenna element, a meander line, inverted-F, normal mode helix, or other planar pattern is used (Figure 23.75). Use of a planar ceramic, on which the antenna element is printed, is typical. However, there is another type: a small but bulky substrate can be structured with very thin multilayered ceramic sheets, in which the antenna element is encapsulated. The antenna elements

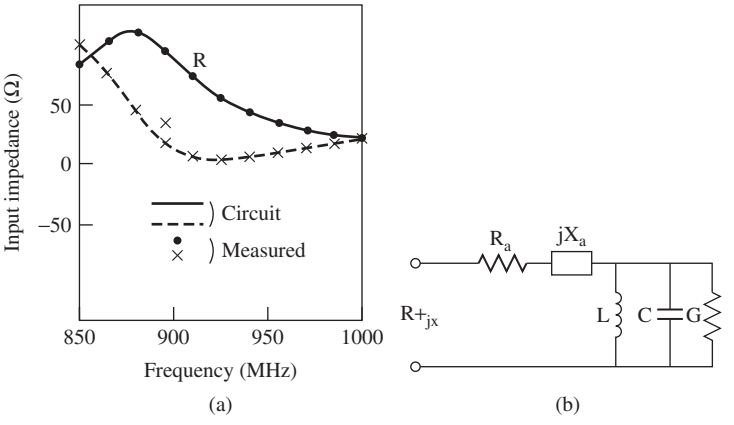


Figure 23.73 Input impedance of an IFA and the equivalent circuit.

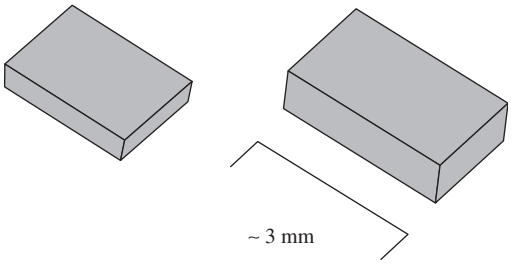


Figure 23.74 Small ceramic chip antennas. (Courtesy of Murata Mfg. Co., Japan.)

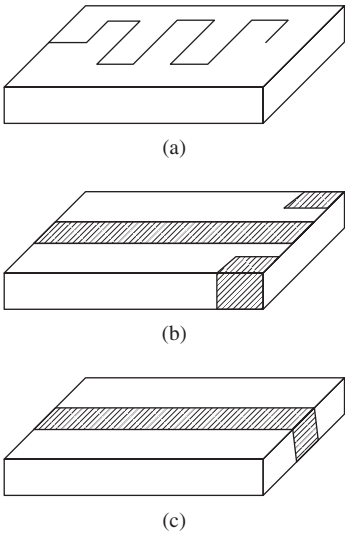


Figure 23.75 Some examples of printed patterns on the substrate surface of a chip antenna. (Courtesy of Murata Mfg. Co., Japan.)

printed on the top and bottom surfaces of the ceramic substrate are connected via thin conducting wires.

One of the smallest ceramic chip antennas is 1.8 mm wide, 3.2 mm long, and 1.3 mm high, as depicted in Figure 23.76. This antenna is one of the world's smallest antennas. Figure 23.77 is an exploded view of the antenna structure. The thin ceramic substrate is composed of multilayered thin ceramic sheets. The antenna element used is a normal mode helix, which is formed in the substrate by connecting printed patterns on the top layer and bottom layer via thin conducting wires. This antenna has been used successfully for small mobile terminals, in which the antenna is placed on a corner of the ground plane (GPL) as shown in Figure 23.78. A small part of the ground plane beneath the antenna element is sometimes taken out in order to avoid excessive loss due to the GPL. The antenna element actually couples with the GPL so that the GPL acts as a radiator. In other words, an antenna element and the GPL are combined to constitute an antenna system. Thus the antenna performance depends on the size and the shape of the GPL and also on the location of the antenna element on the GPL. This type of antenna features a

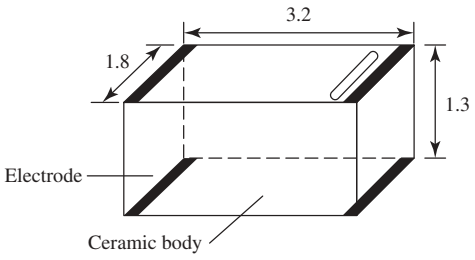


Figure 23.76 A small chip antenna model. (Courtesy of Murata Mfg. Co., Japan.)

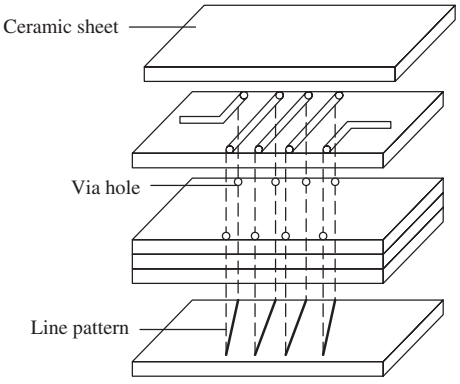


Figure 23.77 Exploded view of a small chip antenna. (Courtesy of Murata Mfg. Co., Japan.)

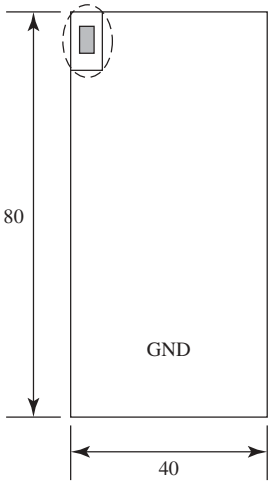
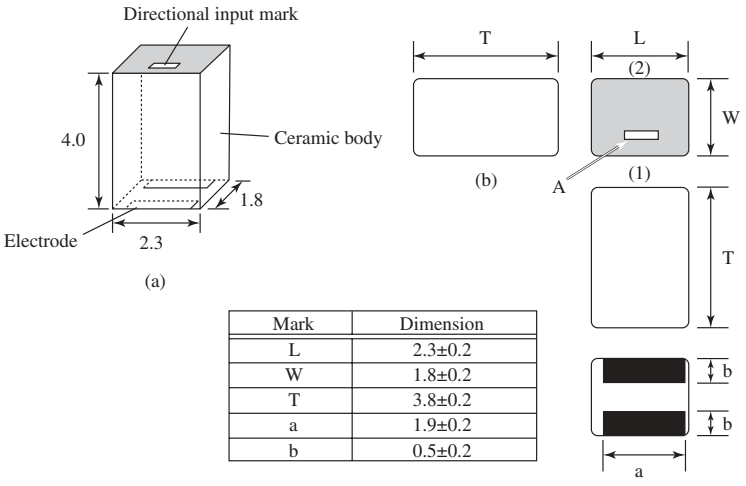


Figure 23.78 Typical location of a chip antenna on the ground plane. (Courtesy of Murata Mfg. Co., Japan.)



Mark	Dimension
L	2.3±0.2
W	1.8±0.2
T	3.8±0.2
a	1.9±0.2
b	0.5±0.2

Mark	Meaning
A	Directional Input Mark

Terminal No.	Terminal Name
(1)	Feeding Point
(2)	NC

Figure 23.79 Small chip antenna: (a) antenna model and (b) dimensions and printed patterns on the substrate surfaces. (Courtesy of Murata Mfg. Co., Japan.)

resonant frequency that can be controlled by its easy-to-adjust structure, reasonably high gain, and very low weight.

Another model having different dimensions, 2.3 × mm 1.8 mm × 4.0 mm, is shown in Figure 23.79a, which depicts the three-dimensional structure. Figure 23.79b shows the printed antenna pattern, with dark color on the bulk surfaces, and the dimensions. This antenna is designed to have a little higher structure so that the base size is made smaller and the space to mount the antenna becomes narrower than for the previous model. Figure 23.80 shows an example of how to place the antenna on the ground plane. The antenna characteristics such as VSWR and the impedance locus on a Smith chart are shown in Figure 23.81. Table 23.6 gives NO ground plane area (a part of the ground plane area is taken out beneath the antenna element), frequency at the lowest VSWR, bandwidth, and gain. Radiation patterns are shown in Figure 23.82. Figure 23.83 illustrates some examples of antenna mounting: on a USB dongle, a cellular phone, a PDA, and a laptop computer.

Here, characteristics of two other types of practical models are introduced. One antenna is a type of ceramic chip: Figure 23.84a shows the printed antenna pattern on the surface of the ceramic substrate and Figure 23.84b shows the antenna placed on the ground plane. Figure 23.85 shows the same type of antenna, but a part of the ground plane beneath the antenna element has been removed. This type of antenna features flexible tuning by means of the feed line and a small circuit component. Antenna placement on the ground plane is shown in Figure 23.86, in which two ways of tuning are illustrated: one by feed line for rough tuning and another fine (precise) tuning with either a capacitive component or an inductive component. Variations of frequency with variation of feed line dimensions,

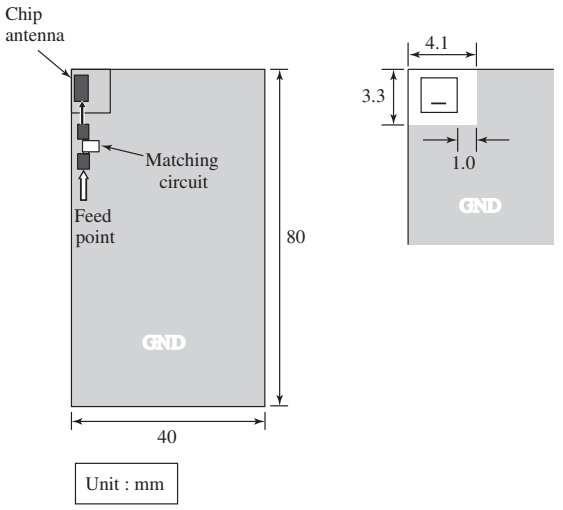


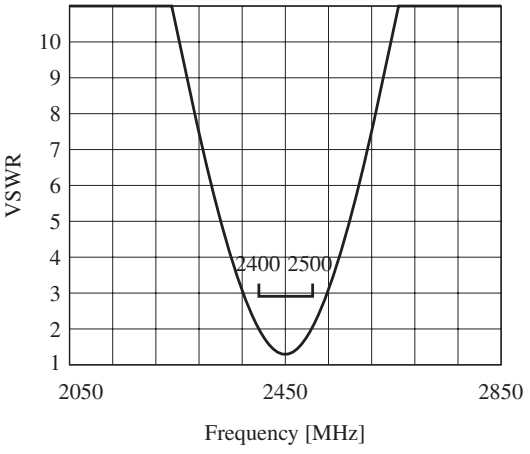
Figure 23.80 Chip antenna on the ground plane: (a) antenna, feed, and matching circuit and (b) location. (Courtesy of Murata Mfg. Co., Japan.)

a capacitance, and an inductance are also shown in the figure. Table 23.6 provides gain values at three frequencies. Figure 23.87 shows radiation patterns. Variations of bandwidth and efficiency, depending on the antenna position on the ground plane, are given in Figure 23.88. In addition to this feature, antenna performance can be varied by the location of the antenna element on the GPL; for example, placing it not only on the corner of the GPL with different directions, but also in other places like near the center. Flexibility in selecting the antenna location suggests that standardized custom-made chip antennas may become available.

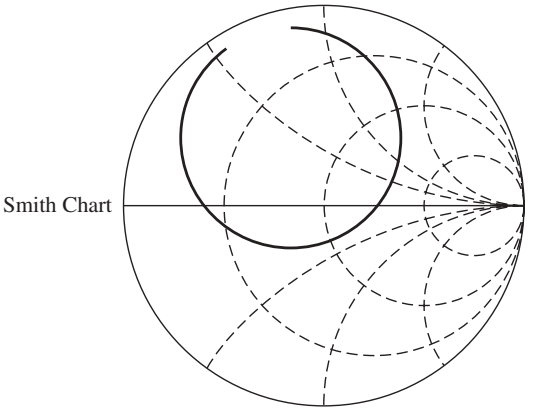
Figure 23.89 shows the antenna structure and dimensions of a composite ceramic–resin type of chip antenna. This type of antenna is designed for three-band operation at the 2.4-GHz band, aiming at application to GSM, PCS, and UMTS. Figure 23.90 illustrates the antenna on the ground plane, which was used for measurement of the antenna characteristics. Figure 23.91 shows return loss characteristics and Table 23.7 provides antenna gain and efficiency for three bands. Figure 23.92 shows radiation patterns in three bands. Figure 23.93 shows a handset model to which this type of antenna is mounted.

23.6.4 Practical Antennas for Small Mobile Terminals

23.6.4.1 Handset Antennas A handset antenna used for the PDC (personal digital cellular) system is an example that employs a monopole (MP), a normal mode helical antenna (NMHA), and a planar inverted-F antenna (PIFA), as shown in Figure 23.59 [1, p. 348]. The NMHA element is placed on top of the MP. The MP element and the NMHA element are separated electrically. In the transmitting/receiving mode, the MP element is extracted outside the handset case and fed at the bottom. The NMHA is not active in this situation; only the MP element radiates. In the receiving mode, as the MP element is retracted inside the handset case and its bottom is disconnected from the feed terminals, only the NMHA acts as an antenna.



(a)



(b)

Antena Type	No GND Area		Peak Freqency [MHz]	Band Width (VSWR<3) [MHz]	Total Average Gain [dBi]		
	[mm]	[mm]			2400 [MHz]	2442 [MHz]	2484 [MHz]
LDA2H Sreies (2.3 × 1.8 × T4.0mm)	3.3 × 4.1	27.1	2443	148	−2.8	−1.6	−1.9

Figure 23.81 Chip antenna characteristics: (a) VSWR and (b) impedance locus on the Smith chart. (Courtesy of Murata Mfg. Co., Japan.)

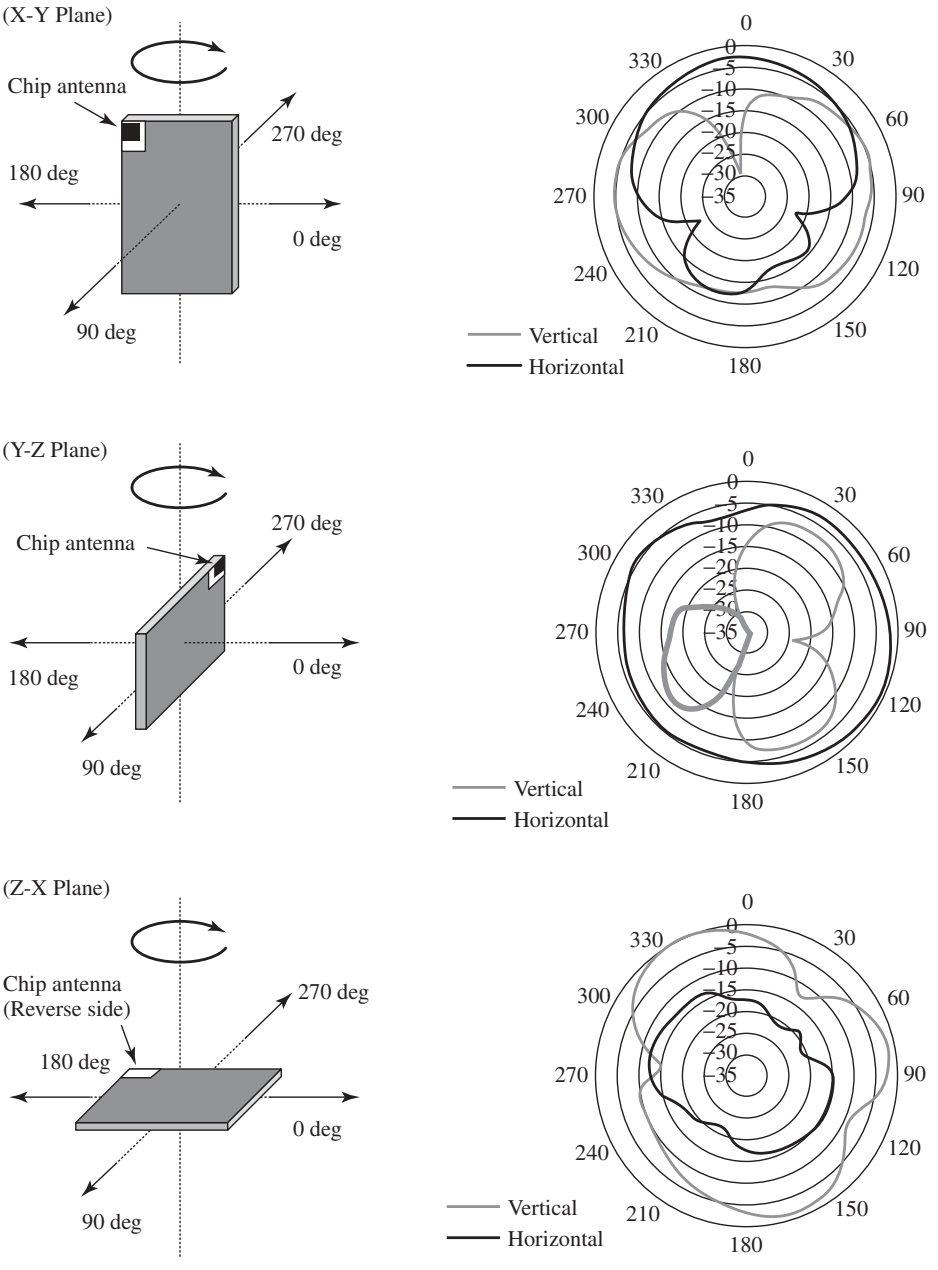


Figure 23.82 Radiation patterns. (Courtesy of Murata Mfg. Co., Japan.)

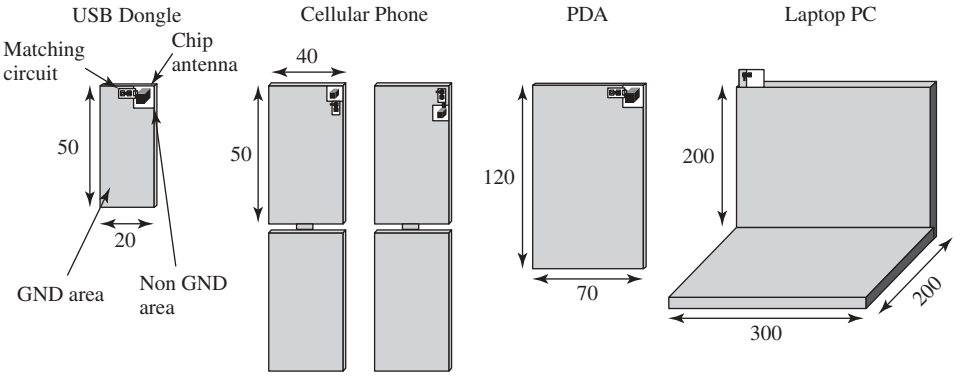
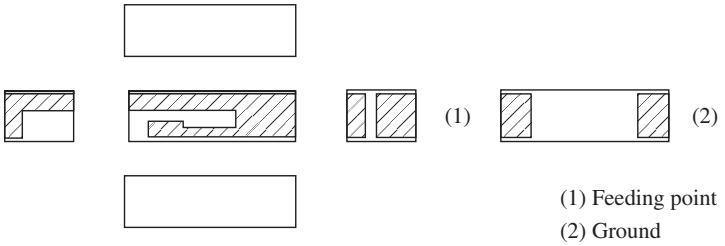
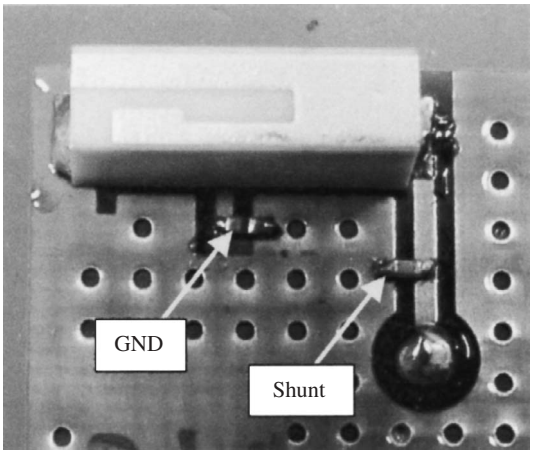


Figure 23.83 Application of chip antenna to various mobile terminals. (Courtesy of Murata Mfg. Co., Japan.)



(a)



(b)

Figure 23.84 A type of chip antenna: (a) the third angle projection of the antenna and printed pattern on the surface of the substrate and (b) antenna and the feed circuit on the ground plane. (Courtesy of Murata Mfg. Co., Japan.)

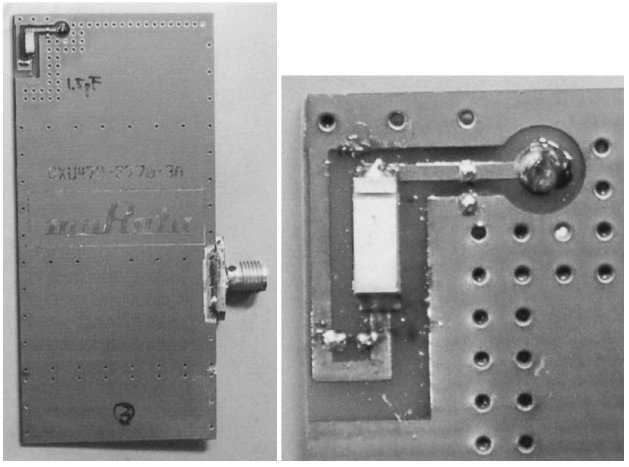


Figure 23.85 A type of chip antenna on the ground plane (GPL): an example of the test model, showing an antenna at the corner of the GPL and the enlarged view, which shows the GPL beneath the antenna partly removed. (Courtesy of Murata Mfg. Co., Japan.)

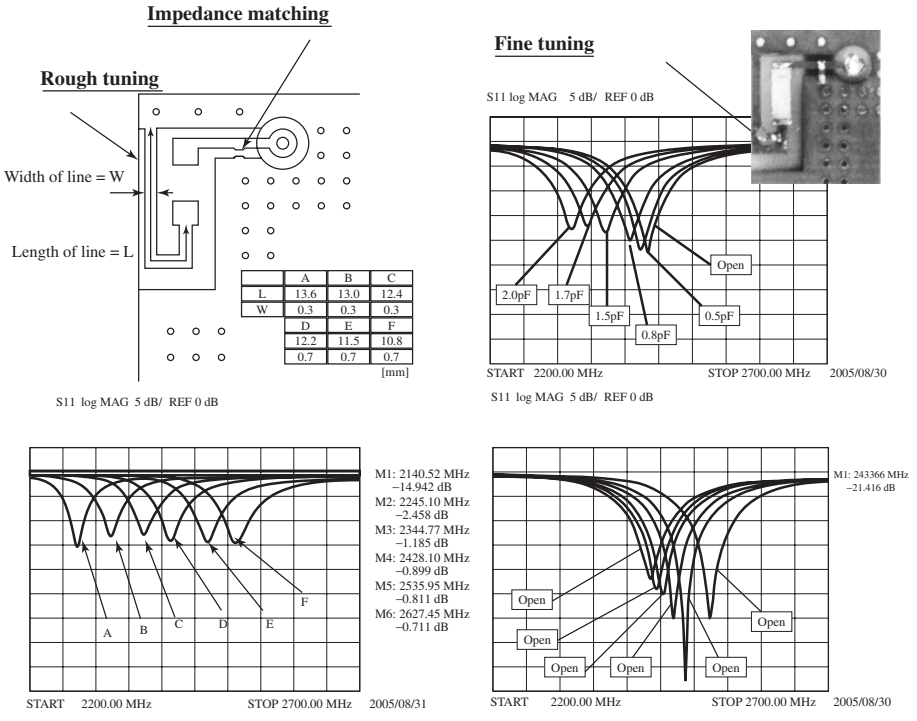
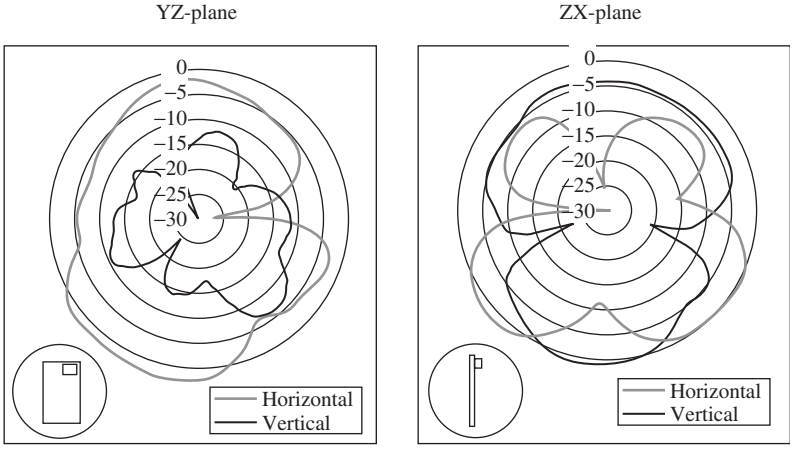


Figure 23.86 Tuning is roughly made with feed circuit arrangement and then precisely made with a reactance component. Variation of resonance frequency is shown with respect to the feed circuit dimensions and either a capacitive or inductive component. (Courtesy of Murata Mfg. Co., Japan.)

TABLE 23.6 Gain and Efficiency at Three Frequencies

[dBi] [dB]

Linear Polarization		YZ-plane hor.	YZ-plane ver.	ZX-plane, hor.	ZX-plane, ver.	Efficiency
2400 MHz	max	1.9	−6.9	2.1	0.9	
	ave	−3.0	−13.6	−3.7	−3.6	−1.7
2442 MHz	**max**	**2.5**	**−6.1**	**2.4**	**1.5**	
	ave	**−2.5**	**−12.9**	**−3.4**	**−3.2**	**−1.4**
2484 MHz	max	2.3	−6.0	2.1	1.0	
	ave	−2.8	−13.1	−3.8	−3.7	−1.7



	HOR.	VER.
MAX	2.5	−6.1
AVE	−2.5	−12.9

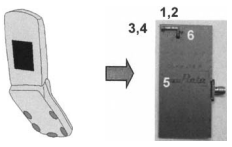
	HOR.	VER.
MAX	2.4	1.5
AVE	−3.4	−3.2

Figure 23.87 Radiation patterns. (Courtesy of Murata Mfg. Co., Japan.)

The typical length of the MP element is either $\frac{3}{8}\lambda$ or $\frac{5}{8}\lambda$, because with these lengths, relatively small current flows on the handset unit and a better input impedance to match the load impedance are achieved than for other lengths [31].

The typical length (extended) of the NMHA is about $\frac{1}{4}\lambda$. Lengths other than $\frac{1}{4}\lambda$ may be used if, for instance, a particular input impedance is specified. The gain of this type of antenna is generally about −0.5 dBd at its peak and about −9 dBd average in the horizontal plane, when the length is about $\frac{1}{4}\lambda$. However, it varies depending on the length of the handset unit.

A significant trend in mobile communications is to have multimode handsets that can operate for such frequency bands as 800 MHz (AMPS), 900 MHz (GSM, PDC,) 1.5 GHz (PDC, GPS), 1.9 GHz (PHS, PCS), 2 GHz (GSM, IMT-2000), 2.4 GHz (Bluetooth, WLAN), and 5 GHz (WiMAX, ITS). Thus the development of multiband antennas that can operate in these frequency bands has notably been accelerated. In the GSM handsets, two-band antennas have been used since several systems have started their services and now multiband antennas are being used as one of the key technologies in



PWB SIZE: 40 × 100 mm

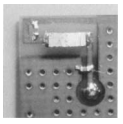
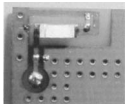
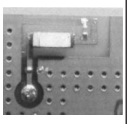
ANT1	ANT2	ANT3	ANT4	ANT5	ANT6
					
BW(VSWR<3)					
84 MHz	129 MHz	128 MHz	94 MHz	132 MHz	132 MHz
Efficiency at 2442 MHz					
–3.2 dB	–1.5 dB	–1.4 dB	–3.2 dB	–0.7 dB	–2.1 dB

Figure 23.88 Placement of antenna on the GPL, and antenna characteristics (bandwidth and efficiency) depending on the antenna placement. (Courtesy of Murata Mfg. Co., Japan.)

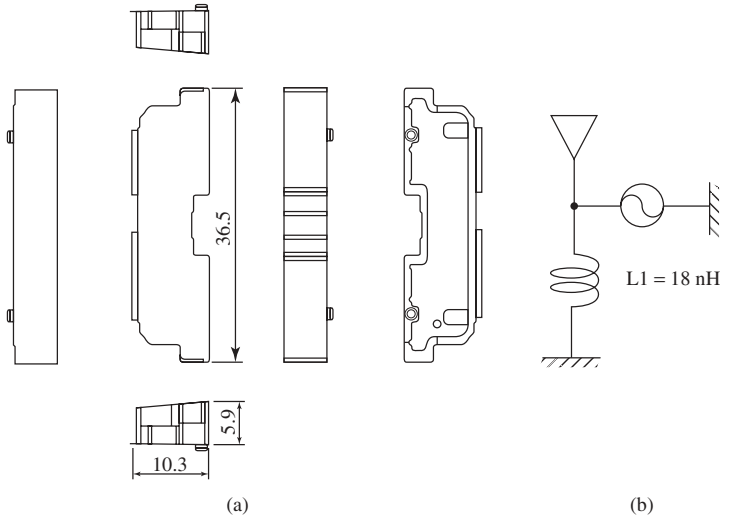


Figure 23.89 A chip antenna of composite resin and ceramic material: (a) the third angle projection and (b) tuning circuit. (Courtesy of Murata Mfg. Co., Japan.)

handset design. Handsets in PDC and the FOMA system in Japan have recently installed dual-band antennas; the lower bands are 810–828 MHz and 940–958 MHz, respectively, for receiving and transmitting, and the higher bands are 1477–1501 MHz and 1429–1453 MHz, respectively, for receiving and transmitting. In IMT-2000, 2-GHz bands are used. In order to achieve antennas that can operate over multifrequency bands, various techniques are used. Lengthening the antenna's electrical length so that the resonance occurs at multifrequency bands is the simplest way. There are several methods for lengthening

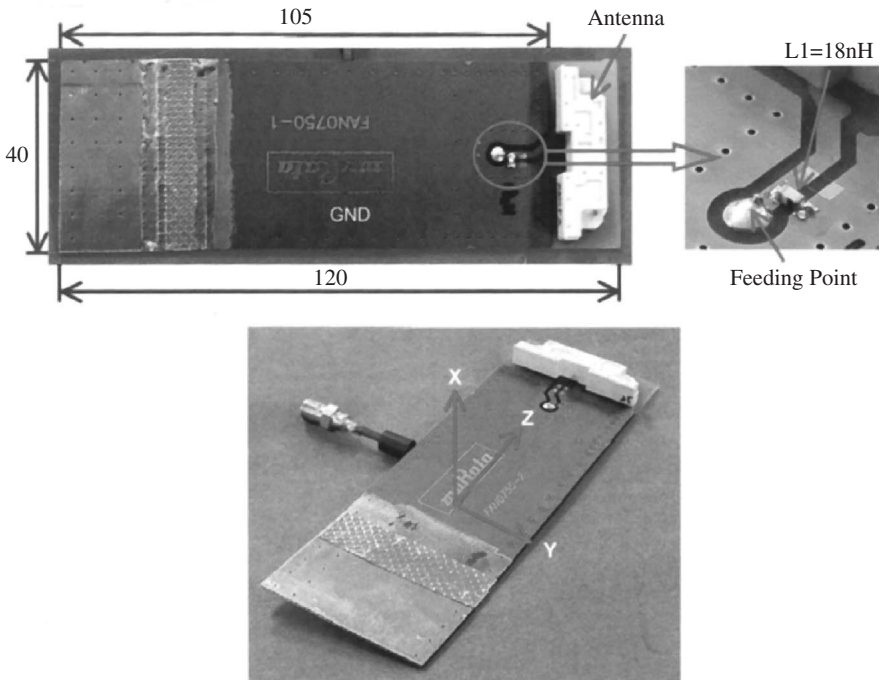


Figure 23.90 A chip antenna: a test model, an antenna on the GPL, the feed line, and the model with the coordinates. (Courtesy of Murata Mfg. Co., Japan.)

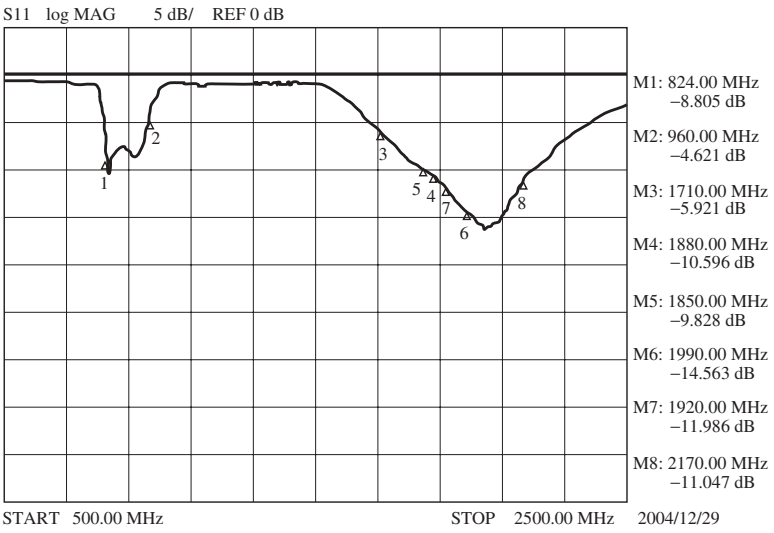


Figure 23.91 Return loss characteristics. (Courtesy of Murata Mfg. Co., Japan.)

TABLE 23.7 Gain and Efficiency at Three Frequency Bands for Three Types of Mobile System

GSM850/EGSM						
Linear Polarization		YZ-plane		ZX-plane (dBi)		Efficiency (dB)
		hor.	ver.	hor.	ver.	
824 MHz	MAX	0.4	−10.4	0.2	−11.9	
	AVE	−3.4	−15.4	−3.1	−14.1	−1.8
892 MHz	**MAX**	**1.5**	**−13.9**	**1.8**	**−16.8**	
	AVE	**−1.9**	**−16.5**	**−1.5**	**−21.6**	**−0.1**
960 MHz	MAX	−0.2	−12.5	0.1	−16.5	
	AVE	−4.4	−16.0	−3.7	−19.2	−2.1

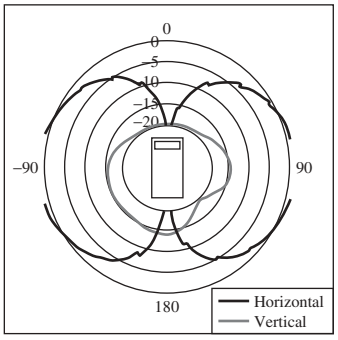
DCS/PCS						
Linear Polarization		YZ-plane		ZX-plane (dBi)		Efficiency (dB)
		hor.	ver.	hor.	ver.	
1710 MHz	MAX	0.2	−5.0	1.5	−7.5	
	AVE	−5.3	−11.0	−3.6	−10.8	−2.6
1850 MHz	**MAX**	**1.3**	**−3.3**	**2.6**	**−7.2**	
	AVE	**−3.7**	**−9.0**	**−2.4**	**−8.2**	**−1.3**
1990 MHz	MAX	1.8	−3.1	3.3	−4.9	
	AVE	−3.2	−8.7	−1.8	−6.5	−0.6

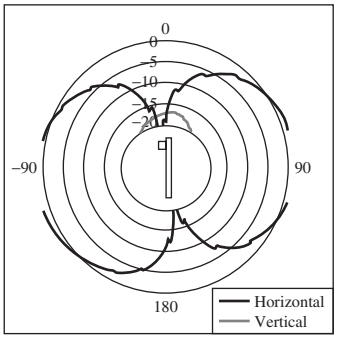
UMTS						
Linear Polarization		YZ-plane		ZX-plane (dBi)		Efficiency (dB)
		hor.	ver.	hor.	ver.	
1920 MHz	MAX	2.1	−3.1	3.6	−5.5	
	AVE	−3.0	−9.2	−1.4	−7.0	−0.9
2045 MHz	**MAX**	**1.7**	**−3.3**	**3.1**	**−4.5**	
	AVE	**−3.1**	**−8.9**	**−1.8**	**−7.0**	**−0.5**
2170 MHz	MAX	0.4	−4.2	2.5	−5.3	
	AVE	−3.9	−9.8	−2.3	−7.6	−1.1

the antenna length: use of a traveling-wave structure, loading of impedance for multiple resonances, and arranging a matching circuit for multifrequency operation. When the bandwidth required is too wide to be covered by a matching circuit, a switching circuit using either a diode or a RF MEMS (microelectromechanical system) is used.

Figure 23.94a shows an example of a triple-band antenna, which has a meander line type element with two coupled parasitic elements. The antenna operates effectively in the AMPS 800 (824–894 MHz), GSM 900 (880–960 MHz), and GSM 1900 (1850–1990 MHz) with VSWR of less than 2.5. The return loss characteristics are shown in Figure 23.94b [43].

Figure 23.95a is another example of a multiband antenna, which is a type of PIFA modified by including a slot on the surface. It works for GSM in three bands—800 MHz, 1800 MHz, and 1900 MHz—and the Bluetooth band. Figure 23.95b depicts the essential part of the antenna.

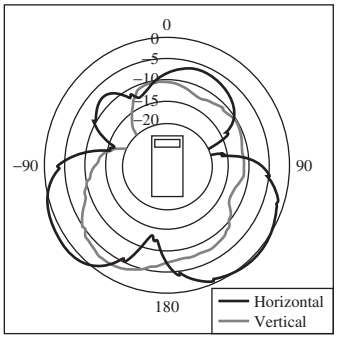


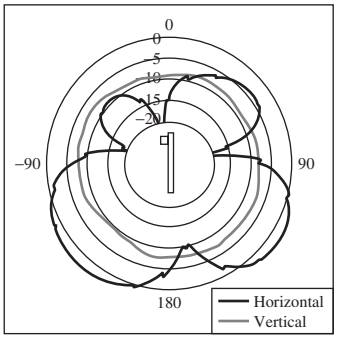


	HOR.	VER.
MAX	1.5	−13.9
AVE	−1.9	−16.5

	HOR.	VER.
MAX	1.8	−16.8
AVE	−1.5	−21.6

(a) [GSM850 / EGSM]

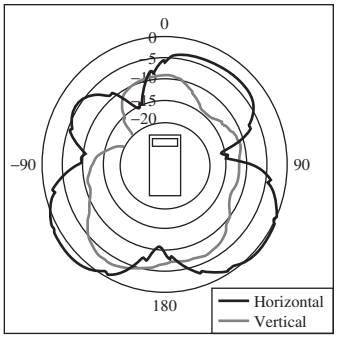


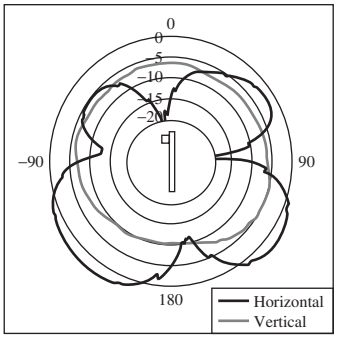


	HOR.	VER.
MAX	1.3	−3.3
AVE	−3.7	−9.0

	HOR.	VER.
MAX	2.6	−7.2
AVE	−2.4	−8.2

(b) [DCS / PCS]





	HOR.	VER.
MAX	1.7	−3.3
AVE	−3.1	−8.9

YZ Plane

	HOR.	VER.
MAX	3.1	−4.5
AVE	−1.8	−7.0

ZX Plane

(c) [UMTS]

Figure 23.92 Radiation patterns. (Courtesy of Murata Mfg. Co., Japan.)

Figure 23.93 Example of a handset on which a chip antenna is installed. (Courtesy of Sanyo Electric Co., Japan.)

An example of a dual-band NMHA, which has a nonuniform helical structure that enables dual-band operation, is shown in Figure 23.96. Figure 23.97 shows its return loss characteristics. Table 23.8 gives maximum and mean gains in free space and at the talk position for various frequency bands.

Antenna performance of a handset usually degrades at the talk position mainly because of the effect of the human body, such as an operator's hand and head. The major reason for the degradation is the variation in the current flow on the handset unit.

It is almost unavoidable that currents on the handset unit are produced by the excitation of the built-in antenna. Since the antenna performance depends on this current flow, antenna performance may vary as the current flow varies due to the body effect. The reduction of such current flow may be advantageous for avoiding variation in the antenna performance. One way to reduce the current flow on the handset unit is to use an antenna system that has balanced terminals, with the antenna fed by a balanced line [44]. An example that uses a folded planar loop antenna placed on a ground plane and fed with a two-wire line is shown in Figure 23.98, in which two types of antenna elements are shown: one is a normal folded loop and the other is a bent-folded loop type. The antenna feed is illustrated in the inset of the figure for (a) an unbalanced system and (b) a balanced system. The current distributions on the ground plane are illustrated in Figure 23.99a for an unbalanced system in free space, in Figure 23.99b for a balanced

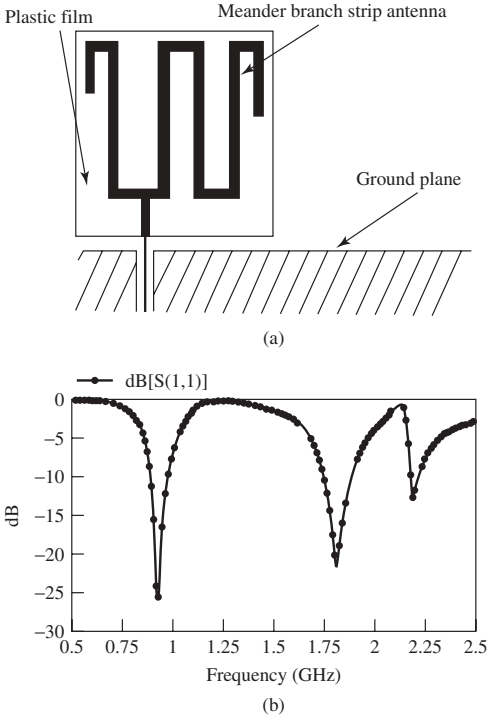


Figure 23.94 Meander line branch antenna: (a) antenna model and (b) return loss characteristics (calculated).

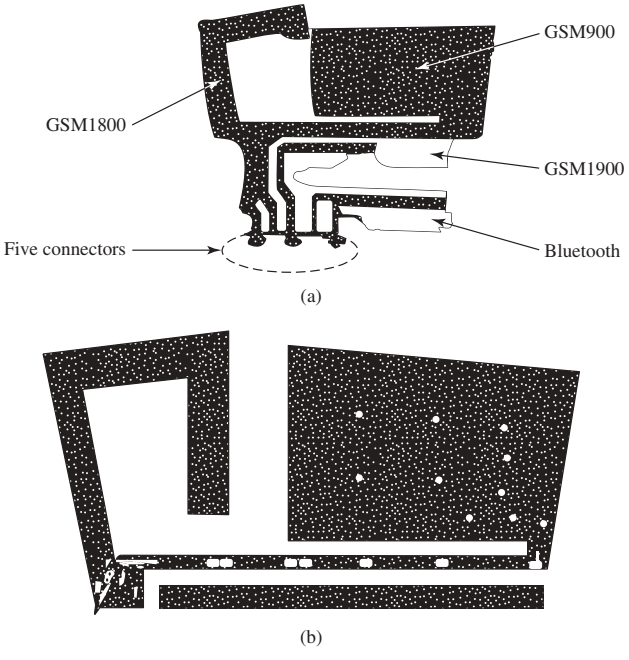
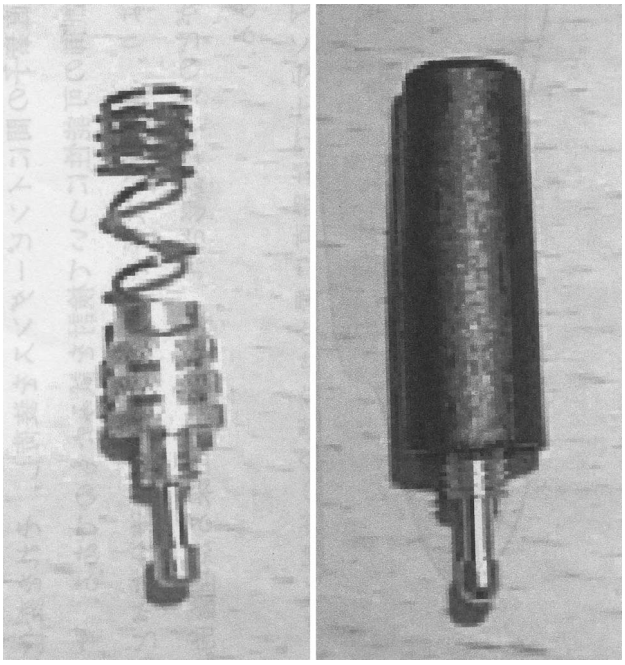


Figure 23.95 Planar multiband antenna: (a) antenna model and (b) antenna pattern.



(a) (b)

Figure 23.96 Nonuniform NMHA: (a) exploded view and (b) plastic covered model.

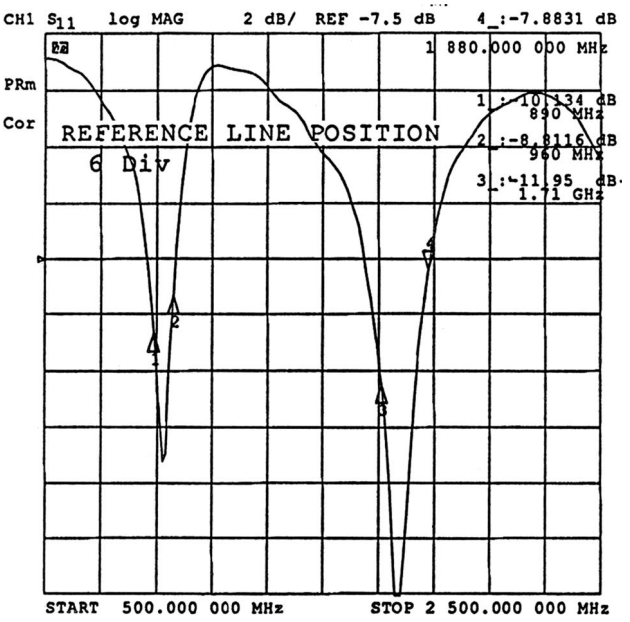


Figure 23.97 Return loss characteristics.

TABLE 23.8 Measured Maximum and Minimum Gain in Free Space and at the Talk Position for Various Frequency Bands

MHz	Maximum Gain Free-Standing (dBd) E_2-plane	Mean Gain Free-Standing (dBd) E_2-plane	Maximum Gain Talk Position (dBd) H-plane	Mean Gain Talk Position (dBd) H-plane
890	0.34	−2.78	−4.82	−9.26
915	0.33	−3.34	−2.66	−7.77
935	0.62	−3.24	−3.04	−8.09
960	−0.82	−4.80	−3.57	−8.01
1710	2.30	−2.34	1.41	−4.66
1785	1.72	−1.81	1.18	−4.66
1805	0.23	−2.84	0.66	−5.33
1880	0.62	−3.38	3.23	−3.35

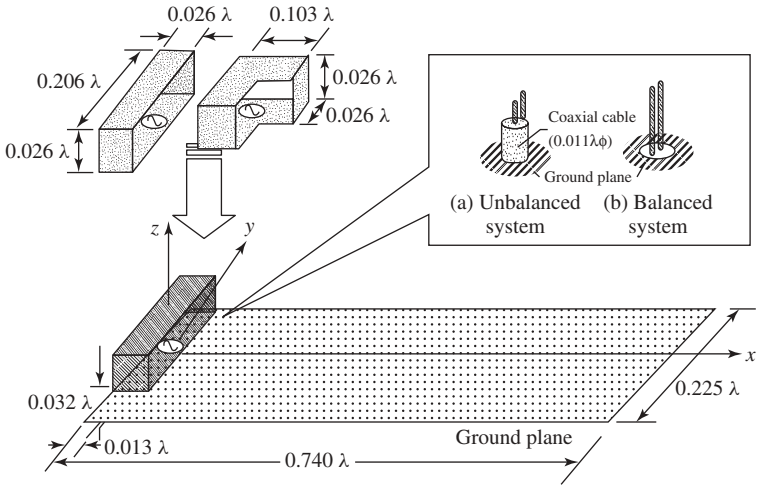


Figure 23.98 Balanced antennas on the ground plane: a folded loop element and bent-folded loop antenna. Feed systems are shown in the inset.

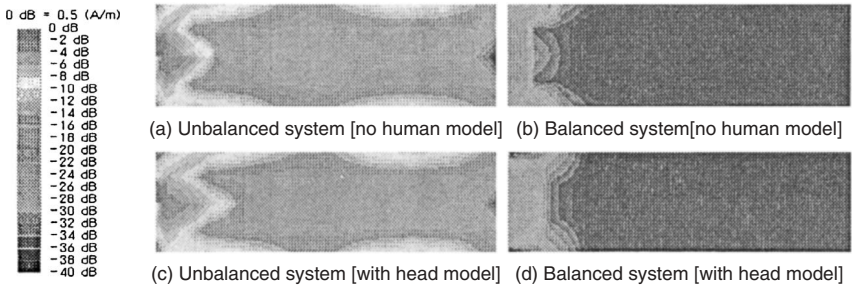


Figure 23.99 Current distributions on the ground plane: (a) unbalanced system without human model, (b) balanced system without human model, (c) unbalanced system with human model, and (d) balanced system with human model.

system in free space, in Figure 23.99c for an unbalanced system with a phantom (human model) beside the antenna, and in Figure 23.99d for a balanced system with a phantom model beside the antenna. From these figures, remarkable reduction of currents on the ground plane can be observed in the balanced system, regardless of the existence of the phantom.

The antenna element is not necessarily a loop for constituting a balanced antenna system; any other element such as NMHA or meander line can be used, whenever they have balanced input terminals and are fed with a balanced line. One should be careful if an antenna system has a very asymmetric structure, even when the terminals are a balanced structure and are fed with a balanced line. With an asymmetric structure, there may be some difficulty in achieving a truly electrically balanced behavior, and thus some unbalanced currents may flow at the terminals.

Practical application of a balanced type antenna is seen in one of the latest handsets. Figure 23.100 depicts a model of a U-dipole with its dimensions [45]. The practical antenna is located on top of the handset as shown in Figure 23.101. The radiation patterns in free space and at a talk position are illustrated in Figures 23.102 and Figure 23.103, respectively. The antenna has a balanced mode and is fed with a balanced line through a balun.

As mobile terminals have become smaller, installation of inner antenna elements is limited to a small area. Thus the antenna element is forced to integrate with nearby circuitry and components. An example is shown in Figure 23.104, where an integrated antenna structure coexists with a speaker and a camera unit. Another one is illustrated in Figure 23.105, which also shows an integrated antenna structure with a speaker unit.

23.6.4.2 Diversity Antenna It is very common to use two antenna elements in a diversity system. However, there has been an exception in a type of PDC handset, where a single antenna element is devised to perform the diversity function [46]. An antenna is applied to a clamshell-type handset (Figure 23.106), the unit of which is composed of two parts: one part flips to either open or close the unit. Each of these two parts has circuitry on it, and some radiation currents flow on the ground plane of these parts. Thus they can be used as part of the radiator as well as the antenna element. In order to achieve the diversity function, the two parts are electrically separated and two impedance components (Z1 and Z2) are located between the two parts. This configuration is simplified: Figure 23.106a illustrates the location of the impedances and Figure 23.106b expresses switching of impedances Z1 and Z2 for the diversity performance. Since the radiation pattern can be varied by changing the value of the impedance, for instance, from Z1 to Z2, pattern diversity is achieved. Changing the value of impedance contributes to varying the phase of the current flowing on the ground plane so that additional lower correlation coefficients can be obtained.

23.6.4.3 Antennas for PHS Terminals PHS (Personal Handy-phone System) has been used in Japan as an extension of the indoor use of the cordless phone system to outdoor application. The communication areas are limited to small zones, for example, several hundred meters in diameter. The transmitter power of the mobile terminals is kept as low as 10 mW. The frequency band is 1895–1918 MHz.

The PHS mobile terminals use a single antenna element, which is typically an MP, an NMHA, or a small chip antenna. Because the operating frequency is higher and the

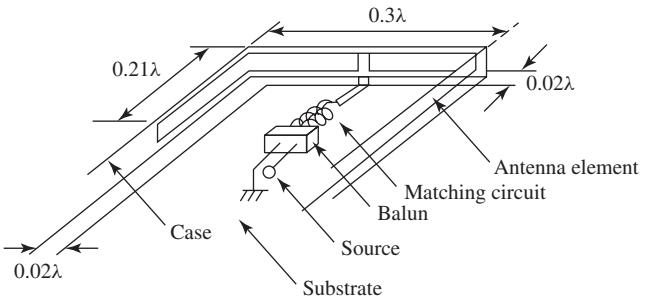


Figure 23.100 Configuration of one-wavelength U-shaped dipole antenna.

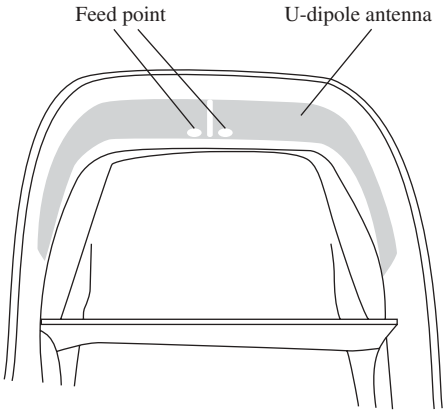


Figure 23.101 U-shaped antenna mounted on top of handset.

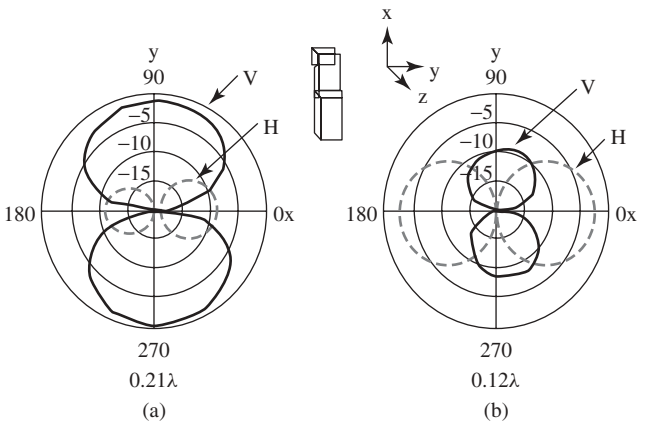


Figure 23.102 Radiation pattern in free space.

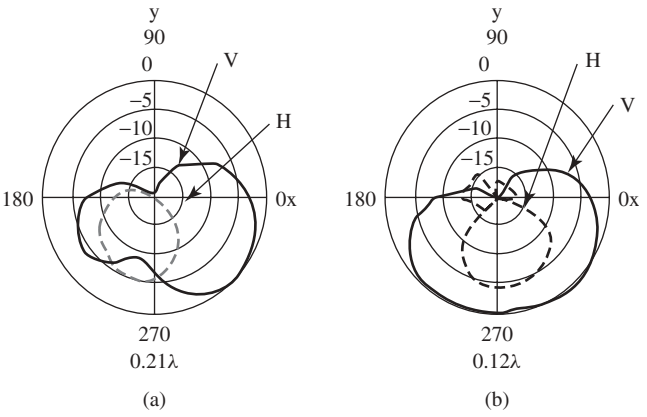


Figure 23.103 Radiation pattern at talk position.

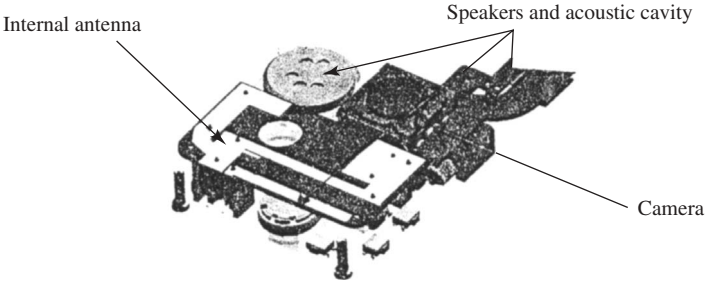


Figure 23.104 Integrated structure of handset antenna with near-by components.

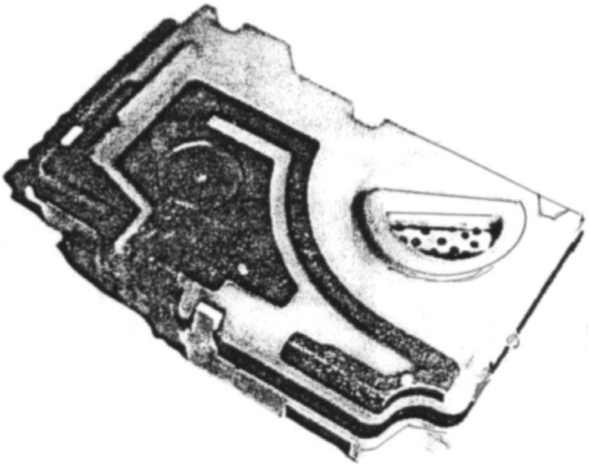


Figure 23.105 Integrated structure of handset antenna with speaker unit.

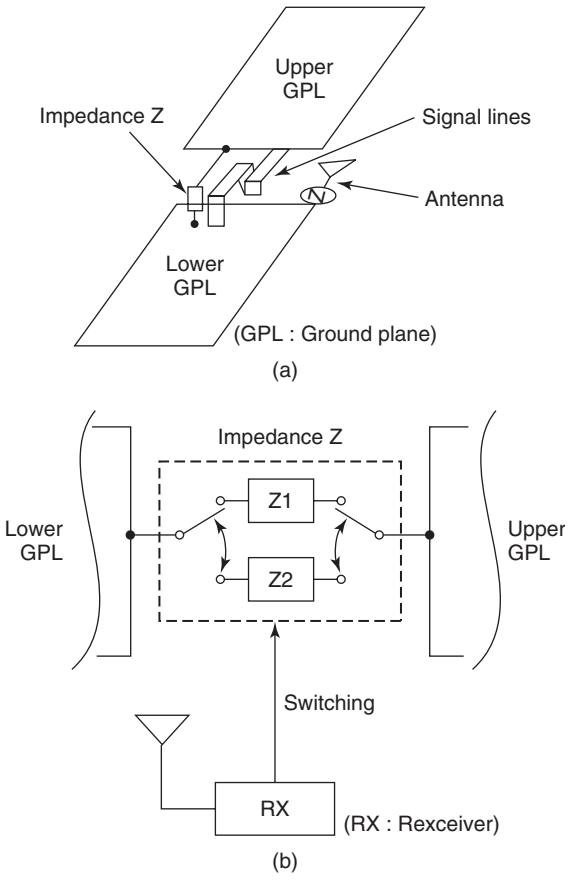


Figure 23.106 Diversity system using a single antenna element: (a) schematic expression and (b) circuitry to function diversity.

bandwidth is narrower than those in the PDC system, a smaller antenna element can be used. Thus the antenna design is easier than that in the PDC systems. The MP element has a similar structure to that in the PDC, but the length is shorter.

Small chip antennas previously introduced in Figure 23.74 also have been used in PHS terminals.

23.7 MEASUREMENT

23.7.1 General

Evaluation of antenna performance is vital for designing antennas and systems to which the antenna is applied, particularly for mobile terminals. Since mobile terminals have small dimensions, and the antenna to be mounted should also be small, various disturbances that affect evaluation of the antenna performance would be encountered when measurements are performed. This is rather natural, because the antenna performance of mobile terminals depends on environmental conditions. There are two possibilities: one

is to eliminate the environmental effects so as to keep antenna characteristics unchanged, thus obtaining correct or precise results; another is to evaluate the antenna performance in a way that includes the effect of the environmental conditions.

The former implies that measurement is performed after the interference parameters are excluded; for example, the influence of a metallic cable connecting the measurement instrument to the antenna under test, which can give rise to disturbance of either the field around the antenna element or the currents flowing on the antenna element and nearby conducting materials.

In the latter case, the measurement is performed even though the nearby effect on the antenna performance cannot be excluded. A typical example is a case where an antenna and the ground plane on which the antenna is mounted cannot be separated electrically. Since some radiation currents induced by the antenna on the ground plane contribute to radiation, this kind of ground plane may intentionally be utilized as a part of an antenna, and so antenna performance is evaluated for the antenna system, in which the ground plane is included.

Difficulty in the measurement of antenna performance may often be encountered when (1) antenna dimensions are small compared to the wavelength, (2) the antenna system including nearby materials has asymmetric structure, and (3) proximity effects should strictly be avoided. Many techniques for evaluating the performance of mobile terminal antennas are common to those of small antennas, because most mobile terminal antenna systems have small structures.

Since the general considerations necessary for measurement can be found in Refs. [47] and [48], some modern technologies used in the measurement of mobile terminal antenna characteristics, such as radiation patterns and radiation efficiency, will be described.

23.7.2 Radiation Patterns

23.7.2.1 Far-Field Measurement The measurement is usually performed in a radio darkroom. It is common that radiation patterns of a small mobile terminal are measured three-dimensionally, typically in the X-Y, Y-Z, and Z-X planes, for both horizontal and vertical polarizations. The reason why all three plane patterns with two polarizations are measured is that a mobile terminal is often operated in a multipath environment, where incident waves vary unpredictably in three-dimensional (3D) space, include cross polarizations, and arrive from indefinite directions. Thus such environmental conditions must be reproduced for evaluating the antenna performance. It usually takes a long time to determine the 3D patterns. In order to perform the 3D pattern measurement efficiently, a device is available that is composed of a sensor array located on a semicircular platform and a rotator, on which an antenna under test (AUT) is placed. Figure 23.107 illustrates this device. The field sensors receive the fields radiated by an AUT located on the center of the turntable (i.e., the center of the semicircular arm). The sensor element is a crossed dipole so that both V and H polarizations can be received in real time. By rotating an AUT on the table, 3D patterns of two polarizations are measured simultaneously. The total time for the measurement of a 3D pattern is very short compared with pattern measurement on each of the three planes separately.

This device can be used not only for measurement of radiation patterns but also for determining radiation efficiency, gain, and so forth. The processing time for obtaining 3D patterns can be as short as 16 seconds, although it depends on the scanning angle.

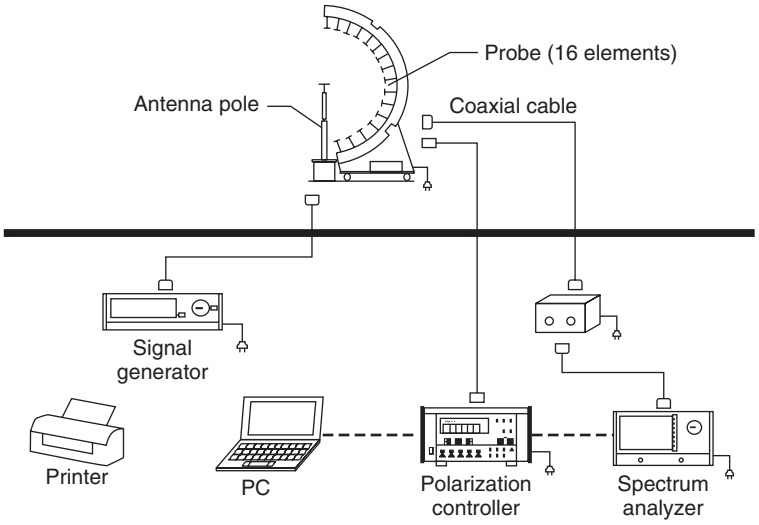


Figure 23.107 Measurement system for 3D radiation patterns.

23.7.2.2 Near-Field Measurement Near-field measurement can be applied to obtain 3D patterns. The measured near-field values are transformed into far-field values by computer processing. There is a device available for measuring near fields, an instrument called the Spherical Near Field Measurement System. The system consists of a spherical scanner, which has an optical field sensor and is capable of scanning spherically around an antenna under test (AUT), and a turntable on which the AUT is mounted. Figure 23.108 illustrates this instrument, which is located in a radio darkroom. The sensor measures the field strength produced by the AUT. The scanner is rotated spherically from the starting position of the sensor up to $\sim 150^{\circ}$ with a precision of 0.1°. By rotating the turntable along with spherical scanning of the scanner, the 3D field strength is obtained in a very short time. A photonic device is used in the sensor in order to transfer directly the electric field strength received by the antenna to the optical strength so as to replace the metallic cable with the optical fiber for connecting the sensor antenna to the instrument. By using optical cable, possible field disturbances that might be caused by the influence of a metallic cable can be avoided. The sensor head is depicted in Figure 23.109, in which a small monopole is shown on the $NbLiO_3$ substrate that transforms electric field strength into optical strength. The near-field strength obtained is transformed into far-field strength by computer processing. The measured data are shown either numerically or graphically on a display.

23.7.3 Radiation Efficiency

There are typically three methods by which radiation efficiency (RE) is evaluated: power integration method, Wheeler Cap method, and random-field method. The most accurate result can be obtained by the power integration method; however, it is the most complicated method and needs rather a large scale.

The Wheeler Cap method is the simplest method among them, although it provides less accurate results than does the power integration method. On the other hand, the

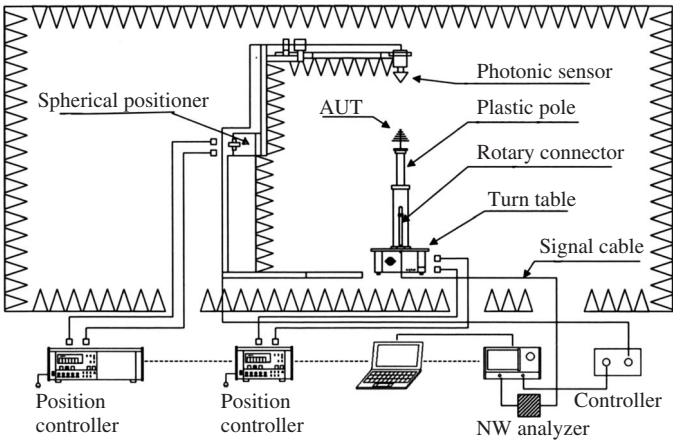


Figure 23.108 Near-field measurement system.

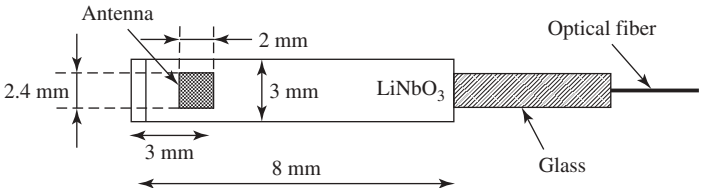


Figure 23.109 Sensor head used in the near-field measurement system.

random-field method is useful, particularly for an antenna system under test located in an environment where random fields exist and where the field is disturbed by the proximity effects, nearby conducting bodies, a human body, and so forth. This situation is often observed in mobile phone terminals in the practical field, operated from a talk position.

RE is defined as the ratio of the radiated power P_{rad} and the input power P_{in} to an antenna and is determined by measuring either these powers or parameters related to these powers.

23.7.3.1 Pattern Integration Method The power P_{in} delivered into an antenna under test (AUT) is expressed by

$$P_{\text{in}} = (1 - |\Gamma|^2) P_{\text{av}} \tag{23.6}$$

where P_{av} and Γ, respectively, are the available power and the reflection factor of the AUT.

Radiated power P_{rad} from the AUT observed at the distance d is given by integration of the radiation powers P_θ and P_φ (vertical and horizontal components, respectively) three-dimensionally as

$$P_{\text{rad}} = \mathrm{d}^2 \int (P_\theta + P_\phi) d\Omega \tag{23.7}$$

$$= \mathrm{d}^2 \int \{W_\theta(\theta, \phi) + W_\phi(\theta, \phi)\} \, d\Omega \tag{23.8}$$

where W_x is the power of the x ($x = \theta, \phi$) component, as a function of θ and ϕ, received at the distance d. W can be expressed by using antenna gain G_r, reflection factor Γ_r, and power density P_x as

$$W_x = P_x(1 - |\Gamma_r|^2)\lambda^2 G_r/4\pi \tag{23.9}$$

Now RE η is given by

$$\eta = P_{\text{rad}}/P_{\text{in}} \tag{23.10}$$

By inserting Eqs. (23.8) and (23.6) into (23.10), the numerical value ofη can be found. A device to perform the pattern integration method is illustrated in Figure 23.110.

23.7.3.2 Wheeler Cap Method In this case, RE can be evaluated by the measurement of input resistances of an AUT for two cases: R_{rad} for a case where the AUT is used for transmitting and R_{loss} for a case where the AUT is used for receiving. The AUT is placed on a ground plane and the resistance $R_r (= R_{\text{rad}} + R_{\text{loss}})$ is measured when it radiates into free space, while R_{loss} is measured when the AUT is covered by a Wheeler Cap, which suppresses radiation of the AUT so that only R_{loss} is observed. R_{rad} can be obtained by taking the difference $(R_r - R_{\text{loss}})$.

Then RE η can be obtained by

$$\eta = R_{\text{rad}}/(R_{\text{rad}} + R_{\text{loss}}) \tag{23.11}$$

This method is the simplest among the above three methods; however, its accuracy is inferior compared to the other two methods.

Care must be taken in use of the Wheeler Cap; losses at the inner surface of the Cap should be kept as low as possible so that no error in the R_{loss} measurement would be included. In addition, measurement should not be performed at resonance frequencies of the Wheeler Cap, as that would contribute serious error in the measured values. Another factor that might give rise to error is alteration of current distributions on the AUT element, which might be caused by covering the AUT with the Wheeler Cap. It

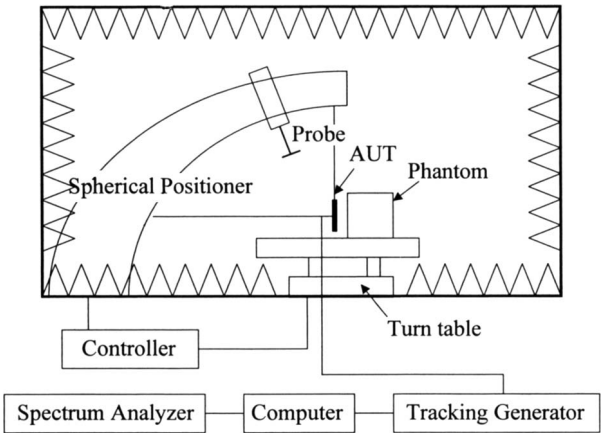


Figure 23.110 Pattern integration measurement system.

is not necessary to use a hemispherical Wheeler Cap, but a rectangular Wheeler Cap is acceptable. Also, the dimensions of the Cap are not necessarily around a radian-sphere size, but may be flexibly chosen with the size a little larger than a radian sphere.

23.7.3.3 Random-Field Method The random-field method is also referred to as the reverberation-field method [49]. The RE is evaluated in an environment where a random field exists as in the practical mobile communication field. The random field is generated in a radio darkroom by means of radio wave scatterers, as shown in Figure 23.111, in which the locations of an AUT, a transmitting antenna, and some scatterers are shown. The power P_{AUT} received by the AUT and the power P_{ref} received by a reference antenna are measured, and the difference between them is used to determine the RE.

Reverberation-field measurement can be applied to measure total power produced by, for example, a mobile phone in practical operation. An example of such a system is illustrated in Figure 23.112, which depicts a reverberation chamber, where evaluation of an antenna system used in a mobile phone at a talk position is performed [49]. This measurement does not imply evaluation of the RE, but gives an indirect measure of

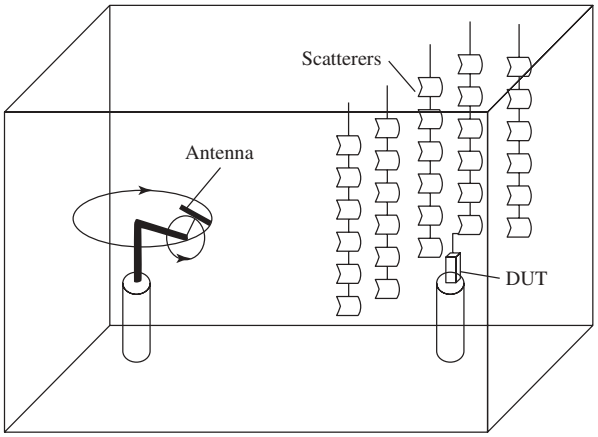


Figure 23.111 Random-field measurement system.

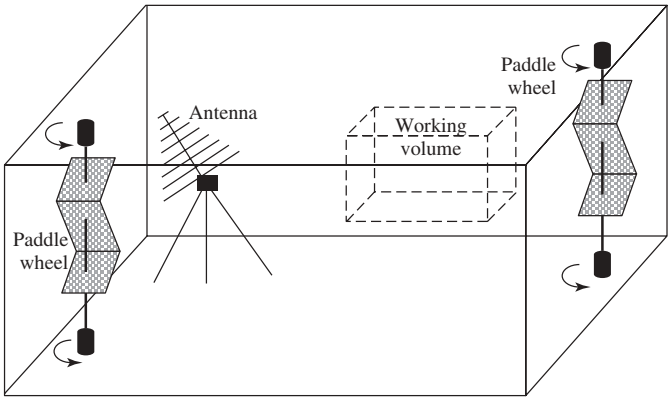


Figure 23.112 Structure of reverberation chamber.

RE, because the total radiated power provides the number from which the loss due to the human body, proximity effects, and so forth can be deduced. The total power is a significant parameter in relation to the SAR value, particularly inside the human head.

REFERENCES

1. K. Fujimoto, Antennas for mobile communications, in *Encyclopedia of Microwave and RF Engineering*, Vol. I, K. Chang (Ed.), John Wiley & Sons, Hoboken, NJ, 2005. K. Fujimoto and J. R. James (Eds.), *Mobile Antenna Systems Handbook*, 2nd ed., Artech House, Norwood, MA, 2000.
2. E. Hawks, *Pioneers of Wireless*, Methuen, London, 1927.
3. S. Isobe, *From Chappe's Semaphore Telegraph to Satellite Communications*, KDD Data Center, 1968 (in Japanese).
4. Historical papers on development in antennas and propagation, 1912 to 1962, *Proc. IRE* (50th Anniversary Issue), pp. 692–717, 1962.
5. J. R. James and G. Andrasic, Multifunction printed antennas, in *Microstrip and Printed Antennas*, H. F. Lee and W. Chen (Eds.), John Wiley & Sons, Hoboken, NJ, 1997, Chap. 6, pp. 273–324.
6. Y. Okumura and A. Akeyama, in *Radio Wave Propagation in Mobile Communications*, Y. Okumura and M. Shinji (Eds.), IEICE Press, 1986, pp. 24–59 (in Japanese).
7. N. Nakajima, H. Mishima, and Y. Yamada, Mobile communications antennas, in *Basic Technology of Mobile Communications*, Y. Okumura and M. Shinji (Eds.), IEICE Press, 1985, Chap. 10 (in Japanese).
8. M. Nakano, T. Satoh and H. Arai, Up-link polarization diversity measurement of hand-held terminal at 900 MHz, Technical Report, AP95-13, IEICE, 1995, pp. 13–19 (in Japanese).
9. K. Tachikawa (Ed.), *W-CDMA Mobile Communication Systems*, Maruzen Co., Japan, 2002, pp. 210–211 (in Japanese).
10. Y. Ebine and M. Sakamorto, System configuration and control, in *Basic Technology of Mobile Communications*, Y. Okumura and M. Shinji (Eds.), IEICE Press, Japan, 1986, Chap. 8, pp. 188–217 (in Japanese).
11. H. Mishima, Y. Ebine, and K. Watanabe, Base station antenna and multiplexer systems for land mobile telephone system, *Electrical Commun. Lab. Tech. J.*, Vol. 26, No. 7, pp. 2011–2036, 1977 (in Japanese).
12. Y. Yamada, T. Nara, S. Kameo, Y. Chitani, and H. Abe, A variable beamwidth corner reflector antenna, *IEICE Natl. Conv. Rec*., p. 694, 1986 (in Japanese).
13. H. Arai and K. Cho, Cellular and PHS base station antenna systems, *Trans. IEICE*, Vol. E86-B, No. 3, p. 982, March 2003.
14. M. Kijima and Y. Yamada, Beam narrowing method for radiation pattern with suppressing some sidelobes, *IEICE Tech Rep. Antennas Propag.*, Vol. AP91, p. 125, 1992 (in Japanese).
15. R. H. Clark, Statistical theory of mobile radio reception, *Bell Syst. Tech. J.*, Vol. 47, No. 6, pp. 957–1000, 1968.
16. Y. Ebine, Design of base station antennas for next generation cellular mobile radios (IMT-2000), *Tech. Rep. IEICE Trans. Commun*. P-2000 (in Japanese).
17. M. Karikomi, Effect of parasitic elements for 0.9/1.5/2 GHz corner-reflector antenna with 120° beamwidth, *IEICE Communication Society Meeting Record* B-I-80, 2002 (in Japanese).
18. Y. Yamada and M. Kijima, A slender two beam base station antenna for mobile radio, *Proc IEEE APS Int. Symp*., pp. 352–355, 1994.

19. Y. Ebine and M. Ito, Design of a dual beam antenna used for base station of cellular mobile radios, *IEICE Trans. Commun.*, Vol. J79-B-II, No. 11, pp. 909–915, 1996 (in Japanese).
20. H. Arai and M. Nakano, Up-link polarization diversity and antenna gain measurement of hand held terminal at 900 MHz, *Proceedings of MDMC'94*, November 1994.
21. M. Nakano, T. Satoh, and H. Arai, Up-link polarization diversity measurement of hand–held terminal at 900 MHz, *Tech Rep. IEICE*, Vol. AP95-13, pp. 13–19, May 1995 (in Japanese).
22. M. Nakano, T. Satoh, and H. Arai, Up-link polarization diversity and antenna gain measurement of hand-held terminal, *IEEE AP-S Int. Symp. Dig*., pp. 1940–1943, 1995.
23. M. Nakano, T. Satoh, and H. Arai, Up-link polarization diversity and antenna gain measurement of hand-held terminal, *IEICE National Convention*, B-11, 1995 (in Japanese).
24. M. Nakano, S. Aizawa, T. Satoh, T. Matsuoka, and H. Arai, Up-link polarization diversity measurement for cellular communication systems using hand-held terminal, *Proc. IEEE AP-S Int. Symp.*, Vol. 66, No. 8, pp. 1360–1363, July 1997.
25. M. Nakano, S. Aizawa, T. Satoh, T. Matsuoka, and H. Arai, Up-link polarization diversity for cellular communication system, *Tech Rep. IEICE*, Vol. AP98-15, pp. 13–20, 1998 (in Japanese).
26. M. Nakano, S. Aizawa, T. Satoh, T. Matsuoka, and H. Arai, Small-sized polarization diversity antenna for cellular base stations, *National Society Convention Record IEICE*, B-1-42, 1997 (in Japanese).
27. M. Nakano and T. Satoh, Polarization diversity for base stations on cdmaOne cellular system, *National Society Convention Record IEICE*, B-1-14, 1998 (in Japanese).
28. Y. Kondo and K. Suwa, Linear predictive transmission diversity for TDMA/TDD personal communication systems, *IEICE Trans. Commun.*, Vol. E79-B, No. 10, pp. 1586–1591, 1996.
29. K. Cho et al., Bidirectional collinear antenna with arc parasitic plates, *Proc. IEEE AP-S Int. Symp*., pp. 1414–1417, 1995.
30. S. Kimura, Kyocera Super PHS-WLL Technology, paper presented at Technical Seminars, 6th Vietnam International Exhibition on Communication, Computer & Office System and 4th Vietnam International Exhibition on Broadcasting Technology, December 1997.
31. H. Morishita, Y. Kim, and K. Fujimoto, Design concept of antennas for small mobile terminals and the future perspectives, *IEEE Antennas Propag. Mag.*, Vol. 44, No. 5, pp. 30–42, 2002.
32. K. Fujimoto, A loaded antenna system applied to VHF portable communication equipment, *IEEE Trans. Vehicular Technol.*, Vol. VT-17, No. 1, pp. 5–12, 1968.
33. K. Hirasawa and K. Fujimoto, Characteristics of wire antennas on a rectangular conducting body, *Trans. IEICE*, Vol. J65-B, No. 9, pp. 1133–1139, 1982 (in Japanese).
34. H. Morishita, H. Furuuchi, and K. Fujimoto, Balance-fed L-type loop antenna system for handset, *Proc. Vehicular Tech. Soc. Int. Conf.*, Vol. 3, pp. 1346–1350, 1999.
35. K. Fujimoto, A. Henderson, K. Hirasawa, and J. R. James, *Small Antennas*, Research Studies Press, 1987, Chap. 4, pp. 89–110.
36. K. Fujimoto, Y. Yamada, and K. Tsunekawa, *Antenna Systems for Mobile Communications*, 2nd ed., Denshi Sohgou Publishing Co., Japan, 1999, p. 55 (in Japanese).
37. J. D. Kraus, *Antennas*, 2nd ed., McGraw-Hill, New York, 1988, pp. 549–558.
38. H. Morishita et al., Small balance-fed helical dipole antenna system for handset, *IEEE VTS International Conference*, May 2000.
39. L. C. Godara (Ed.), *Handbook of Antennas in Wireless Communications*, CRC Press, Boca Raton, FL, 2002, Chap. 12, pp. 12-8–12-10.
40. M. Takiguchi, Miniaturization of meander-line antennas, Master's thesis, National Defense Academy, Japan, 2005 (in Japanese).
41. K. Fujimoto, Antennas for mobile communications, in *Encyclopedia of RF and Microwave Engineering*, Vol. 1, K. Chang (Ed.), John Wiley & Sons, Hoboken, NJ, 2005. p. 351.

42. K. Satoh, K. Matsumoto, K. Fujimoto, and K. Hirasawa, Characteristics of a planar inverted-F antenna on a rectangular conducting body, *Trans. IEICE*, Vol. J71-B, No. 11, pp. 1237–1243, 1988 (in Japanese).
43. Z. Ying, Design of a branch multi-band antenna and efficiency enhancement, *Nordic Antenna Symposium*, Lund, Sweden, 2000.
44. H. Morishita and K. Fujimoto, A balance-fed loop antenna system for handsets, *Trans.* IEICE, Vol. E82-A, No. 7, pp. 1138–1143, 1999.
45. K. Egawa et al., A development of built-in antenna for WCDMA visual terminal, *2002 International Symposium on Antennas and Propgation*, pp. 235–238.
46. T. Fukasawa, K. Kodama, and S. Makino, Diversity antenna for portable telephone with two boxes, *Tech. Rep. IEICE*, Vol. AP2002-78, pp. 87–90, 2002 (in Japanese).
47. K. Fujimoto, Some considerations of small antenna measurements, in *Analysis, Design, and Measurement of Small and Low-Profile Antennas*, K. Hirasawa and M. Haneishi (eds.), Artech House, Norwood, MA, 1992.
48. H. Arai, *Measurement of Mobile Antenna Systems*, Artech House, London, 2001.
49. N. K. Kouveliotis et al., Antenna reverberation chamber, in *Encyclopedia of RF Microwave and Engineering*, Vol. I, K. Chang (Ed.), John Wiley & Sons, Hoboken, NJ, 2005.

CHAPTER 24

Antennas for Mobile Systems

SIMON R. SAUNDERS and ALEJANDRO ARAGÓN-ZAVALA

24.1 INTRODUCTION

Modern mobile systems include cellular radio systems and private land mobile systems. Base stations may be mounted on conventional towers (masts) and rooftops to create macrocells, on street-furniture and building walls to create microcells, or inside buildings to create picocells. In some cases satellites may even be used to create the megacells corresponding to land mobile satellite systems. In all cases, successful provision of user services relies heavily on carefully designed antennas at both the base station and the user terminal (mobile) for efficient operation. It is essential to carefully control radiation patterns to target coverage to desired coverage areas while minimizing interference outside. The antenna structure should minimize interactions with its surroundings such as supporting structures and the human body, while maximizing the efficiency for radiation and reception. The requirements of antennas at both the mobile terminal (Section 24.2) and the base station (Section 24.3) are examined in this chapter, together with an outline of the main structures that are suitable and the key design issues. This chapter is adapted from Chapter 15 of Saunders [1]. For further details, the reader is referred to books such as Stutzmann [2] and Vaughan [3] for detailed design theory and to Fujimoto [4] and Balanis [5] for practical information specific to mobile systems.

24.2 MOBILE TERMINAL ANTENNAS

24.2.1 Performance Requirements

Mobile terminal antennas include those used in cellular phones, walkie-talkies for private and emergency service applications, and data terminals such as laptops and personal digital assistants. Such antennas are subjected to a wide range of variations in the environment that they encounter. The propagation conditions vary from very wide multipath arrival angles to a strong line-of-sight component. The orientation of the terminal is often random, particularly when a phone is in a standby mode. They must be able to operate in close proximity to the user's head and hand. They must also be suitable for

Modern Antenna Handbook. Edited by Constantine A. Balanis

manufacturing in very large volume at an acceptable cost. Increasingly, users prefer that the antenna is fully integrated with the casing of the terminal rather than being separately identifiable. In general, these challenging requirements may be summarized as follows:

- *Radiation Pattern.* Approximately omnidirectional in azimuth and wide beamwidth in the vertical direction, although the precise pattern is usually uncritical given the random orientation, the large degree of multipath, and the pattern disturbance that is inevitable given the close proximity of the user.
- *Input Impedance.* Should be stable and well matched to the source impedance over the whole bandwidth of operation, even in the presence of detuning from the proximity of the user and other objects. Many user terminals now operate over a wide variety of standards, so multiband, multimode operation via several resonances is increasingly a requirement.
- *Efficiency.* Given the low gain of the antenna, it is important to achieve a high translation of input radio frequency (RF) power into radiation over the whole range of conditions of use.
- *Manufacturability.* It should be possible to manufacture the antenna in large volume efficiently, without the need for tuning of individual elements, while being robust enough against mechanical and environmental hazards encountered while moving.
- *Size.* Generally as small as possible, consistent with meeting the performance requirements. Increasingly the ability to adapt the shape to fit into a casing acceptable to a consumer product is important.

This section outlines the most important issues to consider in analyzing this performance and discusses a selection of the various antenna types available and emerging.

24.2.2 Small Antenna Fundamentals

Mobile antennas can often be classified as electrically small antennas, defined as those whose radiating structure can be contained within a sphere of radius r such that $kr < 1$, where k is the wavenumber (i.e., $r \leq \lambda/6$ approximately). There is a body of literature which highlights the performance trade-offs inherent in such antennas. In particular, there is a basic trade-off between size, bandwidth, and directivity. Small antennas, with limited aperture size, cannot achieve high directivity. Similarly, the bandwidth and directivity cannot both be increased if the antenna is to be kept small [6]. As the size decreases, the radiation resistance decreases relative to ohmic losses, thus decreasing efficiency. Increasing the bandwidth tends to decrease efficiency, although dielectric or ferrite loading can decrease the minimum size. This is best illustrated by examining the relationship between the quality factor, Q, of the antenna and the size. The unloaded Q of an antenna is defined as

$$Q = \frac{f_0}{\Delta f_{3\text{dB}}} \tag{24.1}$$

Note that the loaded bandwidth is twice this value. The smallest Q (and hence highest fractional bandwidth) within a given size of enclosing sphere is obtained from a dipole-type field, operating at a fundamental mode of the antenna. Higher-order modes should be avoided as they have intrinsically higher Q. Assuming a resistive matched load,

for a first-order mode, whether electrical or magnetic, the Q is related to the electrical size of the antenna [6] by

$$Q_{01} = \frac{1}{(kr)^3} + \frac{1}{kr} \tag{24.2}$$

If both the electrical and magnetic first-order modes are excited together, then the combined Q is somewhat lower:

$$Q_{E,M01} = \frac{1}{2(kr)^3} + \frac{1}{kr} \tag{24.3}$$

These values are illustrated in Figure 24.1. Practical antennas will always exceed these values. A spherical antenna, such as a spherical helix, approaches the limits most closely but is unlikely to be practical for mobile terminal applications.

This analysis assumed a resistive matched load. The bandwidth can be enhanced using a matching network, such as a bandpass filter, which provides a varying load impedance to the antenna. An even better approach is to make the matching network part of the structure so that the associated currents radiate. This can be achieved via multiple resonant portions to the antenna, such as dipoles of various sizes connected in parallel, or stacked patches of varying sizes.

To illustrate this, near resonance an antenna can be represented by an equivalent series RLC tuned circuit, with an impedance given by

$$\begin{aligned} Z &= R + j\left(\omega L - \frac{1}{\omega C}\right) \\ &= R\left[1 + jQ\left(\frac{f}{f_0} - \frac{f_0}{f}\right)\right] \end{aligned} \tag{24.4}$$

From this we can calculate the voltage standing wave ratio (VSWR). For a simple half-wave dipole with $R = 72\,\Omega$ we obtain the simple dipole characteristics shown in Figure 24.2. By feeding two dipoles with different resonant frequencies at the same input

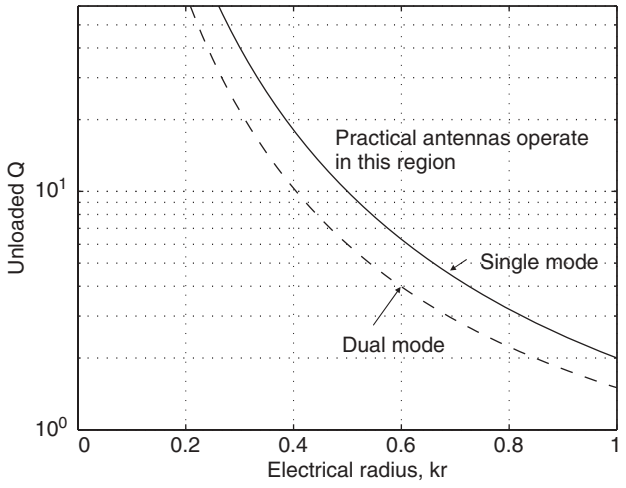


Figure 24.1 Fundamental limits on small antenna efficiency.

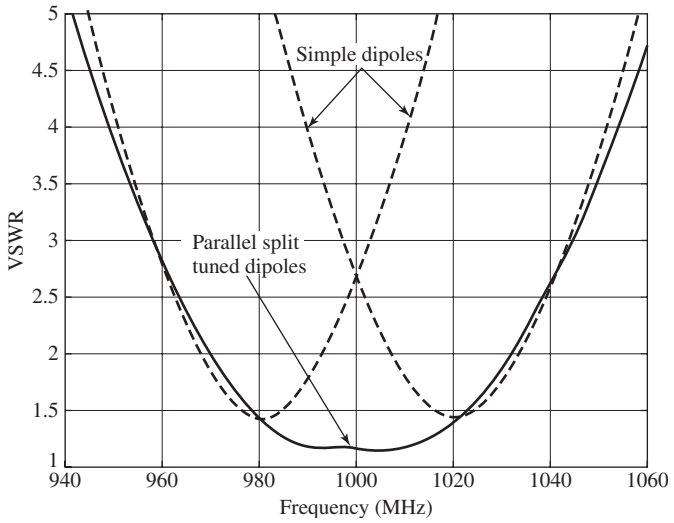


Figure 24.2 Bandwidth enhancement via parallel dipoles. The dipoles are combined in parallel and have resonant frequencies of 980 and 1020 MHz. The source resistance is assumed to be 50 Ω.

terminals, the impedances appear as a parallel combination and the result has a much wider bandwidth (mutual coupling interactions are neglected here).

24.2.3 Dipoles

The most basic starting point for a mobile terminal antenna is a half-wave dipole. Such an antenna has a bandwidth that increases with the diameter of the wire elements in comparison with the overall length. For a length:diameter ratio of 2500 the VSWR = 2 bandwidth is around 8% of the resonant frequency, increasing to 16% for a ratio of 50. A common practical realization of such an antenna is the folded dipole (Figure 24.3a), which increases the impedance fourfold, providing a good match to the 300-Ω balanced

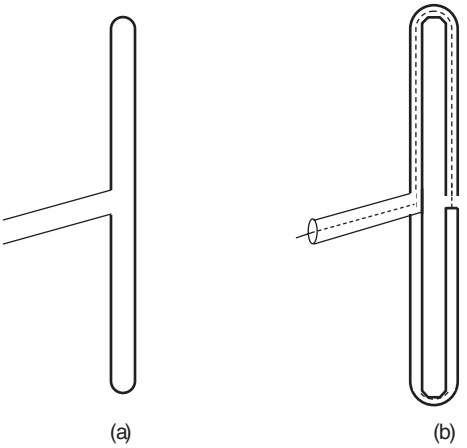


Figure 24.3 Folded half-wave dipoles: (a) balanced feed and (b) coaxial feed.

transmission line and improving the bandwidth. Similarly, Figure 24.3b shows a configuration that provides a coaxial feed. Such configurations are in common practical use as receiving antennas for home FM/VHF radio or as the driven element in Yagi–Uda antennas. The bandwidth may be further increased by broadening or end-loading one or both of the dipole elements, which can produce antennas with bandwidths of many octaves, suitable for ultrawideband systems (Figure 24.4).

Although the half-wave dipole is a versatile and useful antenna in some applications, it is likely to be too large for convenient operation on portable mobiles (e.g., 16.5 cm at 900 MHz). In principle, the lower arm of the dipole could be made the conducting case of the mobile terminal, but then the impedance and the radiation pattern will be severely affected by the interaction between the currents on the case and the user's hand, leading to poor performance.

A quarter-wave monopole (or *quarter-wave whip*), operating over a ground plane, would seem to reduce the antenna size and permit a coaxial feed. However, since the dominant radiation direction of a monopole is along the ground plane, the ground plane size needs to be several wavelengths to produce stable input impedance, rendering the antenna impractical. Nevertheless, such antennas are sometimes used when performance is uncritical.

A common solution to this is to use a *sleeve dipole* as shown in Figure 24.5 [4]. Here the radiating element of the antenna is fed coaxially, with the outer conductor of the coaxial line surrounded by a metal sleeve. The metal sleeve is filled with a cylindrical dielectric insert. If the sleeve is made large in diameter, and the sleeve length and the dielectric constant are chosen appropriately, the sleeve can act as a resonant "choke," which allows RF currents to flow in the outer sleeve but minimizes currents in the terminal case. The antenna is thus fairly robust against variations due to the user's hand on the case and it acts as an asymmetrically fed half-wave dipole with a dipole-like radiation pattern. At frequencies more than around 5% away from the choke resonant frequency, the vertical radiation pattern tends to be multilobed. Sleeve antennas are commonly used for handsets in private mobile radio applications (such as for taxi firms or emergency services) and in older cellular telephones.

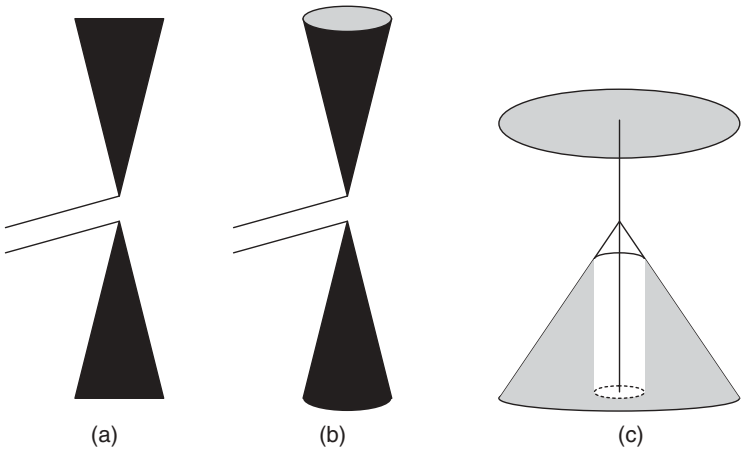


Figure 24.4 Wideband dipole variants: (a) bow-tie antenna, (b) biconical dipole, and (c) top-loaded monopole on conical ground plane or "discone" with coaxial feed.

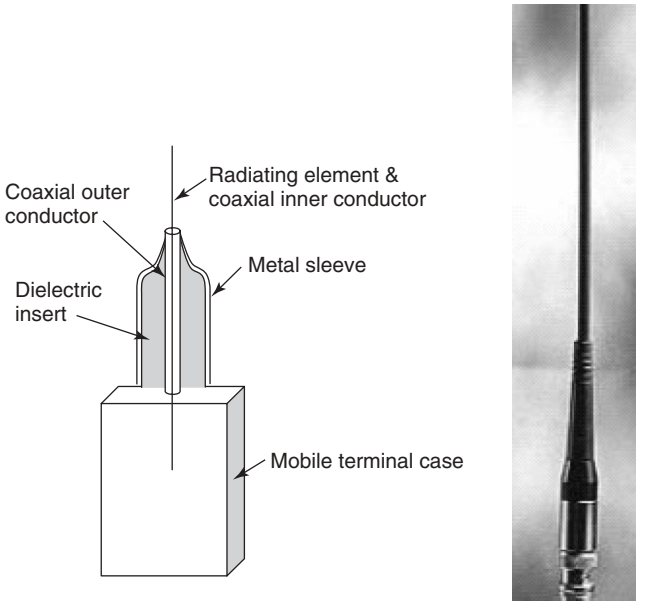


Figure 24.5 Sleeve dipole construction principles and a practical example. (Courtesy of Jaybeam Wireless Ltd.)

24.2.4 Helical Antennas

Given the large size of even a quarter-wave whip at very high frequency (VHF) values for hand-held operation, it is common to reduce the physical size of the radiating element by using a helical antenna radiating in normal mode. However, shortening the element in this way increases the losses, so it is usual to make the element's physical axial length no smaller than approximately $\lambda/12$ (while maintaining its electrical length at $\lambda/4$). Making the element too small also increases the shadowing of the antenna by the user's head. This configuration relies on the use of the case as the balancing element, so it is again strongly influenced by the position of the user's hand on the case. Combining a helical element with a sleeve is one approach, but the sleeve will then dominate the overall size.

24.2.5 Inverted-F Antennas

Given the large ground plane required for efficient operation with a monopole, one solution is to deploy an antenna that produces its maximum radiation normal to the ground plane. The ground plane can then be one side of the terminal case. One such antenna is the inverted-F antenna (Figure 24.6). If the image of this antenna reflected in the ground plane is considered, the antenna appears as a two-wire balanced transmission line with a short circuit at one end. Analysis in this way allows the dimensions to be set to provide resonance and an appropriate impedance match.

A popular development of this antenna type is the *planar inverted-F antenna* (*PIFA*), otherwise referred to as an open microstrip antenna, shown in Figure 24.7 [7]. The wires in the inverted-F antenna are replaced by metal plates, yielding a parallel-plate waveguide between the ground plane and patch, which is often dielectric loaded to reduce the size.

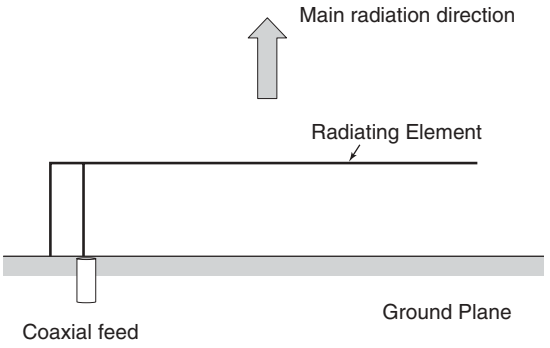


Figure 24.6 Inverted-F wire antenna.

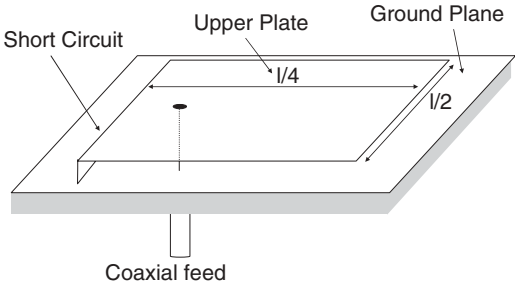


Figure 24.7 Planar inverted-F antenna.

The fields set up in this waveguide are sinusoidal along the length and uniform across the open width, so the radiation pattern is similar to a uniform current dipole in the space above the ground plane. Although the overall dimensions of the antenna are not particularly small, the low profile makes it compact and unobtrusive for cellular telephone applications. The bandwidth is reasonably large, increasing with the height up to around 14%. The bandwidth can be increased still further by adding parasitic conductor layers connected to the upper plate. Its major disadvantage is the awkwardness in manufacture of the short circuit plate.

24.2.6 Patches

Patch antennas are also a common form of portable mobile antenna. They are commonly a half-wavelength square, which is not especially small (depending on the dielectric constant) overall but is compact vertically and avoids the manufacturing awkwardness of the vertical short-circuit in the PIFA. The fields are approximately uniform across the patch width and sinusoidal along the length, resulting in similar patterns to the PIFA. The directivity reduces with increases in the dielectric constant, from around 10 dB for an air substrate to about 5.5 dB for $\varepsilon_r = 10$. The Q is inversely proportional to the patch height and proportional to $\sqrt{\varepsilon_r}$. As with the PIFA, additional layers, producing offset resonances at higher frequencies, can be added to increase the bandwidth, although this increases the manufacturing complexity.

24.2.7 Mean Effective Gain (MEG)

24.2.7.1 Introduction The performance of a practical mobile antenna in its realistic operating environment may be very different from what would be expected for measurements of the gain of the antenna in isolation. This arises because the mobile is usually operated surrounded by scattering objects that spread the signal over a wide range of angles around the mobile. The detailed consequences of this on the fading signal are examined in Saunders [1]. The question arises, given this complexity, as to what value of mobile antenna gain should be adopted when performing link budget calculations. The concept of a *mean effective gain*, which combines the radiation performance of the antenna itself with the propagation characteristics of the surrounding environment, was introduced by Taga [8] to address this and is described in some detail here.

24.2.7.2 Formulation Consider a mobile antenna that receives power from a base station after scattering has occurred through a combination of buildings, trees, and other clutter in the environment. The total average power incident on the mobile is composed of both horizontally and vertically polarized components, P_H and P_V, respectively. All powers are considered as averages, taken after the mobile has moved along a route of several wavelengths. The mean effective gain (MEG) of the antenna, G_e, is then defined as the ratio between the power that the mobile actually receives and the total power available:

$$G_e = \frac{P_{\text{rec}}}{P_V + P_H} \tag{24.5}$$

The received power at the antenna can then be expressed in spherical coordinates taking into account the three-dimensional (3D) spread of incident angles as follows:

$$P_{\text{rec}} = \int_0^{2\pi} \int_0^{\pi} \left[P_1 G_\theta(\theta, \phi)\, P_\theta(\theta, \phi) + P_2 G_\phi(\theta, \phi)\, P_\phi(\theta, \phi) \right] \sin\theta \, d\theta \, d\phi \tag{24.6}$$

where P_1 and P_2 are the mean powers that would be received by ideally θ (elevation) and ϕ (azimuth) polarized isotropic antennas, respectively, G_θ and G_ϕ are the corresponding radiation patterns of the mobile antenna, and P_θ and P_ϕ represent the angular distributions of the incoming waves. The following conditions must be satisfied to ensure that the functions are properly defined:

$$\int_0^{2\pi} \int_0^{\pi} \left[G_\theta(\theta, \phi) + G_\phi(\theta, \phi) \right] \sin\theta \, d\theta \, d\phi = 4\pi \tag{24.7}$$

$$\int_0^{2\pi} \int_0^{\pi} P_\theta(\theta, \phi) \sin\theta \, d\theta \, d\phi = 1 \tag{24.8}$$

$$\int_0^{2\pi} \int_0^{\pi} P_\phi(\theta, \phi) \sin\theta \, d\theta \, d\phi = 1 \tag{24.9}$$

The angular distribution of the waves may be modeled by, for example, Gaussian distributions in elevation and uniform in azimuth as follows:

$$P_\theta(\theta, \phi) = A_\theta \exp\left\{-\left[\theta - \left(\frac{\pi}{2} - m_V\right)\right]^2 \times \frac{1}{2\sigma_V^2}\right\} \tag{24.10}$$

$$P_\phi(\theta, \phi) = A_\phi \exp\left\{-\left[\phi - \left(\frac{\pi}{2} - m_H\right)\right]^2 \times \frac{1}{2\sigma_H^2}\right\} \tag{24.11}$$

where m_V and m_H are the mean elevation angles of the vertically and horizontally polarized components, respectively, σ_V and σ_H are the corresponding standard deviations, and A_θ and A_ϕ are chosen to satisfy Eqs. (24.8) and (24.9). The precise shape of the angular distribution is far less important than its mean and its standard deviation.

24.2.7.3 Example MEG Calculation Since the arrival angle has been assumed uniform in azimuth, any variations from omnidirectional in the radiation pattern will have no impact on the MEG. Although this assumption is likely to be valid in the long term as the mobile user's position changes, there may be short-term cases where this is not so, and the power arrives from a dominant direction. This may particularly be the case in a rural setting where a line-of-sight or near-line-of-sight path exists. Figure 24.8 shows the radiation pattern obtained from a pair of dipoles arranged $\lambda/4$ apart and fed with equal phase and amplitude. This could, for example, represent an attempt to reduce radiation into the human head by placing a null in the appropriate direction. This results in a gain pattern of the form

$$G_\theta = \cos^2\left(\frac{\pi}{4} + \frac{\pi}{4}\cos\phi\right) \tag{24.12}$$

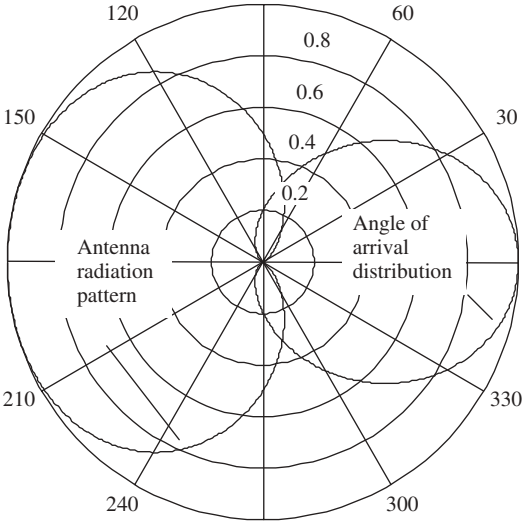


Figure 24.8 Azimuth radiation pattern and angle-of-arrival distribution.

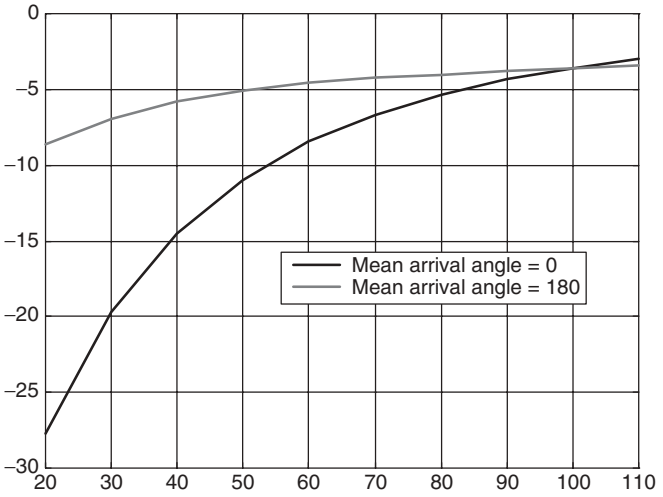


Figure 24.9 Reduction in mean effective gain due to angular spreading.

Also shown is the arrival angle distribution, assumed Gaussian in azimuth and shown with a standard deviation of 50° and a mean of 0°. Calculation of Eq. (24.6) in comparison with the assumption of uniform arrival angle in the azimuth plane yields the results shown in Figure 24.9. This is performed with the center of the arriving waves both within the pattern null and directly opposite and is shown as a function of the standard deviation of the spread relative to the mean. Considerable gain reduction is evident in both cases, particularly when the angular spread is small.

24.2.8 Human Body Interactions and Specific Absorption Rate (SAR)

In the frequency range of interest for practical radio communications, electromagnetic radiation is referred to as "nonionizing radiation" as distinct from the ionizing radiation produced by radioactive sources. The energy associated with the quantum packets or *photons* at these frequencies is insufficient to dissociate electrons from atoms, whatever the power density, so the main source of interactions between nonionizing radiation and surrounding human tissue is simple heating. Nevertheless, given that we are all continually exposed to electromagnetic (EM) radiation from a variety of sources, including mobile phone systems and radio and television broadcasts, there is understandable concern to ensure that human health is not adversely affected.

The potential health impact of EM fields has been studied for many years by both civil and military organizations, as well as the effects and interactions of hand-held antennas with the human head [9]. A number of bodies have commissioned research into such effects and the World Health Organisation has produced guidelines to ensure that this research is conducted according to appropriate standards [10]. The conclusions from these investigations have been used to set regulatory limits on exposure, which reflect a precautionary principle based on the current state of knowledge. Many administrations require equipment manufacturers to ensure that the fields absorbed are below given limits and to quote the values produced by individual equipment under suitable reference conditions. Therefore it is essential at this stage to establish procedures and metrics to assess the impact of antennas on absorption within the body.

24.2.8.1 Formulation The evaluation of human exposure in the near field of RF sources, like portable mobile phones, can be performed by measuring the electric field (E-field) *inside* the body [4]. Given a current density vector $\mathbf{J}$ and an electric field $\mathbf{E}$, the power absorbed per unit volume of human tissue with conductivity σ ($\Omega^{-1} \cdot \mathrm{m}^{-1}$) is

$$P_V = \tfrac{1}{2}\mathbf{J} \cdot \mathbf{E}^* = \tfrac{1}{2}\sigma |\mathbf{E}|^2 \quad (\mathrm{W/m^3}) \tag{24.13}$$

By introducing the density of the tissue ρ (in $\mathrm{kg/m^3}$), the absorption per unit mass is obtained as follows:

$$P_g = \frac{1}{2}\frac{\sigma}{\rho}|E|^2 \quad (\mathrm{W/kg}) \tag{24.14}$$

The term P_g is known as the *specific absorption rate* (SAR), or the rate of change of incremental energy absorbed by an incremental mass contained in a volume of given density. The SAR (divided by the specific heat capacity) indicates the instantaneous rate of temperature increase possible in a given region of tissue, although the actual temperature rise depends on the rate at which the heat is conducted away from the region, both directly and via the flow of fluids such as blood. The power P absorbed by a specific organ is given by

$$P = \int_M P_g dm \tag{24.15}$$

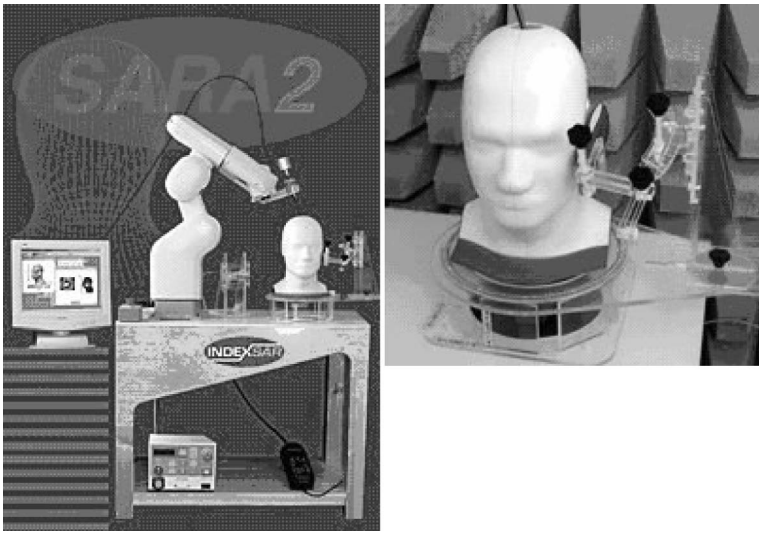
where M is the mass of the object under consideration, which can be the entire human body.

The parameters used in Eqs. (24.13) and (24.14) are complex and depend on many factors. The conductivity of the tissue σ varies with frequency [11], and increases with temperature. The tissue density ρ also changes with tissue and is a function of the water content. The electric field is often determined by the dielectric and physical properties of the tissue, polarization, and exposure environment. The conductivity at various frequencies has been measured and reported in Gabriel [11]. An example of some tissue dielectric values at 900 MHz is shown in Table 24.1.

TABLE 24.1 Tissue Dielectric Properties at 900 MHz

Tissue	Permittivity (ε)	Conductivity (σ)
Bladder	18.93	0.38
Fat (mean)	11.33	0.10
Heart	59.89	1.23
Kidney	58.68	1.39
Skin (dry)	41.40	0.87
Skin (wet)	46.08	0.84
Muscle (parallel fiber)	56.88	0.99
Muscle (transverse fiber)	55.03	0.94
Cerebellum	49.44	1.26
Breast fat	5.42	0.04
Average brain	45.80	0.77
Average skull	16.62	0.24
Average muscle	55.95	0.97

Source: From Ref. 11.



(a) (b)

Figure 24.10 SAR measurement systems: (a) CENELEC compliant system, from *IndexSAR*; and (b) phantom used in SAR tests, with mobile phone head mounting bracket.

24.2.8.2 SAR Measurements A popular method to perform SAR measurements is by logging E-field data in an artificial shape acting as a representation of the human body in normal use, from which SAR distribution and peak averaged SAR can be computed. To do so repeatedly, it is important to use an appropriate body shape and dielectric material. Appropriate "phantoms" for the human head and other body parts are standardized by relevant committees [12, 13], along with the associated positioning of phones and other devices to be tested. Examples of some commercial SAR measurement systems are shown in Figure 24.10. Here a robotic arm is used to hold the E-field probe and scan the whole exposed volume of the phantom, in order to evaluate the 3D field distribution. The use of the robot allows high repeatability and very high position accuracy. A similar procedure can also be used for testing the SAR arising from indoor antennas to establish whether they are "touch safe."

At the early design stages of an antenna, it is also possible to make an assessment of the likely SAR via analytical or numerical calculations. The near-field nature of the problem means that geometrical optical approaches are not appropriate and full-wave solutions of Maxwell's equations must be applied [14].

24.2.8.3 SAR Regulations After characterizing the RF exposure produced through SAR measurements, it is necessary to assess whether this exposure falls beyond acceptable limits. ICNIRP (International Committee on Non-ionising Radiation Protection) is an independent nongovernmental scientific organization, set up by the World Health Organisation and the International Labour Office, responsible for providing guidance and advice on the health hazards of nonionizing radiation exposure [15], and levels based on its recommendations are widely adopted in Europe. On the other hand, the IEEE C95.1–1999 Standard for Safety Levels with Respect to Human Exposure to Radio Frequency Electromagnetic Fields [16] has been adopted in the Americas as a reference, which includes frequencies between 3 kHz and 300 GHz. In 1999, the European

TABLE 24.2 SAR Exposure Limits (W/kg)

Standard	Condition	Frequency	Whole Body	Local SAR (Head and Trunk)	Local SAR (Limbs)
ICNIRP	Occupational	100 kHz to 10 GHz	0.4	10	20
	General public	100 kHz to 10 GHz	0.08	2	4
IEEE	Controlled	100 kHz to 6 GHz	0.4	8	20
	Uncontrolled	100 kHz to 6 GHz	0.08	1.6	4

Source: From Ref. 4.

Committee for Electrotechnical Standardisation (CENELEC) endorsed the guidelines set by ICNIRP on exposure reference levels and recommended that these should form the basis of the European standard [17].

Table 24.2 shows the basic SAR limits for both ICNIRP and IEEE standards. Both standards make a clear distinction between general public (uncontrolled environment) and occupational (controlled environment). For the former, people with no knowledge of or no control over their exposure are included, and hence the exposure limits need to be tighter. However, the general public values are often regarded as representing best practice, whoever the affected parties.

The ICNIRP and IEEE standards also establish field strength and power density limits for far-field exposure, as shown in Tables 24.3 and 24.4. Notice the variations in maximum E-field exposure with frequency. These are the levels typically adopted when testing and predicting the fields around macrocellular base station antennas.

24.2.9 Mobile Satellite Antennas

The key requirements for the antenna on a mobile handset for nongeostationary satellite systems, such as those described in Chapter 14 of Saunders [1], can be summarized as follows:

- *Omnidirectional, Near-Hemispherical Radiation Pattern.* This allows the handset to communicate with satellites receiving from any elevation and azimuth angle, without any special cooperation by the user. The elevation pattern should extend down to at least the minimum elevation angle of the satellite, but should not provide

TABLE 24.3 ICNIRP Reference Field Strength Levels (V/m)

Standard	Condition	>10 MHz, <400 MHz	900 MHz	1.8 GHz	>2 GHz, <300 GHz
ICNIRP	General public	28	41.25	58.3	61
	Occupational	61	90	127.3	137

TABLE 24.4 IEEE Reference E-Field and Power Density Levels

		E-Field (V/m)	Power Density (mW/cm^2)		
Standard	Condition	>30 MHz, <300 MHz	900 MHz	1.8 GHz	>15 GHz, <300 GHz
IEEE	Uncontrolled	27.5	0.6	1.2	10
	Controlled	61.4	3	6	10

too much illumination of angles below the horizon, since this would lead to pick-up of radiated noise from the ground and degradation of the receiver noise figure. The pattern need not necessarily be uniform within the beamwidth; indeed, it may be an advantage in some systems to emphasize lower elevation angles at the expense of higher ones in order to overcome the extra free space and shadowing losses associated with lower satellites [18].

- *Circular Polarization, with Axial Ratio Close to Unity.* This limits the polarization mismatch. A typical specification is that the axial ratio should be no more than 5 dB at elevation angles down to the minimum elevation angle of the satellite constellation.

24.2.9.1 Quadrifilar Helix Antenna (QHA)

A common structure that can be used to meet these requirements is the *quadrifilar helix antenna* (QHA). This was invented by Kilgus [19, 20] and a typical example is shown in Figure 24.11, mounted on top of a conducting box to represent the case of a mobile phone. The four elements of the QHA are placed at 90° to each other around a circle and are fed consecutively at 90° phase difference (0°, 90°, 180°, 270°). The QHA is resonant when the length of each element is an integer number of quarter-wavelengths. When the elements are each one-quarter of a wavelength, for example, the QHA can be considered as a pair of crossed half-wave dipoles, but with the elements folded to save space and to modify the radiation pattern. Adjustment of the number of turns and the axial length allows the radiation pattern to be varied over a wide range, according to the statistics of the satellite constellation and the local environment around the user, while maintaining circular polarization with a small axial ratio from zenith down to low elevation angles [21]. The QHA is relatively unaffected by the presence of the user's head and hand, since the antenna has a fundamentally balanced configuration, leading to relatively little current flowing in the hand-held case, in contrast to terrestrial approaches that use a monopole as the radiating element and rely on the currents in the case to provide a ground plane. The QHA can also be made resonant at multiple frequencies and can be reduced in size via various means [22].

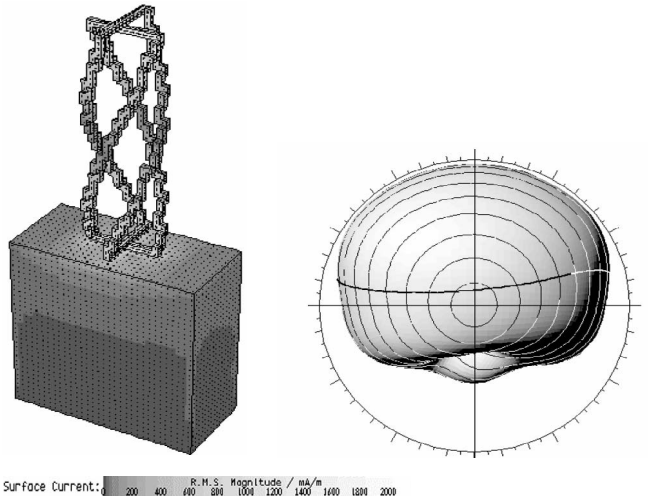


Figure 24.11 Surface current distribution of a quadrifilar helix antenna mounted on a conducting box, plus corresponding (left-hand circularly polarized) radiation pattern.

Although the QHA is dominant for application to hand-held terminals in frequency bands up to around 3 GHz, other antennas are used in other cases. In terminals that are mounted on a vehicle rather than being hand-held, it may be more appropriate to use a circularly polarized patch antenna. For use at higher frequency bands, it is necessary to provide some antenna gain through directivity in order to improve the available fading margin. This necessitates the use of either electrical or mechanical steering in order to ensure the antenna is properly pointed.

An adaptive, intelligent variant of the QHA, the I-QHA, was invented by Agius [23] and described in detail in Leach [24]. The purpose of this configuration is to adapt the antenna to changes in the incoming signal imposed by the environment, the system, and the user handling of the terminal. The I-QHA has been demonstrated to achieve a large potential diversity gain of up to 14 dB [25]. The MEG of the I-QHA has been evaluated in detail by Brown [26].

24.2.9.2 Patch Antennas Patch antennas are also used for satellite mobile terminals and become very attractive due to their low cost and easy manufacture, as well as the reduction in size as a result of the technology used in their construction. Circular polarization may be achieved by dual feeding a square or circular patch at right angles with quadrature phasing, or by perturbing the patch shape (e.g., cutting off one corner) so as to create antiphase currents that produce the same result. A good example of the use of such antennas is in the Global Positioning System (GPS), where patch antennas are the most common configuration, although the QHA is still often applied for high performance requirements.

24.3 BASE STATION ANTENNAS

24.3.1 Performance Requirements in Macrocells

The basic function of a macrocell base station antenna is to provide uniform coverage in the azimuth plane, but to provide directivity in the vertical plane, making the best possible use of the input power by directing it at the ground rather than the sky. If fully omnidirectional coverage is required, vertical directivity is usually provided by creating a vertical array of dipoles, phased to give an appropriate pattern. This is usually called a *collinear* antenna and has the appearance of a simple monopole. A typical radiation pattern for such an antenna is shown in Figure 24.12. Such antennas are commonly used for private mobile radio systems.

More commonly in cellular mobile systems, however, some limited azimuth directivity is required in order to divide the coverage area into sectors. A typical example is shown in Figure 24.13. The choice of the azimuth beamwidth is a trade-off between allowing sufficient overlap between sectors, permitting smooth handovers, and controlling the interference reduction between co-channel sites, which is the main point of sectorization. Typical half-power beamwidths are 85°–90° for 120° sectors. Figure 24.13 also shows some typical patterns.

The elevation pattern of the antenna also has to be carefully designed, as it allows the edge of the cell coverage to be well defined. This can be achieved using either mechanical downtilt, where the antenna is simply pointed slightly downward; or electrical downtilt, where the phase weighting of the individual elements within the panel is

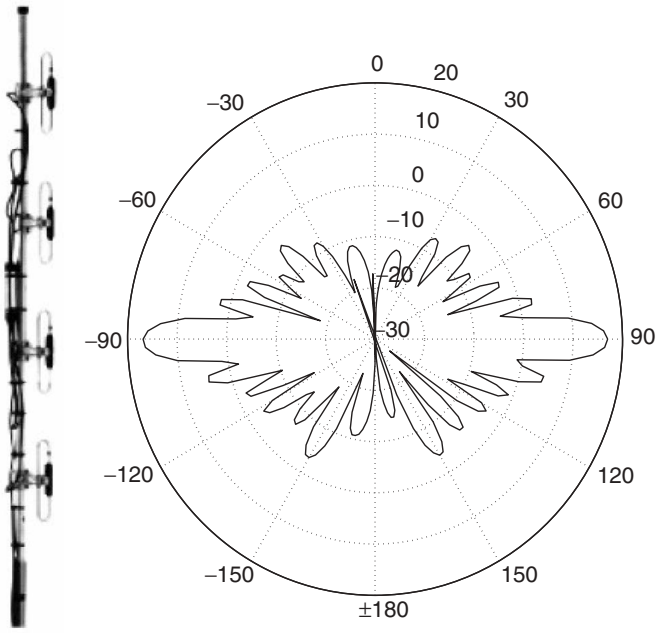


Figure 24.12 Elevation pattern for omnidirectional collinear antenna (courtesy of Jaybeam Wireless Ltd.) and typical elevation radiation pattern: radial axis is gain in dBi. This consists of a set of vertical folded dipoles, cophased by an array of coaxial cables of appropriate lengths to produce the correct phase relationships to synthesize the desired radiation pattern from a single feed.

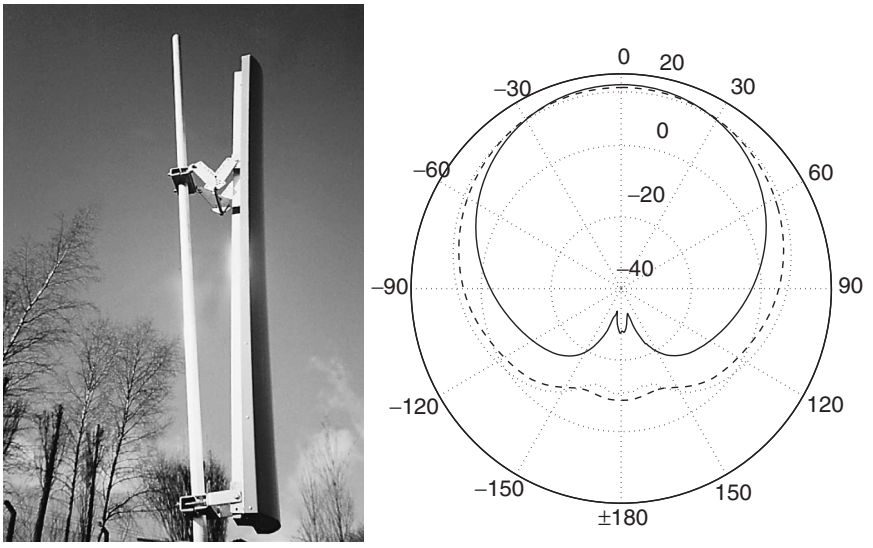


Figure 24.13 Typical macrocell sector antenna and example radiation patterns in azimuth; 3-dB beamwidths are 60° (___), 85° (- - -), and 120° (· · ·).

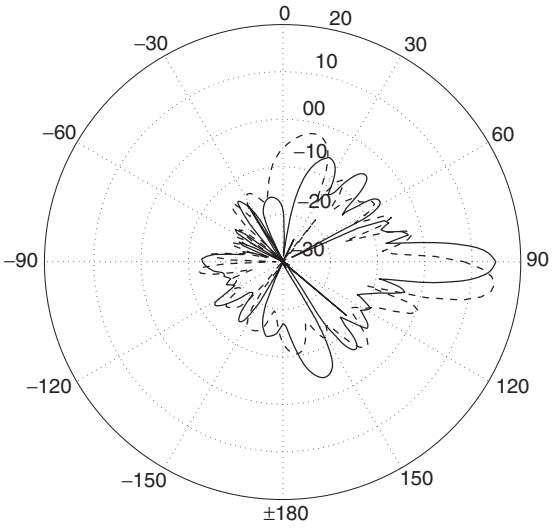


Figure 24.14 Effect of varying electrical downtilt: 0° (___) and 6° (---).

chosen to produce a downward-pointing pattern, with the antenna axis maintained vertical. Electrical downtilt is usually preferred as it can produce relatively even coverage in the azimuth plane. Example elevation patterns produced using electrical downtilt are shown in Figure 24.14.

The combination of a downtilted radiation pattern with the macrocell path loss models can have the effect of increasing the effective path loss exponent, as shown in Figure 24.15 and described in more detail in Saunders [1]. This causes the power received

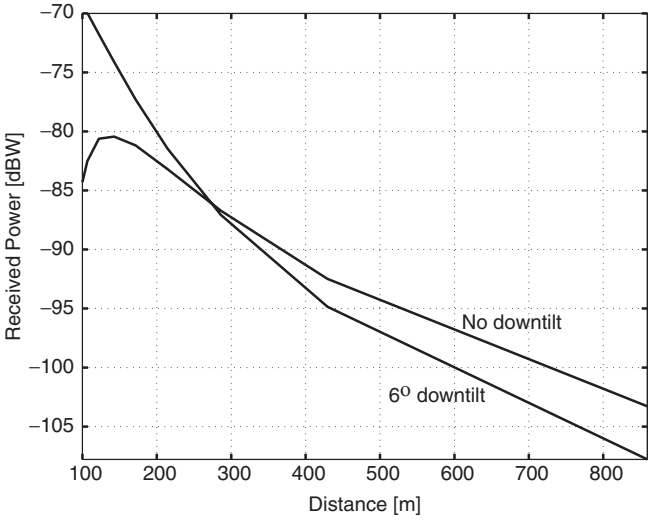


Figure 24.15 Effect of downtilt from the antennas in Figure 24.14 on received signal power: transmitter power 10 W, base station antenna height 15 m, mobile antenna height 1.5 m. Calculated assuming the Okumura–Hata path loss model from Okumura (Ref. 27).

from the base station to fall off more abruptly at the edge of the cell, reducing the impact of interference and permitting the available spectrum to be reused more efficiently.

24.3.2 Macrocell Antenna Design

Modern macrocell antennas usually achieve the desired radiation pattern by creating an array of individual elements in the horizontal and vertical directions, built together into a single panel antenna. The array is typically divided into several subarrays. The array elements are fed via a feed network, which divides power from the feed to excite the elements with differing amplitudes and phases to produce the desired pattern.

The overall elevation pattern,$G(\theta)$, of such an array is given by

$$G(\theta) = g_0(\theta) \sum_{n=1}^{N/M} \sum_{m=1}^{M} I_{nm} \times \exp(j\phi_{nm}) \times \exp(jkd_{nm} \sin\theta) \times \exp(-jkd\phi_r) \quad (24.16)$$

where $g_0(\theta)$ is the radiation pattern of an individual array element (or subarray), I_{nm} and ϕ_{nm} are the amplitude and phase of the excitation of the array element numbered $m + (n-1)M$, $d_{nm} = d(m + (n-1)M)$ (in meters) is the distance along the array of the n,m element, M is the number of element rows in a subarray, m is the row number, N is the total number of rows, n is the subarray number, and

$$\phi_r = m \sin\theta_{\text{sub}} + (n-1)M \sin\theta_{\text{tilt}} \quad (24.17)$$

Here θ_{sub} is the wavefront downtilt associated with a subarray and θ_{tilt} is the desired electrical downtilt of the whole antenna.

The excitation phases and amplitudes are chosen to maximize gain in the desired direction and to minimize sidelobes away from the mainlobe, particularly in directions that will produce interference to neighboring cells. The selection of the excitation coefficients follows generic array principles. In the vertical plane, it is important to minimize sidelobes above the mainlobe, and to ensure that the first null below the mainlobe is filled to avoid the presence of a coverage hole close to the base station. The patterns in Figure 24.14 are good examples.

Since, as illustrated in the previous section, downtilt is such a critical parameter for optimizing the coverage area of a network, it is often desirable to vary the downtilt as the network evolves from providing wide-area coverage to providing high capacity over a limited area. This can be achieved by varying the element feed phases in an appropriate network. For example, the antenna illustrated in Figure 24.16 allows multiple operators to share the same antenna, while being able to independently select from several downtilts according to the needs of their own network.

Array elements can be composed of dipoles operating over corner reflectors over a corner reflector. It is more common in modern antennas, however, to use patch antennas to reduce the antenna panel thickness. These are often created with metal plates suspended over a ground plane rather than printed on a dielectric to maximize efficiency and to avoid arcing arising from the high RF voltages that can develop in high power macrocells. Another important practical consideration in antenna design is to minimize the creation of passive intermodulation products (PIMs). These arise from nonlinearities in the antenna structure, which create spurious transmission products at frequencies that may be far removed from the input frequency. They occur due to rectification of voltages at junctions between dissimilar metals or at locations where metal corrosion has occurred, so the

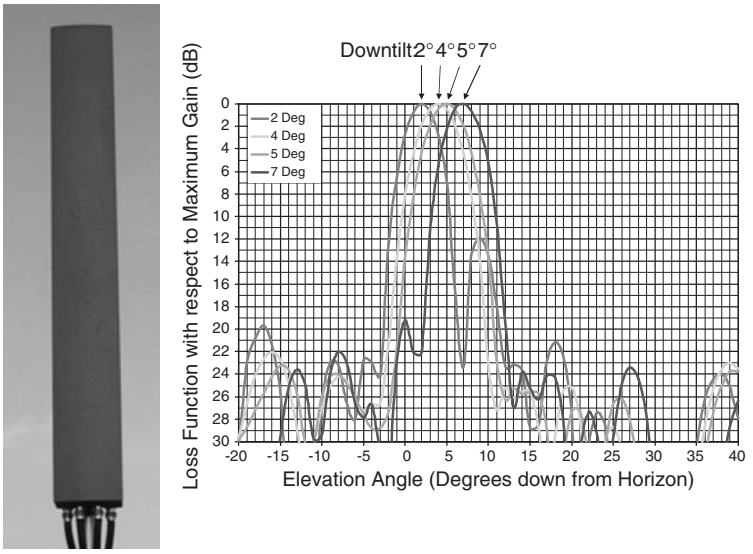


Figure 24.16 A multioperator antenna with independentlyselectable electrical downtilt. (Courtesy of Quintel Ltd.)

choice of metals and the bonding arrangements at junctions must be carefully considered to minimize such issues.

24.3.3 Macrocell Antenna Diversity

Figure 24.17 shows three typical contemporary macrocell antenna installations. The first is a three-sector system, where each sector consists of two panels arranged to provide spatial diversity to overcome narrowband fading. A more compact arrangement is produced by using polarization diversity, where each panel provides two orthogonally polarized outputs. This is typically achieved by dual orthogonal feeds to patch antenna elements. The third case shows how several panel antennas can be mounted onto a building for increased height and reduced visual impact. Figure 24.18 shows some more visionary examples of how macrocell masts might be designed to be a more integral part of the built environment.

24.3.4 Microcell Antennas

The large number of individual rays that can contribute to microcell propagation, as discussed in Saunders [1], make it clear that the cell shape is not determined directly by the radiation pattern of the base station antennas, since each of the rays will emerge with a different power. Nevertheless, it is still important to ensure that power is radiated in generally the right directions so as to excite desirable multipath modes (usually with lobes pointing along streets) and to avoid wasting power in the vertical direction.

In determining the practical antenna pattern, the interactions between the antenna and its immediate surroundings are also very important. These objects may include walls and signs, which may often be within the near field of the antenna, so that accurate prediction and analysis of these effects requires detailed electromagnetic analysis using techniques such as the finite-difference time-domain method [28].

Figure 24.17 Typical macrocell installations.

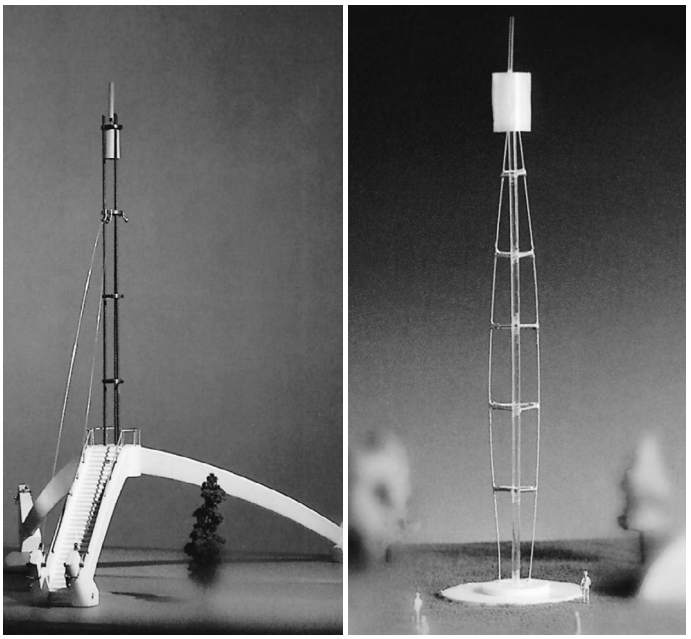


Figure 24.18 Conceptual mast designs for macrocells.

Some typical examples of practical microcell antennas are shown in Figure 24.19. All are designed to be mounted on building walls and to radiate in both directions along the streets they are serving. An alternative approach is to use a directional antenna such as a Yagi–Uda antenna (Figure 24.20), which may help with minimizing interference to other cells. Directional antennas are also useful for containing cell coverage along roads that are not lined with buildings in a sufficiently regular pattern.

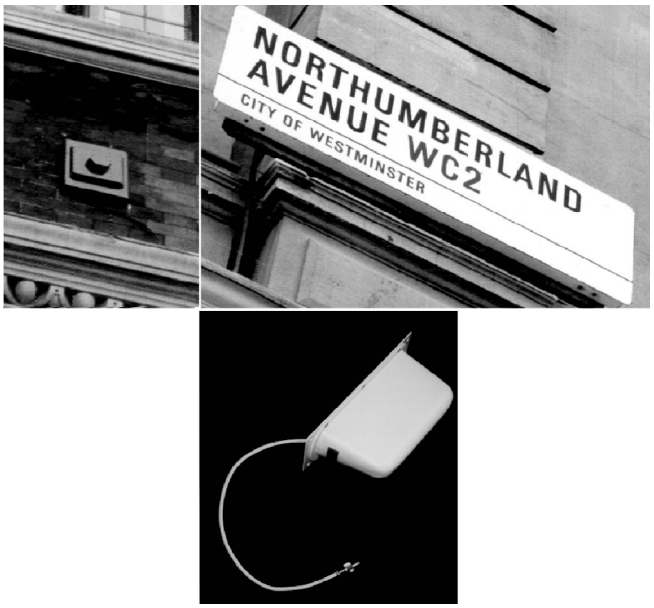


Figure 24.19 Examples of typical microcell antennas.

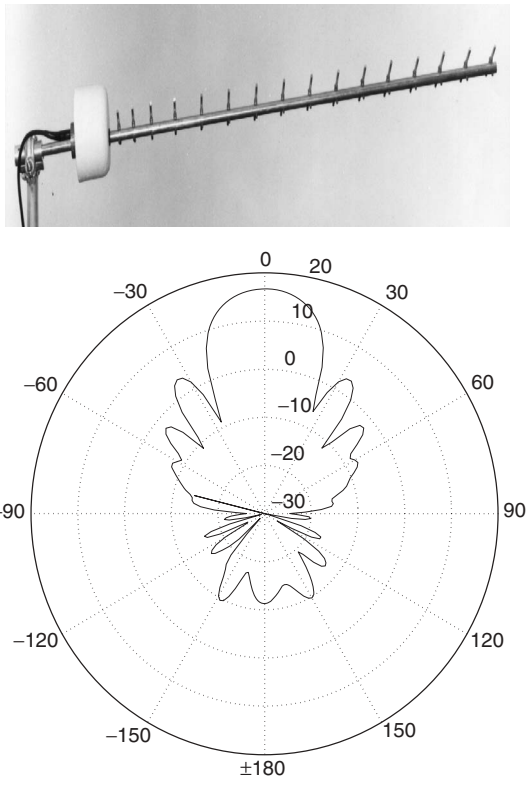


Figure 24.20 Yagi antenna for 900 MHz and its radiation pattern (radial scale in dBi).

It is common practice to use microcell antennas for either outdoor or indoor environments, and often antenna manufacturers do not distinguish between these applications. However, although some microcell antennas can be used indoors, there are other types that due to their size and construction would be aesthetically inappropriate for indoor use. Yagi–Uda, shrouded omnidirectional, ceiling-mount, and monopole antennas are often employed.

24.3.5 Picocell Antennas

A practical ceiling-mounted antenna for indoor coverage at 900 MHz is shown in Figure 24.21. Particular requirements of indoor antennas are very wide beamwidth, consistent with a discrete appearance, so this particular antenna has been designed to look similar to a smoke detector. Linear polarization is currently used almost universally for indoor communications, but there are potential benefits in the use of circular polarization. This has been shown to substantially reduce fade depth and RMS delay spread due to the rejection of odd-order reflections [29] as well as reducing polarization mismatch loss. Similarly, reduction in antenna beamwidth has been shown to substantially reduce the delay spread in line-of-sight situations, but this effect must be traded against the difficulty of providing a reasonably uniform coverage area.

Increasingly, indoor antennas and the associated feed powers also typically have to be compliant with specific requirements on radiated power density and specific absorption rates as described in Section 24.2.8. It is also increasingly desirable to achieve a high level of integration between the systems and technologies deployed by different operators so wideband and multiband indoor antennas are increasingly of interest, providing, for example, wireless local area network (WLAN), second generation (2G), and third generation (3G) technologies in a single antenna housing.

Printed antennas, including microstrip patches, are attractive for indoor antenna designs, with wire antennas being more useful at lower frequencies. Biconical antennas have been proposed for millimeter-wave systems, giving good uniformity of coverage [30].

An issue that has become more important for picocell antennas is that of finding gain values everywhere in space. When performing propagation predictions in outdoor

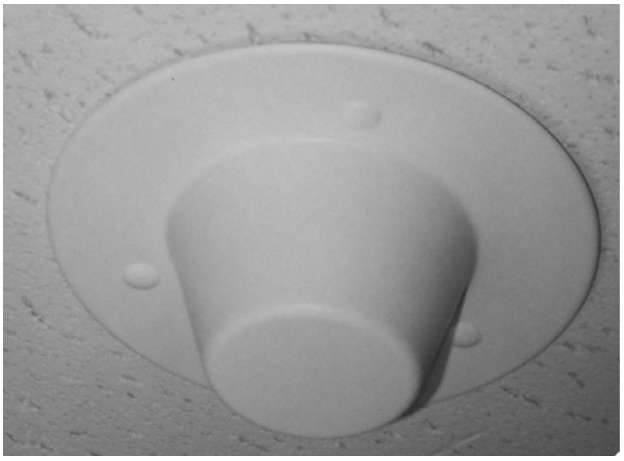


Figure 24.21 Typical ceiling-mounted indoor antenna.

environments, using the models described in Saunders [1], distances between base stations and mobiles are large compared to the base station antenna height, and therefore the signal would be estimated for radiation angles very near to the base station antenna horizontal plane where the manufacturer typically specifies the antenna radiation pattern. However, for picocells, elevation angles typically span the whole range, and hence simple extrapolation methods will lead to unacceptable prediction errors.

A method to overcome such inaccuracies has been proposed by Gil [31] and has been found to give better results. The *angular distance weight model* suggests that the generalized gain $G(\theta, \phi)$ in any direction $P(\theta, \phi)$ is obtained from the previous ones by weighting them with the relative angular distances between the direction of interest and the horizontal (θ_2) and the vertical (θ_1, ϕ_1, ϕ_2) planes, that is, the four points on the sphere closest to the point of interest in Figure 24.22. The basic idea of the model is that the weight by which the value of the gain on a given radiation plane multiplied is inversely proportional to the angular distance, so that the closer the direction of interest is to the given radiation plane, the higher the weight. Therefore continuity of the extrapolation is ensured on each plane.

The final formulation for the angular distance weight model is given by

$$G(\theta, \phi) = \frac{[\phi_1 G_{\phi 2} + \phi_2 G_{\phi 1}]\dfrac{\theta_1 \theta_2}{(\theta_1 + \theta_2)^2} + [\theta_1 G_{\theta 2} + \theta_2 G_{\theta 1}]\dfrac{\phi_1 \phi_2}{(\phi_1 + \phi_2)^2}}{[\phi_1 + \phi_2]\dfrac{\theta_1 \theta_2}{(\theta_1 + \theta_2)^2} + [\theta_1 + \theta_2]\dfrac{\phi_1 \phi_2}{(\phi_1 + \phi_2)^2}} \tag{24.18}$$

where all the angles and gain values are defined in Figure 24.22. This method is recommended as an alternative for pattern extrapolation for indoor environments, because when compared with anechoic chamber measurements taken at various planes, it shows an improvement in accuracy of around 2.5 dB of standard deviation of error.

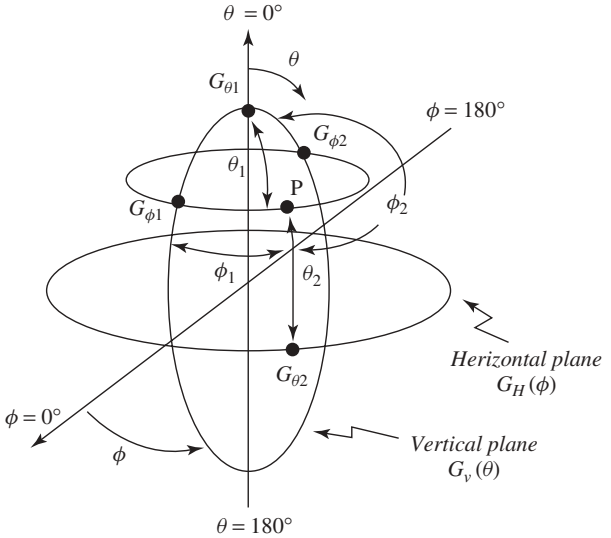


Figure 24.22 Angular distance weight model. (From Ref. 31.)

24.3.6 Antennas for Wireless Local Area Networks (WLANs)

The vast majority of WLANs can be found operating at two frequency bands: the 2.4-GHz ISM band (IEEE 802.11b and 802.11 g standards) and the 5.4-GHz band (IEEE 802.11a standard), with maximum data rates from 11 Mbps (IEEE 802.11b standard) up through to 54 Mbps (IEEE 802.11g/a) and up to over 100 Mbps (IEEE 802.11n, operating in either frequency band).

Spatial diversity is often employed in WLAN access points to overcome multipath fading effects and combine the various replicas of the received signal coherently, achieving substantial spatial diversity gain. Indeed, in the IEEE 802.11n standard, multiple antennas are an absolute requirement to achieve high data rates. Omnidirectional antennas are preferred for some applications, but this depends on whether uniform coverage is required, that is, if the access point and antennas are located in the middle of a room.

Some WLAN access points have integrated antennas, which are often microstrip elements, designed to provide coverage underneath the access point, in an "umbrella" fashion. Floor penetration is sometimes difficult to achieve, especially at the relatively low transmit powers used in access points (50–200 mW EIRP depending on the regulatory regime in the country of use).

When coverage enhancement is required, especially for corridors, tunnels, or to connect two buildings, directional antennas with narrow beamwidth are employed. In this case, parabolic reflectors, Yagi–Uda antennas, and phased-array panels are often used. As the number of channels that can be used is very limited (only three nonoverlapping channels in the 2.4-GHz band in the many countries where 11 or 12 channels are available), interference management and sectorization (also known as zoning for indoor systems) are also important, and hence stringent directional requirements must be enforced to maximize system performance. Such high gain antennas will usually increase effective transmit power beyond the regulatory limits, so transmit power from the access point should be reduced pro-rata; the gain is still effective in increasing the range at the receiver, however.

24.4 SUMMARY

Antennas are the transducer between EM waves carried in a transmission line and coupled into space. The effectiveness of this process depends to a great extent on the efficiency of this coupling, and therefore antenna design becomes essential if system performance is to be maximized. In this chapter, the art of antenna design for mobile systems has been examined, and aspects related to practical antenna configurations have been analyzed and presented. As new wireless services become a reality, antenna design techniques should encompass such improvements, to guarantee that the system design loop can be closed. Stringent requirements such as low cost, small size, and simplicity will still dominate the market requirements and user acceptance for many years to come, and hence formidable and exciting antenna research opportunities are envisaged for the near future.

Both mobile terrestrial and satellite systems increasingly depend on mobile terminal antenna performance to overcome various channel impairments, such as narrowband and wideband multipath fading effects. Diversity and smart antenna techniques can be applied to enhance system performance, at both the base station and the mobile terminal, given appropriate antenna design following the principles in this chapter.

REFERENCES

1. S. R. Saunders and A. Aragon-Zavala, *Antennas and Propagation for Wireless Communication Systems*, 2nd ed., John Wiley & Sons, Hoboken, NJ, 2007.
2. W. L. Stutzman and G. A. Thiele, *Antenna Theory and Design*, 2nd ed., John Wiley & Sons, Hoboken, NJ, 1998.
3. R. Vaughan and J. B. Andersen, *Channels, Propagation and Antennas for Mobile Communications*, Institution of Electrical Engineers, London, 2003.
4. K. Fujimoto and J. R. James (Ed.), *Mobile Antenna Systems Handbook*, 2nd ed., Artech House, London, 2000.
5. C. A. Balanis (Ed.), *Modern Antenna Handbook*, John Wiley & Sons, Hoboken, NJ, 2008.
6. J. S. McLean, A re-examination of the fundamental limits on the radiation Q of electrically small antennas, *IEEE Trans. Antennas Propag.*, Vol. 44, No. 5, 1996.
7. J. R. James and P. S. Hall, *Handbook of Microstrip Antennas*, Peter Peregrinus, London, 1989.
8. T. Taga, Analysis for mean effective gain of mobile antennas in land mobile radio environments, *IEEE Trans. Vehicular Technol.*., Vol. 39, No. 2, pp. 117–131, 1990.
9. P. Suvannapattana and S. R. Saunders, Satellite and terrestrial mobile handheld antenna interactions with the human head, *Proc. IEE Microwaves Antennas Propag.*, Vol. 146, No. 5, pp. 305–310, 1999.
10. M. H. Repacholi and E. Cardis, Criteria for EMF health risk assessment, *Radiat. Prot. Dosim.*, Vol. 72, pp. 305–312, 1997.
11. S. Gabriel, R.W. Lau, and C. Gabriel, The dielectric properties of biological tissues: measurement in the frequency range 10 Hz to 20 GHz, *Phys. Med. Biol.*, Vol. 41, No. 11, pp. 2251–2269, 1996.
12. IEEE SCC34/SC2 Institute of Electrical and Electronics Engineers, *IEEE Recommended Practice for Determining the Spatial-Peak Specific Absorption Rate (SAR) in the Human Body Due to Wireless Communications Devices: Experimental Techniques*, 2000.
13. TC211 WGMBS European Committee for Electrotechnical Standardisation, *Basic Standard for the Measurement of Specific Absorption Rate Related to Human Exposure to Electromagnetic Fields from Mobile Phones* (300 MHz–3 GHz), 2000.
14. P. D. Sewell, D. P. Rodohan and S. R. Saunders, Comparison of analytical and parallel FDTD models of antenna-head interactions, in *IEE International Conference on Antennas and Propagation, ICAP '95*, IEE Conf. Publ. No. 407, pp. 67–71, 1995.
15. International Commission on Non-ionizing Radiation Protection, Guidelines for limiting exposure to time-varying electric, magnetic, and electromagnetic fields (up to 300 GHz), *Health Phys.*, Vol. 75, No. 4, pp. 494–522, 1998.
16. IEEE C95. 1–1991 *IEEE Standard for Safety Levels with Respect to Human Exposure to Radio Frequency Electromagnetic Fields, 3 kHz to 300 GHz*, 1991.
17. European Council recommendation on the limitation of exposure of the general public to electromagnetic fields 0Hz–300 GHz, *Eur. J.*, Vol. 197, 30 July 1999.
18. A. A. Agius, S. R. Saunders, and B. G. Evans, The design of specifications for satellite PCN handheld antennas, *47th IEEE International Vehicular Technology Conference*, Phoenix, Arizona, 5–7 May 1997.
19. C. C. Kilgus, Multielement, fractional turn helices, *IEEE Trans. Antennas Propag.,* Vol. 16, pp. 499–500, 1968.
20. C. C. Kilgus, Resonant quadrifilar helix, *IEEE Trans. Antennas Propag.,* Vol. 17, pp. 349–351, 1969.

21. A. A. Agius and S. R. Saunders, Effective statistical G/T (ESGUT): a parameter describing the performance for non-GEO satellite systems, *Electron. Lett.*, Vol. 34, No. (19), pp. 1814–1816, 1998.
22. D. K. C. Chew and S. R. Saunders, Meander line technique for size reduction of quadrifilar helix antenna, *IEEE Antennas Wireless Propag. Lett.*, Vol. 1, No. 5, pp. 109–111, 2002.
23. A. A. Agius and S. R. Saunders, Adaptive multifilar antenna, U.K. Patent No. WO9941803, August 1999.
24. S. M. Leach, A. A. Agius, and S. R. Saunders, Intelligent quadrifilar helix antenna, *Proc. IEE Microwaves Antennas Propag.*, Vol. 147, No. 3, pp. 219–223, 2000.
25. S. M. Leach, A. A. Agius, S. Stavrou, and S. R. Saunders, Diversity performance of the intelligent quadrifilar helix antenna in mobile satellite systems, *Proc. IEE Microwaves Antennas Propag.*, Vol. 147, No. 4, pp. 305–310, August 2000.
26. T. W. C. Brown, K. C. D. Chew, and S. R. Saunders, Analysis of the diversity potential of an intelligent quadrifilar helix antenna, in *Proceedings of the Twelfth International Conference on Antennas and Propagation—ICAP2003*, 2003, Vol. 1, pp. 194–198.
27. Y. Okumura, E. Ohmori, T. Kawano, and K. Fukuda, Field strength and its variability in VHF and UHF land mobile radio service, *Rev. Electr. Commun. Lab.*, Vol. 16, pp. 825–873, 1968.
28. K. S. Yee, Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media, *IEEE Trans. Antennas Propag.*, Vol. 14, pp. 302–307, 1966.
29. A. Kajiwara, Line-of-sight indoor radio communication using circular polarized waves, *IEEE Trans. Vehicular Technol.*, Vol. 44, No. 3, pp. 487–493, 1995.
30. P. F. M. Smulders and A. G. Wagemans, Millimetre-wave biconical horn antennas for near uniform coverage in indoor picocells, *Electron. Lett.*, Vol. 28, No. 7, pp. 679–681, 1992.
31. F. Gil, A. R. Claro, J. M. Ferreira, C. Pardelinha, and L. M. Correia, A 3-D extrapolation model for base station antennas' radiation patterns, in *IEEE Vehicular Technology Conference*, September 1999, Vol. 3, pp. 1341–1345.

CHAPTER 25

Antenna Array Technologies for Advanced Wireless Systems

MAGDY F. ISKANDER, WAYNE KIM, JODIE BELL, NURI CELIK, and ZHENGQING YUN

25.1 OVERVIEW OF THE PHASED ARRAY ANTENNA (PAA) TECHNOLOGY

Most current scanning antennas utilize mechanical scanning, relying on gimbals, motors, and gear arrangements to physically rotate and/or elevate the antenna. This technique is slow, gravity sensitive, and susceptible to shock and mechanical failure. For airborne radar applications, high speed multiple-target tracking cannot be effectively done by this technique. Electronic scanning using phased array antenna systems overcomes these limitations and allows significantly faster scanning, requires no mechanical/physical rotation of the antennas, and lends itself to multiple-target tracking.

It is known that large antenna systems are difficult to mechanically scan rapidly, and because of this electronically scanned arrays have been developed and often used in modern systems. These arrays, which may have several thousand elements, are scanned by incorporating digital and/or analog phase shifters in each feed line. The electronic control of these phase shifters to produce incremental changes in phase allows very fast scanning of the beam direction in space. The applications for large phased arrays are mostly in advanced radar systems and in radio astronomy. Smaller phased arrays and beam-forming arrays are used as feed systems to illuminate a reflector in satellite communication systems when it is necessary to provide several spot beams and/or wide-angle coverage beams from a one-antenna system [1].

One benefit of phased array antennas in airborne applications is its low radar cross section (RCS) [1]. The phased array antenna has the ability of redirecting enemy radar by electronically scanning the reflected response away from the threat, thereby reducing the overall signature of the aircraft. Another benefit includes advanced beam-forming capabilities. The phased array also has the capability of forming beam nulls or "antijamming" toward the direction of undesired signals in conjunction with maintaining multiple target tracking.

Phased array antenna architectures are reliable systems. The performance of the array degrades slowly if several elements malfunction. The overall system requires less structural intrusion based on the fact that there are no moving parts and it is easier to fit

Modern Antenna Handbook. Edited by Constantine A. Balanis

in confined spaces. Based on these capabilities, electronic scanning systems have many advantages over mechanically scanned architectures. In the following subsections we describe various types of phased array antennas, then include some examples of the antenna elements used in these designs, and finally briefly describe some of the commonly used feed structures.

25.1.1 Examples of Various Phased Array Antenna Designs

25.1.1.1 Phase–Phase Steered Arrays Phase–phase steered arrays are systems that steer the main beam in both azimuth and elevation using phase shifters. The most popular passive phase–phase steered array is the Patriot System [2, 3] illustrated in Figure 25.1.

This array employs over a million phase shifters and has seen very large production runs. Another phase–phase steered array is the ARTHUR Swedish C-band 30-km range, artillery locating system, which uses a traveling-wave tube (TWT) transmitter [4]. Another TWT system is the ship-based European Multifunction Phased Array Radar (EMPAR) C-band passive array [5] illustrated in Figure 25.2. This TWT radar, built by Alenia (Italy) in collaboration with Marconi Co. (U.K.), has a search range of up to 180 km and scans $\pm 45^\circ$ in azimuth and $\pm 60^\circ$ in elevation.

25.1.1.2 Active Phased Array Antennas The first solid-state phase–phase steered array employing bipolar transistors for power amplification is the UHF PAVE PAWS illustrated in Figure 25.3 and the UHF BMEWS radars [2].

Both radar systems employ the same transmit/receive (T/R) modules as well as radiating elements. The radar is used primarily to detect and track sea-launched and intercontinental ballistic missiles. The system also has a secondary mission of Earth-orbiting satellite detection and tracking.

Figure 25.1 The Patriot phased array system. (From Refs. 2, 3.)

Figure 25.2 The EMPAR phased array antenna system. (From Ref. 5.)

Figure 25.3 The PAVE PAWS active phased array antenna system. (From Ref. 2.)

Countries around the world have developed active phased arrays using bipolar transistors. There is the Swedish Erieye airborne early warning radar system that uses an S-band solid-state phased array placed in a dorsal fin over the top of the aircraft. The system has approximately 200 modules, which are shared with the array faces on both sides of the dorsal fin. Another system is the Israeli Phalcon airborne early warning system that uses four solid-state L-band phased arrays with each antenna having approximately 700 T/R modules [4]. The antennas are placed on the left and right sides of the aircraft forward of the wing, and on the underside toward the nose and tail of the aircraft. Israel has also

developed a theater missile defense system that is capable of detection ranges of hundreds of kilometers. This radar also uses an L-band solid-state active phased array [5].

Monolithic microwave integrated circuit (MMIC) phased array systems have been built where the solid-state T/R modules are developed "on-chip" for low cost production. One such system is the theater high altitude area defense (THAAD; formerly called GBR) X-band ground-based radar, which has 25,344 MMIC T/R modules and radiating elements [6]. This is the largest active phased array built in the world in terms of number of modules. It uses subarray time delay steering in order to prevent pulse distortion for its wideband waveforms. Approximately 65,000 MMIC T/R modules have been built for these systems. This program demonstrated that MMIC T/R modules could be manufactured for less than $1000 each at the end of the production run [7].

25.1.1.3 Airborne Phased Array Antennas Figure 25.4 illustrates a phased array antenna system—the Airborne Warning and Control System (AWACS) [1]. From an altitude of 30,000 ft, the radar can detect low altitude and sea-surface targets out to 215 nautical miles (nmi), coaltitude targets out to 430 nmi, and targets beyond the horizon at still greater ranges. The phased array antenna operates at S-band frequencies (typically 3 GHz), is 24 ft $\times$ 5 ft, and is housed in a rotdome that rotates at 6 rpm. The antenna has an extremely narrow beamwidth and is amplitude weighted for sidelobe reduction [1].

Figure 25.5 illustrates the APG-77 active phased array antenna system employed in the F-22 Raptor aircraft. The active phased array radar contains 1000 finger-sized transmit/receive modules. Each module weights 15 g and has a power output of over 4 W. The active phased array provides the frequency agility, low radar cross section, and wide bandwidth required for the fighter's air dominance mission [1].

25.1.2 Examples of Antenna Elements

The most commonly used antennas for large phased array antenna systems are dipoles, slots, open-ended waveguides (or small horns), spirals, microstrip disk or patch elements, and lens systems. The selection of the antenna for a particular application must be based on the following considerations [8]:

1. The required area of the element is small enough to fit within the allowable element spacing and lattice without the formation of grating lobes. In general, this limits the element to an area of little more than $\lambda^2/4$.

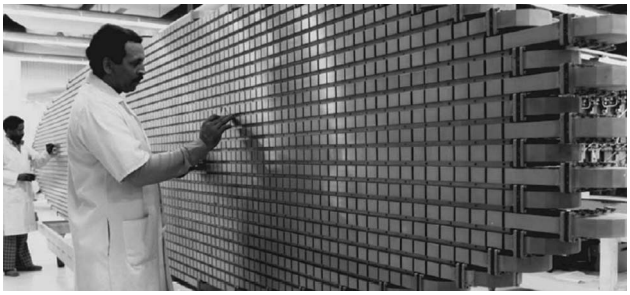


Figure 25.4 Photograph of Airborne Warning and Control System (AWACS) phased array antenna mounted on the E-3 aircraft. (From Ref. 1.)

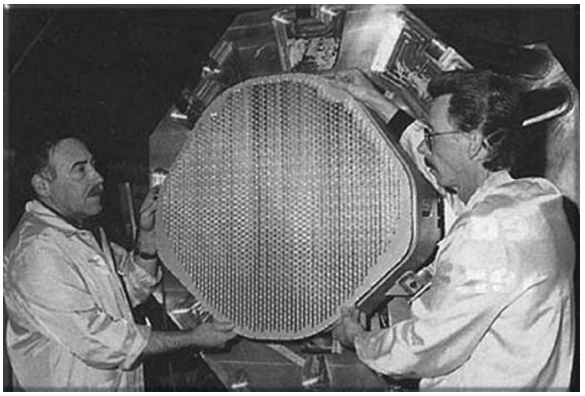


Figure 25.5 Photograph of airborne APG-77 active phased array antenna mounted on the F-22 Raptor aircraft. (From Ref. 1.)

2. The element pattern of the antenna provides the appropriate aperture matching over the required scan coverage.
3. The polarization and power-handling capabilities (both peak and average power) meet the overall system requirements.
4. The physical construction of the radiator must be able to withstand environmental requirements such as thermal, shock, and vibration requirements.

Fixed beam broadside arrays may employ low or moderate gain elements, but most arrays employ low gain elements to minimize the effects of grating and quantization lobes [9]. Moderate gain elements such as the spiral, helix, log-periodic, Yagi–Uda, horn, and backfire generally do not have unique array properties [9, 10].

In practice, the radiation patterns of real antennas are nonisotropic, and the impedance values of the antennas vary as a function of scan angle caused by the mutual coupling between radiators. The pattern of an element in an array configuration is much different from the pattern of an isolated element in amplitude, phase, and in some cases polarization as well.

25.1.3 Examples of Feed Networks

Since the impedance and the radiation pattern of a particular antenna in an array system is determined by the array geometry and environment, the antenna element may be chosen to best suit the feed system and the physical requirements of the antenna [8, 9]. For example, if the antenna is fed from a microstrip phase shifter, a microstrip dipole would be the best choice for the antenna. If, on the other hand, the antenna is fed from a waveguide phase shifter, an open-ended waveguide would be the logical choice for the antenna. For lower frequency applications, a coaxial phase shifter may be employed to feed a dipole radiator, while at higher frequencies, open-ended waveguides and slots are frequently used.

The problem is further complicated in the case of nonuniformly excited elements because of the need to use some form of low loss power-splitting circuit elements to

achieve different amplitude levels at the various antenna elements. It is usually necessary to match each antenna element to the feed structure in order to obtain acceptable performance over a broad frequency range. In typical feed structures the elements are grouped into smaller subgroups or bays according to the symmetry of the array system. In some array applications it may be desirable to use several fixed beams with each covering a particular angular sector. Several beam-forming feed networks have been developed that provide these fixed beam patterns and in the following sections we briefly describe three examples of such networks.

25.1.3.1 Butler Matrix The Butler matrix [8–13] is a multiple beam-forming network that uses a combination of 90° hybrids and fixed-phase shifters to achieve the desired beam pattern. An 8×8 Butler matrix is shown in Figure 25.6. The Butler matrix performs a spatial fast Fourier transform providing 2^n orthogonal beams. These beams are linearly independent combinations of the array element patterns.

The Butler matrix is a simple network that can be implemented in microstrip technology; however, conductor crossovers are required. When used with a linear array, the Butler matrix produces beams that overlap at ~3.9 dB below the beam maxima. A Butler matrix-fed array can cover a sector of up to 360° depending on element patterns and spacing. Each beam can be used by a dedicated transmitter and/or receiver, or a single transmitter and/or receiver can be used, and the appropriate beam can be selected using a radiofrequency (RF) switch. A Butler matrix can also be used to steer the beam of a circular array by exciting the Butler matrix beam ports with amplitude- and phase-weighted inputs followed by a variable uniform phase taper.

25.1.3.2 Blass Matrix The Blass matrix [8, 14] is also a multiple beam-forming network that uses transmission lines and directional couplers to form beams by means of time delays and thus is suitable for broadband operations. Figure 25.7 depicts an example Blass matrix for a three-element array. The Blass matrix may be designed for use with any number of elements. To generate multiple beams, several feed lines are coupled to the branch lines by using directional couplers. Each feed line generates one beam in space. With proper excitation to the input terminals, each coupler leaks off a fraction of the input signal toward the radiating element connected to that junction.

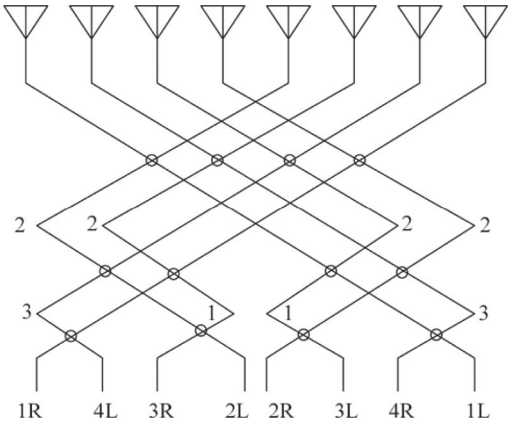


Figure 25.6 An 8×8 Butler matrix feeding an eight-element array. The circles represent 90° hybrids and the numbers represent phase shifts in units of $\pi/8$. (Refs. 8–13.)

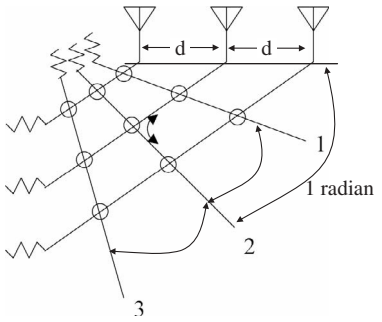


Figure 25.7 Schematic of a Blass matrix where the circles represent directional couplers. (From Refs. 8 and 14.)

The phase shift between the radiating elements is governed by the *tilt* of the feed lines. The branch lines to the radiating elements are circumferential with a particular radius. The primary feed line (port 2) is placed 1 radian from a line parallel to the antenna face and provides equal delays to all elements, resulting in a broadside beam. The other two ports (port 1 and 3) are spaced angularly by multiples of $\Delta\theta$, so that the signal travels a different distance $nd\ \cos(\Delta\theta)$ to reach the nth radiating element. These outer feed lines provide progressive time delays between elements and produce beams that are off broadside. The Blass matrix is lossy due to the resistive terminations. In one recent application [15], a three-element array fed by a Blass matrix was tested for use in an antenna pattern diversity system for a hand-held radio. The matrix was optimized to obtain nearly orthogonal beams.

25.1.3.3 Wullenweber Array A Wullenweber array [8, 16], illustrated in Figure 25.8, is a circular array developed for direction finding at HF. The radiator elements can be either omnidirectional or directional and are oriented radially outward. The array typically consists of 30–100 evenly spaced elements. Approximately one-third of the elements are used at any given time to form a beam that is oriented radially outward from the array. A switching network called a goniometer is used to connect the appropriate elements to the transceiver and may include some amplitude weighting to control the array pattern. Advantages of the Wullenweber array are the ability to scan over 360° with very little change in pattern characteristics. At lower frequencies the Wullenweber array is much smaller than the rhombic antennas that might be employed. Time delays

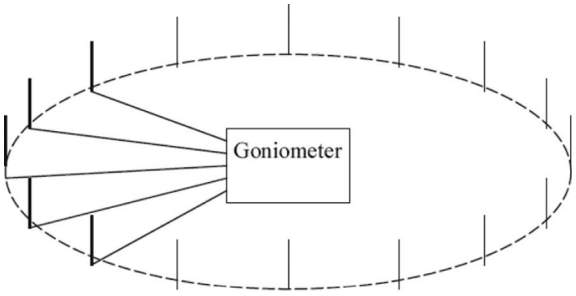


Figure 25.8 Schematic of a Wullenweber array. (From Refs. 8 and 16.)

are used to form beams radial to the array, enabling broadband operation. The bandwidth of a Wullenweber array is limited by the bandwidth and spacing of the elements.

25.1.3.4 Other Fixed Beam-Forming Techniques Fixed beams can also be formed using lens antennas such as the Luneberg or Rotman lenses with multiple feeds. Lenses focus energy radiated by feed antennas that are less directive. Lenses can be made from dielectric materials or implemented as space-fed arrays. Multibeam arrays can be used to feed reflector antennas as well.

25.2 FERROELECTRIC MATERIALS IN PHASED ARRAY ANTENNA SYSTEMS

In this section we describe the different approaches to utilization of ferroelectric materials in the design of high performance phased array antenna systems. The first section is comprised of the integrated ferroelectric materials with radiating elements to form a phased array system, while the second section is based on the use of ferroelectric materials in phase shifters as stand-alone components.

25.2.1 Examples of Ferroelectric Materials Used in Beam Steering

In this section we briefly describe some of the antenna array designs that are based on integrating the array elements with the phase shifters. Focus is placed on approaches that use ferroelectric materials for the phase shifters or to directly achieve the desired beam-steering capability.

25.2.1.1 Ferroelectric Lens Phased Array A ferroelectric lens phased array system with row–column beam-steering capabilities is illustrated in Figure 25.9 [17]. The lens system consists of columns of ferroelectric placed between conducting plates. A dc

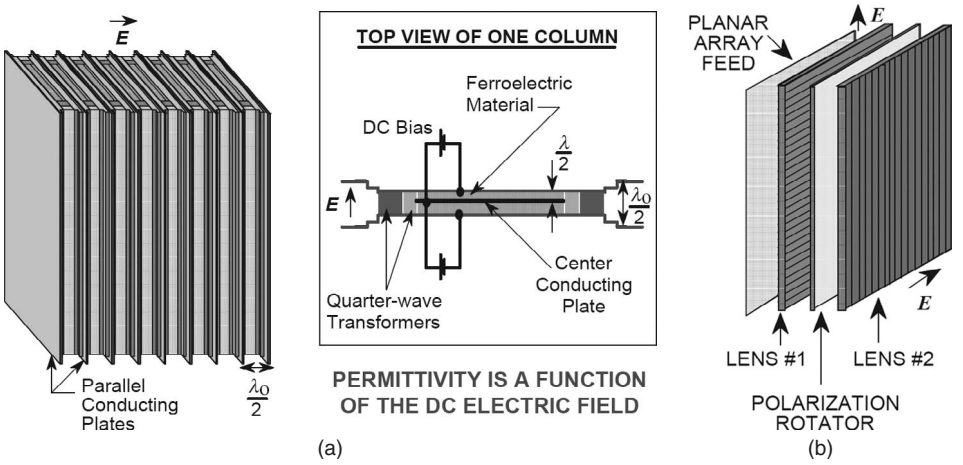


Figure 25.9 Ferroelectric lens antenna developed at the Navy Research Laboratories (NRL) [17]: (a) details of the ferroelectric lens and (b) two lens system employing azimuth and elevation scanning.

voltage is applied across each pair of plates. The permittivity of the ferroelectric material between a pair of plates depends on the dc voltage applied across that pair of plates. As a result, the phase of the electromagnetic signal passing through a ferroelectric column will depend on this dc voltage. Consequently, by applying appropriate dc voltages across the ferroelectric columns, one can create a phase gradient in the horizontal direction for the signal leaving the first lens and thus can scan the beam in azimuth. A second such lens rotated 90° would steer the beam in elevation as illustrated in Figure 25.9b. For a ferroelectric lens, it is necessary that the electric field be linearly polarized and perpendicular to the conducting plates. Hence the first ferroelectric lens requires a horizontally polarized signal while the second lens requires a vertically polarized signal. A 90° polarization rotator is used between the lenses in order to have the horizontally polarized signal out of the first lens become vertically polarized at the input to the second lens.

25.2.1.2 Beam-Steerable Reflectarrays For high data rate, broadband SATCOM applications, a low cost ferroelectric based beam-steerable reflectarray operating at K-band (18–26.5 GHz) and Ka-band (26.5–40 GHz), illustrated in Figure 25.10, was developed by the group at the NASA Research Center [18, 19]. A single high power amplifier feeds RF power to an array panel consisting of microstrip patch antennas. The signal is reradiated from these elements, where beam steering is achieved by adjusting the phase between each element. This is accomplished by modulating the dielectric constant of the ferroelectric film. Another type of phased array antenna that, in this case, incorporates an integrated combination of continuous transverse stub (CTS) radiators along with ferroelectric materials for beam steering is discussed in Section 25.4.

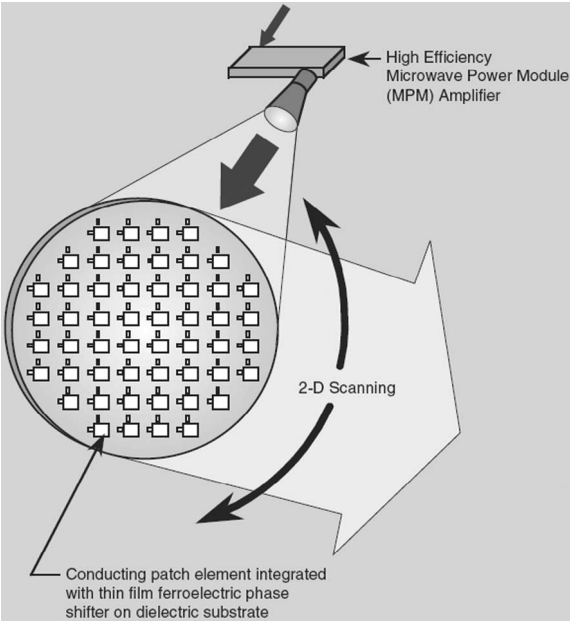


Figure 25.10 Schematic diagram of a reflectarray system based on tunable ferroelectric films for two-dimensional (2D)) beam-scanning capabilities. The array is fed from a high power microwave amplifier. (From Refs. 18 and 19.)

25.2.2 Examples of Phase Shifter Designs Using Ferroelectric Materials Technology

Ferroelectric materials have the potential to overcome all the limitations of microelectromechanical system (MEMS), ferrite, and MMIC true time delay phase shifters. Ferroelectric phase shifters are significantly lower in cost and smaller in size and can also be electronically adjusted in real time for age and environmental stress. A large number of ferroelectric phase shifters can be designed and placed on a single material substrate, thus further reducing the antenna size, weight, and cost.

Several groups have investigated the possibility of implementing phase shifter circuits using barium strontium titanate (BSTO), which has an electric field tunable permittivity. The lack of dielectric dispersion in the "paraelectric" state at frequencies up to $10^{11}-10^{12}$ Hz indicates that the ferroelectric devices are competitive with those based on monolithic type up to submillimeter waves. The response time between permittivity states are on the order of nanoseconds [20]. This makes its application in the design of phased array antennas suitable even for radar-type applications.

The maximum breakdown voltage for the ferroelectric film is >300 MV/m [20]. Therefore, for a 0.5 μm thick substrate, this corresponds to a maximum tolerable voltage of 150 V, and this means that for a 50-Ω transmission line system, the peak power allowable is ($P_{\max} = V_{\text{bk}}^2/2Z_0$) ~ 225 W.

Ferroelectric material typically requires a static electric field of ~0.7–3.0 V/μm to sufficiently change the permittivity. The dc power required to maintain a fixed phase shift is less than a milliwatt due to the high dc resistivity of the ferroelectric film ($\rho \sim 4 \times 10^7 \Omega$m) [21]. The driver energy required to change the properties of the ferroelectric material affects the electrostatic energy and is not dissipated [21]. The following are examples of phase shifter designs using the ferroelectric materials technology, followed by a detailed description of the multidielectric layers coplanar waveguide phase shifter design that has been proposed and developed by our group.

25.2.2.1 Distributed Transmission Line Ferroelectric Phase Shifters

A distributed phase shifter (true time delay) is created by adding tunable reactance to a transmission line. Adjusting the reactance alters the phase velocity of the signal propagating along the line, varying its electrical length and therefore the phase shift. Changing the phase velocity also changes the characteristic impedance of the transmission line, so an impedance mismatch can occur as the circuit is tuned. In general, it should be possible to add both series and shunt tunable reactance to the transmission line to keep an impedance match while achieving the desired tuning; however, a technology for adding tunable series *inductance* has yet to be fully developed. The majority of distributed phase shifters focus on adding tunable shunt capacitances. Ferroelectric varactors, MEMS bridges and switches, and semiconductor diodes are all capable of performing this function. In the majority of cases, the shunt capacitance is added periodically as discrete elements to the transmission line. This capacitance loading makes the distributed phase shifter a periodic structure, with a passband and a stopband. Careful design is necessary to ensure the frequencies of interest fall into the passband, while simultaneously maintaining a high performing, efficient structure.

A ferroelectric phase shifter design based on the distributed transmission line device topology was recently developed [22]. This approach uses periodically loaded tunable BSTO parallel plate capacitors across a coplanar waveguide transmission line. As depicted in Figure 25.11, this approach employs a mixture of the ferroelectric phase

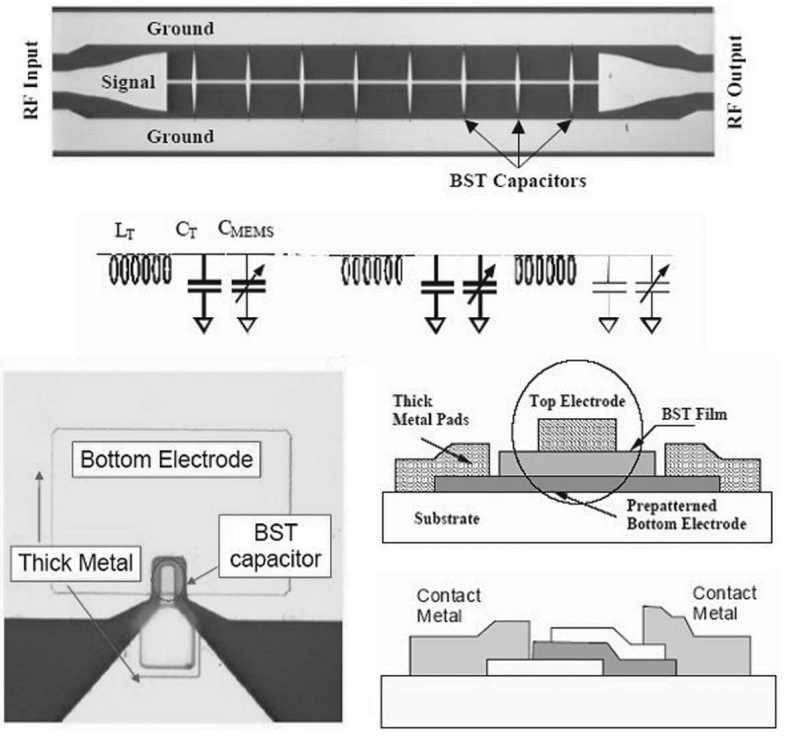


Figure 25.11 Ferroelectric distributed phase shifter and close-up of a single varactor and its equivalent circuit model. (From Ref. 22.)

shifters and MEMS phase shifter technologies. It combines the properties of BSTO at microwave frequencies with the distributed transmission line philosophy of the MEMS phase shifters.

The addition of tunable capacitance alters the effective characteristic impedance, Z_0, and phase velocity, v_{ph}. The addition of the variable capacitance lowers the effective characteristic impedance, and it is therefore necessary that the intrinsic Z_0 of the transmission line be larger than the characteristic impedance of the external circuit in order to maintain an impedance match. A perfect match is not possible under all tuning conditions. The variation in v_{ph} is responsible for the phase shifting behavior of the distributed phase shifter.

One major limitation of this approach is the discontinuities created by the addition of tunable capacitors as reflections result from each of the capacitors distributed along the length of the circuit. As the frequency of the signal approaches a certain value, the phases of the incident and reflected signals interfere destructively, preventing forward propagation. When the signal cannot propagate, the transmission loss increases, and the signal is reflected back toward the source. This frequency is called the Bragg frequency and is defined by

$$f_{\text{Bragg}} = \frac{1}{\pi \Delta l \sqrt{L_0(C_0 + C_{\text{var}})}} \tag{25.1}$$

The Δl parameter represents the spacing between the tuning capacitors and can be adjusted to change the Bragg frequency, independent of the other transmission line parameters. The highest operating frequency of the phase shifter must be significantly below f_{Bragg} to avoid large transmission losses. This relationship limits the high frequency performance of this type of ferroelectric phase shifters.

A loss optimized distributed phase shifter design depends on proper selection of Δl and Z_0. Increasing Δl brings the Bragg frequency closer to the operating frequency and reduces the number of sections required to achieve a given phase shift. Increasing Z_0 lowers C_0 and allows a greater variation of v_{ph}, also reducing the number of sections. This is beneficial if the tunable capacitor is lossy, since fewer are needed in a given design. However, operating closer to the Bragg frequency increases the transmission losses through reflections of the input signal. These conflicting requirements lead to an optimized design that balances the losses, resulting in the lowest loss architecture. As a result, the best design from a loss perspective doesn't necessarily have the shortest length or fewest sections.

25.2.2.2 Transmission Line Ferroelectric Phase Shifter Designs In these true time delay devices, the ferroelectric material (BSTO) forms either the entire microwave substrate on which the conductors are deposited (thick film/bulk crystal) or a fraction of the substrate with thin BSTO film sandwiched between the substrate and the conductors. A microstrip transmission line is shown in Figure 25.12a, [21] and a coplanar waveguide transmission line is shown in Figure 25.12b [23].

These circuits rely on the principle that because part or all of the microwave fields pass through the ferroelectric layer, the phase velocity of waves propagating on these structures can be altered by changing the permittivity of the ferroelectric layer. This approach has the benefits of being less complex compared to the distributed line approach and also allows the possibility of integrating the phase shifter with the antenna elements for a simple compact phased array system. Other benefits include the virtual elimination of the Bragg frequency limitation seen with distributed line phase shifters as described in Section 25.2.2.1. However, one major limitation of the transmission line ferroelectric phase shifter designs is that the conductor losses are unacceptably high due to the high permittivity of the ferroelectric film on which the transmission lines are fabricated.

25.2.2.3 Multi-dielectric Coplanar Phase Shifters The coplanar ferroelectric phase shifter such as the architecture illustrated in Figure 25.12b exhibits high ohmic

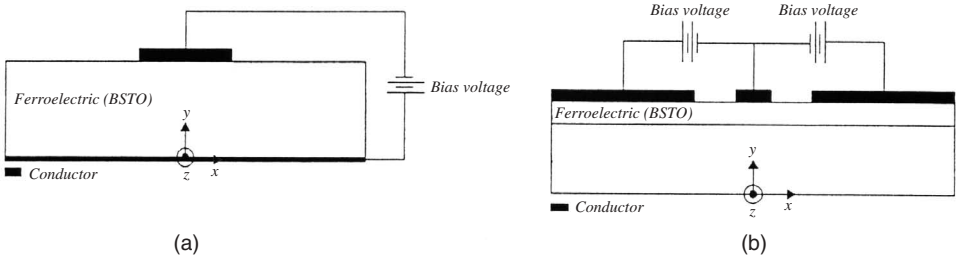


Figure 25.12 (a) A ferroelectric microstrip-based phase shifter (Ref. 21) and (b) a ferroelectric coplanar waveguide microstrip phase shifter. (Ref. 23.)

losses and could be deemed undesirable for practical wireless communication systems, particularly when implemented in phased array antenna systems. A coplanar device possesses low dispersion and good characteristic impedance resiliency and has the architecture that facilitates the integration of circuit elements on the same side of the device [24]. If an improvement could be made to reduce the high insertion losses associated with these transmission line-type ferroelectric phase shifters, the highly anticipated low cost, low profile, and integrated phased array systems may be realized.

Kaydanova [25] reported that, by adding a temperature stable low loss nontunable dielectric layer in between the ferroelectric material and the metallic electrodes associated with interdigitated capacitors, the figure of merit K defined as

$$K = \frac{\varepsilon_{\max} - \varepsilon_{\min}}{\varepsilon_{\max} \tan \delta_{\max}} \tag{25.2}$$

improved by up to 25%. Here $\varepsilon_{\max}$ and $\varepsilon_{\min}$ represent the unbiased and biased permittivity of the ferroelectric material, respectively, and tan $\delta_{\max}$ is the overall loss tangent of the layered structure. Although this approach provided an improvement in insertion loss for interdigitated capacitors, it is not known how it affects the phase shift performance in a coplanar transmission line structure.

In an attempt to examine the trade-offs involved in incorporating a layer of a low loss dielectric in a multidielectric layer coplanar waveguide phase shifter design, our group conducted extensive simulations [26]. Figure 25.13 depicts a multidielectric ferroelectric phase shifter design that employs the implementation of a low loss, nontunable dielectric layer such as silicon dioxide (SiO_2) in between the ferroelectric material and the coplanar electrodes. This design is expected to reduce the attenuation constant and improve the figure of merit (FoM), where the FoM is defined as the amount of phase shift per decibel loss given in units of °/dB. The loss in this case is taken as the attenuation

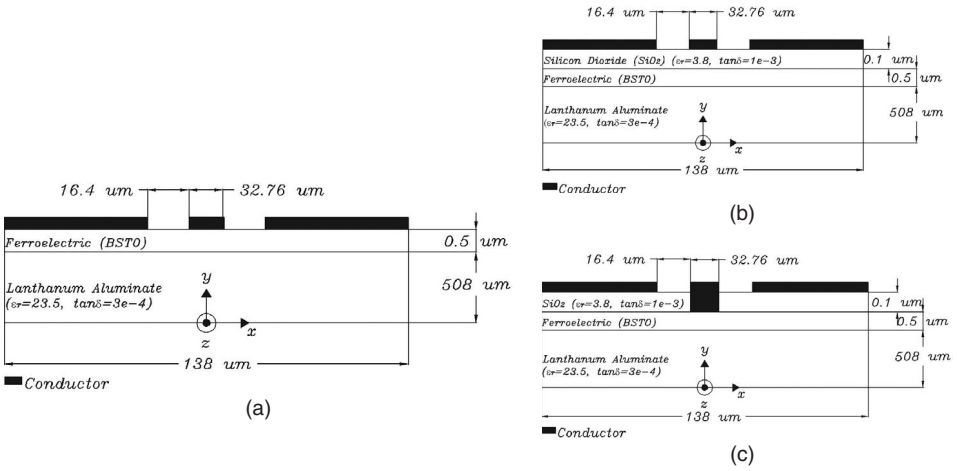


Figure 25.13 Cross section of (a) direct metallized phase shifter, (b) multidielectric ferroelectric coplanar waveguide phase shifter employing very thin, nontunable layer between the ferroelectric film and the coplanar electrodes, and (c) multidielectric design employing a via through the silicon dioxide layer for simple ferroelectric biasing. (Ref. 26.)

constant (dB/mm) for the unbiased ferroelectric condition since this condition possesses the largest permittivity and the higher losses.

The coplanar structure employs barium strontium titanate ($Ba_xSr_{1-x}TiO_3$) with a compositional ratio of $x = 0.6$ over a lanthanum aluminate ($LaAlO_3$) substrate. This composition corresponds to the "paraelectric" state of the material at room temperature ($T \sim 300$ K). The total device width is 138 μm, the ferroelectric layer thickness is 0.5 μm, and the $LaAlO_3$ substrate layer thickness is 508 μm. The relative dielectric constant for the $LaAlO_3$ substrate is $\varepsilon_r = 23.5$ with tan $\delta = 3e\text{-}4$. The tan δ for the $Ba_{0.6}Sr_{0.4}TiO_3$ material is 0.01.

In modeling and examining the microwave performance of the design shown in Figure 25.13b,c, three simulation codes were used including:

1. The method of moments based code (LINPAR) [27] to determine the per unit length transmission line parameters L, C, R, and G.
2. A full 3D method of moments code WIPL-D [28].
3. A spectral matrix based code [23], with a modified Green's function to take into account the finite thickness of the conductors.

Results from these three codes were compared with the experimental measurements reported in Krowne et al. [23].

The following is a brief description of the implementation of each of these codes.

Simulation Using the LINPAR Code LINPAR models 2D multiconductor transmission line structures by defining node coordinates at the interface of two homogeneous boundaries. A pulse is then assigned between two adjacent nodes and the method of moments is employed using pulse basis functions. Higher concentration of nodes is employed where charges undergo rapid changes or where the absolute value of the charge distribution is expected to be high. The frequency-dependent series resistance is modeled using the following relation [27]:

$$R(f) = R(f_{\text{ref}}) \left(\frac{f}{f_{\text{ref}}} \right)^{r_e} \tag{25.3}$$

where f_{ref} is the reference frequency, r_e is the exponential term defining the frequency variation of the series resistance, and $R(f_{\text{ref}})$ is the series resistance evaluated at the reference frequency given by

$$R(f_{\text{ref}}) = \sqrt{\frac{\pi \mu f_{\text{ref}}}{\sigma}} \tag{25.4}$$

where μ is the conductor permeability (in H/m), which is practically always equal to μ_0, and σ is the conductor conductivity given in S/m. For the device in Figure 25.13b, $f_{\text{ref}} = 0.1$ GHz, r_e is 0.917 for the unbiased case, ε_r ($V_{\text{dc}} = 0$V) $= 723(1 - j0.01)$, and r_e is 0.855 for the baised case, ε_r ($V_{\text{dc}} = 40$ V) $= 441.2(1 - j0.01)$. The conductor was made of pure silver with a conductivity of 61.387 MS/m. A total of 982 nodes were used in the model to provide good convergence.

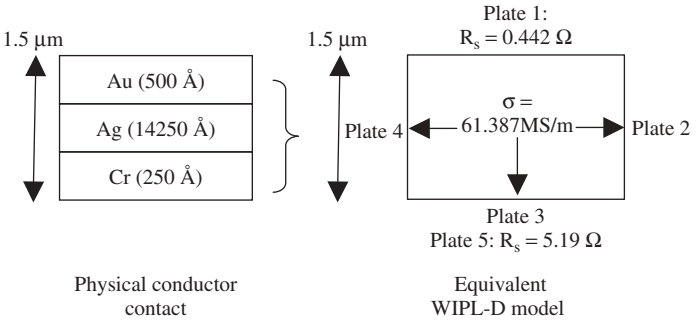


Figure 25.14 Method for modeling metallization contacts in WIPL-D.

Simulation using WIPL-D Code It should be noted that WIPL-D [28] models 3D composite metallic and dielectric structures using wires and plates. The finite conductor thickness is modeled using thin solid boxes. The distributed loading of the metallic plates may be given by the conductivity of the metal *or* as surface impedance given by [28]

$$Z_s = \frac{1}{t\sigma_m} \tag{25.5}$$

where t is the thickness and σ_m is the conductivity of the metal. The multiconductor combination that makes up the ohmic metal loss of the phase shifter device is modeled using metallic plates, each with a unique distributed loading depending on the layer it represents. Figure 25.14 illustrates a method for modeling the metallic layer combination of Cr/Ag/Au (250 Å, 14250 Å, 500 Å). Plate 1 represents the gold (Au) metal layer, lies in the free-space domain, and is assigned a surface impedance of 0.442 Ω, where the conductivity is 45.21 MS/m. Plate 2 represents the silver (Ag) layer, Plate 3 represents the silver (Ag) layer, and Plate 4 represents the silver (Ag) layer. Each of the aforementioned plates is assigned a conductivity of 61.387 MS/m and is in the free-space domain. Plate 5 represents the chromium (Cr) layer and overlaps plate 3. Plate 5 is in the ferroelectric substrate domain and is assigned a surface impedance of 5.19 Ω, where the conductivity is 7.714 MS/m.

The quick variations of the electromagnetic field at the interface between the microstrip conductor and the dielectric cannot be properly approximated by the low order polynomial expansions associated with dielectric plates. This edge effect can be properly taken into account if the edge of a dielectric plate connected to the edge of a microstrip conductor is modeled by a separate narrow strip [28]. The length of the phase shifter is arbitrary since there is no variation in the z-direction. For our design the line length was kept short at 1 mm in order to keep the number of unknowns low while not generating any port-to-port coupling. A total of 299 unknowns were used in the model for good convergence.

Simulation Using the Spectral Matrix Code The spectral matrix code [23] was formulated in MATLAB and the number of basis functions used to expand the x- and z-components of the electric field in the coplanar waveguide slots were set to $n_x = n_z = 3$ and the number of Fourier spectral terms was set to $n = 200$. It is important to note that in the spectral domain approach described in Krowne et al. [23], it was necessary to introduce the following modifying diagonal term in the admittance Green's function to

account for the finite thickness of the conductors. This modification was also implemented in our code.

$$Gs3 = \frac{1}{\sigma w_{\text{eff}}} \left\{ (1+j)\frac{t}{4\delta} + c(f)\frac{w_{\text{cs}}}{w_{\text{cs}} + 2w_{\text{slot}}} \cdot \left[\frac{(1+j)w_{\text{cs}}/\delta}{\tanh\left[(1+j)w_{\text{cs}}/\delta\right]} \right] \right\} \tag{25.6}$$

where

$$w_{\text{eff}} = \delta_{\text{CentCond}} + \delta_{\text{GndCond}} \tag{25.7}$$

$$c(f) = \left(\frac{f}{f_0}\right)^{v}, \quad \delta = \frac{1}{\sqrt{\pi f \mu_0 \sigma}} \tag{25.8}$$

where w_{eff} is the summation of the current crowding depth of the center conductor (δ_{CentCond}) and the ground conductor (δ_{GndCond}) [29]. w_{cs} is the center conductor strip width and w_{slot} is the gap width of the coplanar waveguide device. $f_0 = 10$ GHz, $w_{\text{eff}} = 1.98$ μm, and $v = 0.373$ for the unbiased case and $f_0 = 10$ GHz, $w_{\text{eff}} = 2.97$ μm, and $v = 0.3428$ for the biased case. The values for w_{eff} obtained using the current crowding method [29] were in very good agreement with the values presented in Krowne et al. [23], which were determined based on fitting measured data. The values determined in Ref. 23, were $w_{\text{eff}} = 2.004$ μm for the unbiased case and $w_{\text{eff}} = 3.021$ μm for the biased case. All simulations were performed using a Pentium 4 processor with 3.75 GB of RAM.

Based on these simulation methods, an optimized design was determined and excellent comparison with experimental data in Ref. 23 was achieved. Dimensions and material parameters for this design include a center strip width of 32.76 μm, a gap width of 16.4 μm, and a conductor thickness of 1.5 μm. The metal contact is comprised of the multimetallization system of Cr/Ag/Au (250 Å, 14250 Å, 500 Å). The conductivity of chromium is $\sigma = 7.714$ MS/m, of silver is $\sigma = 61.387$ MS/m, and of gold is $\sigma = 45.21$ MS/m.

To model the anisotropic permittivity of the ferroelectric material as well as the very thin low loss nontunable layer, a new spectral domain approach with a modified Green's function incorporating the current crowding effects [29] was developed [30]. The multidielectric ferroelectric phase shifter was modeled using the newly formed Green's function and was compared with the direct metallization approach described in Ref. 23. The permittivity of the ferroelectric material for the unbiased case was set to

$$\varepsilon_r(V_{\text{dc}} = 0\ \text{V}) = \begin{bmatrix} 723 & 0 & 0 \\ 0 & 723 & 0 \\ 0 & 0 & 723 \end{bmatrix} - j \begin{bmatrix} 723 & 0 & 0 \\ 0 & 723 & 0 \\ 0 & 0 & 723 \end{bmatrix} \tan\delta_{\text{fe}} \tag{25.9}$$

and the permittivity for the biased case was set to

$$\varepsilon_r(V_{\text{dc}} = 40\ \text{V}) = \begin{bmatrix} 441.2 & 0 & 0 \\ 0 & 723 & 0 \\ 0 & 0 & 723 \end{bmatrix} - j \begin{bmatrix} 441.2 & 0 & 0 \\ 0 & 723 & 0 \\ 0 & 0 & 723 \end{bmatrix} \tan\delta_{\text{fe}} \tag{25.10}$$

Figure 25.15 compares the attenuation constant of the direct metallization design illustrated in Figure 25.13a to the multidielectric approach illustrated in Figure 25.13b

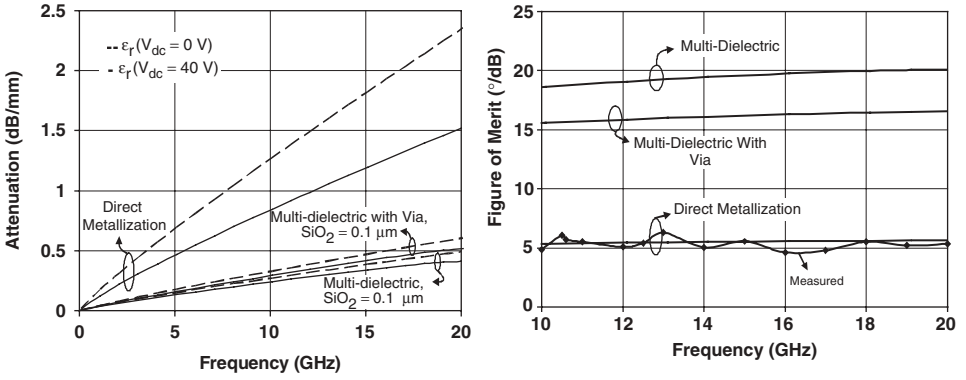


Figure 25.15 (a) Attenuation constant α for the unbiased $\varepsilon_r(V_{\text{dc}} = 0\ \text{V})$ and biased $\varepsilon_r(V_{\text{dc}} = 40\ \text{V})$ cases and (b) figure of merit (FoM) comparing the direct metallization design shown in Figure 25.13a, the multidielectric design shown in Figure 25.13b, as well as the multidielectric with via design shown in Figure 25.13c.

for the unbiased and biased cases. The attenuation constant decreased by ~1.85 dB/mm at 20 GHz for the unbiased case and decreased by ~1.07 dB/mm at 20 GHz for the biased condition. The FoM results in Figure 25.15b demonstrate a significant increase when using the multidielectric approach. The overall phase shift for the multidielectric design decreased by only ~3.6°/mm at 20 GHz compared to the direct metallization case, while there was a large decrease in the attenuation constant, and this is what resulted in the significant overall increase in the FoM (°/dB). The FoM improved from ~6°/dB to ~20°/dB at 20 GHz and illustrates that the multidielectric design improves the overall phase shifter performance.

As expected, the inclusion of the low loss SiO_2 layer would complicate the dc biasing circuit. For the direct metallization approach, the ferroelectric may be tuned using a simple bias tee arrangement. However, in the multidielectric design with a low loss dielectric layer between the signal conductors and ferroelectric film, it is expected that the low permittivity layer would effectively "short out" and E-field in the ferroelectric layer. It is therefore important that we address appropriate dc biasing for the ferroelectric material layer. The multidielectric case in Figure 25.13c, which includes the formation of a via through the silicon dioxide layer, was analyzed to determine the feasibility of effectively polarizing the ferroelectric substrate. In this case biasing would involve the use of a bias tee. As may be seen from Figure 25.15, even with the new biasing approach, the proposed multidielectric phase shifter design continues to provide reduced attenuation as well as increased FoM.

25.3 CONTINUOUS TRANSVERSE STUB (CTS) ANTENNA ARRAY TECHNOLOGY

At microwave frequencies, it is conventional to use slotted waveguide arrays, printed patch arrays, and reflector and lens systems [31–34]. However, as the frequencies increase beyond 20 GHz, it becomes difficult to use these radiating elements. Conventional slotted planar array antennas are difficult to use above 20 GHz due to their complicated design

and expensive fabrication. This, in conjunction with the precision and complexity required in machining, joining, and assembly of these antennas, further limits their use. Printed patch array antennas suffer from inferior efficiency due to their high dissipative losses, particularly at high frequencies and for larger arrays [32]. Frequency bandwidths for such antennas are typically less than that which can be realized with slotted planar arrays. The sensitivity to dimensional tolerances is greater in this type of array due to the dielectric loading and resonant structures inherent in their design [8].

Reflector and lens antennas are generally employed in applications for which planar array antennas are undesirable, and for which the additional bulk and weight of a reflector or lens system is acceptable [33, 34]. The absence of discrete aperture excitation control in reflector and lens systems limits their effectiveness in low sidelobe and shapedbeam applications.

Traditional phased array antennas that employ microstrip patch or slot arrays require many levels of multiport feeding networks, which have tolerance and associated coupling issues. Aperture efficiency is also an issue with increasing frequency. Although reflector and lens antennas are the most popular approach for high frequency applications, some new applications exist such as mobile radar, where lightweight and limited space are among the design requirements and hence the parabolic type is not desirable.

25.3.1 Emerging Low Cost Continuous Transverse Stub Technologies

The continuous transverse stub (CTS) antenna array is a new microwave coupling/radiating element that was patented by Raytheon Systems Company in 1991 [35]. The CTS antenna is useful for frequencies as high as 20 GHz and up to the millimeter wave and quasi-optical frequencies. The CTS array antenna illustrated in Figure 25.16 exploits a unique stub element as its radiator basis. Unlike a typical resonant high-Q radiator such as a dipole, slot, or patch antenna, the CTS exhibits a very low-Q impedance characteristic. This results in a significant frequency bandwidth improvement and a unique dimensional insensitivity, which in turn translates into significant producibility and reliability advantages. The continuous geometry of the stub radiator allows for a dramatic reduction in part count: N by N discrete radiators are replaced by N *continuous* stubs. In addition, the stub radiator's high physical aspect

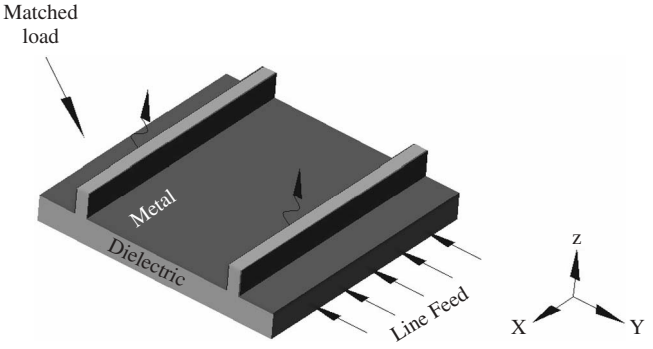


Figure 25.16 Illustration of a parallel plate waveguide continuous transverse stub (CTS) two radiating elements antenna. (Ref. 35.)

ratio (length-to-width) minimizes fringing effects and realizes superior polarization purity over that of competing planar array and reflector-type implementations [36].

Operation of the CTS involves incident parallel plate waveguide modes launched normally via a primary line feed that has longitudinal y-directed electric current components on the conducting walls associated with them. These currents, when interrupted by the presence of a continuous stub, excite a longitudinal, x-directed magnetic current across the stub/parallel plate interface. This induced displacement current, in turn, excites equivalent electromagnetic (EM) waves that radiate into free space. The electric field vector is linearly polarized transverse to the stub elements [35, 36].

One of the advantages of the CTS antenna is its simple design. The antenna consists of a dielectric (a plastic such as rexolite or polypropylene) that is machined or extruded to the shape shown in Figure 25.16. The dielectric is then metal plated to form the final CTS antenna. The CTS thus lends itself to high volume plastic extrusion and metal plating processes, thereby enabling low cost production. Additional advantages include compact size, light weight, low loss, and high directivity.

Pencil beam antenna arrays may be constructed using multiple stub elements. Several prototypes have been manufactured for radar and microwave communications applications including a four subarray system each with 95 elements and independent T/R modules to allow electronic microscanning. The phased array was designed to operate at the 94-GHz frequency band (W-band) with radiation efficiency between 58% and 65% and a nominal directive gain of 35 dBi [37]. The CTS array antenna also finds use in high data rate point-to-point communications in airborne and ground-based systems. The CTS array has also been employed in fixed and mobile terrestrial- and satellite-based broadband access applications at both microwave and millimeter-wave frequencies [36]. The high gain along with its inherently low tolerance to manufacturing errors leads to the low cost of this technology.

25.3.2 Coaxial CTS Antennas

A CTS element with circular radiating stub elements residing in the outer conductor of a coaxial transmission line was invented by our group while at the University of Utah [38, 39]. It is an omnidirectional antenna for operation at microwave frequencies. Purely reactive radiating stub elements are formed when circular stubs of moderate height are open to free space. Precise control of element coupling via coupling of transmission line modes is accomplished through variation of the stub radius, stub width and separation, and the dielectric media that fill the transmission line and stub elements.

The coaxial CTS is fed by a simple coax transmission line to simplify impedance matching and eliminates the complexity of waveguide feeds as is the case for the parallel plate waveguide CTS antennas. The array forms a narrow beam (pencil beam) pattern that is omnidirectional in the plane of the radiating stubs (perpendicular to the axis of the antenna).

Isom et al. [40] developed a dual-frequency band coaxial CTS array antenna. The array is comprised of six elements, where three elements are designed to operate at the lower frequency of 4.2 GHz and three elements are designed to operate at the higher frequency of 19.4 GHz [40]. The high frequency subarray is located near the feed while the low frequency subarray is located between the high frequency section and the load. At lower frequencies, the high frequency subarray is designed to be nonradiating and provides full transmission of microwave power with negligible losses. At higher frequencies,

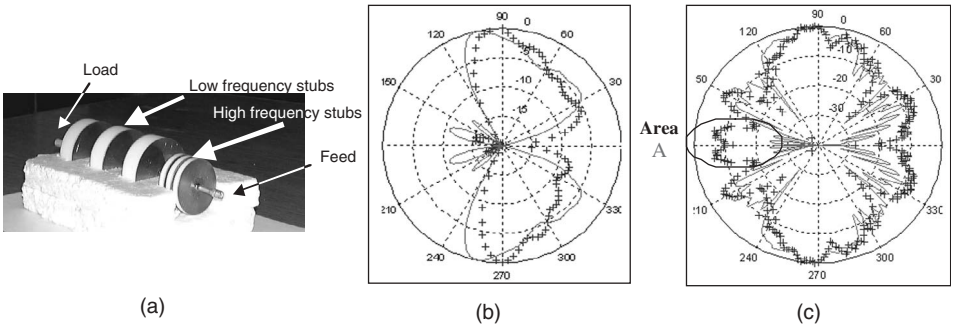


Figure 25.17 (a) Prototype of coaxial CTS array, and E-plane radiation pattern at (b) 4.4 GHz and (c) 19.4 GHz, where the solid lines (—) are simulated and cross lines (+) are measured results. In (c), the reflections observed near "Area A" are attributed to discontinuities in the feed cable (cable near high frequency subarray). (From Ref. 40.)

microwave energy is radiated by the high frequency stubs with little power transferred to the lower frequency stubs at the end of the array. The amount of radiation for both frequency bands is approximately 98%, and good impedance matching is observed (return loss better than −10 dB). Figure 25.17b and 25.17c show measured (+) and simulated (—) E-plane antenna patterns at 4.4 and 19.4 GHz, respectively, exhibiting well-formed main beams and good agreement for both frequency bands.

Thinkom Solutions, Inc. in collaboration with the University of Hawaii through the Small Business Innovation Research (SBIR) government program has expanded the dual-frequency band to a five-band array employing five subarrays for each associated frequency band. The five-band array was formed following the same design concepts of the dual-band array in that the highest frequency subarray was positioned closest to the feed with lower frequency subarrays progressively placed behind it for each subsequent lower frequency band.

From the above discussion regarding the various implementations of the CTS antenna technology, it may be important to point out some of the limitations of the available designs. Although the parallel plate CTS antenna is an innovative architecture for future wireless applications, there is still a need for improvements, specifically in the areas of (1) reducing the production costs associated with waveguide feed technologies and specifically simplifying the feed networks (eliminate the complicated H-plane horn or line feeds), and (2) allowing simple integration with RF front-end circuitry while maintaining the overall CTS performance. The coaxial CTS described earlier overcomes a number of the aforementioned limitations including a simple feed network, thus providing opportunities for low frequency designs, as well as the use of low cost transmission line technology. The coaxial technology, however, is not compatible with planar RF front-end circuitry often based on microstrip technology.

More recently, our group developed yet another version of the CTS technology. The newly developed and patented [41, 42] coplanar waveguide continuous transverse stub (CPW–CTS) antenna makes use of the CTS technology in a planar microstrip configuration and produces a broadside radiation pattern with a maximum in the $+z$-direction, perpendicular to the plane of the antenna as shown in Figure 25.18. The newly developed coplanar CTS phased array antenna would particularly find use where low cost is a requirement, such as in global positioning satellite applications and single-use battlefield

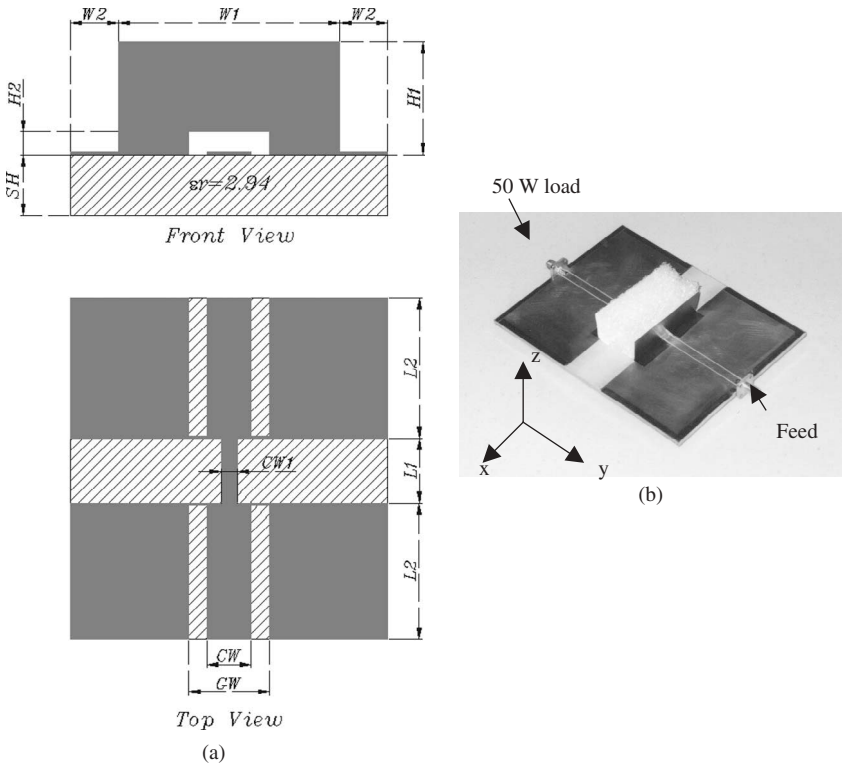


Figure 25.18 (a) Schematic of CPW–CTS and (b) fabricated prototype of the designed CPW–CTS. A styrofoam spacer is inserted in the radiating stub to accurately maintain the stub dimensions. (From Refs. 41 and 42.)

weapon systems. Other applications include airport radar scanning systems, where high speed scan is required for multiple-aircraft detection, and battlefield scenarios, where multiple-target tracking is needed.

The coplanar CTS offers several advantages over previous planar and coaxial CTS designs. These include (1) a unidirectional broadside radiation pattern compared to the omnidirectional pattern associated with the coaxial CTS, (2) simple 50-Ω microstrip transmission line feed compared to the complex waveguide feed networks for the parallel plate design, and (3) simple integration with integrated circuitry in the transceiver's front end, which was not effectively possible in any of the earlier designs. In addition, the CPW–CTS may realize high gain through a series array of elements with a single feed, an advantage over high gain dipole antenna arrays as well as microstrip patch elements, which may require *N feeds for N radiators*. The coplanar implementation also allows for low as well as high frequency use of this technology

25.3.3 Design Procedure for CPW–CTS

The coupling values from the CPW to the radiating stubs are primarily dependent on the height (H1), width (W1), the transmission line width in the section of the radiating elements (CW1), the distance (L1) of the parallel plate stubs, and the transverse stub gap

(H2). The gap distance between the transverse stubs (H2) is used to adjust the coupling capacitance to compensate the inductance of the purely reactive stub elements. The antenna is fed with a simple coplanar waveguide transmission line. The signal conductor width (CW), gap width (GW), and dielectric constant (ε_r) were chosen to provide 50-Ω feed impedance.

The feed point (L2) is positioned away from the edge of the stubs to maintain a good impedance match. L2 is carefully chosen to obtain the desired radiation pattern as a result of the axis of the feed line being oriented in line with the E-plane radiation pattern. The substrate height (SH) is chosen such that it improves the directive gain without severely influencing the incident transmission line modes and increasing reflections.

The length (L1), width (W1), and height (H1) are selected to be approximately one-half wavelength, one wavelength, and one-third wavelength in the dielectric material that fills the stub, respectively. These values of CW and CW1 were carefully chosen after extensive simulations to produce increased radiated power and at the same time maintain good impedance match. The ground width (W2) was approximately one-third wavelength. Large values of W2 are needed to reduce the undesired back lobes, most notably in the $H-$plane radiation measurements.

For simulation purposes WIPL-D, a full 3D electromagnetics simulator based on the method of moments, was used [28]. An antenna operating at C-band was designed and desired features include low input impedance, high radiated power, and a broadside radiation pattern. The obtained design dimensions after extensive simulations were as follows: W1 = 54.5 mm, W2 = 28.3 mm, H1 = 18.2 mm, H2 = 1.5 mm, CW = CW1 = 4.7 mm, GW = 5.97 mm, L1 = 27.3 mm, L2 = 54.5 mm, SH = 3.01 mm, and $\varepsilon_r = 2.94$ (RT/Duroid 6002). Figure 25.19 shows the measured and simulated return and insertion losses. As can be seen, S_{11} is better than -10 dB in the frequency range between 5.2 and 5.6 GHz and the radiated power ratio is greater than 40%. This radiated power ratio is given by

$$Pwr_{\text{rad}} \approx (1 - Pwr_{\text{refl}} - Pwr_{\text{trans}}) * 100\% \tag{25.11}$$

where the reflected power, Pwr_{refl}, and transmitted power, Pwr_{trans}, are given by

$$Pwr_{\text{refl,trans}} \approx 10^{S_{11,21}(\text{dB})/10} \tag{25.12}$$

In Figure 25.19, S_{11} and S_{21} are the reflection and insertion losses, respectively. The term transferred power is used to account for the amount of power received at the end of the antenna, after the radiating stub. Figure 25.20 shows the measured and predicted $E-$and $H-$plane radiation patterns at 5.3 GHz. The measured $E-$plane beam peak was -18° and the measured 3-dB beamwidth was 74.8°. The maximum back lobe was 14 dB lower from the maximum beam peak. The maximum back lobe of the measured $H-$plane was 17 dB down from the maximum beam peak. As may be seen, the measured data and simulated results are in good agreement and validate the predictability of the performance of this new antenna design.

Similar to the cases of parallel plate and coaxial CTS implementations, it is of interest to examine the radiation characteristics of a CPW–CTS array. If the elements are oriented in the form of a single-feed series array, a traveling-wave or a resonant antenna is formed [41, 43]. If the CTS elements are spaced $\lambda_g/2$ apart, the array forms a resonant

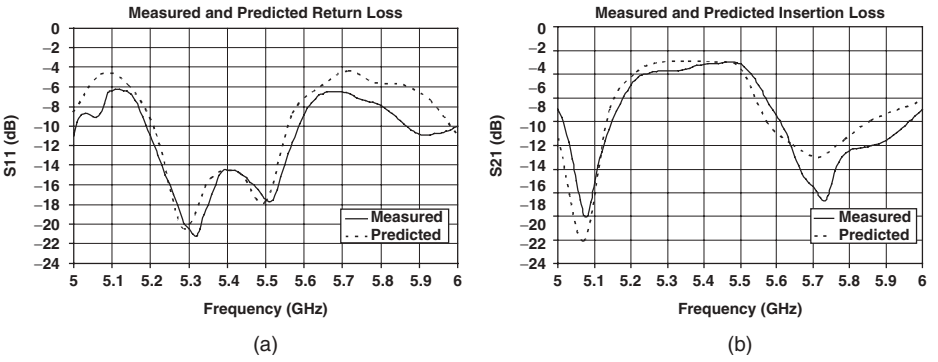


Figure 25.19 Predicted and measured (a) return loss and (b) insertion loss. (From Refs. 41 and 42.)

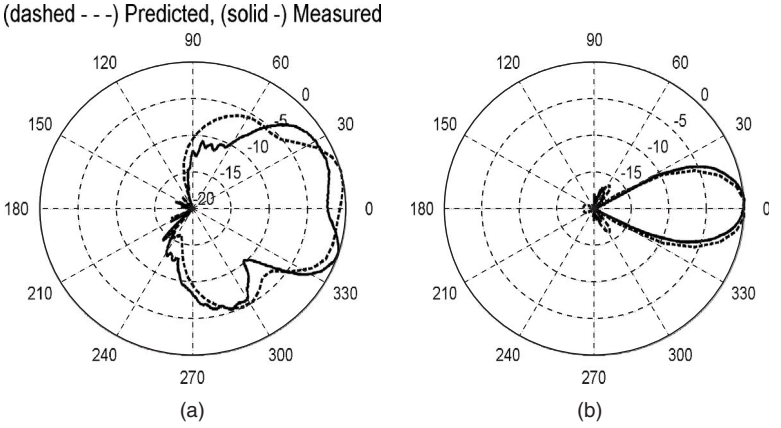


Figure 25.20 Predicted and measured (a) E-plane and (b) H−plane radiation pattern. (From Refs. 41 and 42.)

array; otherwise, it is a traveling-wave array [34]. The single-element CPW−CTS antenna was extended to a resonant array with three elements each spaced $\lambda_g/2$ apart. The three-element array was modeled using WIPL-D and a prototype was developed at X-band. Mutual coupling between elements is appropriately accounted for using the full 3D analysis of WIPL-D. Figure 25.21 illustrates the dimensions of the array and a photograph of the developed prototype. The dimensions of the coplanar waveguide feed and load lines were designed to provide 50-Ω impedance. The selected substrate was Duroid 5870 with a permittivity of 2.3 and tan δ of 0.0005. The center-to-center stub spacings were $\sim\lambda_0/2$ and, because of the low effective permittivity ($\varepsilon_{r,\text{eff}} = 1.4$), it was also $\sim\lambda_g/2$ at the 10-GHz design frequency. The first stub closest to the feed is designed to couple a small amount of power while the last stub is designed to couple relatively more as the total amount of available power is reduced due to radiation from previous stubs. This is done by adjusting the width of each stub. The design was also implemented such that induced currents J_z on each stub are out of phase, which is just the condition required for the electric field E_y radiated by each stub to be in phase

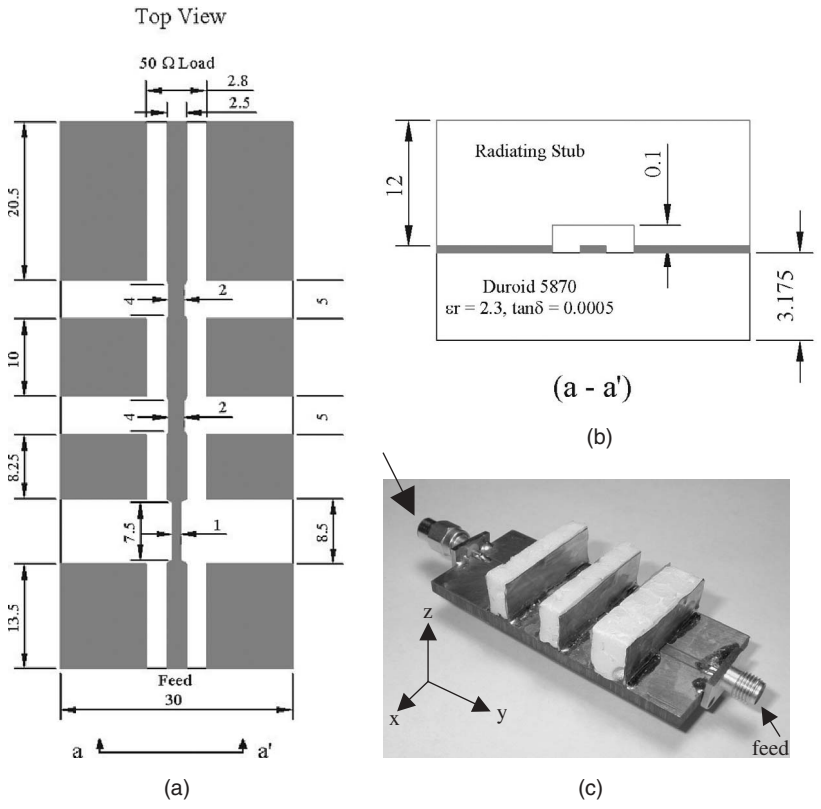


Figure 25.21 Layout of the three- element coplanar waveguide continuous transverse stub (CPW–CTS) array (all dimensions in mm): (a) top view, (b) front view, and (c) fabricated prototype. Conductor thickness was 35 μm (1 oz) electrodeposited copper with conductivity of 58.8 MS/m. (From Ref. 41.)

in the direction of maximum radiation, thus producing linear polarization and superb polarization purity.

The measured and simulated S_{11} of the CPW–CTS array is shown in Figure 25.22a [43]. The 10-dB bandwidth of the CPW–CTS array was ~4% and possesses narrowband characteristics typical of resonant arrays with element spacing of $\sim\lambda_g/2$. The reflections from each stub combine in such a way as to cancel at the input of the array at only one frequency; hence the array has frequency-dependent input impedance. The copolarized and cross-polarized radiation patterns are shown in Figure 25.22b and illustrate that the linearly polarized array produced a broadside radiation pattern where the beam peak is 10.6 dBi at -7.5° from broadside (toward feed direction) and a half-power beamwidth of $\sim 36^\circ$. The measured cross-polarization level is below -20 dBi, illustrating outstanding polarization purity, indicative of the TEM mode propagating in the radiating stubs. The radiated power at the aperture of each stub element is depicted in Figure 25.22c and illustrates nearly equal amount of power from each element. The slight deviation of the main beam from broadside toward the feed direction is a result of the small increase in radiated power by the first stub element.

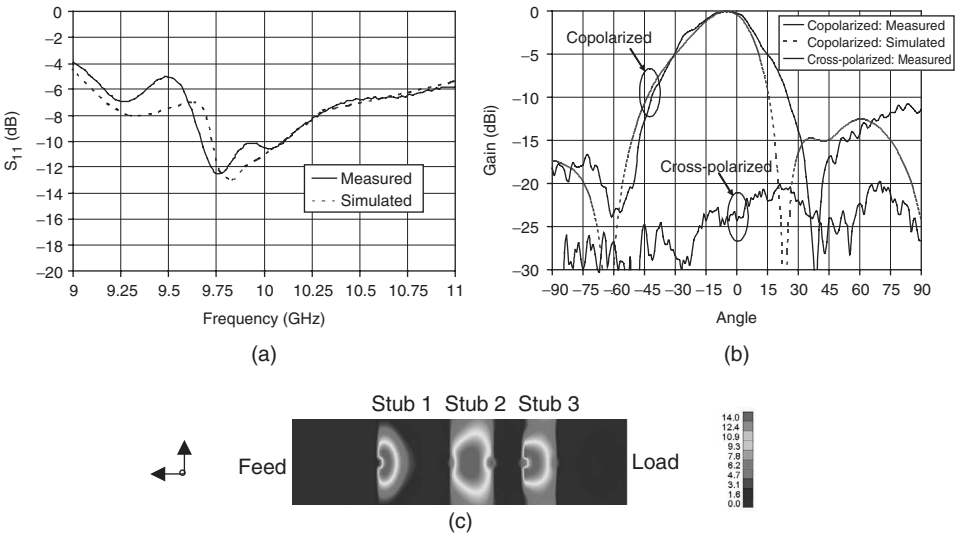


Figure 25.22 (a) Measured results of S_{11} (dB) versus frequency (from Ref. 43), (b) measured results of copolarized (E-plane, y-z plane) and cross-polarized radiation patterns (dBi) at 10 GHz (from Ref. 43), and (c) radiated power (W/m^2) at the aperture of each stub.

25.4 ANTENNA ARRAYS USING INTEGRATED CTS AND FERROELECTRIC MATERIALS TECHNOLOGIES

In this section we provide first a brief overview of some of the research effort aimed at the integration of the CTS and ferroelectric materials technologies in the design of antenna arrays with beam-steering capabilities. Most of these activities were focused on using the parallel plate-type CTS as radiating elements. Even with the invention of the coaxial CTS, it was difficult to expect quick implementation of this integration effort as fabricating ferroelectric rods or depositing ferroelectric cylindrical layers around the center conductor of a coaxial line presents a significant technology challenge. The invention of the coplanar waveguide implementation of the CTS technology, however, renewed hope of a low cost implementation of such a technology, and as a result a significant part of this section is devoted to the simulation of this integrated design. Specifically, Section 25.4.1 provides a general description and a discussion of earlier integration efforts, while Section 25.4.2 focuses on describing the results of a simulation effort that explores the resulting characteristics from integrating the CPW–CTS and the ferroelectric materials technologies.

25.4.1 Overview of Phased Arrays with Ferroelectric-Based Beam-Scanning Capabilities

The Raytheon Company has developed a row–column electronically steered array that employs phase shifters for steering in the H-plane and tunable ferroelectric materials for scanning in the E-plane in the CTS antenna architecture as illustrated in Figure 25.23 [44]. Changing the voltage across the ferroelectric changes its permittivity and, in turn, the velocity of propagation along the parallel plate transmission line. It provides for a

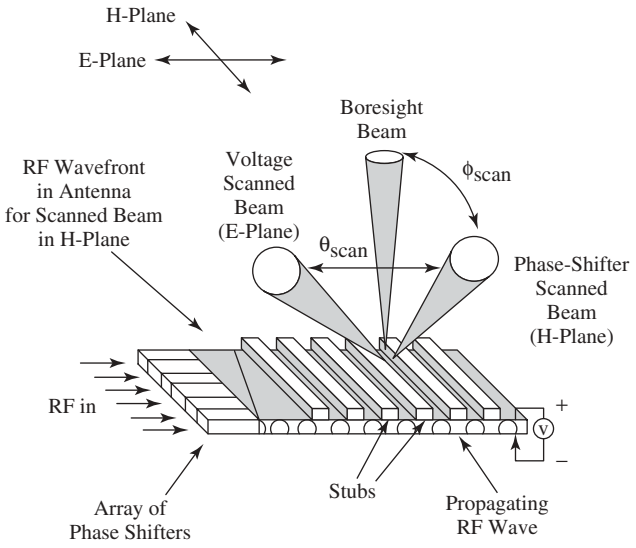


Figure 25.23 Low cost continuous transverse stub (CTS) antenna employing 2D electronic scanning. (From Ref. 44.)

lightweight, low cost, small thickness antenna. These phased array antennas were planned to be used in aircraft radar and mobile multifunction antenna applications. The procurement of this system, however, was eventually cancelled as a result of the impractical (large) scan losses as a result of the large insertion losses (particularly ohmic) associated with the high permittivity ferroelectric material ($\varepsilon_r \sim 100-1000$).

Iskander et al. [45] developed a ferroelectric-filled CTS antenna with 15 radiating elements, illustrated in Figure 25.24 at K-band for electronic scanning across the E-plane (y-z plane). The antenna possessed a variable waveguide height to allow appropriate impedance matching between radiating stub sections. Approximately $\pm 60^{\circ}$ of beam scanning was achieved with a ferroelectric tunability of 19.81%, where the tunability is given by

$$\text{Tunability} \approx \frac{\varepsilon_1 - \varepsilon_2}{\varepsilon_1} \tag{25.13}$$

where ε_1 is the permittivity of the unbiased case and ε_2 is the permittivity of the biased case. Although the array produced excellent scanning capabilities, the ohmic losses were very high as well, particularly as the height of the parallel plate waveguide was decreased to reduce the required biasing voltage. For example, for $f = 30$ GHz and the thickness of the conductor equal to 0.1 mm, the conductor loss of the selected ferroelectric material is 23.35 dB (unbiased) and 21.44 dB (biased). This high value may be reduced by increasing the height of the waveguide, and hence increasing the guide impedance; however, this is at the expense of increased biasing requirements and the potential for dielectric breakdown.

To overcome conductor losses, a novel multidielectric approach was developed, whereby multiple dielectric layers were employed in the parallel plate CTS antenna (see Figure 25.25) to reduce the ohmic losses without sacrificing tunability, and thus maintain the beam-scanning performance [46]. The phase shifter section was based

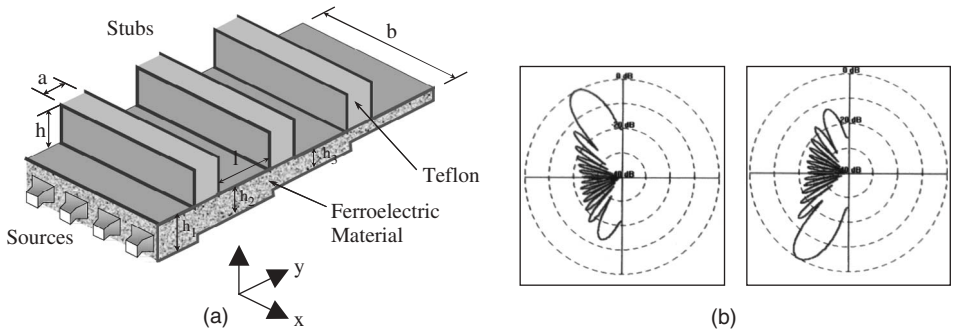
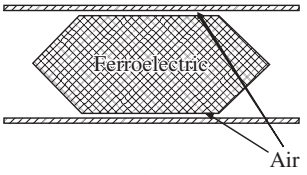


Figure 25.24 (a) Continuous transverse stub (CTS) antenna employing electronic scanning (y-z plane) and (b) -60° (l) and $+60^\circ$ (r) electronic scanning based on 19.81% of ferroelectric tunability. (From Ref. 45.)



(a)

ferroelectric	Normalized beam steering	Normalized Loss	Z_0(100 mm width)
0.16 mm	0.19	0.05	0.1434
0.26 mm	0.43	0.02	0.4145
0.46 mm	0.84	0.001 ± 0.003	1.3663

(b)

Figure 25.25 (a) Schematic of multidielectric layers in a parallel plate configuration, where the ferroelectric is insulated from the conductors by means of a thin layer of air. The ferroelectric is wedge shaped to improve the impedance matching characteristics. (b) Results of beam steering, conductor losses, and characteristic impedance normalized to that of parallel plate structures with fully filled ferroelectric as a function of varying the height of the ferroelectric film. (From Ref. 46.)

on a section of transmission line filled with multiple dielectric layers, including the ferroelectric material. Figure 25.25a depicts a schematic of the multidielectric arrangement, whereby the ferroelectric material is separated from the transmission line conductors by a low permittivity, low loss material such as air. A wedge-shaped ferroelectric slab was incorporated to help improve the impedance match. As described in the table shown in Figure 25.25, by implementing multidielectric materials with a relatively thicker ferroelectric film ($\sim$0.46 mm) and a high ferroelectric permittivity of ε_r ($V_{\text{dc}} = 0$ V) $\sim 400 - j0.36$, it is possible to maintain a high percentage of the tunability ($\sim$85%) compared to the fully filled ferroelectric case.

As previously mentioned, the characteristic impedance of the transmission line may be increased by increasing the height of the parallel plate waveguide; however, this is done at the expense of increasing the ferroelectric biasing requirement. By implementing

the multidielectric transmission line architecture, the characteristic impedance increases as a direct result of the lower *effective* permittivity of the overall system, and without severely increasing the ferroelectric bias requirements.

Figure 25.26a illustrates a schematic of a parallel plate-type CTS array incorporating the multidielectric phase shifter, with an H-plane horn as the feed. This array consists of multiple antenna elements, which are series fed, and the use of variable phase or true time-delay control between each element to scan the beam. Figure 25.26b illustrates

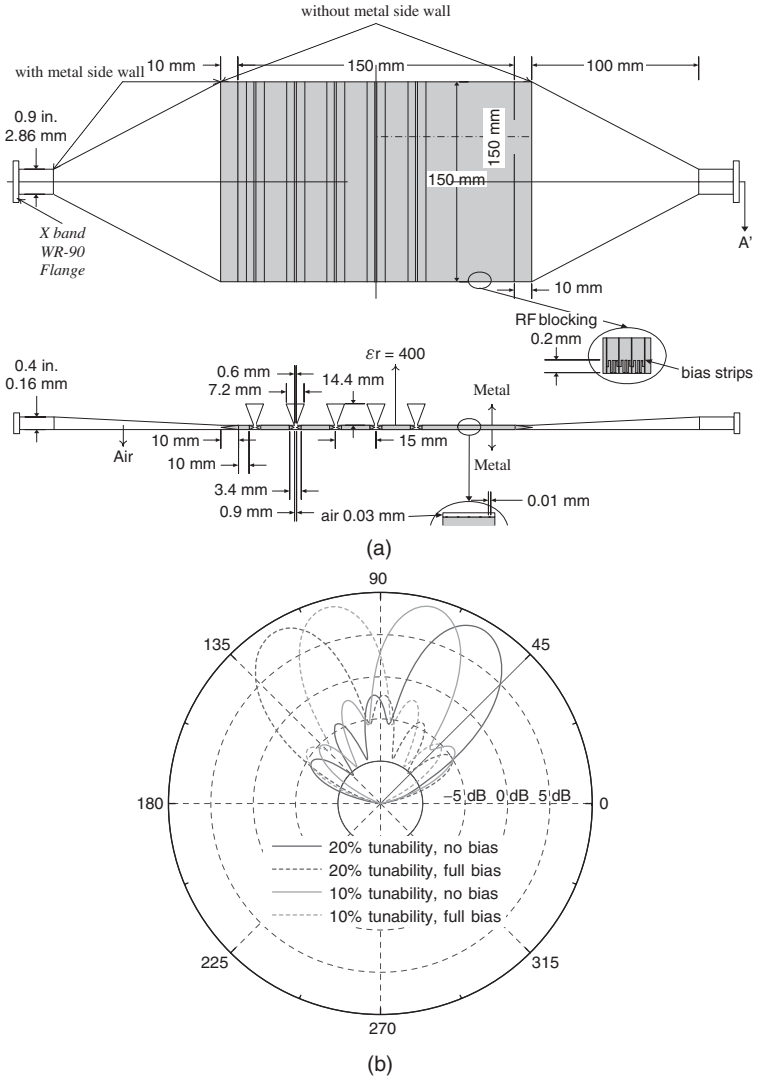


Figure 25.26 (a) Schematic of CTS antenna incorporating a multidielectric parallel plate configuration, where the ferroelectric is insulated from the conductors by means of a thin layer of air. (b) Radiation pattern at 10 GHz illustrating $\sim\pm30^\circ$ of scanning with 20% tunability and $\sim\pm15^\circ$ of scanning with 10% tunability. (From Ref. 46.)

that with a ferroelectric permittivity of ε_r ($V_{dc} = 0$ V) $\sim 400 - j0.36$, the phased array antenna produced $\sim\pm30^\circ$ of scanning with $\sim$20% ferroelectric tunability and $\sim\pm15^\circ$ of scanning with $\sim$10% ferroelectric tunability.

To further investigate the feasibility of the multidielectric system, a simple microstrip transmission line phase shifter was developed [46]. It was found that the phase shift decreased with increased thickness of the low permittivity layer. Other observations included an increase in the characteristic impedance (Z_0) as well as a decrease in insertion losses (S_{21}) with increased low permittivity layer thickness. From these results it may be noted that if one is willing to sacrifice tunability for low insertion loss, higher impedance, and reduction in ferroelectric biasing (result of increasing low permittivity layer thickness), the multidielectric ferroelectric approach is an attractive solution for high efficiency phased array systems.

25.4.2 Integrated Coplanar Waveguide CTS and Ferroelectric Phase Shifters

Based on the earlier discussions, one may point out some limitations in integrating parallel plate waveguide-type CTS antennas with ferroelectric phase shifters. As a phased array antenna system with integrated ferroelectric phase shifters, the parallel plate CTS developed in Refs. 44–46 employs thick (bulk) (>0.1 mm) ferroelectric material that has the limitations of possessing difficult phase, composition, and morphology control, and large control voltages (>100 V_{dc}). This results in a low radiation efficiency and is also considered undesirable for implementation in practical wireless communication systems [20].

Although the coaxial CTS described earlier overcomes a number of the aforementioned limitations, it is also difficult to implement with ferroelectric phase shifters due to the complexity of fabricating ferroelectric materials on cylindrical geometry. In the following section, we describe new and low cost CTS designs as stand-alone antennas, antenna arrays, and as phased array antennas addressing a greater part of the limitations associated with the other available CTS technologies.

25.4.2.1 Design Considerations and Simulation Model of the CPW–CTS Antennas The newly designed ferroelectric phase shifter may be integrated with the coplanar waveguide continuous transverse stub array to form a phased array antenna with electronic beam-scanning capabilities. The CPW–CTS allows less complex feeds associated with Butler, Blass, and Wullenweber networks described earlier. As noted previously, ferroelectric phased array antennas utilize the change in ferroelectric material permittivity with an applied dc control voltage to create phase differences between the radiation elements and hence achieve either beam-steering capability or variations in the radiation patterns [17, 18, 20, 21].

To help fully evaluate this integrated technology and identify the trade-offs involved, our group simulated an integrated two-element CPW–CTS and ferroelectric antenna array system. A detailed drawing of the developed phased array antenna employing two radiating stubs is illustrated in Figure 25.27 [47, 48]. The phased array antenna forms a series-based phase shifter geometry, and benefits include low loss scanning [9]. The multidielectric ferroelectric phase shifter allows appropriate tuning of the phase velocity, hence varying the guide wavelength and enabling electronic beam-steering capabilities. As discussed in earlier sections, the multidielectric coplanar waveguide phase shifter

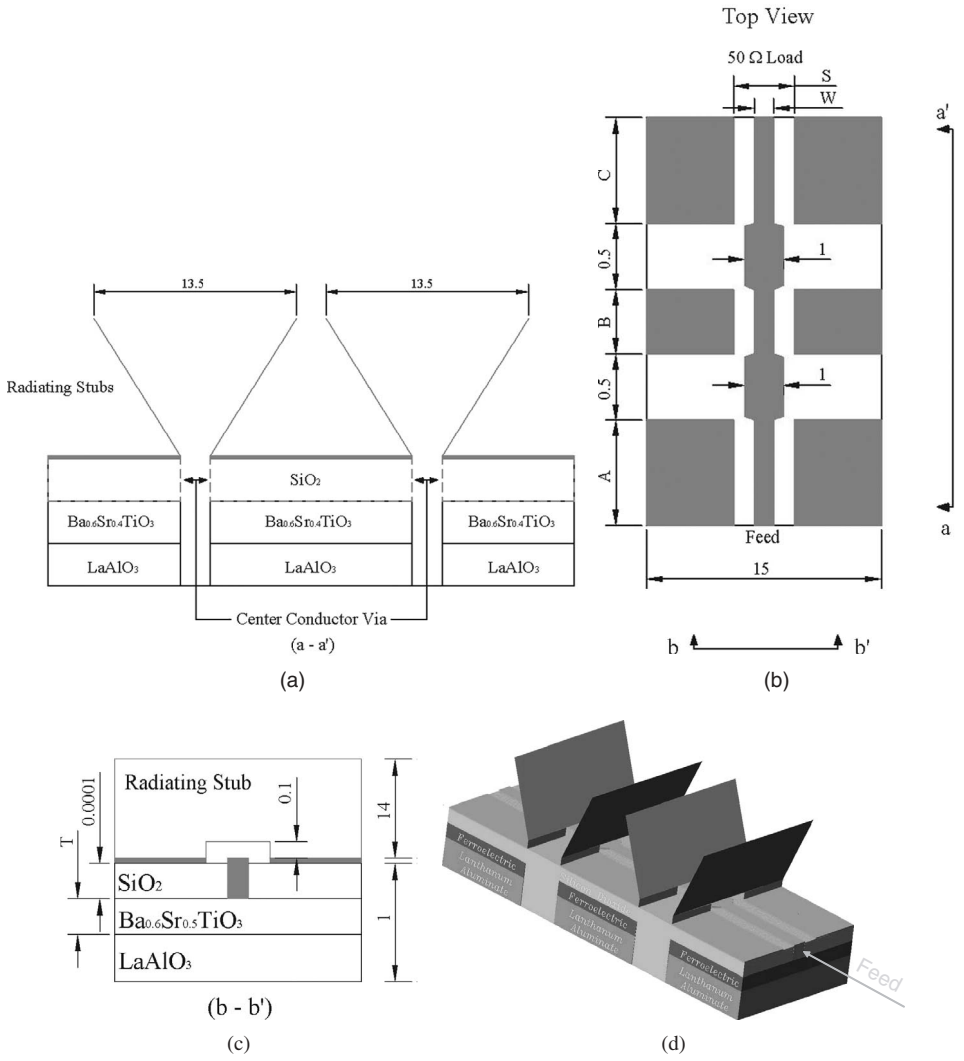


Figure 25.27 Layout of the CPW–CTS phased array antenna: (a) top view, (b) side view, (c) front view, and (d) perspective view (all dimensions in mm). The conductor thickness (above SiO_2 layer) is 1.5 μm and comprised of the multiconductor combination of Cr/Ag/Au (250 Å, 14250 Å, 500 Å.) (From Ref. 43.)

provides adequate tunability, good impedance matching, and low attenuation and hence is an attractive candidate for phased array applications.

Specific design considerations in the implemented design include the following:

1. The ferroelectric material is removed from under the radiating stubs as illustrated in Figure 25.27a. The section under the radiating stubs is filled with low permittivity SiO_2. This is done to enhance the coupling from the coplanar microstrip feed to the radiating stubs. This procedure is also expected to further reduce the sensitivity of the integrated design to manufacturing tolerances.

2. The radiating stub elements are designed in the form of V-shaped stubs, which are narrow at the base and wider toward the top. The tapered geometry allows longer phase shifters, thus allowing larger phase shifts between stubs.
3. The narrow stub section is located so as to couple power near the current maximum on the transmission line feed.
4. To help simplify the dc biasing procedure, the center conductor of the coplanar waveguide penetrates the silicon oxide layer and is in direct contact with the ferroelectric layer (i.e., using the via biasing approach in Section 25.2.2.3) as shown in Figure 25.27c.

25.4.2.2 Simulation Results Simulations were performed using WIPL-D [28] with mutual coupling fully accounted for. Due to the very thin layers as well as the ferroelectric layer, the computation "plates" run the risk of forming "nonconvex" geometries, where one dimension of the plate is much smaller than the other. This error can be overcome if the model contains plates with small dimensions, resulting in high concentration of plates in the ferroelectric sections and an overall increase in the number of unknowns. It was found that 12 plates for the multidielectric phase shifter provided the necessary convergence and accurate solution. Figure 25.28 shows the WIPL-D model with a conductor thickness of 1.5 μm, where the ohmic metal loss of the phase shifter is modeled according to the methods described in Figure 25.15. The radiating stubs were made of copper foil with a conductivity of 58.8 MS/m.

The number of unknowns used in the simulation was 5690. Results in Figure 25.29 show that with the center conductor width $W = 13$ μm, gap width $S = 300$ μm (to provide a 50-Ω transmission line impedance), $A = 9.8$ mm, $B = 13$ mm, and $C = 8.0$ mm, a ferroelectric thickness $T = 5$ μm, SiO_2 thickness of 0.1 μm, and a ferroelectric permittivity modulation between ε_r ($V_{dc} = 0$ V) $= 723(1 - j0.01)$ and ε_r ($V_{dc} = 40$ V) $= 441.2(1 - j0.01)$, approximately $\pm 20^\circ$ of electronic beam scanning is observed. The gain is ~5 dBi for the unbiased case and ~2.8 dBi for the biased case, resulting in ~2.2 dB of scan loss at 10 GHz. The half-power beamwidth is ~47° for the unbiased case and ~46° for the biased case. As shown in Figure 25.29b, the cross polarization remains below −20 dBi for both ferroelectric conditions and demonstrates good polarization purity, similar to that observed with the CPW–CTS array.

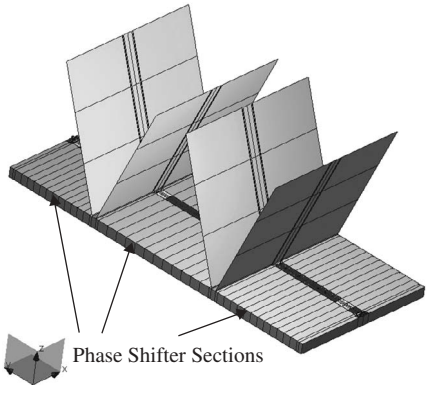


Figure 25.28 WIPL-D model of two-element CPW–CTS phased array, illustrating the high concentration of plates in the multidielectric phase shifter sections.

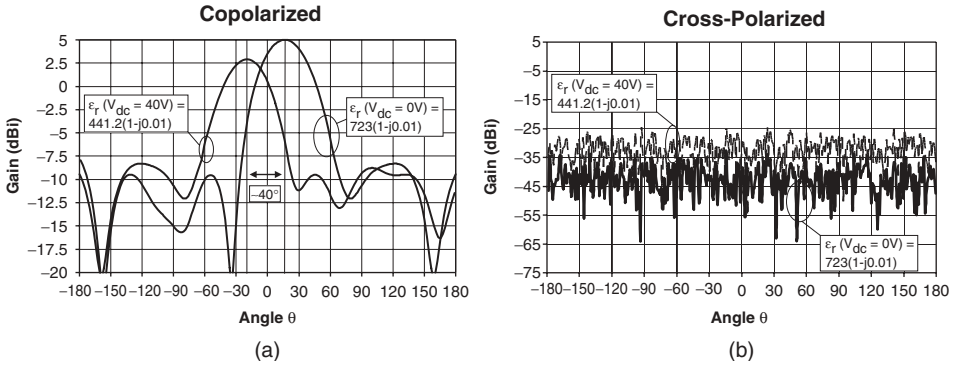


Figure 25.29 Radiation pattern for a 5-μm ferroelectric film: (a) copolarized and (b) cross-polarized radiation pattern of the phased array antenna shown in Figure 25.27 at 10 GHz demonstrating $\pm 20^{\circ}$ of beam scanning. (From Ref. 43.)

The CPW–CTS phased array antenna is a nonresonant antenna array. The distance between radiating stub centers is not equal to $\lambda_g/2$ under either biased or unbiased ferroelectric states. It was designed to provide a broadside radiation pattern at an intermediate value of λ_g between the biased and unbiased cases. Thus a broadside radiation pattern is not observed under either the biased or unbiased ferroelectric cases. Based on the condition that the stubs are not spaced $\lambda_g/2$ apart, the reflections from the different stubs do not add up in phase and the total reflection coefficient at the input to the array is relatively small. Specifically, S_{11} for the unbiased case (ε_r ($V_{\text{dc}} = 0$ V) $= 723(1 - j0.01)$) was -12 dB, and for the biased case (ε_r ($V_{\text{dc}} = 40$ V) $= 441.2(1 - j0.01)$) was -18 dB.

While these results illustrate the effectiveness of the proposed concept of an integrated low cost phased array antenna, it should be noted that the proposed design provides ample opportunity for alternatives and optimizations to fit specific applications and needs. For example, a thinner ferroelectric substrate may be implemented to reduce the V_{dc} requirement of the phase shifter. In this case, we simulated another design with the thickness of the ferroelectric layer reduced to $T = 0.5$ μm, instead of $T = 5$ μm, and the obtained results are shown in Figure 25.30. For this design, the dimensions are $W = 10$ μm, $S = 50$ μm, $A = 9$ mm, $B = 17$ mm, and $C = 8.5$ mm, and the thickness of the SiO_2 layer is 0.1 μm. It may be seen that approximately $\pm 15^{\circ}$ of electronic beam scanning was achieved. This is smaller than what was achieved when a thicker ferroelectric layer was used, but the gain in this case is ~5.3 dBi for the unbiased case and ~5.0 dBi for the biased case, resulting in only ~0.3 dBi of scan loss, instead of 2.2 dBi scan loss for the earlier design, at 10 GHz. The half-power beamwidth is $\sim 42^{\circ}$ for the unbiased case and $\sim 41^{\circ}$ for the biased case. As shown in Figure 25.30b, the cross polarization remains below -20 dBi for both ferroelectric conditions and also demonstrates good polarization purity. S_{11} for the unbiased case (ε_r ($V_{\text{dc}} = 0$ V) $= 723(1 - j0.01)$) was -27 dB, and for the biased case (ε_r ($V_{\text{dc}} = 40$ V) $= 441.2(1 - j0.01)$) was -12 dB.

These and similar design examples show that a thinner ferroelectric layer may be employed, thus leading to higher directive gain and lower scan loss, but would require an increase in the array length to achieve the desirable scan range. A thicker ferroelectric substrate, on the other hand, may be employed to reduce the overall array length; but this will result in a lower directive gain and an increase in the scan loss and will require larger

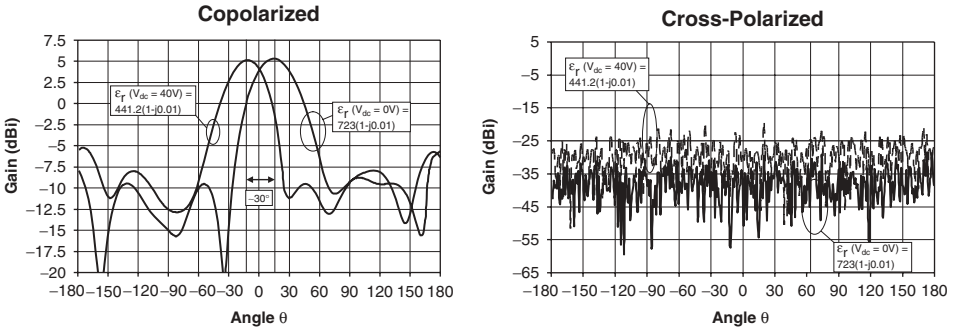


Figure 25.30 Radiation pattern for a 0.5-μm ferroelectric film: (a) copolarized and (b) cross-polarized radiation pattern of the phased array antenna shown in Figure 25.27 at 10 GHz demonstrating ±15° of beam scanning. (From Ref. 43.)

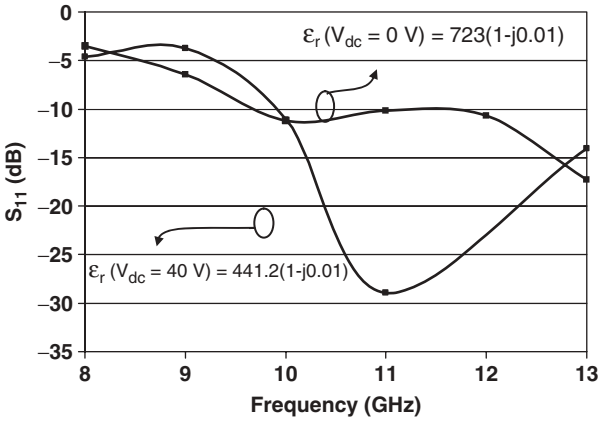


Figure 25.31 Return loss for the case of 0.5-μm ferroelectric film.

dc bias, V_{dc}. Figure 25.31 illustrates the return loss of the phased array system. The return loss is below −10 dB between 10 and 13 GHz, thus depicting good impedance match conditions. The stubs are not spaced $\lambda_g/2$, thus it is not a resonant array as was seen with the three- element CPW–CTS array. As a result, the reflections between elements do not add in phase and therefore the reflection coefficient at the input is small.

25.5 LOW COST PHASED ARRAY ANTENNAS USING DIGITAL BEAM-FORMING TECHNIQUES

In the previous sections we described phased array antennas with analog beam steering using analog phase shifters. Analog phase shifters provide a continuously variable phase. Electrically controlled analog phase shifters can be realized with either varactor diodes that change capacitance with voltage, or as described in Section 25.2.2 of this chapter, by

forming an entire substrate such as ferroelectrics comprised of barium strontium titanate and controlling the phase by applying variable dc voltage. With the broad availability and extended capabilities of digital devices, however, satisfactory beam steering can also be achieved using low cost digital devices. There are several basic designs that are used to build digital-type phase shifters. In all designs, MMIC, MEMS, and/or PIN diodes are typically employed to switch the transmission line in and out of the transmission path. Digital phase shifter types include switched, loaded (constant-phase), and reflection line devices, each with a different number of bits. Analog phase shifters have many advantages over digital phase shifters in phased array antenna systems. In the following section we first briefly review (Section 25.5.1) various digital beam-steering approaches and then describe in more details a cost effective digital beam-forming approach using a hybrid smart antenna technique [49]. In this case (Sections 25.5.2–25.5.3), both simulation results and data from experimental measurements on a developed prototype are presented [50, 51]. This section concludes (Section 25.6) with and a brief performance comparison between analog and digital beam-forming approaches.

25.5.1 Digital Beam-Steering and Beam-Forming Techniques

Beam steering can be achieved using various mechanisms such as adaptive beam-forming, beam switching, and antenna switching. System parameters such as available computing power, number of array elements, desired tracking speed, and required hardware cost are very important in selecting the steering approach. These various beam-forming mechanisms are explained and compared, including their advantages and limitations, in the following sections.

25.5.1.1 Antenna Switching This scheme employs directional antennas as array elements. These elements are placed in such a way that the combined array pattern covers the region of interest. At an instant the antenna with highest receive signal strength is selected by a switch as shown in Figure 25.32.

One of the advantages of this system is the computational simplicity; the switching system scans the antennas and selects the antenna with highest receive strength, which also results in a higher tracking speed. Low cost is another advantage of this system because a down converter and an analog-to-digital converter (ADC) with one channel is sufficient at the receiver side. As the beam pattern of an element is fixed, a disadvantage arises when the desired signal and interference happen to be in the beam pattern of the same antenna. Moreover the received signal strength varies when the receiver moves due to scalloping effects, which are related to the roll-off of the antenna pattern as a function of angle [52]. As only a single antenna is active at a given time, this system cannot take advantage of the multipaths whenever the arriving signal is not in the coverage area of the selected antenna and the system is unable to steer the maximum of the beam pattern in the user direction, which may also cause significant losses.

25.5.1.2 Beam Switching In the beam-switching approach, the received signals at each array element are multiplied by a set of precalculated weights and the output with the highest power is selected. This principle is illustrated in Figure 25.33. One advantage of beam switching is that it does not require digital processing. The multiplication with weights can be accomplished in the RF domain using fixed feed networks as discussed in Section 25.1.3. As the weights are fixed, there is negligible computational power

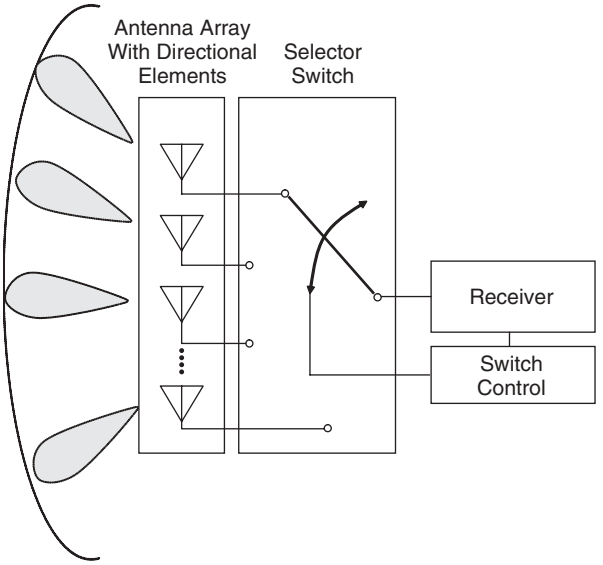


Figure 25.32 Antenna switching system.

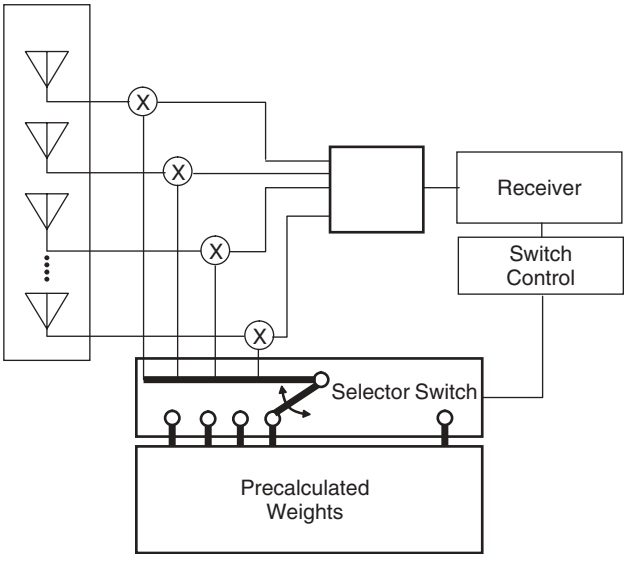


Figure 25.33 Beam switching system.

required for this mechanism, which is one of its main advantages. On the other hand, this mechanism suffers from scalloping effects similar to the antenna switching approach. The other disadvantages are inability to utilize multipath effects and mismatch of user location and beam peak. In Figure 25.34 we show an example set of beam patterns obtained using orthogonal weights for an eight-element standard linear array (SLA). In orthogonal weighting, the weight vector corresponding to each beam is determined so as

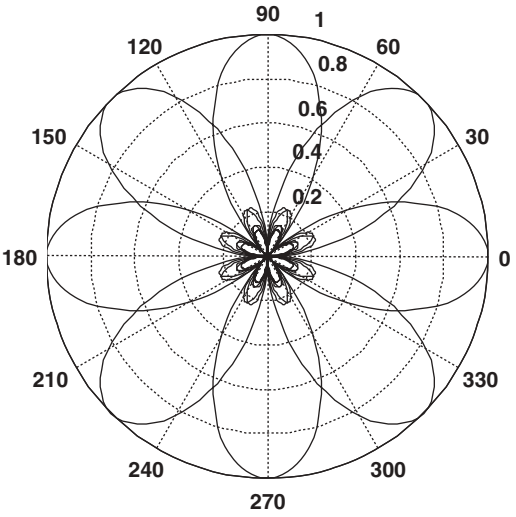


Figure 25.34 Orthogonal beam set for an eight-element array.

to be orthogonal to the weight vectors of all the remaining beams. This results in having the peak of a particular beam correspond to the nulls of the other beams, as shown in Figure 25.34.

25.5.1.3 Adaptive Beam Forming In adaptive beam forming, the weights used for combining the antennas can be adaptively calculated and adjusted to achieve greater performance improvements than what is possible when using switched beam and switched antenna systems. An adaptive beam-former block diagram is shown in Figure 25.35. In this figure, the desired signal corresponds to a local generated replica of the user's signal obtained using a predetermined training sequence.

As can be seen from Figure 25.35, the hardware requirements of adaptive arrays are much larger than switched antenna and switched beam approaches as the baseband converter and analog-to-digital converter (ADC) have to have multiple channels. For increased beam resolution, the number of antennas should be increased, which also increases the computational complexity as well as hardware cost. On the other hand, this approach has an unmatched ability to place the maximum of the beam at the user location and place nulls at interferer locations. If the processing speed is low, however, the combining weights take longer to converge and this can cause tracking errors if the mobile or target is moving with a high speed. For weight calculation, a least mean squares (LMS) algorithm that is very stable and simple to implement is mostly used, which has a computational complexity of $O(N^3)$, where N is the number of antenna elements on the array [53].

25.5.1.4 Adaptive Beam Forming with Reconfigurable Antennas It has recently been reported [54–56] that digital beam-forming techniques can be used to properly modulate (with appropriate weights) the ferroelectric permittivity associated with each microstrip antenna to provide shaped patterns that simultaneously suppress interference signals. This architecture is in the class of reconfigurable antennas. Figure 25.36a

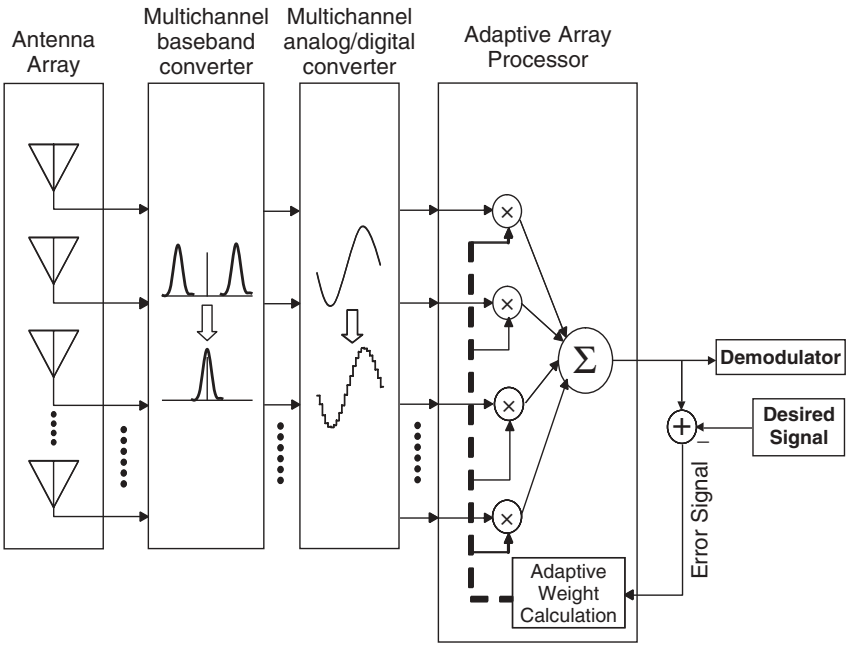


Figure 25.35 Block diagram of adaptive beam-former. Note that the number of channels that the baseband converter and analog-to-digital converter have is equal to number of antenna elements.

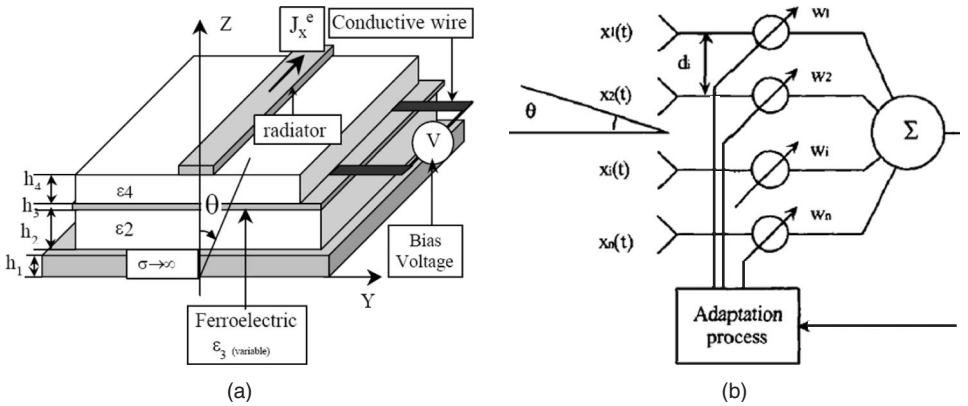


Figure 25.36 (a) Ferroelectric-based reconfigurable microstrip antenna and (b) N-element adaptive array structure. (From Refs. 54–56.)

illustrates a schematic of the antenna element employing a microstrip radiator over a multidielectric ferroelectric substrate. For frequencies just above cutoff for the first higher order mode on the microstrip line, the electromagnetic fields are loosely coupled, resulting in a considerable amount of radiation. Figure 25.36b shows a block diagram of the adaptive array process.

25.5.2 Hybrid Smart Antenna and Beam-Forming Technique

Taking the advantages and limitations of existing beam-steering mechanisms into account, a hybrid smart antenna system that combines the advantages of the adaptive and the antenna switching mechanisms for beam steering is proposed in Zhang et al. [51]. The hybrid beam-forming mechanism relies on the use of directional antenna elements in an array and the employment of a smart switch that connects the elements with the highest receive signal strength to the receiver as shown in Figure 25.37. This way, the narrow beamwidth of directional antennas is combined with the advantages of adaptive beam forming.

In traditional antenna arrays, the number of elements is often increased to narrow the width of the main lobe resulting in increased computational complexity and cost of the antenna array system. The hybrid beam-forming mechanism avoids this by employing directional antennas with high directivity and the smart switch does not include the antennas with small received signal in the weight calculation process. The reduced number of antennas included in the weight calculation results in tremendous savings in cost by decreasing the number of ADCs and the required down conversion channels. Moreover, the computational complexity is highly reduced, as the complexity of the LMS algorithm is proportional to the third power of the number of antennas selected for adaptive beam forming.

In the antenna selection process, the smart switch selects the subset consisting of neighbor elements with highest total power. It was suggested [49, 51] that a subset of elements that receive the strongest signals should be selected but, to prevent possible beam splitting and the formation of high levels of sidelobes, it was necessary to limit selection to neighboring elements with highest total power.

The hybrid beam-forming mechanism combines the advantages of the low computational power requirement of antenna switching systems with the increased beam-steering capacity of adaptive beam-forming systems to provide better tracking speed, decreased

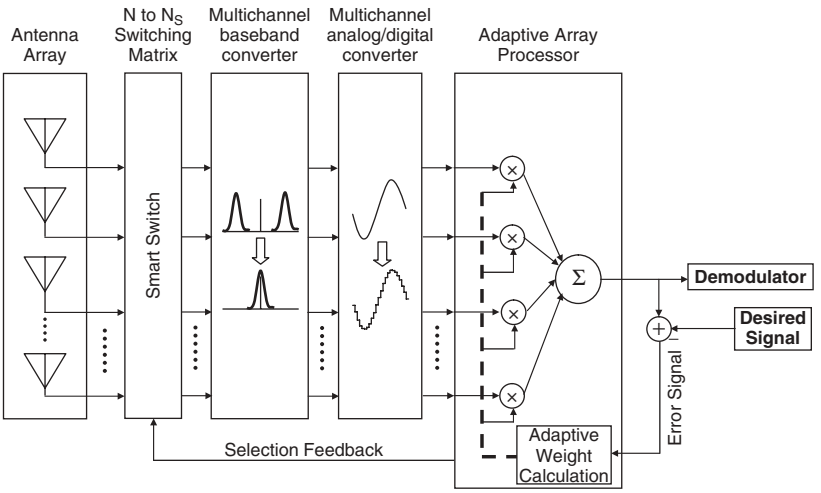


Figure 25.37 Block diagram of hybrid beam former. Note that the number of channels that baseband converter and analog-to-digital converter have (N_S) is lower than the number of antenna elements (N).

beamwidth, and lower hardware cost. In the following sections the formulation of the hybrid beam-forming algorithm (Section 25.5.2.1); the performance comparison of the hybrid beam-forming algorithm with switched beam and adaptive beam-forming mechanisms are described with the help of computer simulations done in an additive white Gaussian noise (AWGN) wireless propagation environment. Moreover, the simulated performance of an antenna array employing CPW–CTS as array elements is discussed and compared with an antenna array using omnidirectional array elements. The experimental results of a smart antenna system testbed using the hybrid beam-forming algorithm for beam steering are included and compared with the simulation results in Section 25.5.2.3.

25.5.2.1 Formulation for Hybrid Beam Forming For mathematical formulation, a linear antenna having N elements spaced by Δx meters is considered. A plane wave $d(t)$ impinges on the antenna array at an angle as shown in Figure 25.38. We assume that the incoming wave $d(t)$ is narrowband such that its complex envelope does not change throughout the array. Thus beam forming can be accomplished by simply adjusting the phases of antenna outputs using complex antenna weights. If interference and multipath are neglected, the received signals at the N antenna elements can be represented in vector form as follows:

$$\mathbf{u}(t) = [u_1(t) \quad u_2(t) \quad \cdots \quad u_N(t)]^T \tag{25.14}$$

where $u_k(t)$ represents the signal received at the kth antenna.

By choosing the first element as the reference antenna, the received signal $\mathbf{u}(t)$ can be written

$$\mathbf{u}(t) = \mathbf{g}(\theta, \phi) \times \mathbf{a}(\Delta x, k_x) \times e^{-j\omega\tau} \times d(t) + \mathbf{n}(t) \tag{25.15}$$

Here $\mathbf{n}(t)$ is the $N \times 1$ vector of noise samples, τ is the path delay of the reference element from the transmitter, ω is the angular frequency of the transmitted signal,

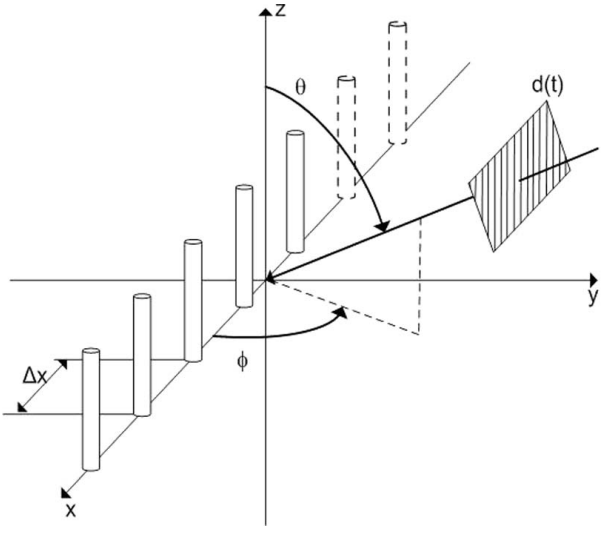


Figure 25.38 Antenna array and plane-wave configuration.

$k_x = -(2\pi/\lambda)\sin\theta\cos\phi$ is the wavenumber in the x-direction, and $\mathbf{g}(\theta, \phi)$ is the $N \times N$ diagonal array gain matrix,

$$\mathbf{g}(\theta,\phi) = \begin{bmatrix} g_1(\theta,\phi) & 0 & 0 & \dots & 0 \\ 0 & g_2(\theta,\phi) & 0 & \cdots & 0 \\ 0 & 0 & g_3(\theta,\phi) & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & g_N(\theta,\phi) \end{bmatrix} \tag{25.16}$$

Each diagonal element in the matrix is the antenna gain of the corresponding array element in the denoted arrival direction. $\mathbf{g}(\theta, \phi)$ is the identity matrix and is not dependent on the arrival direction for an antenna array with omnidirectional elements. $\mathbf{a}(\Delta x, k_x)$ is called the steering vector, which can be expressed as [56]

$$\mathbf{a}(\Delta x, k_x, \theta, \phi) = [1 \quad e^{j\Delta x k_x} \quad e^{j2\Delta x k_x} \quad \dots \quad e^{j(N-1)\Delta x\ k_x}]^T \tag{25.17}$$

As mentioned previously, the hybrid approach employs a smart switch for selecting a subset of antennas with maximum signal strength and doing beam forming only using this subset. To prevent possible beam splitting, the formation of high levels of sidelobes, and to keep antenna spacing constant, the subset should consist of neighboring elements and the smart switch selects the subset with highest total power. Assuming that the smart switch selects N_S out of N array elements for further processing, the received signals at the selected array elements can be expressed as

$$\mathbf{u}_h(t) = \mathbf{S} \cdot \mathbf{u}(t) = \mathbf{S} \cdot [\mathbf{g}(\theta, \phi) \times \mathbf{a}(\Delta x, k_x) \times u_1(t) + \mathbf{n}(t)] \tag{25.18}$$

where $\mathbf{S}$ is an $N \times M$ selection matrix with elements

$$s_{kl} = \begin{cases} 1, & l = p_k \\ 0, & l \neq p_k \end{cases}$$

and $\mathbf{P} = [p_1 \quad p_2 \quad \cdots \quad p_N$ is the vector holding the indices of selected antennas.

The output of the beam former is an estimate of the transmitted signal $d(t)$ and can be written

$$\hat{\mathbf{d}}(t) = \mathbf{w}^H(t) \times \mathbf{u}_h(t) \tag{25.19}$$

where the superscript H is the Hermitian operator and $\mathbf{w}(t)$ is the $N_S \times 1$ weight vector. The complexity and convergence speed of the adaptive algorithm used to calculate the weight vector depends on N_S. Therefore the convergence speed is greatly improved by not including the elements with small receive signal strength in the process of weight calculation. In the LMS algorithm the weights are updated as follows:

$$\mathbf{w}(t + \Delta t) = \mathbf{w}(t) + \mu \times e^*(t) \times \mathbf{u}_h(t) \tag{25.20}$$

$$e(t) = d(t) - \mathbf{w}^H(t) \times \mathbf{u}_h(t) \tag{25.21}$$

In Eq. (25.20), μ is the parameter determining the speed of convergence of the LMS algorithm; the higher the μ the faster is the convergence. On the other hand, there is an

upper limit of μ that is determined by the largest eigenvalue of input correlation matrix; if μ exceeds this value the algorithm becomes unstable.

25.5.2.2 Simulation Procedure and Results The following is a step-by-step description of the simulation procedure.

A. Obtaining weights:
   1. Select an arrival angle ϕ_a, and set $\theta = \pi/2$ for simplicity.
   2. Define antenna radiation patterns $g_k(\theta, \phi)$ for $k = 1, 2 \ldots, N$.
   3. Select the transmitted waveform $d(t)$ and the SNR.
   4. Generate $\mathbf{u}(t)$ according to Eq. (25.15).
   5. Measure the signal powers received at each element by correlating $u_k(t)$ with $d(t)$.
   6. Select the neighboring M elements with highest total signal power for beam forming.
   7. Apply adaptive LMS algorithm given in Eq. (25.20) to the selected subset of receive signals starting with some arbitrary initial weights (such as zero) and find the combining weights $\mathbf{w}_{\phi a}$.

B. Plotting the beam pattern:
   1. Define a set of equally spaced arrival angles ϕ_k between 0 and π.
   2. For each ϕ_k:

      Generate $\mathbf{u}(t)$ according to Eq. (25.15) with $\mathbf{n}(t) = \mathbf{0}$.
      Calculate $\hat{\mathbf{d}}_{\phi_k}(t)$ using $\mathbf{w}_{\phi a}$ as the combining weights.
   3. Plot the powers of $\hat{\mathbf{d}}_{\phi_k}(t)$ versus ϕ_k.
   4. The ϕ_k corresponding to the highest $\hat{\mathbf{d}}_{\phi_k}(t)$ power is $\hat{\phi}_a$

To compare the performance of the hybrid, switched, and fully adaptive smart antenna systems, the SNR is assumed fixed at 20 dB throughout the simulations. It is expected and it was proved through simulations that as the SNR decreases, the performances of both the hybrid and the adaptive smart antenna systems degrade. However, since adaptive beam-forming algorithms are used, the effect observed was not as significant as that noticed if other nonadaptive digital beam formers were utilized [57]. It is worth mentioning that the frequency at which a new subarray selection decision is applied for the hybrid case does not change throughout simulations. Results show that a full steering capability within the range of interest can be achieved using different adaptive beam-forming techniques. However, in order to illustrate the comparisons drawn in the figures that follow, the location of the mobile station (MS) is assumed at $\pm 10^\circ$ off the broadside of the array. Unless otherwise mentioned, the radiation pattern of the antennas involved is considered to take the form of $\sin(\phi)$ (HPBW $= 90^\circ$). Additional simulations were made where the measured radiation pattern from the CPW–CTS antenna was taken into account and some obtained results are also reported later in this chapter.

Figure 25.39 illustrates the performance using the recursive least squares (RLS) adaptive algorithm. In this simulation, an adaptive antenna array system with eight elements is compared to a switched system in addition to a hybrid smart antenna system utilizing two and six elements out of the total eight elements. Since, in general, similar results were

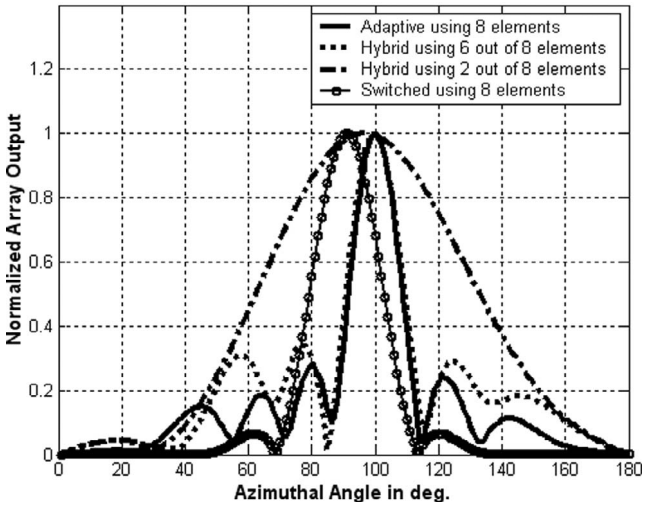


Figure 25.39 Comparison between the adaptive, switched, and hybrid systems using RLS algorithm for an AWGN channel and a MS located at 100° with respect to the array axis (10° from broadside). Elements have 90° HPBW.

obtained using different adaptive beam-forming algorithms assuming the same simulation scenarios, results from the LMS beam-forming algorithm are presented.

In Figure 25.39, it is shown that with only two elements out of an eight-element array, the actual location of the MS can be correctly detected and hence a good BER that compares well with that when using the fully adaptive antenna array structure can be obtained [49, 51]. It is worth emphasizing that the observed overall radiation pattern is often formed by neighboring elements and this leads to a radiation pattern characterized by a dominant main beam. Otherwise, the neighboring element with the next strongest signal is selected in order to help reduce possible beam splitting, avoid large sidelobes, and maintain a dominant main beam pattern. It may be noted, however, that with only two elements, a relatively broad main beamwidth is obtained using the proposed hybrid approach, which in turn would allow interferences and multipath components to be captured, resulting in decreased overall system performance and a BER that compares well with the adaptive system only in low interference and multipath signal environments. This agrees with and explains the system level simulation results previously conducted [49]. With more than half the elements selected in the hybrid approach (six out of eight), however, less beamwidth, as narrow as that of the adaptive, is achieved. An increased side lobe level (SLL) is also observed as another drawback of the hybrid approach. Moreover, for this simulation scenario, one can observe that while the proposed hybrid approach converged to the correct direction of the user's location, the best beam selected by the switched system approach was not in the actual location of the incident direction. This in turn affects the BER and the quality of the provided wireless service and points to a drawback when using the switched beam smart antenna approach.

For high resolution applications where a narrow beam is required—and hence the number of array elements needs to be increased—the hybrid and the adaptive systems perform nearly the same with only half the number of elements being selected in the

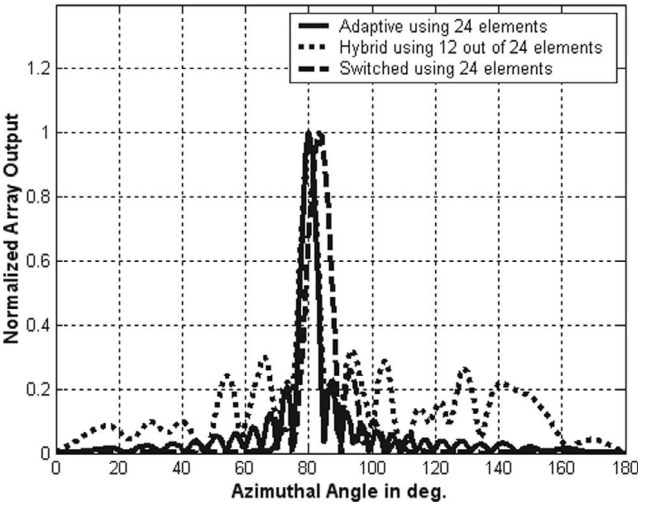


Figure 25.40 Comparison between the adaptive, switched, and hybrid smart antenna systems using the LMS algorithm for an AWGN channel and a MS located at 80° off the array axis. Elements have 90° HPBW.

proposed hybrid case. This result is illustrated in Figure 25.40. Fortunately, this observation is significant because as the number of antenna elements being selected for the adaptation process decreases, latency and cost of implementation reduce proportionally as a result of the reduced amount of hardware needed, including the number of ADCs involved.

More interestingly, that if more directional elements are being used in the antenna array being deployed, the increased SLL problem, which may be observed in Figure 25.40, can be considerably overcome as shown in Figure 25.41. For a hybrid system utilizing 6 out of 12 elements, it is shown that with the reduction in the beamwidths of the radiating elements from 90° to 40°, significant reduction in the SLL may be achieved. Similar results can be observed by comparing the dotted curves in both Figures 25.40 and 25.42 where different HPBWs are utilized for the same number of selected elements (12 out of 24). It is worth mentioning at this point that as the total number of elements used in deploying the antenna array increases (i.e., narrow beam applications), the performance of the hybrid system utilizing the same corresponding ratio of elements (e.g., one-half) will also improve.

Furthermore, simulation results showed that for high resolution applications using highly directional elements (∼40° HPBW or less), the performance of the hybrid system compares well with that of the fully adaptive using only about one-third of the antenna array elements in the adaptation process. An example is illustrated in Figure 25.42, where the dashed curve shows the performance using 8 out of 24 elements with ∼40° HPBW. Comparing with the dotted curve in Figure 25.40, one can observe that the good performance of the hybrid array, compared to that of the fully adaptive system, could be achieved by selecting a subset of only one-third of the elements and using higher directional elements. This in turn implies that the hybrid smart antenna system indeed has the potential to further reduce the cost with careful design of the involved antenna array elements.

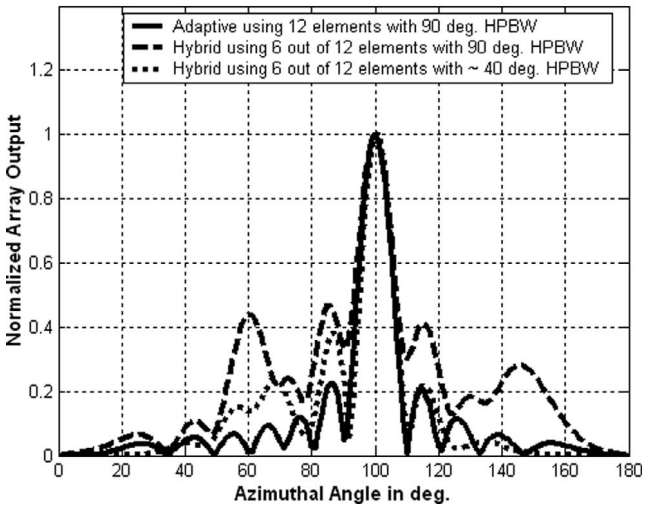


Figure 25.41 Comparing the adaptive system with twelve 90° HPBW elements to that of the hybrid system utilizing 6 out of 12 elements with HPBW of 90° and 40°. An AWGN channel and the LMS algorithm are used.

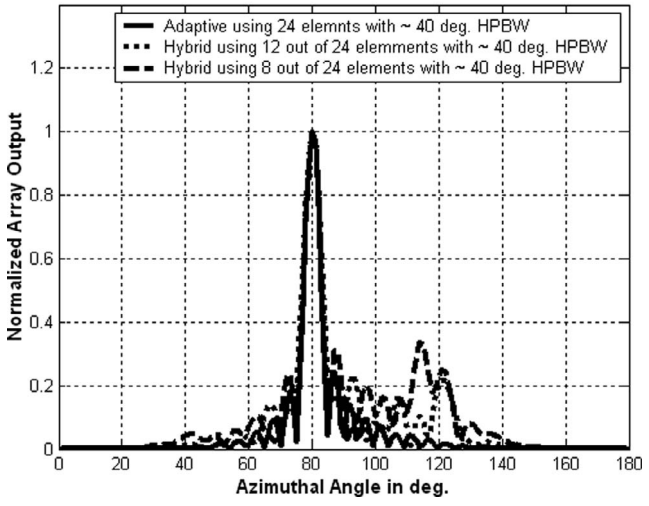


Figure 25.42 Comparing the adaptive system utilizing 24 element with 40° HPBW to that of the hybrid system utilizing 12 and 8 elements out of the 24 elements. An AWGN channel and the LMS algorithm are used.

From the results discussed thus far, it may be observed that a low cost, highly directional, completely scalable antenna design provides the best candidate for implementing the newly proposed hybrid smart antenna system [49, 51]. An avenue for implementing directive elements in the hybrid smart antenna approach is to use CPW–CTS antennas. In this case, it may be possible to use a single stub or an array of multiple stubs to

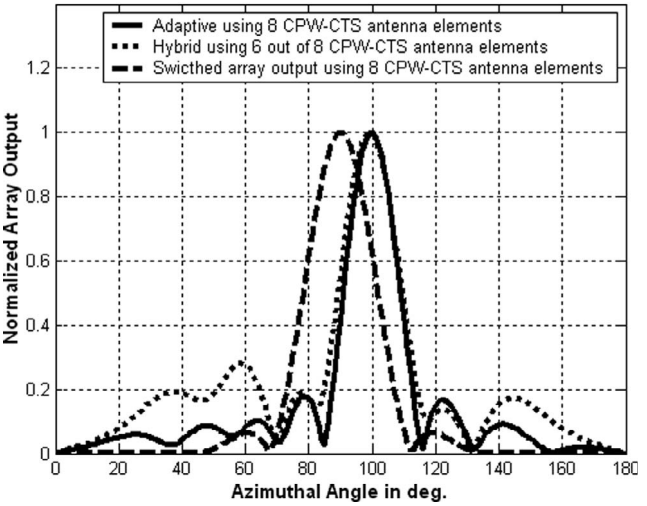


Figure 25.43 Comparing the adaptive, switched, and hybrid systems using CPW–CTS antenna elements and RLS algorithm. A signal incident from a mobile station (MS) located at 10° off the array broadside and AWGN channel are assumed.

increase directivity without complicating the feed structure as in both cases a single feed line will be used. The advantages, features, measured versus simulated radiation pattern, return loss, and insertion loss of the CPW–CTS antenna are described in more detail in Section 25.5.3. Simulations using different adaptive beam-forming algorithms taking into account the measured radiation pattern of the CPW–CTS prototype were therefore carried out to illustrate the implementation procedure. Some of the obtained results are illustrated in Figures 25.43 and 25.44.

Figure 25.43 shows that the hybrid and the fully adaptive results are almost identical in the main lobe and differences only appear in the SLL. The performance of the hybrid approach is much better than that using a CPW–CTS switched beam array approach, where the best selected beam was 10° off the actual direction of the incident signal. Since CPW–CTS antennas provides a more directive pattern ($\sim 74.8^\circ$ HPBW) than the assumed $\sin(\phi)$ one (HPBW $= 90^\circ$), a more confined array output pattern in the range of interest with slight reduction in SLL can be achieved. This result is illustrated by comparing Figures 25.43 and 25.38 or by comparing the two curves in Figure 25.44 for 10 CPW–CTS elements and 10 $\sin(\phi)$ elements selected out of 20 in the hybrid approach. Furthermore, for narrow beam applications, simulations showed a hybrid system performance that compares well with that of a fully adaptive can be achieved by selecting a subset of less than half the number of elements when utilizing CPW–CTS antennas in building the array (e.g., 6 out of 16 or 8 out of 20 as shown in Figure 25.44). This indeed emphasizes the potential advantage of reducing the implementation cost of smart antenna systems using the hybrid approach, especially for narrow beam applications.

25.5.2.3 Prototype Implementation and Experimental Results A smart antenna receiver prototype is built to verify the accuracy of the proposed hybrid beam-forming algorithm. This prototype consists of an eight-channel down converter, which converts 2.4-GHz RF frequency signals to an IF signal with a bandwidth of

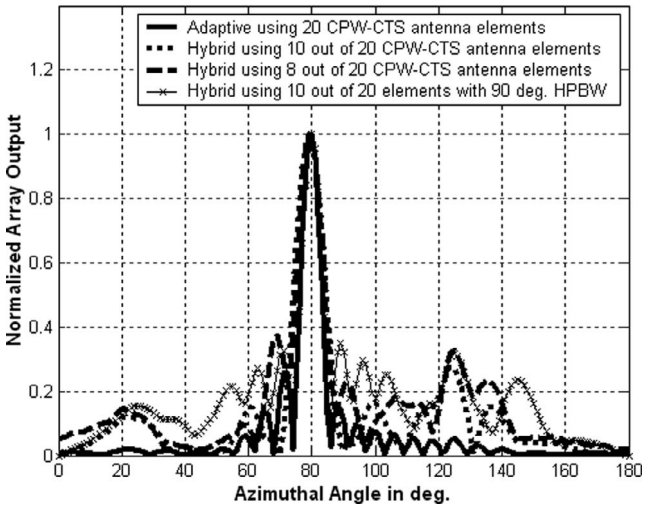


Figure 25.44 Comparing adaptive system using 20 CPW–CTS antenna elements to hybrid system utilizing 8 and 10 CPW–CTS elements or 10 elements with HPBW $= 90^{\circ}$ out of 20. AWGN channel and LMS beam-forming algorithm are used and MS is assumed at 80° with respect to the array axis.

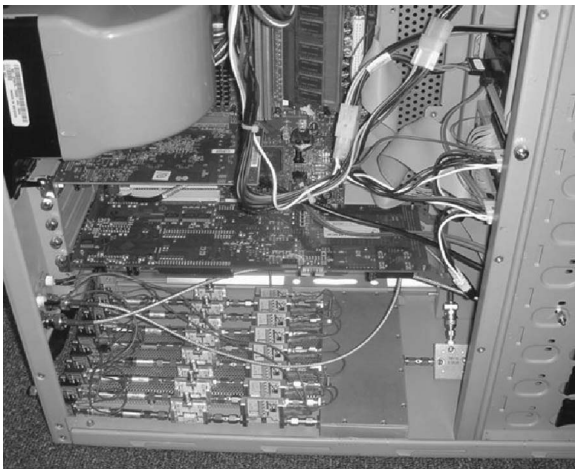


Figure 25.45 Inside view of the case housing the eight-channel RF–IF chain, ADC, and the data capturing and processing computer.

10 MHz. This IF signal is then digitized by the eight-channel analog-to-digital converter (ADC) card using a 25-MHz sampling rate. Then these digital signals are fed to the beam-forming algorithm to obtain the weight vector and corresponding AoA. As shown in Figure 25.45, the digital processing system (a PC running Matlab) and the down-converter system along with the ADC are put into a custom computer case for the sake of compactness. In the design of the down converter, the same local oscillator

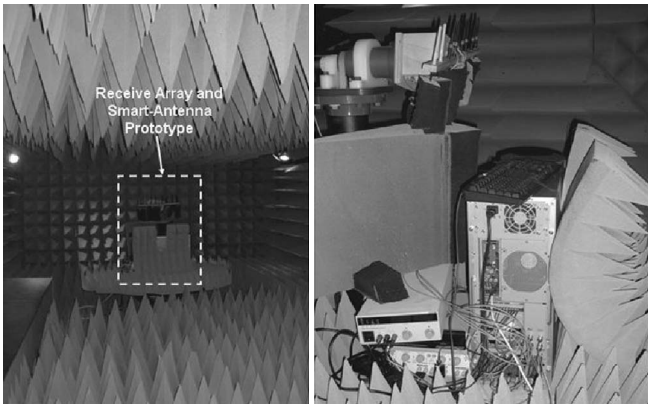


Figure 25.46 Receive array and the smart antenna array prototype with eight-channel RF–IF chain, ADC, and the data capturing and processing computer inside the indoor antenna range.

is used for all channels through the use of a power splitter to overcome the phase and frequency discrepancies. For a correct calculation of the AoA, the phase differences due to different element delays are canceled by calibration weights. These weights are obtained by sending signals in the $(\theta, \phi) = (\pi/2, \pi/2)$ direction and adaptively calculating the beam-forming weights. The calibration is done inside an anechoic chamber (see Figure 25.46), as the presence of multipath will adversely affect the accuracy of the system. The following is the step-by-step description of the experimental measurement procedure:

A. Obtaining weights
   1. Set up the system in a controlled multipath environment, for example, an antenna range with a positioning system for precise measurement of the arrival angle.
   2. Generate the data signal $d(t)$ at the transmitter.
   3. Calibrate the system.
      (i) Set the arrival angle $\phi_a = \pi/2$ and $\theta = \pi/2$ by rotating either the transmitter or receiver.
      (ii) Receive, down convert, and digitize $\mathbf{u}(t)$ and select all antennas.
      (iii) Find amplitude normalization coefficient c_k for each antenna such that $\|u_k(t)/c_k\| = g_k(\pi/2, \pi/2)$ to compensate the nonuniform amplifier gains in the system.
      (iv) Apply adaptive LMS algorithm given in Eq. (25.20) with some initial weights to find the optimal combining weights $\mathbf{w}_c$.
      (v) Calculate the phases $\angle \mathbf{w}_c$ needed to compensate the phase differences in the down-converter system.
   4. Change the arrival angle ϕ_a to a desired value.
   5. Receive, down convert, and digitize $\mathbf{u}(t)$.
   6. Calibrate the system by dividing each $u_k(t)$ with $c_k \times e^{j\angle w_{ck}}$.

7. Measure the signal powers received at each element by correlating $u_k(t)$ with $d(t)$.
   (i) Apply adaptive LMS algorithm given by Eq. (25.20) starting with some initial set of weights and find the combining weights $\mathbf{w}_{\phi a}$.

B. Plotting the beam pattern: see description at the end of Section 25.5.2.

In conducting the experiment, a continuous wave (CW) of frequency 2.4 GHz is transmitted by a horn antenna and the receiving array has eight isotropic (sleeve antenna) elements. In all of the experiments, the elevation angle θ is selected to be $\pi/2$ and the azimuth angle was kept as the angle variable ϕ, which is calculated from the obtained combining weights and compared to the actual angle. The received bandpass signal is converted to its lowpass equivalent by calculating its Hilbert transform. The least mean square (LMS) algorithm is used for adaptive calculation of the beam-forming weights.

Figure 25.47 shows the experimental results for the beam pattern generated when the AoA of the incoming wave is 100°. Results are shown for different values of the selected number of elements (N) in the subarray. As may be seen, the AoA is correctly found even with 3 out of 8 elements selected. It may also be noted that the width of the main lobe as well as the number of sidelobes increase when the number of antennas selected is decreased, as predicted in our earlier discussion of the simulation results. Figure 25.48 shows the polar plot of the beam pattern when all array elements are selected, together with the polar plot for the case when only four elements of the array are selected. Figures 25.49 and 25.50 show the beam patterns when the AoA of the incoming signal is changed to 125°. Again, the AoA is correctly found by the weights. The effect of decreasing the number of array elements is the same as in Figure 25.47, which is consistent with the simulation results. These experiments have further shown that the developed prototype has the accuracy of distinguishing small changes in AoA within a few degrees as shown in Figure 25.51.

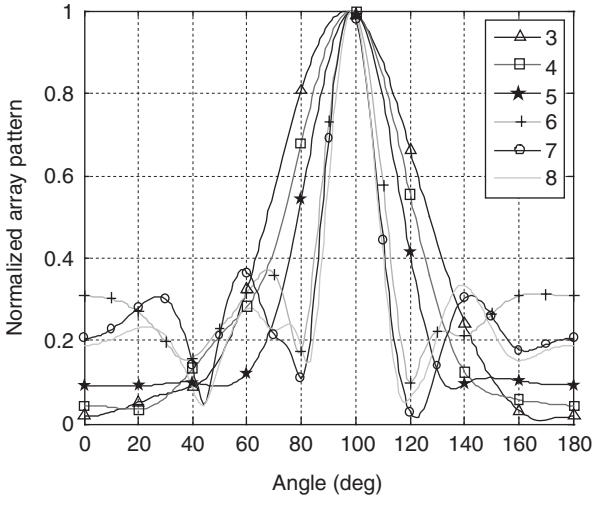


Figure 25.47 Array pattern for AoA = 100° with different number (N) of antennas used in the subarray.

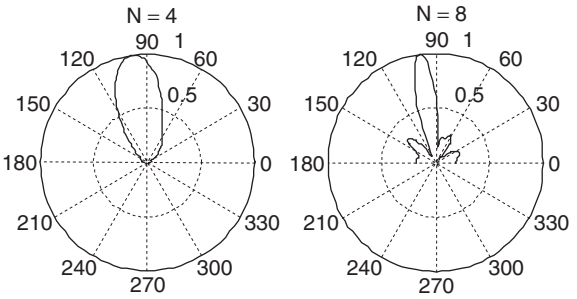


Figure 25.48 Polar plot of array pattern for AoA = 100° with 4 and 8 antennas used in the subarray.

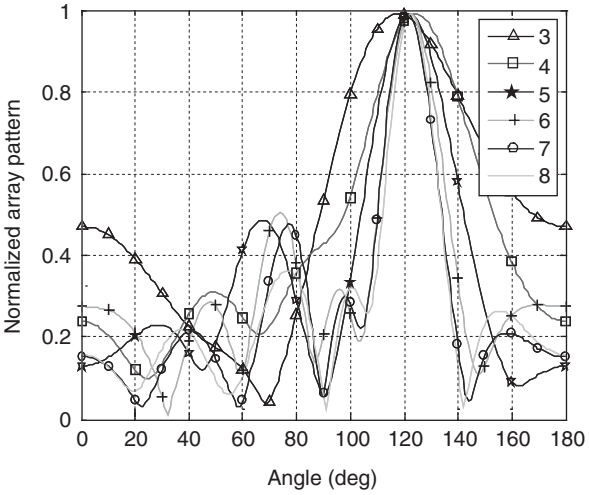


Figure 25.49 Array pattern for AoA = 125° with different number (N) of antennas used in the subarray.

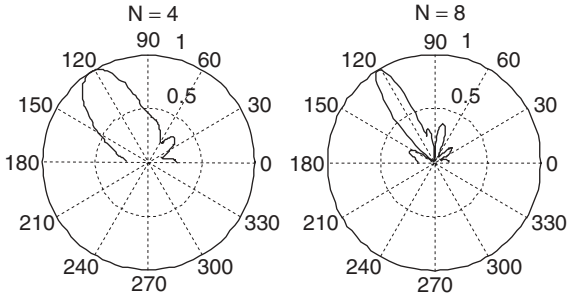


Figure 25.50 Polar plot of array pattern for AoA = 125° with 4 and 8 antennas used in the subarray.

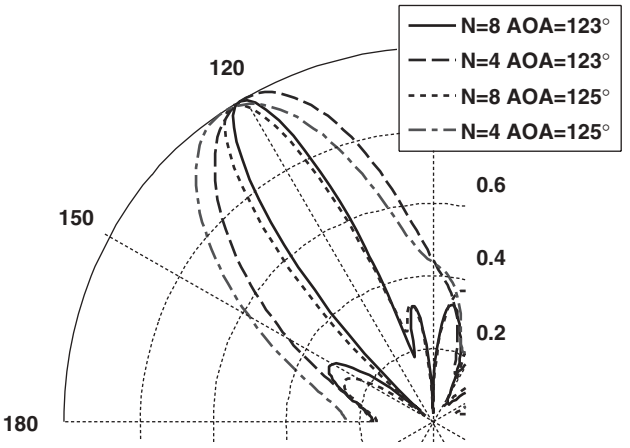


Figure 25.51 Polar plot of array patterns for AoA = 123° and AoA = 125° with different number (N) of antennas used in the subarray.

25.5.3 Comparative Study of Analog and Digital Beam-Forming Techniques

This section briefly highlights trade-offs between analog and digital beam-forming techniques. In essence, it is shown that an optimum performance could be achieved using a hybrid approach, whereby large beam-steering angles could be done digitally and the fine angle tuning/small angle beam steering could be accomplished using analog methods. The following is a brief discussion of some of the comparative issues.

25.5.3.1 Quantization Effects in Digital Beam-Forming Systems This section examines the effect on the directivity and array factor performance of linear phased arrays for different numbers of phase shifter bits. Table 25.1 summarizes the average array-factor characteristics for an eight-element array utilizing one to five bit phase shifters. As shown in Table 25.1 [58, 59], the maximum sidelobe level (SLL) as well as the directive loss improves with increasing phase shifter bits (approaching analog). The average scan angle deviation, $\Delta\theta$, also improves with increasing phase shifter bits. This, however, will be at the expense of cost as well as complexity and increased losses.

TABLE 25.1 Average Results of Maximum SLL, Directivity Loss, and Scan Angle Deviation for an Eight-Element Array Using Phase Shifters of Different Number of Bits[a]

	Number of Phase Shifter Bits				
Parameter	1 Bit	2 Bits	3 Bits	4 Bits	5 Bits
Maximum SLL (dB)	−6.89	−6.97	−10.19	−11.62	−12.18
Directivity loss (dB)	3.15	0.82	0.2	0.047	0.012
Scan angle deviation, $\Delta\theta$ (degrees)	2.95	1.76	0.98	0.46	0.23

[a] See Ref. 58.

TABLE 25.2 Average Results of Maximum SLL, Directivity Loss, and Scan Angle Deviation for a 64-Element Array Using Phase Shifters of Different Number of Bits[a]

	Number of Phase Shifter Bits			
Parameter	2 Bits	3 Bits	4 Bits	5 Bits
Maximum SLL (dB)	−9.18	−12.72	−12.97	−13.13
Directivity loss (dB)	0.9	0.22	0.054	0.013
Scan-angle deviation, $\Delta\theta$ (degrees)	0.064	0.02	0.014	0

[a]See Ref. 58.

Table 25.2 summarizes the phased array performance for a 64-element phased array for various phase shifters [58]. As mentioned with the smaller array, the overall characteristics are improved with an increase in the number of bits, reaching the performance of analog phase shifters. The average maximum sidelobe level stays under −12.5 dB from the three bit phase shifter. The average directivity losses drop below 0.1 dB from four bits.

As the size of the array increases, the number of bits may be reduced to maintain adequate beam-steering performance [9]; therefore a design trade-off is required to meet the specific phased array antenna parameters. For reduced antenna elements, increasing bit size is required which results in increased phase shifter complexity. For switched line digital phase shifters, the real estate of the integrated circuit increases due to the large amount of switches and the large area required for lines of different lengths.

25.5.3.2 Frequency Dependence Digital constant-phase and switched line phase shifters bring a frequency dependence to the scan angle, each type generating additional sidelobes (as well as elevated levels) as the frequency deviates from the design frequency. When broadband signals are to be radiated by a phased array system, the use of different phases for the excitation of the various elements in the array will generally result in some signal distortion unless the phase shifters are true time-delay (analog) devices [59–61]. The use of true time-delay phase shifters in the feed network gives superior performance, keeping the scan angle constant over the frequency range of the array elements.

As may be seen from the above discussion, using a combination of digital phase shifters for each element and a true time-delay unit at the subarray level considerably increases the operating bandwidth compared to that of employing purely digital phase shifters [58].

25.6 ULTRAWIDEBAND HYBRID EBG/FERRITE GROUND PLANE FOR LOW PROFILE ARRAY ANTENNAS

Ultrawideband and low profile phased array antennas are considered a most critical component in the development of lowband systems such as airborne foliage penetrating radar and EW applications. It is imperative that these antennas are ultrawideband beginning in the hundreds of megahertz to provide the necessary foliage penetration and low profile to allow the systems to fit compactly in the aircraft. Additionally, it is desirable that the antenna systems provide unidirectional radiation to prevent the radiation back into the aircraft that is seen with bidirectional radiating antennas. For these applications it

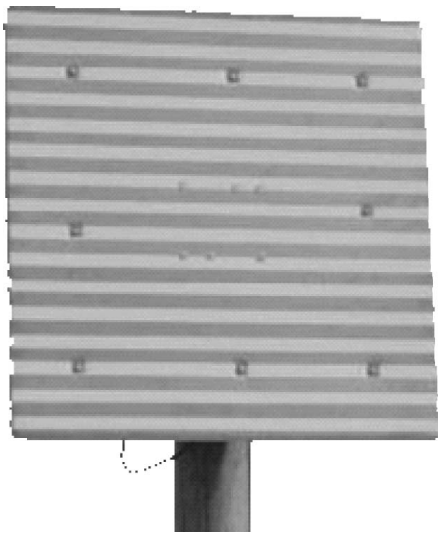


Figure 25.52 Raytheon's long slot array antenna [62].

is clear that an ultrawideband and low profile ground plane needs to be developed and implemented to achieve the desired specifications and required performance.

Recently, an ultrawideband long slot array antenna was invented, analyzed, and experimentally tested at Raytheon. Figure 25.52 shows a photo of this antenna design, which reported a potential 100:1 bandwidth beginning in the hundreds of megahertz if no back plane is attached to the bidirectional aperture [62–65]. Like many antenna systems that operate at these relatively lower frequencies, the radiation pattern of the developed long slot array antenna is bidirectional. Therefore to provide the unidirectional radiation pattern desirable for airborne types of applications, a ground plane that is low profile while still maintaining the ultrawideband operation beginning in the hundreds of megahertz needs to be developed.

Currently, a variety of ground plane options exist. Yet each option comes with its own set of advantages and limitations. These options include electric conductors, absorbing materials, and artificial magnetic conductors (AMCs). Electric conductor ground planes are the traditional choice for ground plane implementation and are still widely used today in many antenna applications. When using an electric conductor ground place with antennas that have currents parallel to the ground plane, the ground plane must be spaced at a distance of $\lambda/4$ (midband) behind the antenna to provide constructive interference between the fields radiated from the antenna and the fields reflected off the ground plane. As it is clear to see, this spacing requirement comes with limitations. Two areas where electric conductor implementation limits antenna performance include profile reduction and the operational bandwidth. Absorbing material is sometimes placed between the ground plane and the antenna. This absorbing material offers profile reduction without introducing excessive reflections and destructive interference. The implementation of absorbing material, however, comes at the cost of energy waste (half of the power radiated from the antenna is lost to the absorbing material). More importantly, the bandwidth where the absorbing material performs optimally is limited and often occurs at lower frequencies. A current widely discussed research area is related

to the design and optimization of electromagnetic bandgap (EBG) structures, particularly those that emulate magnetic conductors, referred to as AMCs. AMC ground planes offer the most promising design for low profile ground planes for antennas that have currents parallel to the ground plane. However, as with other ground plane options, AMC ground planes have their limitations. AMC ground planes are designed using various types of structures that provide a high impedance surface. These designs have inherently limited bandwidths because the frequency region where the structure offers high impedance and thus emulates a magnetic conductor is achieved in a limited range around a resonant frequency.

As described previously, each of the aforementioned ground plane options has limitations. To achieve desirable performance with minimal limitations, it is logical to lean toward a hybrid design that implements the combination of more than one ground plane option. In this section a novel ultrawideband and low profile hybrid EBG/ferrite ground plane is presented. The presented hybrid ground plane design offers ultrawideband operation beginning in the hundreds of megahertz with one design version offering an operational bandwidth exceeding 40:1 starting at 120 MHz. The use of this ground plane will allow for the practical and effective implementation of the new highly sophisticated and ultrawideband antenna systems developed at Raytheon.

This section begins with the descriptions of the design and background of four versions of the hybrid EBG/ferrite ground plane technology. Following this, a description of the simulation procedures and considerations using Ansoft's HFSS software are presented. This is followed by the presentation of the reflectivity and phase analyses simulation results for each of the four hybrid ground plane design versions. Next, to verify the effectiveness of the hybrid structure to perform as an ultrawideband ground plane, simulation results of the hybrid ground plane implemented with two bidirectional radiating antennas, including dipole antennas and Raytheon's long slot array antenna, are presented. Then the procedure for calculating the effective complex permittivity and permeability of the hybrid ground plane is described. Finally, descriptions of the experimental testing procedures as well as the experimental results are presented.

25.6.1 Hybrid EBG/Ferrite Ground Plane Design

The design of the novel hybrid EBG/ferrite ground plane consists of the implementation of ETS-Lindgren's FT-2000 ferrite tile absorber [66] and the mushroom EBG structure described in Ref. 67. It should be noted that an AMC design presented in Ref. 68 also implements ferrite material in an EBG structure. However, the basis behind this implementation as well as the resultant effects are quite different from that of the hybrid ground plane presented here. ETS-Lindgren's ferrite absorber is limited to operation in the low frequency range of hundreds of megahertz and the mushroom EBG structure is limited to operation around a certain bandwidth that does not exceed 3:1. The hybrid ground plane design works by using the ferrite absorber component of the design to absorb the antenna's back radiation in the lower frequency band while the EBG structure component of the design is used to provide a desirable phase for the reflected radiation in the frequency band beyond the operational region of the ferrite absorber. As a result, unidirectional radiation is obtained without performance degradation over a bandwidth that is larger than what is provided by either the ferrite absorber or the EBG structure alone.

The type of EBG structure implemented in the hybrid ground plane is the mushroom structure presented in Ref. 67. It consists of a substrate that is backed by an electric

conductors ground plane with a lattice of electric conductive patches on the surface of the substrate that are connected to the ground plane with conductive vias. The design of this structure creates an effective parallel combination of inductances and capacitances. Therefore a high impedance results at the resonant frequency of the equivalent LC circuit. In a certain band around the resonant frequency of the equivalent LC circuit, the structure maintains its high impedance and provides the reflected radiation with a phase in the range between $\pm 90^{\circ}$ and hence does not cause destructive interference with the forward radiation. The structure's resonant frequency and the frequency band surrounding the resonant frequency depend on various factors including the thickness of the structure, the permittivity and of the substrate, the surface patch dimensions and spacings, and the radii of the vias.

To determine the effective operational bandwidth of ETS-Lindgren's FT-2000 ferrite tile absorber alone with an electric conductor ground plane backing, simulations were conducted using FT-2000 material properties provided by ETS-Lindgren in the frequency range from 100 MHz to 2 GHz. The material data provided by ETS-Lindgren for a few frequency points including 100 MHz, 500 MHz, 1 GHz, and 2 GHz are given in Table 25.3. The reflectivity and phase analyses results of Ansoft's HFSS simulations for ETS-Lindgren's FT-2000 alone with an electric conductor ground plane backing are shown in Figure 25.53. As may be seen in Figure 25.53, ETS-Lindgren's FT-2000 alone offers desirable performance (reflectivity less than -14 dB) over a 9.5:1 bandwidth ending at 950 MHz. The -14-dB criteria level was set to ensure that the reflected

TABLE 25.3 Material Properties of ETS-Lindgren's FT-2000 Ferrite Absorber

Frequency (GHz)	Relative Permittivity	Relative Permeability	Dielectric Loss Tangent	Magnetic Loss Tangent
0.1	10.919	15.957	0.00220	4.1560
0.5	10.948	0.213	0.00311	69.1549
1	10.732	0.511	0.00671	14.2759
2	10.727	0.468	0.01417	6.9252

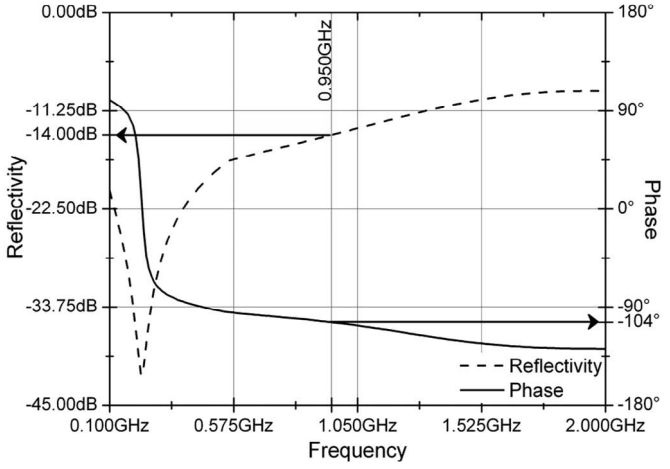


Figure 25.53 Reflectivity and phase analyses of ETS-Lindgren's FT-2000 ferrite absorber.

radiation will not cause noticeable interference with the forward radiation. At 950 MHz, Figure 25.53 shows that the magnitude of the reflected radiation exceeds −14 dB, while the phase of the reflected radiation is -104°, which is expected to result in undesirable destructive interference between forward and reflected radiation. From the simulation results it can be concluded that the ferrite absorber alone does not offer a desirable ground plane above 950 MHz.

In the hybrid EBG/ferrite ground plane design, an EBG structure design is implemented to correct the phase of the reflected signal beyond the desirable performance region of the ferrite absorber. To provide the necessary EBG structure design for the hybrid ground plane, an EBG structure is designed alone to provide desirable operation beginning at around 1 GHz to achieve a starting point for the geometry of the hybrid EBG/ferrite ground plane. The performance and operation of the combined ferrite absorber and EBG structure are expected to be different when the two are implemented together. For that reason, it is important to note that the initial design of the EBG structure is to provide a starting point for the overall optimization of the structure when implemented with the ferrite absorber. For the first version of the hybrid ground plane, the initial EBG structure design is implemented with a plywood backed slab of ETS-Lindgren's FT-2000 ferrite absorber resting on the PEC ground plane of the EBG structure. This is done because ETS-Lindgren offers the option of delivering the ferrite absorber with a plywood backing for additional support. It should be noted that, in implementing the integrated EBG/ferrite ground plane, a small portion of the ferrite slab is removed around each via to allow the EBG structure to behave as desired. The design of the other three versions is similar to the first one with the ferrite slab being implemented without the plywood backing to reduce the overall profile of the hybrid ground plane. By altering the thickness of the structure, the surface patch sizes and spacings, as well as the dimensions of the removed sections of the ferrite material, the hybrid structure is optimized to provide the desired operational performance. The desired operational performance for each of the four versions of the hybrid ground plane is described next.

For all four versions, the ferrite region of operation, which occurs at the lower frequencies of the hybrid ground plane's operational bandwidth, is defined where less than −14 dB of reflected radiation is seen from the hybrid ground plane. The EBG region of operation for the first, second, and fourth designs is defined where the phase of the reflected radiation is between $\pm 90^{\circ}$. This is done to provide for the largest operational bandwidth. The EBG region of operation for the third design is defined where the phase of the reflected radiation is between $\pm 45^{\circ}$. This is done to provide increased constructive interference at the edges of the EBG operational region. However, this enforced enhancement gain design (reflection angle within $\pm 45^{\circ}$) comes at the cost of a reduced operational bandwidth. The four optimized hybrid EBG/ferrite ground plane design versions are shown in Figure 25.54 with their dimensions being shown in Table 25.4.

25.6.2 Simulation Procedures and Considerations

When simulating the hybrid EBG/ferrite ground plane design, much attention was paid to the simulation procedures and many considerations were taken into account to ensure accurate simulations. These procedures and considerations are described in detail next.

The first conducted simulation analyses were done to determine the reflectivity of the hybrid ground plane and the phase of radiation reflected off the ground plane. To

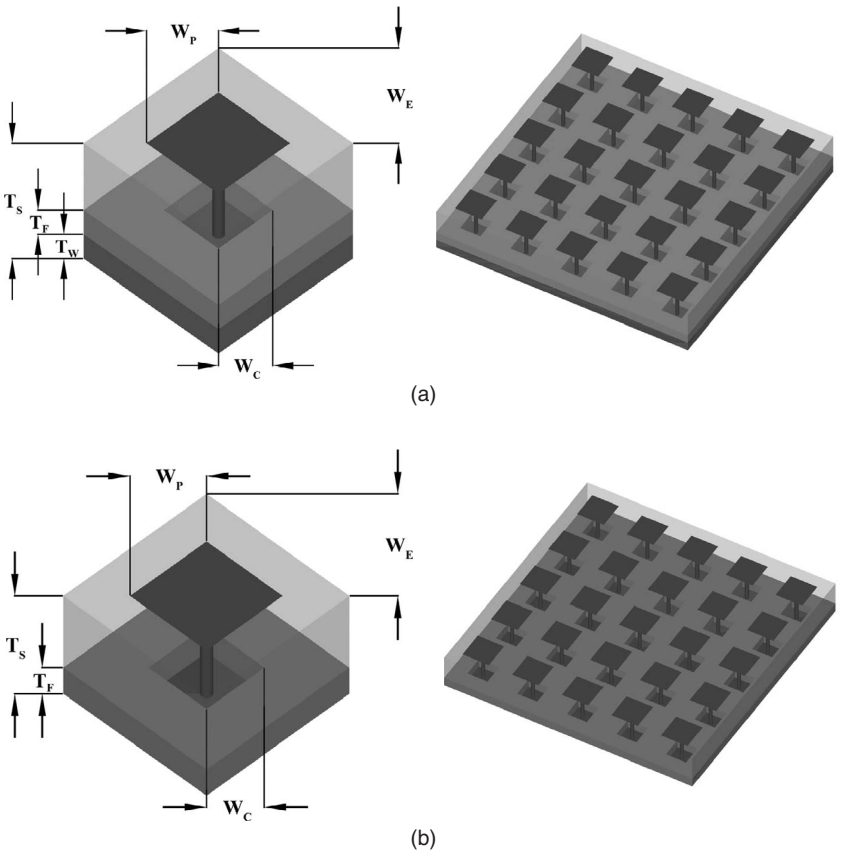


Figure 25.54 Geometries of (a) first and (b) second, third, and fourth hybrid EBG/ferrite ground plane design versions.

TABLE 25.4 Dimensions of the Four Hybrid Ground Plane Design Versions

Version	T_S (cm)	T_F (mm)	T_W (mm)	W_E (cm)	W_P (cm)	W_C (cm)
1	4.92	6.7	6.7	3.5	1.5	1.5
2	3.9	6.7	—	3.5	1.5	1.5
3	5.2	6.7	—	3.5	1.5	1.5
4	2.5	6.7	—	3.7	1.975	1.5

perform these analyses, a plane wave was directed at the surface of the hybrid ground plane and the magnitude and phase of the reflected radiation were calculated. To reduce simulation complexity and hence increase the probability of convergence while also reducing simulation run time, a unit cell of the hybrid ground plane was used for these simulations. To ensure accurate simulation results, careful attention was paid to how the boundaries were set up for the simulation space. To provide an infinite array of the hybrid ground plane unit cells, master and slave boundaries were used in Ansoft's HFSS. Ansoft states that master and slave boundaries are useful for simulating infinite arrays.

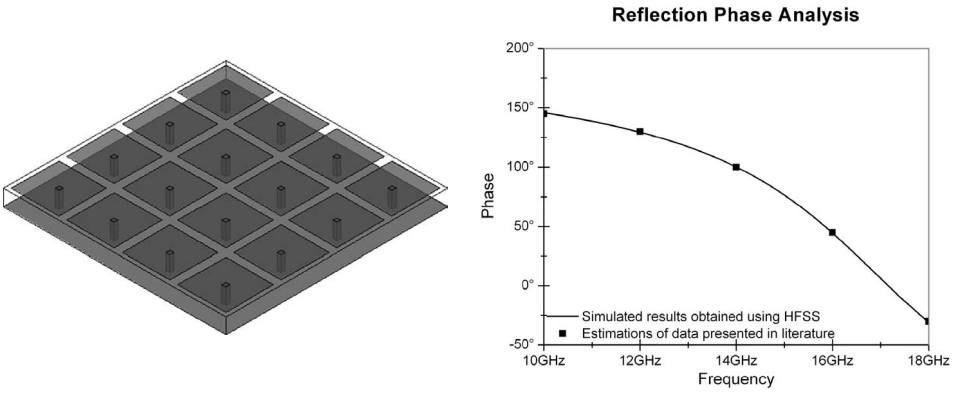


Figure 25.55 (a) EBG ground plane design presented in Ref. 69 and (b) comparison of reflection phase results of presented EBG ground plane obtained using Ansoft's HFSS and estimations of data presented in Ref. 69.

They force the E-field at each point on the slave boundary to match the E-field at the corresponding point on the master boundary to within a user-defined phase difference. To verify the setup of the simulation space, a simulation for comparison purposes was performed on an EBG structure presented in the literature. The presented EBG structure design, shown in Figure 25.55a, consists of surface patches with widths and lengths of 3 mm, patch spacings of 0.5 mm, and a substrate thickness of 1 mm. As Figure 25.55b depicts, the results obtained using Ansoft's HFSS match the results presented in Ref. 69.

The next simulation analyses conducted were done to determine the effectiveness of the hybrid structure to perform as an ultrawideband ground plane when implemented with bidirectional radiating antennas. In the simulation case where the antenna placed above the hybrid ground plane is not an array antenna, it is not possible to implement master and slave boundaries (master and slave boundaries are used for simulating infinite arrays). Ansoft states that the radiation boundaries need to be at least $\lambda/4$ from the radiating element. Consequently, the simulation space needed to be sufficiently large so that the radiation boundaries could be greater than $\lambda/4$ from the radiating element at the lower frequencies. To meet these criteria, the simulation space was set to have a length, width, and height of a little over λ at the lowest frequency simulated (providing for a spacing from the radiating element of approximately $\lambda/2$ for the lowest frequency). To verify the accuracy of the simulations over the desired frequencies for the set simulation space, $\lambda/2$ dipoles were simulated at various frequencies. Much attention was paid to the simulation mesh to provide the correct directivity for the dipole antennas at each frequency point. The verified simulation space and mesh seeding were then used for the simulation of bidirectional radiating antennas placed above a hybrid ground plane with 11×11 elements. An 11×11 element hybrid ground plane was chosen because the ground plane placed below the antenna needed to be large enough to provide unidirectional radiation without too much fringing around the edges of the ground plane, while also being a size that allowed the simulations to converge. It should also be noted that symmetry boundaries were used to reduce the simulation space to one-quarter of its size, thus reducing simulation complexity and increasing the probability of convergence while

also reducing simulation run time. In the case where the antenna placed above the hybrid ground plane is an array antenna, master and slave boundaries could be implemented to once again reduce simulation complexity and hence increase the probability of convergence while also reducing simulation run time. Antenna array elements were placed above a single element of the hybrid ground plane and master and slave boundaries were used to simulate an infinite array of the combination. The return loss of the combination was calculated for an infinite array of the unit cells. The directivity of the combination was initially calculated for a single unit cell. Therefore the directivity of any size array of these unit cells can be calculated by multiplying the single unit cell directivity by an array factor. To verify the accuracy of the return loss results as well as the directive gain results of this setup, a simulation for comparison purposes was performed on an ultrawideband antenna array, more specifically Raytheon's long slot array antenna [64, 65] placed above an electric conductor ground plane. The element widths and lengths of the long slot array presented in Ref. 64 were 6.25 cm and the slot widths were 2.5 cm while the element widths and lengths of the long slot array presented in Ref. 65 were 27.94 cm and 28.03 cm, respectively, and the slot widths were 13.97 cm. As shown in Figure 25.56b,d, the results obtained using Ansoft's HFSS match the results presented in Refs. 64 and 65.

25.6.3 Reflectivity and Phase Analyses of Hybrid EBG/Ferrite Ground Plane

The first simulation analyses conducted were done to determine the reflectivity of the hybrid ground plane and the phase of radiation reflected off the ground plane. These analyses were conducted with each of the four versions of the hybrid ground plane design depicted in Figure 25.54. To perform these simulation analyses, a plane wave was directed at the surface of a unit cell of the hybrid ground plane and the magnitude and phase of the reflected radiation were calculated. As described previously, periodic boundary conditions (master and slave boundaries) were used in Ansoft's HFSS to simulate an infinite array of the hybrid ground plane.

The reflectivity and phase analyses for the four versions of the hybrid ground plane design can be seen in Figure 25.57. From the results in Figure 25.57, the two areas of operation for the hybrid structure can be seen. The ferrite absorber region of operation is defined where less than -14 dB of reflectivity is achieved. For the first, second, and fourth design versions, the EBG region of operation is defined where the phase of the reflected radiation is between $\pm 90^\circ$ regardless of the magnitude of the reflected radiation. This is done to provide the largest operational bandwidth without introducing destructive interference between the reflected and forward radiation. For the third design version, the EBG region of operation is defined where the phase of the reflected radiation is between $\pm 45^\circ$ regardless of the magnitude of the reflected radiation. As mentioned previously, this is to provide for increased constructive interference at the edges of the EBG operation region.

From the reflectivity and phase analyses shown in Figure 25.57, the bandwidth performance of each hybrid ground plane design version can readily be determined. Displayed in Table 25.5 is a comparison of the bandwidth performance results for the four hybrid ground plane design versions. To give an overall understanding of the bandwidth capabilities of each design version for the two operation regions of the hybrid ground plane, the ferrite operation region and the EBG operation region, the bandwidth of each region of operation is presented in Table 25.5 along with the overall bandwidth of the hybrid ground plane. As may be seen, the fourth design version of the hybrid ground plane

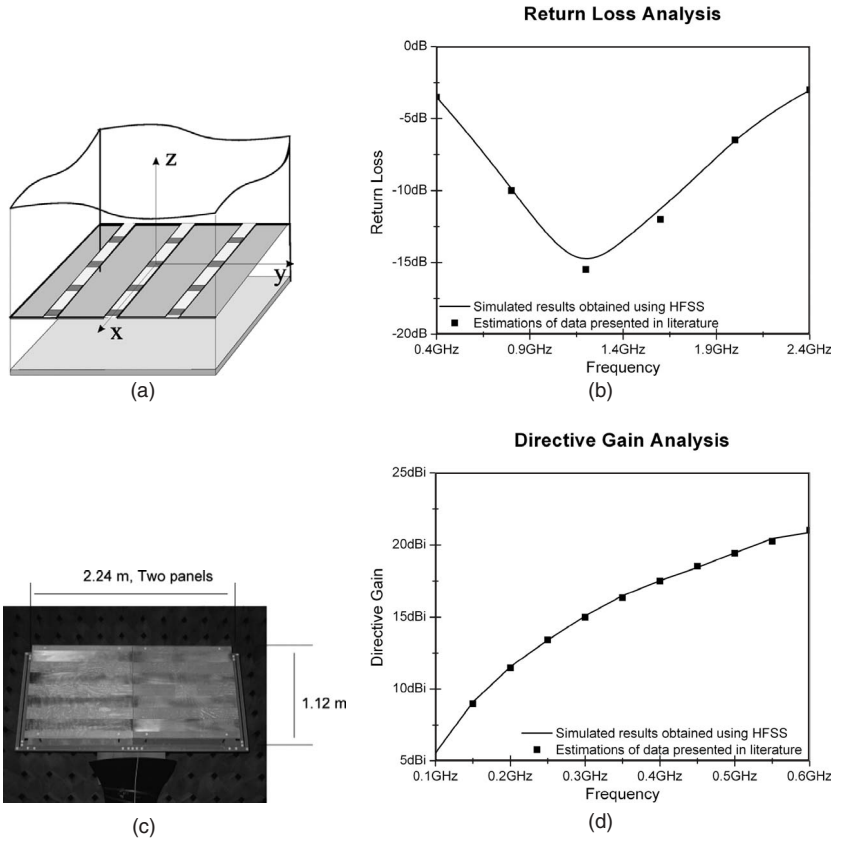


Figure 25.56 (a) Long slot array antenna above PEC ground plane as presented in Ref. 64, (b) comparison of return loss results of presented long slot array antenna above PEC ground plane obtained using Ansoft's HFSS and estimations of data presented in Ref. 64, (c) long slot array antenna above PEC ground plane as presented in Ref. 65, and (d) comparison of directive gain results of presented long slot array antenna above PEC ground plane obtained using Ansoft's HFSS and data presented in Ref. 65.

offers the greatest operational bandwidth and as described in Table 25.4 the lowest profile. Additionally, as expected, the stricter constraints on the EBG region of operation for the third design version cause it to have the smallest operational bandwidth. From the reflectivity and phase analyses results, it can also be seen that implementing a plywood backing, as seen in the first design version, provides enhancement for the ferrite region of operation while at the same time reducing the EBG region of operation and hence results in a reduction of the overall operational bandwidth.

25.6.4 Antenna and Hybrid EBG/Ferrite Ground Plane Implementation Analyses

The next simulation analysis performed was to verify the effectiveness of the hybrid EBG/ferrite ground plane in providing the desired unidirectional radiation for a bidirectional radiating antenna. For this analysis two dipole antennas were simulated above the

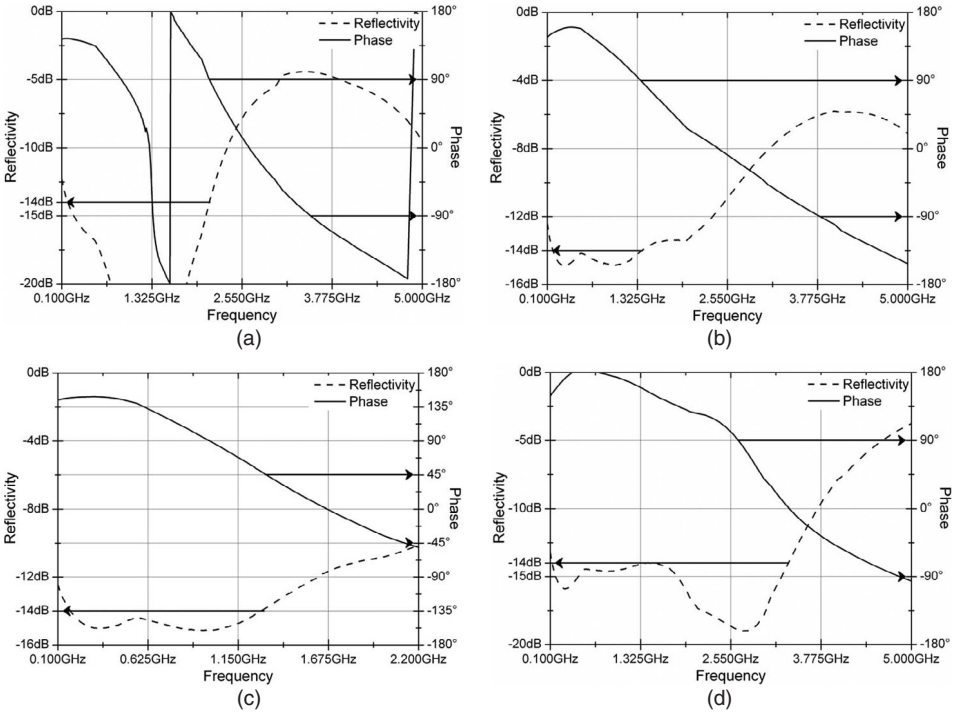


Figure 25.57 Reflectivity and phase analyses of (a) first, (b) second, (c) third, and (d) fourth hybrid EBG/ferrite ground plane design versions.

TABLE 25.5 Hybrid EBG/Ferrite Ground Plane Bandwidth Performance Comparison

Design	Parameter	Ferrite Operation Region	EBG Operation Region	Hybrid Ground Plane Operation
First design version	Frequency range	170 MHz to 2.1 GHz	2.1–3.5 GHz	170 MHz to 3.5 GHz
	Bandwidth	12.4:1	1.7:1	20.6:1
Second design version	Frequency range	170 MHz to 1.39 GHz	1.39–3.8 GHz	170 MHz to 3.8 GHz
	bandwidth	8.2:1	2.7:1	22.4:1
Third design version	Frequency range	170 MHz to 1.28 GHz	1.28–2.15 GHz	170 MHz to 2.15 GHz
	Bandwidth	7.5:1	1.7:1	12.6:1
Fourth design version	Frequency range	120 MHz to 3.33 GHz	2.65–4.86 GHz	120 MHz to 4.86 GHz
	Bandwidth	27.8:1	1.8:1	40.5:1

second version of the hybrid ground plane, shown in Figure 25.54b. The dipole antennas were placed above and parallel to the hybrid ground plane as depicted in Figure 25.58a. The hybrid ground plane implemented in this analysis was an 11×11 element array. One of the dipole antennas was designed to operate in the ferrite region of operation at 0.5 GHz and the other was designed to operate in the EBG region of operation at

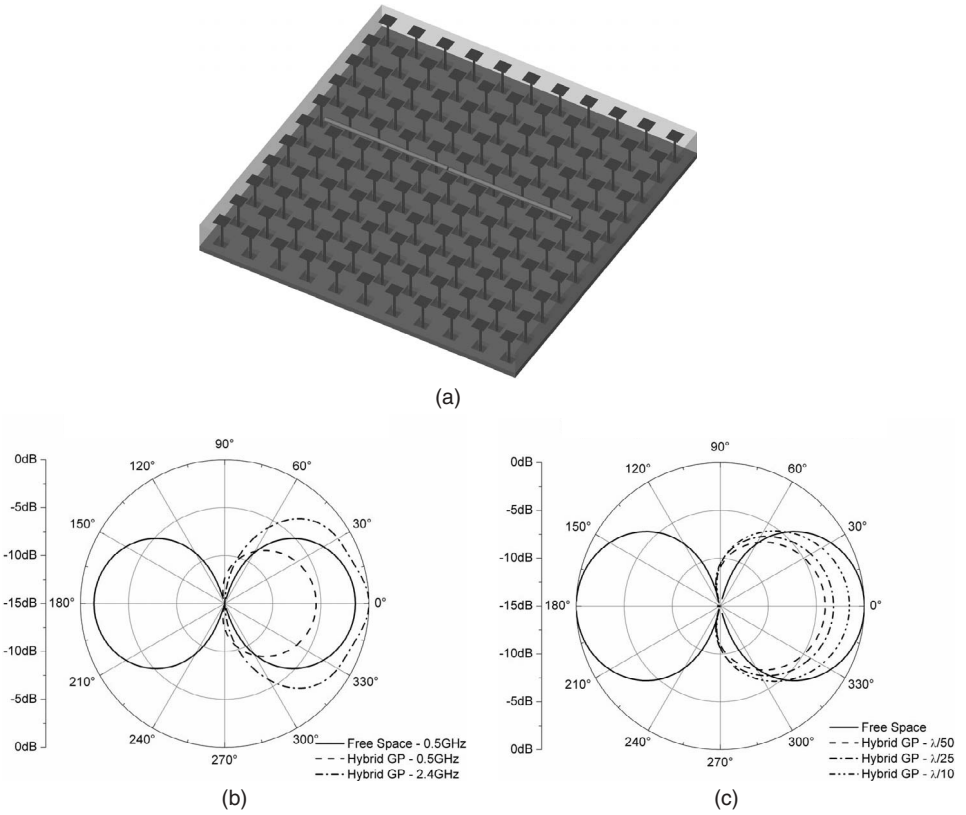
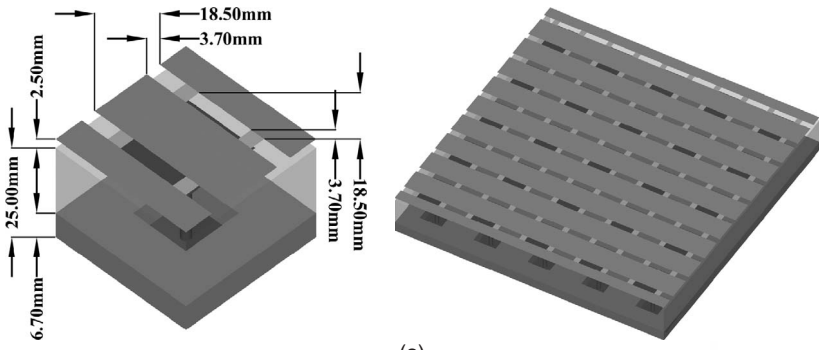


Figure 25.58 (a) Dipole antenna (0.5-GHz dipole) implemented with hybrid ground plane, (b) radiation patterns of dipole antennas (0.5- and 2.4-GHz dipoles) in free space and above hybrid ground plane, and (c) radiation patterns of dipole antenna (0.5-GHz dipole) in free space and above hybrid ground plane.

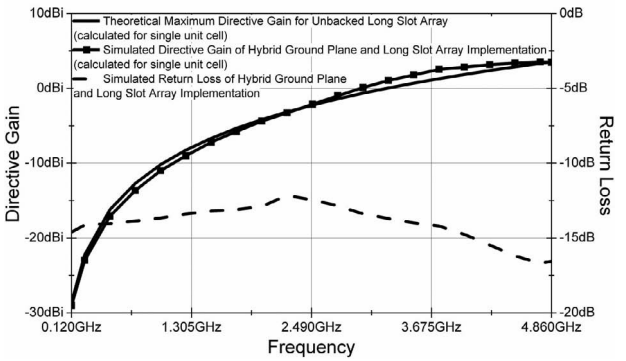
2.4 GHz. Additionally, for comparison purposes, a dipole antenna designed to operate at 0.5 GHz was simulated in free space. The radiation pattern of the dipole antenna in free space as well as the radiation patterns of the dipole antennas placed at $\lambda/50$ above the hybrid ground plane can be seen in Figure 25.58b. As Figure 25.58b depicts, the dipole antennas above the hybrid ground plane produce unidirectional radiation patterns. From this it can be seen that the novel hybrid EBG/ferrite ground plane performs effectively as a ground plane for bidirectional radiating types of antennas. It should also be noted that there is an increase in the directive gain of the dipole antenna above the hybrid ground plane operating in the EBG region of operation over that of the dipole antenna in free space. This is due to the significant amount of reflection from the hybrid ground plane. Because the phase of the reflected radiation is maintained in the range between $\pm 90^{\circ}$ in this region of operation, the interference between the reflected and forward radiation, though not totally constructive, is not destructive. From Figure 25.58b it can also be seen that the directive gain of the dipole antenna above the hybrid ground plane operating in the ferrite region of operation is lower than that of the dipole antenna in free space. Through simulations it was determined that this is due to interactions

between the antenna and the surface of the hybrid ground plane. It should be noted that this effect is not due to an impedance mismatch at the feed of the dipole antennas. The return loss at the feed of a dipole antenna above the hybrid ground plane at 0.5 GHz was −38 dB. To illustrate the effect of these interactions, a comparative analysis of a dipole antenna designed for 0.5 GHz was performed for the antenna in free space and at λ/50, λ/25, and λ/10 above the hybrid ground plane. The simulation results of this analysis, depicted in Figure 25.58c, show that the detrimental interactions are reduced as the antenna is moved further away from the hybrid ground plane. Therefore it can be seen that there is a trade-off between the directive gain in the ferrite region of operation and the overall profile of an antenna implemented with the hybrid ground plane.

The next simulation analysis performed was to verify the effectiveness of the hybrid EBG/ferrite structure to perform as a ground plane when implemented with an ultra-wideband antenna, specifically Raytheon's long slot array antenna. This analysis was performed by simulating Raytheon's long slot array antenna at 0.25 cm above the fourth version of the hybrid EBG/ferrite ground plane, as depicted in Figure 25.59a, and observing various antenna performance parameters over the operational bandwidth of the hybrid ground plane. It should be noted that Raytheon's long slot array antenna was fed for the simulations using Ansoft's HFSS with lumped port excitation sources, which represent



(a)



(b)

Figure 25.59 (a) Long slot array antenna implemented with hybrid ground plane and (b) directive gain and return loss of long slot array antenna above hybrid ground plane.

the δ gap type of feeds. The first performance parameter observed was the radiation characteristics of a single unit cell of the long slot array and hybrid ground plane combination that consisted of a 4×4 element long slot array antenna simulated above a single element of the hybrid ground plane. As described previously, periodic boundary conditions were used in Ansoft's HFSS to simulate an infinite array of this combination and the radiation characteristics were calculated for a single unit cell. The directive gain versus frequency for the long slot array antenna above the hybrid ground plane compared to the theoretical maximum directive gain for an unbacked long slot array, $G_{D\max} = 2\pi A/\lambda^2$ where A is the area of the aparture and λ is the wavelength of the frequency of operation, can be seen in Figure 25.59b. As was seen with the case of the dipole antenna implementation, there is an increase in the broadside directivity of the long slot array antenna above the hybrid ground plane operating in the EBG region of operation over that of the long slot array antenna in free space. This is again due to the significant amount of reflection from the hybrid ground plane. Because the phase of the reflected radiation is maintained in the range between $\pm 90^\circ$ in this region of operation, the interference between the reflected and forward radiation, though not totally constructive, is not destructive. Figure 25.59b also shows that, in the ferrite region of operation, the broadside directivity of the long slot array antenna and hybrid ground plane implementation is slightly less than that of the long slot array antenna in free space. As in the case of the dipole antenna, this is due to the close proximity of the antenna to the hybrid ground plane in the ferrite region of operation. The second performance parameter observed was the return loss characteristics, seen in Figure 25.59b, for the feed system of the long slot array antenna above the hybrid ground plane. The feed system uses a push–pull feed setup [65] with a 50-Ω input impedance for both the push and pull feed arms. This push–pull feed setup provides for a match to the input impedance seen at the slot of 100 Ω. The return loss characteristics verify that the implementation of Raytheon's long slot array antenna and the hybrid EBG/ferrite ground plane provides acceptable operational performance over a 40.5:1 bandwidth ranging from 120 MHz to 4.86 GHz. From the simulation results it can be seen that the novel hybrid EBG/ferrite structure performs as an effective ground plane when implemented with ultrawideband antennas.

25.6.5 Effective ε^* and μ^* of the Hybrid EBG/Ferrite Ground Plane

As described in Ref. 64, the slots of Raytheon's long slot array antenna are excited by an array of feed elements that are spaced at the "Nyquist" interval, which is equal to one-half of the wavelength at the highest frequency of operation. Therefore it is critical to determine the effective complex permittivity and permeability of the structure being used to back the long slot array antenna. Knowledge of these important permittivity and permeability properties can ensure the structure being used to back the long slot array antenna will not load the antenna in a way as to require an inefficient number of feed elements to meet the "Nyquist" interval spacing requirement.

The first method used to determine the effective complex permittivity and permeability of the fourth version of the hybrid EBG/ferrite ground plane entailed an analytical approach. First, using a volume mixture theory and optimal fit method, the effective complex permittivity and permeability of the air/ferrite substrate supporting the hybrid EBG/ferrite structure were calculated. Once these properties were obtained, the effective complex permittivity and permeability of the hybrid ground plane were calculated using a combination of volume mixture theory and Clavijo's approach presented in Ref. 70.

For this process the effective complex permittivity and permeability of the patch portion of the hybrid ground plane (ε_r^Peff and μ_r^Peff) were calculated using Eqs. (25.22) and (25.23), where p is the patch array periodicity, g is the gap width between patches, t is the thickness of the patches, ε_r^Pavg is the average relative complex permittivity of the materials surrounding the patches, and μ_r^S is the effective relative complex permeability of the air/ferrite substrate supporting the hybrid EBG/ferrite structure. Then the effective complex permittivity and permeability of the air/ferrite substrate and via portion of the hybrid ground plane (ε_r^Seff and μ_r^Seff) were calculated using Eqs. (25.24) and (25.25), where ε_r^S is the effective relative complex permittivity of the air/ferrite substrate supporting the hybrid EBG/ferrite structure and α is the ratio of the via's cross-sectional area to the unit cell area. Finally, by using volume mixture theory the effective complex permittivity and permeability of the overall hybrid EBG/ferrite ground plane were calculated.

$$\varepsilon_r^P\text{eff} = \frac{2p}{\pi t}\ln\left(\frac{2p}{\pi g}\right)\varepsilon_r^P\text{avg} \tag{25.22}$$

$$\mu_r^P\text{eff} = \mu_r^S \tag{25.23}$$

$$\varepsilon_r^S\text{eff} = \varepsilon_r^S\left(\frac{1+\alpha}{1-\alpha}\right) \tag{25.24}$$

$$\mu_r^S\text{eff} = \frac{\varepsilon_r^S\mu_r^S}{\varepsilon_r^S\text{eff}} \tag{25.25}$$

The second method used to determine the effective complex permittivity and permeability of the fourth version of the designed hybrid ground plane entailed an optimization approach. For this method, the global optimization software package provided by Tomlab as an add-on for Matlab was used [71]. To determine the effective complex permittivity and permeability of the hybrid ground plane using the optimization software package, the surface impedance of the hybrid ground plane was first calculated from the reflectivity and phase analyses results determined using Ansoft's HFSS. Then using the optimization software with starting values equal to the analytically calculated complex permittivity and permeability values, the optimal complex permittivity and permeability of the hybrid ground plane were determined from the calculated surface impedance values of the hybrid ground plane.

Figure 25.60 shows a comparison of the effective complex permittivity and permeability determined using the two previously described methods. As the comparison shows, the results for both methods match well, thus verifying the accuracy of both methods. More importantly, the low values of the complex permittivity and permeability shown in Figure 25.60 are to be noted. These results indicate that the hybrid EBG/ferrite ground plane will not significantly load the long slot array antenna—thus requiring an unacceptably large number of feed elements. Through determination of the effective complex permittivity and permeability of the hybrid EBG/ferrite ground plane, it can be seen that the hybrid ground plane will not load antennas in an undesirable manner.

25.6.6 Experimental Testing of the Hybrid EBG/Ferrite Ground Plane

Experimental reflectivity and phase analyses were performed on the hybrid EBG/ferrite ground plane to verify the previously presented simulation results. To perform the experimental analyses on the hybrid ground plane, a TEM cell was used to emulate a plane

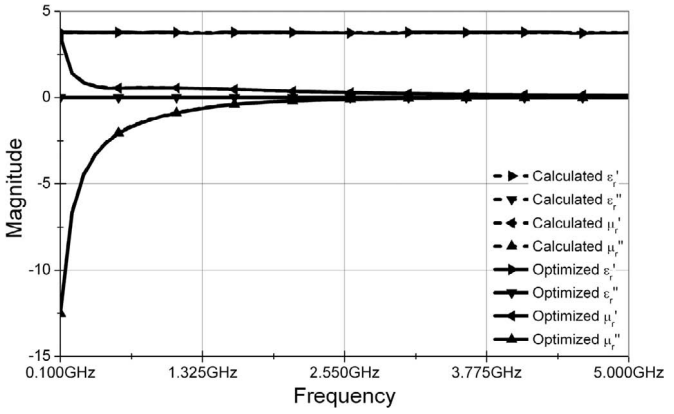
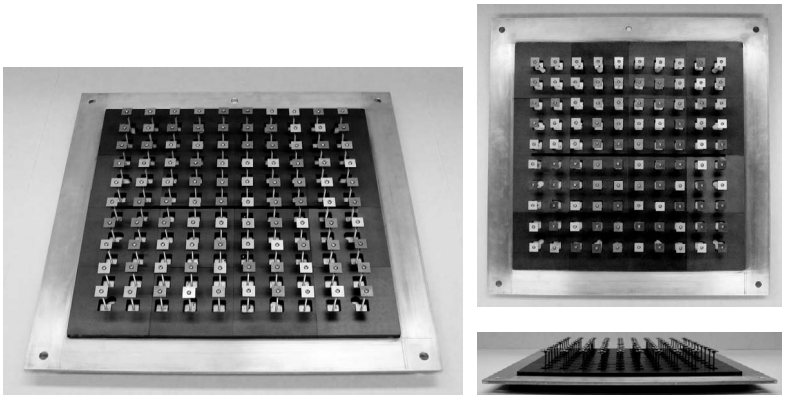


Figure 25.60 Effective complex permittivity and permeability of the hybrid EBG/ferrite ground plane.

wave incident on the ground plane. ETS-Lindgren's FT-2000 ferrite absorbing tiles, square stainless steel washers, stainless steel machine screws, and a sheet of aluminum were used to fabricate the second design version of the hybrid ground plane as shown in Figure 25.61a. The hybrid ground plane was fabricated using a 4×4 array of 10-cm $\times$ 10-cm ferrite tiles, which allowed for an array of 10×10 unit cells with each unit cell being 35 mm $\times$ 35 mm as depicted in Figure 25.54b while giving a ferrite cushion of 2.5 cm on each edge. The 4×4 array of ferrite tiles was mounted on the 49-cm $\times$ 49-cm sheet of aluminum and a 10×10 array of the "mushroom" configured stainless steel washer and screw combinations was mounted on the aluminum-backed array of ferrite tiles. The TEM cell consists of a flared square coaxial line that supports a TEM wave with no low frequency cutoff. Use of the TEM cell allows for low frequency measurements using a surface under test (SUT) that is of reasonable dimensions (free-space low frequency measurements require the SUT to be unreasonably large for accurate results). The experimental setup for measuring the reflectivity and phase of a SUT using the TEM cell is shown in Figure 25.61b.

The IEEE Standard 1128–1998 [72] was referenced for the test procedure used to perform the experimental measurements on the hybrid EBG/ferrite ground plane [72]. To start the experimental measurements, a full one-port S_{11} calibration is performed on the network analyzer at the cable connector that connects to the excitation port of the TEM cell. The TEM cell is then terminated with a metallic shorting plate as the SUT. Time gating is then used to discriminate the reflection from the SUT from other undesired reflections (time gated and ungated results for the metallic shorting plate can be seen in Figure 25.62a). The time gated reflection coefficient measurements for the metallic plate are stored as the reference for subsequent result calculations. The metallic shorting plate is then removed and the TEM cell is terminated with the hybrid ground plane as the SUT. As with the metallic shorting plate, time gating is used to discriminate the reflection from the SUT from other undesired reflections (time gated and ungated results for the hybrid ground plane can be seen in Figure 25.62b). The ratio of the magnitude of the two reflection coefficients is calculated to provide the reflectivity of the hybrid ground plane. To determine the phase offset at the surface of the hybrid ground plane, twice the electrical distance to the surface of the hybrid ground plane (taking into account for the



(a)

(b)

Figure 25.61 (a) Fabricated hybrid EBG/ferrite ground plane and (b) experimental test setup using a TEM cell.

round trip to and from the SUT) for each frequency point is subtracted from the phase of the measured reflection coefficient for the hybrid ground plane. The physical distance to the surface of the hybrid ground plane is determined by subtracting the thickness of the hybrid ground plane from the physical distance to the metallic shorting plate which is determined from the time domain reflection coefficient results for the metallic shorting plate.

Results of the experimental reflectivity and phase analyses for the hybrid ground plane are shown in Figure 25.63. These presented results include the simulation results of a simulated infinite array of the hybrid ground plane and a simulated experimental test setup including the TEM cell as well as the experimental results. The results show good agreement between the measured and simulated results, thus experimentally verifying the performance of the hybrid ground plane.

25.7 SUMMARY

In this chapter we presented an overview of some of the recent advances in developing high performance and low cost phased array antennas with beam-steering capabilities for advanced wireless communications and radar systems. There is no question that phased array antenna design technology has achieved remarkable performance successes for DoD applications. For the consumer-based wireless communication systems and advanced

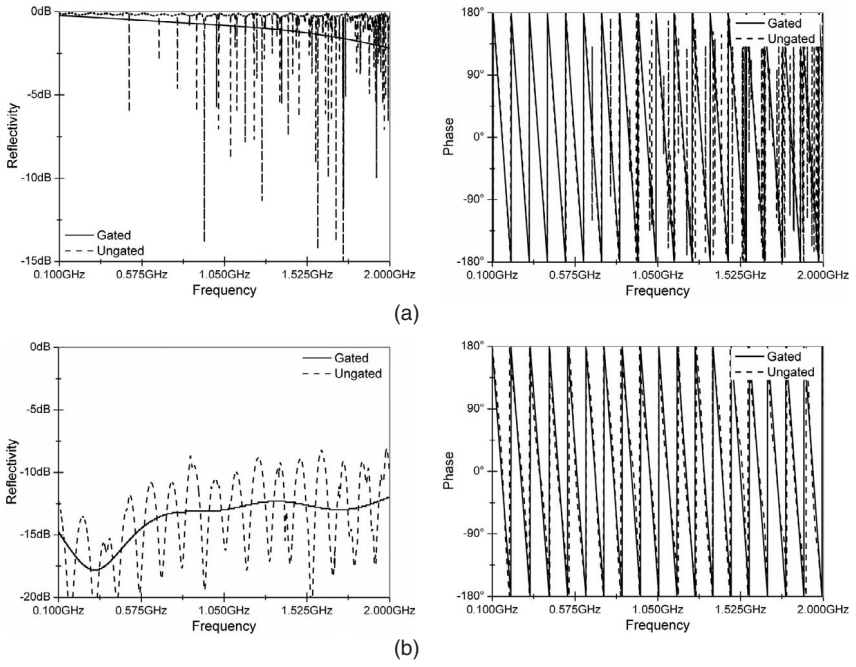


Figure 25.62 (a) Measured results for the metallic shorting plate (gated and ungated) and (b) measured results for the hybrid EBG/ferrite ground plane (gated and ungated).

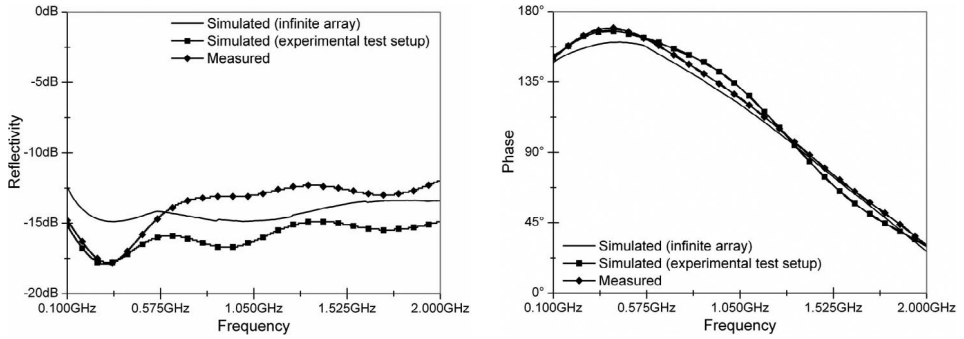


Figure 25.63 Experimental reflectivity and phase analyses results for the hybrid EBG/ferrite ground plane.

communications, radar, and EW implementation in DoD systems, however, the low cost issue emerged as a vital parameter that requires serious consideration in the design and implementation processes. It is these kinds of phased array antennas that the discussion in this chapter was focused on. Phase shifters using ferroelectric materials and multidielectric layer design procedure present a highly promising technology for realizing low cost and high performance phase shifters.

Simulation results examining the feasibility of integrating the CTS antenna array design with the ferroelectric materials technology are also very encouraging and may lead to breakthroughs in developing low cost, high performance phased array antenna systems. An integrated approach that combines low cost antenna array systems, such as the CTS technology, together with low cost phase shifter designs that may include the combination of digital and analog phase shifter technologies may ultimately lead to the realization of this ambitious objective of developing high performance low cost phased array systems. The new designs of the ultrawideband hybrid EBG/ferrite ground plane are exciting and may lead to new innovations to reduce losses while maintaining the ultrawideband characteristics.

REFERENCES

1. G. Stimson, *Introduction to Airborne Radar*, George Stimson III, 1998.
2. E. Brookner, *Aspects of Modern Radar*, LexBook (formerly published by Artech House), Lexington, MA, 1988.
3. E. Brookner, *Radar Technology*, LexBook (formerly published by Artech House), Lexington, MA, 1977.
4. E. Brookner, Major advances in phased arrays: Part I, *Microwave J.*, pp. 288–294, May 1997.
5. B. Palumbo, Some examples of system developments in Italy based on phased array technology, in *1996 IEEE International Symposium on Phased Array Systems and Technology*, Boston, MA, 15–18 October, 1996, pp. 444–449.
6. M. Sarcione, J. Mulcahey, D. Schmidt, K. Chang, M. Russell, R. Enzmann, P. Rawlinson, W. Gluzak, R. Howard, and M. Mitchell, The design, development and testing of the THAAD (Theater High Altitude Area Defense) solid state phased array (formerly ground based radar), in *1996 IEEE International Symposium on Phased Array Systems and Technology*, Boston, MA, 15–18 October 1996, pp. 260–265.
7. E. D. Cohen, Trends in the development of MMICs and packages for active electronically scanned arrays (AESAs), in *1996 IEEE International Symposium on Phased Arrays Systems and Technology*, Boston, MA, 15–18 October 1996, pp. 1–4.
8. R. C. Johnson, and H. Jasik, *Antenna Engineering Handbook*, McGraw-Hill, New York, 1984.
9. R. Hansen, *Phased Array Antennas*, John Wiley & Sons, Hoboken, NJ, 1998.
10. I. Bahl and P. Bhartia, *Microstrip Antenna Arrays*, Artech House, Norwood, MA, 1980.
11. J. L. Butler, and R. Lowe, Beam forming matrix simplifies design of electronically scanned antennas, *Electronic Design*, Vol. 9, pp. 170–173, 12 April, 1961.
12. J. L. Butler, Digital, matrix, and intermediate-frequency scanning, in *Microwave Scanning Antennas*, Vol. III, R. C. Hansen (Ed.), Academic Press, San Diego, 1966, Chap. 3.
13. J. P. Shelton, and K. S. Kelleher, Multiple beams from linear arrays, *Trans. IRE*, Vol. AP-9, pp. 154–161, March 1961.
14. J. Blass, Multidirectional antenna: a new approach to stacked beams, *IRE Int. Conf. Rec.*, Vol. 8, Part 1, 1960.
15. S. Mano et al., Application of planar multibeam array antennas to diversity reception, *Electron. Commun. Japan Part 1*, Vol. 79, No. 11, pp. 104–112, 1996.
16. A. W. Rudge et al. (Eds.), *The Handbook of Antenna Design*, Vol. 2, Peter Peregrinus Ltd., London, 1983.
17. J. B. Rao, P. K. Hughes III, G. V. Trunk, and J. C. Sureau, Affordable phased-array for ship self-defense engagement radar, in *Proceedings of the 1996 IEEE National Radar Conference*, Ann Arbor, MI, 13–16, May 1996, pp. 32–37.

18. C. Mueller, R. Romanofsky, and F. Miranda, Ferroelectric thin film and broadband satellite systems, *IEEE Potentials*, Vol. 20, No. 2, pp. 36–39, April/May 2001.
19. J. Pelton, Telecommunications for the 21st century, *Sci. Am.*, pp. 80–85, April. 1998.
20. B. York, *Thin-film ferroelectrics: deposition methods and applications*, Internal Presentation at University of California, Santa Barbara, CA.
21. F. Deflaviis, N. Alexopoulos, and O. Stafsudd, Planar microwave integrated phase-shifter design with high purity ferroelectric material, *IEEE Microwave Theory Tech.*, Vol. 45, No. 6, pp. 963–969, June 1997.
22. B. Acikel, R. Taylor, P. Hansen, J. Speck, and R. York, A New High Performance Phase Shifter Using $Ba_xSr_{1-x}TiO_3$ thin films, *IEEE Microwave Wireless Components Lett.*, Vol. 12, No. 7, pp. 237–239, July 2002.
23. C. Krowne, M. Daniel, S. Kirchoefer, and J. Pond, Anisotropic permittivity and attenuation extraction from propagation constant measurements using an anisotropic full-wave Green's function solver for coplanar ferroelectric thin- film devices, *IEEE Microwave Theory Tech.*, Vol. 50, No. 2, pp. 537–548, February 2002.
24. R. Collin, *Foundations for Microwave Engineering*, 2nd ed., John Wiley & Sons, Hoboken, NJ, 2001.
25. T. Kaydanova, Improved performance in BST-based tunable circuits employing low-loss non-tunable dielectric, *Integrated Ferroelectrics*, No. 56, 2003.
26. W. Kim, M. Iskander, and C. Tanaka, High-performance low-cost phase-shifter design based on the ferroelectric materials technology, *Electron. Lett.*, Vol. 40, No. 21, pp. 1345–1347, October 2004.
27. A. Djordjevic, M. Bazdar, T. Sarkar, and R. Harrington, *LINPAR for Windows, Artech House, Norwood, MA*, 1999.
28. WIPL-D Pro software, version 6.0, www.wip1-d.com.
29. E. Carlsson and S. Gevorgian, Effect of enhanced current crowding in a CPW with a thin ferroelectric film, *Electron. Lett.*, Vol. 33, No. 2, pp. 145–146, 1997.
30. W. Kim, M. Iskander, and C. Krowne, Modified Green's function and spectral domain approach for analyzing anisotropic and multi-dielectric layers coplanar waveguide ferroelectric phase shifters, *IEEE Trans. Microwave Theory Tech.*, Vol. 55, No. 2, pp. 402–409, 2007.
31. R. S. Elliot, *Antenna Theory and Design,* Prentice Hall, Englewood Cliffs, NJ, 1981.
32. I. J. Bahl and P. Bhartia, *Microstrip Antennas*, Artech House, Norwood, MA, 1980.
33. J. Kraus, *Electromagnetics with Applications*, 5th ed., McGraw-Hill, New York, 1999.
34. R. Collin, *Antennas and Radiowave Propagation*, McGraw-Hill, New York, 1985.
35. W. Milroy, Continuous transverse stub element devices and methods of making same, U.S. Patent No. 5266961.
36. W. Milroy, Advanced broadband access applications of the continuous transverse stub (CTS) array, *International Conference on Electromagnetics in Advanced Applications*, Toronto, Italy, September 2001.
37. A. Lemons, R. Lewis, W. Milroy, R. Robertson, S. Coppedge, and T. Kastle, W-band CTS planar array, *IEEE Microwave Theory Tech. Symp.*, Vol. 2, pp. 651–654, 13–19 June 1999.
38. Z. Zhang, M. Iskander, and Z. Yun, Coaxial continuous transverse stub element device antenna array and filter, U.S. Patent No. 6201509.
39. M. Iskander, Z. Zhang, Z. Yun, and R. Isom, Coaxial continuous transverse stub (CTS) array, *IEEE Microwave Wireless Components Lett.*, Vol. 11, No. 12, pp. 489–491, 2001.
40. R. Isom, M. Iskander, Z. Yun, and Z. Zhang, Design and development of multiband coaxial continuous transverse stub (CTS) antenna arrays, *IEEE Trans. Antennas Propag.*, Vol. 52, No. 8, pp. 2180–2184, August 2004.

41. M. Iskander and W. Kim, Coplanar waveguide continuous transverse stub (CPW–CTS) antenna for wireless communications, U.S. Patent No. 7079082.
42. W. Kim and M. Iskander, A new coplanar waveguide continuous transverse stub (CPW–CTS) antenna for wireless communications, *IEEE Antennas Wireless Propag. Lett.*, Vol. 4, pp. 172–174, 2005.
43. W. Kim, M. F. Iskander, and W. D. Palmer, An integrated phased array antenna design using ferroelectric materials and the continuous transverse stub technology, *IEEE Trans. Antennas Propag.*, Vol. 54, No. 11, pp. 3095–3105, November, 2006.
44. T. W. Bradley, Development of a voltage-variable dielectric (VDD), electronic scan antenna, in *1997 IEE International Radar Conference*, Edinburgh, Scotland, 1997, pp. 383–385.
45. M. Iskander, Z. Yun, Z. Zhang, R. Jensen, and S. Redd, Design of a low-cost 2-D beam-steering antenna using ferroelectric material and the CTS technology, *IEEE Microwave Theory Tech.*, Vol. 49, pp. 1000–1003, May 2001.
46. M. Iskander, Z. Zhang, Z. Yun, R. Isom, M. Hawkins, R. Emrick, B. Bosco, J. Synowczynski, and B. Gersten, New phase shifters and phased antenna array designs based on ferroelectric materials and CTS technologies, *IEEE Microwave Theory Tech.*, Vol. 12, pp. 2547–2553, December 2001.
47. W. Kim and M. Iskander, A multi-element coplanar waveguide continuous transverse stub (CPW–CTS) antenna for wireless communications, *IEEE Antennas and Propagation Society*, Washington DC, July 2005.
48. W. Kim and M. Iskander, Integrated phased antenna array design using ferroelectric materials and the coplanar waveguide continuous transverse stub technologies, *IEEE Antennas and Propagation Society*, Albuquerque, NM, July 2006.
49. M. Rezk, W. Kim, Z. Yun, and M. Iskander, Performance comparison of a novel hybrid smart antenna system versus the fully adaptive and switched beam arrays, *IEEE Antennas Wireless Propag. Lett.*, Vol. 4, pp. 285–288, 2005.
50. N. Celik, W. Kim, M. Demirkol, M. Iskander, and R. Emrick, Implementation and experimental verification of hybrid beamforming algorithm, *IEEE Antennas Wireless Propag. Lett.*, Vol. 5, No. 1, pp. 280–283, December 2006.
51. Z. Zhang, M. F. Iskander, Z. Yun, and A. Host-Madsen, Hybrid smart antenna system using directional elements—performance analysis in flat Rayleigh fading, *IEEE Trans. Antennas Propag.*, Vol. 51, No. 10, pp. 2926–2935, October 2003.
52. J. C. Liberti and T. S. Rappaport, *Smart Antennas for Wireless Communications: IS-95 and Third Generation CDMA Applications*, Prentice Hall, Englewood Cliffs, NJ, 1999.
53. T. S. Rappaport, *Wireless Communications*, Prentice Hall, Englewood Cliffs, NJ, 1999.
54. J. Modelski and Y. Yashchyshyn, New smart beamforming—analysis and designs, in *5th International Conference on Antenna Theory and Techniques*, May 2005, pp. 23–28.
55. J. Modelski and Y. Yashchyshyn, Null pattern synthesis of ferroelectric smart antennas, *IEEE Microwave Theory Tech. Symp.*, Vol. 1, pp. 467–470, May 2001.
56. J. Modelski and Y. Yashchyshyn, Voltage-controlled ferroelectric microstrip antenna for phased arrays, *IEEE Antennas Propag. Symp.*, Vol. 2, pp. 506–509, July 2000.
57. Lal C. Godara, Applications of antenna arrays to mobile communications, Part II: Beam-forming and direction-of-arrival considerations, *Proc. IEEE*, Vol. 85, no. 8, pp. 1195–1245, August 1997.
58. M. Clenet and G. A. Morin, Visualization of radiation—pattern characteristics of phased arrays using digital phase shifters, *IEEE Antennas Propag. Mag.*, Vol. 45, No. 2, pp. 20–35, April 2003.

59. C. J. Miller, Minimizing the effects of phase quantisation errors in an electronically scanned array, in *Proceedings of the 1964 Symposium on Electronically Scanned Phased Arrays and Applications*, RADC-TDR-64–225, Griffith AFB, Vol. 1, pp. 17–38.

60. R. J. Mailloux, *Phased Array Antenna Handbook*, Artech House, Norwood, MA. 1994.

61. M. Clenet and G. A. Morin, Graphical investigation of quantization effects of phased shifters on array patterns, Internal Report, DREO TR 2000-92, November 2000.

62. J. J. Lee, S. Livingston, and R. Koenig, Wide band long slot array antennas, *IEEE Antennas Propag. Soc. Int. Symp.*, Vol. 2, pp. 452–455, June 2003.

63. A. Neto and J. J. Lee, Infinite bandwidth long slot array antenna, *Antennas Wireless Propag. Lett.*, Vol. 4, pp. 75–78, 2005.

64. A. Neto and J. J. Lee, Ultrawide-band properties of long slot arrays, *IEEE Trans. Antennas Propag.*, Vol. 54, pp. 534–543, February 2006.

65. J. J. Lee, S. Livingston, R. Koenig, D. Nagata, and L. L. Lai, Compact light weight UHF arrays using long slot apertures, *IEEE Trans. Antennas Propag.*, Vol. 54, pp. 2009–2015, July 2006.

66. ETS-Lindgren/Rantec, EMC anechoic absorber ferrite tile FT- 2000, http://www.ets-lindgren.com/PDF/FT-2000.pdf.

67. D. Sievenpiper, High-impedance electromagnetic surfaces, Ph.D. dissertation, University of California, Los Angeles, 1999.

68. R. Diaz, Magnetic loading of artificial magnetic conductors for bandwidth enhancement, *IEEE Antennas Propag. Soc. Int. Symp.*, Vol. 2, pp. 431–434, June 2003.

69. F. Yang and Y. Rahmat-Samii, Reflection phase characterizations of the EBG ground plane for low profile wire antenna applications, *IEEE Trans. Antennas Propag.*, Vol. 51, pp. 2691–2703, October 2003.

70. S. Clavijo, R. E. Diax, and W. E. McKinzie III, Design methodology for Sievenpiper high-impedance surfaces: an artificial magnetic conductor for positive gain electrically small Antennas, *IEEE Trans. Antennas Propag.*, Vol. 51, pp. 2678–2690, October 2003.

71. Tomlab Optimiztion Inc., TOMLAB Optimization Environment [Online]. Available at http://tomopt.com/tomlab.

72. IEEE Standard 1128-1998, *IEEE Recommended Practice for Radio-Frequency (RF) Absorber Evaluation in the Range of 30MHz to 5GHz*, April 1998.

73. J. M. Bell and M. F. Iskander, *Equivalent Circuit Model of an Untrawideband Hybrid EBG/Ferrite Structure*, IEEE Antennas and Wireless Propagation Letters, accepted for Publication, April 2008.

74. J. M. Bell and M. F. Iskander, *Experimental Analysis of an Ultrawideband Hybrid EBG/Ferrite Ground Plane*, IEEE Transactions on Instrumentation and Measurement, accepted for publication, March 2008.

75. J. M. Bell and M. F. Iskander, *Effective Propagation Properties of an Enhanced Hybrid EBG/ferrite Ground Plane*, IEEE Antennas and Wireless Propagation Letters, accepted for publication November 2007.

76. J. M. Bell and M. F. Iskander, Ultrawideband hybrid EBG/ferrite ground plane for low-profile array antennas, *IEEE Trans. Antennas Propag.*, Vol. 55, pp. 4–12, January 2007.

77. J. M. Bell and M. F. Iskander, A low-profile Archimedean spiral antenna using an EBG ground plane, *Antennas Wireless Propag. Lett.*, Vol. 3, pp. 223–226, 2004.

CHAPTER 26

Antenna Design Considerations for MIMO and Diversity Systems

MICHAEL A. JENSEN and JON W. WALLACE

26.1 INTRODUCTION

Because of the signal fading induced by multipath propagation, it can be difficult to offer reliable wireless communication in many practical environments. However, because multipath propagation is characterized by electromagnetic waves that depart at a variety of different transmit angles and arrive at different angles at the receiver, it is theoretically possible to achieve high capacity in these scenarios by exploiting these spatial propagation characteristics. For example, if the transmit and receive antennas were to provide infinite spatial selectivity (antenna beams that could excite a single plane wave), unique data streams could be transmitted on each multipath component. While practical considerations clearly make this infeasible, this simple conceptual illustration reveals the high potential capacity available in a rich scattering environment.

While infinite antenna resolution is infeasible, some degree of spatial selectivity can be obtained using multiple antenna elements at the transmit and receive ends of a link. Such multiple-input multiple-output (MIMO) wireless communications exploit the multipath channel characteristics to provide a new resource, namely, spatial processing, that allows improvement of the system performance. This new resource can be used to increase data throughput, improve signal reliability, or reduce transmitted power (leading to extended battery life in mobile devices), all without requiring an increase in the spectrum used for communication. Given these benefits, MIMO has received considerable attention in the research community and is now integrated into emerging communications standards.

This chapter focuses on antenna issues related to MIMO systems [1]. Specifically, we explore the interaction of the antenna elements with the electromagnetic propagation, a study that reveals principles that aid in the design of antennas suitable for MIMO systems. Antenna radiation characteristics such as pattern shape and polarization as well as array configuration naturally represent an important part of the discussion. We also focus on mutual coupling and array supergain, which represent significant issues when implementing MIMO technology on mobile devices where the array must be compact. Finally, the discussion turns to the synthesis of optimal antennas for MIMO communication and provides a way to compare practical designs to this optimal benchmark.

Modern Antenna Handbook. Edited by Constantine A. Balanis

26.2 MIMO SYSTEMS

It is difficult to fully appreciate the impact of the antenna properties on MIMO communication performance without at least a basic understanding of a MIMO radio architecture as well as the underlying principles that enable improved communication performance using appropriate MIMO signal processing. While this discussion focuses on a system with multiple antennas at both the transmitter and receiver, single-input multiple-output (SIMO) and multiple-input single-output (MISO) systems, also known respectively as receive and transmit diversity systems, are simply special cases of the full MIMO representation. The only fundamental difference between MIMO, SIMO, and MISO systems is in the types of signaling and detection algorithms employed.

For this discussion and throughout this chapter, boldface uppercase and lowercase letters represent matrices (matrix $\mathbf{H}$ with mnth element H_{mn}) and column vectors (vector $\mathbf{h}$ with mth element h_m), respectively. When dealing with electromagnetic polarizations, we use the notation $\overline{\mathbf{x}}$ and $\overline{\overline{\mathbf{G}}}$ for vectors and dyads, respectively, to clearly distinguish between polarization vectors and array signal vectors. Finally, we use the explicit functional notation $\mathbf{x}(t)$ and $\mathbf{x}(\omega)$ to indicate that a waveform is to be interpreted in the time domain with time variable t or in the frequency domain with radian frequency ω. If a waveform is given with no functional notation, it can be assumed to be represented in the frequency domain.

26.2.1 MIMO System Model

Figure 26.1 depicts a portion of a radio appropriate for MIMO communication, where the remainder of the radio is typically identical to what one would find in a traditional single-input single-output (SISO) link. To simplify the discussion, we assume that the transmitter has performed the bit-to-symbol mapping so that the input consists of a discrete sequence of complex numbers or communication *symbols*. Each block of Q symbols is formed into a $Q \times 1$ vector $\mathbf{b}^{(k)}$, where k is an integer time index, which is fed into the space–time encoder to create the $N_T \times 1$ complex output vector $\mathbf{x}^{(k)}$, where N_T is the number of transmit antennas. Each element of this vector is fed into a pulse shaping circuit that transforms the discrete sequence into a time-domain signal, resulting in the $N_T \times 1$ vector function $\mathbf{x}(t)$. This function is upconverted to the appropriate carrier frequency to form the vector function $\mathbf{x}_A(t)$, which drives the transmit array to generate the vector radiated field $\overline{\mathbf{x}}_P(t, \Omega_T)$, where $\Omega_T = (\theta_T, \phi_T)$ is a solid angle

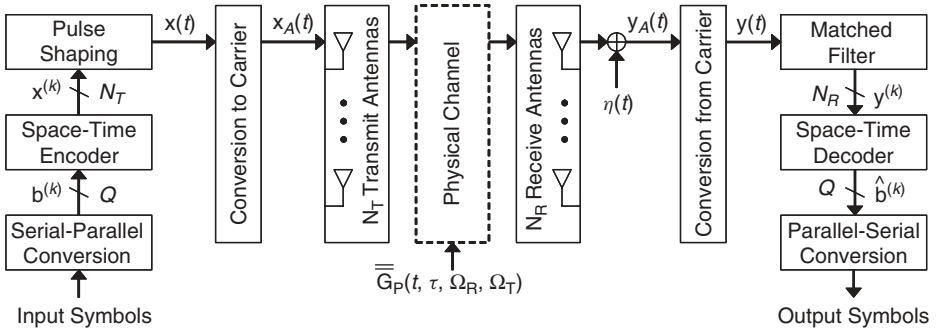


Figure 26.1 A generic MIMO communication system.

coordinate expressed in the transmit array coordinate frame with θ_T and ϕ_T representing the elevation and azimuth angles, respectively.

The dyadic function $\overline{\overline{\mathbf{G}}}_P(t, \tau, \Omega_R, \Omega_T)$, where $\Omega_R = (\theta_R, \phi_R)$ is the solid angle referenced to the receive array coordinate frame, represents the impulse response relating field radiated by the transmit array to the field incident on the receive array. The dependence on time t explicitly shows that this impulse response can be time variant due to motion of the transmitter and receiver or changing scatterers in the environment. The variable τ represents the time delay relative to the excitation time t. We make two assumptions relative to this impulse response: (1) $\overline{\overline{\mathbf{G}}}_P(t, \tau, \Omega_R, \Omega_T) = 0$ for $\tau > \tau_0$ (finite impulse response), and (2) $\overline{\overline{\mathbf{G}}}_P(t, \tau, \Omega_R, \Omega_T)$ remains constant over a time interval (in t) of duration at least as long as the larger of τ_0 and the symbol duration T_{symbol}. The second assumption allows the physical channel to be considered a linear, time-invariant system over each symbol transmission. As we consider only a single symbol time in the following, we drop the explicit dependence of $\overline{\overline{\mathbf{G}}}_P$ on t, understanding, however, that the channel can change from symbol to symbol.

At the receive array, the field distribution $\overline{\mathbf{y}}_P(t, \Omega_R)$ is constructed by a convolution integration in delay and superposition integration in transmit angle, or

$$\overline{\mathbf{y}}_P(t, \Omega_R) = \int_{\Omega_T} \int_{-\infty}^{\infty} \overline{\overline{\mathbf{G}}}_P(\tau, \Omega_R, \Omega_T) \cdot \overline{\mathbf{x}}_P(t - \tau, \Omega_T)\, d\tau\, d\Omega_T \tag{26.1}$$

where $\{\cdot\}$ represents a vector dot product. The N_R-element receive array spatially samples this field to generate the $N_R \times 1$ signal vector $\mathbf{y}'_A(t)$ at the array terminals. Noise in the system is typically generated in the physical propagation channel (interference) and the receiver front-end electronics (thermal noise). To simplify the discussion, we lump all additive noise into a single contribution represented by the $N_R \times 1$ vector $\boldsymbol{\eta}(t)$, which is injected at the receive antenna terminals. The resulting signal plus noise vector $\mathbf{y}_A(t)$ is downconverted to produce the $N_R \times 1$ baseband output vector $\mathbf{y}(t)$, which is passed through a matched filter and then sampled once per symbol to create $\mathbf{y}^{(k)}$. The space–time decoder produces the $Q \times 1$ vector $\hat{\mathbf{b}}^{(k)}$ representing estimates of the originally transmitted symbols, which can then be converted to a serial symbol stream.

26.2.2 Antennas

The characteristics of the propagation channel fundamentally control the potential MIMO performance for a given environment. However, the transmit and receive antenna arrays must ultimately enable the system to properly interact with the propagation channel in order to achieve this performance. Since it is traditional to represent antenna radiation and impedance properties in the frequency domain, we take the Fourier transform of the relevant signals (note that the transform of $\overline{\overline{\mathbf{G}}}_P$ is performed with respect to the delay variable τ). If $\overline{\mathbf{e}}_{T,n}(\omega, \Omega_T)$ represents the vector radiation pattern of the nth transmit antenna, then the field radiated by the transmit array can be expressed as

$$\overline{\mathbf{x}}_P(\omega, \Omega_T) = \sum_{n=1}^{N_T} \overline{\mathbf{e}}_{T,n}(\omega, \Omega_T) x_{A,n}(\omega) \tag{26.2}$$

where $x_{A,n}(\omega)$ represents the antenna driving point excitation.

The choice of which type of driving point excitation (voltage, current, incident voltage wave) should be used in this analysis can vary depending on the situation. However, it must be recognized that the choice of excitation will impact the radiation pattern $\overline{\mathbf{e}}_{T,n}(\omega, \Omega_T)$ used. For example, if $x_{A,n}(\omega)$ represents the antenna driving point voltage, current, or incident voltage wave, then Eq. (26.2) implies that $\overline{\mathbf{e}}_{T,n}(\omega, \Omega_T)$ must be the radiation pattern for the nth element with all other elements terminated, respectively, in a short-circuit, open circuit, or the system characteristic impedance. For many antenna types, the radiation pattern of the driven element with all others terminated in an open circuit can be closely approximated by the radiation pattern of the element in isolation, since the open-circuit termination does not allow terminal currents on adjacent elements to flow. More discussion on this particular point appears in Section 26.4.3. For now we assume that these open-circuit patterns are used in the analysis, so that $x_{A,n}(\omega)$ is the antenna terminal current. It should also be mentioned that while here we are using generic input and output variable designations x and y, we will alter this notation to explicitly represent voltages, currents, voltage waves, or electric fields as we transition into more detailed analysis.

Each element in the receive array samples the incident field $\overline{\mathbf{y}}_P(\omega, \Omega_R)$, leading to the received signal vector $\mathbf{y}_A(\omega)$. If $\overline{\mathbf{e}}_{R,m}(\omega, \Omega_R)$ is the radiation pattern of the mth receive element with all other elements terminated in an open circuit, then the open-circuit voltage on the mth receive antenna is

$$y_{A,m}(\omega) = \int_{\Omega_R} \overline{\mathbf{e}}_{R,m}(\omega, \Omega_R) \cdot \overline{\mathbf{y}}_P(\omega, \Omega_R)\, d\Omega_R \tag{26.3}$$

Combinining Eqs. (26.1), (26.2), and (26.3) leads to

$$\mathbf{y}_A(\omega) = \mathbf{H}_A(\omega)\mathbf{x}_A(\omega) + \boldsymbol{\eta}(\omega) \tag{26.4}$$

where the *channel transfer matrix* (or simply *channel matrix*) has elements

$$H_{A,mn}(\omega) = \int_{\Omega_R} \int_{\Omega_T} \overline{\mathbf{e}}_{R,m}(\omega, \Omega_R) \cdot \overline{\overline{\mathbf{G}}}_P(\omega, \Omega_R, \Omega_T) \cdot \overline{\mathbf{e}}_{T,n}(\omega, \Omega_T)\, d\Omega_R\, d\Omega_T \tag{26.5}$$

Equation (26.4) provides a simple transfer relationship for the MIMO system with Eq. (26.5) explicitly revealing that the physical channel includes the impact of both the propagation environment and the antennas.

It is important to recognize that since the radiation patterns are computed with other elements present but open-circuited, this analysis includes mutual coupling in the sense that it incorporates all appropriate electromagnetic boundary conditions for the antenna pattern determination. However, at this point Eq. (26.4) simply relates transmit currents to received open-circuit voltages. In Section 26.4, we formulate the channel to include at least a portion of the radiofrequency (RF) electronics, providing the opportunity to introduce the impact of the full antenna impedance matrix into the anlaysis.

While the waveform-based transfer relationship in Eq. (26.4) is useful for some analyses, signal-processing analyses often go one step further by considering the channel to relate the input $\mathbf{x}^{(k)}$ to the pulse shaping block and the sampled output $\mathbf{y}^{(k)}$ of the matched filter. Assuming that the frequency-domain channel transfer matrix remains approximately constant over the signal bandwidth, a scenario often described as *frequency nonselective* or *flat fading*, the frequency-domain transfer functions are treated as complex constants

that scale the complex input symbols. The discrete-time transfer relationship assumes the form

$$\mathbf{y}^{(k)} = \mathbf{H}^{(k)}\mathbf{x}^{(k)} + \boldsymbol{\eta}^{(k)} \tag{26.6}$$

where $\boldsymbol{\eta}^{(k)}$ contains the noise that has passed through the receiver and has been sampled at the matched-filter output. Despite the fact that for notational simplicity we sometimes drop the superscript k representing the symbol index, it is important to emphasize that the channel matrix $\mathbf{H}^{(k)}$ can change over time. It is also important to remember that $\mathbf{H}^{(k)}$ is based on the value of $\mathbf{H}_A(\omega)$ evaluated at the carrier frequency, but includes the additional effects of the transmit and receive electronics. For much of this chapter, we assume a narrowband system with ideal electronics, and therefore Eq. (26.6) is a relevant description.

26.2.3 Representative Propagation Channel

Because the MIMO physical channel depends critically on both the antennas and the propagation environment, it is important to discuss briefly how the propagation environment can be described mathematically. It is very common in signal-processing analyses to treat the physical channel transfer matrix $\mathbf{H}_A$ (or discrete-time channel matrix $\mathbf{H}$) as a random matrix whose elements are random variables, typically drawn from a Gaussian distribution. While this is convenient for analysis, incorporation of the antenna radiation properties requires a more descriptive physical channel representation to allow explicit formulation of the interaction between the antennas and the electromagnetic fields responsible for the propagation.

In such cases, it is common to assume that all scattering in the propagation channel is in the far field of the arrays and that a finite number of discrete propagation "paths" (plane waves) connects the transmit and receive arrays. (It is important to recognize that the far-field scattering and plane wave assumptions are not explicitly true for many environments. However, the operation of a MIMO system does not depend critically on this assumption being true. In fact, the rapid variation of near fields can enhance MIMO performance.) If L denotes the number of propagating plane waves that link the transmitter to the receiver, the physical channel response can be expressed as

$$\overline{\overline{\mathbf{G}}}_P(\tau, \Omega_R, \Omega_T) = \sum_{\ell=0}^{L-1} \overline{\overline{\boldsymbol{\beta}}}_\ell \delta(\tau - \tau_\ell)\delta(\Omega_T - \Omega_{T,\ell})\delta(\Omega_R - \Omega_{R,\ell}) \tag{26.7}$$

where $\delta(\cdot)$ represents the Dirac delta function and $\overline{\overline{\boldsymbol{\beta}}}_\ell$ is the dyadic complex gain of the ℓth path with angle of departure (AoD) $\Omega_{T,\ell}$, angle of arrival (AoA) $\Omega_{R,\ell}$, and time delay of arrival (TDoA) τ_ℓ. Note that the dyadic gain $\overline{\overline{\boldsymbol{\beta}}}_\ell$ incorporates polarization changes of the field due to scattering as the wave propagates from the transmitter to the receiver. Any channel time variation is included by making the multipath parameters (L, $\overline{\overline{\boldsymbol{\beta}}}_\ell$, τ_ℓ, $\Omega_{T,\ell}$, $\Omega_{R,\ell}$) time dependent. With this channel model, the channel matrix has elements

$$H_{A,mn}(\omega) = \sum_{\ell=0}^{L-1} e^{-j\omega\tau_\ell} \overline{\mathbf{e}}_{R,m}(\omega, \Omega_{R,\ell}) \cdot \overline{\overline{\boldsymbol{\beta}}}_\ell \cdot \overline{\mathbf{e}}_{T,n}(\omega, \Omega_{T,\ell}) \tag{26.8}$$

In many situations, it is useful to describe the propagation environment stochastically so that it is possible to assess the statistical behavior of a given MIMO system over an

ensemble of representative propagation environments. The mechanism for accomplishing this is to treat each multipath parameter as a random variable drawn from a specified distribution. Most current models are based on initial work in outdoor environments by Turin et al. [2], which demonstrated that multipath components are generally grouped into clusters that decay exponentially with increasing delay. Later work extended the model to indoor scenarios [3] and added directional information [4–7]. Provided that the underlying statistical distributions are properly specified, these models can offer highly accurate channel representations (in a statistical sense). In this section we detail one implementation of such a model which extends the well-known Saleh–Valenzuela model [3] to include AoA/AoD in addition to TDA and multipath amplitude. This model is referred to as the *Saleh–Valenzuela model with angle* or simply the *SVA model*. Naturally, there are numerous other mechanisms for describing the channel characteristics, and therefore this should be considered an introduction to the considerations relevant for incorporating the effect of the antenna in the MIMO system analysis rather than a comprehensive discussion on the topic. More details on this subject can be found in the scientific literature [1, 8].

The SVA model is based on the experimentally observed phenomenon that multipath arrivals appear at the receiver in clusters in both space and time. We refer to arrivals within a cluster as *rays* and restrict our discussion to the horizontal plane for simplicity ($\theta_T = \theta_R = \pi/2$). The model also assumes a single electromagnetic polarization, although it has been extended to include vector fields [4].

Assuming we have N_c clusters with N_r rays per cluster, then the scalar directional channel impulse response of Eq. (26.7) can be written

$$\begin{aligned} G_P(\tau, \phi_R, \phi_T) &= \frac{1}{\sqrt{N_c N_r}} \sum_{\ell=0}^{N_c-1} \sum_{k=0}^{N_r-1} \beta_{k\ell} \delta(\tau - T_\ell - \tau_{k\ell}) \\ &\quad \times \delta(\phi_T - \Phi_{T,\ell} - \phi_{T,k\ell}) \delta(\phi_R - \Phi_{R,\ell} - \phi_{R,k\ell}) \end{aligned} \tag{26.9}$$

where the summation explicitly reveals the concept of clusters (index ℓ) and rays within the cluster (index k). The parameters T_ℓ, $\Phi_{T,\ell}$, and $\Phi_{R,\ell}$ represent the initial arrival delay, mean departure angle, and mean arrival angle, respectively, of the ℓth cluster. The kth ray arrival delay $\tau_{k\ell}$, departure angle $\phi_{T,k\ell}$, and arrival angle $\phi_{R,k\ell}$ are taken relative to the corresponding cluster values.

The SVA model specifies the statistical distribution of the various multipath parameters. The delay of each cluster conditioned on the delay of the previous cluster satisfies the probability density function (PDF)

$$p\left(T_\ell | T_{\ell-1}\right) = \Lambda_T e^{-\Lambda_T (T_\ell - T_{\ell-1})} \tag{26.10}$$

where Λ_T controls the cluster arrival rate for the environment and typically $T_0 = 0$. Similarly, the arrival time for the kth ray in the ℓth cluster obeys the conditional PDF

$$p\left(\tau_{k\ell} | \tau_{k-1,\ell}\right) = \lambda_\tau e^{-\lambda_\tau (\tau_{k\ell} - \tau_{k-1,\ell})} \tag{26.11}$$

where λ_τ controls the ray arrival rate.

The complex gain $\beta_{k\ell}$ has a magnitude that obeys the Rayleigh distribution with the expected power (or variance) satisfying

$$\mathrm{E}\left\{|\beta_{k\ell}|^2\right\} = \mathrm{E}\left\{|\beta_{00}|^2\right\} e^{-T_\ell/\Gamma_T} e^{-\tau_{k\ell}/\gamma_\tau} \tag{26.12}$$

which makes the amplitudes of the clusters as well as the amplitudes of the rays within the clusters decay exponentially with the time constants Γ_T and γ_τ, respectively. The phase of the complex gain is assumed uniformly distributed on $[0, 2\pi]$.

Finally, the angles of departure and arrival must be specified. For indoor and dense urban areas where scattering tends to come from all directions, the cluster departure and arrival angles can be modeled as uniformly distributed random variables on $[0, 2\pi]$. Based on measured data taken in Ref. 5, a two-sided Laplacian distribution is assumed for the ray AoA/AoD distribution with PDF given by

$$p(\phi_P) = \frac{1}{\sqrt{2}\sigma_{P,\phi}} \exp(-|\sqrt{2}\phi_P/\sigma_{P,\phi}|) \tag{26.13}$$

where $P \in \{T, R\}$ and $\sigma_{P,\phi}$ is the standard deviation of angle in radians.

When this SVA model is used in simulations for this chapter, statistical parameters appropriate for an indoor environment, as measured in Ref. 5, are used. These parameters are given as follows:

Parameter	Value	Parameter	Value
$\sigma_{T,\phi}$	26°	$\sigma_{R,\phi}$	26°
$1/\Gamma_T$	34 ns	$1/\Lambda_T$	17 ns
$1/\gamma_\tau$	29 ns	$1/\lambda_\tau$	5 ns

26.2.4 Channel Capacity

In SISO systems, antenna designers typically focus only on the system gain (or received power) enabled by the antenna. However, in a MIMO system, the multivariate nature of the problem complicates the description, and therefore the community has resorted to the capacity as a metric for defining channel quality [9, 10]. This quantity represents the highest error-free MIMO transmission rate for a given transfer matrix and signal-to-noise ratio (SNR) under optimal space–time coding and modulation. Because capacity has become the standard evaluative metric for MIMO systems, it is constructive to discuss this parameter here.

We apply the following analysis to the discrete-time channel in Eq. (26.6) (we drop the time index k). In this case, the capacity represents the maximum number of bits conveyed per time slot (bits/use). If the capacity for the physical channel in Eq. (26.4) is computed, it gives the transmission rate per hertz of bandwidth (bits/s/Hz). Under ideal processing, the two capacities are the same provided that proper signal and noise power definitions are used.

Optimal communication performance is enabled when the transmitter and receiver both know the channel transfer matrix. Let the singular value distribution (SVD) of the channel matrix be denoted by $\mathbf{H} = \mathbf{USV}^\dagger$, where $\{\cdot\}^\dagger$ denotes a conjugate transpose, the matrices $\mathbf{U}$ and $\mathbf{V}$ of singular vectors are unitary, and the diagonal matrix $\mathbf{S}$ contains the real, nonnegative singular values ordered from largest to smallest. If $\mathbf{V}_Q$ consists of

the first Q columns of $\mathbf{V}$, we can encode the input data (space–time encoding) using $\mathbf{x} = \mathbf{V}_Q\mathbf{Pb}$, where $\mathbf{P}$ is a diagonal $Q \times Q$ matrix of real, positive values, which specifies the power allocated to each symbol. The receiver performs the operation (space–time decoding) $\hat{\mathbf{b}} = \mathbf{P}^{-1}\mathbf{S}_Q^{-1}\mathbf{U}_Q^{\dagger}\mathbf{y}$, where $\mathbf{S}_Q$ is the upper left $Q \times Q$ block of $\mathbf{S}$. Since $\mathbf{U}$ and $\mathbf{V}$ are unitary, the resulting output is

$$\begin{aligned}\hat{\mathbf{b}} &= \mathbf{P}^{-1}\mathbf{S}_Q^{-1}\mathbf{U}_Q^{\dagger}\left[\mathbf{Hx} + \boldsymbol{\eta}\right] \\ &= \mathbf{P}^{-1}\mathbf{S}_Q^{-1}\mathbf{U}_Q^{\dagger}\left[\left(\mathbf{USV}^{\dagger}\right)\mathbf{V}_Q\mathbf{Pb} + \boldsymbol{\eta}\right] \\ &= \mathbf{b} + \mathbf{P}^{-1}\mathbf{S}_Q^{-1}\mathbf{U}_Q^{\dagger}\boldsymbol{\eta}\end{aligned} \tag{26.14}$$

Our signaling strategy has resulted in a received signal vector that is a noisy replica of the transmitted signal vector. Figure 26.2 shows a block diagram of this signaling approach.

This signaling strategy is important as it provides physical intuition regarding MIMO communication. Since each (power-weighted) element of $\mathbf{b}$ multiplies the corresponding column of $\mathbf{V}_Q$, this operation suggests that each column of $\mathbf{V}_Q$ represents the array weights for the input symbol. Since the operation $\mathbf{U}_Q^{\dagger}$ creates a vector whose elements are scaled copies of the original transmitted vector, each row of $\mathbf{U}_Q^{\dagger}$ (column of $\mathbf{U}_Q^{*}$) represents the receive array weights for the corresponding symbol. We can think of the array patterns associated with each symbol as *eigenpatterns* that create independent (spatially orthogonal) parallel communication channels in the multipath environment (see Ref. 11 for detailed discussion). Note that each of the created *eigenchannels* (or *eigenmodes* or *modes*) in general excites all multipath components.

For a specific channel, Q and $\mathbf{P}$ can be chosen to achieve the channel capacity, which mathematically is the maximum mutual information between the transmit and receive signal vectors. If we assume that $\boldsymbol{\eta}$ has independent zero-mean Gaussian-distributed elements with equal variance σ_η^2, then the capacity achieving signal vector $\mathbf{x}$ has elements that are zero-mean complex Gaussian-distributed random variables with covariance $\mathbf{R}_x = \mathrm{E}\left\{\mathbf{xx}^{\dagger}\right\}$, where $\mathrm{E}\{\cdot\}$ represents the expectation. The expression for channel capacity is

$$C = \max_{\{\mathbf{R}_x : \mathrm{Tr}[\mathbf{R}_x] \le P_T\}} \log_2\left|\mathbf{I} + \mathbf{HR}_x\mathbf{H}^{\dagger}/\sigma_\eta^2\right| \tag{26.15}$$

where $\mathbf{I}$ is the identity matrix, $|\cdot|$ is the determinant, $\mathrm{Tr}[\cdot]$ is the trace, and P_T is a measure of the total power. Note that $\mathrm{Tr}\left[\mathbf{R}_x\right]$ constrains the sum of the squares of the antenna excitation signals, which in our case represents currents. We later discuss how this does not constrain the power radiated by the transmit array. However, this constraint is commonly used in practice, and therefore we use it for this brief capacity discussion. The off-diagonal elements of $\mathbf{R}_x$ represent the correlation between the transmitted signal streams, with increased correlation generally resulting in decreased capacity.

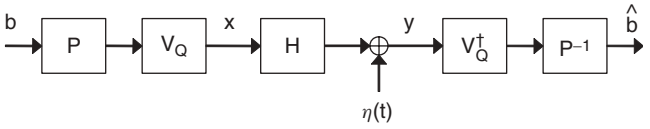


Figure 26.2 Signaling strategy for optimally communicating over the channel eigenmodes.

Our problem therefore becomes that of determining the covariance $\mathbf{R}_x$ that maximizes the right-hand side of Eq. (26.15) subject to the constraint. First, it can be shown that $\mathbf{H}\mathbf{R}_x\mathbf{H}^\dagger$ must be diagonal to maximize this expression [12]. Using the SVD of $\mathbf{H}$ and defining $\mathbf{\Lambda}_x = \mathbf{V}^\dagger\mathbf{R}_x\mathbf{V}$, we can write

$$\left|\mathbf{I} + \mathbf{U}\mathbf{S}\mathbf{\Lambda}_x\mathbf{S}^T\mathbf{U}^\dagger\right| = \left|\mathbf{U}\left(\mathbf{I} + \mathbf{S}\mathbf{\Lambda}_x\mathbf{S}^T\right)\mathbf{U}^\dagger\right| = \left|\mathbf{I} + \mathbf{S}\mathbf{\Lambda}_x\mathbf{S}^T\right| \tag{26.16}$$

where $\{\cdot\}^T$ is a transpose and we have used that $\mathbf{U}$ is unitary (recall that $\mathbf{S}$ is diagonal and real). Note that, given our discussion above, $\mathbf{\Lambda}_x$ represents the covariance of the power-scaled vector $\mathbf{b}$, and since $\mathbf{\Lambda}_x$ is diagonal (to maximize the capacity expression) the input streams should be uncorrelated. Our capacity expression becomes

$$C = \max_{\{\mathbf{\Lambda}_x : \sum_m \Lambda_{x,mm} \le P_T\}} \sum_{m=1}^{N_R} \log_2\left(1 + S_{mm}^2\Lambda_{x,mm}/\sigma_\eta^2\right) \tag{26.17}$$

where we have used that $\mathrm{Tr}[\mathbf{R}_x] = \mathrm{Tr}[\mathbf{\Lambda}_x]$ since $\mathbf{V}$ is unitary. The terms $\Lambda_{x,mm}$ and S_{mm}^2 represent, respectively, the power to be allocated to the mth symbol (and therefore mth channel eigenmode) and the power gain of the mth eigenmode. Our problem is now reduced to finding the real, nonnegative values of the diagonal matrix $\mathbf{\Lambda}_x$. This can be accomplished using Lagrange multipliers to obtain the *water-filling solution* [9, 12–14], which allocates power to the high gain modes and may not use weaker channels, so that Q becomes the number of nonzero values of $\Lambda_{x,mm}$.

While this strategy is optimal, it requires that the transmitter be aware of the channel matrix $\mathbf{H}$, something that may not be practical in systems where the channel state changes rapidly due to node motion or changes in scatterers. If the transmitter has no knowledge of the channel state information (CSI), then each equally weighted data stream can be transmitted from one of the antennas, yielding a transmit covariance of $\mathbf{R}_x = (P_T/N_T)\mathbf{I}$. Substitution into Eq. (26.15) results in the uninformed transmit capacity [15]

$$C_{UT} = \log_2\left|\mathbf{I} + \frac{P_T}{N_T\sigma_\eta^2}\mathbf{H}\mathbf{H}^\dagger\right| \tag{26.18}$$

For full-rank channel matrices at high SNR, the penalty paid for an uninformed transmitter is relatively small [14]. Note that if $N_R \ge N_T$, then the maximum likelihood estimate of the vector $\mathbf{x}$ can be obtained (under the assumption of spatially white Gaussian noise) using

$$\hat{\mathbf{x}} = \mathbf{H}^+\mathbf{y} \tag{26.19}$$

where $\{\cdot\}^+$ represents a pseudo-inverse operation. This is the basic principle behind the well publicized VBLAST algorithm [16, 17].

Since capacity depends on receive SNR, it is important to properly normalize channel matrices for correct interpretation of results. For channel matrices $\mathbf{H}^{(k)}$, $1 \le k \le K$, normalized channel matrices are computed as

$$\mathbf{H}_{\text{norm}}^{(k)} = \mathbf{H}^{(k)}\left[\frac{1}{KN_TN_R}\sum_{k=1}^{K}\|\mathbf{H}^{(k)}\|_F^2\right]^{-1/2} \tag{26.20}$$

where $\|\cdot\|_F$ is the Frobenius norm. When normalized matrices are used in the capacity expressions, $\rho = P_T/\sigma_\eta^2$ represents the average SNR of a single antenna system and is referred to as the SISO SNR. The differences in path loss among a number of channel matrices may be removed by normalizing each matrix independently, or $K = 1$. Using $K > 1$ preserves the relative power levels among the K channels.

26.2.5 Diversity and Spatial Multiplexing

One of the most important conclusions drawn from the analysis of this section is that MIMO systems exploit the channel spatial degrees of freedom to increase communication performance. The two signaling strategies discussed in Section 26.2.4 are focused on using the degrees of freedom to simultaneously communicate multiple independent data streams over the channel. This technique, referred to as *spatial multiplexing*, is only possible in a true MIMO system with multiple antennas at both ends of the link. Note that while this seems the most efficient use of the channel, accurate detection of the received streams is sensitive to noise and the accuracy of the CSI at the receiver (and possibly the transmitter). Therefore it is typically necessary to use a lower rate error control code to obtain reliable transmission.

An alternate approach is to transmit a single data symbol each symbol period, but encode it so that the spatial degrees of freedom provide redundancy. At the transmitter, this is equivalent to spatial error control coding, although we typically refer to this as *transmit diversity*. With multiple antennas at the receiver, further redundancy is assured using *receive diversity*. When the system relies on diversity rather than multiplexing, a higher-rate temporal error control code can typically be employed, and it is technically possible to achieve the same throughput using the two strategies. More generally, the temporal and spatial codes can be combined into a single matrix so that a single symbol is sent over multiple antennas and multiple symbol periods, leading to the concept of *space–time codes*. In this case, the discussion of the operations depicted in Figure 26.1 must be modified to allow transmission of multiple vectors in time, all based on a single input symbol.

The advantage of a diversity system is that its performance is often more robust to situations where one or more of the eigenmodes becomes unusable due to changes in the scattering environment. This is particularly important if the transmitter is unaware of the channel or unable to adapt the number of transmitted streams based on the channel quality. Furthermore, MISO and SIMO signaling are transmit and receive diversity, respectively, indicating that diversity signaling is relevant for systems with only a single antenna at one end of the link. It is also possible to combine multiplexing and diversity [18], which is particularly useful when $N_T \neq N_R$ so that the extra antennas at one end of the link can be used to advantage in a diversity fashion.

Despite the fundamental differences between these two signaling approaches, the goal in each is to use the antennas to exploit the spatial degrees of freedom inherent in the multipath propagation to increase communication performance. Therefore the physical antenna radiation characteristics impact the performance of SIMO, MISO, and MIMO systems in essentially the same manner. In fact, despite the meaning of the word *diversity* as a communication concept, in the context of antennas the word *diversity* simply means that each element of the array possesses unique radiation characteristics that allow it to excite (or sample) the multipath field in a unique way. Therefore, as this discussion transitions to focus additional attention on the role of antennas in determining system

performance, we often remove the complexity associated with considering both transmit and receive antenna arrays and focus only on the receive diversity performance of the array. Naturally, at other times, we show how the analysis applies to MIMO systems as well.

26.3 ANTENNA DIVERSITY

Armed now with a basic understanding of the principles governing MIMO operation, we can focus additional detail on the role of the antennas in determining system performance. Consider a simple example of two receive antennas with vector field patterns $\overline{\mathbf{e}}_1(\theta, \phi)$ and $\overline{\mathbf{e}}_2(\theta, \phi)$ and placed at the coordinates $(-d/2, 0, 0)$ and $(d/2, 0, 0)$ in Cartesian space. A set of L plane waves, with the ℓth plane wave characterized by complex strength E_ℓ, arrival angles (θ_ℓ, ϕ_ℓ), and electric field polarization $\hat{e}_\ell$, impinges on the antenna array. If each pattern is obtained with the other antenna terminated in an open circuit, then the open-circuit signal voltages on each antenna are given by

$$\begin{aligned} \hat{v}_1 &= \sum_{\ell=0}^{L-1} E_\ell \left[\overline{\mathbf{e}}_1(\theta_\ell, \phi_\ell) \cdot \hat{e}_\ell\right] e^{-j(\pi d/\lambda)\sin\theta_\ell \cos\phi_\ell} \\ \hat{v}_2 &= \sum_{\ell=0}^{L-1} E_\ell \left[\overline{\mathbf{e}}_2(\theta_\ell, \phi_\ell) \cdot \hat{e}_\ell\right] e^{j(\pi d/\lambda)\sin\theta_\ell \cos\phi_\ell} \end{aligned} \tag{26.21}$$

where λ is the free-space wavelength and we have explicitly represented the phase terms which depend on position. For the MIMO system to work effectively, the signals $\hat{v}_1$ and $\hat{v}_2$ must be unique, despite the fact that both antennas observe the same set of plane waves, which can be accomplished when each antenna provides a unique weighting to each of the plane waves. Equation (26.21) reveals that this can occur in three different ways:

1. Based on the different antenna element positions, each antenna places a unique phase on each multipath component based on its arrival angle. This is traditional *spatial diversity*.
2. If the radiation patterns $\overline{\mathbf{e}}_1(\theta, \phi)$ and $\overline{\mathbf{e}}_2(\theta, \phi)$ are different, then each multipath is weighted differently by the two antennas. When the antennas share the same polarization but have different magnitude and phase responses in different directions, this is traditional *angle diversity*.
3. If the two antennas have different polarizations, the dot product leads to a unique weighting of each multipath component. This is traditional *polarization diversity*.

It is noteworthy that both angle and polarization diversity are subsets of the more inclusive *pattern diversity*, which simply implies that the two antenna radiation patterns (magnitude, phase, and polarization) differ to create the unique multipath weighting.

26.3.1 Antenna Diversity Performance

While saying that an array design provides good diversity is qualitatively helpful, it does nothing to quantify the array performance. It should be clear based on this discussion that

the performance depends on the propagation environment as well as the antenna radiation properties. Because it is convenient to be able to discuss the performance of an array at one end of the link independent of the design at the other, we can consider a framework where a complex vector electromagnetic wave $\overline{\mathbf{e}}_{\mathrm{s}}(\Omega_R)$ impinges on a receiving array. We assume that this field is represented as a zero-mean Gaussian stochastic process, which means that the complex field envelope varies slowly in time (slow relative to the sinusoidal variation of the carrier frequency). We also assume that the field arriving at one angle is uncorrelated with that arriving at another angle, or

$$\mathrm{E}\left\{\overline{\mathbf{e}}_{\mathrm{s}}(\Omega_R)\overline{\mathbf{e}}_{\mathrm{s}}^{*}(\Omega_R')\right\} = \mathrm{E}\left\{\overline{\mathbf{e}}_{\mathrm{s}}(\Omega_R)\overline{\mathbf{e}}_{\mathrm{s}}^{*}(\Omega_R)\right\}\delta\left(\Omega_R - \Omega_R'\right) = \overline{\overline{\mathbf{P}}}_{\mathrm{s}}\left(\Omega_R\right)\delta\left(\Omega_R - \Omega_R'\right) \tag{26.22}$$

where the vector–vector product is a dyadic multiplication (essentially a vector outer product) and $\overline{\overline{\mathbf{P}}}_{\mathrm{s}}\left(\Omega_R\right)$ is the power angular spectrum, or the average power per unit angle, for the environment.

Again letting $\overline{\mathbf{e}}_{R,m}(\Omega_R)$ represent the field radiation pattern of the mth receive antenna with all other elements terminated in an open circuit, the random variable representing the open-circuit signal voltage on the mth receive antenna in this environment is

$$\hat{v}_{\mathrm{s},m} = \varphi \int_{\Omega_R} \overline{\mathbf{e}}_{R,m}(\Omega_R) \cdot \overline{\mathbf{e}}_{\mathrm{s}}(\Omega_R)\, d\Omega_R \tag{26.23}$$

where the integration is over a full 4π steradians and φ is a complex constant. The covariance of the random signal vector $\hat{\mathbf{v}}_{\mathrm{s}}$ has elements

$$R_{\mathrm{s},mp} = \mathrm{E}\left\{\hat{v}_{\mathrm{s},m}\hat{v}_{\mathrm{s},p}^{*}\right\} \tag{26.24}$$

$$= |\varphi|^2 \int_{\Omega_R'} \int_{\Omega_R} \overline{\mathbf{e}}_{R,m}(\Omega_R) \cdot \mathrm{E}\left\{\overline{\mathbf{e}}_{\mathrm{s}}(\Omega_R)\overline{\mathbf{e}}_{\mathrm{s}}^{*}(\Omega_R')\right\} \cdot \overline{\mathbf{e}}_{R,p}^{*}(\Omega_R')\, d\Omega_R\, d\Omega_R' \tag{26.25}$$

$$= |\varphi|^2 \int_{\Omega_R} \overline{\mathbf{e}}_{R,m}(\Omega_R) \cdot \overline{\overline{\mathbf{P}}}_{\mathrm{s}}(\Omega_R) \cdot \overline{\mathbf{e}}_{R,p}^{*}(\Omega_R)\, d\Omega_R \tag{26.26}$$

where we have used the result in Eq. (26.22).

Any analysis of multiantenna system performance must incorporate the receiver noise. If the noise is generated by an interfering incident field $\overline{\mathbf{e}}_{\mathrm{i}}(\Omega_R)$, then the open-circuit noise voltage on the mth antenna is given by

$$\hat{v}_{\mathrm{i},m} = \varphi \int_{\Omega_R} \overline{\mathbf{e}}_{R,m}(\Omega_R) \cdot \overline{\mathbf{e}}_{\mathrm{i}}(\Omega_R)\, d\Omega_R \tag{26.27}$$

We assume that the interference is a zero-mean Gaussian random process and that the interference field arriving at one angle is uncorrelated with that arriving at another angle. The covariance of the open-circuit noise voltage is

$$R_{\mathrm{i},mp} = |\varphi|^2 \int_{\Omega_R} \overline{\mathbf{e}}_{R,m}(\Omega_R) \cdot \overline{\overline{\mathbf{P}}}_{\mathrm{i}}(\Omega_R) \cdot \overline{\mathbf{e}}_{R,p}^{*}(\Omega_R)\, d\Omega_R \tag{26.28}$$

where $\overline{\overline{\mathbf{P}}}_{\mathrm{i}}(\Omega_R)$ is the interference power angular spectrum and the matrix $\mathbf{R}_{\mathrm{i}}$ is in general full.

An alternate source of noise is thermal noise generated by the receiver front-end amplifiers. In this case, accurate injection of the noise into the model is somewhat complicated and depends on the receiver architecture and nature of the noise source. For simplicity, it is often assumed that the noise voltage referred to the open-circuit antenna terminals is a zero-mean Gaussian process with covariance $\mathbf{R}_\mathrm{t} = \sigma_\mathrm{t}^2\mathbf{I}$, where σ_t^2 is the thermal noise variance. A more detailed method of incorporating realistic thermal noise models is discussed in Section 26.4.2.

We generally assume that the noise sources are independent of each other and are both independent of the signal, such that the overall open-circuit voltage and its covariance are

$$\hat{\mathbf{v}} = \hat{\mathbf{v}}_\mathrm{s} + \underbrace{\hat{\mathbf{v}}_\mathrm{i} + \hat{\mathbf{v}}_\mathrm{t}}_{\hat{\mathbf{v}}_\eta} \tag{26.29}$$

$$\mathrm{E}\left\{\hat{\mathbf{v}}\hat{\mathbf{v}}^\dagger\right\} = \mathbf{R}_\mathrm{s} + \underbrace{\mathbf{R}_\mathrm{i} + \mathbf{R}_\mathrm{t}}_{\mathbf{R}_\eta} \tag{26.30}$$

where the subscript η indicates that the quantity represents the total of all noise sources.

Since $\mathbf{R}_\mathrm{s}$ is in general a full matrix, the open-circuit signal voltages on the receive antenna terminals are correlated, a fact that complicates the analysis of the system performance. To facilitate the analysis, we run the received open-circuit voltage through two *beam formers*. First, we route the signal through a spatial noise *prewhitening* filter expressed as

$$\mathbf{y}' = \mathbf{R}_\eta^{-1/2}\hat{\mathbf{v}} \tag{26.31}$$

which has covariance

$$\mathbf{R}'_\mathrm{y} = \mathbf{R}_\eta^{-1/2}\mathbf{R}_\mathrm{s}\mathbf{R}_\eta^{-1/2} + \mathbf{I} = \mathbf{R}'_\mathrm{s} + \mathbf{I} \tag{26.32}$$

Next, because the covariance $\mathbf{R}'_\mathrm{s}$ is positive semidefinite and Hermitian, it can be expressed in terms of its eigenvalue decomposition (EVD) $\mathbf{R}'_\mathrm{s} = \boldsymbol{\xi}_\mathrm{s}\boldsymbol{\Lambda}_\mathrm{s}\boldsymbol{\xi}_\mathrm{s}^\dagger$, where $\boldsymbol{\xi}_\mathrm{s}$ is a unitary matrix of eigenvectors and $\boldsymbol{\Lambda}_\mathrm{s}$ is a diagonal matrix with real, nonnegative entries. If we pass the signal $\mathbf{y}'$ through a second beam former to obtain $\mathbf{y} = \boldsymbol{\xi}_\mathrm{s}^\dagger\mathbf{y}'$, the signal $\mathbf{y}$ has covariance

$$\mathbf{R}_\mathrm{y} = \boldsymbol{\Lambda}_\mathrm{s} + \mathbf{I} \tag{26.33}$$

Figure 26.3 illustrates the beam-forming process described.

The convenience of this beam-forming process is that we have transformed our correlated signals to a set of uncorrelated ones, where the average SNR of the mth uncorrelated signal is given by the real value $\Lambda_{\mathrm{s},mm}$. The number and magnitudes of the nonzero eigenvalues are an indicator of the system performance. More precisely, we can assume that

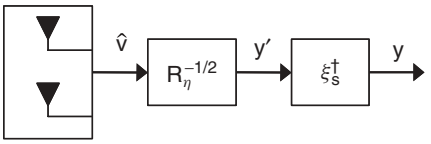


Figure 26.3 Beam-forming approach for uncorrelating the noisy output of the receive antenna array.

we will combine the signals on the independent branches in an appropriate manner. There are several methods for performing this combining, which in order of increasing performance and implementation complexity include the following:

1. *Switched Combining:* Only a single branch is selected for connection to the receiver electronics at any given time. The currently selected branch is used until its SNR drops below a predefined threshold, at which time the next branch is connected to the electronics.
2. *Selection Combining:* All branches are constantly monitored, and the one with the highest SNR is selected for connection to the receiver electronics.
3. *Equal Gain Combining:* The phase of the carrier on each branch is shifted such that the signals on all branches share the same phase. The signals on all branches are then added together.
4. *Maximal Ratio Combining:* As in equal gain combining, the phase of the carrier on each branch is shifted such that the signals on all branches share the same phase. When the branch signals are added together, however, those with the highest SNR receive more weight, with the exact weight chosen according to an optimal algorithm. This approach is optimal in terms of maximizing the SNR at the output of the combining circuitry.

A more comprehensive examination of these diversity combining strategies can be found in Ref. 19. For the purposes of this chapter, we assume that the receiver uses the optimal maximal ratio combining strategy. If all of the branches have equal average SNR or $\mathbf{\Lambda}_s = \mathbf{\Lambda}_0\mathbf{I}$ and if the SNR on each branch satisfies a Rayleigh distribution, then the instantaneous SNR γ at the output of the maximal ratio combiner satisfies the cumulative distribution function (CDF) [20]

$$P_{\mathrm{MR}}(\gamma \leq x) = 1 - e^{-x/\Lambda_0} \sum_{m=1}^{N_T} \frac{(x/\Lambda_0)^{m-1}}{(m-1)!} \tag{26.34}$$

In the case of distinct values of average SNR (all elements of $\mathbf{\Lambda}_s$ are unique), the instantaneous SNR at the combiner output satisfies the CDF [20]

$$P_{\mathrm{MR}}(\gamma \leq x) = \sum_{m=1}^{N_T} \frac{1}{\epsilon_m} \left(1 - e^{-x/\Lambda_{s,mm}}\right) \tag{26.35}$$

$$\epsilon_m = \prod_{\substack{k=1 \\ k \neq m}}^{N_T} \left(1 - \frac{\Lambda_{s,kk}}{\Lambda_{s,mm}}\right) \tag{26.36}$$

Figure 26.4 illustrates the resulting CDF of the instantaneous SNR normalized by the average SNR (γ/Λ_0) at the combiner output for maximal ratio combining with up to four different branches with equal average SNR. The curves are plotted on a Rayleigh scale such that when the CDF represents a Rayleigh distribution (one branch), the curve is a straight line. The curve for a single diversity branch indicates that the normalized SNR will be below -20 dB 1% of the time, while for two branches the normalized SNR will

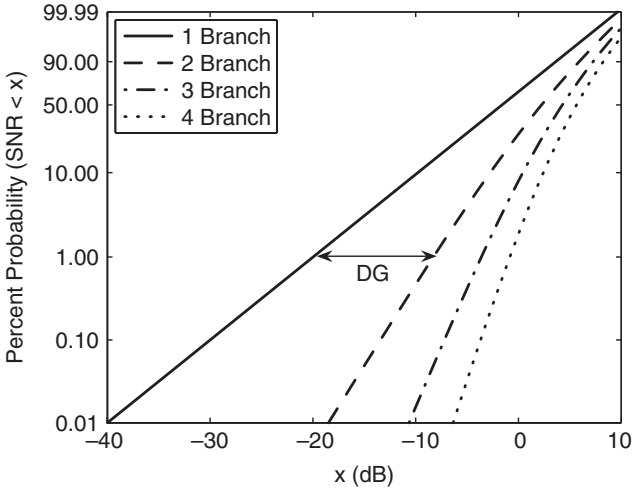


Figure 26.4 Cumulative distribution function of the instantaneous SNR (normalized by the average SNR) for maximal ratio diversity combining with different numbers of branches.

be below −9 dB 1% of the time. Effectively, the use of two antennas and maximal ratio combining represents a gain of 11 dB at the 1% probability level. This gain is referred to as the *diversity gain*, designated as "DG" in the figure. It is noteworthy that the value of the diversity gain depends on the probability level assumed.

While the diversity gain represents a useful metric for quantifying the performance of a diversity system, it is also possible to characterize performance using what is arguably a more intuitive quantity. Consider Figure 26.5, which shows the CDF curves for maximal ratio combining (two and three branches) when all branches have unity average SNR. This figure also shows the CDF for a system with three branches with average SNR values of 1, 0.7, and 0.5. Again considering the 1% probability level, the *number of effective diversity branches* represents the fractional number of uncorrelated antennas with equal average SNR required to achieve the same output SNR as that obtained with the three-branch system with unequal average SNR values. For example, this number of effective branches would be roughly 2.7 for the example case in Figure 26.5. We use this metric as a means for comparing the performance of different antenna topologies.

26.3.2 Pattern (Angle and Polarization) Diversity

As a first study of the diversity potential for different array topologies, it is interesting to explore the fundamental limits of achievable performance with a multiport antenna located at a single point in space. Clearly, this means that all diversity must be pattern diversity, which as previously mentioned includes both angle and polarization diversity. Polarization diversity can be particularly intriguing, since it can potentially lead to low correlation on at least two branches even when the channel is characterized by little or no multipath scattering. However, proper implementation of polarization diversity for MIMO systems requires understanding of the physics involved.

This analysis begins with infinitesimal electric and magnetic current elements (dipoles) radiating into free space. For a three-dimensional coordinate frame, we use three of each

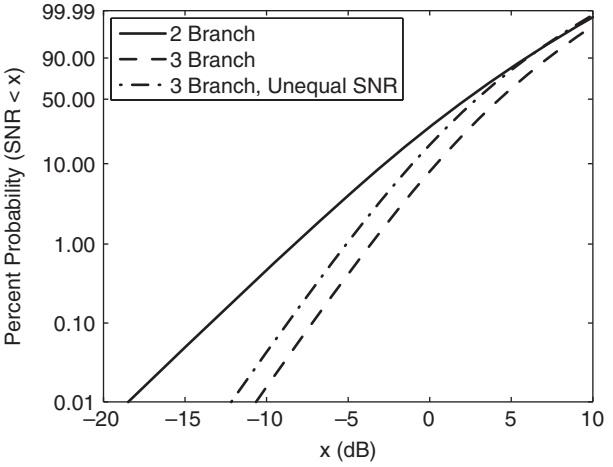


Figure 26.5 Cumulative distribution function of the instantaneous SNR (normalized by the average SNR) for maximal ratio diversity combining with two or three equal gain branches as well as three branches with unequal SNR.

current type (six total dipoles) oriented in the $\hat{x}$, $\hat{y}$, and $\hat{z}$ directions. Each current creates a unique vector far-field radiation pattern given as follows:

Current Orientation	Pattern: Electric Current	Pattern: Magnetic Current
$\hat{x}$	$\mathbf{e}_{R,1} = -\hat{\theta}\cos\theta_R\cos\phi_R + \hat{\phi}\sin\phi_R$	$\mathbf{e}_{R,4} = \hat{\theta}\sin\phi_R + \hat{\phi}\cos\theta_R\cos\phi_R$
$\hat{y}$	$\mathbf{e}_{R,2} = -\hat{\theta}\cos\theta_R\sin\phi_R - \hat{\phi}\cos\phi_R$	$\mathbf{e}_{R,5} = -\hat{\theta}\cos\phi_R + \hat{\phi}\cos\theta_R\sin\phi_R$
$\hat{z}$	$\mathbf{e}_{R,3} = \hat{\theta}\sin\theta_R$	$\mathbf{e}_{R,6} = -\hat{\phi}\sin\theta_R$

These results reveal that polarization diversity *cannot* be completely independent of angle diversity since the radiation pattern shape depends on the orientation (polarization) of the radiating current.

With the dipole patterns defined by this radiation analysis, we now consider all six antennas located at the same point in space and receiving an incident multipath field. We assume the power angular spectrum is

$$\overline{\overline{\mathbf{P}}}_{\mathrm{s}}(\Omega_R) = \begin{cases} (1/\Delta\Omega)\overline{\overline{\mathbf{I}}}, & |\theta_R - \pi/2| \le \Delta\theta,\ |\phi_R| \le \Delta\phi \\ \overline{\overline{\mathbf{0}}}, & \text{otherwise} \end{cases} \tag{26.37}$$

where the notation $\overline{\overline{\mathbf{0}}}$ simply means a dyad of all zeros and

$$\Delta\Omega = \int_{-\Delta\phi}^{\Delta\phi}\int_{\pi/2-\Delta\theta}^{\pi/2+\Delta\theta} \sin^2\theta\ d\theta\ d\phi \tag{26.38}$$

We further assume that there is no external interference and that the thermal noise has covariance $\mathbf{R}_{\mathrm{t}} = \sigma_{\mathrm{t}}^2\mathbf{I}$. With this incident field and the radiation patterns specified, the covariance matrix elements can be computed in closed form [21]. For computation of

the number of effective diversity branches, all eigenvalues of $\mathbf{\Lambda}_s$ are normalized by the largest eigenvalue, and the average SNR used for construction of the CDF for systems with equal branch SNR is $\Lambda_0 = 1$. This normalization makes the absolute value of σ_t^2 irrelevant.

Figure 26.6 plots the effective number of branches as a function of the angular extent of the power angular spectrum. As can be seen, for zero angle spread (which corresponds to a single plane wave), the system offers two effective branches corresponding to the two orthogonal polarizations available in the plane wave. As the angle spread increases, so does the number of diversity branches, with a maximum value of six resulting when the multipath arrives from all angles on the sphere [22]. One interpretation of this full angle spread result is that, under these conditions, the three electric and three magnetic field vector components are all mutually uncorrelated. Another equivalent interpretation is that the three orthogonal radiation patterns times two polarizations ($\hat{\theta}$ and $\hat{\phi}$) leads to the six degrees of freedom. Either way, it is important to recognize that practical environments are rarely characterized by full angle spread, with most of the energy confined to the horizontal plane. Therefore the achievement of six uncorrelated channels using a point sensor is not realized in most practical systems.

There are other practical challenges associated with this multipolarization concept. First, in most environments, transmission of the vertical polarization results in a difference of 3 to 10 dB in the received power levels for the vertical and horizontal polarizations. Mathematically, this means that, if we use multiple polarizations at both ends of the link in a MIMO system, the elements of the channel matrix corresponding to reception that is cross-polarized to the transmission will be weak relative to the copolarized reception [23]. This produces a decrease in the overall receiver SNR, which tends to compensate for the reduced signal correlation enabled by the multipolarized signaling.

Furthermore, while this analysis using infinitesimal dipoles is instructive, realizing the performance suggested in Figure 26.6 with practical antennas is difficult. For example, using half-wavelength dipoles and full-wavelength loops leads to strong mutual coupling and nonideal pattern characteristics that can reduce the number of independent channels.

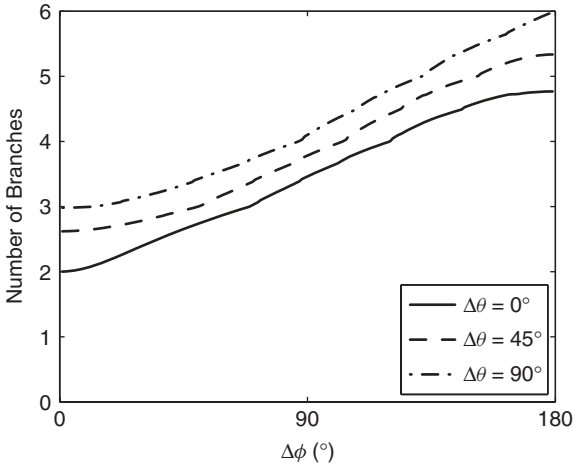


Figure 26.6 Number of effective diversity branches versus incident field angle spread parameters assuming ideal point sensors.

For this reason, alternate topologies such as the *MIMO cube* antenna have been proposed, which achieve a combination of pattern and spatial diversity (see Section 26.3.3) [24]. This antenna consists of 12 dipoles placed along the edges of a cube, as depicted in Figure 26.7. Computational studies show that the pattern diversity allows this antenna to provide good performance even when the side length is as small as $\lambda/20$.

There are of course other ways of achieving pattern diversity. For example, it is possible to obtain different magnitude and polarization responses by exciting multiple modes of an antenna. Figure 26.8 illustrates the concept of using different coaxial feeds in a circular microstrip patch antenna positioned so that different relative excitations on the two ports excite different patch modes. The resulting radiation patterns have been shown to exhibit low correlation and reasonable capacity for many practical propagation scenarios [25]. However, when using this multimode principle, it is important to use modes that not only provide high diversity but also properly direct their energy in angular regions where multipath power is the highest. Failure to do so can reduce the effective received signal power so severely that the benefit gained by the low correlation created by the patterns can be outweighed by the loss in SNR due to poor excitation/reception within the environment of interest [26]. Naturally, another option is simply to use different antenna element types in the design or placing the same element with different orientations on the communications device. Regardless of the approach, the framework considered here provides a method for determining the performance of the array based on knowledge of the element radiation patterns.

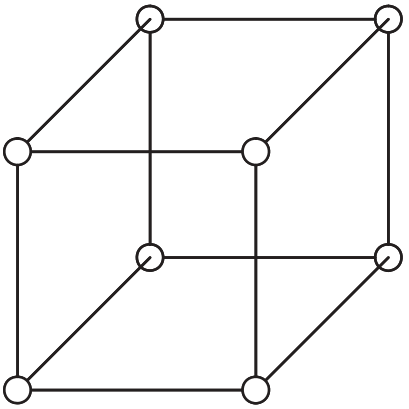


Figure 26.7 MIMO cube antenna consisting of 12 dipole antennas placed along the edges of a cube.

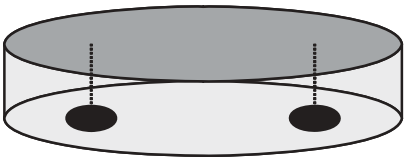


Figure 26.8 Circular microstrip patch antenna with two different coaxial feed points positioned to excite two different radiation modes.

26.3.3 Spatial Diversity

The pattern diversity discussion in Section 26.3.2 is concerned with the types of elements that should be used in a multiantenna system. The concept of spatial diversity, on the other hand, emphasizes the array topology that should be used. The most straightforward illustration of spatial diversity assumes the propagation environment given in Eq. (26.37) assuming $\Delta\theta = 0$ and $\Delta\phi = \pi$ and with two Hertzian dipoles separated by a distance d and placed along the x axis. The covariance matrix for this scenario is given by

$$R_{\mathrm{s},mm} = \frac{1}{2\pi}\int_0^{2\pi} d\phi_R = 1 \tag{26.39}$$

$$R_{\mathrm{s},mp} = \frac{1}{2\pi}\int_0^{2\pi} e^{jk(d/2)\cos\phi_R} e^{jk(d/2)\cos\phi_R} = J_0(kd) \tag{26.40}$$

where k is the free-space wavenumber and $J_0(\cdot)$ is the Bessel function of the first kind of order 0.

Figure 26.9 plots the number of effective diversity branches as a function of the dipole spacing for this configuration using a reference SNR of $\Lambda_0 = 1$ and thermal noise power of $\sigma_{\mathrm{t}}^2 = 1$. It is rather intuitive that the number of branches is small at close spacing and increases to the value of two at larger spacing where the correlation is small. What may be surprising, however, is that the analysis predicts a number of diversity branches that is larger than unity as the spacing approaches zero. This phenomenon is a result of the fact that the analysis ignores electromagnetic effects (electromagnetic coupling), and therefore the two antennas both receive a full share of power even when the spacing is small. Naturally, a more comprehensive analysis, such as the one provided in Section 26.4.2, would limit the effective aperture of the coupled antennas so that this nonphysical behavior would be removed from the computation. Despite this limitation, this analysis reveals that as long as the antenna spacing is larger than roughly $\lambda/4$, the two dipoles essentially behave independently in terms of providing diversity performance.

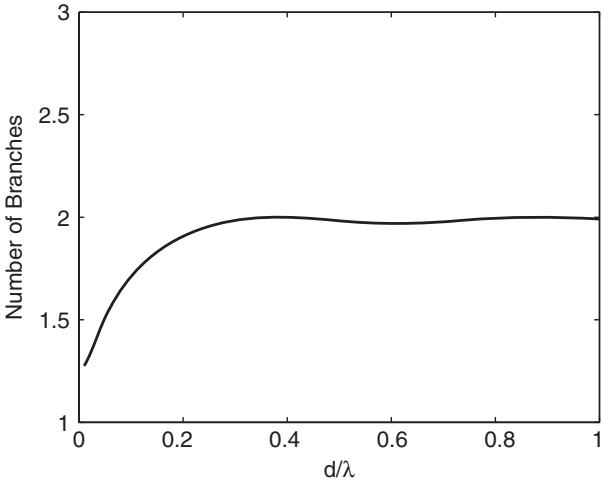


Figure 26.9 Number of effective diversity branches versus element spacing for two Hertzian dipoles. The incident power is uniformly distributed in angle in the horizontal plane.

While this simple analysis provides insight into some considerations relevant to array design, the actual problem of array synthesis is generally complicated. There has been one notable study where several different array configurations were explored for both the base station and the mobile unit in an outdoor environment [27]. The base station antennas included single and dual polarization array and multibeam structures. The arrays on the mobile were constructed from monopoles to achieve spatial, angle, and/or polarization diversity. All of the array configurations provided very similar performance, with the exception of the multibeam base station antennas, which resulted in a 40–50% reduction in measured capacity since generally only one of the beams pointed in the direction of the mobile. These results suggest that average capacity is relatively insensitive to array configuration provided the signal correlation is adequately low and that all elements in the array have patterns that are effective for the environment of interest. More on the issue of antenna synthesis is found in Section 26.6.

26.4 MUTUAL COUPLING

Up to this point in the discussion, the concept of antenna electromagnetic coupling, generally referred to as mutual coupling, has been neglected in the analysis. In systems that can allow relatively large electrical spacing between the antenna elements, ignoring the impact of coupling is justified. However, in many compact devices, the antenna elements must be closely spaced, and the resulting antenna mutual coupling can impact communication performance. It is therefore important to have a framework for analyzing the impact of coupling on the system operation.

Before embarking on a detailed analysis of mutual coupling in MIMO and diversity systems, it is important to first understand the impact of coupling on the antenna radiation characteristics. We can divide the overall effect into two different physical phenomena:

1. Consider an antenna in isolation and driven by a terminal current $i_1(\omega)$. Electromagnetic analysis can be used to construct the radiation pattern for this system. However, if a second antenna is terminated in an open circuit and brought into proximity of the driven element, the boundary conditions used in the electromagnetic analysis change, thereby changing the effective radiation pattern of the driven element. We note, however, that the open-circuit termination does not allow terminal current to flow in the parasitic element, and for many types of elements (e.g., half-wave dipoles) the impact of the altered boundary conditions on the pattern is often quite minor. As a result, in some analyses, the radiation pattern for the element in isolation (isolated element pattern) is used in place of the pattern for the driven element with other elements present but terminated in an open circuit (open-circuit pattern). In the following, we assume that this open-circuit pattern with unit driving current for the nth element is denoted as $\overline{\mathbf{e}}_{T,n}(\Omega_T)$.
2. The open-circuit voltage $\hat{v}_2$ induced at the coupled antenna terminals is related to the current i_1 in the driven element according to $\hat{v}_2 = Z_{21} i_1$, where Z_{21} is the mutual impedance. If the coupled antenna is now terminated with a load impedance Z_L, the induced voltage will create a current in the coupled antenna that depends on the termination impedance. Therefore the effective radiation pattern

is the superposition

$$\overline{\mathbf{e}}_T(\Omega_T) = i_1 \left[\overline{\mathbf{e}}_{T,1}(\Omega_T) + \frac{Z_{21}}{Z_L} \overline{\mathbf{e}}_{T,2}(\Omega_T) \right] \tag{26.41}$$

The composite pattern therefore depends on the load attached to the coupled element. Naturally, by reciprocity, the analysis at the receiver parallels that at the transmitter.

Item 2 makes it clear that ambiguity exists in defining the pattern to use in the correlation analysis expressed in Eq. (26.26). Furthermore, an impedance matching network can have impact on the ultimate superposition of the signals at the antenna terminals. Finally, the way in which this superposition is performed has impact on both the signal and the noise, and therefore any thorough analysis of diversity and MIMO systems with coupled antennas must consider the noise sources involved. In this section, we present approaches for incorporating antenna mutual coupling and realistic noise sources into the analysis. We begin our analysis by focusing on the diversity performance of a receiver subsystem. This is then extended in Section 26.4.3 to incorporate the transmitter and a MIMO propagation channel.

26.4.1 Receiver Modeling: Terminated Antennas

The traditional approach for incorporating mutual impedance into the analysis of receive diversity is to first compute the voltage $\hat{v}_{s,m} + \hat{v}_{i,m}$ at the open-circuited terminals of the mth receive antenna with all other elements terminated in an open circuit using the integrations in Eqs. (26.23) and (26.27). Since the open circuit patterns are specified as the patterns with all other elements open-circuited, this step incorporates the impact of adjacent elements on the electromagnetic boundary conditions of the problem. The challenge now is simply to incorporate the impact of the antenna array impedance matrix into the analysis.

Figure 26.10 shows a block diagram of the receiver model, which includes the incident field, coupled receive array with full impedance matrix $\mathbf{Z}_R$, and a termination characterized by the impedance matrix $\mathbf{Z}_L$. Note that we are not necessarily constraining the load impedance matrix $\mathbf{Z}_L$ to be diagonal, implying the possibility of a coupled load network in addition to the coupled antenna array. In this model, any thermal noise generated within the receiver network should be represented as an open-circuit noise voltage at

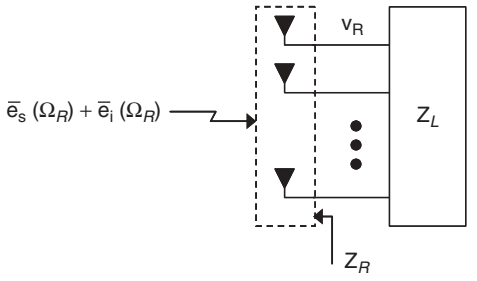


Figure 26.10 Block diagram of a receiver diversity system including the incident field, mutually coupled array, and loads.

the antenna terminals represented as $\hat{\mathbf{v}}_t$. Using simple network analysis, the vector $\mathbf{v}_R$ of voltages at the terminals of the load network can be expressed in terms of the total open-circuit voltage vector $\hat{\mathbf{v}}$ according to [28–33]

$$\mathbf{v}_R = \mathbf{Z}_L \left(\mathbf{Z}_L + \mathbf{Z}_R\right)^{-1} \hat{\mathbf{v}} = \mathbf{C}_R \hat{\mathbf{v}} \tag{26.42}$$

where $\mathbf{C}_R$ is a coupling matrix.

With this representation, the covariance matrix of the voltage across the receiver terminations is given as

$$\mathrm{E}\left\{\mathbf{v}_R \mathbf{v}_R^{\dagger}\right\} = \mathbf{C}_R \mathbf{R}_s \mathbf{C}_R^{\dagger} + \mathbf{C}_R \mathbf{R}_\eta \mathbf{C}_R^{\dagger} = \widetilde{\mathbf{R}}_s + \widetilde{\mathbf{R}}_\eta \tag{26.43}$$

where the signal and noise covariance matrices are defined in Eqs. (26.26) and (26.28), respectively. The transformed covariance matrices $\widetilde{\mathbf{R}}_s$ and $\widetilde{\mathbf{R}}_\eta$ can then be used in the framework of Section 26.3.1 to analyze the system performance. This approach, which has been used to analyze the performance of adaptive array [28] and diversity [29] systems, effectively captures the physics of electromagnetic coupling into the formulation.

It should be emphasized that some analysis is required to reflect receiver front-end noise to an effective open-circuit antenna voltage. Furthermore, while it is tempting to simply use this *coupling matrix* formulation to analyze the impact of coupling termination on the signal and then assume that the noise voltages on the output branches are uncorrelated with each other and equal power ($\widetilde{\mathbf{R}}_\eta = \sigma_\eta^2 \mathbf{I}$), this approach does not yield the correct noise properties. Finally, this approach does not conveniently allow incorporation of more sophisticated noise models such as are encountered in practical front-end low noise amplifiers.

26.4.2 Receiver Modeling: Front-End Network

Because the diversity or MIMO capacity performance depends critically on the SNR observed on each diversity branch or MIMO subchannel, properly incorporating appropriate thermal noise models into the analysis framework is important if the results are to properly reflect the achievable system performance. When considering such an analysis framework, it is important to remember that the noise performance of a transistor-based amplifier depends on the impedance of the source driving the amplifier. Therefore the model must properly include the impedance interface between the antenna terminals and the low noise amplifiers and accurately represent the nature of the amplifier thermal noise. Unfortunately, it is not obvious how the coupling matrix analysis approach discussed previously can easily be modified to include these effects.

Naturally, when increasing the level of modeling detail, one could include additional levels of complexity such as component nonlinearity, conversion loss, matched-filter imperfections, and sampling. While all of these issues can be important, we limit our focus to determining the branch SNR levels achieved by the antenna interfaced to the RF front-end. Figure 26.11 shows a block diagram of the resulting system model, which includes the incident field, coupled receive array, an impedance matching network, noisy amplifiers (used to model the source of thermal front-end noise), and a termination. Since impedance parameters do not lend themselves well to the analysis of cascaded system blocks, we instead use scattering parameters (S-parameters) referenced to a real impedance Z_0 [34] to describe the network signal flow wherein the incident and reflected

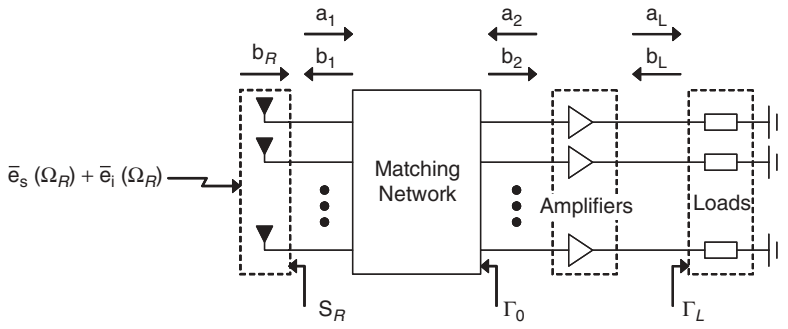


Figure 26.11 Block diagram of a receiver diversity system including the incident field, mutually coupled array, matching network, receiver amplifiers, and loads.

traveling waves are denoted as **a** and **b**, respectively. The flow diagram for this network, with the various system blocks delineated by dashed lines, is shown in Figure 26.12. The specific traveling-wave vectors, S-parameter matrices (symbol **S**), and reflection coefficient matrices (symbol $\boldsymbol{\Gamma}$) appearing in Figure 26.11 are used in the analysis.

26.4.2.1 Network Analysis The first step in our analysis is to properly transform the open-circuit antenna terminal voltage representing the signal plus external interference to a voltage across the receiver terminations and add to this the contribution due to amplifier thermal noise. We begin with the vector $\mathbf{b}_R$, which represents the traveling wave resulting from the incident signal plus interference that will be delivered by the receive antenna terminals to a set of independent loads of resistance Z_0 (so that $\mathbf{b}_1 = \mathbf{0}$). Given this as a *source* wave, the general expression for the wave $\mathbf{a}_1$ incident on the matching network input is given by

$$\mathbf{a}_1 = \mathbf{b}_R + \mathbf{S}_R\mathbf{b}_1 \tag{26.44}$$

where $\mathbf{S}_R$ represents the S-parameter matrix for the coupled array. If the antenna array is terminated in an open circuit, then $\mathbf{b}_1 = \mathbf{a}_1$. Substituting this observation into Eq. (26.44) produces

$$\mathbf{b}_R = (\mathbf{I} - \mathbf{S}_R)\mathbf{a}_1 \tag{26.45}$$

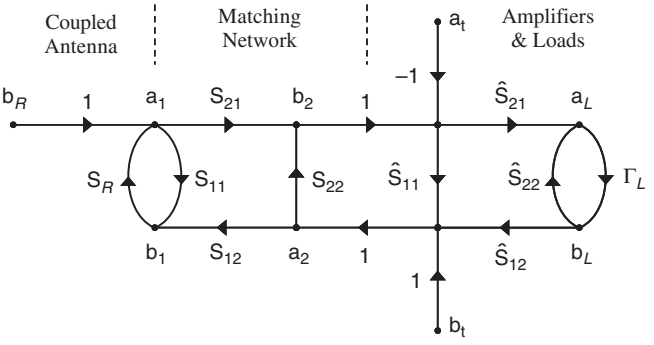


Figure 26.12 Flow diagram for the receiver block diagram shown in Figure 26.11.

However, since we already know that the voltage $\hat{\mathbf{v}}_s + \hat{\mathbf{v}}_i$ represents the open-circuit voltage at the antenna terminals and since $\hat{\mathbf{v}}_s + \hat{\mathbf{v}}_i = Z_0^{1/2}(\mathbf{a}_1 + \mathbf{b}_1) = 2Z_0^{1/2}\mathbf{a}_1$ (again using that $\mathbf{b}_1 = \mathbf{a}_1$ for an open-circuit termination), we can use this expression in Eq. (26.45) to obtain

$$\mathbf{b}_R = \frac{1}{2Z_0^{1/2}} (\mathbf{I} - \mathbf{S}_R) (\hat{\mathbf{v}}_s + \hat{\mathbf{v}}_i) \tag{26.46}$$

This gives us a mechanism for computing $\mathbf{b}_R$ based on our knowledge of the open-circuit antenna voltage vector and the S-parameter matrix of the coupled array.

The multiport matching network is described by the block S-parameter matrix

$$\mathbf{S}_M = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{bmatrix} \tag{26.47}$$

where 1 and 2 refer to input and output ports, respectively. With this representation, we use Eq. (26.44) with $\mathbf{b}_1 = \mathbf{S}_{11}\mathbf{a}_1 + \mathbf{S}_{12}\mathbf{a}_2$ to obtain

$$\mathbf{a}_1 = (\mathbf{I} - \mathbf{S}_R\mathbf{S}_{11})^{-1} (\mathbf{b}_R + \mathbf{S}_R\mathbf{S}_{12}\mathbf{a}_2) \tag{26.48}$$

Since $\mathbf{b}_2 = \mathbf{S}_{21}\mathbf{a}_1 + \mathbf{S}_{22}\mathbf{a}_2$ we have

$$\mathbf{b}_2 = \mathbf{S}_{21} (\mathbf{I} - \mathbf{S}_R\mathbf{S}_{11})^{-1} \mathbf{b}_R \underbrace{\left[\mathbf{S}_{22} + \mathbf{S}_{21} (\mathbf{I} - \mathbf{S}_R\mathbf{S}_{11})^{-1} \mathbf{S}_R\mathbf{S}_{12}\right]}_{\boldsymbol{\Gamma}_0} \mathbf{a}_2 \tag{26.49}$$

where we have used $\boldsymbol{\Gamma}_0$ to represent the reflection coefficient at the matching network output (see Figure 26.11).

The amplifier block introduces noise into the system, with the mth amplifier contributing forward and reverse noise waves $a_{t,m}$ and $b_{t,m}$, respectively, at the amplifier input [35]. The negative branch gain used to connect $\mathbf{a}_t$ into the flow diagram of Figure 26.12 is simply a convention used commonly in noise analysis [35]. Furthermore, the amplifier S-parameters are represented using the block matrix structure in Eq. (26.47), but with the blocks denoted by $\widehat{\mathbf{S}}_{ij}$. Using this notation, the amplifier output waves are expressed as

$$\mathbf{a}_2 = \widehat{\mathbf{S}}_{11}\mathbf{b}_2 + \widehat{\mathbf{S}}_{12}\mathbf{b}_L - \widehat{\mathbf{S}}_{11}\mathbf{a}_t + \mathbf{b}_t \tag{26.50}$$

$$\mathbf{a}_L = \widehat{\mathbf{S}}_{21}\mathbf{b}_2 + \widehat{\mathbf{S}}_{22}\mathbf{b}_L - \widehat{\mathbf{S}}_{21}\mathbf{a}_t \tag{26.51}$$

Inserting Eq. (26.50) with $\mathbf{b}_L = \boldsymbol{\Gamma}_L\mathbf{a}_L$ into Eq. (26.49) and rearranging leads to

$$\mathbf{b}_2 = \left(\mathbf{I} - \boldsymbol{\Gamma}_0\widehat{\mathbf{S}}_{11}\right)^{-1} \left[\mathbf{S}_{21} (\mathbf{I} - \mathbf{S}_R\mathbf{S}_{11})^{-1} \mathbf{b}_R + \boldsymbol{\Gamma}_0 \left(\mathbf{b}_t - \widehat{\mathbf{S}}_{11}\mathbf{a}_t + \widehat{\mathbf{S}}_{12}\boldsymbol{\Gamma}_L\mathbf{a}_L\right)\right] \tag{26.52}$$

Similarly, using $\mathbf{b}_L = \boldsymbol{\Gamma}_L\mathbf{a}_L$ in Eq. (26.51) leads to

$$\mathbf{a}_L = \left(\mathbf{I} - \mathbf{S}_{A,22}\boldsymbol{\Gamma}_L\right) \left(\widehat{\mathbf{S}}_{21}\mathbf{b}_2 - \widehat{\mathbf{S}}_{21}\mathbf{a}_t\right) \tag{26.53}$$

Substituting Eq. (26.52) into this result gives us the voltage $\mathbf{v}_L$ across the loads as

$$\begin{aligned} \mathbf{v}_L &= Z_0^{1/2} (\mathbf{I} + \boldsymbol{\Gamma}_L) \mathbf{a}_L \\ &= \mathbf{Q} \left[\mathbf{G} (\hat{\mathbf{v}}_s + \hat{\mathbf{v}}_i) + \boldsymbol{\Gamma}_0\mathbf{b}_t - \mathbf{a}_t\right] \end{aligned} \tag{26.54}$$

where

$$\mathbf{G} = \frac{1}{2Z_0^{1/2}}\mathbf{S}_{21}\left(\mathbf{I} - \mathbf{S}_R\mathbf{S}_{11}\right)^{-1}\left(\mathbf{I} - \mathbf{S}_R\right) \tag{26.55}$$

$$\mathbf{Q} = Z_0^{1/2}\left(\mathbf{I} + \boldsymbol{\Gamma}_L\right)\left[\left(\mathbf{I} - \boldsymbol{\Gamma}_0\widehat{\mathbf{S}}_{11}\right)\widehat{\mathbf{S}}_{21}^{-1}\left(\mathbf{I} - \widehat{\mathbf{S}}_{22}\boldsymbol{\Gamma}_L\right) - \boldsymbol{\Gamma}_0\widehat{\mathbf{S}}_{12}\boldsymbol{\Gamma}_L\right]^{-1} \tag{26.56}$$

We are now equipped with an expression representing the total signal plus noise across the loads given the signal and interference incident on the array and the thermal noise generated in the low noise amplifiers.

26.4.2.2 Matching Network Specification Armed with this expression for the output signal and noise voltages, we are prepared to discuss specification of the matching network, which is in reality coupled with a discussion on amplifier design. Practical amplifier design involves specifying a performance goal and synthesizing the source and load terminations that achieve this goal. Signal amplifiers are typically designed to provide minimum noise figure, optimal power gain, or some compromise between the two [36]. Our task is to define a desired value of $\boldsymbol{\Gamma}_0$, which is the source termination seen by the amplifier, and use this value to determine $\mathbf{S}_M$ for the matching network.

To begin this analysis, we take the SVD of each subblock using the representation $\mathbf{S}_{ij} = \mathbf{U}_{ij}\boldsymbol{\Lambda}_{ij}^{1/2}\mathbf{V}_{ij}^{\dagger}$, where the square root on the diagonal matrix of real singular values is used for notational convenience. Assuming that the matching network is lossless ($\mathbf{S}_M^{\dagger}\mathbf{S}_M = \mathbf{I}$) and reciprocal ($\mathbf{S}_M = \mathbf{S}_M^T$), it can be shown that the subblocks can be written

$$\begin{aligned}
\mathbf{S}_{11} &= \mathbf{U}_{11}\boldsymbol{\Lambda}_{11}^{1/2}\mathbf{U}_{11}^T \\
\mathbf{S}_{12} &= j\mathbf{U}_{11}\left(\mathbf{I} - \boldsymbol{\Lambda}_{11}\right)^{1/2}\mathbf{U}_{22}^T \\
\mathbf{S}_{21} &= j\mathbf{U}_{22}\left(\mathbf{I} - \boldsymbol{\Lambda}_{11}\right)^{1/2}\mathbf{U}_{11}^T \\
\mathbf{S}_{22} &= \mathbf{U}_{22}\boldsymbol{\Lambda}_{11}^{1/2}\mathbf{U}_{22}^T
\end{aligned} \tag{26.57}$$

With $\boldsymbol{\Gamma}_0 = \mathbf{U}_0\boldsymbol{\Lambda}_0^{1/2}\mathbf{U}_0^T$ representing the SVD of the reciprocal matrix $\boldsymbol{\Gamma}_0$, we use the expression for $\boldsymbol{\Gamma}_0$ in Eq. (26.49) and the relations in Eq. (26.57) to obtain

$$\boldsymbol{\Gamma}_0 = \mathbf{U}_{22}[\boldsymbol{\Lambda}_{11}^{1/2} - (\mathbf{I} - \boldsymbol{\Lambda}_{11})^{1/2}\mathbf{T}(\mathbf{I} - \boldsymbol{\Lambda}_{11})^{1/2}]\mathbf{U}_{22}^T \tag{26.58}$$

$$\mathbf{T} = \mathbf{U}_{11}^T(\mathbf{I} - \mathbf{S}_R\mathbf{U}_{11}\boldsymbol{\Lambda}_{11}^{1/2}\mathbf{U}_{11}^T)^{-1}\mathbf{S}_R\mathbf{U}_{11} \tag{26.59}$$

We have flexibility in specifying the singular vectors $\mathbf{U}_{11}$ and $\mathbf{U}_{22}$ and therefore choose representations that lead to mathematical simplicity. First, we see that if $\mathbf{S}_R = \mathbf{U}_R\boldsymbol{\Lambda}_R^{1/2}\mathbf{U}_R^T$ represents the SVD of the reciprocal matrix $\mathbf{S}_R$, then by choosing $\mathbf{U}_{11} = \mathbf{U}_R^*$ we obtain

$$\mathbf{T} = (\mathbf{I} - \boldsymbol{\Lambda}_R^{1/2}\boldsymbol{\Lambda}_{11}^{1/2})^{-1}\boldsymbol{\Lambda}_R^{1/2} \tag{26.60}$$

which is diagonal. If we further choose $\mathbf{U}_{22} = \mathbf{U}_0$, we can solve Eq. (26.58) to obtain

$$\boldsymbol{\Lambda}_{11}^{1/2} = (\boldsymbol{\Lambda}_0^{1/2} + \boldsymbol{\Lambda}_R^{1/2})(\mathbf{I} + \boldsymbol{\Lambda}_0^{1/2}\boldsymbol{\Lambda}_R^{1/2})^{-1} \tag{26.61}$$

We now have identified the values of all matrices in Eq. (26.57), which means that the matching network has been specified to achieve the design goals.

We generally assume that the amplifiers and loads are uncoupled ($\widehat{\mathbf{S}}_{ij}$ and $\mathbf{\Gamma}_L$ are diagonal), so that typical design goals are achieved for diagonal $\mathbf{\Gamma}_0$. If Γ_{opt} and Γ_{MS} represent the (scalar) source reflection coefficient for achieving amplifier minimum noise figure and maximum power gain [36], respectively, then these goals are achieved by setting $\mathbf{\Gamma}_0 = \Gamma_{\text{opt}}\mathbf{I}$ and $\mathbf{\Gamma}_0 = \Gamma_{\text{MS}}\mathbf{I}$. Since the performance of a multiantenna system depends on SNR, we expect a design for minimum noise figure to outperform one for maximum power gain. We also point out that if $\mathbf{\Gamma}_0 = \mathbf{0}$, then any reverse traveling noise from the transistor will not be reflected back into the amplifier, leading to perhaps reduced SNR compared to a design for optimal power gain. This condition can be obtained by setting $\mathbf{S}_{11} = \mathbf{S}_R^{\dagger}$.

To achieve diagonal $\mathbf{\Gamma}_0$, the matching network must be coupled to "undo" the coupling created by the antenna, and it therefore acts as an *array combining network* as well as an impedance transforming network. This is an important observation, since the linear combination of the signals in the matching network is performed *before* injection of the thermal noise by the amplifiers, implying that the matching network can have significant impact on the final system SNR. It is also interesting that if the network produces diagonal $\mathbf{\Gamma}_0$, the effective radiation patterns observed at the output of the matching network are mutually orthogonal, indicating that they can achieve perfect radiation efficiency as demonstrated in Ref. 37 as well as zero correlation in an environment where the multipath field is equally likely to arrive at any angle on the sphere [29]. These features of a coupled antenna along with a coupled matching network can therefore result in a diversity performance that is higher than what could be obtained with uncoupled antennas.

Finally, while optimal matching networks must be coupled, practically speaking it is easier to design an uncoupled network. We assume that the coupled antenna impedance can be represented using the diagonal matrix $\widetilde{\mathbf{Z}}_R$ containing only the diagonal elements of the full impedance matrix $\mathbf{Z}_R$. The resulting diagonal S-parameter matrix $\widetilde{\mathbf{S}}_R$ has elements $\widetilde{S}_{R,ii} = (\widetilde{Z}_{R,ii} - Z_0)/(\widetilde{Z}_{R,ii} + Z_0)$. This value of $\widetilde{\mathbf{S}}_R$ is then used in place of $\mathbf{S}_R$ when formulating the matching network S-parameter matrix. However, when analyzing the performance of this *self-impedance* match, the nondiagonal form of $\mathbf{S}_R$ must be used in Eq. (26.54). We should also mention that while the self-impedance matching approach provides a mechanism for creating an uncoupled matching network for a coupled antenna array, this does not generally represent the optimal uncoupled matching network in terms of diversity performance. Analytic and numerical optimizations can yield matching networks that outperform this self-impedance match for highly coupled antenna arrays [38].

26.4.2.3 Covariance Formation

Computing the impact of coupling and the matching network on the diversity performance requires that we formulate the covariance matrices for the signal and noise voltage vectors based on our transfer function of Eq. (26.54). For the thermal noise, we assume that the noise in each amplifier is uncorrelated with that of all other amplifiers, leading to the expressions

$$\begin{aligned} \text{E}\{\mathbf{a}_t\mathbf{a}_t^{\dagger}\} &= k_B B T_{\alpha}\mathbf{I} \\ \text{E}\{\mathbf{b}_t\mathbf{b}_t^{\dagger}\} &= k_B B T_{\beta}\mathbf{I} \\ \text{E}\{\mathbf{a}_t\mathbf{b}_t^{\dagger}\} &= k_B B T_{\Gamma}^{*}\mathbf{I} \end{aligned} \tag{26.62}$$

where k_B is Boltzmann's constant, B is the system noise power bandwidth, and T_α, T_β, and $T_\Gamma = T_\gamma e^{j\phi_\gamma}$ are amplifier effective noise temperatures that can readily be computed from other transistor noise parameters [35]. The covariance of the thermal noise voltage at the terminations is therefore

$$\widetilde{\mathbf{R}}_{\mathrm{t}} = \mathbf{Q} \underbrace{\mathrm{E}\left\{(\boldsymbol{\Gamma}_0\mathbf{b}_{\mathrm{t}} - \mathbf{a}_{\mathrm{t}})(\boldsymbol{\Gamma}_0\mathbf{b}_{\mathrm{t}} - \mathbf{a}_{\mathrm{t}})^\dagger\right\}}_{\widehat{\mathbf{R}}_{\mathrm{t}}} \mathbf{Q}^\dagger \tag{26.63}$$

$$\widehat{\mathbf{R}}_{\mathrm{t}} = k_B B(T_\alpha \mathbf{I} + T_\beta \boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0^\dagger - T_\Gamma \boldsymbol{\Gamma}_0 - T_\Gamma^* \boldsymbol{\Gamma}_0^\dagger) \tag{26.64}$$

For the signal and external interference, the output voltage covariances become

$$\widetilde{\mathbf{R}}_{\mathrm{s}} = \mathbf{Q}\mathbf{G}\mathbf{R}_{\mathrm{s}}\mathbf{G}^\dagger\mathbf{Q}^\dagger \tag{26.65}$$

$$\widetilde{\mathbf{R}}_{\mathrm{i}} = \mathbf{Q}\mathbf{G}\mathbf{R}_{\mathrm{i}}\mathbf{G}^\dagger\mathbf{Q}^\dagger \tag{26.66}$$

where the signal and noise covariance matrices are defined in Eqs. (26.26) and (26.28), respectively. These transformed covariance matrices can then be used to compute the diversity performance using the framework of Section 26.3.1.

26.4.2.4 Computational Example To demonstrate application of this framework and quantify the impact of coupling on the diversity performance, we conduct a simple computational example. The antennas used in this representative computation are two z-oriented half-wave (total length) dipoles, separated by the distance d, and each with a wire radius of 0.01λ. To accurately characterize these coupled antennas, we use the finite-difference time-domain (FDTD) [39, 40] computational approach with single-frequency antenna excitation. The FDTD grid uses 80 cells per wavelength in the z direction and 200 cells per wavelength in the x and y directions to adequately model the azimuthal current variations for close antenna spacings. A buffer region of a quarter-wavelength is placed between the antennas and the terminating eight-cell perfectly matched layer (PML) absorbing boundary condition (ABC). Pattern computations are performed when one antenna is excited while the second is terminated in an open circuit. The antenna S-parameter matrix $\mathbf{S}_R$ is computed with the antennas terminated in Z_0.

The incident field is vertically polarized with a power angular spectrum given by Eq. (26.37) with $\Delta\theta = 0$ and $\Delta\phi = \pi$, which means that the field is equally likely to arrive from any angle in the horizontal plane. We also assume that there is no external interference present in the environment. To maintain a consistent reference as antenna parameters (such as spacing) are swept, we characterize an isolated dipole using the FDTD approach and then compute the scalar variance $\widetilde{R}_{\mathrm{s}}$ for this single antenna in the propagation environment. For simplicity, we use a matching network corresponding to $\Gamma_0 = 0$ in this computation, leading to an average SNR for the single isolated dipole of $\widetilde{R}_{\mathrm{s}}/k_B B T_\alpha$. The reference CDF curves for the independent branch signals used in computation of the effective number of branches for the coupled system are then constructed based on this value of SNR.

The transistor used for the amplifier is a BJT taken from an application note [41]. At a collector–emitter bias voltage of 10 V, collector current of 4 mA, frequency of 4 GHz, and reference impedance of $Z_0 = 50\ \Omega$, the S-parameters and noise parameters are

given as

$$\begin{aligned} \widehat{\mathbf{S}}_{11} &= 0.552\angle 169^\circ \quad \widehat{\mathbf{S}}_{12} = 0.049\angle 23^\circ \\ \widehat{\mathbf{S}}_{21} &= 1.681\angle 26^\circ \quad \widehat{\mathbf{S}}_{22} = 0.839\angle -67^\circ \\ F_{\min} &= 2.5 \text{ dB} \quad \Gamma_{\text{opt}} = 0.475\angle 166^\circ \\ R_{\text{n}} &= 3.5\ \Omega \end{aligned} \tag{26.67}$$

where $F_{\min}$, Γ_{opt}, and R_{n} are the device minimum noise figure, optimal source termination for noise figure, and effective noise resistance, respectively. These parameters are converted to the effective noise temperatures T_α, T_β, and T_Γ using algebraic relations [35].

We now explore the impact of matching on the diversity performance of the coupled antennas as a function of the dipole spacing. In the examples, we use matching networks designed to achieve optimal amplifier noise figure and optimal power gain. Matching network synthesis is based on the full antenna coupling matrix $\mathbf{S}_R$ as well as the diagonal coupling matrix $\widetilde{\mathbf{S}}_R$ as discussed in Section 26.4.2.2 (self-impedance match).

Figure 26.13 plots the number of effective diversity branches as a function of dipole spacing for matching networks achieving optimal noise figure and power gain. Several observations regarding these results deserve attention. First, for very close antenna spacings, the two antennas behave largely as a single element, resulting in approximately one effective diversity branch. This is in contrast to the plot in Figure 26.9, where the fact that coupling was neglected resulted in the number of effective branches being larger than unity for zero spacing. The low diversity performance for close spacing increases rapidly with separation and for certain moderate spacings can actually exceed the diversity gain achieved for large element separation. This peak in the diversity performance stems from the pattern orthogonality created by the coupled antennas and matching network as discussed in Section 26.4.2.2, which leads to higher diversity than can be achieved with the uncoupled dipoles with the same spacing used as a reference in the diversity

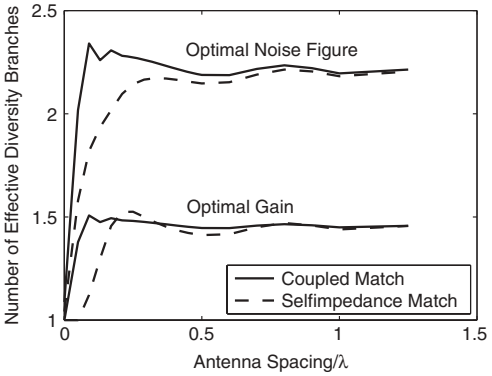


Figure 26.13 Number of effective diversity branches versus dipole spacing for matching networks that achieve optimal noise figure or optimal power gain for the amplifier. Curves are for optimal matching or for a matching network realized assuming the antenna impedance matrix is diagonal (SI = self-impedance).

gain computation. The height of this peak is also influenced by the fact that the reference antenna used for this computation has a suboptimal match of $\Gamma_0 = 0$, which means that the optimal match of $\mathbf{\Gamma}_0 = \Gamma_{\text{opt}}\mathbf{I}$ for the coupled antennas should outperform the reference dipole for larger element spacings.

The results of Figure 26.13 also show that matching to the self-impedance creates relatively little degradation in performance, particularly for element spacings larger than about $\lambda/4$. It is also particularly revealing that while optimal power transfer is a typical design goal, it is dramatically suboptimal in terms of diversity performance. This is an intuitive result, since matching for maximum power transfer neglects the impact of the match on amplifier noise figure, which directly controls the received SNR, the key parameter in determining the overall communication performance. This superiority of matching for minimum noise figure is therefore general for any receiving system equipped with practical noisy amplifiers.

26.4.3 MIMO System Modeling

To extend our discussion of antenna mutual coupling to the case of MIMO communication, we must augment our modeling framework by adding the transmitter and properly characterizing the MIMO communication channel. We use the same framework to perform this augmentation for the models in Sections 26.4.1 and 26.4.2, and then use this augmentation to explore some implications of antenna coupling on MIMO system performance.

26.4.3.1 Network Analysis Figure 26.14 illustrates an augmentation of the diversity block diagram from Figure 26.10 appropriate for characterizing a MIMO system. In this representation, the transmit array is characterized by the impedance matrix $\mathbf{Z}_T$ and is fed through a source impedance network by a set of generators creating the voltage $\mathbf{v}_S$. The antenna terminal voltage is related to the generator voltage vector as [31]

$$\mathbf{v}_T = \mathbf{Z}_T \left(\mathbf{Z}_T + \mathbf{Z}_S\right)^{-1} \mathbf{v}_S = \widetilde{\mathbf{C}}_T \mathbf{v}_S \tag{26.68}$$

where $\widetilde{\mathbf{C}}_T$ is a coupling matrix.

We can now combine this treatment with the analysis of Section 26.4.1 to arrive at a MIMO system transfer function. For example, if we assume that $H_{A,mn}$ represents the ratio of the open-circuit voltage $\hat{v}_m$ on the mth receive antenna to the driving point

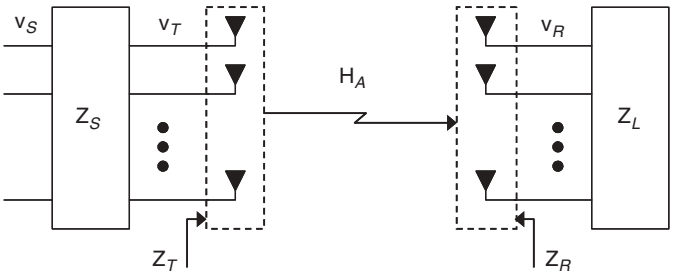


Figure 26.14 Augmentation of the block diagram of Figure 26.10 to represent a MIMO system with coupled transmit and receive arrays.

voltage $v_{T,n}$ on the nth transmit antenna, then the transfer function becomes simply (see Eq. (26.42))

$$\mathbf{v}_R = \mathbf{C}_R \mathbf{H}_A \widetilde{\mathbf{C}}_T \mathbf{v}_S \tag{26.69}$$

At this point, we must revisit the more general analysis of Sections 26.2.2 and 26.2.3, where we chose to formulate the channel transfer matrix $\mathbf{H}_A$ using transmit radiation patterns constructed with all other elements terminated in an *open circuit*. This allowed the superposition in Eq. (26.2) to properly represent the radiated field provided that the excitation vector $\mathbf{x}_A$ represents a driving point current. If, instead, we let the excitation vector represent a driving point voltage, then the radiation patterns used in the construction of $\mathbf{H}_A$ must be constructed with all other elements terminated in a *short circuit*. This approach is possible, but it has some practical disadvantages:

1. It is not particularly standard to characterize antenna patterns in this fashion, and in fact it can be challenging to obtain accurate numerical computations of such a scenario.
2. Since this short-circuit pattern will likely differ significantly from the pattern of the antenna in isolation, this approach does not lend itself to the approximation that the coupled element pattern is approximately the same as the isolated pattern.
3. It is common to assume that $\mathbf{H}_A$ consists of independent and identically distributed (i.i.d.) Gaussian random variables, which has been shown to be a reasonable approximation for dipole antennas characterized by their isolated element patterns. The radiation patterns of the elements in the presence of other short-circuited elements will generally be significantly perturbed from their isolated values, which indicates that the i.i.d. Gaussian assumption may not be valid under these circumstances.
4. The pattern characterization at the receiver (which requires open-circuit patterns) differs from that at the transmitter, which may be inconvenient in practical scenarios.

We can, instead, assume that $H_{A,mn}$ represents the ratio of the open-circuit voltage $\hat{v}_m$ on the mth receive antenna to the driving point current $i_{T,n}$ on the nth transmit antenna. We must modify Eq. (26.68) as

$$\mathbf{i}_T = (\mathbf{Z}_T + \mathbf{Z}_S)^{-1} \mathbf{v}_S = \mathbf{C}_T \mathbf{v}_S \tag{26.70}$$

The transfer matrix $\mathbf{H}_A$ now can be constructed using patterns with all other elements terminated in an open circuit, and the MIMO transfer relationship becomes

$$\mathbf{v}_R = \mathbf{C}_R \mathbf{H}_A \mathbf{C}_T \mathbf{v}_S \tag{26.71}$$

There have been a variety of studies based on this approach [30, 31, 33]. Perhaps most notable is the study of Janaswamy [31], which uses a detailed analysis based on the MIMO transfer relationship of Eq. (26.69). This analysis reveals that increasing the number of antennas in a uniform linear array of dipoles (of fixed aperture size) cannot increase the capacity beyond a certain limit due to the impact of coupling. This is intuitive,

since mutual coupling limits the ability of closely spaced elements to independently excite or sample the propagation environment.

The same approach can be used to augment the model of Section 26.4.2, which includes a detailed description of the receiver front-end network. Figure 26.15 shows the block diagram representation for the MIMO system, where the receiver architecture is identical to that detailed in Figure 26.11. Again assuming that the transmit element radiation patterns are obtained with all other elements terminated in an open circuit and that the elements of the transfer matrix $\mathbf{H}_A$ represent the ratio of the open-circuit receive antenna voltages to the driving point currents, the model of Eq. (26.54) is simply augmented by recognizing that the open-circuit signal voltage is $\hat{\mathbf{v}}_\mathrm{s} = \mathbf{H}_A\mathbf{C}_T\mathbf{v}_S$. The voltage transfer function therefore becomes

$$\mathbf{v}_L = \mathbf{Q}\left[\mathbf{G}\mathbf{H}_A\mathbf{C}_T\mathbf{v}_S + \mathbf{G}\hat{\mathbf{v}}_\mathrm{i} + \boldsymbol{\Gamma}_0\mathbf{b}_\mathrm{t} - \mathbf{a}_\mathrm{t}\right] \tag{26.72}$$

We use this expression for the remainder of this discussion on mutual coupling for MIMO systems since it provides a detailed representation of the interference and thermal noise.

26.4.3.2 MIMO System Capacity Computing the capacity for the MIMO signal relationship in Eq. (26.72) requires that we once again formulate the covariance of $\mathbf{v}_L$ assuming the input signal $\mathbf{v}_S$, interference, and noise are drawn from zero-mean complex Gaussian distributions. The interference and noise covariances are given by Eq. (26.63) and (26.66), respectively. If we let $\mathbf{R}_S = \mathrm{E}\{\mathbf{v}_S\mathbf{v}_S^\dagger\}$ represent the covariance of the input signal, then the covariance of the signal at the receiver loads is given by

$$\widetilde{\mathbf{R}}_\mathrm{s} = \mathbf{Q}\mathbf{G}\mathbf{H}_A\mathbf{C}_T\mathbf{R}_S\mathbf{C}_T^\dagger\mathbf{H}_A^\dagger\mathbf{G}^\dagger\mathbf{Q}^\dagger = \mathbf{Q}\widehat{\mathbf{H}}_A\mathbf{R}_S\widehat{\mathbf{H}}_A^\dagger\mathbf{Q}^\dagger \tag{26.73}$$

The capacity for this MIMO system is obtained by determining the matrix $\mathbf{R}_S$ that maximizes the mutual information expression

$$I(\mathbf{v}_L, \mathbf{v}_S) = \log_2 \frac{\left|\mathbf{Q}\left[\widehat{\mathbf{H}}_A\mathbf{R}_S\widehat{\mathbf{H}}_A^\dagger + \widehat{\mathbf{R}}_\eta\right]\mathbf{Q}^\dagger\right|}{\left|\mathbf{Q}\widehat{\mathbf{R}}_\eta\mathbf{Q}^\dagger\right|} \tag{26.74}$$

$$= \log_2\left|\widehat{\mathbf{R}}_\eta^{-1/2}\widehat{\mathbf{H}}_A\mathbf{R}_S\widehat{\mathbf{H}}_A^\dagger\left(\widehat{\mathbf{R}}_\eta^{-1/2}\right)^\dagger + \mathbf{I}\right| \tag{26.75}$$

where $\widehat{\mathbf{R}}_\eta = \mathbf{G}\mathbf{R}_\mathrm{i}\mathbf{G}^\dagger + \widehat{\mathbf{R}}_\mathrm{t}$ and we have assumed that the matrix $\mathbf{Q}$ is nonsingular.

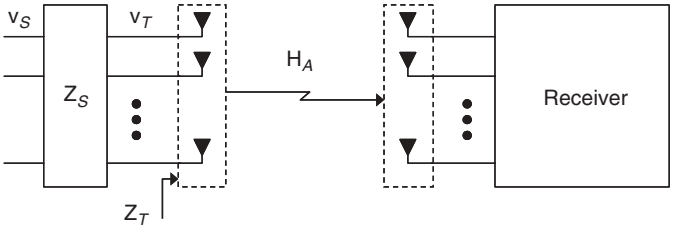


Figure 26.15 Block diagram of a MIMO system including the coupled transmit and receive arrays, MIMO channel, and the receiver network (detailed in Figure 26.11).

Equation (26.75) is now in a form that can be maximized using the water-filling procedure referred to in Section 26.2.4. We note, however, that this optimization is coupled with a power constraint that limits the diagonal elements of the transmit signal covariance matrix $\mathbf{R}_S$. While this is a reasonable constraint for uncoupled antennas, it should be noted that for coupled antennas this will not limit the amount of power radiated by the array. Since this radiated power is what is specified by regulating authorities and is the power that must be provided by the generators, it is logical to formulate the capacity based on the power radiated.

Assuming lossless antennas, the power radiated by the array averaged over one sinusoidal cycle can be computed from

$$p_{\text{rad}} = \tfrac{1}{2}\text{Re}\{\mathbf{i}_T^\dagger \mathbf{Z}_T \mathbf{i}_T\} = \tfrac{1}{4}\{\mathbf{i}_T^\dagger \mathbf{Z}_T \mathbf{i}_T + \mathbf{i}_T^T \mathbf{Z}_T^* \mathbf{i}_T^*\} \tag{26.76}$$

$$= \tfrac{1}{4}\mathbf{i}_T^\dagger \{\mathbf{Z}_T + \mathbf{Z}_T^\dagger\}\mathbf{i}_T \tag{26.77}$$

$$= \tfrac{1}{2}\mathbf{i}_T^\dagger \text{Re}\{\mathbf{Z}_T\}\mathbf{i}_T = \mathbf{v}_S^\dagger \mathbf{A} \mathbf{v}_S \tag{26.78}$$

where we have used Eq. (26.70) and $\mathbf{A} = \frac{1}{2}\mathbf{C}_T^\dagger \text{Re}\{\mathbf{Z}_T\}\mathbf{C}_T$. In this formulation, we have taken the transpose of the second (scalar) term on the right-hand side of Eq. (26.76) and have used that $\mathbf{Z}_T$ is symmetric for reciprocal antennas. For zero-mean signals, the average radiated power is given by

$$P_{\text{rad}} = \text{E}\{p_{\text{rad}}\} = \text{Tr}[\mathbf{R}_S \mathbf{A}] = \text{Tr}\left[\mathbf{A}^{1/2}\mathbf{R}_S\mathbf{A}^{1/2}\right] \tag{26.79}$$

where we have used the facts that $\mathbf{A}$ is real and positive semidefinite and that matrices commute under the trace. Defining $\widehat{\mathbf{R}}_S = \mathbf{A}^{1/2}\mathbf{R}_S\mathbf{A}^{1/2}$ and substitution into Eq. (26.75) leads to

$$I(\mathbf{v}_L, \mathbf{v}_S) = \log_2 \det\left\{\mathbf{H}\widehat{\mathbf{R}}_S\mathbf{H}^\dagger + \mathbf{I}\right\} \tag{26.80}$$

where $\mathbf{H} = \widehat{\mathbf{R}}_\eta^{-1/2}\widehat{\mathbf{H}}_A\mathbf{A}^{-1/2}$. The radiated power constraint is simply $\text{Tr}\left[\widehat{\mathbf{R}}_S\right] < P_T$. The problem is now in a form for which the capacity can be computed using the framework of Section 26.2.4.

26.4.3.3 Computational Example We can demonstrate the impact of coupling on the capacity using the receive antenna and amplifier configuration used in Section 26.4.2.4 with the addition of a transmit array, which is also a two-element array of half-wave dipoles with a half-wave element spacing. Once again, the interference is assumed to be zero. The channel matrix $\mathbf{H}_A$ is constructed as a random realization of the SVA channel model discussed in Section 26.2.3, with the capacity results representing an average over 5000 such channel realizations.

For each physical channel realization, a single transmit and a single receive dipole are used in conjunction with a lossless receive matching network with $S_{11} = S_R^*$ so that $\Gamma_0 = 0$ (all terms are scalars). The single-input single-output (SISO) SNR is then taken as the square of the received signal voltage to the variance of the thermal noise, or

$$\text{SNR} = \frac{|G H_A C_T v_S|^2}{k_B B T_\alpha} = \frac{|(1 - S_R) H_A|^2}{2Z_0\left(1 - |S_R|^2\right)\text{Re}\{Z_T\}}\frac{P_T}{k_B B T_\alpha} \tag{26.81}$$

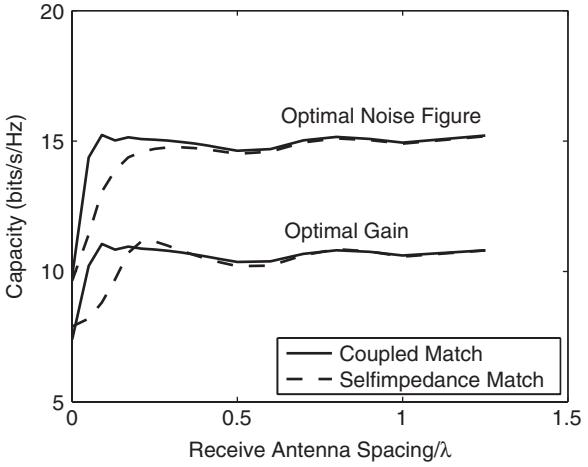


Figure 26.16 Average capacity as a function of receive dipole separation with mutual coupling (optimal and self-impedance match). Matching for both optimal noise figure and optimal power gain are considered.

where we have used that $|S_{21}|^2 = 1 - |S_{11}|^2$. This SNR value is then averaged in space by moving each dipole in 0.1λ steps over a linear range of 1.5λ. For a given transmit power, the value of $k_B B$ can be computed to achieve an average SISO SNR (20 dB in this work) for the channel realization.

For each receive dipole spacing, we construct the matching network to achieve the specified design goal as outlined in Section 26.4.2.2. Figure 26.16 plots the capacity as a function of receive dipole spacing for matching networks that achieve minimum noise figure and maximum amplifier gain. Results for a coupled match and a simpler self-impedance match are included. We first observe that the match achieving minimum amplifier noise figure (noise figure of $F = F_{\min} = 2.5$ dB) produces notably higher capacity than the match providing maximum power transfer, which generates a much higher noise figure of $F = 7.2$ dB. This result is intuitive, since ultimately capacity depends on SNR as opposed to absolute signal strength, as discussed in Section 26.4.2.4. We also observe that for close antenna spacings with high coupling, the shortcomings of the self-impedance match are evident. However, once the spacing reaches approximately $d = \lambda/4$, this match provides near optimal performance.

26.5 SUPERDIRECTIVITY IN MIMO SYSTEMS

The discussion in Section 26.4.3.2 revealed that constraining the square of the transmit antenna excitations, as is traditionally done in MIMO system capacity formulations, does not limit the power radiated by the antenna. Since it is this radiated power that is limited by regulatory agencies and since the transmit generators (amplifiers) must supply this power, it is logical that the capacity formulation should directly constrain the radiated power. This section discusses this issue in additional detail and demonstrates that when the radiated power is constrained the capacity solution specifies superdirective excitations [42–46] characterized by very high array directivity in preferred directions.

It also illustrates that external interference can lead to superdirective array weights at the receiver.

While the phenomenon of superdirectivity theoretically allows the system to advantageously exploit the propagation channel spatial characteristics, its implementation is typically considered impractical for a variety of reasons [43]. Therefore this section provides two mechanisms for computing the capacity when the level of superdirectivity is limited.

26.5.1 Capacity Expression

The goal of this formulation is to provide a simpler capacity expression than that obtained for a MIMO system with coupled antennas while exposing the mechanisms that lead to array superdirectivity. For this development, we assume that the excitation signal is the antenna driving point current $\mathbf{i}_T$. As outlined in Section 26.4.3.2, one mechanism for computing the radiated power (averaged over one sinusoidal cycle) is to use the mutual impedance matrix $\mathbf{Z}_T$ for the transmit array as

$$p_{\text{rad}} = \tfrac{1}{2}\mathbf{i}_T^{\dagger}\text{Re}\{\mathbf{Z}_T\}\mathbf{i}_T = \mathbf{i}_T^{\dagger}\mathbf{A}\mathbf{i}_T \tag{26.82}$$

where $\mathbf{A} = \frac{1}{2}\text{Re}\{\mathbf{Z}_T\}$.

In situations where the impedance matrix is unknown, it is also possible to compute the radiated power based on the transmitted field. If $\overline{\mathbf{e}}_{T,n}(\Omega)$ represents the radiation pattern of the nth transmit element with all other elements terminated in an open circuit, the total transmitted field is

$$\overline{\mathbf{e}}_T(\Omega_T) = \sum_{n=1}^{N_T} \overline{\mathbf{e}}_{T,n}(\Omega_T) i_{T,n} \tag{26.83}$$

The radiated power is

$$\begin{aligned} p_{\text{rad}} &= \frac{1}{2\eta_0}\int_{\Omega_T} \overline{\mathbf{e}}_{\text{t}}^{*}(\Omega_T)\cdot\overline{\mathbf{e}}_{\text{t}}(\Omega_T)\, d\Omega_T \\ &= \sum_{n=1}^{N_T}\sum_{q=1}^{N_T} i_{T,n}^{*} \underbrace{\left[\frac{1}{2\eta_0}\int_{\Omega_T} \overline{\mathbf{e}}_{T,n}^{*}(\Omega_T)\cdot\overline{\mathbf{e}}_{T,q}(\Omega_T)\, d\Omega_T\right]}_{A_{nq}} i_{T,q} \end{aligned} \tag{26.84}$$

where η_0 is the intrinsic impedance of free space. In either case, the average radiated power is

$$P_{\text{rad}} = \text{Tr}\left[\mathbf{R}_T\mathbf{A}\right] \tag{26.85}$$

where $\mathbf{R}_T = \text{E}\{\mathbf{i}_T\mathbf{i}_T^{\dagger}\}$ is the covariance of the transmit currents. If we represent our transmit currents instead as $\mathbf{i}_T = \mathbf{A}^{-1/2}\hat{\mathbf{i}}_T$, where $\hat{\mathbf{i}}_T$ has covariance $\widehat{\mathbf{R}}_T = \text{E}\{\hat{\mathbf{i}}_T\hat{\mathbf{i}}_T^{\dagger}\}$, then we can recast this power constraint to

$$P_{\text{rad}} = \text{Tr}\left[\mathbf{A}^{-1/2}\widehat{\mathbf{R}}_T\mathbf{A}^{-1/2}\mathbf{A}\right] = \text{Tr}\left[\widehat{\mathbf{R}}_T\right] \tag{26.86}$$

where we have used the fact that $\mathbf{A}$ is real and positive semidefinite. It is this quantity that will be constrained in the capacity formulation.

We assume that the noise consists of external interference and that the element H_{mn} of the channel matrix represents the ratio of the open-circuit voltage on the mth receive antenna to the driving point current on the nth transmit antenna. In this case, the transfer relationship between the open-circuit receive voltage and the transmit current is simply

$$\hat{\mathbf{v}} = \mathbf{H}\mathbf{i}_T + \hat{\mathbf{v}}_\mathrm{i} = \hat{\mathbf{v}}_\mathrm{s} + \hat{\mathbf{v}}_\mathrm{i} \tag{26.87}$$

Following the development of Eq. (26.28), the covariance of the open-circuit interference voltage $\hat{\mathbf{v}}_\mathrm{i}$ at the receive antenna terminals is given by

$$R_{\mathrm{i},mp} = |\varphi|^2 \int \bar{\mathbf{e}}_{R,m}(\Omega_R) \cdot \overline{\overline{\mathbf{P}}}_\mathrm{i}(\Omega_R) \cdot \bar{\mathbf{e}}^*_{R,p}(\Omega_R)\, d\Omega_R \tag{26.88}$$

where $\overline{\overline{\mathbf{P}}}_\mathrm{i}(\Omega_R)$ represents the interference power angular spectrum. Using a spatial prewhitening filter for the noise and introducing our modified current definition transforms the transfer relationship to

$$\hat{\mathbf{v}}_0 = \mathbf{R}_\mathrm{i}^{-1/2}\mathbf{H}\mathbf{A}^{-1/2}\hat{\mathbf{i}}_T + \mathbf{R}_\mathrm{i}^{-1/2}\hat{\mathbf{v}}_\mathrm{i} \tag{26.89}$$

The mutual information for this transfer relationship can be maximized subject to a constraint on the radiated power defined in Eq. (26.86), as discussed in Section 26.2.4.

The effect of the radiated power constraint and interference can be observed from this mutual information expression. Small eigenvalues in $\mathbf{A}$ or $\mathbf{R}_\mathrm{i}$ will lead to spatial subchannels with high gain in the effective channel matrix $\mathbf{R}_\mathrm{i}^{-1/2}\mathbf{H}\mathbf{A}^{-1/2}$. When the transmit covariance $\mathbf{R}_T$ is constructed in the capacity solution, it will tend toward a solution that exploits these high gain channels. It is therefore instructive to understand the physical mechanisms leading to this high channel gain.

26.5.2 Superdirectivity

The matrix quantities $\mathbf{A}$ and $\mathbf{R}_\mathrm{i}$ can be linked to the phenomenon of array superdirectivity. We assume that all of the elements are identical, allowing us to construct the pattern overlap matrices, which for the transmitter and receiver as

$$\mathbf{A}_T = \mathbf{A}/A_{11} \tag{26.90}$$

$$\mathbf{A}_R = \mathbf{R}_\mathrm{i}^*/R_{\mathrm{i},11},^* \qquad \text{with } \mathbf{R}_\mathrm{i} \text{ computed using } \overline{\overline{\mathbf{P}}}_\mathrm{i}(\Omega) = 1 \tag{26.91}$$

The superdirectivity Q factors [44–46] for the transmit array with a vector of transmit currents $\mathbf{i}_T$ and for the receive array with applied beam-forming weights $\mathbf{w}_R$ are then given as

$$Q_T = \frac{\mathbf{i}_T^\dagger \mathbf{i}_T}{\mathbf{i}_T^\dagger \mathbf{A}_T \mathbf{i}_T} \tag{26.92}$$

$$Q_R = \frac{\mathbf{w}_R^\dagger \mathbf{w}_R}{\mathbf{w}_R^\dagger \mathbf{A}_R \mathbf{w}_R} \tag{26.93}$$

The product of the array Q factor and the quality factor Q_{e} of the individual array elements approximates the quality factor of the antenna array for the excitation or weight vector [42]. Therefore a high Q factor corresponds to a small usable bandwidth. For example, suppose we use an element that has $Q_e = 10$ when operating in isolation. This corresponds to a 10% frequency bandwidth, something easily obtainable by practical elements such as a half-wave dipole. If the array configuration leads to a modest Q factor of 10, the overall array quality factor will be 100, leading to a frequency bandwidth of only 1%. Therefore the attempt to use superdirectivity to enhance system performance will in most cases fail due to this bandwidth reduction. Since the goal of using MIMO technology is to obtain high spectral efficiency, this severe bandwidth reduction can be considered counterproductive to this fundamental goal.

When the array is used for information communication, the excitation and weighting and therefore the array Q factors are time variant. We can, however, gain insight into the value of the Q factor from $\mathbf{A}_T$ or $\mathbf{A}_R$ (which depend only on the array properties). Considering the transmit array, we let the EVD of $\mathbf{A}_T$ be represented by $\mathbf{A}_T = \boldsymbol{\xi}_T \boldsymbol{\Lambda}_T \boldsymbol{\xi}_T^{\dagger}$ [13], where $\boldsymbol{\xi}_T$ is a unitary matrix of eigenvectors and $\boldsymbol{\Lambda}_T$ is a diagonal matrix of real eigenvalues (since $\mathbf{A}_T$ is Hermitian). Suppose that at one instant in time, the excitation is the pth eigenvector (column of $\boldsymbol{\xi}_T$), which can be expressed mathematically as $\mathbf{i}_T = \boldsymbol{\xi}_{T,p}$. The Q factor for this excitation will is $Q_T = 1/\Lambda_{T,pp}$. The array Q factor is therefore large (indicative of superdirectivity) when the current is aligned with eigenvectors associated with small eigenvalues and reaches a maximum value of $Q_{T,\max} = 1/\Lambda_{T,\min}$, where $\Lambda_{T,\min}$ represents the smallest eigenvalue in $\boldsymbol{\Lambda}_T$. Referring to our discussion in Section 26.5.1, where we suggested that the small eigenvalues of $\mathbf{A}$ create spatial subchannels with high gain, it becomes clear that this high gain is created by transmit array superdirectivity.

Naturally, an analogous discussion is possible for the receive array. We must recognize, however, that strictly speaking the matrix $\mathbf{R}_{\mathrm{i}}$, which accounts for the high gain discussed in Section 26.5.1, is not the same as the matrix $\mathbf{A}_R$ unless the average multipath power arriving at the receiver is uniformly distributed in angle. However, the similarity between the definitions of $\mathbf{A}_R$ and $\mathbf{R}_{\mathrm{i}}$ suggests that the small eigenvalues in $\mathbf{R}_{\mathrm{i}}$ leading to high gain are also representative of superdirective effects.

It is important to recognize that this array superdirectivity, while intriguing from a theoretical standpoint, is associated with several practical problems. For example, superdirective excitations are characterized by large antenna current magnitudes (which leads to high ohmic loss), extreme sensitivity to the excitation weights, and narrow operating bandwidth [43]. To gain a feel for the level of superdirectivity that is achievable for compact arrays, consider a uniform circular transmit array of 16 Hertzian dipoles, where the array diameter is $D = \lambda/2$. Figure 26.17 plots the inverse of the eigenvalues of $\mathbf{A}_T$ for this array. These results clearly show that the array Q factors can become very large. This motivates identifying a mechanism for limiting array superdirectivity in the capacity solution to enable identification of the maximum *achievable* capacity given practical constraints.

26.5.3 Limiting Superdirectivity: Beam Formers

One method for limiting the influence of superdirectivity on the capacity solution is to use transmit and receive beam formers that disallow superdirective weights [47, 48]. For the following, we define Q_{t} and Q_{r} as the highest allowable Q factor for the transmit

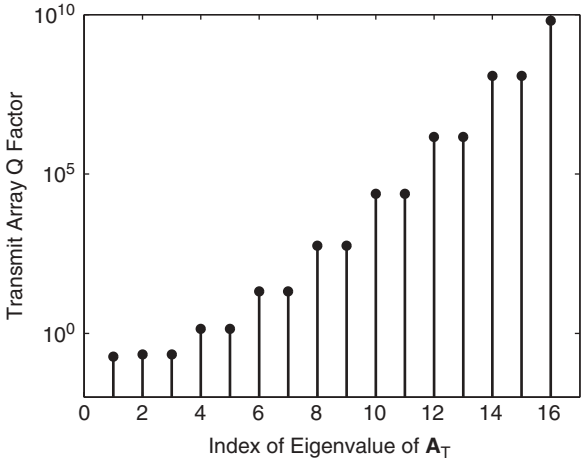


Figure 26.17 Q factors (inverse eigenvalues) associated with each of the eigenvectors of $\mathbf{A}_T$ for a 16-element uniform circular array with diameter $D = \lambda/2$.

and receive arrays, respectively. The matrix $\widehat{\boldsymbol{\xi}}_T$ represents the eigenvectors in $\boldsymbol{\xi}_T$ associated with eigenvalues in $\boldsymbol{\Lambda}_T$ that are greater than $1/Q_{\mathrm{t}}$. Similarly, $\widehat{\boldsymbol{\xi}}_R$ represents the eigenvectors in $\boldsymbol{\xi}_R$ associated with eigenvalues in $\boldsymbol{\Lambda}_R$ that are greater than $1/Q_{\mathrm{r}}$.

We can limit the transmit superdirectivity to have a Q factor below Q_{t} by requiring the excitation $\mathbf{i}_T$ to remain within the subspace spanned by $\widehat{\boldsymbol{\xi}}_T$. We therefore can modify our representation of the transmit current to (see the discussion leading up to Eq. (26.86))

$$\mathbf{i}_T = \mathbf{A}^{-1/2}\widehat{\boldsymbol{\xi}}_T\hat{\mathbf{i}}_T \tag{26.94}$$

The resulting modification to the power constraint becomes

$$\mathrm{Tr}\left[\mathbf{R}_T\mathbf{A}\right] = \mathrm{Tr}\left[\mathbf{A}^{-1/2}\widehat{\boldsymbol{\xi}}_T\widehat{\mathbf{R}}_T\widehat{\boldsymbol{\xi}}_T^{\dagger}\mathbf{A}^{-1/2}\mathbf{A}\right] = \mathrm{Tr}\left[\widehat{\mathbf{R}}_T\right] \tag{26.95}$$

where we have used $\widehat{\boldsymbol{\xi}}_T^{\dagger}\widehat{\boldsymbol{\xi}}_T = \mathbf{I}$.

At the receiver, we apply the beam former represented by $\widehat{\boldsymbol{\xi}}_R^{\dagger}$ to create the signal

$$\hat{\mathbf{v}}' = \widehat{\boldsymbol{\xi}}_R^{\dagger}\hat{\mathbf{v}} = \widehat{\boldsymbol{\xi}}_R^{\dagger}\mathbf{H}\mathbf{A}^{-1/2}\widehat{\boldsymbol{\xi}}_T\hat{\mathbf{i}}_T + \widehat{\boldsymbol{\xi}}_R^{\dagger}\hat{\mathbf{v}}_{\mathrm{i}} \tag{26.96}$$

where the transformed noise has covariance $\widehat{\mathbf{R}}_{\mathrm{i}} = \widehat{\boldsymbol{\xi}}_R^{\dagger}\mathbf{R}_{\mathrm{i}}\widehat{\boldsymbol{\xi}}_R$. With this projection, any subsequently applied receive beam-forming weights characterized by a Q factor above Q_{r} will lie in the null space of $\hat{\mathbf{v}}'$ and therefore will not contribute to the capacity. Application of the prewhitening filter gives

$$\hat{\mathbf{v}}_0 = \widehat{\mathbf{R}}_{\mathrm{i}}^{-1/2}\hat{\mathbf{v}}' = \underbrace{\widehat{\mathbf{R}}_{\mathrm{i}}^{-1/2}\widehat{\boldsymbol{\xi}}_R^{\dagger}\mathbf{H}\mathbf{A}^{-1/2}\widehat{\boldsymbol{\xi}}_T}_{\widehat{\mathbf{H}}}\hat{\mathbf{i}}_T + \underbrace{\widehat{\mathbf{R}}_{\mathrm{i}}^{-1/2}\widehat{\boldsymbol{\xi}}_R^{\dagger}\hat{\mathbf{v}}_{\mathrm{i}}}_{\hat{\mathbf{v}}_{\mathrm{i}}'} \tag{26.97}$$

where the noise $\hat{\mathbf{v}}_{\mathrm{i}}'$ has covariance $\mathbf{I}$. The mutual information between the transmit current and the open-circuit received voltage is given by

$$I(\hat{\mathbf{v}}_0, \hat{\mathbf{i}}_T) = \log_2 \left| \widehat{\mathbf{H}} \widehat{\mathbf{R}}_T \widehat{\mathbf{H}}^{\dagger} + \mathbf{I} \right| \tag{26.98}$$

This expression can be maximized subject to the constraint $\mathrm{Tr}[\widehat{\mathbf{R}}_T] \leq P_T$ to obtain the capacity using the water-filling solution discussed in Section 26.2.4.

We assume that the interference power is uniformly distributed in the horizontal plane or $\overline{\overline{\mathbf{P}}}_{\mathrm{i}}(\Omega_R) = \sigma_{\mathrm{i}}^2 \delta(\theta_R - \pi/2)$. To specify the channel average SISO SNR, we construct the scalars A and R_{i} for a single dipole at transmit and receive. The average radiated power is $P_T = |i_T|^2 A$ and the received noise power is given by R_{i}. If the channel transfer function for this SISO case is the scalar H, then the received SISO SNR is

$$\mathrm{SNR}_{\mathrm{isotropic}} = \frac{|H i_T|^2}{R_{\mathrm{i}}} = \frac{|H|^2 P_T}{A R_{\mathrm{i}}} \tag{26.99}$$

The value of σ_{i}^2 can then be adjusted to produce the desired SISO SNR for each channel realization.

The transmit and receive arrays are the same 16-element circular array of Hertzian dipoles with an array diameter of $D = \lambda/2$, as was considered for Figure 26.17. We compute the capacity averaged over 500 channel realizations of the SVA channel model as a function of the threshold Q factors $Q_{\mathrm{t}} = Q_{\mathrm{r}}$. Figure 26.18 plots this capacity for the isotropic external noise field using the water-filling and uninformed transmitter capacity solutions. The jumps in the capacity occur when the threshold is increased enough to increase the dimensionality of $\widehat{\boldsymbol{\xi}}_T = \widehat{\boldsymbol{\xi}}_R$. As expected, the water-filling solution, which exploits channel state information at the transmitter, is larger than the capacity for the uninformed transmitter, although the difference at this large SNR of 20 dB is relatively small [14]. This result illustrates the dramatic impact of superdirectivity on the capacity performance.

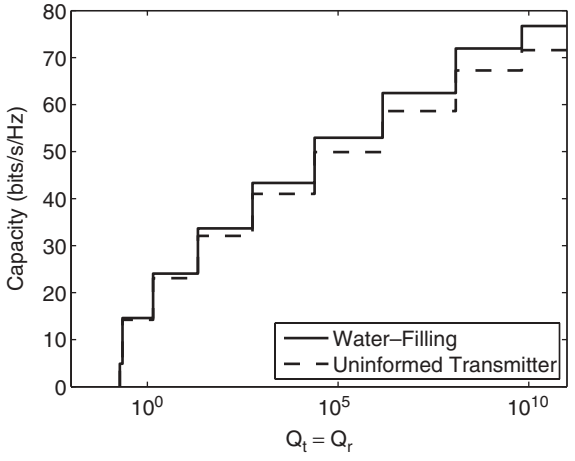


Figure 26.18 Capacity (averaged over 500 channel realizations) for a 16-element circular array with diameter $D = \lambda/2$ as a function of $Q_{\mathrm{t}} = Q_{\mathrm{r}}$ for different capacity solutions.

26.5.4 Limiting Superdirectivity: Antenna Loss

While the beam-forming approach discussed in Section 26.5.3 clearly is effective for limiting the level of superdirectivity used by the MIMO system, this solution is not optimal. Specifically, it is possible to form currents from a linear combination of the vectors from the superdirective and nonsuperdirective spaces that achieve an overall Q factor that is below the given threshold. Stated another way, the beam former limits the excitation currents or receive weights to lie in a subspace, while the actual constraint should limit these vectors to an ellipsoid in the multidimensional space. Unfortunately, there does not appear to be an obvious way to achieve the optimal solution using the beam-forming approach combined with the capacity solution.

Because transmit superdirectivity solutions are characterized by high current magnitudes (for a given radiated power), the loss introduced by even a small antenna resistance makes superdirective excitations inefficient and unfavorable relative to nonsuperdirective ones. At the receiver, ohmic loss leads to spatially white thermal noise that will remove receive superdirective solutions. Because of the difference in the effects at the transmitter and receiver, we develop the impact of antenna loss at each end of the link separately.

26.5.4.1 Transmitter Incorporating transmit antenna loss as part of the channel implies that the capacity formulation power constraint must limit the power *delivered* to the transmit array rather than the power *radiated*, since some of the power is consumed by antenna loss. We formulate the capacity under this delivered constraint in this section and then discuss an approach for compensating for the reduction in radiated power in Section 26.5.4.3. Recognizing that the diagonal elements of $\mathbf{A}$ represent the radiation resistance of each antenna, we can write that the average power delivered to the antenna array:

$$P_{\text{in}} = \text{Tr}\left[\mathbf{R}_T \widehat{\mathbf{A}}\right] \tag{26.100}$$

where $\widehat{\mathbf{A}} = \mathbf{A} + \mathbf{L}_T$. The nth element of the diagonal matrix $\mathbf{L}_T$ represents one-half the antenna ohmic loss resistance for the nth transmit antenna. We emphasize that this is the physical antenna resistance, which can be obtained from radiation efficiency measurements for practical scenarios [49]. For arrays constructed of identical elements, this matrix is $\mathbf{L}_T = L_T\mathbf{I}$, where L_T is half the loss resistance of each element.

If the delivered power is constrained in the capacity formulation, then $\widehat{\mathbf{A}}$ replaces $\mathbf{A}$ in Eq. (26.89). The addition of the diagonal matrix $\mathbf{L}_T$ in $\widehat{\mathbf{A}}$ eliminates the very small eigenvalues associated with superdirectivity and therefore regularizes the matrix inverse $\widehat{\mathbf{A}}^{-1/2}$ even when the antenna loss is modest. This is a mathematical indication of the fact that superdirective solutions exhibit high loss and become unfavorable relative to more traditional excitations.

For the example computations that follow, all array elements are assumed identical so that A_{nn} is the same for all n. We can therefore rearrange $\widehat{\mathbf{A}}$ as

$$\widehat{\mathbf{A}} = \mathbf{A} + A_{11}\,(1/\mu_T - 1)\,\mathbf{I} \tag{26.101}$$

where $\mu_T = A_{11}/(A_{11} + L_T)$ is the single element efficiency [49]. This allows demonstration of the impact of loss on superdirectivity as a function of this practical efficiency parameter.

26.5.4.2 Receiver At the receiver, ohmic loss does not explicitly change the possibility of observing receive superdirectivity in the capacity solution since the loss operates identically on the signal and the external interference. However, in this case the resistance adds a thermal noise component to the received signal that must be modeled correctly. Specifically, if the receive array is characterized by a diagonal ohmic loss resistance matrix $\mathbf{L}_R$ (where each matrix element represents half the ohmic loss of the corresponding antenna), then an open-circuit noise voltage vector $\hat{\mathbf{v}}_\mathrm{t}$ is introduced so that the received signal becomes

$$\hat{\mathbf{v}} = \mathbf{H}\mathbf{i}_T + \hat{\mathbf{v}}_\mathrm{i} + \hat{\mathbf{v}}_\mathrm{t} \tag{26.102}$$

Since the noise on each antenna is assumed independent of the noise on all other antennas, the covariance of this noise is [35]

$$\mathbf{R}_\mathrm{t} = 4k_B BT(2\mathbf{L}_R) \tag{26.103}$$

Given this spatially white thermal noise contribution, which is assumed independent of the external interference, the total interference plus noise has covariance

$$\mathbf{R}_\eta = \mathbf{R}_\mathrm{i} + \mathbf{R}_\mathrm{t} \tag{26.104}$$

The addition of the diagonal matrix $\mathbf{R}_\mathrm{t}$ provides the regularization required to avoid receive superdirectivity. Once again, for the computations shown in this chapter, all array elements are assumed identical so that $R_{\mathrm{i},mm}$ is the same for all m and $\mathbf{L}_R = L_R\mathbf{I}$. We can therefore write

$$\mathbf{R}_\eta = \mathbf{R}_\mathrm{i} + \frac{R_{\mathrm{i},11}}{\mathrm{INR}}\mathbf{I} \tag{26.105}$$

where $\mathrm{INR} = R_{\mathrm{i},11}/R_{\mathrm{t},11}$ is the interference-to-noise ratio.

26.5.4.3 Capacity With the antenna ohmic loss now properly included, we can follow the procedure used to obtain Eq. (26.89) to generate our modified transfer relationship

$$\hat{\mathbf{v}}_0 = \mathbf{R}_\eta^{-1/2}\mathbf{H}\widehat{\mathbf{A}}^{-1/2}\hat{\mathbf{i}}_T + \mathbf{R}_\eta^{-1/2}(\hat{\mathbf{v}}_\mathrm{i} + \hat{\mathbf{v}}_\mathrm{t}) \tag{26.106}$$

The mutual information for this model is given by

$$I(\hat{\mathbf{v}}_0; \hat{\mathbf{i}}_T) = \log_2\left|\widehat{\mathbf{H}}\widehat{\mathbf{R}}_T\widehat{\mathbf{H}}^\dagger + \mathbf{I}\right| \tag{26.107}$$

where $\widehat{\mathbf{H}} = \mathbf{R}_\eta^{-1/2}\mathbf{H}\widehat{\mathbf{A}}^{-1/2}$. The capacity can be determined subject to the power constraint $\mathrm{Tr}\left[\widehat{\mathbf{R}}_T\right] \le P_T$. Because the antenna loss results in reduced radiated power, we can construct $\mathbf{R}_T$ from the formulation and then scale it by α so that $P_\mathrm{rad} = \alpha\mathrm{Tr}[\mathbf{R}_T\mathbf{A}] = P_T$. Using this scaled version when evaluating Eq. (26.107) then provides the capacity bound under the delivered power constraint (to suppress transmit superdirectivity) but with the impact of the reduced radiated power removed.

Figure 26.19 shows the water-filling and uninformed transmit capacity as a function of the radiation efficiency of the transmit elements in isolation for an interference-to-noise ratio (INR) of 10 dB and circular array diameter $D = \lambda/2$. Most apparent is the fact that the optimal water-filling capacity is somewhat larger than the corresponding value

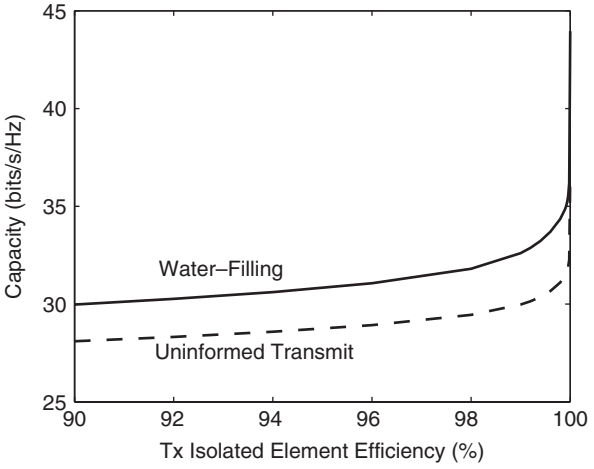


Figure 26.19 Capacity (averaged over 500 channel realizations) versus transmit isolated element efficiency for a 16-element circular array with diameter $D = \lambda/2$ for INR = 10 dB and different capacity solutions.

obtained using the suboptimal uninformed transmitter solution. Otherwise, the curves show similar trends, implying that the same physical phenomena apply to both capacity solutions. These curves reveal that as the antenna efficiency is increased, the capacity increases, with the most dramatic impact occuring around 99% efficiency, where superdirective excitations dominate the solution. It is noteworthy that the 99% threshold efficiency is very difficult to achieve in practice, suggesting that true transmit superdirective behavior would not be observed in a practical system.

Figure 26.20 shows the water-filling capacity as a function of the INR for an isolated transmit element efficiency of 95% and three different circular array diameters (antenna

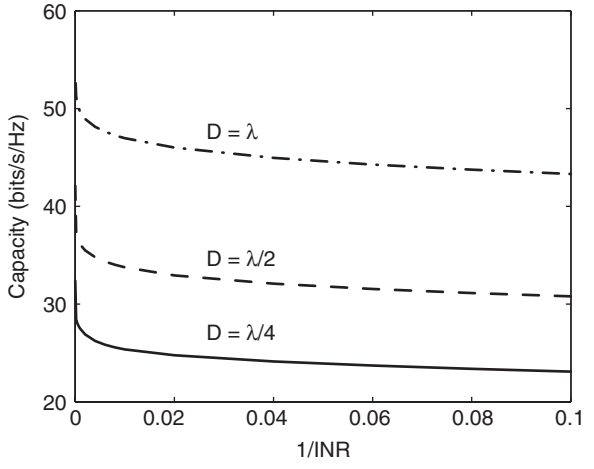


Figure 26.20 Water-filling capacity (averaged over 500 channel realizations) versus INR for a 16-element circular array for $\mu_T = 95\%$ and different array radii.

spacing reduces with diameter). The horizontal axis is actually expressed as 1/INR to emphasize the dramatic change in capacity as the thermal noise goes to zero (antenna becomes lossless). For small arrays, the close element spacing enables superdirectivity, which accounts for the sharp capacity increase for small antenna loss. The performance of the largest array, on the other hand, in less impacted by the reduced loss since the increased element spacing results in reduced superdirective effects.

26.6 MIMO ANTENNA SYNTHESIS

All of the tools we have discussed represent a framework for analyzing the impact of antennas on the performance of MIMO systems. However, this discussion has not touched on the issue of synthesizing antennas that are appropriate for MIMO communications. Naturally, any concept of optimality will be tied to specific characteristics of the propagation environment, although these characteristics can be specified stochastically to ensure that the final design is appropriate over an ensemble of channels. To explore this concept, we consider the covariance matrix computation specified by Eq. (26.26). An optimal set of antennas for maximizing diversity performance should generate a covariance matrix that satisfies two criteria:

1. $R_{s,mm}$ large, indicating that each antenna element receives a large amount of signal power.
2. $R_{s,mp} = 0$ for $m \neq p$, indicating that each antenna element samples the incident signal field in a unique way. In other words, if two antennas sample the field in largely the same way, then there is little diversity offered by the second antenna. With reference to Eq. (26.26), we see that this means that the radiation patterns are orthogonal with respect to the power angular spectrum of the incident field.

These observations indicate that the radiation patterns of the antennas effectively form a basis that should be able to reasonably represent the power angular spectrum of the field. The goal is to design the optimal set of antennas that accomplish these goals.

Simply defining the radiation patterns that optimally accomplish our goals is overly simplistic since the achievable patterns depend on practical issues such as the volumetric aperture in which the antennas reside. We therefore must formulate the problem by incorporating practical constraints to ensure that the antennas provide a reasonable design benchmark against which the performance of practical implementations can be compared. This section offers a framework for accomplishing this optimal antenna synthesis.

26.6.1 Pattern Synthesis

The first step in this formulation is to relate the radiation patterns used in Eq. (26.26) to the physical aperture to which the antennas are restricted. Patterns can be defined by considering either radiating currents (transmit perspective) or the way in which the fields incident on the aperture are sampled and weighted before they are added together (receive perspective), with reciprocity being a mechanism to tie these two perspectives into a single framework. While our discussion on diversity has been focused on receiving incident fields, we define our radiation patterns in terms of radiation currents as this is arguably a more intuitive perspective when considering the antenna synthesis problem.

Therefore consider a volume V containing an electric current distribution. While we could also consider magnetic current distribution, we ignore such currents for the sake of simplicity. We represent the current distribution as a sum of vector functions $\bar{\mathbf{j}}_m(\bar{\mathbf{r}})$, with the radiation pattern for the mth current function being given by

$$\bar{\mathbf{e}}_m(\Omega) = \int_V \bar{\bar{\mathbf{G}}}(\Omega, \bar{\mathbf{r}}) \cdot \bar{\mathbf{j}}_m(\bar{\mathbf{r}})\, d\bar{\mathbf{r}} \tag{26.108}$$

where $\bar{\bar{\mathbf{G}}}(\Omega, \bar{\mathbf{r}})$ is the dyadic Green's function relating the currents to the far-field radiation. The goal is therefore to determine the optimal functions $\bar{\mathbf{j}}_m(\bar{\mathbf{r}})$ that create the desired radiation patterns.

To simplify this synthesis problem, we first represent each current function as a weighted sum of vector basis functions $\bar{\mathbf{f}}_n(\bar{\mathbf{r}})$, or

$$\bar{\mathbf{j}}_m(\bar{\mathbf{r}}) = \sum_n B_{nm} \bar{\mathbf{f}}_n(\bar{\mathbf{r}}) \tag{26.109}$$

where B_{nm} represents the unknown weighting coefficient for the nth basis function and the mth current function. Substitution of this expansion into Eq. (26.108) yields

$$\bar{\mathbf{e}}_m(\Omega) = \sum_n B_{nm} \int_V \bar{\bar{\mathbf{G}}}(\Omega, \bar{\mathbf{r}}) \bar{\mathbf{f}}_n(\bar{\mathbf{r}})\, d\bar{\mathbf{r}} = \sum_n B_{nm} \bar{\mathbf{z}}_n(\Omega) \tag{26.110}$$

where the function $\bar{\mathbf{z}}_n(\Omega)$ physically represents the radiation pattern due to the nth basis function. Finally, substitution of this result into Eq. (26.26) gives

$$R_{\mathrm{s},mp} = \sum_n \sum_q B_{nm} \underbrace{\int \bar{\mathbf{z}}_n(\Omega) \cdot \bar{\bar{\mathbf{P}}}_{\mathrm{s}}(\Omega) \cdot \bar{\mathbf{z}}_q^*(\Omega)\, d\Omega}_{C_{nq}} B_{qp}^* \tag{26.111}$$

or

$$\mathbf{R}_{\mathrm{s}} = \mathbf{B}^T \mathbf{C} \mathbf{B}^* \tag{26.112}$$

Our problem has now been reduced to identifying the coefficients contained in $\mathbf{B}$ which accomplish our goal. However, we must recognize that even if the current basis functions $\bar{\mathbf{f}}_n(\bar{\mathbf{r}})$ are normalized, each resulting radiation pattern $\bar{\mathbf{e}}_m(\Omega)$ can have a unique total radiated power unless the pattern itself is properly normalized. This normalization is accomplished by constraining each resulting radiation pattern to satisfy

$$\frac{1}{2\eta_0} \int \bar{\mathbf{e}}_m^*(\Omega) \cdot \bar{\mathbf{e}}_m(\Omega)\, d\Omega = P_{\mathrm{rad}} \tag{26.113}$$

where P_{rad} is the desired total radiated power for each pattern. If we let the vector $\mathbf{b}_m$ represent the mth column of the matrix $\mathbf{B}$, then using Eq. (26.110) in this constraint leads to $\mathbf{b}_m^\dagger \mathbf{A} \mathbf{b}_m = P_{\mathrm{rad}}$, where

$$\mathbf{A}_{nq} = \frac{1}{2\eta_0} \int \bar{\mathbf{z}}_n^*(\Omega) \cdot \bar{\mathbf{z}}_q(\Omega)\, d\Omega \tag{26.114}$$

We must also consider that it is possible for the solution to use superdirective excitations to optimize the radiation patterns to the environment. To avoid this, we use

the approach detailed in Section 26.5.4, wherein antenna loss is introduced to avoid superdirective excitations. To this end, we introduce loss by assuming that the material in which the transmit currents flow is characterized by a conductivity σ_T. The loss for the nth basis function is then

$$P_{\text{loss},n} = \frac{1}{\sigma_T} \int \left| \bar{\mathbf{f}}_n(\bar{\mathbf{r}}) \right|^2 d\bar{\mathbf{r}} \tag{26.115}$$

where the integration is over the antenna aperture. Since A_{nn} represents the radiated power for the nth basis function, we can choose σ_T such that the radiation efficiency for the nth mode $\mu_{T,n} = A_{nn}/(P_{\text{loss},n} + A_{nn})$ achieves a specified value. For many basis functions, a single value of σ_T may result in different values of radiation efficiency for each basis function, and therefore the value of σ_T can be chosen to set the radiation efficiency for a single basis function, likely the lowest order one, to a specified value. We then use $\widehat{\mathbf{A}} = \mathbf{A} + \mathbf{L}_T$, where $\mathbf{L}_T$ is a diagonal matrix with $L_{T,nn} = P_{\text{loss},n}$.

To satisfy both the supergain and the radiated power constraints, we can parameterize $\mathbf{b}_m$ using

$$\mathbf{b}_m = P_{\text{rad}}^{1/2} \widehat{\mathbf{A}}^{-1/2} \hat{\mathbf{b}}_m \tag{26.116}$$

which leads to $\mathbf{b}_m^\dagger \mathbf{A} \mathbf{b}_m = P_{\text{rad}} \hat{\mathbf{b}}_m^\dagger \hat{\mathbf{b}}_m$ so that $\hat{\mathbf{b}}_m \hat{\mathbf{b}}_m^\dagger = 1$. Since this also implies that

$$\mathbf{B} = P_{\text{rad}}^{1/2} \widehat{\mathbf{A}}^{-1/2} \widehat{\mathbf{B}} \tag{26.117}$$

our covariance matrix in Eq. (26.112) can be written

$$\mathbf{R}_{\text{s}} = \widehat{\mathbf{B}}^T \underbrace{P_{\text{rad}} \widehat{\mathbf{A}}^{-1/2} \mathbf{C} \widehat{\mathbf{A}}^{-1/2}}_{\widehat{\mathbf{C}}} \widehat{\mathbf{B}}^* \tag{26.118}$$

We are now prepared to determine the coefficients that generate the optimal radiation patterns according to our criteria. Because $\mathbf{C}$ and therefore $\widehat{\mathbf{C}}$ are positive semidefinite and Hermitian, the EVD of $\widehat{\mathbf{C}}$ has the structure $\widehat{\mathbf{C}} = \boldsymbol{\xi}_C \boldsymbol{\Lambda}_C \boldsymbol{\xi}_C^\dagger$. It is important to recognize that the number of basis functions may (and generally should) be larger than the number of desired antennas. Therefore if M represents the desired number of antennas, we let $\boldsymbol{\xi}_M$ contain the columns of $\boldsymbol{\xi}_C$ corresponding to the M largest eigenvalues in $\boldsymbol{\Lambda}_C$. If we therefore choose $\widehat{\mathbf{B}} = \boldsymbol{\xi}_M^*$ then the covariance matrix becomes

$$\mathbf{R}_{\text{s}} = \boldsymbol{\xi}_M^\dagger \boldsymbol{\xi}_C \boldsymbol{\Lambda}_C \boldsymbol{\xi}_C^\dagger \boldsymbol{\xi}_M = \boldsymbol{\Lambda}_M \tag{26.119}$$

where $\boldsymbol{\Lambda}_M$ represents the $M \times M$ diagonal matrix containing the M dominant eigenvalues in $\boldsymbol{\Lambda}_C$. We have therefore satisfied our criteria. Furthermore, since the eigenvectors are orthonormal, we also satisfy the constraint that $\hat{\mathbf{b}}_m \hat{\mathbf{b}}_m^\dagger = 1$. The final coefficients are then constructed from Eq. (26.117).

26.6.2 Computational Example

We restrict ourselves to a two-dimensional scenario where the incident field and the antenna aperture lie in the horizontal plane and a single electromagnetic polarization

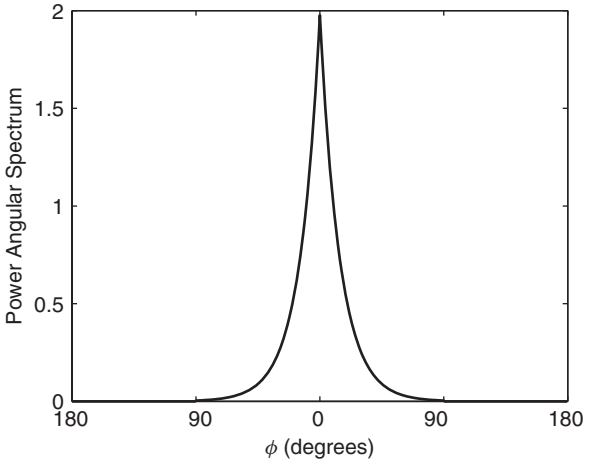


Figure 26.21 Laplacian power angular spectrum used in the generation of the optimal radiation patterns.

is present. The signal power angular spectrum is defined by the truncated Laplacian distribution shown in Figure 26.21. For a rectangular aperture of side lengths L_x and L_y in the x and y dimensions, respectively, the scalar basis functions used for the computation are "Fourier functions" defined by

$$f_n(x, y) = e^{j2\pi(n_x x/L_x + n_y y/L_y)} \tag{26.120}$$

where $-N_x \leq n_x \leq N_x$, $-N_y \leq n_y \leq N_y$, and a unique value of n is assigned to each unique pair (n_x, n_y). For simplicity in the following computations, $L_x = L_y = 1\lambda$ and $N_x = N_y = 10$. Furthermore, we use the basis function corresponding to $n_x = n_y = 0$ in Eq. (26.115) for computing the value of loss to achieve the specified radiation efficiency.

Figure 26.22 shows the four best current distributions and the resulting radiation patterns for this scenario when the basis function radiation efficiency is set to $\mu_T = 99\%$.

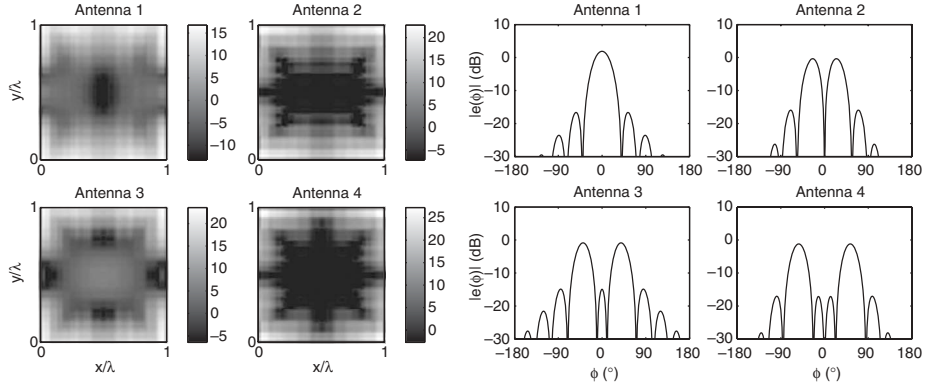


Figure 26.22 Optimal four current distributions and radiation patterns (all in dB) for an environment described by a Laplacian power angular spectrum if $\mu_T = 99\%$.

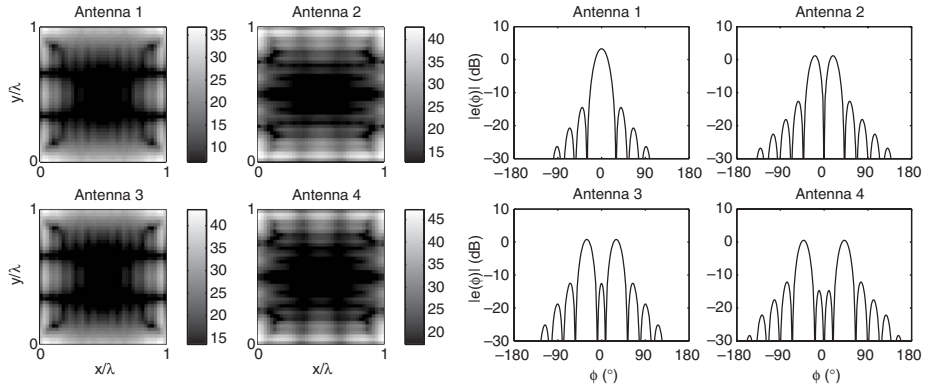


Figure 26.23 Optimal four current distributions and radiation patterns (all in dB) for an environment described by a Laplacian power angular spectrum if $\mu_T = 99.99\%$.

For comparison, a computation is also performed using an array of four filamentary currents placed at the corners of the aperture. In this case, the diversity gain of the optimal antennas is 4.8 dB higher than that of the array of filamentary currents (at the 1% probability level). Figure 26.23 shows the same results when the basis function radiation efficiency is set to $\mu_T = 99.99\%$. In this case, it is interesting that the currents are more concentrated near the edges of the aperture and have much larger magnitudes. Furthermore, the optimal radiation patterns have narrower beams enabled by the array superdirectivity for this low loss scenario. In this case, the diversity gain of the optimal antennas is 6.7 dB higher than that of the array of filamentary currents.

26.7 SUMMARY

This chapter discusses the role of antennas in determining the communication performance of MIMO and diversity systems. After a brief introduction to MIMO system architectures and the signal processing used to enhance communication performance using antenna arrays in multipath propagation environments, the discussion turns to the physical mechanisms through which the antenna radiation characteristics impact the system performance. Issues such as achievable performance with polarization and spatial diversity, antenna mutual coupling, and array superdirectivity are incorporated in the analysis. The chapter concludes with a discussion on the synthesis of optimal antenna patterns for maximizing diversity performance.

REFERENCES

1. M. A. Jensen and J. W. Wallace, A review of antennas and propagation for MIMO wireless communications, *IEEE Trans. Antennas Propag.*, Vol. 52, pp. 2810–2824, November 2004.
2. G. L. Turin, F. D. Clapp, T. L. Johnston, B. F. Stephen, and D. Lavry, A statistical model of urban multipath propagation, *IEEE Trans. Vehicular Technol.*, Vol. 21, pp. 1–9, February 1972.

3. A. A. M. Saleh and R. A. Valenzuela, A statistical model for indoor multipath propagation, *IEEE J. Selected Areas Commun.*, Vol. 5, pp. 128–137, February 1987.
4. J. W. Wallace and M. A. Jensen, Modeling the indoor MIMO wireless channel, *IEEE Trans. Antennas Propag.*, Vol. 50, pp. 591–599, May 2002.
5. Q. H. Spencer, B. D. Jeffs, M. A. Jensen, and A. L. Swindlehurst, Modeling the statistical time and angle of arrival characteristics of an indoor multipath channel, *IEEE J. Selected Areas Commun.*, Vol. 18, pp. 347–360, March 2000.
6. A. F. Molisch, A generic model for MIMO wireless propagation channels, in *Proceedings of the 2002 IEEE International Conference Communication*, New York, 28 April to 2 May 2002, Vol. 1, pp. 277–282.
7. C.-C. Chong, C.-M. Tan, D. I. Laurenson, S. McLaughlin, M. A. Beach, and A. R. Nix, A new statistical wideband spatio-temporal channel model for 5-GHz band WLAN systems, *IEEE J. Selected Areas Commun.*, Vol. 21, pp. 139–150, February 2003.
8. M. A. Jensen and J. W. Wallace, MIMO wireless channel modeling and experimental characterization, in *Space–Time Processing for MIMO Communications*, A. B. Gershman and N. D. Sidiropoulos (Eds.), John Wiley & Sons, Hoboken, NJ, 2005, Chap. 1.
9. T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Hoboken, NJ, 1991.
10. D. Gesbert, M. Shafi, D. Shan Shiu, P. J. Smith, and A. Naguib, From theory to practice: an overview of MIMO space–time coded wireless systems, *IEEE J. Selected Areas Commun.*, Vol. 21, pp. 281–302, April 2003.
11. J. B. Andersen, Array gain and capacity for known random channels with multiple element arrays at both ends, *IEEE J. Selected Areas Commun.*, Vol. 18, pp. 2172–2178, November 2000.
12. G. G. Raleigh and J. M. Cioffi, Spatio-temporal coding for wireless communication, *IEEE Trans. Commun.*, Vol. 46, pp. 357–366, March 1998.
13. T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. Prentice Hall, Englewood Cliffs, NJ, 2000.
14. M. A. Khalighi, J. Brossier, G. Jourdain, and K. Raoof, Water filling capacity of Rayleigh MIMO channels, in *Proceedings of the 2001 IEEE 12th International Symposium on Personal, Indoor and Mobile Radio Communication*, San Diego, CA, 30 September to 3 October 2001, Vol. 1, pp. 155–158.
15. G. J. Foschini and M. J. Gans, On limits of wireless communications in a fading environment when using multiple antennas, *Wireless Personal Commun.*, Vol. 6, pp. 311–335, March 1998.
16. G. D. Golden, C. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, Detection algorithm and initial laboratory results using V-BLAST space–time communication architecture, *Electron. Lett.*, Vol. 35, pp. 14–16, January 1999.
17. P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel, in *URSI ISSSE'98*, Pisa, Italy, 29 September to 2 October 1998, pp. 295–300.
18. L. Zheng and D. Tse, Diversity and multiplexing: a fundamental tradeoff in multiple antenna channels, *IEEE Trans. Inf. Theory*, Vol. 49, pp. 1073–1096, May 2003.
19. W. C. Jakes, *Microwave Mobile Communications*, IEEE Press, New York, 1993.
20. O. Nørklit, P. D. Teal, and R. G. Vaughan, Measurement and evaluation of multi-antenna handsets in indoor mobile communication, *IEEE Trans. Antennas Propag.*, Vol. 49, pp. 429–437, March 2001.
21. T. Svantesson, M. A. Jensen, and J. W. Wallace, Analysis of electromagnetic field polarizations in multi-antenna systems, *IEEE Trans. Wireless Commun.*, Vol. 3, pp. 641–646, March 2004.

22. R. A. Andrews, P. P. Mitra, and R. de Carvalho, Tripling the capacity of wireless communications using electromagnetic polarization, *Nature*, Vol. 409, pp. 316–318, January 2001.
23. J. W. Wallace, M. A. Jensen, A. L. Swindlehurst, and B. D. Jeffs, Experimental characterization of the MIMO wireless channel: data acquisition and analysis, *IEEE Trans. Wireless Commun.*, Vol. 2, pp. 335–343, March 2003.
24. B. N. Getu and J. B. Andersen, The MIMO cube—a compact MIMO antenna, *IEEE Trans. Wireless Commun.*, Vol. 4, pp. 1136–1141, May 2005.
25. T. Svantesson, Correlation and channel capacity of MIMO systems employing multimode antennas, *IEEE Trans. Vehicular Technol.*, Vol. 51, pp. 1304–1312, November 2002.
26. C. Waldschmidt, T. Fugen, and W. Wiesbeck, Spiral and dipole antennas for indoor MIMO-systems, *IEEE Antennas Wireless Propag. Lett.*, Vol. 1, No. 1, pp. 176–178, 2002.
27. C. Martin, J. Winters, and N. Sollenberger, MIMO radio channel measurements: performance comparison of antenna configurations, in *Proceedings of the 2001 IEEE 54th Vehicular Technology Conference*, Atlantic City, NJ, 7–11 October 2001, Vol. 2, pp. 1225–1229.
28. I. J. Gupta and A. A. Ksienski, Effect of mutual coupling on the performance of adaptive arrays, *IEEE Trans. Antennas Propag.*, Vol. 31, No. 5, pp. 785–791, 1983.
29. R. G. Vaughan and J. B. Andersen, Antenna diversity in mobile communications, *IEEE Trans. Vehicular Technol.*, Vol. 36, pp. 147–172, November 1987.
30. T. Svantesson and A. Ranheim, Mutual coupling effects on the capacity of multielement antenna systems, in *Proceedings of the 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Salt Lake City, UT, 7–11 May 2001, Vol. 4, pp. 2485–2488.
31. R. Janaswamy, Effect of element mutual coupling on the capacity of fixed length linear arrays, *IEEE Antennas Wireless Propag. Lett.*, Vol. 1, No. 1, pp. 157–160, 2002.
32. C. Waldschmidt, J. v. Hagen, and W. Wiesbeck, Influence and modeling of mutual coupling in MIMO and diversity systems, in *Proceedings of the 2002 IEEE Antennas and Propagation Society International Symposium*, San Antonio, TX, 16–21 June 2002, Vol. 3, pp. 190–193.
33. B. Clerckx, D. Vanhoenacker-Janvier, C. Oestges, and L. Vandendorpe, Mutual coupling effects on the channel capacity and the space–time processing of MIMO communication systems, in *Proceedings of the 2003 IEEE International Conference on Communication*, Anchorage, AK, 11–15 May 2003, Vol. 4, pp. 2638–2642.
34. D. M. Pozar, *Microwave Engineering*, John Wiley & Sons, Hoboken, NJ, 1998.
35. J. Engberg and T. Larsen, *Noise Theory of Linear and Nonlinear Circuits*, John Wiley & Sons, Hoboken, NJ, 1995.
36. G. Gonzalez, *Microwave Transistor Amplifiers*, Prentice Hall, Englewood Cliffs, NJ, 1997.
37. S. Stein, On cross coupling in multiple-beam antennas, *IRE Trans. Antennas Propag.*, Vol. 10, pp. 548–557, September 1962.
38. J. B. Andersen and B. K. Lau, On closely coupled dipoles in a random field, *Antennas Wireless Propag. Lett.*, Vol. 5, pp. 73–75, 2006.
39. M. A. Jensen and Y. Rahmat-Samii, Performance analysis of antennas for hand-held transceivers using FDTD, *IEEE Trans. Antennas Propag.*, Vol. 42, pp. 1106–1113, August 1994.
40. A. Taflove and S. C. Hagness, *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, 2nd ed., Artech House, Boston, 2000.
41. Hewlett-Packard, A low noise 4 GHz transistor amplifier using the HXTR-6101 silicon bipolar transistor, *Hewlett-Packard Application Note 967*, May 1975.
42. M. Uzsoky and L. Solymar, Theory of super-directive linear antennas, *Acta Phys. Hungarica*, Vol. 6, No. 2, pp. 185–205, 1956.
43. R. F. Harrington, Antenna excitation for maximum gain, *IEEE Trans. Antennas Propag.*, Vol. 13, pp. 896–903, November 1965.

44. Y. T. Lo, S. W. Lee, and Q. H. Lee, Optimization of directivity and signal-to-noise ratio of an arbitrary antenna array, *Proc. IEEE*, Vol. 54, pp. 1033–1045, August 1966.

45. S. M. Sanzgiri and J. K. Butler, Constrained optimization of the performance indices of arbitrary array antennas, *IEEE Trans. Antennas Propag.*, Vol. 19, pp. 493–498, July 1971.

46. L. P. Winkler and M. Schwartz, A fast numerical method for determining the optimum SNR of an array subject to a Q factor constraint, *IEEE Trans. Antennas Propag.*, Vol. 20, pp. 503–505, July 1972.

47. M. L. Morris and M. A. Jensen, Impact of supergain in multi-antenna systems, in *Proceedings of the 2005 IEEE Antennas and Propagation Society International Symposium*, Washington DC, 3–8 July 2005.

48. M. L. Morris, M. A. Jensen, and J. W. Wallace, Superdirectivity in MIMO systems, *IEEE Trans. Antennas Propag.*, Vol. 53, pp. 2850–2857, September 2005.

49. C. A. Balanis, *Antenna Theory: Analysis and Design*, John Wiley & Sons, Hoboken, NJ, 1997.

CHAPTER 27

Antennas for Medical Therapy and Diagnostics

JAMES C. LIN, PAOLO BERNARDI, STEFANO PISA, MARTA CAVAGNARO, and EMANUELE PIUZZI

27.1 INTRODUCTION

Advances in the biomedical use of electronic technology associated with electromagnetic (EM) fields and waves have been enhanced tremendously by a better understanding of biophysical interactions of EM fields in living tissues. A critical system in the wide range of applications in biology and medicine is the antenna system employed to deliver EM energy to and receive EM energy from the target tissue. Thus, antennas in biomedical applications have similar roles to those in telecommunication systems—transferring the EM energy between a confined guiding structure and its surrounding medium. However, there are substantial differences in the way they are deployed, which make them distinct. Antennas for biomedical applications are used in close proximity to, or even inside, the human body or specific organs, so that emitting and receiving characteristics of a given antenna are different from those of the same antenna when operating in free space.

In the case of antennas for therapeutic applications, some well known parameters of typical antennas, such as the radiation pattern which is defined in the antenna's far field have minor practical value, while evaluation of the EM field produced in the immediate surroundings of the antenna often becomes of crucial importance. Another key parameter in the design of antennas for biomedical applications is the input impedance, which must be taken into account in order to optimize power delivery from the antenna to biological tissues. Obviously, both the near field and the input impedance are strongly influenced by the electrical and geometrical characteristics of the biological body or specific organ in which the antennas are used.

Depending on the particular purpose, antennas for biomedical applications may be classified into two major categories: antennas for therapeutic applications and antennas for diagnostic imaging and sensing. In both categories, antennas may be used either outside or inside the human or animal body (or a specific organ); however, it is reasonable

Modern Antenna Handbook. Edited by Constantine A. Balanis

to state that antennas for diagnostic applications are typically employed outside the body or in direct contact with the body surface (noninvasive applications), while antennas for therapeutic applications are used mostly in direct contact with the body surface or even inside the body (implanted antennas).

Organs and tissues function most efficiently at the normal range of body temperature. For humans the norm is maintained at a relatively stable temperature near 37 °C. Temperature elevations above this norm are associated with varying levels of biological responses. Hyperthermia is the term used to describe significant elevation of tissue temperature above the usual limit (40 °C) encompassed by routine thermoregulatory activity. Its use for therapeutic purposes has expanded in recent years to include a variety of abnormal conditions. Investigations to date have shown that while hyperthermia can produce local tissue modifications for effective therapy, temperatures at which the desired tissue response occurs vary over a wide range. Current medical applications fall into two broad categories: hyperthermia cancer therapy and coagulative ablation treatment. An important aspect of these developments is the production of adequate temperature rise and distribution in the target tissue, superficial or deep seated. Moreover, successful hyperthermia and ablative therapies require not only a suitable energy source for heat production but an understanding of the underlying pathological condition being treated to define the critical target tissue temperature and the ability of the therapeutic EM energy to reach the target tissue.

In diagnostic applications, EM fields and waves at higher frequencies can provide a convenient approach to detect and monitor physiological movements without compromising the integrity of the underlying physiological events. In this case, antennas are used to direct EM waves to the target and the reflected wave is processed to yield information on the organ of interest or the physiological event under interrogation. This noninvasive technique enables continuous monitoring as well as quantifying time-dependent changes in cardiovascular and respiratory systems. Electromagnetic energy is also used in some clinically important diagnostic imaging modalities. In particular, magnetic resonance imaging (MRI) relies on radiofrequency (RF) magnetic fields for excitation of nuclear resonance and has been shown to offer a distinct advantage in a multitude of disease processes when compared to other imaging modalities, including ionizing radiations such as X-rays.

There are many different antenna types for biomedical applications. The choice depends on frequency, power density, duration of exposure, the desired specific absorption rate (SAR) and its distribution, the biological target involved, and the domain of application. The various antennas and application scenarios would depend on whether the exposure is whole body or partial body or inside the tissue target. Each of these approaches allows an abundant array of antenna configurations. Beginning with antennas for therapeutic applications, the various types of antenna and application scenarios are presented in this chapter, along with discussions on techniques for design and performance evaluation. Antennas for therapeutic applications are divided into two parts: external antennas designed to be placed in close proximity or in direct contact with the body and implanted and embedded antennas to be inserted into the human body using catheters. Antennas for diagnostic applications follow, with the discussions being divided into categories of noninvasive sensing of cardiovascular and pulmonary signatures and microwave and magnetic resonance imaging.

27.2 ANTENNAS FOR THERAPEUTIC APPLICATIONS

27.2.1 Introduction

As mentioned earlier, current therapeutic applications of electromagnetic fields can be divided into two broad categories: hyperthermia cancer treatment and thermal ablation therapy. This section discusses antennas that have been designed to facilitate these applications.

Hyperthermia cancer therapy is a treatment procedure in which tumor temperatures are elevated to the range of 42–45 °C. The rationale for hyperthermia therapy is related to the ability of elevated temperature to selectively destroy malignant cells [1, 2]. While tumor cells exhibit inherent temperature sensitivity, their response is characterized by the hypoxic, acidic (pH is usually below 7.4, i.e., the value of normal tissue), and nutritionally deprived environment often found in the interior of various tumors. The typically poor blood perfusion found in the interior of a large tumor also facilitates heat accumulation. Moreover, from the EM point of view, the higher levels of water and ionic contents present in most tumor cells give rise to a higher conductivity compared to normal cells [3], which helps to enhance absorption of EM energy. The combination of higher sensitivity to heat and higher EM energy absorption by tumor cells further facilitates the efficacy of hyperthermia therapies for the treatment of cancer.

Hyperthermia can be used either alone or as an adjuvant to other anticancer treatments, since it enhances the cytotoxic effects of some antitumor drugs and potentiates the cell-killing ability of ionizing radiation. The synergism of hyperthermia and ionizing radiation is particularly noteworthy since it is accomplished by thermal killing of hypoxic cells and cells in S phase (DNA synthesis), which are resistant to ionizing radiation.

An important aspect of hyperthermia treatment is the production of adequate temperature distribution in the target tumor tissue, superficial or deep seated. Another important specification for the hyperthermia system is its ability to heat the entire tumor volume without overheating the surrounding normal tissue. Thus successful hyperthermia therapy requires not only a suitable energy source for heat production but also the ability to reach the target tissue with the required heating modality.

As already cited, RF and microwave antennas for hyperthermia can be divided by their intended use into external or internal applicators. External applicators are placed externally to the human body—in close proximity or in direct contact. Internal applicators are deployed by using interstitial, intraluminal, or intracavitary approaches for insertion into the body to the treatment zone, through blood vessels, natural orifices, or by using catheters and needles.

While an ideal hyperthermia system would be noninvasive, invasive systems have several technical advantages over noninvasive ones, particularly with reference to the possibility of better controlling the temperature increments and distribution within the tissue (both malignant and healthy).

The second broad category of therapeutic applications considered here is ablation therapies for treating cardiac arrhythmia and endometrium disorders. It should be noted that ablative thermal therapies are stand-alone therapies and use temperatures up to 50–90 °C, applied for short durations (a few minutes). The treatment of cardiac arrhythmia represents one of the early applications of ablation therapy; however, recently, ablation

approaches have been considered also for treating endometrium disorders and as a stand-alone cancer therapy. This is in contrast to traditional hyperthermia, where tissue temperatures are elevated to 42–45 °C, and is usually used as an adjuvant therapy [2, 4, 5]. Ablation therapies are based on the use of interstitial antennas, since the direct application of ablative heating at high temperatures has the advantage of minimizing damage to the surrounding healthy tissue.

Antennas for therapeutic applications are presented in two parts: in Section 27.2.2 antennas designed to be placed externally in close proximity or in direct contact with the human body are discussed; while in Section 27.2.3 antennas to be inserted into the human body through blood vessels and natural orifices or by using catheters and needles are presented. Often, the two categories of external and embedded antennas correspond also with the two types of applications considered here; that is, external antennas are usually designed for hyperthermia therapies, while embedded antennas are used for ablation therapies, but not always.

The design and performance of antennas for hyperthermia or ablation therapy involve analytic and/or numerical solution of the EM and thermal problems, that is, solution of Maxwell's equations and the bioheat equation (BHE), by considering the EM field source together with the target biological body. The field distribution within the body is often reported using the metric of specific absorption rate (SAR), which is defined as the power or time rate of change of energy absorbed in the body per unit mass, and is given by the following formula:

$$SAR(x, y, z) = \frac{\sigma(x, y, z)|E(x, y, z)|^2_{\text{rms}}}{\rho(x, y, z)} \quad \text{(W/kg)} \tag{27.1}$$

where σ represents the conductivity (S/m), and ρ the density of the tissue (kg/m^3). A more detailed discussion of SAR is provided in Chapter 28. Numerical methods to solve Maxwell's equations and BHEs are presented in Section 27.2.4.

27.2.2 External Antennas

External microwave antennas used in hyperthermia treatment of cancer have been designed as a single radiating antenna or in array configurations. Typically, such antennas are positioned near the body, and a bolus full of deionized water is used to fill the gap between the radiating antenna and the body. The water bolus has a dual purpose: to improve the matching between the antenna and the body in order to optimize the EM power transfer, and to provide cooling to help maintain the temperature of the most superficial body layers at normal levels.

In the following, external antennas designed as hyperthermia sources are described and illustrated, and their main properties are reviewed. Included also are the theoretical approaches used in their design and evaluations of SAR distributions in the biological body. In particular, with reference to the design procedure, simple analytical approaches are discussed since they provide fast and easy ways to assess the principal antenna performances and help understand the fundamental EM principles underlying their performances. Numerical methods that can be used to design antennas for hyperthermia systems and to evaluate the SAR distribution and corresponding temperature increments are given at the end of this section. For interested readers, references are provided to more detailed discussions of the subject.

To properly evaluate the antenna's performance and to compare different antennas at a given design frequency, antenna matching and performance in bodies with different sizes and compositions are important considerations. The better the match, the higher the power transmitted from the antenna to the target biological tissue. In addition, power absorption, penetration depth of the radiation, and the corresponding temperature increase should be examined with particular regard to the dimensions of the area where the temperature elevation is to reach its goal of 42–45 °C.

27.2.2.1 Antenna Type, Configuration, and Performance The most common types and configurations of antennas placed externally to the human body for hyperthermia are inductive applicators, patch antennas, wave guide antennas, and phased array antennas.

Inductive Applicators At low and intermediate frequencies, where tissue resistive losses are more predominant than dielectric ones [6], inductive applicators are often used. In particular, current loops with various configurations have been studied and applied in clinical settings [7]. For example, an inductive applicator operating at 150 MHz was developed from four parallel electric-current-carrying conductors, which used a capacitor to tune the applicator and a reflector to provide a current return path. The antenna is relatively small in dimension (about 10 cm) but is able to produce a penetration depth of 3–4 cm [8].

Patch Antennas Microstrip applicators or patch antennas have been designed for operation at microwave frequencies; they utilize low loss materials with high relative permittivity as substrates (i.e., ε_r between 10 and 30). The design procedure is based on the classical microstrip formula [9]. Antennas designed to operate at 915 and 433 MHz have shown penetration depths between 2 and 3 cm, while designs for operation at 120 or 190 MHz have yielded penetration depths higher than 4 cm. In general, these microstrip applicators have been tested to give good impedance matching characteristics.

For the microstrip applicator shown in Figure 27.1, the excitation line was immersed in water and radiation into the tissue medium was achieved through an aperture in the ground plane. For control of superficial tissue temperature, water was circulated within the radiator during treatment. The microstrip feed line was terminated with a rectangular patch (Figure 27.1b). The antenna is found to resonate at two frequencies, 434 and 915 MHz, with a penetration depth of about 2 cm at the lower frequency [10]. The design was facilitated by the use of the finite-difference time-domain (FDTD) numerical algorithm. Subsequent designs of the same applicator have employed differently shaped radiating aperture (annular and horseshoe) and multiple applicators with two or four antennas, located side by side. The multiple-applicator configuration, with each antenna fed in phase and with the same amplitude, was able to heat a greater volume of tissue. A comparison between the "ring" and "horseshoe" radiating slots showed similar heating volume but with slightly different shaped heating regions [11].

Waveguide Antennas Waveguide antennas are often used in superficial hyperthermia treatments and are designed from waveguides of rectangular, coaxial, or circular cross section [12–14]. In the case of rectangular cross section, the waveguide is often terminated with a horn antenna filled with water. Figure 27.2 shows two box-horn type applicators: one made by tapering only the walls parallel to the electric field vector

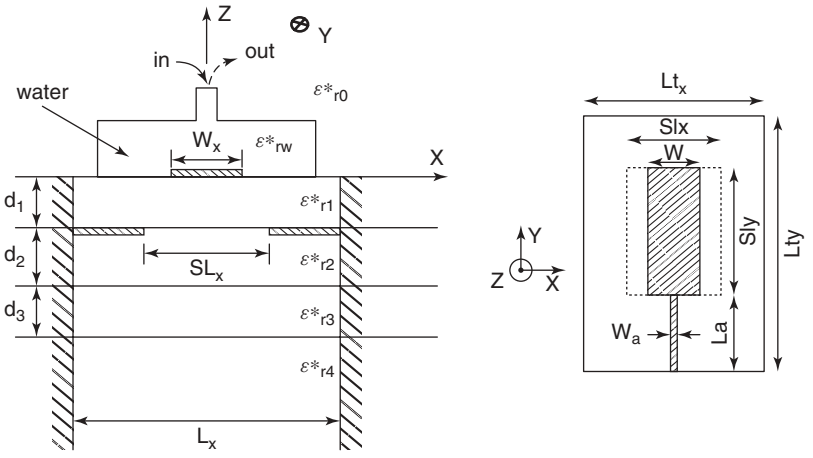


Figure 27.1 Microstrip–microslot applicator: (a) cross section and (b) top view. $W_x = 20$ mm, Slx = 35 mm, Sly = 38 mm, Ltx = 70 mm, Lty = 100 mm, $d_1 = 7.6$ mm, $d_2 = 1.58$ mm, and $d_3 = 35.1$ mm. (From Ref. 10.)

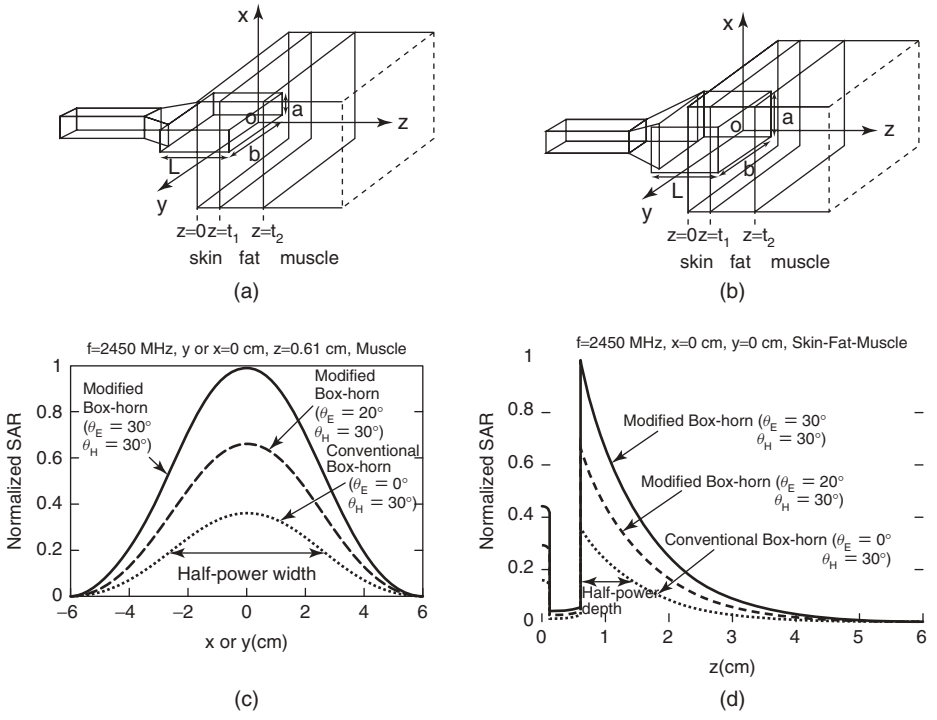


Figure 27.2 Water-loaded box-horn applicator: (a) conventional applicator and (b) modified applicator. Comparison of SAR distribution between the modified and conventional box-horn applicators: (c) transversal plane and (d) penetration depth. (Adapted from Ref. 15.)

("conventional" applicator—Figure 27.2a) and the other by tapering all the guide walls ("modified" applicator—Figure 27.2b). A comparison between the two showed that the modified design gives higher SAR values than the conventional one (Figure 27.2c,d). As expected, the penetration depth is higher for designs operating at lower frequencies (433 versus 2450 MHz); however, the resolution is lower also [15].

A still different modified horn applicator design (Lucite cone applicator—LCA) was made by replacing the diverging metal walls of the horn parallel to the electric field with Lucite walls, and by inserting a PVC cone at the center of the applicator aperture [16]. The Lucite walls impose a nonzero E-field condition at the boundary of the aperture, while the PVC cone is used as a field adapter. FDTD analysis and experimental measurements of the LCA applicator, operating at 433 MHz, showed that in comparison with a standard horn applicator the LCA could provide a greater penetration depth and a higher "effective field size," defined as the area enclosed within the 50% iso–SAR curve at 1-cm depth inside a flat homogeneous phantom [16, 17].

Phased Arrays The use of phased arrays in hyperthermia treatments allows the focusing of heating patterns deep inside the human body [18]. The focal point can be shifted by simply changing the amplitudes and phases feeding the different antennas in the array. A phased array of 16 radiating apertures, located along an octagonal cylindrical structure operating in the frequency range between 50 and 110 MHz was designed, tested, and put into clinical trials [19, 20]. The patient body was placed inside a cylindrical structure with a water bolus used to fill the space between the antennas and the human body. Several other phased array applicators used in hyperthermia treatments have been implemented using a similar structure.

A conformal array applicator, which adapts its shape to the body and is made using dual-mode microstrip antenna elements operating at 915 MHz, has been developed to heat superficial tumors that cover large surface areas [21]. The antenna has a lightweight structure and consists of a square annular slot dual concentric conductor (DCC) aperture (Figure 27.3a). It has been demonstrated that it can provide uniform heating to a large area of superficial tissue [22]. The array configuration can be made by etching the radiating apertures from a single layer of flexible copper foil. Each antenna in the array can be noncoherently driven at 915 or 433 MHz [24].

Although the ability to independently change the amplitude and phase of the array elements represents a considerable advantage of phased arrays, particularly for focusing the absorbed power in different body parts, difficulties in monitoring power deposition and, more importantly, temperature elevation, have hindered the effectiveness and clinical acceptance of phased arrays [25, 26]. Currently, a great deal of research is underway on techniques for noninvasive temperature estimation, and on techniques for model-based optimization of the pattern of temperature increases.

For example, the DCC antenna has been combined with an Archimedean spiral element for use as a radiometer [23]. The spiral element was inserted into the solid copper patch that forms the inner conductor of the DCC aperture, etching away its central portion (Figure 27.3b). The high gain Archimedean spiral has a Gaussian-shaped radiation pattern, which allows the antenna to receive blackbody radiation from deep-seated tissues. Through the use of an intervening bolus layer, centrally located along the boresight direction, sampling of tissue temperature under the DCC aperture is guaranteed. The resulting antenna, and the corresponding array configuration, was able to monitor tissue temperatures during treatment.

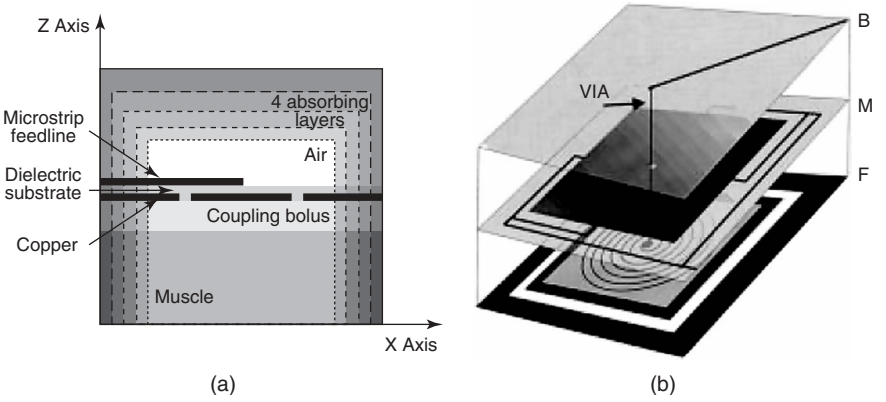


Figure 27.3 (a) Microstrip dual-slot aperture antenna (from Ref. 22) and (b) dual-mode antenna. (From Ref. 23.)

Recently, a clinical applicator designed for hyperthermia treatment in conjunction with noninvasive MRI temperature monitoring has been presented [27]. The applicator was made by using hermetically closed cassettes containing specially shaped bent dipole structures arranged in two transversal subarrays with six water-coated-antenna (WACOA) modules each. A modular water-bolus system was used to cool the superficial tissue layers.

27.2.2.2 Simple Formulas for Antennas Design: As a general statement, to evaluate the performances of an antenna placed close to a biological (i.e., scattering) body, Maxwell's equations should be solved by taking into account the actual configurations. However, both the antenna and the biological body have a complex structure from the geometrical and electrical point of view. This complex structure renders the solution of the EM problem very difficult, which makes it very difficult to understand the phenomena underlying EM power deposition into a biological body. However, as a first approximation, the EM problem may be solved by using both simplified EM field source and biological body to obtain simple formulas that can help to understand the fundamental parameters governing the process under study.

Thus a simple model to study the EM energy absorption by a body placed in direct contact with an antenna applicator for hyperthermia systems, is obtained by considering a plane wave normally incident on a one-dimensional layered tissue with infinite extent in the transverse plane (Figure 27.4a).

The power transmitted from the incident wave to the different layers can be evaluated by using the equivalence between the geometry and wave propagation in transmission lines. As a consequence, the layered structure of Figure 27.4a can be studied as successive sections of transmission lines with different properties corresponding to different body layers (Figure 27.4b). The wave propagating in the ith layer is constituted by a direct and a reflected wave according to the following formula:

$$\left.\begin{aligned} V_i &= V_i^+ e^{-jk_i z} + V_i^- e^{jk_i z} \\ I_i &= \frac{V_i^+}{Z_i} e^{-jk_i z} - \frac{V_i^-}{Z_i} e^{jk_i z} \end{aligned}\right\} \tag{27.2}$$

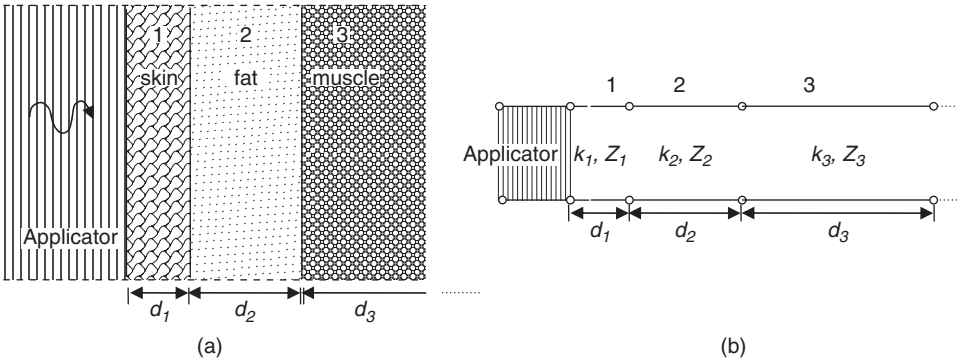


Figure 27.4 (a) Plane wave incident on a layered tissue and (b) equivalent transmission line.

where the propagation constant k_i and the characteristic impedance Z_i can be obtained from the corresponding dielectric properties as

$$k_i = \sqrt{\omega^2 \mu_i \varepsilon_{ci}} \tag{27.3}$$

$$Z_i = \sqrt{\mu_i / \varepsilon_{ci}} \tag{27.4}$$

where $\omega = 2\pi f$ is the angular frequency, μ_i is the magnetic permeability, and ε_{ci} is the complex relative dielectric constant, which also takes into account the electrical conductivity. V_i^+ and V_i^- are constants obtained by applying the boundary conditions between layer i and the next layer, using the equation

$$V = ZI \tag{27.5}$$

where V and I are the voltage and current at the boundary between layer i and the following one, and Z corresponds to the load impedance at the same position. In particular, according to the transmission line equivalence, Z represents the input impedance of layer $i + 1$ and is given by

$$Z = Z_{\text{in}}^{(i+1)} = Z_{i+1} \frac{Z_{\text{in}}^{(i+2)} - jZ_{i+1} \tan(k_{i+1} d_{i+1})}{Z_{i+1} - jZ_{\text{in}}^{(i+2)} \tan(k_{i+1} d_{i+1})} \tag{27.6}$$

with d_{i+1} as the thickness of the $(i+1)$th layer.

The preceding equations completely define the problem, which can be solved to obtain the ratio of the voltage transmitted into layer i as a function of the incident voltage in layer 1, representing the applicator. Once the ratio is obtained, it can be used to study the influence of applicator material properties on power absorption and the effects of thicknesses and dielectric properties of the different layers constituting the body. Consequently, the power transfer from the applicator to the target tissue can be optimized.

To extend this simple model for studying the power transfer to a finite-size therapeutic applicator, the incident wave may be decomposed into a spectrum of plane waves and a set of different angles of incidence. Studies on monopole and dipole antennas showed that the simple analysis, based on plane wave incidence, can give useful preliminary

results, so that an optimization process based on the simple model can be developed to provide fairly accurate results [3, 9].

A slightly more complicated analysis can be performed by considering tumors with simple shapes inserted into healthy tissue, with the incident EM wave represented as a plane wave [3]. The condition under which the maximum absorbed power ($P = \sigma |E|^2_{\text{rms}}$) is found within the malignant tissue has been studied via this simple analytical approach for four possible situations, namely:

1. Wave propagating parallel or orthogonal to the interface between tissues, with the electric field vector parallel to it.
2. Wave propagating parallel to the tissue interface, with the electric field vector orthogonal to it.
3. An embedded malignant tissue sphere (tissue 2) with a small radius compared to the EM field wavelength in the surrounding normal tissue (tissue 1).
4. An embedded malignant tissue cylinder with a small radius compared to the EM field wavelength in the surrounding normal tissue (tissue 1). Electric field vector and propagating direction are perpendicular to the axis of the cylinder.

In the first situation, the boundary condition on the electric field (parallel to the interface between two different tissues) leads to

$$E_2 = E_1 \Rightarrow \frac{P_2}{P_1} = \frac{\sigma_2}{\sigma_1} \tag{27.7}$$

Consequently, if σ_2 is higher than σ_1, the absorbed power is greater in tissue 2.

When the electric field vector is normal to the interface between tissues (second situation considered), the boundary condition is imposed on the displacement and conduction currents, as

$$(\sigma_2 + j\omega\varepsilon_2)E_2 = (\sigma_1 + j\omega\varepsilon_1)E_1 \tag{27.8}$$

Correspondingly, the ratio between the powers absorbed in the two tissues is

$$\frac{P_2}{P_1} = \frac{\sigma_2}{\sigma_1}\left|\frac{\sigma_1 + j\omega\varepsilon_1}{\sigma_2 + j\omega\varepsilon_2}\right|^2 \Rightarrow \frac{P_2}{P_1} = \frac{\sigma_2}{\sigma_1}\left(\frac{\varepsilon_1}{\varepsilon_2}\right)^2\left[\frac{1 + (\sigma_1/\omega\varepsilon_1)^2}{1 + (\sigma_1/\omega\varepsilon_1)^2}\right] \tag{27.9}$$

The behavior of the ratio between the two powers as a function of the frequency can easily be investigated starting from the dependence on frequency of the relative permittivity and conductivity.

If the malignant tissue is a sphere surrounded by normal tissue, it can be shown [3, 28, 29] that the ratio between the electric fields is given by

$$\frac{E_2}{E_1} = \frac{3}{2 + \dfrac{\sigma_2 + j\omega\varepsilon_2}{\sigma_1 + j\omega\varepsilon_1}} \tag{27.10}$$

while that for a cylinder is given by

$$\frac{E_2}{E_1} = \frac{2}{1 + \dfrac{\sigma_2 + j\omega\varepsilon_2}{\sigma_1 + j\omega\varepsilon_1}} \tag{27.11}$$

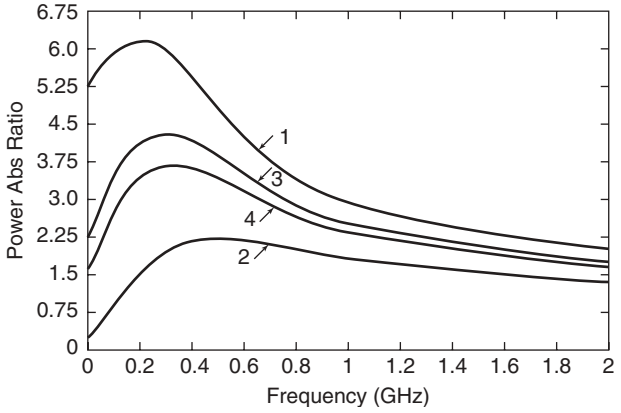


Figure 27.5 Ratio of power absorbed in SMT-2A mammary gland tumor over normal mammary tissue for four possible geometries. (From Ref. 3.)

As an example, the ratio of power absorption in tissue 2 (tumor—SMT-2A mammary gland) over that in tissue 1 (normal tissue) is shown in Figure 27.5 for the above mentioned four conditions, as a function of frequency [3].

The approach of a plane wave incident on a layered tissue has been used recently to study the performances of a modified box-horn applicator [15]. The incident and transmitted waves in a three-layered body model were represented through their plane wave spectra and determined by applying boundary conditions between the different media.

When a single external applicator is used for hyperthermia treatments, the choice of the operating frequency and applicator structure must be made according to the depth and size of the target tumor. In general, the higher the frequency used, the shorter the EM field penetration depth. Thus lower frequencies could produce greater penetration depth, but the applicator size would become correspondingly larger in order to provide a reasonably localized heating and avoid damaging healthy tissues surrounding the tumor. Alternatively, greater penetration depths can be obtained by using an array of applicators and by feeding the antennas with appropriate phase and amplitude.

Simple formulas to design phase arrays for near-field focusing can be found in published literature [30, 31].

When an array of EM sources is used, the incident radiation can be analyzed through geometrical optics. The EM field at a generic observation point can be written as a weighted sum of contributions coming from all the antennas in the array such that

$$\mathbf{E}_{\text{tot}}(A) = \sum_m W_m \mathbf{E}_m(A) \tag{27.12}$$

where $W_m = |W_m| e^{j\alpha_m}$ represents the weighting factor of the mth element, optimized to focus the EM field to a desired point location. Gee et al. [30] considered an array in contact with a water bolus on the top of a tissue surface. Transmission coefficients for the vertical and parallel polarization were used for evaluating the field transmitted into the tissue. A good agreement was obtained between the theory predicted and experimentally

measured data. Specifically, focusing of the field at 11.4 cm from the array plane, when water and muscle tissue were considered, was obtained with a 19-element hexagonal array operating at 2450 MHz.

The study of antennas and arrays can also be accomplished by using numerical and experimental techniques. For numerical studies, the FDTD and finite-element method (FEM) algorithms are well suited for solving the EM problem [32]. Moreover, a finite-difference solution of the bioheat equation can be implemented to obtain the corresponding temperature increase [33]. Numerical techniques can also be used for developing patient-specific hyperthermia models, starting from tomographic images of the treatment area. In these body models, the EM and thermal analyses can be used to optimize power absorption in the specific area of interest [34]. Some discussion of the numerical methods is given later.

27.2.3 Implanted and Embedded Antennas

The technical difficulties confronted by external applicators in heating deep-seated tumors without overheating adjacent normal tissue may be partially overcome by interstitial array techniques as a treatment modality. Interstitial techniques have the capacity to adapt the SAR distribution to an irregularly shaped tumor volume and to provide uniform temperature in deep-seated tumors. Also in combination with brachytherapy, interstitial hyperthermia renders a treatment modality for malignancies with little additional risk to the patient [35, 36].

In addition to hyperthermia and ablative treatment of liver cancer, prostate cancer, and brain tumors, interstitial antennas are employed in ablation therapies to treat cardiac arrhythmia and endometrium disorders in the form of transcatheter and intracavitary applications.

Ablation therapies based on implanted microwave antennas represent a recent development of therapies based on RF sources. Comparisons between microwave ablation techniques and RF sources have found that microwaves can produce larger induced thermal lesion (necrotic tissue), due to their differences in interaction mechanisms with biological tissues. In fact, when RF energy is used, a current is induced to flow between the active electrode and the ground one; this current leads to a resistive heating, rapidly decreasing in amplitude with the distance from the active electrode, according to a fourth-power law. On the other hand, when microwaves are used, the propagating EM field dielectrically couples with the tissue constitutive molecules. As a consequence, power deposition inside the tissue, and correspondingly the temperature increase, decays with distance from the microwave source by following a second-power law, compared to the fourth-power dependence of RF ablation. Thus microwave allows deeper lesions, that is, regions in which the temperature exceeds the minimum value necessary for ablating or killing the abnormal cells [37–39]. However, this deeper lesion can be achieved only if the antennas are properly designed, especially with regard to their matching with the external (tissue) environment. An improper antenna matching, in fact, can lead to currents flowing along the antenna feed line with consequent heating of healthy tissues [38].

The performances of antennas designed for ablation therapies may be evaluated by considering matching characteristics of the antenna at the design frequency and its behavior when the surrounding dielectric properties change (as during the treatment or from human to human), and by considering the shape and dimension of the lesion produced (i.e., the region where the temperature reaches at least 50 °C).

27.2.3.1 Antenna Type, Configuration, and Performance Several types of antennas have been proposed and applied for microwave thermal treatments. These antennas can be classified according to their structure as monopoles, dipoles, coaxial sleeve dipoles, cap-choke antennas, slot antennas, and helical or spiral antennas. Frequently, these antennas are fabricated from coaxial cables (or triaxial cables) by terminating them with the specific antenna design. Finally, dielectric waveguide antennas and array configurations have also been proposed for interstitial therapies.

Monopole Antennas Simple monopole antennas in an array arrangement were the first to be used in the 1980s for hyperthermia treatment of cancer. They are commonly made by removing the outer conductor of a miniature, flexible or semirigid coaxial cable and then completely stripping the central conductor to a given length. These simple antennas were poorly matched and had the tendency to produce a cold spot or low heating zone near the distal tip of the antenna [40–42]. Moreover, the requirement of uniform temperature distribution throughout the entire tumor volume necessitated insertion of the tip of the antenna well beyond the tumor boundary [43–45], which created an unnecessary situation for potential damage to normal tissues. One of the first interstitial antennas with a design suitable to overcome these difficulties is the sleeve-slot antenna (see later discussion) [43].

Several small catheters are implanted inside the tumor with a distance between the catheters of 10–18 mm in brachytherapy. These catheters can be used to insert interstitial antennas to combine brachytherapy with hyperthermia treatments. For example, monopole antennas were obtained from a UT-34 semirigid coaxial cable (0.85-mm outer diameter) [46]. Different antenna lengths were chosen in order to match the antenna for use both in the thermal treatment at the lower microwave frequency and as a radiometer for temperature measurements at higher microwave frequencies. In the array configuration, these antennas with different phase shifts were able to provide larger heated volumes within the array. However, the simple monopole antennas have a reflection coefficient whose value depends on the antenna insertion depth into the tissue, so that antenna performances are degraded under operating conditions [47]. This finding is related to the improper matching of the simple monopole antenna, which causes reflected current to flow along the antenna feeding line. As mentioned earlier, these problems are avoided by using a sleeve or choke design. Several of the antennas incorporating such designs have been studied for both hyperthermia and ablation therapies and are discussed in later sections.

Some investigators have compared reflection coefficient and heating pattern of different monopoles—open tip, dielectric tip, and metal tip [48]. The study confirmed that a metal tip allows the antenna current to reach the maximum at the antenna tip, and this, in turn, shapes the absorbed power in the pear-shaped pattern around the antenna tip. However, an attempt to optimize both the lengths of the monopole and the metal tip to obtain the best SAR distribution and the lowest reflection coefficient failed. In particular, it was found that short monopoles were well matched but had a highly nonuniform SAR pattern, while longer monopoles had a poorer reflection coefficient. A similar optimization was tried, with limited success, to obtain minimum reflection coefficient by changing the active length of a monopole antenna with a metallic cap to be used in microwave cardiac ablation [49].

The influence of the thickness of the insulation layer around the monopole on antenna matching and on power deposition has also been considered. It was found that thicker insulation layers can broaden the power deposition pattern and increase the antenna

resonant frequency [50]. Note that multisection monopole antennas have also been investigated for interstitial hyperthermia treatments [50].

Dipoles Dipole antennas for interstitial hyperthermia therapies have been made by connecting the inner conductor of a coaxial cable with the outer one with a finite length. The other arm of the dipole consists of the feeding coaxial line, whose length corresponds to the insertion depth of the line in the biological body (Figure 27.6). As a consequence, the performance of these antennas depends greatly on the length of the two arms of the dipole and on the antenna's operating frequency.

A comparison among three designs for dipole antennas for use in tumor hyperthermia has been conducted by considering resonant versus nonresonant dipoles, the influence of the thickness of the antenna catheter wall (1.2- versus 1.8-mm inner diameter, with 2.2-mm outer diameter), and asymmetry [52]. Results showed that in these antenna designs the maximum SAR is always obtained at the antenna junction, irrespective of the antenna insertion depth. However, for radial SAR distribution that is independent of the antenna length, dipoles with resonant length yield higher SAR values than those of nonresonant length, due to antenna input impedance mismatches. As for the catheter thickness, no significant difference was obtained between thin-walled and thick-walled catheters [53].

Coaxial Sleeve Dipoles Like monopoles, dipoles have a reflection coefficient value that depends on the antenna insertion depth. Similarly, SAR deposition extends over the entire antenna active length. In order to eliminate the reflection coefficient dependence on insertion depth and concentrate the SAR distribution, sleeve (or choke or balun, the same concept by different names) could be used [53, 54]. For example, in one design, the two arms of the dipole are connected to the inner and outer conductor of the feeding coaxial cables. A conductor sleeve, one-quarter wave in length, is attached to the outer conductor of the input coaxial cable to form a sleeve balun for matching [54]. Thus the effective impedance between the balun and the outside of the transmission line at the dipole junction is made sufficiently high to permit efficient power transmission and eliminate standing waves on the coaxial line.

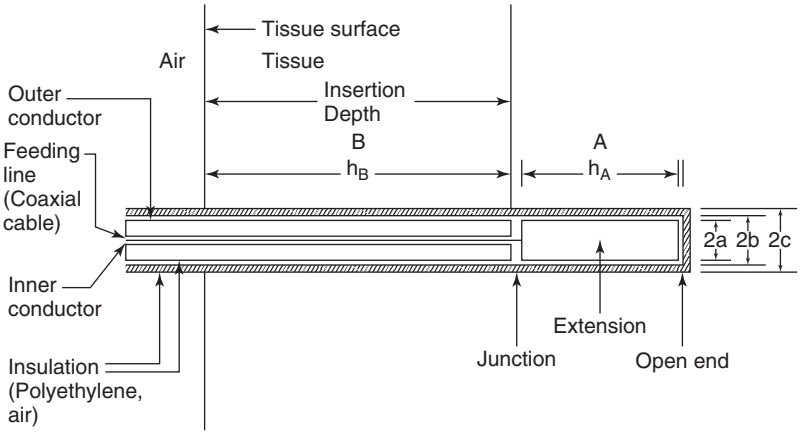


Figure 27.6 Dipole antenna. (From Ref. 51.)

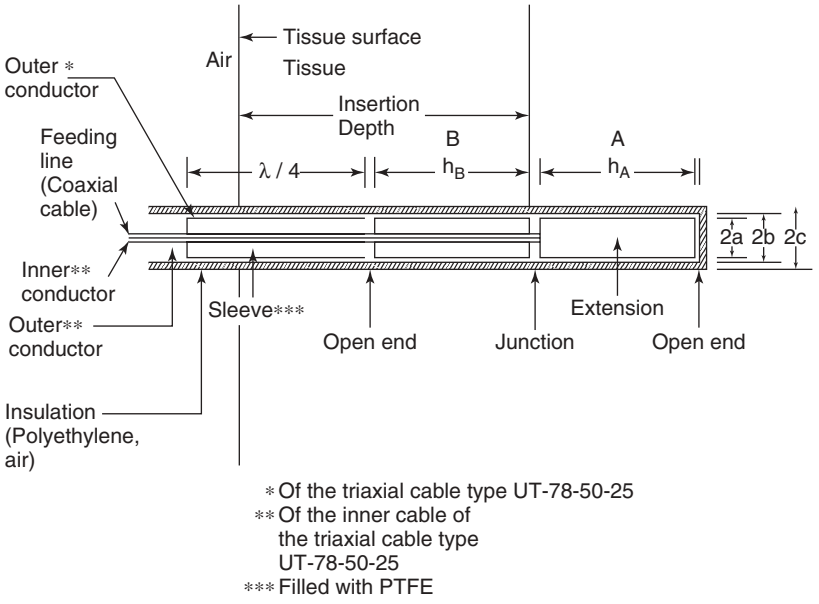


Figure 27.7 Dipole antenna with a λ/4 sleeve. (From Ref. 51.)

As another example, a λ/4 sleeve ($\pi/2$ transformer) was added to the lower arm of a dipole (Figure 27.7); this arm sees an open-end terminating impedance, and thus the input impedance of the applicator becomes independent of the insertion depth of the antenna [51]. In this manner the applicator is matched irrespective of its insertion depth, and its radiation is localized.

In a different design, a dipole (or split-dipole) antenna with a coaxial choke has been obtained by connecting the two arms of a dipole (3-mm length) to the inner and outer conductor of an RG/178BU flexible coaxial cable and connecting the choke to the outer conductor [55]. The antenna has a hemi-ellipsoidal form and the two arms of the dipole can also be used as electrodes for bipolar endocardiac electrogram recording during cardiac ablation procedures (see Figure 27.8).

Finally, a different structure for a coaxial sleeve dipole was proposed by starting from a triaxial cable and short circuiting the inner and outer coaxial cable at the distal end of the antenna. The other arm of the dipole is obtained by short circuiting the outer coaxial cable with a length different from that of the first arm of the dipole (nonsymmetrical antenna)

Figure 27.8 A dipole antenna with a coaxial choke. The two hemi-ellipsoidal caps make up the two arms of the dipole. (From Ref. 55.)

and connecting a coaxial choke to the outer conductor. By appropriately choosing the lengths of the two arms of the dipole and of the choke, it is possible to match the antenna at a desired frequency and localize the radiated field to the tip of the antenna [56, 57].

It should be noted that the design of an ablation or hyperthermia applicator with a coaxial choke or sleeve allows the matching of the antenna input impedance but, on the other hand, by using existing coaxial cables, this approach enlarges the antenna radial dimensions [44]. To minimize such enlargements, a choked coaxial antenna of the asymmetric dipole type has been proposed [58]. The choke section for the antenna was formed by the lateral wall of the metallic biopsy needle used to introduce the antenna into the tissue to be ablated; in such a way a minimum overall transversal size was obtained. Moreover, this technique permits some limited amount of real-time adjustments of the choke length to compensate for minor impedance mismatches during the operation. The applicator was designed to fit into a 14-gauge (14-G) metal biopsy needle (inner diameter, 1.78 mm; outer diameter, 2.05 mm).

A similar design but without metallic contact between the outer conductor of the coaxial cable and the metallic biopsy needle has also been studied [59]. This antenna, called a triaxial antenna, has dimensions that are small enough to allow insertion into a 17-G or 18-G needle.

Cap-Choke Antennas The cap-choke antenna design comprises an annular cap connected to the enlarged inner conductor of a coaxial cable, and a choke to the coaxial outer conductor (Figure 27.9) [60]. The annular cap, as well as the enlargement of the inner conductor, increases the capacitance at the antenna end, letting the radiated power be more concentrated near the antenna tip. The coaxial choke provides antenna match and prevents the reflected current from flowing up the transmission line along the external surface of the coaxial outer conductor. This, in turn, makes the radiating properties of the antenna independent of the insertion depth of the catheter, and the corresponding SAR more concentrated on the active length of the antenna. By appropriately choosing the coaxial cable, the antenna can be designed and constructed to fit the physical constraints of the particular application.

The effect of enlarging the inner conductor has been studied and compared with simple monopole antennas [62]. Specifically, the enlarged antenna tip (Figure 27.10a—reactive load) has been compared with a simple monopole (Figure 27.10b), a monopole with a metallic tip (Figure 27.10c), and a "modified reactive load" (Figure 27.10d). The

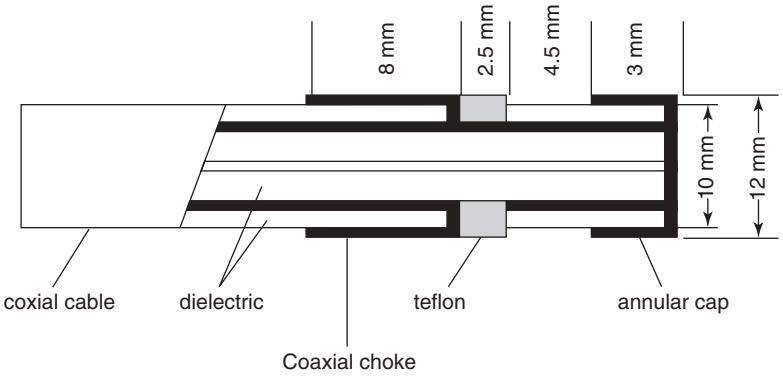


Figure 27.9 Cap-choke antenna. (From Ref. 61.)

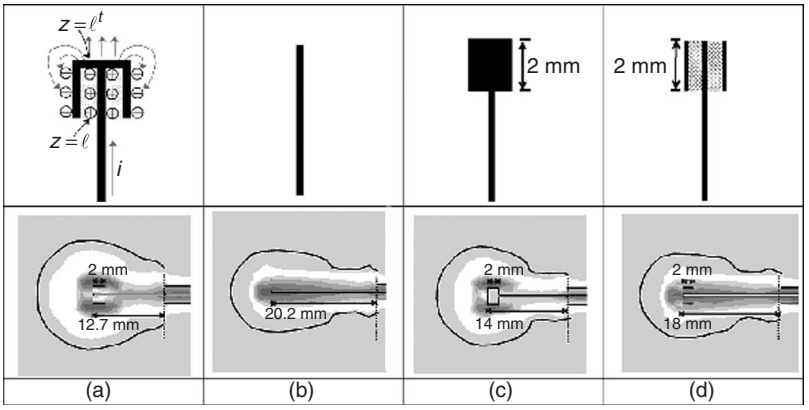


Figure 27.10 Comparison of the SAR distribution obtained from different monopole terminations: (a) enlarged cap (reactive load), (b) open end, (c) metallic cap, and (d) "modified reactive load." (From Ref. 62.)

antennas dimensions were optimized with reference to the reflection coefficient using an analytical theory [63]. The performances of the antennas have been studied in terms of SAR distribution in the surrounding tissue. The reactively loaded antenna was found to have the largest region of power absorption.

Slot Antennas Slot antennas have been proposed since the late 1980s for hyperthermia treatment of cancer [43]. For these applications, the antenna had a rigid structure with a square tip. An improved version of these antennas for use in cardiac ablation was formed by mounting the antenna at the end of an RF/178 BU coaxial cable (Figure 27.11). A flexible choke is connected to the outer conductor, the protruding inner conductor is enlarged in order to form a curved cap, and it has two slots. The first slot is positioned between the cap and the terminating plane of the coaxial outer conductor, and the other inside the coaxial choke. This antenna showed good matching performances (power reflection coefficient less than 4% at 2450 MHz) and a power deposition pattern concentrated near the antenna tip [64].

A simpler coaxial slot antenna design for use in hyperthermia treatment was realized by starting from a thin semirigid coaxial cable, short-circuiting its distal end, and cutting some ring slots on the outer conductor [65]. Simulations showed that if only one slot is made on the outer conductor, the antenna performance is dependent on the antenna

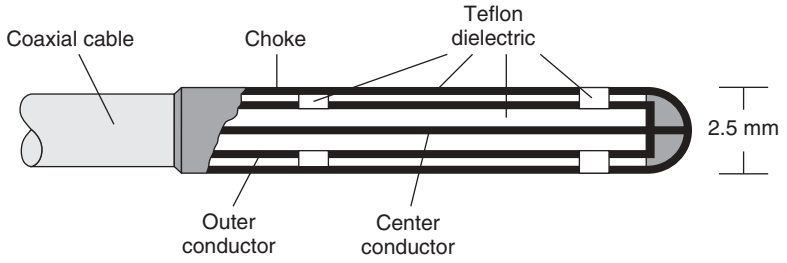


Figure 27.11 Design of a flexible catheter slot antenna. (From Ref. 55.)

insertion depth and a considerable amount of power is dissipated in the most superficial tissue layers near the applicator insertion point. This could be avoided if two separate slots are included. Such findings can easily be explained in light of the preceding discussion, by considering the single-slot antenna equivalent to a simple dipole antenna, and the two-slot antenna as a dipole with a matching choke.

The coaxial slot antennas have been used in clinical trials both as a single applicator and in an array configuration formed by four antennas placed at the corners of a rectangle.

Loop, Spiral, and Helical Antennas A loop-shaped microwave antenna was tested for hyperthermia treatment of liver tumors [66]. The antenna was inserted first into the liver through a 13-gauge needle and the loop was then deployed. To put the antenna into operation, during deployment, 60–70 W of power was applied to assist the loop in "cutting" through hepatic tissue in order to place the antenna with minimal shape distortion. Three different configurations were tested: single loop, two parallel loops, and two orthogonal loops. Mean lesion volume and mean maximum lesion diameter were found to be largest with the parallel loop configuration; in contrast, the orthogonal configuration showed shortest mean time to reach 60 °C and the highest mean maximum temperature (97.2 °C) recorded in 7 min. Theoretical advantages of the loop antenna include the ability to deliver large amounts of precisely targeted microwave energy to the tumor with minimal collateral damage to normal structures. The most obvious disadvantage is the more complicated process needed in targeting tumors [66].

A spiral antenna for use in ventricular tachyarrhythmia ablation therapies was proposed with the aim of producing a radiating aperture as large as possible [67, 68]. This antenna exhibited good matching at the design frequency (−24.1 dB at 915 MHz) and a circular power deposition pattern.

A theoretical study of the heating pattern produced by helical antennas immersed in a dissipative medium showed a limited depth of heating, far less than that of a simple dipole of comparable dimensions [69]. Consequently, helical antennas were deemed not suitable for applications such as hyperthermia, where considerable depth of heating must be obtained. However, they could be used in catheter ablation or angioplasty applications, where the required depth of heating may be shorter and the widest possible lesion is desired. Note that a 12-mm helical antenna terminating a coaxial cable, tested for hyperthermia treatments, showed poor antenna matching and consequently heating along the shaft of the catheter [70].

Dielectric Waveguide Antennas A dielectric loaded waveguide was designed and constructed as an interstitial applicator for use in endometrial ablation treatments [71]. The applicator was made from a metal pipe circular waveguide filled with a high permittivity ceramic in order to reduce the transverse dimensions of the applicator. The ceramic was extended beyond the metal pipe to form a dielectric tip. The final applicator has a diameter of 8 mm for operation at 9.2 GHz. A thermocouple on the applicator tip is used to measure the endometrium temperature in order to control the ablation process. The applicator has been studied both numerically (by using the HFSS^{TM} commercial code) and experimentally (*in vitro* by using egg white and excise uteri) before *in vivo* trials.

Antenna Arrays In many hyperthermia treatments, interstitial applicators are applied in an array configuration. In array configurations, the number of elements, the spacing among them, the amount of input power to each element, and, finally, the phase difference of the excitation must be determined for optimal heating.

Studies on array configurations showed that a four-dipole-antenna rectangular array, with a spacing of about 1–2 cm between the antennas, could generate a fairly uniform SAR distribution in the plane perpendicular to the antenna axis at the level of the antenna feed [72]. Similar conclusions were reached in investigations using equilateral triangular and hexagonal arrays of sleeved slot antennas [73]. For both triangular and hexagonal arrays, it was shown that an additional antenna at the center of these array configurations does not improve the SAR distribution. Similarly, the SAR distribution is not improved using an independent stub on each antenna to match the antenna itself. This effect is related to the different E-field values obtained from the different antennas and to the phase relationship among them [42, 74].

A computer simulation of a three-element array of sleeved slot antennas has been conducted for treatment of brain tumors [33]. The analysis showed that equilateral triangular arrays with 10-, 15-, and 20-mm spacing and a uniform excitation could produce SAR and temperature distributions appropriate to treat tumors of 10–40 mm in diameter. By using input powers ranging from 2 to 32 W, it is possible to reach therapeutic temperatures in about 3 minutes. Moreover, it has been found that taking into account a 50% reduction in blood perfusion rate, typical in the core of large tumors, the region where temperature is above 43 °C can be enlarged by about 40%, for a given input power.

27.2.3.2 Simple Antenna Design Formulas Embedded, implanted, or interstitial antennas for therapeutic applications are typically designed and studied by considering their unique situation of being completely immersed in biological materials. In the past, antenna design has been conducted both numerically and experimentally, by directly measuring antenna performances in homogeneous phantoms filled with tissue-equivalent material or with egg white. In both types of study, the antenna design follows successive steps in a trial-and-error procedure. Recently, the analytical technique developed by King and colleagues [63, 75–77] was applied successfully to study the return loss of interstitial antennas for ablation therapies [78].

King's theory provides a procedure for the evaluation of the input impedance and the EM field radiated by an insulated linear antenna immersed in a dense, homogeneous medium with infinite extension [75]. In particular, the theory demonstrated the equivalence, under some conditions, between the embedded antenna and a coaxial line, in which the antenna conductor is represented by the coaxial inner conductor, the insulation layer(s) is the dielectric filling the coaxial line, and the external medium is represented by the outer conductor of the coaxial line. The analogy between the embedded antenna and the coaxial line is illustrated in Figure 27.12. The characteristic impedance and propagation constant of the coaxial line are evaluated as

$$k_L^2 = k_2^2 \left(\frac{\ln(d/b)}{\ln(c/b) + n_{23}^2 \ln(d/c)} \right) \left(\frac{\ln(d/b) + F}{\ln(d/b) + n_{24}^2 F} \right) \tag{27.13}$$

$$Z_C = \frac{\omega \mu_0 k_L}{2\pi k_2^2} (\ln(c/b) + n_{23}^2 \ln(d/c) + n_{24}^2 F) \tag{27.14}$$

where b, c, and d are radii of the different layers constituting the antenna, as illustrated in Figure 27.12, $n_{23} = k_2/k_3$, $n_{24} = k_2/k_4$, $k_i = \omega\sqrt{\mu_0 (\varepsilon_0 \varepsilon_i - j\sigma_i/\omega)}$ is the wavenumber for region i $(i = 2, 3)$, $k_4 = \omega\sqrt{\mu_0 (\varepsilon_0 \varepsilon_4(\omega) - j\sigma_4(\omega)/\omega)}$ is the wavenumber for region 4 (the external medium), j is the imaginary unit $(=\sqrt{-1})$,

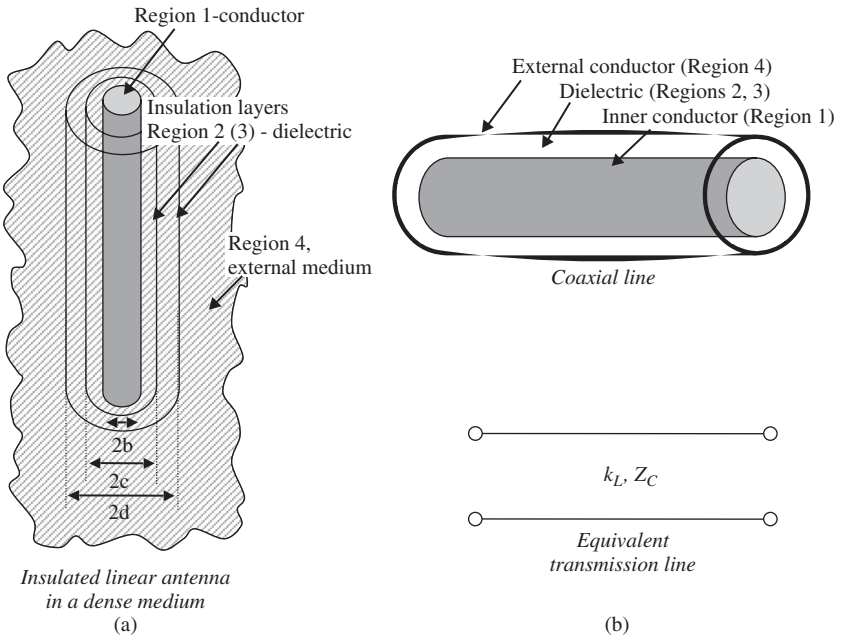


Figure 27.12 King's equivalence between a linear insulated antenna immersed in (a) a relatively dense medium and (b) a coaxial line.

and $F = H_0^{(2)}(k_4 d)/(k_4 d)H_1^{(2)}(k_4 d)$, where $H_i^{(2)}(x)$ are the Hankel functions of the second type.

It should be noted that King and colleagues used a different notation for the definition of the complex constant j and, as a consequence, in their works they define the parameter F by using the Hankel functions of the first type. The constraints for using King's theory are that the insulated layers must be electrically thin; that is, $|k_2 c| \ll 1$ and $|k_3 d| \ll 1$.

Once the antenna immersed in a medium has been substituted with a coaxial line, and its characteristic impedance and propagation constant are defined, the line is completely characterized. The evaluation of its input impedance can be conducted by using the standard transmission line formula. Considering a dipole antenna as an example (see Figure 27.13a), the equivalence approach leads to a two-arm transmission line connected in series, each arm terminated in an open circuit (Figure 27.13b).

The transmission line input impedance is twice that of the single line, which, in turn, is evaluated as $Z_{\text{in}} = -jZ_C/[\tan(k_L h)]$, where h is the arm length.

This approach has been applied to an interstitial antenna for ablation therapy [78]. The input impedance of the antenna was characterized as a function of the dielectric properties of both antenna and surrounding medium, and various antenna elements such as conducting cap and choke. Comparisons between analytical results and measurements show very accurate theoretical predictions, despite the simple formulation.

In addition to antenna matching when immersed in a dissipative and dispersive medium, it is necessary to evaluate the antenna's performance in terms of the pattern of EM power deposition and the corresponding temperature distribution. Such a study

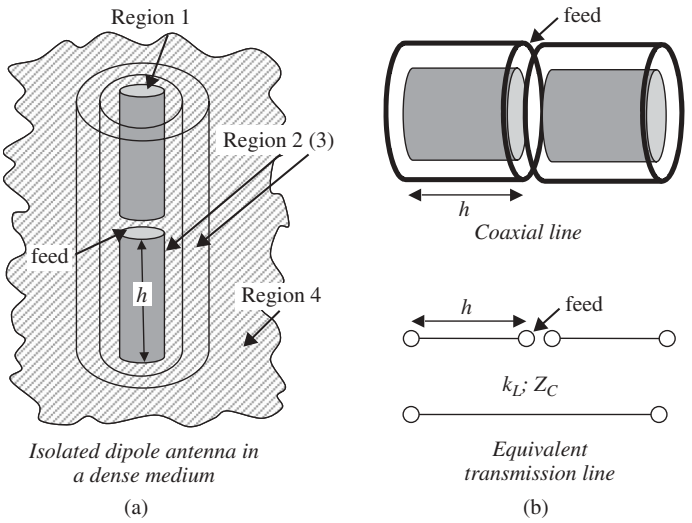


Figure 27.13 Equivalence between (a) an isolated dipole antenna immersed in a dense medium and (b) a two-section coaxial line.

may be conducted experimentally with *in vitro* or *in vivo* techniques, or numerically by solving the EM and thermal problems. Numerical studies have the added advantage of being able to study different operating conditions, both with antenna positioning in the biological sample and influence of biological parameters (e.g., blood perfusion).

However, when evaluating antenna performances with respect to distributions of temperature increase, it is necessary to distinguish among the different applications in thermal therapies. For example, in ablative hyperthermia treatment, it is necessary to obtain a lesion (i.e., a region where the temperature exceeds at least 50 °C) in order to achieve cell death due to necrosis [52]. Typically, a spherical shaped lesion extending over the tumor area and beyond (i.e., a few centimeters around the antenna active length) is obtained with the antenna inserted into the tumor. On the other hand, in cardiac ablation it is necessary to concentrate temperature increase (final temperature up to about 65 °C) in a small area beyond the antenna tip with some depth into the cardiac muscle (about 5 mm). Finally, in endometrial ablation therapies, the temperature to be reached to achieve a successful treatment is above 75 °C over a very wide area, 6 mm in depth (sometimes all the endometrium tissue must be ablated) [71].

Other considerations on antenna performances, particularly in regard to hyperthermia treatments, are related to the different sites in which a tumor is to be treated. In fact, tumor sites vary from the skin surface to deep within the brain or pelvis. The difference in location, and consequently the variations in antenna insertion depth into the biological tissue, can influence antenna performance if the antenna is not properly matched. A final consideration, related to the different applications of interstitial thermal therapy, is the constraint on the physical dimensions of the antenna. For an antenna to be used in the cervix, its maximum diameter would have to be about 1 cm, while an antenna to be inserted through the femoral vein up to the heart (cardiac catheter antenna) would have a diameter smaller than a few millimeters.

27.2.4 Numerical Methods for Antenna Design

Numerical studies on performances of antennas for hyperthermia or ablation therapies have been conducted using the FEM [79, 80] and FDTD [61] methods. Both simple homogeneous models and more complex ones have been considered. Moreover, numerical solutions of the bioheat equation have been used to obtain temperature increase distributions [81]. Results have been reported for antenna reflection coefficients, SAR values, and temperature distributions within various tissues.

The following presents a brief description of the FDTD method. The FDTD method has the advantage over the FEM of being very easy to implement and it is in the time domain. It thus permits studies of time evolution of the power absorption inside the tissue. Moreover, explicit solution of the BHE can be coupled to the FDTD solution to obtain the time evolution of temperatures inside the tissue.

The FDTD method [82] solves Maxwell's equations in the time domain by using the finite-difference approach. This approach corresponds to the substitution of the partial derivatives in the equations with the corresponding finite-difference formula, according to

$$\frac{\partial f(x, y, z, t)}{\partial x} = \frac{f\left(x + \frac{\Delta x}{2}, y, z, t\right) - f\left(x - \frac{\Delta x}{2}, y, z, t\right)}{\Delta x} \tag{27.15}$$

where f represents one of the components of the electric or magnetic field, and similar formulas are used when other spatial variables or time are considered. In order to apply Eq. (27.15) to the electromagnetic field, the domain under study must be subdivided into elementary cells (whose sides are $\Delta x, \Delta y, \Delta z$), and, in each cell, the EM field must be considered uniform, and similarly for the time (Δt). Once the time and the space have been discretized, applying Eq. (27.15) to all six scalar equations representing Maxwell's equations in the time domain, an explicit scheme is found in which the time evolution of the electromagnetic quantities can be obtained. In order to converge to a stable solution and have accurate results, the temporal and spatial increments must satisfy the following equations, respectively:

$$\Delta t \leq \frac{1}{c} \frac{1}{\sqrt{\frac{1}{\Delta x^2} + \frac{1}{\Delta y^2} + \frac{1}{\Delta z^2}}} \tag{27.16}$$

$$\max(\Delta x, \Delta y, \Delta z) \leq \lambda/10 \tag{27.17}$$

where c represents the wave velocity in vacuum (about 3×10^8 m/s), and λ equals the minimum wavelength into the body.

Finally, absorption boundary conditions (ABCs) must be imposed on the FDTD solution to bound the domain under study. The most widely used ABCs are those proposed by Berenger [83], known as PML-ABC. Several improvements have been developed starting with the basic FDTD scheme in books and papers (e.g., see Ref. 84). In ablation studies, bidimensional FDTD codes have been developed using cylindrical coordinates in order to exploit the symmetry typically found in ablation antenna designs and in *in vitro* experimental studies [81].

It should be noted that an important component of numerical simulation is the dielectric properties of the biological tissues used as inputs for the modeling. The dielectric properties of normal tissues have been investigated extensively [85] and are available on-line

(http://www.fcc.gov/fcc-bin/dielec.sh). However, information on the dielectric properties of tumorous tissues are more scattered in the literature. It is reported that tumors have an excess of sodium ions, which causes malignant tissue to retain more fluid than normal tissue in the form of bound water. The higher water content and sodium in tumors lead to a higher conductivity and dielectric permittivity than the healthy surrounding tissue [3].

Once the EM field distribution within the biological target is known, the corresponding temperature increment can be determined by solving the BHE [86]. The BHE relates the time rate of heat accumulated (or lost) per unit volume at a point inside the body to the time rate of temperature increase (or decrease). Specifically,

$$\nabla \cdot [K(\mathbf{r})\nabla T] + A_0(\mathbf{r}, T) + Q_v(\mathbf{r}) - R_L(\mathbf{r}) - B_0(\mathbf{r}, T)(T - T_b) = C(\mathbf{r})\rho(\mathbf{r})\frac{\partial T}{\partial t} \quad (\mathrm{W/m^3}) \tag{27.18}$$

where $K(\mathbf{r})$ is the tissue thermal conductivity (W/(m $\cdot$ °C)), A_0 and Q_v are the volumetric heat sources (W/m^3) due to metabolic processes and EM power deposition, respectively, R_L represents the respiratory heat losses in the lungs (W/m^3) in the case of whole body exposure, $B_0(\mathbf{r},T)$ is a parameter proportional to the blood perfusion (W/(m^3 $\cdot$ °C)), T_b is the blood temperature, and, finally, C (J/(kg $\cdot$ °C)) and ρ (kg/m^3) are the tissue-specific heat and density, respectively.

At the interface between the biological tissue and the air, a boundary condition must be applied that imposes the continuity of the perpendicular heat flow (convective boundary condition). This convective boundary condition can be expressed as

$$-K(\mathbf{r})(\nabla T \cdot \mathbf{n}_0)_S = H(T_s - T_a) + SW(T) \quad (\mathrm{W/m^2}) \tag{27.19}$$

where S is the external surface (skin when the whole human body model is studied), $\mathbf{n}_0$ is the outward unit vector normal to S, H is an equivalent convection coefficient taking into account convective heat exchange and radiative heat exchange (W/m^2 $\cdot$ °C), T_s and T_a represent the temperature of the body surface and of the surrounding air, respectively, and SW represents heat loss due to sweating. Equation (27.18) can also be solved for simple phantom models by neglecting the metabolic processes, sweating, and so on.

As for Maxwell's equations, the BHE can be solved numerically by using a finite-difference approach, leading to an explicit finite-difference scheme linked with the FDTD method [32, 87]. For ablation studies, the BHE has been solved with a semi-implicit approach [33] and with an explicit two-dimensional scheme by taking advantage of the cylindrical symmetry [81].

Thermal properties of normal and malignant biological tissues are not as well established as the dielectric properties. Nevertheless, the values of specific heat C and thermal conductivity K are available for most human body tissues [88–90]. The missing values have been estimated based on the percentage water content (w) of the tissue by using the following equations [90]:

$$C = 1670 + 25.1w \quad \mathrm{J/(kg \cdot ^\circ C)}$$

$$K = 0.0502 + 0.00577w \quad \mathrm{J/(s \cdot m \cdot ^\circ C)}$$

The values for blood perfusion (B_0) may be found in Williams and Leggett [91] and Van Den Berg et al. [92]. Since there is a paucity of data on metabolic heat generation, the values for A_0 often are assumed to be proportional to blood perfusion [93].

TABLE 27.1 Thermal Parameters of Human Body Tissues Under Basal Conditions

Tissue/Organ	C [J/(kg · °C)]	K [J/(s · m · °C)]	A_0 [J/(s · m^3)]	B_0 [J/(s · °C · m^3)]
Bile	3900	0.55	0	0
Bladder	3300	0.43	1600	9000
Blood	3900	0.0	0	0
Bone marrow	2700	0.22	5700	32000
Cancellous bone	1300	0.40	590	3300
Cartilage	3500	0.47	1600	9000
Cerebellum	3700	0.57	7100	40000
Cerebrospinal fluid	4200	0.62	0	0
Colon	3700	0.56	9500	53000
Cornea	4200	0.58	0	0
Cortical bone	1300	0.40	610	3400
Eye humour	4000	0.60	0	0
Eye lens	3000	0.40	0	0
Eye tissue (sclera)	4200	0.58	0	0
Fat	2500	0.25	300	1700
Gallbladder	3500	0.47	1600	9000
Glands	3600	0.53	64000	360000
Grey matter	3700	0.57	7100	40000
Heart	3700	0.54	9600	54000
Internal air	1000	0.03	0	0
Kidney	3900	0.54	48000	270000
Liver	3600	0.51	12000	68000
Lung	3600	0.14	1700	9500
Mucous membrane	3300	0.43	1600	9000
Muscle	3600	0.50	480	2700
Nerve—spinal cord	3500	0.46	7100	40000
Pancreas	3500	0.54	7300	41000
Skin	3500	0.42	1620	9100
Small intestine	3700	0.56	13000	71000
Spleen	3700	0.54	15000	82000
Stomach	3600	0.53	5200	29000
Tendon	3300	0.41	1600	9000
Testis	3800	0.53	64000	360000
White matter	3600	0.50	7100	40000

Source: From Ref. 94.

Table 27.1 presents values of the thermal parameters for most of the body tissues [94].

27.2.5 Summary

Antennas for therapeutic applications of EM fields have been presented. Therapeutic applications fall into two broad categories: hyperthermia cancer treatment and thermal ablation therapy. Both therapies aim at provoking an opportune temperature increment in the biological target tissue (up to 42 °C in hyperthermia, at least 50 °C in ablation therapies), by way of EM energy absorption.

The discussions of antennas for therapeutic applications have been divided into two major groups: external antennas for use near or in direct contact with the human body,

and embedded antennas used for insertion into the body to target tissues to be ablated through blood vessels or natural orifices or by using catheters and needles. The salient features and performances of these antennas are described; simple analytical approaches have been presented to allow simple and quick analysis for preliminary designs of the antennas. In conclusion, numerical methods for simulations and completion of the antenna design are summarized.

Clearly, a great deal has been accomplished in the design of antennas for therapeutic applications of EM fields. However, some challenges still remain that would help the therapeutic applications of EM fields to be more fully developed. For example, with regard to both external and embedded antennas temperature increments obtained in tissues during treatment. Some of the latest concepts for developments in external systems include antennas for use in conjunction with MRI scanning systems, and for embedded systems, the use of thermosensors embedded in the patient body close to the active antenna to measure temperature distributions.

27.3 ANTENNAS FOR SENSING AND DIAGNOSTIC IMAGING

27.3.1 Introduction

Increasingly, RF and microwave radiation is being explored and applied for wireless, noninvasive, transcutaneous sensing and interrogation of deep-lying body organs. RF and microwave energy possess some unique features that have allowed them in some cases to become as useful as many established modalities for diagnostic sensing and imaging. EM radiation provides a simple approach to detect and monitor physiological movements without compromising the integrity of such physiological events as heart rate, respiration, ventricular movement, and pressure pulse. A beam of microwave energy (such as from a horn antenna) may be directed to the target organ or body and the reflected signal processed to yield information about the target organ under interrogation [95, 96]. Likewise, the use of narrow nanosecond pulses to noninvasively probe the motion of the internal organs of the human body has been proposed [97, 98]. In principle, chest, heart, lung, vocal cord, blood vessel, bowel, bladder, fetus, and other body movements may be monitored using nanosecond pulse systems or narrowband microwave systems.

As mentioned, reflection takes place when an incident microwave encounters a biological target. Moreover, the reflected microwaves experience a Doppler effect, which shifts its frequency either up or down from the frequency of the impinging microwave, depending on the direction of the movement with respect to the microwave source [99]. Thus microwaves bounced back from moving organs provide an approach to noninvasively, and even remotely, sense organ movements inside the body. The advantage afforded by noninvasive and distant sensing suggests, in addition to vital signs such as blood flow, heart beat, and respiration, the potential use of this technology for monitoring frail and elderly patients or patients with premature development, monitoring of unrestrained persons in holding areas or retention facilities, and detection of unauthorized personnel or intruders [100].

The ability to detect distantly such vital signs as heart beat and respiration rate is particularly attractive in situations where direct contact with the subject is either impossible or impractical, for example, persons who fell prey to such hazardous scenarios as explosion, fire, chemical or nuclear contamination, and natural or terrorism-created disasters. Indeed, heart beat and respiration have been detected at distances of a few to

tens of meters, with or without intervening material barriers, using off-the-shelf antennas designed for wireless communication uses [99, 101, 102].

For example, the small displacement of the precordium overlying the apex of the heart is related to the low frequency movement in the left ventricle and echoes the hemodynamic events within the left ventricle. Microwave apexcardiograms using 2.45 GHz showed close correlation to the hemodynamic events occurring within the left ventricle [103]. They involve detecting the reflected Doppler signals using an antenna located a few centimeters over the apex of the heart. This approach has several advantages over more conventional techniques because it does not require any physical contact with the subject. Problems such as skin irritation, restriction of breathing, and electrode connections are eliminated.

As another example, Doppler microwaves have been employed to interrogate the wall properties and pressure pulse characteristics at a variety of arterial sites, including the carotid, brachial, radial, and femoral arteries, using a waveguide aperture antenna [104–106]. Microwave-sensed carotid pulse waveforms have been obtained in patients using contact application of 25-GHz energy along with simultaneously recorded intra-aortic pressure waves [105]. The resemblance of the microwave-sensed arterial pulse and the invasively recorded pressure wave was remarkable and the results confirm that a noninvasive Doppler microwave sensor can successfully and reproducibly detect pressure pulse waveforms of diagnostic quality. Other interesting applications are the use of microwaves for sensing cerebral edemas [107, 108] and for speech articulator measurement [109].

The majority of current imaging systems used to diagnose diseases and tissue anomalies are based on sources that emit ionizing EM energy. Nuclear magnetic resonance imaging or MRI relies on nonionizing static and RF magnetic fields and has been shown to offer a distinct advantage in a multitude of disease processes when compared to ionizing modalities. The success of MRI has prompted investigations into several emerging near-field modalities in the form of microwave and microwave-induced thermoelastic tomographic imaging [110, 111]. The wide range of dielectric property variations offers a potential for higher contrast and better tissue characterization. Microwave tomography has been explored to reconstruct images associated with dielectric property variations in body cross section [112–117]. Microwave thermoelastic imaging uses a microwave pulse-induced thermoelastic pressure wave to form planar or tomographic images [111, 118–121]. Since the generation of thermoelastic (or thermoacoustic) pressure waves depends on permittivity, specific heat, and acoustic properties of tissue, microwave thermoelastic imaging possesses some unique features that are being explored as an imaging modality for noninvasive characterization of tissues. Note that nanosecond pulse-based systems have also been proposed for imaging breast cancer [110].

The following sections describe antennas used for wireless sensing and diagnostic imaging, including birdcage and TEM coils specifically designed for use in MRI.

27.3.2 Antennas for Wireless Noninvasive Sensing

Many types of antennas are available for sensing applications. They can be classified into two general categories: close range or distant approaches. Close range approaches may be further divided into contact or noncontact, although both applications are in the near field of the antenna. In this case, the sensing environment is akin to partial-body exposure for therapeutic purposes; thus the same types of antennas are applicable (see Section

27.2.2.1 for patch and waveguide antennas) at lower power levels. Similarly, since the sensing applications involve the detection of transmitted or reflected microwaves, which is common to both sensing and imaging, the resistively loaded dipoles and bow-tie antennas described in Section 27.3.3.2 for diagnostic microwave imaging could serve equally well in near-field sensing applications.

The present discussion focuses on waveguide aperture and horn antennas. It should be noted that, for the same reason stated previously, waveguide aperture and horn antennas have also been used for imaging applications (see Section 27.3.3.3).

27.3.2.1 Waveguide Aperture Antennas Common EM waveguides have either rectangular or cylindrical configurations. They are used as the transmission line for propagation of EM waves. The types of fields that can be propagated in these waveguides are transverse electric (TE) and transverse magnetic (TM) modes. The essential wave propagation properties of rectangular and cylindrical waveguides are the same. In the case of common rectangular waveguides, the dominant TE_{10} mode field is associated with a characteristic cut-off wavelength $\lambda_c = 2b$, below which the wave can propagate, and b is the broad width of the waveguide (also see waveguide antennas in Section 27.2.2.1).

If the waveguide is terminated, the fields at the open end are approximately the same as they would be if the waveguide does not terminate there, but continues to infinity. The open-ended waveguide forms an aperture through which it radiates an EM field, with a characteristic sine-wave pattern dependent on the waveguide dimension a, especially in the immediate vicinity of the waveguide aperture. The waveguide aperture can also serve as a receiving antenna; the total power it extracts from the reflected wave is proportional to the area of the open waveguide aperture. Since the effective aperture for receiving EM waves is not the same as the physical aperture, the efficiency of waveguide aperture antennas (and horn antennas) typically ranges from 50% to 80% [122]. Rectangular waveguide apertures, with or without dielectric loading, are the most commonly used antennas for microwave imaging. Dielectric loading serves to reduce the required waveguide dimensions, which helps to increase spatial resolution of the image.

27.3.2.2 Horn Antennas A variation of the waveguide aperture antenna is obtained by flaring out the waveguide walls to form an EM horn. The horn antennas can provide a greater directivity to radiated fields (or higher gain) for some of the noninvasive and distant sensing applications. It is noted that the directivity or gain can also be increased by using structures of other antenna designs, such as corner reflectors and parabolic dish and dielectric lens antennas, designed for wireless communications applications (e.g., see Ref. 122 and other chapters in this book).

An antenna formed by a rectangular waveguide flared out in one field plane is called a sectorial horn antenna, whereas an antenna formed by a waveguide flaring out in both field planes is referred to as a pyramidal horn antenna. For best operation of a sectorial horn antenna flared out in the plane of the electric field (Figure 27.14), the flare angle (ψ) must be sufficiently small such that the area of the wavefront is approximately equal to the area of the aperture [123]. In this case, the field is constant over the aperture along the direction of flare but will vary as $\cos(\pi x_1/a)$ in the other direction or plane and a is the height of the waveguide. The square of the absolute magnitude of the electric field strength in the forward direction (proportional to gain) is given by

$$|E|^2 = \frac{2L}{\lambda}\left(\frac{2aE_y^o}{\pi r}\right)^2\left[C^2\left(\frac{b}{\sqrt{2\lambda L}}\right) + S^2\left(\frac{b}{\sqrt{2\lambda L}}\right)\right] \tag{27.20}$$

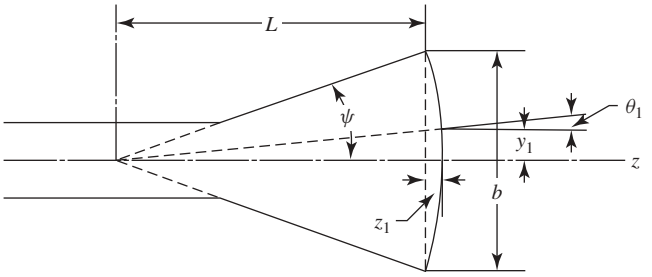


Figure 27.14 Schematic diagram of a sectorial EM horn antenna. (From Ref. 123.)

where λ is the wavelength, r is the radial distance, and C and S are the Fresnel integrals giving rise to the Cornu's spiral [123]. Equation (27.20) indicates that if b is increased for a given L, the field strength will first rise to a maximum and then fall, rising again to a secondary maximum that is smaller than the first maximum. A similar situation results if b and L are both increased to keep the flare angle constant. This suggests that as b is increased, contributions from some of the secondary sources on the wavefront are out of phase with others and tend to decrease instead of increase the field strength in the forward direction.

27.3.2.3 Ultrawideband Antennas For applications where narrow nanosecond pulses are employed to noninvasively sense the motion of the internal organs of the human body, the radiation and reception of the EM signals require an antenna with ultrawideband performance. Since the antennas must be sufficiently small to be placed on or near the body, small resistively loaded dipoles and bow-tie antennas have been proposed. These antennas are similar to those used for confocal microwave imaging and are discussed in more detail in the next section.

27.3.3 Antennas for Imaging Applications

Several tomographic techniques for tissue imaging are currently under investigation; indeed, considerable efforts have been directed toward breast cancer detection. It is noteworthy that the contrast in terms of dielectric constant and electrical conductivity between normal and malignant tissue is higher for the breast than many other organs. However, the same high contrast in tissue permittivity induces diffraction in the propagation of microwaves in tissue, which greatly complicates its use as a source for tissue imaging. It has frustrated many attempts to date to develop a clinically feasible imaging modality.

27.3.3.1 Microwave Tomography Microwave tomography has been explored to reconstruct images associated with variations of permittivity in body tissues. The prototypes have employed organ-in-water phantoms, that is, model targets are immersed in a tank of water. The most commonly used antennas are rectangular waveguide apertures, with or without dielectric loading. As mentioned previously, dielectric loading reduces the required waveguide dimensions, which helps to increase spatial resolution of the image. Some interesting results, among others, have been obtained for isolated kidneys [112] and heart tissues [115, 117].

27.3.3.2 Confocal Microwave Imaging Confocal microwave imaging (CMI) focuses backscattered signals to create images that indicate regions of significant permittivity contrast [124]. The higher permittivity of malignant tumors gives rise to a larger microwave scattering cross section than normal breast tissues. Range-gated spatial focusing overcomes the problems of attenuation and background clutter of breast tissues and is achieved through the use of electronically scanned antenna arrays. These arrays use antennas such as resistively loaded monopoles or dipoles as radiators for temporally short, broad-bandwidth pulses [125–128]. Resistively loaded bow-tie antennas have also been proposed [124, 129].

Resistively Loaded Dipoles A loaded dipole antenna is realized by resistively loading a dipole. This can be accomplished, for example, by using as dipole arms two tubes whose conductivity varies along their length, as shown in Figure 27.15.

For a loaded dipole antenna the conductivity variation along the dipole arms for optimal broadband behavior was proposed by Wu and King [125] (Figure 27.15). According to their results, using a thin-walled metallic tube of outer radius a_m and wall thickness $d \ll a_m$ with conductivity varying along z as (see Figure 27.15 right)

$$\sigma(z) = \frac{\sigma_0}{h}(h - |z|) \tag{27.21}$$

The current along the antenna is given by

$$I_z(z) = \frac{I_{z0}}{h}(h - |z|)e^{-jk_0|z|} \tag{27.22}$$

In Eq. (27.21), $\sigma_0 = (h/a_m)/d\zeta_0\psi_0$, where ψ_0 is the zero frequency value of

$$\psi = \frac{A_z(z)}{I_z(z)}\frac{4\pi}{\mu_0}$$

with typical values $6.0|\psi| < 12.0$ [130].

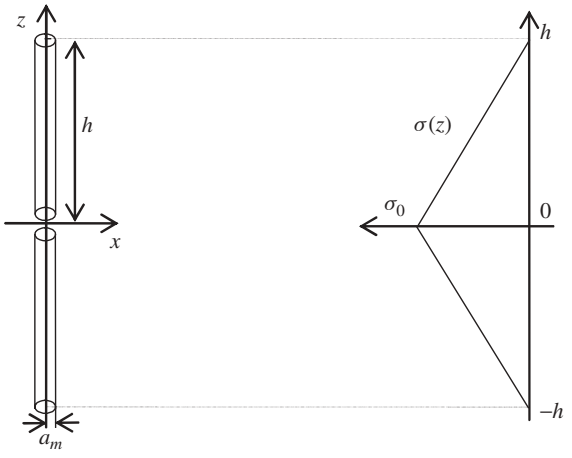


Figure 27.15 Resistively loaded dipole antenna. $\sigma(z)$ is the conductivity profile along the dipole.

Equation (27.22) represents a current wave traveling from the generator ($z = 0$) toward both ends of the dipole. Accordingly, a loaded dipole can be used efficiently as an antenna for UWB signals. In fact, when a UWB pulse is applied to the feed of an ideal dipole, two current impulses travel along the antenna elements until they reach the ends where the current goes to zero, giving rise to a reflected wave traveling back to the antenna feed. This process continues until radiation or ohmic losses bring the current to zero. In contrast, when an optimally loaded dipole is used, only a forward wave propagates along the arms with no reflections. However, this result is accomplished at the expense of efficiency that, for the loaded dipole antenna, varies between 10% and 40% depending on its length [131]. A further drawback of this type of antenna is that its input impedance is usually very high, $Z_i \cong 600\ \Omega$. Thus significant reflections ($\Gamma_i \cong 0.8$) occur at the drive point of the Wu–King dipole when a 50-Ω coaxial feed line is used. These reflections are reduced if the dipole is fed from a two-wire line.

In practical realizations, a resistively loaded dipole can be built by depositing a resistive film of variable thickness on a dielectric rod [132] or by soldering together a set of precision high frequency resistors [130, 133]. To improve the directivity of these antennas, "V"-shaped dipole arms, obtained by rotation through a given angle (α in Figure 27.16a) can be used [128]. An optimized version of this antenna has been obtained by curving the dipole arms and modifying the Wu–King loading profile (see Figure 27.16b) [134]. This geometry has been manufactured by printing the arms on a thin substrate and loading them with surface-mount resistors. Some problems in this structure arise from the need of a balun to transform the 50-Ω unbalanced feed line to the 200-Ω balanced line feeding the antenna. A bandwidth with VSWR lower than 2.0 between 1.2 and 8 GHz has been obtained [134].

Bow-Tie Antennas A bow-tie antenna is obtained by enlarging the arms of a planar dipole to form triangles. Hagness et al. [124] reported the design of a wideband bow-tie antenna suitable for near-surface biological sensing. The design was based on a modification of the continuous resistive loading proposed by Shlager et al. [129], in combination with the location of the antenna at the interface of the biological tissue. Figure 27.17 shows the antenna configuration. A bow-tie antenna with a flare length of $h = 4$ cm and

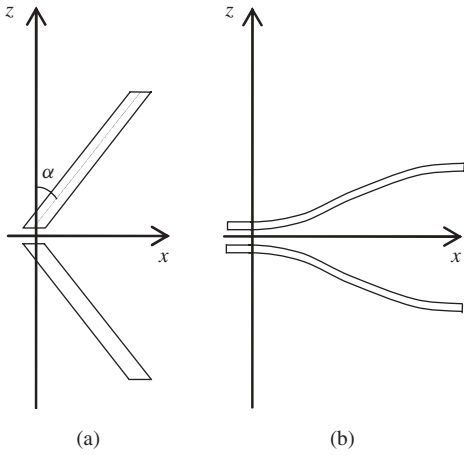


Figure 27.16 The V dipole: (a) linear profile and (b) curved taper.

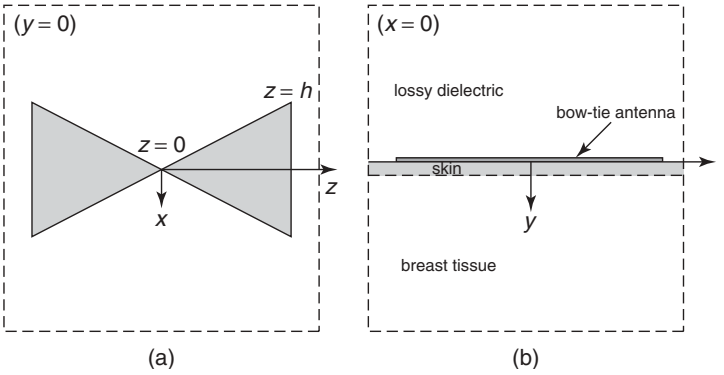


Figure 27.17 Bow-tie antenna backed with a lossy dielectric slab and located at the surface of the breast tissue: (a) top view and (b) lateral view. (From Ref. 124.)

a flare angle of 53° is located at the surface of the breast. The breast model is comprised of a 1-mm thick layer of skin ($\varepsilon_r = 36$, $\sigma = 4$ S/m) and a half-space of normal breast tissue ($\varepsilon_r = 9$, $\sigma = 0.4$ S/m). The antenna consisted of a material with the following conductivity:

$$\sigma(z) = \sigma_0 \left[\frac{1 - z/h}{1 + (\sigma_0/\sigma_{1/2} - 2)(z/h)} \right] \tag{27.23}$$

where z/h is the normalized axial distance along the bow-tie, σ_0 is determined by the metal used at the feed point, and $\sigma_{1/2}$ is chosen to give the desired level of suppression of the reflected pulse. A 3D FDTD simulation shows that the UWB end reflections for the bow-tie element can be reduced to 60 dB with proper resistive loading.

This optimal loading reduces the radiated pulse into the breast by only about 1 dB. For this optimized element, the backscatter responses of spherical tumors of 1.7–5.3-mm diameter embedded within the breast at depths of 3–5 cm are clearly detected [124].

To avoid the need of a balun for the coaxial feed of bow-tie antennas, bow-tie monopoles, also called tab monopoles, have been realized. The tab monopole consists of a subwavelength tapered radiating element fed by a suitable transmission structure (Figure 27.18a) [136].

The sizes of the tab monopole elements are approximately 0.12λ high and 0.22λ wide, with a 20°–40° taper. The tab monopole shown in Figure 27.18a is fed by a grounded coplanar waveguide line but other feeding configurations are possible including microstrips, striplines, and coaxial cables. A matching of the input impedance better than −10dB is obtained over an operational bandwidth better than 50% [136].

The broadband performance of tab monopoles can be improved further by transforming the tab in a staircase manner [135] as shown in Figure 27.18b. In this manner, a bandwidth of 77% (VSWR ≤ 2) was achieved. By reducing the tab dimensions, these tab antennas can cover bandwidths from 4 to 10 GHz with a VSWR less than 3 [137].

The loaded bow-tie antenna shown in Figure 27.19a is based on the use of continuous resistive loading and of a tab shape as reported by Shlager et al. [129]. This antenna is fed from a 50-Ω coaxial transmission line, through an image plane. The half-angle of the cone is $\alpha/2 = 78.1°$; this makes the characteristic impedance of the infinitely long, perfectly conducting bow-tie, equal to 50 Ω. The reflections of the metallic bow–tie have

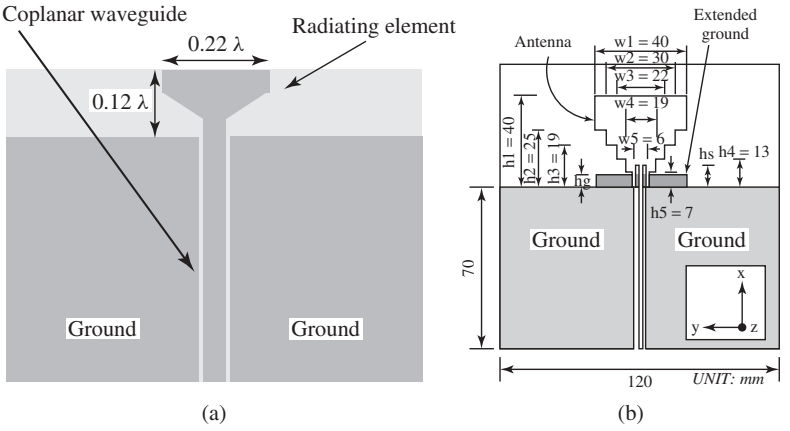


Figure 27.18 (a) Tab monopole and (b) staircase tab monopole. (From Ref. 135.)

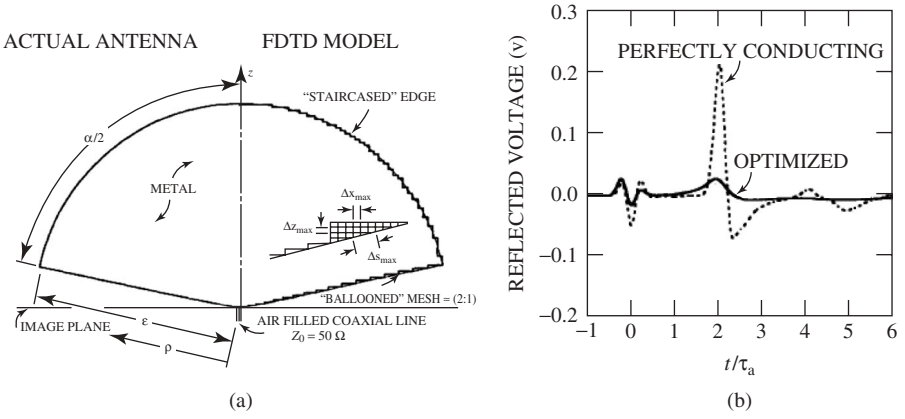


Figure 27.19 (a) Loaded bow-tie antenna and (b) comparison of the computed reflected voltage for the optimized bow-tie with that of a metallic bow-tie. (From Ref. 129.)

been reduced by adding a discrete capacitance ($C \approx 0.08$ pF) at the driving point, and by using resistive loading along its length.

The internal resistance per unit length of the bow-tie was chosen to have the same dependencies as proposed by Wu and King [125] and, consequently, the resistance per square of the thin sheet that forms the antenna has the following dependence on ρ/s:

$$R(\rho/s) = R(1/2)\left[\frac{\rho/s}{1-\rho/s}\right], \quad 0 \le \rho/s \le 1 \tag{27.24}$$

The reflected voltage in the feeding transmission line for this case is graphed in Figure 27.19b (solid line) and compared with that obtained with a perfectly conducting bow-tie (dashed line). The figure shows that the presence of resistive loading strongly reduces the reflections.

27.3.3.3 Microwave Thermoelastic Tomography A high peak-power microwave pulse, upon absorption by soft tissues in the body, launches a thermoelastic wave of acoustic pressure in the tissue medium, which propagates away from the launching site in all directions. The thermoelastic pressure can be detected with an acoustic or ultrasound sensor or transducer. Thus it provides a dual modality for tissue imaging that was first proposed in the 1980s to form planar and/or tomographic images of body tissues. Most feasibility studies were conducted using rectangular waveguide antennas, with or without dielectric loading [111, 118–121]. For example, an open-waveguide applicator delivered the microwave pulses to the hand-in-water phantom at the surface of the water tank. Microwave-induced thermoelastic pressure waves were detected using a square of piezoelectric transducer array located at the bottom of the tank [118]; whereas a microwave thermoelastic tomography system specifically designed for human breast imaging consisted of a hemispherical, 300-mm diameter bowl in which 64 piezoceramic transducers were attached in a spiral pattern [120, 121]. The bowl was mounted on a cylindrical shaft and could be rotated 360° about its axis. The entire device was immersed in water contained in a cylindrical tank. For human breast imaging, the tank itself was placed beneath an examination table, in which a circular hole was cut and mated to the top of the imaging tank. Initially, a helical RF antenna was installed at the base of the transducer array.

A semirigid coaxial cable, which passed through the center of the shaft, was used to feed microwave energy to the helical antenna. This antenna arrangement produced circularly polarized microwaves, and a slowly varying spatial intensity pattern within the imaging volume. An updated version of the system replaced the helical antenna with waveguide antennas [138]. The waveguide antennas produced linear polarization and averted the strong, thermoelastic signals generated at the surface of the helical antenna. These unwanted thermoelastic signals overlapped in time with the much weaker thermoelastic signals from the microwave absorbing tissues. A schematic diagram of the latter system is shown in Figure 27.20. Note that this system has eight waveguide antennas and 128 ultrasonic transducers arrayed on the hemispherical surface.

27.3.3.4 Magnetic Resonance Imaging Current generations of clinical MRI scanners are based on induced nuclear magnetic resonance of water molecules. An applied static magnetic field induces the magnetic dipole moments associated with hydrogen protons in tissue to be aligned in the direction of the applied field. Each proton or dipole moment precesses about the static field, $\mathbf{B}_0$, in the transverse plane at the Larmor frequency f_0. An RF magnetic field that oscillates at f_0 is applied using an appropriate antenna in the direction transverse to $\mathbf{B}_0$, and then by choosing the pulse duration of the excitation, the magnetic dipole moments can be tipped into the transverse direction. The decaying signal following the pulse excitation is detected using an RF receiving antenna. For an image of high quality, the applied RF magnetic field should be uniform over all regions of the body being imaged, since a nonuniform field introduces distortion into the image. Moreover, for maximum energy coupling into the hydrogen nuclei, it is necessary to have an antenna (coil) that can generate circularly polarized fields. Birdcage and TEM coils have been designed to satisfy both of these requirements and they are described next.

Birdcage Coils Birdcage coils typically are made by using two conducting end-rings (ERs) connected by a number of conducting legs (rungs) distributed at equal spacing

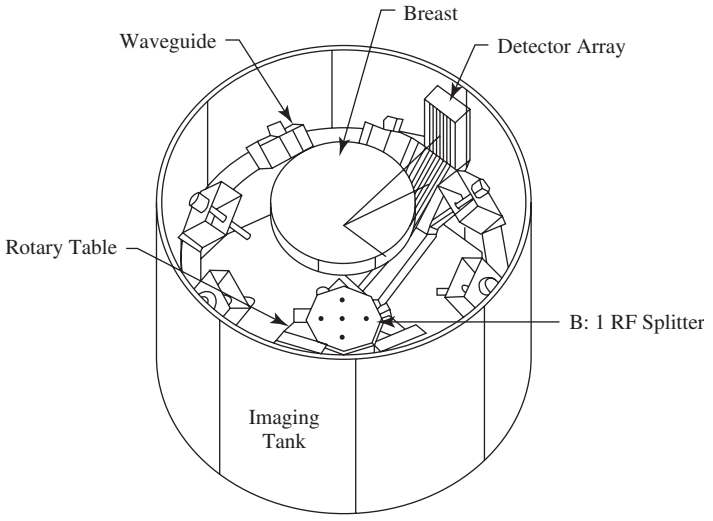


Figure 27.20 Schematic diagram of the thermoelastic tomography system. This system has eight waveguide antennas and 128 ultrasonic transducers arrayed on the hemispherical surface. (From Ref. 138.)

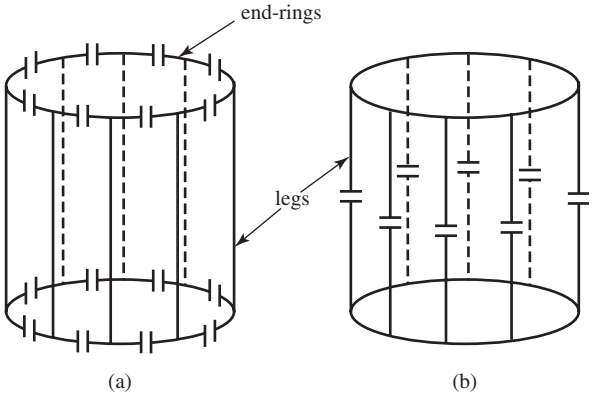


Figure 27.21 (a) High pass birdcage coil and (b) low pass birdcage coil. (From Ref. 143.)

around the perimeter of the ERs [139]. Capacitors are inserted into legs or ERs to achieve fine tuning of the desired resonant mode (see Figure 27.21). A high pass (HP) coil has the capacitors situated in ER segments (Figure 27.21a), while a low pass (LP) coil has capacitors placed in the legs (Figure 27.21b). A bandpass (BP) coil has capacitors at both locations.

Birdcage coils are resonators characterized by more than one resonant mode: in fact, an N-leg coil has $(N/2) + 1$ distinct resonant modes [139]. Conductors can be modeled as inductances in an equivalent circuit in order to predict the complete resonance spectrum [140–143]. According to the analysis described in Jin [142] and Giovannetti et al. [143], a high pass birdcage coil can be represented as shown in Figure 27.22, where Lleg

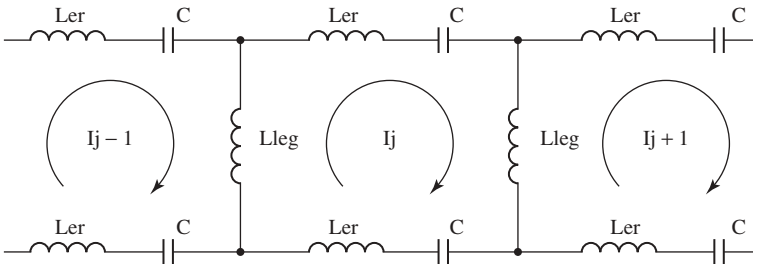


Figure 27.22 Section of the equivalent circuit for a high pass birdcage coil. (From Ref. 143.)

represents the self-inductance of one leg, C the capacitance of the capacitor between two legs, and Ler the self-inductance of the end-ring segments.

For a more rigorous analysis of this structure, the mutual inductances between conductors should be taken into account. Several approximate formulas have been proposed in the literature for computing self- and mutual inductances for strip and cylindrical conductors [141–143].

By applying Kirchhoff's voltage law to each loop with current I_j, and $\lambda = 1/\omega^2$, the following matrix equation can be obtained:

$$\boldsymbol{K}\mathbf{I} = \lambda \boldsymbol{H}\mathbf{I} \tag{27.25}$$

where $\mathbf{I}$ represents a column vector given by $\mathbf{I} = [I_1, I_2, \ldots, I_N]^{\mathrm{T}}$; $\boldsymbol{K}$ and $\boldsymbol{H}$ are $N \times N$ square matrices whose elements are given by

$$K_{j,k} = L\mathrm{leg}_{j,k} - L\mathrm{leg}_{j+1,k} - L\mathrm{leg}_{j,k+1} + L\mathrm{leg}_{j+1,k+1} + 2(L\mathrm{er}_{j,k} - \mathrm{L'er}_{j,k}) \tag{27.26}$$

$$H_{j,k} = 2\delta_{j,k}/C \tag{27.27}$$

where $\delta_{j,k}$ indicates the Kronecker delta.

For a nontrivial solution of Eq. (27.25), it is necessary that

$$\det\{\boldsymbol{K} - \lambda \boldsymbol{H}\} = 0 \tag{27.28}$$

The N-degree polynomial equation (27.28) has N solutions $(\lambda_1, \lambda_2, \ldots, \lambda_N)$, which correspond to the N resonant frequencies. For each λ_k, a solution $\mathbf{I}_k$ exists for $\mathbf{I}$.

The determination of the birdcage resonance spectrum can be simplified by neglecting the mutual inductances [142]. For a birdcage coil with N legs, the current I_j must satisfy the periodic condition $I_{j+N} = I_j$. Therefore the N linearly independent normalized currents have the form

$$I_{jk} = \begin{cases} \cos\dfrac{2\pi jk}{N}, & k = 0, 1, 2, \ldots, N/2 \\ \sin\dfrac{2\pi jk}{N}, & k = 1, 2, \ldots, (N/2) - 1 \end{cases} \tag{27.29}$$

where I_{jk} represents the value of the current in the jth loop for the kth solution (kth mode). The resonant frequencies for a low pass birdcage are given by

$$\omega_k = \left[C\left(L\mathrm{er} + L\mathrm{leg}/2 \sin^2 \frac{\pi k}{N} \right) \right]^{-1/2} \qquad (k = 0, 1, 2, \ldots, N/2) \tag{27.30}$$

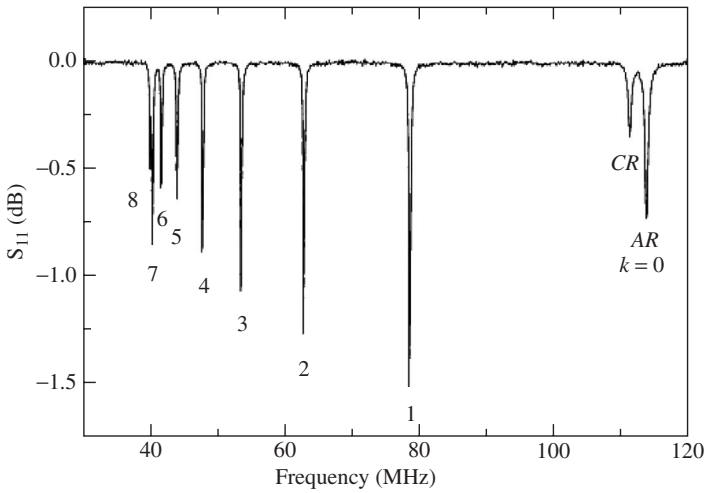


Figure 27.23 The resonance spectrum of a 16-leg high pass birdcage coil. (From Ref. 141.)

Similarly, for a high pass coil the resonant frequencies are

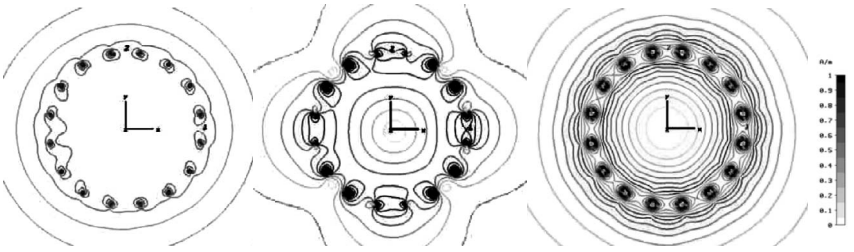
$$\omega_k = \left[C \left(L_{\text{er}} + 2L_{\text{leg}} \sin^2 \frac{\pi k}{N} \right) \right]^{-1/2} \qquad (k = 0, 1, 2, \ldots, N/2) \tag{27.31}$$

Figure 27.23 shows the measured resonance spectrum of a high pass birdcage coil ($N = 16$).

From the figure it can be seen that (1) the highest resonance frequency occurs at $k = 0$, which means constant and opposite currents in the two end-rings and a null current in the legs, giving rise to an antiresonant (AR) mode; (2) the two solutions in Eq. (27.29) ($\cos(2\pi jk/N)$ and $\sin(2\pi jk/N)$ with the same k ($k = 1, \ldots, 7$) have the same resonant frequency (degenerate modes); and (3) the first two solutions, for $k = 1$, produce a circularly polarized uniform magnetic field—a highly desirable feature for medical applications of MRI.

In designing an MRI apparatus, shielding is used to minimize the interaction of the RF field generated by the birdcage with other parts of the system. The effect of shielding on the resonance frequency and field distribution in the coil can be studied by the image method, which assumes that the field produced by the induced current on the shield is identical to the field produced by image currents radiating in the absence of the shield. Because the birdcage coil conductors are parallel to the shield, the image currents have to flow in the opposite direction in order to satisfy the boundary conditions. The use of shields reduces the inductances and increases the resonance frequency [140–143].

Another important aspect of MRI antennas is their ability to produce a uniform magnetic field inside the cage. Using the Biot–Savart equation, the magnetic field distribution can be computed from conductor currents associated with a given mode among the various modes of excitation sustained by the birdcage coil. Specifically, in order to calculate the magnetic field generated by the currents, a coil can be subdivided into leg and end-ring segments in order to separately study their contribution to the total field. Note that the use of Biot–Savart equation implies a quasistatic magnetic field assumption, and it is



(a) (b) (c)

Figure 27.24 Contour plots of the H_1 field of a 16-leg low pass birdcage coil for mode (a) $k = 1$, (b) $k = 2$, and (c) $k = 3$.

valid when the coil dimensions are much smaller than the wavelength. In practice, this assumption is satisfied at frequencies used in most MRI scanners. For example, at 64 MHz ($B_0 = 1.5$ T), the wavelength is 4.7 m; thus the quasistatic assumption holds for practical birdcage dimensions. This approximation also holds when increasing the static magnetic field strength, since the coil dimension decreases accordingly. Figure 27.24 shows the contour plot of H_1 fields for a 16-leg low pass birdcage coil [142]. A higher field uniformity for mode $k = 1$ compared to mode $k = 3$ is clearly visible.

TEM Coils The use of higher static magnetic fields B_0 and, consequently, higher Larmor frequencies for better MR images has introduced systems operating at 3–9.4 T (128–400 MHz). This development requires new types of antennas. One of the most commonly used is the transverse electromagnetic (TEM) coil proposed by Röschmann [144] and Bridges [145]. These coils are based on TEM transmission lines with capacitive loads of various configurations and have become known as "TEM resonators" [146].

The geometries of two TEM resonators are shown in Figure 27.25. These resonators consist of multiple longitudinal conductors arranged in a circular cylindrical pattern and enclosed by a cylindrical shield with end plates that form a cavity. Figure 27.25a shows the most commonly used design in which the inner conductors are hollow tubes that do not connect directly to the end plates but are separated from them by short air gaps. The interior of each tube contains another conducting cylinder whose length can be adjusted, giving rise to an open-circuited transmission line that acts as a capacitive tuning stub. Figure 27.25b shows a design that uses microstrip conductors, supported by a hollow cylinder made of dielectric material, and connected to the shield with commercially available fixed and trimmer capacitors. These coils are driven by sources connected to the end of one or more conductors.

Many theoretical studies on TEM resonators have been conducted [147–151]. Some investigations calculated resonance frequencies in terms of equivalent total lumped inductance and capacitance [149, 151]; others derived expressions for field distribution and resonance frequencies using transmission line theory [147, 148, 150]. In particular, it has been shown that the TEM resonator behaves like a loaded multiconductor transmission line (MTL) capable of supporting standing waves, shown in Figure 27.26 [148].

The voltages $\mathbf{V}_n(z)$ between the generic and the reference conductor and the currents $\mathbf{I}_n(z)$ satisfy the vector transmission line equations:

$$\frac{d}{dz}\mathbf{V} = -\mathbf{Z}\,\mathbf{I} \quad \frac{d}{dz}\mathbf{I} = -\mathbf{Y}\,\mathbf{V} \tag{27.32}$$

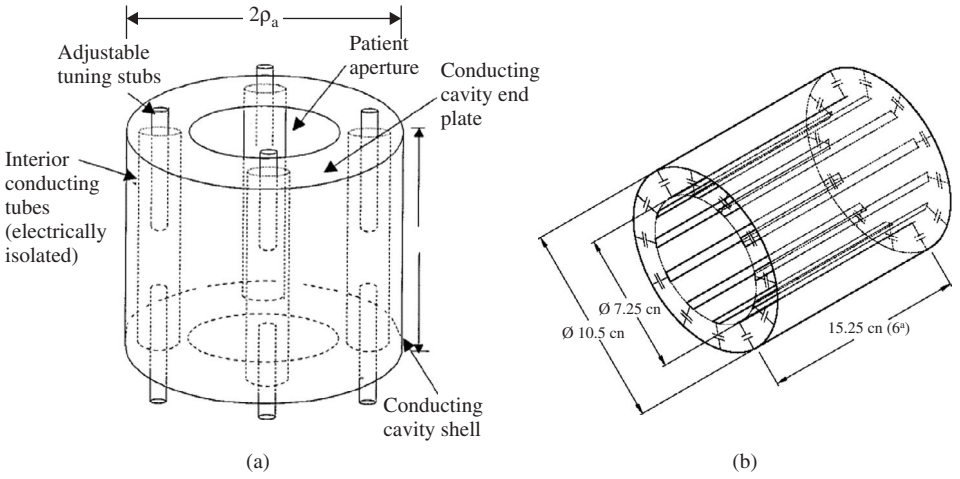


Figure 27.25 TEM resonators. (a) Configuation of coaxial TEM coil. In practice, a large even number of elements (e.g., 16) is used, but only four elements are shown in the figure for simplicity (from Ref. 147). (b) Configuration of microstrip TEM coil (from Ref. 148.)

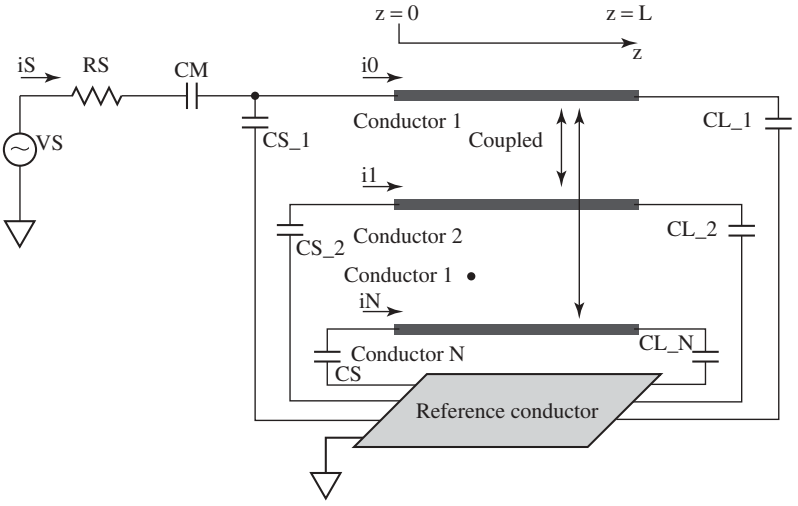


Figure 27.26 Schematic of the MTL model for the TEM resonator. (From Ref. 148.)

where $\boldsymbol{Z}$ and $\boldsymbol{Y}$ are the per-unit-length impedance and admittance matrices of the line, which for physical reasons are symmetric. These matrices can, in turn, be expressed as $\boldsymbol{Y} = \boldsymbol{G} + j\omega\boldsymbol{C}$ and $\boldsymbol{Z} = \boldsymbol{R} + j\omega\boldsymbol{L}$, which involve the per-unit-length resistance ($\boldsymbol{R}$), inductance ($\boldsymbol{L}$), conductance ($\boldsymbol{G}$), and capacitance ($\boldsymbol{C}$). For an empty coil the conductance matrix $\boldsymbol{G}$ vanishes and it is also possible to set the resistance matrix $\boldsymbol{R} = 0$, since the losses are dominated by dissipation inside the tissues.

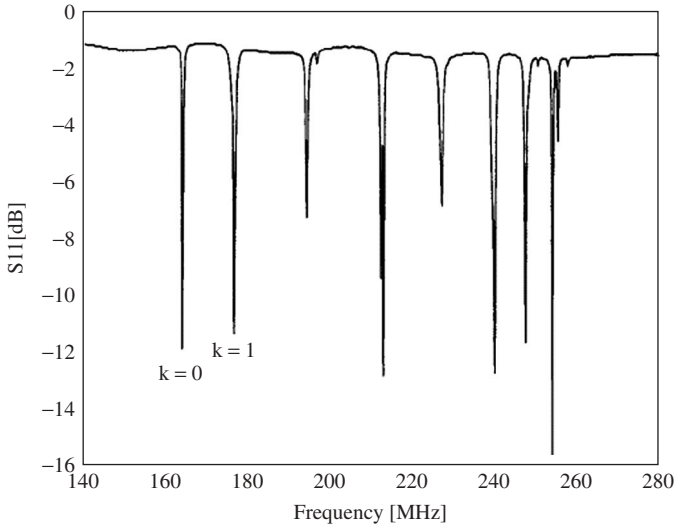


Figure 27.27 Return loss measured for a 16-element coil for head imaging. (From Ref. 147.)

In the absence of excitation, and for symmetrically placed tubes, the condition for resonance results in

$$\prod_{n=1}^{N} (1-\Gamma_n^+ \Gamma_n^- e^{-2jk_n l}) = 0 \tag{27.33}$$

where Γ_n^+ and $\Gamma_n^-$are the modal decompositions of the reflection coefficients at the end of the lines. The frequency response of an $N = 16$ element coil for head imaging was measured experimentally [147]. The return loss S_{11} measured at the input of this coil is shown in Figure 27.27. It can be seen that $N/2 + 1 = 9$ modes are present. The desired $m = 1$ mode appears near 177 MHz.

The magnetic field distribution inside the coil is determined by the conductor currents according to

$$\mathbf{H} = \sum_n \lfloor \mathbf{I}_n^{m+} e^{-jk_n z} - \mathbf{I}_n^{m-} e^{+jk_n z} \rfloor \mathbf{h}_n(x, y) \tag{27.34}$$

where k_n is the wavenumber of mode n, which for TEM modes and lossless conductors is given by $k_n = \omega_{n/c}$, and $\mathbf{h}_n$ is the transverse vector mode function.

Using Eq. (27.34) it is possible to compute field distributions for modes of all orders. In particular, Figure 27.28 presents the intensity of the transverse magnetic field at the midplane, for mode $m = 1$, in a 16-element coil [147]. A single source was used to excite the coil, resulting in linear polarization over the head-accessible portion of the coil. It can be seen that a major portion of the magnetic field energy resides in inaccessible regions. This aspect of the coil's operation is shared not only by the TEM resonators but also by conventional birdcage coils. Because circular polarization is most effective in exciting the MR signal, coils are commonly driven in a quadrature fashion to produce circularly polarized fields. When driven in this manner, the field is highly homogeneous over the central region.

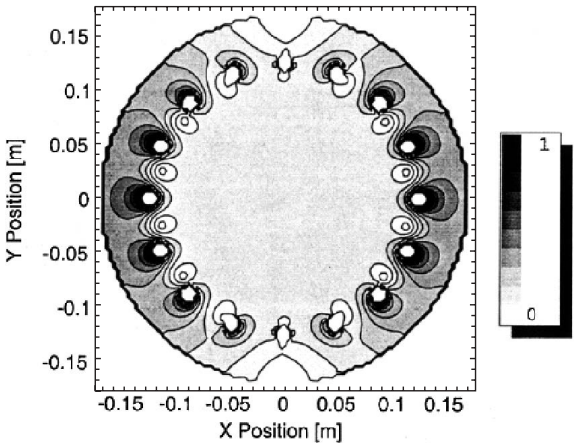
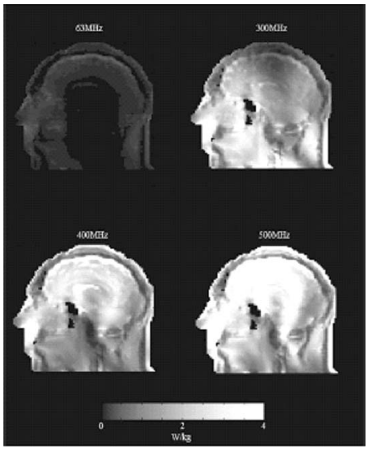


Figure 27.28 The B_1 field intensity for mode 1 in a 16-element TEM coil with a single drive point. (From Ref. 147.)

Patient's Safety Aspects The design of MRI coils must also take into account, together with resonance frequency evaluation, the patient's safety aspects. In fact, during MRI patients are exposed to relatively large radiofrequency pulses that determine energy absorptions inside the body that, once converted into heat, can give rise to adverse thermal effects. In order to assess the exposure, SAR and temperature distributions have to be evaluated and the obtained peak values compared with limits reported in existing international guidelines.

The finite-difference time-domain (FDTD) method has been used to determine the SAR distribution during an MRI head examination for frequencies ranging from 50 to 400 MHz and by using anatomically realistic human models [152–155]. Detailed thermal models including conductive and convective heat transfer mechanisms have been used to predict temperature elevations as a function of various exposure parameters in the leg due to an RF surface coil [156] and in a realistic head model [155, 157]. Figure 27.29 shows the results obtained by Nguyen et al. [157].

In particular, Figure 27.29a shows the SAR distributions in the transverse sagittal plane while Figure 27.29b lists the maximum 1-g SAR values for specific tissues within the head and the maximum temperature increments in the head at frequencies of 63, 300, 400, and 500 MHz. Figure 27.29b shows that the maximum SAR is frequency dependent and, in general, increases as the frequency increases. However, at 300 MHz the SAR values are somewhat higher than expected. Collins et al. [152] and Gandhi and Chen [153] reported this same phenomenon near 300 MHz. These authors hypothesized that this is an indication of a resonance occurring near 300 MHz due to the relative dimensions of the head in comparison to the wavelength. In fact, at 300 MHz the head used in the study is nearly half a wavelength in diameter. Another observation based on these results is that the maximum SAR absorption occurs mostly within the skin. This high absorption rate, in and near the skin (including muscle), is not surprising since the electric field magnitude within the birdcage increases as one approaches the wire of the coil. By solving the thermal problems, these authors found a maximum temperature increase of 1.34 °C at 300 MHz (see Figure 27.29b). On the basis of the ICNIRP recommendation, no adverse health effects are expected if the temperature increase in any part of the head



(a)

Frequency	Tissue Type	Max 1g SAR (W/kg)	ΦT_{MAX} (°C)
63 MHz	Cornea/Sclera	0.54	0.03
	Muscle	0.51	
	Tendon	0.43	
300 MHz	Skin	22.30	1.34
	Cerebellum	8.39	
	Tendon	8.34	
400 MHz	Skin	11.31	0.53
	Cerebellum	6.55	
	Brain (White)	5.71	
500 MHz	Skin	16.92	0.93
	Brain (White)	8.47	
	Tendon	8.39	

(b)

Figure 27.29 (a) Sagittal cross section of SAR distribution and (b) list of SARs and temperature increments at various frequencies. (From Ref. 157.)

does not exceed about 1 °C [158]. Hence the simulation results that may have clinical significance are those at 300 and 500 MHz. However, it must be noted that the thermal model presented does not account for all thermal regulatory responses of the body and thus represents a worst case scenario. The actual temperature increases within the head are likely to be less than those reported.

The current U.S. Food and Drug Administration (FDA) and International Electrotechnical Commission (IEC) "normal mode" limits for MRI are, respectively, 3.0 and 3.2 W/kg averaged over the head. Recently, numerical computations for SARs of a human head model inside TEM coils at frequencies from 64 to 400 MHz have been compared to the safety limits [155]. A comparison of the ratios of maximum local to whole-head average SAR with the ratios in the safety limits is shown in Table 27.2 for six MRI frequencies. It can be seen that, at every frequency, the ratio of maximum local SAR in either 1-g or 10-g regions to the head-average SAR is higher than that in the safety limits. These results suggest that local SAR in the human head will nearly always exceed existing limits before whole-head SAR does. Because only whole-head or whole-body SAR is typically monitored in clinical MRI scans, this also indicates that existing limits on local SAR are likely exceeded often in practice. Numerically computed 10-g SAR ratios are closer to the ratios of corresponding IEC limits than 1-g SARs are to the ratios of corresponding FDA limits. The IEC SAR limits appear much less restrictive overall than FDA SAR limits.

MRI is a widely accepted tool for the diagnosis of a variety of disease states. Today, however, MRI is contraindicated for patients implanted with pacemakers (PMs). The most adverse effect seems to be the heating of the catheter produced by the high currents induced by the RF field used in the MRI procedure [159]. This problem has been studied, both experimentally and numerically, by considering a thorax model placed in a birdcage coil [160]. A pacemaker equipped with a 60-cm long catheter has been inserted inside the thorax. The presence of this implant gives rise to peak SAR averaged over 1 mm^3 of about 1800 W/kg close to the catheter tip, while the average SAR is about 1 W/kg—the same value obtained in the absence of the pacemaker. The obtained SAR

TABLE 27.2 Ratio of Maximum Local to Whole-Head Average SAR for a Human Head Model in a TEM Coil at Several Frequencies Compared to the Ratios in the Regulatory Limits

Frequency (MHz)	SAR_{1g}/SAR_W[c]	SAR_{10g}/SAR_W
64	6.1	3.4
128	5.2	3.2
200	5.6	3.5
300	5.4	4.1
340	6.7	5.0
400	6.0	4.4
FDA[a]	2.7	—
IEC[b]	—	3.12

[a]FDA, Food and Drug Administration limits.
[b]IEC, International Electrotechnical Commission limits.
[c]SAR_w, whole-head average SAR.
Source: From Ref. 155.

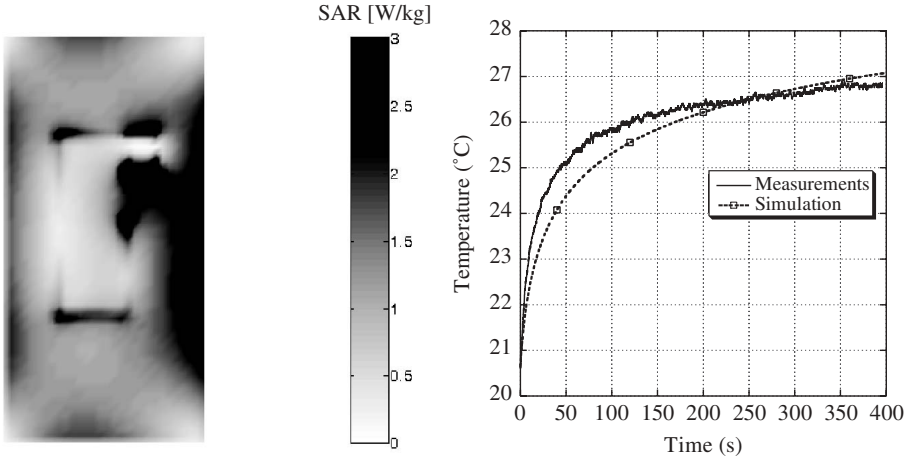


Figure 27.30 (a) SAR distribution in the coronal section passing through the catheter and (b) time behavior of the temperature, at the catheter tip, on the simulated and physical phantoms. (From Ref. 160.)

distribution (Figure 27.30a) gives rise to temperature increments up to 6 °C for 6-min MRI investigations (Figure 27.30b). However, also in this case, lower temperature increments are expected if more realistic human body models together with blood perfusion and thermoregulation are considered.

27.3.4 Summary

There are many different types of antennas for biomedical applications. The choice depends on frequency, power, SAR and its distribution, intended biological target, and the domain of application. The various antennas and application scenarios would depend

on whether the exposure is whole body or partial body, or the target organ or tissue. Each of these approaches allows an abundant array of antenna configurations. Beginning with antennas for noninvasive sensing applications, the various types of antenna and application scenarios are presented in this chapter, along with discussions on techniques for design and their expected performance. Antennas for microwave and magnetic resonance imaging then followed. Antennas can also be divided into categories on the basis of whether the antenna is designed to be used in close proximity to or in direct contact with the body. Because imaging and sensing both involve signal detection, the same type of antenna can often be employed for both sensing and imaging purposes and they include waveguide aperture and horn antennas. Likewise, resistively loaded dipoles and bow-tie antennas can be used to transmit and receive UWB signals found in confocal microwave imaging and internal organ sensing applications.

Clinical MRI scanners are based on induced nuclear magnetic resonance of water molecules. For images of high quality, the applied RF magnetic field should be uniform over all regions of the body being imaged since a nonuniform field introduces distortion into the image. Moreover, for maximum energy coupling into the hydrogen nuclei it is necessary to have an antenna that can generate circularly polarized fields in the range of 30–500 MHz. Birdcage and TEM coils designed to satisfy both of these requirements are described, along with their performance characteristics.

In addition, the patient's safety during MRI procedures is of interest, especially for some newer imaging protocols involving levels of applied magnetic fields as high as 10 T. In practice, whole-head or whole-body SAR and temperature are typically monitored during clinical MRI scans. Under worst case scenarios, computer simulations have shown that the temperature increase in the human head can approach or exceed $1\,^{\circ}$C (the current ICNIRP Guideline) at 300 and 500 MHz. Recent results indicate that existing USFDA and IEC safety limits on local SAR are likely exceeded often under most circumstances in practice.

The MRI procedure is contraindicated for patients implanted with pacemakers. It is interesting to note that computer simulation has shown a catheter inserted inside the thorax can give rise to a peak local SAR of about 1800 W/kg near the catheter tip, while the average SAR is about 1 W/kg,—the same value in the absence of the pacemaker. This SAR can produce a temperature increase of $6\,^{\circ}$C in 6 min of MRI imaging, under the worst conditions.

Antennas have been an important part of the research and development conducted to enhance medical applications in diagnostic sensing and imaging, as well as in therapeutic treatments. A number of successful clinical applications in imaging, especially MRI, have bolstered confidence in the potential of nonionizing RF and microwave energy. Clearly, research and development in biomedical applications of EM systems will continue to expand, and with it our understanding of biological interactions and progress in new biomedical applications of electromagnetic theory and technology.

REFERENCES

1. G. M. Hahn, Hyperthermia for the engineer: a short biological primer, *IEEE Trans. Biomed. Eng.*, Vol. 31, No. 1, pp. 3–8, 1984.
2. J. C. Lin, Hyperthermia therapy, in *Encyclopaedia of Electrical and Electronics Engineering*, J. G. Webster (Ed.), John Wiley & Sons, Hoboken, NJ, 1999, pp. 450–460.

3. W. T. Joines, Frequency-dependent absorption of electromagnetic energy in biological tissue, *IEEE Trans. Biomed. Eng.*, Vol. 31, No. 1, pp. 17–20, 1984.
4. M. G. Skinner, M. N. Iizukay, M. C. Koliosz, and M. D. Sherary, A theoretical comparison of energy sources—microwave, ultrasound and laser—for interstitial thermal therapy, *Phys. Med. Biol.*, Vol. 43, pp. 3535–3547, 1998.
5. C. J. Diederich, Thermal ablation and high temperature thermal therapy: overview of technology and clinical implementation, *Int. J. Hyperth.*, Vol. 21, No. 8, pp. 745–753, 2005.
6. C. Gabriel, Dielectric properties of biological materials. Bioengineering and biophysical aspects of electromagnetic fields, in *Handbook of Biological Effects of Electromagnetic Fields*, 3rd ed., F. Barnes and B. Greenebaum (Eds.), CRC Press, Boca Raton, FL, 2007, pp. 51–100.
7. J. R. Oleson, A review of magnetic induction methods for hyperthermia treatment of cancer, *IEEE Trans. Biomed. Eng.*, Vol. 31, No. 1, pp. 91–97, 1984.
8. J. Bach Andersen, A. Baun, K. Harmark, L. Heinzl, P. Raskmark, and J. Overgaard, A hyperthermia system using a new type of inductive applicator, *IEEE Trans. Biomed. Eng.*, Vol. 31, No. 1, pp. 21–27, 1984.
9. R. H. Johnson, J. R. James, J. W. Hand, J. W. Hopewell, P. R. C. Dunlop, and R. J. Dickinson, New low-profile applicators for local heating of tissue, *IEEE Trans. Biomed. Eng.*, Vol. 31, No. 1, pp. 28–37, 1984.
10. P. Y. Cresson, C. Michel, L. Dubois, M. Chive, and J. Pribetich, Complete three-dimensional modeling of new microstrip-microslot applicators for microwave hyperthermia using the FDTD method, *IEEE Trans. Microwave Theory Tech.*, Vol. 42, No. 12, pp. 2657–2666, 1994.
11. J. Carlier, V. Thomy, J. C. Camart, L. Dubois, and J. Pribetich, Modeling of planar applicators for microwave thermotherapy, *IEEE Trans. Microwave Theory Tech.*, Vol. 50, No. 12, pp. 3036–3042, 2002.
12. A. W. Guy, J. F. Lehmann, and J. B. Stonebridge, Therapeutic applications of electromagnetic power, *Proc. IEEE*, Vol. 62, pp. 55–75, 1974.
13. J. C. Lin, G. Kantor, and A. Ghods, A class of new microwave therapeutic applicators, *Radio Sci.*, Vol. 17, pp. 119S–123S, 1982.
14. S. M. Michaelson and J. C. Lin, *Biological Effects and Health Implications of Radiofrequency Radiation*, Plenum, New York, 1987, Chap. 3.
15. R. C. Gupta and S. P. Singh, Analysis of the SAR distributions in three-layered bio-media in direct contact with a water-loaded modified box-horn applicator, *IEEE Trans. Microwave Theory Tech.*, Vol. 53, No. 9, pp. 2665–2671, 2005.
16. G. C. Van Rhoon, P. J. M. Rietveld, and J. Van der Zee, A 433 MHz Lucite cone applicator for superficial hyperthermia, *Int. J. Hyperthermia*, Vol. 14, No. 1, pp. 13–27, 1998.
17. T. Samaras, P. J. M. Rietveld, and G. C. Van Rhoon, Effectiveness of FDTD in predicting SAR distributions from the Lucite cone applicator, *IEEE Trans. Microwave Theory Tech.*, Vol. 48, No. 11, pp. 2059–2033, 2000.
18. J. C. Lin (Ed.), *Special Issue on Phased Arrays for Hyperthermia Treatment of Cancer, IEEE Trans. Microwave Theory Tech.*, Vol. 34, pp. 481–482, 1986.
19. P. F. Turner, Regional hyperthermia with an annular phased array, *IEEE Trans. Biomed. Eng.*, Vol. 31, No. 1, pp. 106–114, 1984.
20. F. A. Gibbs, Jr., M. D. Sapozink, K. S. Gates, and J. R. Stewart, Regional hyperthermia with an annular phased array in the experimental treatment of cancer: report of work in progress with a technical emphasis, *IEEE Trans. Biomed. Eng.*, Vol. 31, No. 1, pp. 1115–1119, 1984.
21. F. Sterzer, Microwave medical devices, *IEEE Microwave Mag.*, Vol. 3, No. 1, pp. 65–70, 2002.

22. P. R. Stauffer, F. Rossetto, M. Leoncini, and G. Biffi Gentilli, Radiation patterns of dual concentric conductor microstrip antennas for superficial hyperthermia, *IEEE Trans. Biomed. Eng.*, Vol. 45, No. 5, pp. 605–613, 1998.

23. S. Jacobsen, P. R. Stauffer, and D. G. Neuman, Dual-mode antenna design for microwave heating and noninvasive thermometry of superficial tissue disease, *IEEE Trans. Biomed. Eng.*, Vol. 47, No. 11, pp. 1500–1509, 2000.

24. F. Rossetto and R. Stauffer, Theoretical characterization of dual concentric conductor microwave array applicators for hyperthermia at 433 MHz, *Int. J. Hyperthermia*, Vol. 17, No. 3, pp. 258–270, 2001.

25. K. S. Nikita, N. G. Maratos, and N. K. Uzunoglu, Optimization of the deposited power distribution inside a layered lossy medium irradiated by a coupled system of concentrically placed waveguide applicators, *IEEE Trans. Biomed. Eng.*, Vol. 45, No. 7, pp. 909–920, 1998.

26. M. E. Kowalski and J. M. Jin, Model-based optimization of phased arrays for electromagnetic hyperthermia, *IEEE Trans. Microwave Theory Tech.*, Vol. 52, No. 8, pp. 1964–1977, 2004.

27. J. Nadobny, W. Wlodarczyk, L. Westhoff, J. Gellermann, R. Felix, and P. A. Wust, Clinical water-coated antenna applicator for MR-controlled deep-body hyperthermia: a comparison of calculated and measured 3-D temperature data sets, *IEEE Trans. Biomed. Eng.*, Vol. 52, No. 3, pp. 505–519, 2005.

28. J. C. Lin, A. W. Guy, and C. C. Johnson, Power deposition in a spherical model of a man exposed to 1–20MHz electromagnetic fields, *IEEE Trans. Microwave Theory Tech.*, Vol. 21, No. 12, pp. 791–797, 1973.

29. W. T. Joines, C. F. Blackman, and M. A. Hollis, Frequency broadening of the RF power-density window for calcium-ion efflux from brain tissue, *IEEE Trans. Biomed. Eng.*, Vol. 28, No. 8, pp. 568–573, 1981.

30. W. Gee, S. W. Lee, N. K. Bong, C. A. Cain, R. Mittra, and R. L. Magin, Focused array hyperthermia applicator: theory and experiment, *IEEE Trans. Biomed. Eng.*, Vol. 31, No. 1, pp. 38–46, 1984.

31. G. Arcangeli, P. P. Lombardini, G. A. Lovisolo, G. Marsiglia, and M. Piattelli, Focusing of 915 MHz electromagnetic power on deep human tissues: a mathematical model study, *IEEE Trans. Biomed. Eng.*, Vol. 31, No. 1, pp. 47–52, 1984.

32. J. C. Lin and P. Bernardi, Computer methods for predicting field intensity and temperature change, in *Handbook of Biological Effects of Electromagnetic Fields*, 3rd ed., F. Barnes and B. Greenebaum (Eds.), CRC Press, Boca Raton, FL, 2007, pp. 293–380.

33. S. Pisa, M. Cavagnaro, E. Piuzzi, P. Bernardi, and J. C. Lin, Power density and temperature distributions produced by interstitial arrays of sleeved-slot antennas for hyperthermic cancer therapy, *IEEE Trans. Microwave Theory Tech.*, Vol. 51, No. 12, pp. 2418–2426, 2003.

34. M. J. Piket-May, A. Taflove, W. C. Lin, D. S. Katz, V. Sathiaselan, and B. B. Mittal, Initial results for automated computational modelling of patient-specific electromagnetic hyperthermia, *IEEE Trans. Biomed. Eng.*, Vol. 39, No. 3, pp. 226–237, 1992.

35. D. J. Lee, R. Mayer, and L. Hallinan, Outpatient interstitial thermoradiotherapy, *Cancer*, Vol. 77, pp. 2363–2370, 1996.

36. T. Nakajima, D. W. Roberts, T. P. Ryan, P. J. Hoopes, C. T. Coughlin, B. S. Trembly, and J. W. Strohbehn, Pattern of response to interstitial hyperthermia and brachytherapy for malignant intracranial tumour: a CT analysis, *Int. J. Hyperthermia*, Vol. 9, pp. 491–502, 1993.

37. J. C. Lin, Biophysics of radiofrequency ablation, in *Radiofrequency Catheter Ablation of Cardiac Arrhythmias: Basic Concepts and Clinical Applications*, 2nd ed., S. K. S. Huang and D. J. Wilber (Eds.), Futura, Armonk, NY, 2000, pp. 13–24.

38. J. C. Lin, Y. L. Wang, and R. J. Heriman, Comparison of power deposition patterns produced by microwave and radio frequency cardiac ablation catheters, *Electron. Lett.*, Vol. 30, No. 12, pp. 922–923, 1994.
39. T. L. Wonnel, P. R. Stauffer, and J. J. Langberg, Evaluation of microwave and radio-frequency catheter ablation in a myocardium-equivalent phantom model, *IEEE Trans. Biomed. Eng.*, Vol. 39, No. 10, pp. 1086–1095, 1992.
40. J. W. Strohbehn, B. S. Trembly, E. B. Douple, and D. C. de Sieyes, Evaluation of an invasive microwave antenna system for heating deep-seated tumor, *J. Nat. Cancer Inst. Monogr.*, Vol. 61, pp. 489–491, 1982.
41. P. F. Turner, Interstitial equal-phased arrays for EM hyperthermia, *IEEE Trans. Microwave Theory Tech.*, Vol. 34, pp. 572–578, 1986.
42. T. Z. Wong, J. W. Strohbehn, K. M. Jones, J. A. Mechling, and B. S. Trembly, SAR patterns from an interstitial microwave antenna-array hyperthermia system, *IEEE Trans. Microwave Theory Tech.*, Vol. 34, No. 5, pp. 560–567, 1986.
43. J. C. Lin and Y. J. Wang, Interstitial microwave antennas for thermal therapy, *Int. J. Hyperthermia*, Vol. 3, pp. 37–47, 1987.
44. G. Schaller, J. Erb, and R. Engelbrecht, Field simulation of dipole antennas for interstitial microwave hyperthermia, *IEEE Trans, Microwave Theory Tech.*, Vol. 34, No. 6, pp. 887–895, 1996.
45. Y. Wang and J. C. Lin, A comparison of microwave interstitial antennas for hyperthermia, *Proc. IEEE Eng. Med. Biol. Conf.*, pp. 1463–1466, 1986.
46. J. C. Camart, J. J. Fabre, B. Prevost, J. Pribetich, and M. Chive, Coaxial antenna array for 915 MHz interstitial hyperthermia: design and modelization—power deposition and heating pattern—phased array, *IEEE Trans. Microwave Theory Tech.*, Vol. 40, No. 12, pp. 2243–2250, 1992.
47. J. C. Camart, D. Despretz, M. Chive, and J. Pribetich, Modeling of various kinds of applicators used for microwave hyperthermia based on the FDTD method, *IEEE Trans. Microwave Theory Tech.*, Vol. 44, No. 10, pp. 1811–1818, 1996.
48. S. Labonté, A. Blais, S. R. Legault, H. O. Ali, and L. Roy, Monopole antennas for microwave catheter ablation, *IEEE Trans. Microwave Theory Tech.*, Vol. 44, No. 10, pp. 1832–1840, 1996.
49. Z. Kaouk, A. Khebir, and P. Savard, A finite element model of a microwave catheter for cardiac ablation, *IEEE Trans. Microwave Theory Tech.*, Vol. 44, No. 10, pp. 1848–1854, 1996.
50. M. F. Iskander and A. M. Tumeh, Design optimization of interstitial antennas, *IEEE Trans. Biomed. Eng.*, Vol. 36, No. 2, pp. 238–246, 1989.
51. W. Hurter, F. Reinbold, and W. J. Lorenz, A dipole antenna for interstitial microwave hyperthermia, *IEEE Trans. Microwave Theory Tech.*, Vol. 39, No. 6, pp. 1048–1054, 1991.
52. K. M. Jones, J. A. Mechling, B. S. Trembly, and J. W. Strohbehn, SAR distributions for 915 MHz interstitial microwave antennas used in hyperthermia for cancer therapy, *IEEE Trans. Biomed. Eng.*, Vol. 35, No. 10, pp. 851–857, 1988.
53. R. D. Nevels, G. D. Arndt, G. W. Raffoul, J. R. Carl, and A. Pacifico, Microwave catheter design, *IEEE Trans. Biomed. Eng.*, Vol. 45, No. 7, pp. 885–890, 1998.
54. J. C. Lin and Y. J. Wang, An implantable microwave antenna for interstitial hyperthermia, *Proc. IEEE*, Vol. 75, No. 8, pp. 1132–1133, 1987.
55. J. C. Lin, Catheter microwave ablation therapy for cardiac arrhythmias, *Bioelectromagnetics*, Vol. 20, pp. 120–132, 1999.
56. L. K. Wu and D. W. F. Su, Design and analysis of an asymmetrically fed insulated coaxial slot antenna with enhanced tip-heating performance, *IEEE Trans. Microwave Theory Tech.*, Vol. 48, No. 8, pp. 1369–1374, 2000.

57. H. M. Chiu, A. Sanagavarapu Mohan, A. R. Weily, D. J. R. Guy, and D. L. Ros, Analysis of a novel expanded tip wire (ETW) antenna for microwave ablation of cardiac arrhythmias, *IEEE Trans. Biomed. Eng.*, Vol. 50, No. 7, pp. 890–899, 2003.

58. I. Longo, G. Biffi Gentili, M. Cerretelli, and N. Tosoratti, A coaxial antenna with miniaturized choke for minimally invasive interstitial heating, *IEEE Trans. Biomed. Eng.*, Vol. 50, No. 1, pp. 82–88, 2003.

59. C. L. Brace, P. F. Laeseke, D. W. van der Weide, and F. T. Lee, Jr., Microwave ablation with a triaxial antenna: results in *ex vivo* bovine liver, *IEEE Trans. Microwave Theory Tech.*, Vol. 53, No. 1, pp. 215–220, 2005.

60. J. C. Lin and Y. J. Wang, The cap-choke catheter antenna for microwave ablation treatment, *IEEE Trans. Biomed. Eng.*, Vol. 43, No. 6, pp. 657–660, 1996.

61. S. Pisa, M. Cavagnaro, P. Bernardi, and J. C. Lin, A 915-MHz antenna for microwave thermal ablation treatment: physical design, computer modeling and experimental measurement, *IEEE Trans. Biomed. Eng.*, Vol. 48, No. 5, pp. 599–601, 2001.

62. H. R. Ahn and K. Lee, Interstitial antennas tipped with reactive load, *IEEE Microwave Wireless Components Lett.*, Vol. 15, No. 2, pp. 83–85, 2005.

63. R. W. P. King, B. S. Trembly, and J. W. Strohbehn, The electromagnetic field of an insulated antenna in a conducting or dielectric medium, *IEEE Trans. Microwave Theory Tech.*, Vol. 31 No. 7, pp. 574–583, 1983.

64. J. C. Lin and Y. J. Wang, A catheter antenna for percutaneous microwave therapy, *Microwave Optical Technol. Lett.*, Vol. 8, No. 2, pp. 70–72, 1995.

65. K. Saito, H. Yoshimura, K. Ito, Y. Aoyagi, and H. Horita, Clinical trials of interstitial microwave hyperthermia by use of coaxial-slot antenna with two slots, *IEEE Trans. Microwave Theory Tech.*, Vol. 52, No. 8, pp. 1987–1991, 2004.

66. S. A. Shock, K. Meredith, T. F. Warner, L. A. Sampson, A. S. Wright, T. C. Winter III, D. M. Mahvi, J. P. Fine, and F. R. Lee, Jr., Microwave ablation with loop antenna: *in vivo* porcine liver model, *Radiology*, Vol. 231, No. 1, pp. 143–149, 2004.

67. Z. Gu, C. M. Rappaport, P. J. Wang, and B. A. VanderBrink, A $2\frac{1}{4}$-turn spiral antenna for catheter cardiac ablation, *IEEE Trans. Biomed. Eng.*, Vol. 46, No. 12, pp. 1480–1482, 1999.

68. Z. Gu, M. Rappaport, P. J. Wang, and B. A. VanderBrink, Development and experimental verification of the wide-aperture catheter-based microwave cardiac ablation antenna, *IEEE Trans. Microwave Theory Tech.*, Vol. 48, No. 11, pp. 1892–1900, 2000.

69. M. S. Mirotznik, N. Engheta, K. R. Foster, Heating characteristics of thin helical antennas with conducting cores in a lossy medium—I: Noninsulated antennas, *IEEE Trans. Microwave Theory Tech.*, Vol. 41, No. 11, pp. 1878–1886, 1993.

70. S. S. Hsu, L. Hoh, R. M. Rosenbaum, A. Rosen, P. Walinsky, and A. J. Greenspon, A method for the *in vitro* testing of cardiac ablation catheters, *IEEE Trans. Microwave Theory Tech.*, Vol. 44, No. 10, pp. 1841–1847, 1996.

71. I. B. Feldberg and N. J. Cronin, A 9.2 GHz microwave applicator for the treatment of menorrhagia, in *IEEE MTT-S International Microwave Symposium Digest*; Baltimore, MD, 7–12 June 1998, pp. 755–758.

72. K. L. Clibbon, A. McCowen, and J. W. Hand, SAR distributions in interstitial microwave antenna arrays with a single dipole displacement, *IEEE Trans. Biomed. Eng.*, Vol. 40, No. 9, pp. 925–932, 1993.

73. J. C. Lin, S. Hirai, C. L. Chiang, W. L. Hsu, J. L. Su, and Y. J. Wang, Computer simulation and experimental studies of SAR distributions of interstitial arrays of sleeved-slot microwave antennas for hyperthermia treatment of brain tumors, *IEEE Trans. Microwave Theory Tech.*, Vol. 48, No. 11, pp. 2191–2197, 2000.

74. K. M. Jones, J. A. Mechling, J. W. Strohbehn, and B. S. Trembly, Theoretical and experimental SAR distributions for interstitial dipole antenna arrays used in hyperthermia, *IEEE Trans. Microwave Theory Tech.*, Vol. 37, No. 8, pp. 1200–1209, 1989.
75. T. T. Wu, R. W. P. King, and D. V. Giri, The insulated dipole antenna in a relatively dense medium, *Radio Sci.*, Vol. 8, No. 7, pp. 699–709, 1973.
76. R. W. P. King, K. M. Lee, S. R. Mishra, and G. S. Smith, Insulated linear antenna: theory and experiment, *J. Appl. Phys.*, Vol. 45, No. 4, pp. 1688–1697, 1974.
77. K. M. Lee, and R. W. P. King, The terminated insulated antenna, *Radio Sci.*, Vol. 11, No. 4, pp. 367–373, 1976.
78. D. W. F. Su and L. K. Wu, Input impedance characteristics of coaxial slot antennas for interstitial microwave hypertermia, *IEEE Trans. Microwave Theory Tech.*, Vol. 47, No. 3, pp. 302–307, 1999.
79. S. Labonté, Numerical model for radio-frequency ablation of the endocardium and its experimental validation, *IEEE Trans. Biomed. Eng.*, Vol. 41, No. 2, pp. 108–115, 1994.
80. S. Labonté, A computer simulation of radio-frequency ablation of the endocardium, *IEEE Trans. Biomed. Eng.*, Vol. 41, No. 9, pp. 883–890, 1994.
81. P. Bernardi, M. Cavagnaro, J. C. Lin, S. Pisa, and E. Piuzzi, Distribution of SAR and temperature elevation induced in a phantom by a microwave cardiac ablation catheter, *IEEE Trans. Microwave Theory Tech*. (*Special Issue on Medical Applications and Biological Effects of RF/Microwaves*, Vol. 52, No. 8, pp. 1978–1986, 2004.
82. K. S. Yee, Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media, *IEEE Trans. Antennas Propag.*, Vol. 14, No. 3, pp. 302–307, 1966.
83. J. P. Berenger, A perfectly matched layer for the absorption of electromagnetic waves, *J. Comput. Phys.*, Vol. 114, pp. 185–200, 1994.
84. A. Taflove and S. C. Hagness, *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, Artech House, Boston, 2000.
85. C. Gabriel, Compilation of the dielectric properties of body tissues at RF and microwave frequencies, *Tech. Rep. AL/OE-TR-1996-0037*, Brooks Air Force Base, Brooks AFB, TX, 1996.
86. H. H. Pennes, Analysis of tissue and arterial blood temperatures in resting forearm, *J. Appl. Physiol.*, Vol. 1, pp. 93–122, 1948.
87. P. Bernardi, M. Cavagnaro, S. Pisa, and E. Piuzzi, SAR distribution and temperature increase in an anatomical model of the human eye exposed to the field radiated by the user antenna in a wireless LAN, *IEEE Trans. Microwave Theory Tech.*, Vol. 46, pp. 2074–2082, 1998.
88. J. J. W. Lagendijk, A mathematical model to calculate temperature distributions in human and rabbit eyes during hyperthermic treatment, *Phys. Med. Biol.*, Vol. 27, No. 11, pp. 1301–1311, 1982.
89. J. A. Scott, A finite element model of heat transport in the human eye, *Phys. Med. Biol.*, Vol. 33, No. 2, pp. 227–241, 1988.
90. F. A. Duck, *Physical Properties of Tissue*, Academic Press, San Diego, CA, 1990.
91. L. R. Williams and R. W. Leggett, Reference values for resting blood flow to organs of man, *Phys. Med. Biol.*, Vol. 10, No. 3, pp. 187–217, 1989.
92. P. M. Van Den Berg, A. T. De Hoop, A. Segal, and N. Praagman, A computational model of the electromagnetic heating of biological tissue with application to hyperthermic cancer therapy, *IEEE Trans. Biomed. Eng.*, Vol. 30, pp. 797–805, 1983.
93. R. G. Gordon, R. B. Roemer, and S. M. Horvath, A mathematical model of the human temperature regulatory system—transient cold exposure response, *IEEE Trans. Biomed. Eng.*, Vol. 23, pp. 434–444, 1976.

94. P. Bernardi, M. Cavagnaro, S. Pisa, and E. Piuzzi, Specific absorption rate and temperature elevation in a subject exposed in the far-field of radio-frequency sources operating in the 10–900-MHz range, *IEEE Trans. Biomed. Eng.*, Vol. 50, No. 3, pp. 295–304, 2003.

95. J. C. Lin, Microwave sensing of physiological movement and volume change: a review. *Bioelectromagnetics*, Vol. 13, pp. 557–565, 1992.

96. J. C. Lin, Biomedical applications of electromagnetic engineering, in *Engineering Electromagnetics: Applications*, Vol. 7, R. Bansal, (Ed.), CRC Press, Boca Raton, FL, 2006, pp. 211–233.

97. T. E. McEwan, Ultra-wideband radar motion sensor. US Patent 5,361,070 dated November 1, 1994.

98. J. D. Taylor, *Introduction to Ultra-Wideband Radar Systems*, CRC Press, Boca Raton, FL, 1995.

99. J. C. Lin, Microwave propagation in biological dielectrics with application to cardiopulmonary interrogation, in *Medical Applications of Microwave Imaging*, L. E. Larsen and J. H. Jacobi (Eds.), IEEE Press, New York, 1986, pp. 47–58.

100. J. C. Lin, Studies on microwaves in medicine and biology: from snails to humans, *Bioelectromagnetics*, Vol. 25, pp. 146–159, 2004.

101. K. M. Chen, D. Misra, H. Wang, H. R. Chuang, and E. Postow, An X-band microwave life-detection system, *IEEE Trans. Biomed. Eng.*, Vol. 33, pp. 697–701, 1986.

102. K. H. Chan and J. C. Lin, Microprocessor based cardiopulmonary rate monitor, *Med. Biol. Eng. Computing*, Vol. 25, pp. 41–44, 1987.

103. J. C. Lin, J. Kiernicki, M. Kiernicki, and P. B. Wollschlaeger, Microwave apexcardiography, *IEEE Trans. Microwave Theory Tech.*, Vol. 27, pp. 618–620, 1979.

104. J. Y. Lee and J. C. Lin, A microprocessor based non-invasive pulse wave analyzer, *IEEE Trans. Biomed. Eng.*, Vol. 32, pp. 451–455, 1985.

105. M. A. Papp, C. Hughes, J. C. Lin, and J. M. Pouget, Doppler microwave: a clinical assessment of its efficacy as an arterial pulse sensing technique, *Invest. Radiol.*, Vol. 22, pp. 569–573, 1987.

106. A. Thansandote, S. S. Stuchly, and A. M. Smith, Monitoring variations of biological impedances using microwave Doppler radar, *Phys. Med. Biol.*, Vol. 28, pp. 983–990, 1983.

107. J. C. Lin and M. J. Clarke, Microwave imaging of cerebral edema, *Proc. IEEE*, Vol. 70, pp. 523–524, 1982.

108. M. J. Clarke and J. C Lin, Microwave sensing of increased intracranial water content, *Invest. Radiol.*, Vol. 18, pp. 245–248, 1983.

109. F. Holzrichter, G. C. Burnett, L. C. Ng, and W. A. Lea, Speech articulator measurements using low power EM-wave sensors, *J. Acoust. Soc. Am.*, Vol. 103, pp. 622–625, 1998.

110. E. C. Fear, S. C. Hagness, P. M. Meaney, M. Okoniewski, and M. A. Stuchly, Enhancing breast tumor detection with near field imaging, *IEEE Microwave Mag.*, Vol. 3, pp. 48–56, 2002.

111. J. C. Lin, Microwave thermoelastic tomography and imaging, in *Advances in Electromagnetic Fields in Living Systems*, Vol. 4, Springer, New York, 2005, pp. 41–76.

112. L. E. Larsen and J. H. Jacobi, Microwave scattering parameter imagery of an isolated canine kidney, *Med. Phys.*, 6, pp. 394–403, 1979.

113. J. L. Guerquin-Kern, M. Gautherie, G. Peronnet, L. Jofre, and J. C. Bolomey, Active microwave tomographic imaging of isolated, perfused animal organs, *Bioelectromagnetics*, Vol. 6, pp. 145–156, 1985.

114. P. M. Meaney, K. D. Paulsen, A. Hartov, and R. K. Crane, Microwave imaging for tissue assessment: initial evaluation in multitarget tissue-equivalent phantoms, *IEEE Trans. Biomed. Eng.*, Vol. 43, pp. 878–890, 1996.

115. S. Y. Semenov, R. H. Svenson, A. E. Boulyshev, A. E. Souvorov, V. Y. Borisov, Y. Sizov, A. N. Starostin, K. R. Dezern, G. P. Tatsis, and V. Y. Baranov, Microwave tomography—two-dimensional system for biological imaging, *IEEE Trans. Biomed. Eng.*, Vol. 43, pp. 869–877, 1996.

116. A. Franchois, A. Joisel, C. Pichot, and J. C. Bolomey, Quantitative microwave imaging with a 2.45-GHz planar microwave camera, *IEEE Trans. Med. Imaging*, Vol. 17, pp. 550–561, 1998.

117. S. Y. Semenov, R. H. Svenson, A. E. Bulyshev, A. E. Souvorov, A. G. Nazarov, Y. E. Sizov, V. G. Posukh, A. Pavlovsky, P. N. Repin, A. N. Starostin, B. A. Voinov, M. Taran, G. P. Tatsis, and V. Y. Baranov, Three-dimensional microwave tomography: initial experimental imaging of animals, *IEEE Trans. Biomed. Eng.*, Vol. 49, pp. 55–63, 2002.

118. R. G. Olsen and J. C. Lin, Acoustical imaging of a model of a human hand using pulsed microwave irradiation, *Bioelectromagnetics*, Vol. 4, pp. 397–400, 1983.

119. J. C. Lin and K. H. Chan, Microwave thermoelastic tissue imaging-system design, *IEEE Trans. Microwave Theory Tech.*, Vol. 32, pp. 854–860, 1984.

120. R. A. Kruger, D. R. Reinecke, and G. A. Kruger, Thermoacoustic computed tomography—technical considerations, *Med. Phys.*, Vol. 26, pp. 1832–1837, 1999.

121. R. A. Kruger, K. D. Miller, H. E. Reynolds, W. L. Kiser, D. R. Reinecke, and G. A. Kruger, Breast cancer *in vivo*: contrast enhancement with thermoacoustic CT at 434 MHz —feasibility study, *Radiology*, Vol. 216, pp. 279–283, 2000.

122. J. D. Kraus and R. J. Marhefka, *Antennas*, McGraw-Hill, Boston, 2002.

123. E. C. Jordan and K. G. Balmain, *Electromagnetic Waves and Radiating Systems*, Prentice-Hall, Englewood Cliffs, NJ, 1968.

124. S. C. Hagness, A. Taflove, and J. E. Bridges, Three-dimensional FDTD analysis of pulsed microwave confocal system for breast cancer detection: design of an antenna-array element, *IEEE Trans. Antennas Propag.*, Vol. 47, pp. 783–791, 1999.

125. T. T. Wu and R. W. P. King, The cylindrical antenna with non-reflecting resistive loading, *IEEE Trans. Antennas Propag.*, Vol. 13, pp. 369–373, 1965.

126. M. Kanda, Time domain sensors for radiated impulsive measurements, *IEEE Trans. Antennas Propag.*, Vol. 31, pp. 438–444, 1983.

127. K. Esselle and S. S. Stuchly, Pulse receiving characteristics of resistively loaded dipole antennas, *IEEE Trans. Antennas Propag.*, Vol. 38, pp. 1677–1683, 1990.

128. T. Montoya and G. Smith, A study of pulse radiation from several broad-band loaded monopoles, *IEEE Trans. Antennas Propag.*, Vol. 44, pp. 1172–1182, 1996.

129. K. L. Shlager, G.S Smith, and J. G. Maloney, Optimization of bow-tie antennas for pulse radiation, *IEEE Trans. Antennas Propag.*, Vol. 42, pp. 975–982, 1994.

130. J. Maloney and G. Smith. A study of transient radiation from the Wu–King monopole—FDTD analysis and experimental measurements, *IEEE Trans. Antennas Propag.*, Vol. 41, pp. 668–676, 1993.

131. L. C. Shen and R. W. P. King, Corrections to: The cylindrical antenna with nonreflective resistive loading, *IEEE Trans. Antennas Propag.*, Vol. 13, pp. 998, 1965.

132. M. Kanda, A relatively short broadband antenna with tapered resistive loading for picosecond pulse measurements. *IEEE Trans. Ant. Propag.*, Vol. 26, pp. 439–447.

133. E. C. Fear and M. A. Stuchly, Microwave detection of breast cancer, *IEEE Trans. Microwave Theory Tech.*, Vol. 48, pp. 1854–1863, 2000.

134. K. Kim and W. R. Scott, Design of a resistively loaded vee dipole for ultrawide-band ground penetrating radar applications, *IEEE Trans. Antennas Propag.*, Vol. 53, pp. 2525–2532, 2005.

135. J. P. Lee, S. O. Park, and S. K. Lee, Bow-tie wide-band monopole antenna with the novel impedance-matching technique, *Mic. Opt. Tech. Lett.*, Vol. 33, pp. 448–452, 2002.

136. J. M. Johnson and Y. Rahmat-Samii, The tab monopole, *IEEE Trans. Antennas Propag.*, Vol. 45, pp. 187–188, 1997.

137. H. R. Chuang, C. C. Lin, and Y. C. Kan, A printed UWB triangular monopole antenna, *Microwave. J.*, pp. 108–120, 2006.

138. R. A. Kruger, K. M. Stantz, and W. L. Kiser, Jr., Thermoacoustic CT of the breast, *Proc. SPIE*, Vol. 4682, pp. 521–525, 2002.

139. C. E. Hayes, W. A. Edelstein, J. F. Schenck, O. M. Mueller, and M. Eash, An efficient, highly homogeneous radiofrequency coil for whole-body NMR imaging at 1.5 T, *J. Magn. Reson.*, Vol. 63, pp. 622–628, 1985.

140. J. Tropp, The theory of the bird-cage resonator, *J. Magn. Reson.*, Vol. 82, pp. 51–62, 1989.

141. M. C. Leifer, Resonant modes of the birdcage coil, *J. Magn. Reson.*, Vol. 124, pp. 51–60, 1997.

142. J. Jin, *Electromagnetic Analysis and Design in Magnetic Resonance Imaging*, CRC Press, Boca Raton, FL, 1999.

143. G. Giovannetti, L. Landini, M. F. Santarelli, and V. Postano, A fast and accurate simulator for the design of birdcage coils in MRI, *Magn. Reson. Mat. Phys. Biol. Med.*, Vol. 15, pp. 36–40, 2002.

144. P. Röschmann, High-frequency coil system for a magnetic resonance imaging apparatus, U. S. Patent 4 746 866, 1988.

145. J. F. Bridges, Cavity resonator with improved magnetic field uniformity for high frequency operation and reduced dielectric heating in NMR imaging devices, U. S. Patent 4 751 464, 1988.

146. P. Röschmann, Radiofrequency penetration and absorption in the human body: limitations to high-field whole-body nuclear magnetic resonance imaging, *Med. Phys.*, Vol. 14, pp. 922–932, 1987.

147. B. A. Baertlein, O. Ozbay, T. Ibrahim, R. Lee, Y. Yu, A. Kangarlu, and P. M. L. Robitaille, Theoretical model for an MRI radio frequency resonator, *IEEE Trans. Biomed. Eng.*, Vol. 47, pp. 535–546, 2000.

148. G. Bogdanov and R. Ludwig, Coupled microstrip line transverse electromagnetic resonator model for high-field magnetic resonance imaging, *Magn. Res. Med.*, Vol. 47, pp. 579–593, 2002.

149. J. T. Vaughan, H. P. Hetherington, J. O. Otu, J. W. Pan, and G. M. Pohost, High frequency volume coils for clinical NMR imaging and spectroscopy, *Magn. Reson. Med.*, Vol. 32, pp. 206–218, 1994.

150. G. C. Chingas and N. Zhang, Design strategy for TEM high field resonators, in *Proceedings of the 4th Annual Meeting of the International Society of Magnetic Resonance in Medicine*, New York, 1996, p. 1426.

151. J. Tropp, Mutual inductance in the bird cage resonator, *J. Magn. Reson.*, Vol. 126, pp. 9–17, 1997.

152. C. M. Collins, S. Li, and M. B. Smith, SAR and B_1 field distributions in a heterogeneous human head model within a birdcage coil, *Magn. Res. Med.*, Vol. 40, pp. 1–10, 1998.

153. O. P. Gandhi and X. B. Chen, Specific absorption rates and induced current densities for an anatomy-based model of the human for exposure to time varying magnetic fields of MRI, *Magn. Res. Med.*, Vol. 41, pp. 816–823, 1999.

154. T. S. Ibramhim, A. M. Abduljalil, B. A. Baertlein, R. Lee, and P. M. L. Robitaille, Analysis of B_1 field profiles and SAR values for multi-strut transverse electromagnetic RF coils in high field MRI applications, *Phys. Med. Biol.*, Vol. 46, pp. 2545–2555, 2001.

155. Z. W. Wang, J. C. Lin, W. H. Mao, W. Z. Liu, M. B. Smith, and C. M. Collins, SAR and temperature: simulations and comparison to regulatory limits for MRI, *J. Magn. Resonance Imaging*, Vol. 26, pp. 437–441, 2007.

156. J. W. Hand, R. W. Lau, J. Lagendijk, J. Ling, M. Burl, and I. R. Young, Electromagnetic and thermal modeling of SAR and temperature fields in tissue due to an RF decoupling coil, *Magn. Res. Med.*, Vol. 42, pp. 183–192, 1999.

157. U. D. Nguyen, S. Brown, I. A. Chang, J. K. Krycia, and M. S. Mirotznik, Numerical evaluation of heating of the human head due to magnetic resonance imaging, *IEEE Trans. Biomed. Eng.*, Vol. 51, pp. 1301–1309, 2004.

158. ICNIRP, Statement on medical magnetic resonance (MR) procedures: protection of patients, *Health Phys.*, Vol. 87, pp. 197–216, 2004.

159. T. Sommer, C. Vahlhaus, G. Lauck, A. Von Smekal, M. Reinke, U. Hofer, W. Bloch, F. Traber, C. Schneider, J. Gieseke, W. Jung, and H. Schild, MR imaging and cardiac pacemaker: *in-vitro* evaluation and *in-vivo* studies in 51 patients at 0.5 T, *Radiology*, Vol. 215, pp. 869–879, 2000.

160. S. Pisa, G. Calcagnini, M. Cavagnaro, E. Piuzzi, M. Triventi, and P. Bernardi, SAR and temperature elevations in pacemaker holders exposed to EM fields produced by MRI apparatus, in *2006 IEEE MTT-S International Microwave Symposium Digest*, San Francisco, pp. 1754–1757.

161. T. Montoya and G. Smith (1996). Resistively-loaded vee antennas for short-pulse ground penetrating radar. *Proc. IEEE Ant and Propag Symp.*, Baltimore, pp. 2068–2071.

CHAPTER 28

Antennas for Biological Experiments

JAMES C. LIN, PAOLO BERNARDI, STEFANO PISA, MARTA CAVAGNARO, and EMANUELE PIUZZI

28.1 INTRODUCTION

The rapid development of wireless radio communication systems during the past decade, with the resulting widespread use of cellular mobile telephones, has raised public concerns about possible adverse health effects resulting from exposure of the human body and especially the human head to the electromagnetic (EM) field radiated by such devices. Indeed, with the advent of cellular mobile telephones, a radiofrequency (RF) source is placed close to the human head or in direct contact with the human body for the first time in human history. In addition, power lines for transmission of electrical power have been in use around the world long before cellular communication systems were introduced. They operate in the 50–60-Hz range of extremely low frequencies (ELFs) and produce EM fields along their rights-of-way, exposing the population in the path of transmission. The potential adverse health effects of ELF fields have been a subject of public concern and scientific investigation for more than two decades. Indeed, one focus of the interdisciplinary field of bioelectromagnetics has been the study of interactions of EM fields with biological systems and possible health effects of EM fields.

The methodologies employed in bioelectromagnetics to study the interaction of EM fields with biological systems and possible health effects of EM fields whether at extremely low frequencies or radiofrequencies can be divided into three categories: *in vitro* biological experiments, *in vivo* experiments, or epidemiological studies. *In vitro* biological experiments usually involve biological entities constituted by cells contained within flasks or Petri dishes and are exposed to a well-defined EM field. These experiments are most suited to study the possible effects of exposure on specific biological targets or to study postulates and verify proposed interaction mechanisms aimed toward explaining some observed biological effects. However, the implications of any effect observed at the cellular level with *in vitro* studies are not always obvious in terms of health effects on whole organisms. Thus it becomes necessary to conduct *in vivo* experiments, where whole animals such as mice and rats are directly exposed to EM fields and the potential for induction of specific health effects is investigated. While *in vivo* experiments are important to assess the existence of possible health effects and the thresholds for their

Modern Antenna Handbook. Edited by Constantine A. Balanis

induction, they are also useful in allowing extrapolation of animal observations to human subjects. Epidemiological studies can offer the most direct evidence on the health effects from human exposure to EM fields. However, a major limitation of epidemiological studies, apart the exposure assessment of the EM field (i.e., the quantification of exposure levels), which is rather difficult if not impossible to obtain with an acceptable level of accuracy, is the typical prolonged period of time required for observation in most cases.

Unlike epidemiological studies, where investigated subjects are exposed daily to real sources encountered in everyday life, both *in vitro* and *in vivo* biological experiments require the design and realization of appropriate systems allowing exposure, under controlled laboratory conditions and with reproducible exposure and specific absorption rate (SAR) or induced field levels of the biological target. Whether it is for a cell culture or a whole animal, the complex systems are generally referred to as exposure systems. Antennas used for biological experiments are a part of exposure systems whose aim is to produce a field distribution inside the system suitable for the exposure of the biological target at the required levels and with a given uniformity over the whole target or a part of it. Therefore these antennas must generally produce well-defined field distributions over a limited and confined region of space where the targets will be placed and, as such, they are often quite different from a conventional antenna, designed to radiate EM fields in free space.

Of course the antenna, on which the present chapter is focused, is but one piece of the whole exposure system. Indeed, besides the generation of the EM field and the production of the desired SAR or fields within the biological subject, the exposure system is also required to maintain physiological and environmental conditions appropriate for the experimental preparation under investigation and allow, when necessary, the monitoring of biological parameters of interest during exposure (e.g., access for visual observation by a microscope). Therefore it also consists of devices for maintaining the optimal environmental conditions, such as incubators and ventilation systems, and sensors for monitoring various physiological parameters. Moreover, appropriate holders or retention mechanisms for the precise positioning of the biological subject within the exposure field must also be included. Figure 28.1 shows two examples of complete exposure systems: Figure 28.1a is an *in vitro* system based on rectangular waveguides for exposure of cell cultures contained inside Petri dishes, and Figure 28.1b is an *in vivo* system using loop antennas for localized head exposure of mice or rats.

The rest of this chapter proceeds with a brief description of the organization and approaches for interdisciplinary research in bioelectromagnetics and provides a summary of the main features characterizing an exposure system, including a definition of the dosimetric quantities used to evaluate the performances of an antenna in producing the desired field distributions within the biological body. It is followed by a discussion of antennas used in recent *in vitro* and *in vivo* experiments. The antennas are divided according to their intended application in antennas for *in vitro* experiments and antennas for *in vivo* studies. The descriptions are further subdivided into categories of antennas for whole-body and localized exposures. For each antenna a description of its main characteristics and a few examples to illustrate types of applications are given. Where possible, some general design rules are also provided. It is noted that, for the most part, these antennas are designed and optimized through a trial-and-error approach making use of numerical simulations and only in some rare cases analytical design formulas exist. It must also be emphasized that the operation of the overall exposure system poses many different, and often contrasting, constraints on the design of the antenna itself.

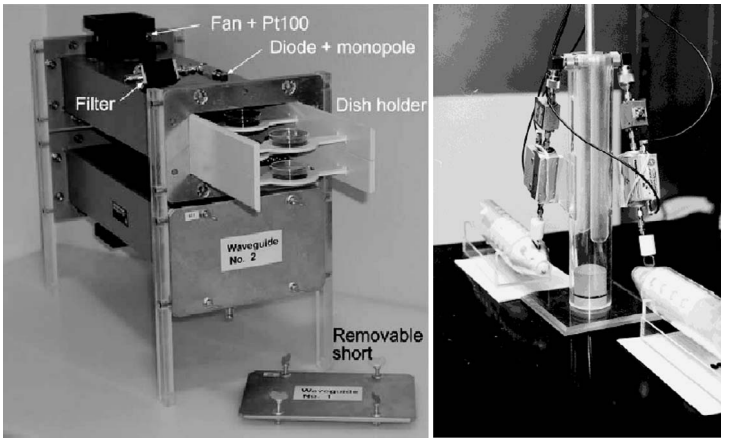


Figure 28.1 Example of exposure systems: (a) *in vitro* system (from Ref. 1) and (b) *in vivo* system (from Ref. 2).

For example, the incubator dimensions limit the maximum size of the antenna, while the subject dimensions limit its minimum size. Moreover, electromagnetic compatibility (EMC) problems pose a limit on the maximum leakage field. This means that general rules for antenna design cannot be given without exact knowledge of the global specifications of the overall exposure system where they will be used, the protocols planned, and the specific biological preparation involved.

It is worth mentioning that an antenna for RF biological experiments, aiming at the production of a prescribed induced field or SAR within a biological subject, essentially possesses the same goals as antennas for hyperthermia treatment of cancer (see Chapter 27), and therefore, especially in the past, external hyperthermia applicators have often been used as antennas for biological experiments. However, nowadays, it is well recognized that, in order to produce meaningful and reproducible results, special care must be taken to properly characterize the exposure system from the dosimetric point of view. Therefore most recent biological experiments make use of specifically designed and well-characterized exposure systems, employing antennas specifically tailored for the particular application.

Obviously, the principal application of materials discussed in this chapter is in bioelectromagnetics research, where the interaction of EM fields with biological systems and possible health effects of EM fields are investigated using antennas and exposure systems. The following section provides a discussion on organizing interdisciplinary bioelectromagnetics research.

28.2 ORGANIZED INTERDISCIPLINARY BIOELECTROMAGNETICS RESEARCH

Research investigations in bioelectromagnetics are interdisciplinary and complex. A well-organized and expertly executed multidisciplinary program is often necessary to

study interactions of EM fields with biological systems and to provide meaningful results and help advance our understanding of the health effects.

28.2.1 Organization of Research

Research programs may be organized in various ways, depending on their type and level of complexity. There are two major categories of organized research: exploratory and mission-oriented. The strategies through which the research may be organized and supported are distinct. While exploratory research may be hypothesis driven, its objective, in general, is to gain knowledge and learn about the interaction of various forces in natural or under artificial environments. Mission-oriented research has a specific goal, such as landing on Mars or finding cure for a particular disease. Most sponsored research to ascertain if there is any linkage between the EM fields from mobile phones or power lines and their health effects, to date, falls under the category of mission-oriented research.

The key steps in mission-oriented research or finding are (1) definition of issues, (2) identification a research agenda, (3) dissemination of the agenda, (4) call for proposals, (5) selection of projects, (6) execution of studies, and (7) reporting of results. Each step is crucial for successful completion of the research. A leadership style and management to ensure proper execution of each step are fundamental for success. Clearly, deficiencies or blunders in steps 1 to 4 would preordain failure. However, faults or oversights in steps 5 to 7 could reduce the value of the research.

A key component in enhancing the perceived credibility of sponsored research is the "arm's-length" or "fire-wall" arrangement, in which the sponsor defines issues, identifies the research agenda, disseminates the agenda, and provides the funding. This approach leaves the call for proposals, selection of projects, and program management through an independent organization, and the execution of projects and reporting of results to be done by project investigators, under an arm's-length arrangement. The fire wall is used for sponsors, to provide funding for the selected projects but to prevent their direct participation in the selection and management of research projects or determination of their outcome.

The public is often apprehensive and skeptical of research results and conclusions, if they suspect breaching of the arm's-length or fire-wall protocol. Likewise, if there is a lack of demonstrated transparency in the seven key steps outlined above, public perception and acceptance of the resulting research or conclusion would be checked and restrained.

It is appropriate for sponsors to conduct in-house research and/or contract the research to universities or other independent organizations. This strategy would serve the sponsor's need to know and could provide the sponsor with first-hand information on any linkage between the EM fields used for the product or service and their health effects. Obviously, the arm's-length and fire-wall arrangements do not apply under this scenario.

Given the widely circulated experiences involving research on cigarette smoking and leaded gasoline, however, the public has become very skeptical about results from such research. The recent record for public acceptance of any conclusions derived from such research has been dismal, because of perceived challenges in establishing any arm's-length or fire-wall type of arrangement. Nevertheless, "corporately conducted" research is important from a utilitarian perspective and has its place in the overall strategy of a given corporate entity.

28.2.2 Interdisciplinary Team Approach

It goes without saying that the research projects must be carried out with well-known scientific methodology, regardless of whether the research is exploratory or mission-oriented, or in-house or sponsored. However, research into the biological effects of EM fields, from wireless personal communication systems and power lines, is complicated by the interdisciplinary nature of the subject. It places requirements on investigators that are beyond their usual training in fields of particular disciplines.

The subtle interactions that take place between the EM fields and any observed biological response may easily elude the eyes of one trained in a specific field of study. An electrical engineer may not recognize complications imposed by the exposure of biological systems that are beyond ordinary electromagnetic interference phenomena. Likewise, an investigator trained in biological science, using well-established techniques, may not realize that the response may be an artifact from the exposure system. Thus the outcome and significance of an investigation may often depend on the knowledge and experience of the investigator or investigators in the specialized bioelectromagnetics field of study.

The interdisciplinary team approach is essential to the investigation of health effects of RF/microwave radiation from cellular-mobile-telephone operations and from power-line fields. It demands the collective efforts of investigators who are knowledgeable and experienced with specific training in this specialized field of study. This is not to suggest that new or experienced investigators from other fields should be excluded from conducting research in this area. (In fact, they may make valuable contributions to the investigation.) Instead, they could be a part of an interdisciplinary team, where there are individuals with expertise and experience in this specialized field.

It might be expedient to recruit a team of investigators with associated knowledge and experience. Given the complications of this field, the outcome would be uncertain at best. One consequence is research results that are of questionable value, creating more controversy. It may prolong the search in darkness instead of shining more light on the subject. The situation can become more tenuous if the team has a total lack of experience in research involving EM exposure of biological systems. Therefore, attention needs to be given to both experimental and management factors that could have independently provoked or prevented the observation of a biological response.

28.2.3 Need for Confirmation and Replication Studies

Unlike physical systems, where the behavior of the system, under most circumstances, especially in its designed range of operation is deterministic, the responses of biological systems and organisms are notoriously variable. Their behavior is often uncertain, even under similar circumstances of EM field exposure. A given observation must be replicated by the same investigators. Thus a large number of experiments would normally be required to establish a reliable cause-and-effect relationship between the exposure and biological response. If the number of experiments is limited, it would be difficult to detect a true association.

Because a single study or report seldom provides definitive evidence for or against a biological response, several independent studies are needed to arrive at a statistically significant association, or at a convincing answer to the health effect question. It should be noted that the scientific approach and acquisition of scientific knowledge are predicated on replication and confirmation by repeated experimentations. The problem of health

implications can only be met with the presence of a critical mass of scientists working on issues of the interaction of EM fields with biological systems. The uncertainties on the linkage between the wireless mobile communication systems and power lines and their health effects persist today, in part, because of the limited number and scope of replication or confirmation studies that have been conducted or completed.

28.3 INDUCED FIELDS AND SAR

As mentioned previously, the performances of antennas and exposure systems for biological experiments typically are evaluated on the basis of their ability to produce a prescribed induced field or SAR within a biological object. Indeed, SAR in biological systems or tissue models has been adopted as the dosimetric quantity of choice, especially at radiofrequencies. The metric SAR expressed in watts per kilogram (W/kg) is defined as the time rate of change of the incremental energy absorbed by or deposited in an incremental mass contained in a volume of a specific density. Knowledge of SAR or induced field is of interest because it may serve as an index for comparison and extrapolation of experimental results from tissue to tissue, from animal to animal, from animal to human, and from human to human exposures. It is also useful in analyzing relationships among various observed biological effects in different experimental models and subjects. This is in clear contrast to incident field or any other external measures of exposure, which often can produce different induced fields and SARs inside different biological systems, even for the same values of incident field. Also, the whole-body absorption of RF electromagnetic energy by humans and laboratory animals, as specified by specific absorption (SA), is of interest because it is related to the energy required to alter the thermoregulatory response of the exposed subject [3].

It should be noted that the quantity of induced field is the primary driving force underlying the interaction of electromagnetic energy with biological systems. The induced field in biological tissue is a function of body geometry, tissue permittivity and conductivity, and the exposure conditions. Moreover, a determination of the induced field is recommended because (1) it relates the field to specific responses of the body, (2) it facilitates understanding of biological phenomena, and (3) it is independent of any mechanism of interaction. Once the induced field is known, quantities such as SAR can be derived from it by a simple conversion formula. For example, from an induced electric field E (in V/m), the SAR can be derived as

$$SAR(x, y, z) = \frac{\sigma(x, y, z)|E(x, y, z)|^2_{\text{rms}}}{\rho(x, y, z)} \quad \text{(W/kg)} \tag{28.1}$$

where σ represents the conductivity (S/m), and ρ is the density of the tissue (kg/m^3).

However, in the absence of a small, isotropic, implantable electric-field sensor with sufficient sensitivity, a common practice in experimental dosimetry relies on the use of small field-insensitive temperature probes (such as fiberoptic sensors) to measure the temperature elevation produced under a short-duration (< 30 s), high intensity exposure condition (sufficiently high to produce an easily measurable tissue temperature elevation). The short duration is not enough for significant conductive or convective heat transfer to contribute to tissue temperature rise. In this case, the time rate of initial rise in temperature

(slope of transient temperature response curve) is linear and can be related to SAR through a secondary procedure:

$$SAR = c\frac{\Delta T}{\Delta t} \tag{28.2}$$

where ΔT is the temperature increment (°C), c is the specific heat capacity of the tissue (J/kg·°C), and Δt is the duration(s) over which ΔT is measured. Thus the rise in tissue temperature during the initial transient period of RF energy absorption is linearly proportional to the value of SAR. It is important to distinguish the use of SAR and its derivation from temperature measurement. The quantity of SAR is merely a metric for energy deposition or absorption and it should not be construed to imply any mechanism of interaction, thermal or otherwise. However, it is a quantity that pertains to a macroscopic phenomenon by virtue of the use of bulk electrical conductivity in its derivation (Eq. (28.1)) and the use of specific heat capacity in Eq. (28.2).

The definition of SAR and its relation to induced fields puts us now in a position to examine more closely antennas and exposure systems. We first discuss the salient features used to characterize exposure systems; indeed, these features, for the most part, are unique to this particular application. Discussions on antennas for biological experiments then follow.

28.4 CHARACTERISTIC FEATURES OF EXPOSURE SYSTEMS

The first important feature of an exposure system is the homogeneity of the induced field or SAR distribution for RF exposures or the EM field for ELF exposures within the target, which may be quantified by the standard deviation of the distribution.

Another essential but nonelectromagnetic feature is the volume efficiency, which is defined as the ratio between the subject volume and the space required for the entire exposure environment. Volume is an issue since if the volume efficiency is poor, a large space may be needed to accommodate the exposure system, which would increase the amount of effort needed for environmental control. It would be impractical for use with a small incubator. Among the more classical antenna parameters, one that must always be taken into account for RF antennas is the input impedance that must be achieved to ensure a good match with the power source. Finally, one more parameter characterizing the performance of an exposure system is its efficiency, which may be defined as the average SAR level produced within the target for 1 W of input power. A low efficiency leads to high power requirements for the amplifier, in order to reach the prescribed absorption. Since RF power amplifiers are very expensive, this can considerably increase the total system cost.

Although well-defined exposure conditions for biological experiments are an obvious and indispensable prerequisite for interpretation and repeatability of results, the difficulties involved in obtaining such conditions have been severely underestimated by many groups conducting RF experiments in the past [4]. Consequently, design and characterization of exposure systems have become a top priority for most research addressing the health effects of RF exposures. Indeed, the design and implementation of exposure systems is an extensive engineering and scientific challenge requiring profound knowledge of the biological experiments, open and closed electromagnetic transmission systems, near-field measurement techniques, animal behavior, of biological preparation,

dosimetry, material science, and numerical simulation methods and especially their application in bioelectromagnetics research.

Parameters that must be carefully defined and considered for a proper design of the exposure system include the following:

- Frequency, polarization, and wavelength of the electromagnetic field with respect to the biological object must be ascertained.
- For *in vivo* experiments, the target tissue must be defined and must take into account whether whole body, partial body, or local exposure is required.
- The required SAR levels and field amplitudes must be known at the location of the cell, tissue preparation, or inside the animal's body.
- The desired field distribution at the site of interest, for example, the maximum tolerable deviation from homogeneous field distribution or the degree of correspondence with the actual exposure in human subjects, must be known.

The design and optimization of the exposure system with respect to these parameters requires an efficient and powerful numerical analysis tool and a careful experimental verification of the resulting system. The evaluation must also include a sensitivity analysis for any parameters that might vary during the experiments, for example, size, posture, and movements of animals or amount of medium, location variations of the cells within the flasks, and the position of the flasks within the exposure system. Based on the results of the performance of the system, some changes in the plan might be necessary, which involves a return to the design and optimization stage. Several such cycles may be necessary until a final design of the exposure system is attained. The exposure systems should not be used for biological experiments before the numerical results have been subjected to thorough experimental validation.

In the dosimetric evaluation of an exposure system, particular attention must be paid to the uniformity of the exposure. Although the geometry of the cell culture may be simple, true field homogeneity is quite difficult to achieve for *in vitro* RF studies. Experience has shown that an overall standard deviation from homogeneity of less than 30% is realistic [4]. In practice, a homogeneous field distribution is often not achievable for *in vivo* experiments. However, the system should enable an accurate description of the distribution and magnitude of the induced fields inside different tissues and organs under exposure.

28.5 ANTENNAS FOR *IN VITRO* EXPERIMENTS

As already mentioned in the Introduction, well controlled *in vitro* exposure of individual cells or aggregates of cells to ELF or RF fields is a necessary protocol for determining the induced field or SAR thresholds at the cellular level for biological effects and damage [5–7].

In Vitro exposures to ELF magnetic fields are usually accomplished by using coil antennas of different geometries, capable of producing predictable and uniform magnetic field levels within a given volume where cell cultures are placed. The presence of biological samples does not significantly alter the magnetic field distribution, as is well known from electromagnetic theory.

Exposures of *in vitro* preparations to RF fields are more complex because the biological sample alters the EM field distribution, making it difficult to maintain induced fields and SARs as uniform as desirable, while at the same time the temperature and other environmental conditions must be kept under precise control to prevent electrical artifacts and thermal hot spots from arising. In the past, a common system for RF exposure frequently consisted of a microwave horn antenna used to expose cells in ordinary culture dishes or flasks suspended in free space or within an anechoic chamber in front of the antenna. Under these exposure conditions, the magnitude and uniformity of the SAR in the cells are highly dependent on the size and shape of containers. There may be sharp discontinuities between the cell medium, the air, and the container's wall and the polarization of the incident radiation.

The relationship between the average SAR in test tubes and tissue culture flasks filled with biological preparations and the polarization of the exposure fields follows the well known general characteristics in exposing whole animals. For example, field coupling is strongest when the electric field is aligned with the long axis of an exposed medium and the magnetic field is perpendicular to the side of the exposed medium with the broadest cross-sectional area. Conversely, the weakest coupling of energy occurs if the E-field is aligned with the shortest dimension and the magnetic field is perpendicular to the side with the smallest area. Exposed objects with radii of curvature on the order of a wavelength inside the object can experience internal SAR hot spots that greatly exceed values at the surface.

The effect of field polarization on SAR values and distributions within cell cultures has been investigated extensively [8]. Results based on the commonly used median-sized (60-mm) Petri dishes and rectangular T-75 flasks showed that if the dimensions of the medium are much smaller than the wavelength of the incident field, the total electric field inside the medium is quasistatic in nature—an observation that had been made for whole-body exposures also using spherical models [9]. The induced electric field can be separated into a capacitively coupled electric field E_{cap}, proportional to the incident electric field E_{inc}, and an inductively coupled electric field E_{ind}, proportional to the incident magnetic field H_{inc}. Therefore the SAR can be written

$$SAR = \frac{\sigma}{2\rho}(\mathbf{E}_{\text{ind}} + \mathbf{E}_{\text{cap}}) \cdot (\mathbf{E}_{\text{ind}} + \mathbf{E}_{\text{cap}})^* \tag{28.3}$$

where σ and ρ are the conductivity and density of the medium and the asterisk denotes the complex conjugate of field phasors.

The magnitude of $\mathbf{E}_{\text{cap}}$ is larger if $\mathbf{E}_{\text{inc}}$ is tangential to the boundary of the medium and smaller if $\mathbf{E}_{\text{inc}}$ is normal to the boundary of the medium, due to the high relative permittivity of the aqueous media. For example, in the frequency range used for cellular communication (from 800 MHz to 2.5 GHz) the relative permittivity is about 75. The magnitude of $\mathbf{E}_{\text{ind}}$ is larger if $\mathbf{H}_{\text{inc}}$ intercepts a larger cross section of the medium. Moreover, $|\mathbf{E}_{ind}|$ is proportional to the frequency of the incident field.

From the above discussion, for dishes with horizontal dimensions considerably larger than their vertical dimensions but still considerably smaller than the wavelength, the coupling can be expected to be strongest for H polarization (largest cross section of H_{inc} with the medium and E_{inc} parallel to the longest dimension of the medium), intermediate for K polarization, and weakest for E polarization (small cross section of H_{inc} with the medium and E_{inc} parallel to the smallest dimension of the medium). The influence of the different polarizations is summarized in Figure 28.2. Note that the SAR caused by

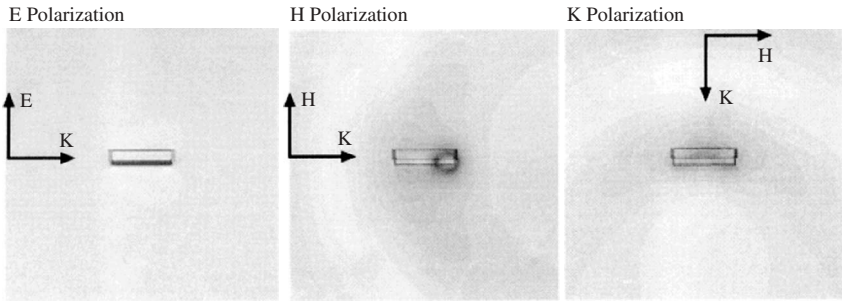


Figure 28.2 Possible field polarizations with reference to a Petri dish. (From Ref. 8.)

H_{inc} is proportional to the square of frequency. As discussed earlier, the coupling is expected to be weak for E polarization. However, an E polarization would produce a homogeneous induced field distribution for a thin monolayer of cells at the bottom of the container. Instead, H and K polarizations would yield better homogeneity for cells in suspension [10].

Expanding on the brief introduction to some of the general aspects concerning *in vitro* experiments, the following sections discuss in detail antennas used for *in vitro* experiments, starting with those for ELF magnetic field exposures and proceeding to antennas used for RF exposures.

28.5.1 ELF Coils

For ELF magnetic field exposures, coils are the preferred antennas. Indeed, coils are used to produce uniform ac or dc magnetic fields within volumes of exposure, which are large compared to those of the coil windings. In most cases, the stray fields outside the coils are designed to rapidly decrease with distance, for EMC purposes. It is worth mentioning that coils used for ELF magnetic field exposure are very similar to those employed as transmitting and receiving antennas in MRI systems, presented in Chapter 27. Even though the classical Helmholtz coil pair is the simplest system to generate a uniform magnetic field in a given volume, another system with an excellent uniform-field to coil-volume ratio, consisting of four appropriately spaced square windings [11], has become very popular as an ELF antenna for biological experiments, both *in vitro* and *in vivo*. The square-shaped coils in this system (see Figure 28.3) have several advantages over the circular coils of the classic Helmholtz coil pair. Namely, the square coils could create a uniform field of approximately square boundaries, which are better suited for laboratory objects such as cages with animals. Furthermore, the rectangular coils and their support structure are easy to fabricate and, if uniformity over a large volume is required, more square coils may be used. Complete details for design of Merritt's coils can be found in Merritt et al. [11], where solutions for the design parameters are obtained by considering a Taylor series expansion for the axial field $B_x(x)$. For maximal homogeneity on the axis, as many terms in the expansion as possible are imposed to vanish at the center. Since $B_x(x)$ will be an even function of x by symmetry, all odd-order terms vanish.

For a three-coil system (see Figure 28.3a), a numerical optimization yielded $s/d = 0.821$ for the ratio of the distance between the two outside coils s to the length d

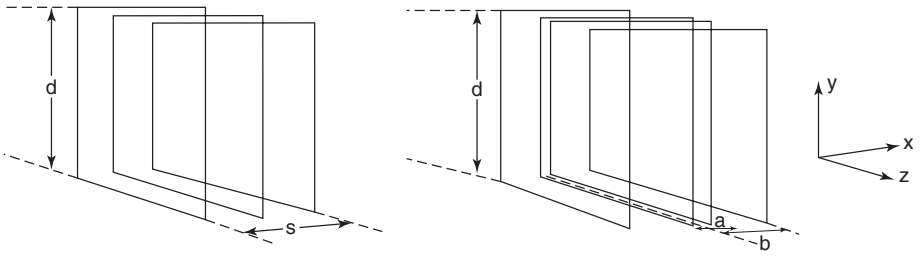


Figure 28.3 Merritt's coil system: (a) three-coil system, and (b) four-coil system. (From Ref. 11.)

of the sides of the coils, and $I'/I = 0.513$ for the ratio of the current in the center coil, I', to that of the outer coils, I [11]. With the coils connected in series, the ratio I'/I can be approximated by integral ratios of turns such as 17/33 or 20/39, and perfected with a shunt across the central coil. For this system the axial field at the center was found to be $B(0) = 1.75I/d$ μT, with I in amperes and d in meters.

The four-coil system (see Figure 28.3b) has one additional degree of freedom and is usually the preferred choice because of its better performance in terms of field homogeneity. A similar numerical solution yielded the ratio of the distance a from the center to the inner pair of coils and d, the side length of the coils:

$$a/d = 0.128 \tag{28.4}$$

The ratio of the distance b from the center to the outer pair of coils and d was

$$b/d = 0.505 \tag{28.5}$$

The ratio of the current in the inner coil pair I' to that in the outer coil pair I is

$$I'/I = 0.424 \tag{28.6}$$

The axial field at the center of the system is given by

$$B(0) = 1.80I/d \quad (\mu\text{T}) \tag{28.7}$$

Merritt et al. [11] also calculated the dimensions of spherical and cylindrical regions in which the axial field does not differ from its value at the center by more than 1 part in 10^6, 10^5, 10^4, 10^3, and 10^2, respectively. For example, for a four-coil system the spherical region in which the axial field does not deviate from its value at the center by more than 10^{-6} has a diameter equal to $0.16d$.

The four-coil Merritt system has been investigated further using a field-modeling program, which was based on the Biot–Savart law [12]. The required ratio between inner coil and outer coil currents was obtained by using different numbers of turns on the two coil groups. In this way, an empirical formula was established, which gives the magnetic flux density B of the uniform field in the center of Merritt's system. This flux

density depends on the total number of turns n, the current I, and the side length d in meters of the square coils, such that

$$B = 0.63\frac{nI}{d} \quad (\mu\text{T}) \tag{28.8}$$

A general expression for the inductance of a Merritt system is obtained as

$$L = An^2d \tag{28.9}$$

where A is an experimentally determined constant [12]. Using a measured inductance of a Merritt system, with coil side $d = 0.42$ m and total number of turns $n = 240$, of $L = 22$ mH, the following empirical formula for the inductance of the Merritt coil system was deduced:

$$L = 0.0009\,n^2d \quad (\text{mH}) \tag{28.10}$$

Other examples of different ELF exposure systems based on Merritt's coil systems can be found in Caputa and Stuchly [12]. These systems allow, for example, exposure of cultured cells in suspension in circular dishes of 100-mm diameter with a magnetic flux density of up to 1 mT and field uniformity better than ±5%. The coil systems can be accommodated inside an incubator with internal dimensions of 1.83 m by 0.91 m by 0.66 m.

An example of Merritt's coil-based cell exposure system capable of generating rotating and linearly polarized magnetic fields of intensities up to 10 mT at 50 Hz and variable frequency can be found in Yamazaki et al. [13], where concentric compensation coils are employed to reduce stray fields. An important feature of this system is the width of the uniform field (400-mm cube with field deviation less than ±3%) within which the cell incubator was installed. To design the coil exposure system, additional calculations of the field uniformity in a cubic volume rather than on an axis or plane have been performed for the 4-square multiple coil system. The result showed that the side length of the cube with field uniformity within ±3% deviation was about 58% of that of the square coil.

Although Merritt's coils are probably the most common solution for ELF exposure systems, other types of antenna configurations also are employed. For example, the classic Helmholtz coils are often used because of their simplicity [14] and the possibility of combining multiple coil pairs to give simultaneous ac and dc magnetic fields at any arbitrary angle. Helmholtz coils can provide field uniformity of better than 1% over a substantial fraction of the volume between the coils. Note that Helmholtz coils are two circular coils spaced by a distance equal to their radius and carry currents in the same sense. They can provide a very uniform field over a substantial part of the region between the coils with a magnetic flux density given by

$$B = \frac{\mu_0 NI}{\left(\frac{5}{4}\right)^{3/2} a} \tag{28.11}$$

where I is the current through each coil, N is the number of turns of each coil, and a is the coil radius.

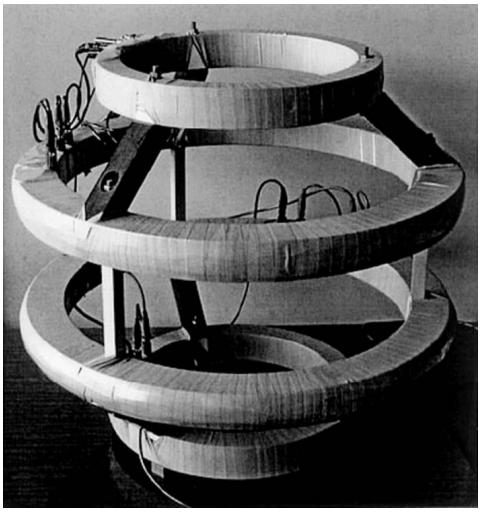


Figure 28.4 Tetracoil prototype. (From Ref. 15.)

Several more complex ELF antenna systems have also been proposed. One such system is the four-circular-coil (Tetracoil) system in which the coils are geometrically constrained by a sphere [15]. In this case, a uniform spherical current shell produces a completely uniform magnetic field throughout the enclosed volume [46]. The Tetracoil exposure system with four coaxial coils inscribed in a sphere showed very good performance compared to most systems used for exposure of biological samples in terms of uniformity of the field, accessibility for the introduction of biological samples, and ratio between overall dimensions and uniformity region. The Tetracoil system, due to its spherical constraint, can easily be extended as a building block for a nested system composed of four coils along each perpendicular direction, in order to completely control the magnetic field within the coils, including the ambient magnetic field. A prototype of the Tetracoil system is shown in Figure 28.4.

Improved versions of Merritt's square coil system have been developed, with optimal uniform B-field exposure; this system also minimizes the difference between the coils for exposure and sham groups [16]. The apparatus was based on two identical coil systems, consisting of four square loops, that are placed inside a commercial incubator. The coils were optimized to produce a homogeneous, linearly polarized B-field perpendicular to the plane of the Petri dishes and placed inside a μ-metal box. The field distribution was optimized numerically through a trial-and-error approach. The degrees of freedom were the number of windings and the positions of the outer and inner loops of the coil. The performance of the optimized system thus obtained has been compared to that of a classic four-coil Merritt system. Results showed that the usable volume with less than 1% field deviation from the coil center was 41% of the total coil volume for the improved four-coil system but only 27% for Merritt coils. Therefore, in addition to better shielding of exposure and sham, the μ-metal casing considerably enlarges the 1% tolerance volume. However, if the field tolerance is reduced below 0.1%, the Merritt system could provide higher uniformity.

28.5.2 TEM Chambers

A wide variety of antenna types have been designed for use with RF exposure systems. One of the first to be adopted for this purpose was the so-called TEM chamber. In particular, the transverse electromagnetic (TEM) chamber or Crawford cell is capable of producing exposure conditions mimicking the free-space environment and thus has been widely used in the past for *in vitro* exposures. Well matched TEM chambers provide a reasonably homogeneous field distribution in the area surrounding its center. Moreover, the power density can be regulated easily and monitored with the help of standard equipment. The TEM chamber is able to simulate free-space exposures in a relatively small space with considerably less power requirements than for plane wave exposures. Other advantages include fields confined within the TEM chamber, and design features that facilitate efficient forced-air cooling, relative affordability, ease of use, and well-controlled sham exposures.

TEM chambers were first developed at the National Bureau of Standards (NBS) (now National Institute for Standards and Technology—NIST) for establishing standard, uniform, EM fields in a shielded environment, to be used essentially for calibrating EM field probes and for emission and susceptibility testing of small to medium sized equipment [17]. The technique employs TEM transmission chambers, which operate as 50-Ω impedance-matched systems. A cross-sectional view of a typical chamber is shown in Figure 28.5.

The Crawford cell consists of a section of rectangular coaxial transmission lines tapered at each end to adapt to standard coaxial connectors. The line and tapered transitions have a nominal characteristic impedance of 50 Ω to ensure minimum voltage standing wave ratio (VSWR) values. The EM field is developed inside the cell when RF energy is coupled to the line from a transmitter connected at the input port. A matched

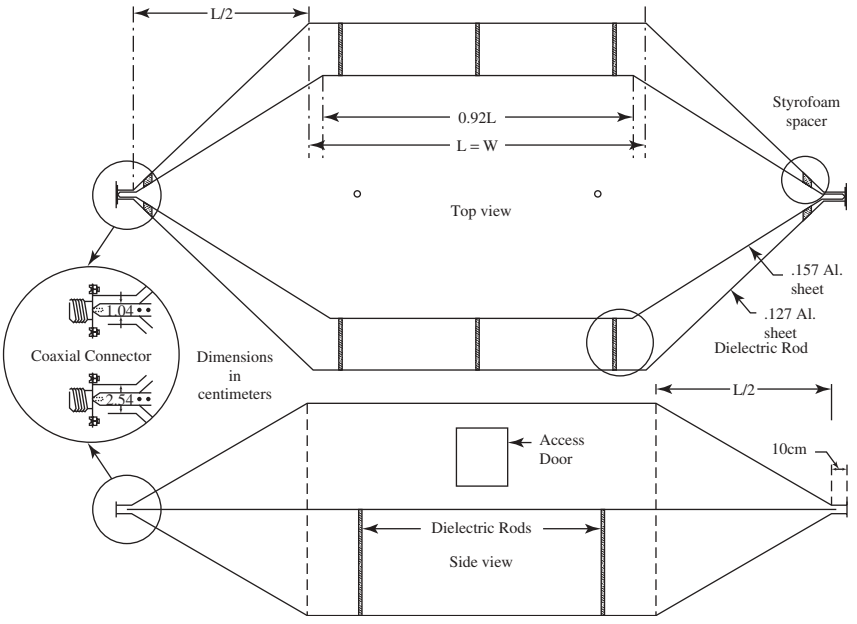


Figure 28.5 Crawford TEM cell sectional views. (From Ref. 17.)

50-Ω termination is connected to the output port. The major design considerations are (1) maximize usable test cross-sectional area; (2) maximize upper useful frequency limit; (3) minimize chamber impedance mismatch or voltage standing wave ratio; and (4) maximize uniformity of EM field pattern characteristics of the chamber.

The chambers are designed using experimental modeling and the approximate equation for the characteristic impedance of a shielded strip line [17], which is

$$Z_0 \cong \frac{94.15}{(\varepsilon_r)^{1/2}\left[\dfrac{w}{b(1-t/b)} + \dfrac{C'_f}{\varepsilon_0 \varepsilon_r}\right]} \quad (\Omega) \tag{28.12}$$

where ϵ_r is the relative dielectric constant of the medium between the conductors, C'_f is the fringing capacitance per unit length (in farads per meter), w and t are the width and thickness of the metal strip, respectively, and b is the total height of the strip line (distance between the outer shielding conductors).

The challenge of modifying the shielded strip line into a "rectangular coaxial" line is primarily one of determining either experimentally or numerically the value of C'_f. This was done by Crawford experimentally by using a time-domain reflectometer (TDR) and an approximate value of C'_f equal to 0.087 pF/cm was found for the cell shown in Figure 28.5 [17]. To meet design considerations 1, 3, and 4, dimensions of b and a (the total width of the cell) are selected to provide as much as one-third of the volume between the septum and outer plates to be occupied by the equipment under test (EUT). However, there is an upper limit for a in order to prevent the excitation of higher-order modes in the cell. Therefore a must be chosen so that the cutoff frequency f_c of higher-order modes is below the maximum frequency at which the chamber is to be used. In particular, the cutoff frequency of the first higher mode can be approximately obtained from

$$f_c \cong \frac{c}{2a} \tag{28.13}$$

where c is the speed of light in vacuum.

Once b is calculated, an experimental or numerical estimate of C'_f can be made, and once an available metal thickness t is selected, w is calculated from the approximate Eq. (28.13), assuming a nominal 52 Ω for the line characteristic impedance, to allow for some impedance loading effect when inserting the EUT inside the chamber.

The absolute electric field strength E_v at the center of the chamber between the upper wall and the center conductor is determined using the equation

$$E_v = (P_n R_c)^{1/2}/d \tag{28.14}$$

where P_n is the net power flowing through the chamber, R_c is the real part of the chamber's complex characteristic impedance, and d is the separation distance between the cell's upper wall and its center plate or septum.

Rather than performing TEM chamber design from scratch, many such chambers are commercially available and can be used as building blocks around which the exposure system may be assembled. Indeed, many examples exist in the literature for applying commercially available TEM chambers to perform RF exposure of cell cultures. For instance, biological experiments have been performed by using standard Petri dishes,

with a given volume of isotonic medium added to the layer of cells lining the bottom of the dish [18]. The dishes were exposed to a narrowband signal at 835 MHz in the center of the standard TEM chamber (IFI-CC 110), which measured $18 \times 18 \times 18$ cm^3. It is interesting to note that an attempt was made to derive an approximation formula for the SAR distribution in the middle of the Petri dish along the vertical axis (z-axis), starting from results obtained through a numerical calculation. Because the coupling of the field with the object (medium) is very poor and the numerical results suggest a quasistatic approach (i.e., separation between "capacitive" and "inductive" coupling), the induced SAR can be estimated as

$$SAR(z) = \frac{\sigma}{2\rho}|\mathbf{E}_{\text{inc}}|^2 \left[\left(\frac{\mu\omega z}{Z_0} \right)^2 + \frac{1}{|\varepsilon_k|^2} \right] \tag{28.15}$$

where $z = 0$ corresponds to the dish center, $Z_0 = 377\ \Omega$, $\varepsilon_k = \varepsilon_r - j\sigma/\varepsilon_0\omega$, and ρ is the mass density of the medium.

In specific cases, commercially available TEM chambers have been modified to make them better suited for particular experiments. For example, two modified TEM chambers (model IFI-CC 110), housed in a water-jacketed incubator, have been used [19]. The TEM chambers were modified as follows:

1. Shelves for placing four T-25 flasks (25-cm^2 area for the bottom of each flask) at the needed location within each TEM chamber were added. On the long sides, the shelves were provided with slots assisting repeatable positioning of the flasks and enabling cooling of the medium.
2. The doors were enlarged for easy access to both sides of the central conductor.
3. Forced airflow through the TEM chambers was achieved by introducing ventilators on the top of each cell.
4. Power sensors were mounted at the output of each TEM chamber for improved exposure control.

The efficiency of this exposure system was evaluated, and the overall average SAR for all flasks was determined to be 6.0 W/kg at 1-W input power, with a standard deviation of less than 52%.

28.5.3 Rectangular Waveguides

Apart from TEM chambers, another device that is widely employed to perform *in vitro* RF exposures is the rectangular waveguide. In fact, as indicated earlier, when exposing cell cultures inside Petri dishes, the polarization of the incident electric field parallel to the cylindrical symmetry axis of a Petri dish (E polarization) provides the highest degree of homogeneity. Moreover, E polarization causes the lowest disturbances to the incident field; that is, it would allow several Petri dishes to be positioned in the same system. But this will result in low efficiency for an individual dish, while having minimal effect on the overall efficiency.

The rectangular waveguide is a simple way to obtain an exposure system with an ensured E polarization [20, 21]. For example, a waveguide system based on standard waveguides was evaluated and found to be well suited with respect to efficiency, size, and cost, to perform cell exposures if the carrier frequency is restricted to the band between

1.5 and 2.5 GHz [22]. An interesting aspect of rectangular waveguides is that their efficiency can be increased by using a short-circuit termination. Indeed, it has been shown that inductive coupling is the dominant coupling mechanism in the case of suspension and plated cell cultures (z in Eq. (28.15) is maximized), while capacitive coupling is the dominant factor for hanging drops ($z \cong 0$ in Eq. (28.15)) [18]. Therefore efficiency can be increased by terminating the waveguides with a short-circuiting plate and placing the Petri dishes in the maxima of the standing wave's magnetic field, if exposure of cell suspensions or plated cultures is required, or in the maxima of the standing wave's electric field, if hanging cells in drops are exposed. Note that a condition for sufficient homogeneity of the SAR distribution is that the wavelength must be significantly larger than the diameter of the Petri dishes. Another advantage of terminating the waveguide with a short-circuiting plate is the reduction of the overall length of the exposure system. Additionally, a removable short-circuiting plate allows easy and fast access to the Petri dishes.

There are two basic criteria to consider when choosing a rectangular waveguide exposure system: the operating frequency of the waveguide must cover the range required by the specific experiment and the wavelength must be sufficiently large compared to the diameter of the exposed sample. If a and b are the width and height of the waveguide, respectively, for $a > 2b$, the ideal operating frequency range of the waveguide is given by

$$f_{c1} \leq f \leq f_{c2}; \quad f_{c1} = c/2a \quad \text{and} \quad f_{c2} = c/a \tag{28.16}$$

where c is the speed of light in vacuum. For a given frequency f within the operating range of the waveguide, the corresponding wavelength λ can be computed as

$$\lambda = \frac{\lambda_0}{\sqrt{1-\nu^2}}; \quad \lambda_0 = \frac{c}{f} \quad \text{and} \quad \nu = \frac{f_{c1}}{f} \tag{28.17}$$

Like TEM chambers, commercially available rectangular waveguides in the frequency range of interest can be used to build the exposure system. For example, commercially available R14 waveguides (cross section 165.10 mm $\times$ 82.55 mm) have wavelengths inside the waveguide of about 200 mm at 1.71 GHz, which means that the neighboring maxima of the standing wave's electric (or magnetic) field are separated by approximately 100 mm. Of course, assessment of SAR distribution within the sample is mandatory. For example, one such system was optimized using numerical techniques [22]. The number of Petri dishes per waveguide can be increased by stacking the dishes on top of one another. The numerical analysis showed that the homogeneity of the SAR distribution was greatly impaired in stacks of three dishes, whereas there was little loss in performance if two dishes were stacked together. The coupling efficiency was found to be 1.5–2 W/kg per watt of input power for the plated and suspension cultures. Depending on the culturing method and medium used, the standard deviation of inhomogeneity was between 25% and 35%.

The performance of the exposure system described above could be further improved by simply changing the transition from the feeding coax cable to the rectangular waveguide from a monopole to a flat loop [1]. Indeed, the loop coupler, compared to the monopole coupler, has the advantage of reducing the evanescent mode region inside the waveguide and therefore enabling a larger proportion of the waveguide to be excited.

28.5.4 Radial Transmission Lines

Both TEM chambers and rectangular waveguides would only allow exposure of a limited number of Petri dishes or flasks. In many cases, simultaneous exposure of many samples under similar conditions is required. The radial transmission line (RTL) offers a convenient structure to provide uniform irradiation of many preparations simultaneously. The RTL is a waveguide formed by two parallel metal plates into which a transverse electromagnetic wave is launched by an appropriate radially symmetric antenna at its center. The entire unit is appropriately terminated at the lateral plate boundaries. Locally, the incident field is a free-space-like TEM plane wave, as long as circumferential or longitudinal higher-order mode excitation is not very significant (see later discussion). Typically, after a short distance ($\rho > \lambda_0/2$), the cylindrical wave impedance would approach the plane wave impedance for propagating waves. Therefore an exposure in the radial waveguide is similar to exposure in free space, provided the cross section of the exposed body is much smaller than its distance from the center so that the impinging wavefront can be considered locally flat and uniform. A shorting electric wall around the circular perimeter converts the radial waveguide into a cavity. Alternatively, RF absorbers (i.e., matched loads) can be placed around the perimeter. Some distinct advantages deriving from terminating a radial waveguide in a matched load include good impedance matching to the RF source, which is typically broadband and, to some extent, insensitive to the position and size of the exposed bodies.

The theory of the RTL and its practical realization are available in the literature [23]. By considering the propagation in an RTL of plate spacing b (see Figure 28.6), the following can be shown:

1. The fundamental mode is transverse electromagnetic (TEM), the **E** vector having only a vertical, z-component that varies in neither the vertical (z) nor the angular (ϕ) directions and, in the lossless case, falls off asymptotically in the radial (ρ) direction as $1/\sqrt{\rho}$. At sufficiently low frequencies, only this mode will propagate. The electric and magnetic field components for this mode are given by

$$E_z(\rho) = A_f H_0(\beta_0 \rho) \quad \text{and} \quad H_\phi(\rho) = \frac{1}{-j\eta_0} A_f H_1(\beta_0 \rho) \tag{28.18}$$

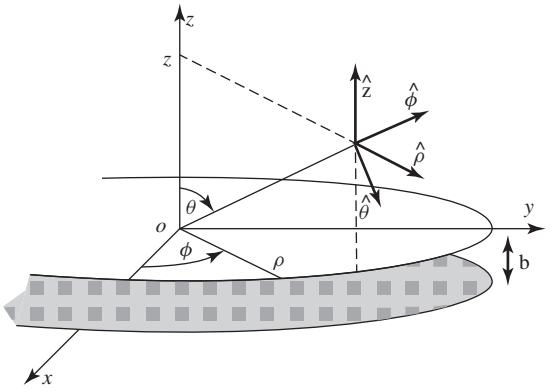


Figure 28.6 Reference frames associated with the RTL. (Adapted from Ref. 24.)

where A_f is an amplitude constant of the forward-propagating wave, H_0 and H_1 are Hankel functions of zeroth and first order, respectively, $\beta_0 = 2\pi f\sqrt{\mu_0\varepsilon_0}$ is the free-space propagation constant, and $\eta_0 = \sqrt{\mu_0/\varepsilon_0}$ is the intrinsic impedance of free space.

2. As the frequency increases, higher-order modes with z or ϕ variation can develop. The lowest frequency at which a given order (m) of angular mode can propagate is given by

$$f_{m0} = cm/2\pi\rho \tag{28.19}$$

while the lowest frequency at which a given order (n) of vertical mode can propagate is

$$f_{0n} = cn/2b \tag{28.20}$$

The importance of a gradual transition from the feeding coaxial line to the RTL is discussed in Moros et al. [23]. A cross section through the center of the transition is shown in Figure 28.7a. Back reflection from the lateral boundaries was suppressed by a layer of microwave-absorbing foam (see Figure 28.7b).

In the system described in Moros et al. [23], 16 T-75 flasks, centered on a radius of approximately $\rho_0 = 292$ mm, extending radially over a distance $\delta\rho = \pm 46$ mm, could be exposed simultaneously. At 835.62 MHz and for a nominal net forward power of 1 W, the average SAR among all locations inside T-75 flasks filled with 40 mL of culture medium was 16.0 ± 2.5 mW/kg (mean $\pm$ standard deviation). The nominal SAR variation was ± 0.69 dB inside the culture; but one naturally expects this nominal value to be modified in practice by scattering at the plastic (e.g., see Lin and Wu [25] and Lin et al. [26]) and at the culture–medium interfaces.

28.5.5 Wire Patch Cells

An interesting alternative to the RTL for simultaneous exposure of multiple samples, which has recently been developed, is the "wire patch cell" (WPC). This exposure system, based on the wire patch antenna, was designed for operation at 900 MHz [27]. Compared to TEM chambers, this device can generate higher E-field levels with the same input

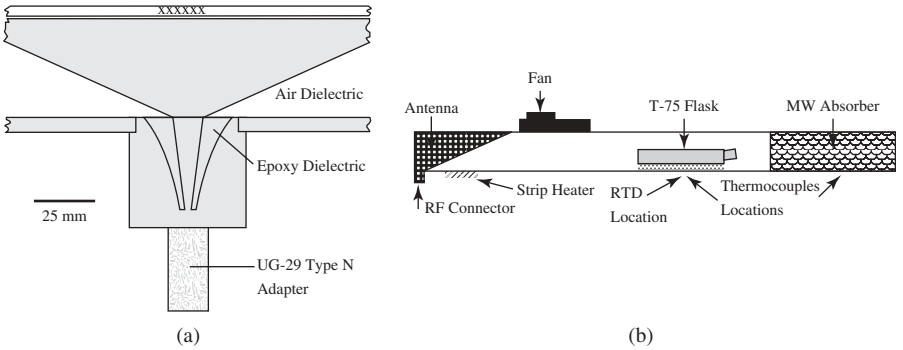


Figure 28.7 Cross section through the antenna of the RTL: (a) detail of the antenna region and (b) half-structure view. (From Ref. 23.)

power. The wire patch cell can easily be built and used because of its small size and simple structure. This open device can be placed into an incubator, leading to better ventilation to avoid the possible temperature increase intrinsic to closed systems. Similar to the RTL, the simple structure allows exposure of several biological samples to the same field level, thus enhancing the statistical power of biological studies.

The wire patch antenna was developed from the concept of miniature planar antennas. It is an antenna having the global structure of a conventional coaxially fed microstrip antenna with one or more ground wires connecting the patch to the ground plane. This structure may be used to expose cell cultures since it provides a homogeneous 900-MHz field between the patch and the ground plane. In addition, the field distribution between these planes is independent of the antenna's shape, and it is possible to properly match the antenna to 50 Ω at 900 MHz by an appropriate modification of its physical parameters. Specifically, the miniature antenna was widened to allow the insertion of sample holders. A prototype of the empty device (see Figure 28.8) consists of two metallic planes (the ground plane and the roof patch), four shorting posts, and one coaxial probe located at the center of the device, going through the ground plane and connected to the roof. Both metallic planes are of the same size in order to increase the electric field amplitude under the roof. The four metallic grounding contacts are located at each corner of the device roof in order to maintain a large free area at the center of the device, where the electric fields remain homogeneous. The coaxial probe is located at the center of the device in order to avoid the excitation of the microstrip antenna cavity fundamental resonance mode, which would disturb the uniformity of the electric field distribution under the roof. Consequently, the antenna structure is symmetric, enabling it to irradiate several sample holders equally.

The exposure device was matched at 900 MHz. However, the matching of the test cell is sensitive to many parameters, such as the height of the medium present in the sample holders. Consequently, to maintain the test match at 900 MHz, a matching system was integrated into the wire patch antenna, using one of its intrinsic parameters. The matching process consisted of simultaneously moving the four props (contacts) along the device diagonals. The final design, obtained after numerical optimization, was a 15-cm × 15-cm × 2.9-cm device capable of exposing up to eight 35-mm Petri dishes (two stacks of four per side). The mean SAR values and their maximum deviations, for 1 W of input power, found in the 35-mm Petri dishes were as follows [27]:

- First level dishes: 575 ± 66 mW/kg
- Second level dishes: 560 ± 68 mW/kg

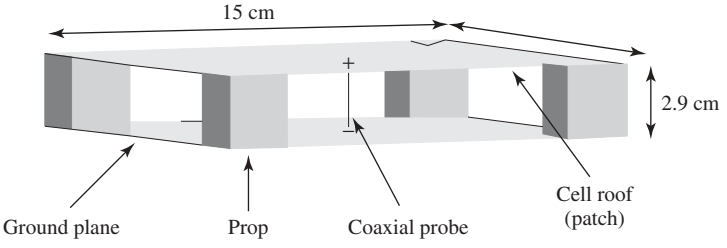


Figure 28.8 Wire patch cell. (From Ref. 27.)

Unfortunately, the WPC is a rather complex system for which no simple analytical design formula is available. Once the dimensions of the structures are obtained, based on constraints deriving from dimensions of the samples to be exposed, numerical simulations must be performed in order to optimize the position of the shorting posts so as to achieve an acceptable impedance match. Of course, further simulations must then be performed with the samples in place to assess SAR levels and SAR uniformity within the samples themselves.

Based on the original 900-MHz design, a 1800-MHz WPC device was developed [28]. Since the dimensions of a radiating structure are related to the signal wavelength λ, it would have been possible to use half-scaled dimensions of the 900-MHz WPC for the realization of the 1800-MHz WPC (i.e., $7.5 \times 7.5 \times 1.5$ cm^3). However, such a device would be too small to contain sample holders like Petri dishes with a diameter of either 3.5 or 5 cm. Therefore the WPC was redesigned using a numerical approach, leading to a prototype of the device consisting of two 20-cm $\times$ 20-cm plates (the ground plane and the roof, placed at a distance of 2 cm apart), four metallic grounding contacts (props) having a diameter of 1 cm, and one coaxial probe with an inner diameter of 3.8 mm (located at the center of the device, passing through the roof and connecting to the ground plane). This symmetric structure allows easy insertion of four 3.5- or 5-cm Petri dishes and fits adequately into an incubator. The power coupling efficiency of the system was 1.2 W/kg per watt of input power.

28.5.6 Coplanar Waveguides

As already stated in the Introduction, often the need for monitoring during exposure of the biological sample imposes specific constraints on the exposure system, which may result in the need for nonconventional antennas. One such example is when ionic currents through the cell membrane must be recorded during exposure, through the so-called patch-clamp technique [29].

Indeed, the intrinsic nature of the patch-clamp experimental technique requires the use of a glass microelectrode for current recording (see Figure 28.9) and the use of a microscope for cell identification, sealing, and observations. One solution is the adoption of the coplanar waveguide (CPW) structure used for the guided propagation of electromagnetic fields [30]. The CPW is an open propagating structure with a dielectric substrate on which three metallic strips are positioned. Figure 28.10 shows the cross section of the structure, together with E-field flux lines, where w represents the distance between

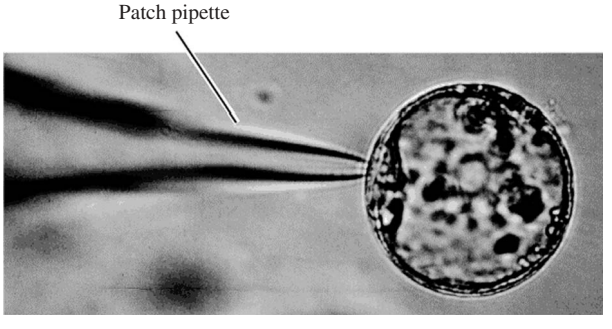


Figure 28.9 Patch-clamp technique.

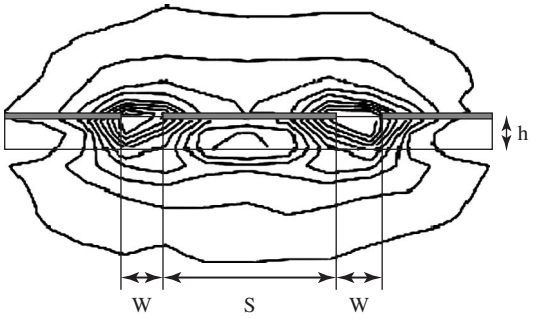


Figure 28.10 Cross section of a CPW. (From Ref. 30.)

the central conducting metallic strip and the two lateral reference planes, S represents the width of the central conducting strip, and h represents the height of the dielectric substrate.

Replacement of the microscope head stage with a CPW on a glass substrate and positioning of Petri dishes in the visibility windows (width w) between the strips permits the insertion of a microelectrode and, at the same time, guarantees the concentration and focusing of EM field lines close around the exposure zone. In particular, in a CPW the E- and H-fields of the propagating mode have a quasi-TEM configuration, where the direction of the E-field is almost parallel to the Petri dish surface and, as a consequence, orthogonal to the glass microelectrode used for current measurements. In this way, it is possible to minimize the interference of the field with the electrode.

The biological and EM requirements to be satisfied in the design of the CPW are the following:

- Glass substrate in order to achieve visibility of the sample, $\varepsilon_r = 5.5$.
- Finite visibility window to insert the external electrode, that is, $w \geq 0.5$ cm.
- Maximum optical length, that is, $h \leq 1$ cm.
- Avoiding losses due to higher modes and dissipation in the dielectric substrate, that is, $h \leq 0.1\lambda$.
- Avoiding losses due to radiation, that is, $2\ w + S \leq \lambda/\ 2$.
- Characteristic impedance of the structure, Z_0 (which is a complex function of dielectric constant ε_r, transversal dimensions w and S, and vertical dimension h) equal to 50 Ω, in order to achieve a good matching of the structure when connecting it to a typical coaxial cable. In particular, the characteristic impedance Z_0 can be obtained from the following formula [31]:

$$Z_0 = \frac{30\pi}{\sqrt{\varepsilon_{re}}} \frac{K'(k_1)}{K(k_1)} \tag{28.21}$$

$$\varepsilon_{re} = 1 + \frac{\varepsilon_r - 1}{2} \frac{K(k_2)}{K'(k_2)} \frac{K'(k_1)}{K(k_1)} \tag{28.22}$$

where K and K' are the complete elliptic integrals of the first kind and their complements, respectively, with $k_1 = S/(S + 2w)$ and $k_2 = \sin \text{h}\ (\pi S/4\ h)$.

Choosing an operating frequency range between 800 and 2200 MHz, and fulfilling the geometric constraints from both sets of requirements, it is not possible to achieve a value for the characteristic impedance lower than 59 Ω. Therefore, in order to obtain a match between the structure and the feeding coaxial cable, it was necessary to use a linear impedance taper. Moreover, a single impedance match is insufficient to obtain a low and uniform reflection coefficient over the entire operating band. In addition, care must also be taken to avoid field propagation and to reduce reactive effects at discontinuities. Therefore, starting from classical coax-to-CPW transitions, the CPW conductors were tapered to the dimensions of the coaxial cable. The resulting system showed an efficiency for Petri dishes, placed in the two visibility windows for different points within the dish, ranging from 13 to 20 W/kg per watt of input power at 900 MHz, and from 33 to 48 W/kg per watt of input power at 1750 MHz.

28.6 ANTENNAS FOR *IN VIVO* EXPERIMENTS

For an assessment of the impact on humans of exposures to EM fields in the ELF and RF ranges, it would be most relevant to conduct studies of the distribution of any disease and the extent of ELF and RF exposures in human populations. However, the rapid development and widespread utilization of EM devices and systems and the retrospective nature of this type of studies often prompt the use of alternative approaches. Moreover, the wide variety of exposure scenarios makes it difficult to ascertain true exposure with an acceptable degree of certainty. While there are protocols in epidemiology which may help compensate for this deficiency, more often, laboratory experiments are carried out *in vivo*, employing living animals to investigate potential effects in humans or related issues. Likewise, while *in vitro* studies using cell and tissue preparations may provide insights into biological response mechanisms and thresholds of interaction, its utility is limited in that the potential implications of any observed effect or threshold from cell cultures may not easily be translated into humans. Thus experiments are typically carried out in laboratories involving living animals. The most commonly employed species in these studies are mice and rats, but rabbits and monkeys are not unusual. It is important to emphasize that the same exposure parameters do not usually produce the same SAR and SAR distributions in cells *in vitro*, in animals *in vivo*, or in humans.

In general, exposure systems for *in vivo* experiments are very different from those employed for *in vitro* studies. Some of the reasons include:

- The volume occupied by exposed animals is generally much larger than that required by Petri dishes or flasks.
- Animals tend to move inside the exposure system, unless they are restrained, but in such case possible confounding effects resulting from animal stress must be carefully investigated, usually through a sham-exposed control group.
- No incubator is needed, but on the other hand, proper ventilation, feeding, watering, grooming, and so on must be provided.

While the abovementioned factors make exposure systems for *in vivo* studies different from those employed for *in vitro*, the types of antennas used can be similar or even identical. One important distinction that must be made is between antennas designed to achieve localized (partial-body) exposure and those designed for whole-body exposure.

28.6.1 Localized and Partial-Body Exposures

The main objective of antennas for localized exposures is to induce energy deposition in a confined region of the exposed animal, while keeping SAR levels in the remaining regions as low as possible. This attribute is indeed equivalent to that required for external hyperthermia applicators (see Chapter 27). It is therefore not surprising that many experiments involving localized exposures of animals had been conducted in the past using similar types of applicators. For example, several head exposure studies were conducted with waveguide aperture or contact applicators on cats, rats, and guinea pigs to study microwave effects on the nervous systems [32–36]. Similarly, many microwave cataractogenesis studies have been carried out using a diathermy system, 2450-MHz C-director applicator, to irradiate rabbit and monkey eyes [37–39]. However, in more recent experiments, aimed at studying the biological effects of localized head exposure resulting from the use of cellular phones, it was realized that a better animal exposure system simulating human exposure was required. The two most commonly adopted solutions consist of using loop antennas or carousels, as described in the following.

28.6.1.1 Loop Antennas The initial specifications that led to the use of loop antennas were intended to develop a system capable of producing energy absorption in a rat brain comparable to that in a human brain and an order of magnitude higher than those experienced by wireless phone users [2]. A numerical approach was used to model dipole and loop antennas and determine whether these antennas were appropriate for localized rat head exposure. A few short dipoles, 3 cm long, positioned tangentially along the surface of a skull model, with a separation distance of 1 cm, were simulated. Results showed localized intense absorptions near the ends of the dipole due to high electric fields. It was also determined that short dipoles were inefficient radiators; therefore they were inappropriate for this application. The investigation continued with different sized small loop antennas positioned in various orientations with respect to a model of the skull. It was determined that a 3-cm × 1-cm rectangular loop aligned with the loop outer edge parallel to the skull would maximize magnetic field coupling to brain tissue while providing SAR distributions similar to those reported in human model studies.

A schematic drawing of a loop antenna, designed to operate at 837 MHz, is given in Figure 28.11a. Unfortunately, small loops behave as very low impedance antennas. Indeed, the radiation impedance of a small loop can be obtained from

$$R_r \cong 31.2\left(S^2/\lambda^4\right) \tag{28.23}$$

where S is the loop area and λ is the free-space wavelength. Application of the formula to the 3-cm^2 area loop antenna would yield, at the operating frequency of 837 MHz, a low radiation resistance of less than 1 Ω.

This low impedance antenna was matched to a 50-Ω source impedance through a section of capacitively coupled transmission line embedded in a Teflon rod. This loop is tunable for the cellular phone frequency band (800–900 MHz) by adjusting the loop feed-line length and depth in the Teflon rod. By extending the Teflon rod and adjusting the lengths of the capacitively coupled feed lines through a numerical optimization procedure, the basic design was modified to produce a loop antenna for operation at 1.8–2 GHz (see Figure 28.11b). Note that a half-wavelength longer feed line and Teflon rod were used because the shorter version was difficult to tune. Dosimetric studies have shown that SAR was higher near the top of the brain, below the skull, with values ranging between

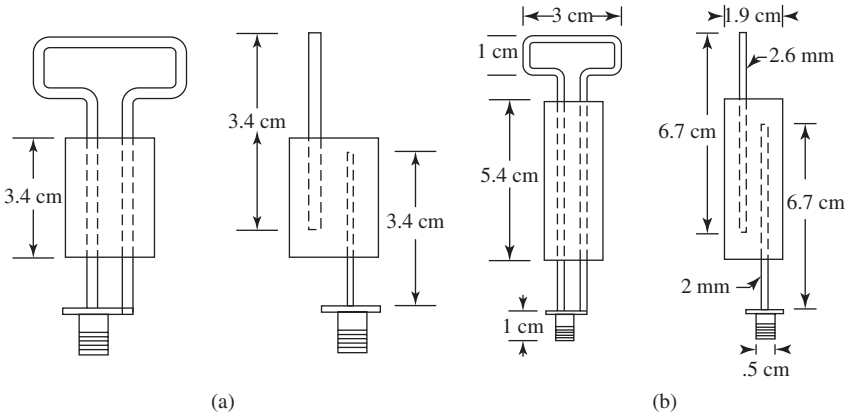


Figure 28.11 Loop antennas for localized head exposure: (a) 900 MHz and (b) 1800 MHz. (From Ref. 2.)

60 and 90 W/kg per watt of input power [2]. In the sagittal plane, the SAR was 30–60 W/kg per watt of input power in about one-third of the brain and 5–30 W/kg per watt of input power in the remaining two-thirds of the brain.

The original loop design was subsequently modified in a new development [40]. Starting from the wire loop antenna, the new system was designed, consisting of a printed circuit with two metallic lines printed on a dielectric substrate (Figure 28.12). The dielectric substrate was composed of epoxy resin. One end of the metallic structure was made of a 1-mm wide metallic line in the shape of a loop. This antenna may be placed close to the head of the animal inside a restrainer. A coaxial connector was soldered to the metallic lines in order to feed the antenna. A classical movable stub, made out of two microstrips, was added for matching. The correct position of the stub can be found through numerical optimization or experimentally.

28.6.1.2 Carousels The loop antenna was designed for exposure of a single rat. One of the solutions proposed for achieving simultaneous RF head exposure of several restrained animals is the so-called carousel [41], configured for irradiation in small shielded enclosures called "chamberettes." The chamberette is quasi-anechoic and has a 40-rat capacity (four 10-rat units). The aluminum sheathing allowed them to function as Faraday cages, which electromagnetically isolated the interior from the exterior. A chamberette is shown schematically in Figure 28.13.

Each 10-rat unit was assembled on a device called a carousel. Each rat was restrained using a transparent tube, which keeps the rat head directly facing the radiating antenna, thus causing exposure to take place mainly in the animal head. A large aluminum box containing 40 rats, a small quantity of low loss plastic, and four antennas potentially could resonate as a microwave reverberant chamber. It would be prone to hot spot generation and nonuniform exposure. To suppress this tendency, the lateral interior faces of the chamberette (including the door) were lined with microwave absorber. A standard telecommunications antenna, namely, a half-wave sleeve dipole with a rated operating range of 806–870 MHz and a nominal input impedance of 50 Ω, was chosen for the radiating antenna. It must be stressed that the resulting antenna system was composed of a vertical collinear array of four linear antennas. However, since each antenna was

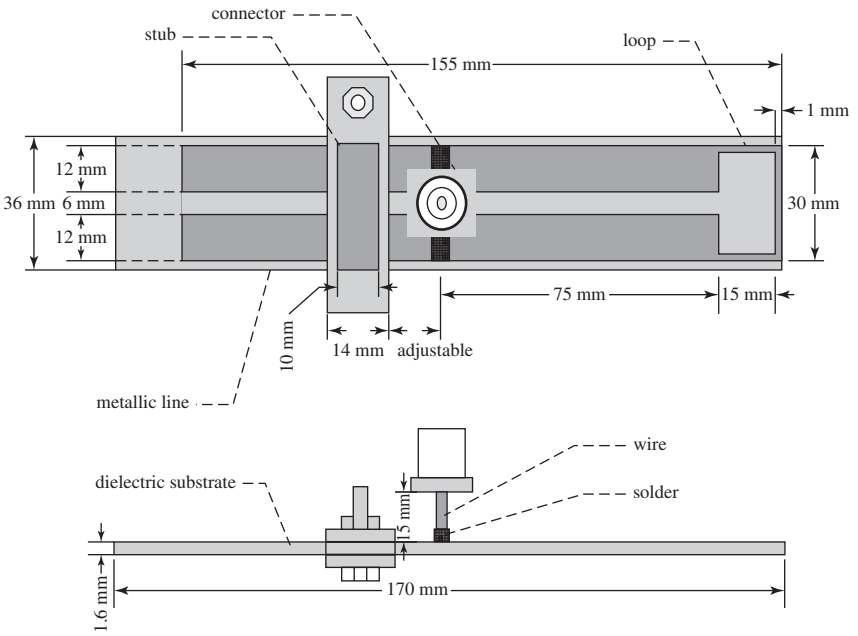


Figure 28.12 Printed loop antenna for localized head exposure. (From Ref. 40.)

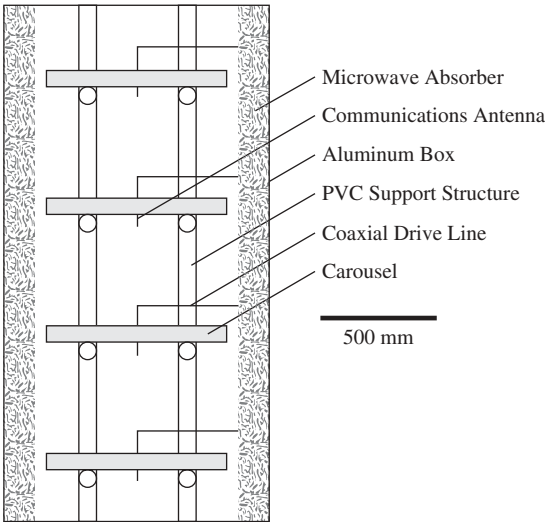


Figure 28.13 Schematic view of a "chamberette" with carousels inside. (From Ref. 41.)

positioned in the minima of the other antenna's field, the antennas had little influence on each other.

The carousel exposure system was subsequently reexamined in order to analyze the suitability of the apparatus for the higher frequency bands used for wireless communications, that is, 1.4–2 GHz [42]. Dosimetric analysis performed at 1.62 GHz indicated a decrease of about 5% of the average SAR in the brain when the distance between the antenna and the nose of all rats was increased from 30 to 37.5 mm. Vertical movements of ±10 mm of the head, up or down, changed the average SAR in the brain in the range of about ±15%, with the lowest values occurring for an animal in the up position. The same simulations also predicted that the system's efficiency is highest when about 60% of the emitted power is absorbed by the animals. The average SAR in the brain was found higher by a factor of about 7–9 than the average SAR in the body. Finally, for an adult animal, the power conversion efficiency was about 2.8 W/kg per watt of antenna input power for a 50-Ω matched sleeve dipole antenna.

28.6.2 Whole-Body Exposures

Exposure systems designed to provide whole-body animal exposure employ antennas that are, in principle, very similar to those used for plane wave *in vitro* exposures. In the past, one of the most commonly adopted irradiation techniques was to use a microwave anechoic chamber. Animals were normally arranged in a planar array in the far field, at one end of the chamber, with a horn antenna located at the opposite end. Various polarizations of the field can be used, but all require moderately high irradiation powers because the power density falls off as the area subtended by the microwave beam increases. Moreover, the uniformity of the power density declines rapidly off axis.

One possible alternative is the use of a microwave reverberant (or mode stirred) chamber in which the test animals are placed. This device is similar to a microwave oven with mode stirrers [43]. The principal advantages of a reverberation chamber are its potential for field uniformity and its ability to accommodate multiple tests to be performed in parallel. This can significantly increase the statistical sample size and therefore increase the reliability of experimental results. Moreover, the possibility of a quick determination of the total absorbed power, by measuring the energy balance of the chamber, allows easy evaluation of whole-body averaged SAR. However, the reverberation chamber does not lend itself favorably to ascertaining SAR distributions inside the specimen or experimental animals. EM fields in reverberation chambers are known to interfere with many of the required animal keeping and housing functions in a long-term study and are prone to produce SAR artifacts. The logistics involved in supporting a chronic animal investigation are extensive. Thus, in most recent studies, the antennas encountered often are coils for ELF exposures and radial waveguides or TEM chambers for microwave exposures.

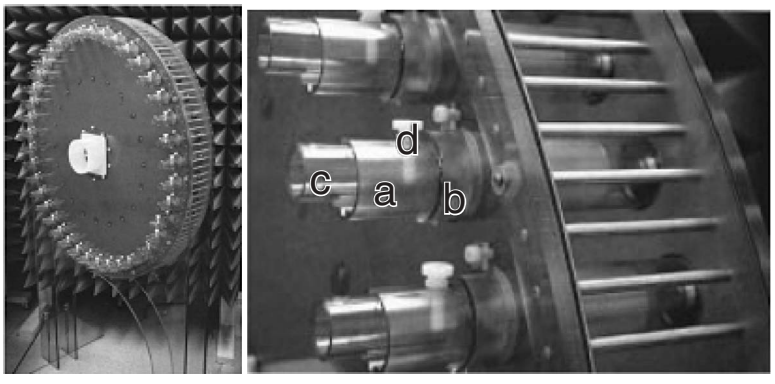
28.6.2.1 ELF Coils ELF coils do not differ from those already described with respect to *in vitro* systems, and therefore the reader can refer to the previous section for information concerning their design and utilization. An example of an *in vivo* ELF exposure system making use of Merritt's coils can be found in Caputa and Stuchly [12].

28.6.2.2 Radial Waveguides Radial waveguides (also called radial transmission lines—RTLs) are one of the preferred solutions for simultaneous, whole-body RF exposure of a large number of living animals. Their basic operation has already been discussed

with respect to *in vitro* exposure systems, where design guidelines for such systems have been given. For example, a radial waveguide was developed for simultaneous exposure of nonrestrained hamsters placed in cages arranged on a circle around the center of the waveguide. Sufficient spacing and absorbers (behind the cages) were provided to ensure no higher-order waveguide modes could occur in the vertical direction for operations up to 900 MHz. A dosimetric study revealed that, depending on the relative position of the animals, the field distribution could change due to mutual shadowing effects [44]. From these calculations, an average SAR of 80 mW/kg per hamster was achieved at 900 MHz for a total input power of only 2.9 W. The standard deviation of SAR was 30%, accounting for the presence of absorber materials and the radial field decay along the cages. Of these, approximately 25% are ascribable to the biological design of the experiment investigating nonrestrained animals.

Starting from the radial waveguide idea, a new exposure system, nicknamed the "Ferris wheel" for its appearance (Figure 28.14), was developed [24]. The exposure system consists of a radial electromagnetic cavity formed by parallel circular plates mounted on a polycarbonate frame, joined around the perimeter by an array of shorting posts. A tunable transition from a 50-Ω coaxial feed line excites a cylindrical TEM wave that propagates in a carousel of 40 symmetrically arranged mice, equidistant from the excitation. The mice, restrained in plastic tubes inserted through circular holes in the plates, are held copolarized with the incident electric field (E polarization) to maximize the absorption of RF energy. Compared with the RF-absorber-terminated waveguide, the use of shorting posts makes impedance matching more sensitive to the cavity load (e.g., the size of mice). The loaded cavity has high stored reactive energy so that narrowband impedance matching to the source is typically achieved. Thus a tunable coax-to-cavity transition is needed to enable good matching at the desired operating frequency. The transition was formed by a top-loaded monopole antenna, with a capacitively coupled passive counterpoise. Tuning of the loaded cavity was performed through adjusting the capacitive coupling, which was done by moving the counterpoise closer to or farther away from the monopole, and was accomplished by threads on its arm.

The dosimetric analysis of the exposure system revealed that the distance of the lateral wall from the mice was the principal parameter to be optimized for efficiency. In particular, a periodic dependence was found, indicating that the best matching is achieved



(a) (b)

Figure 28.14 Ferris wheel exposure system: (a) general view and (b) mouse restrainer detail. (From Ref. 24.)

when the short-circuited radial waveguide section acts as an inductive load, in contrast to the common assumption that the highest absorption would occur at the peak of the electric field at about a quarter-wave from the short. In terms of efficiency, the study revealed that more than 80% of the input power was absorbed by the exposed animals. Moreover, improved uniformity of the SAR distribution with respect to typical planewave incidence was obtained due to the animal's exposure to waves reflected from the lateral cavity walls.

28.6.2.3 TEM Chambers An alternative solution to radial waveguides for the simultaneous exposure of a large number of animals is the use of TEM chambers. A system based on an extended length TEM chamber was proposed [45]. It allows exposure of up to 12 mice to a uniform plane wave at 900 MHz, with a higher value of power per unit of mass induced in the biological target. The TEM chamber had equal transversal dimensions with a broad and thin center conductor, to produce a large uniform electric field region. A transverse dimension of 12 cm and a 10-cm wide inner septum was utilized to ensure a pure TEM mode propagation up to 1.25 GHz, and to allow 50-Ω matching at the chamber ports. The length of the transmitting part of the cell was fixed to 120 cm (i.e., 4λ at the frequency of 900 MHz) for the simultaneous exposure of several animals under the same conditions. Numerical simulations have shown that, in the transversal section of the chamber, a region of uniform electric field indeed exists. The region was large enough to position a couple of mice on each side with respect to the chamber inner septum. Moreover, the caudal axis must be parallel to the field propagation direction in order to have the best condition for homogeneity of exposure. The average power efficiency for the central group of mice was found to be about 0.4 W/kg per watt of input power.

28.7 SUMMARY

Recent interest in the interaction of electromagnetic fields with biological systems has focused on the frequency regions used for cellular mobile communications and electric power transmission and distribution. Aside from their intended roles in communication, entertainment, and power delivery, the EM energy associated with these applications produces biological effects that have been shown to influence the life processes of living organisms. While the subject of biological effects and health implications of EM fields remains controversial, clearly, research and development in applications of electromagnetic systems will continue to expand, and with it our need for better understanding of biological interactions.

It is emphasized that progress in this area can come about only from an appreciation that this field of study—bioelectromagnetics—is an interdisciplinary endeavor. Indeed, it places requirements on investigators that are beyond their usual training in fields of particular disciplines. The subtle interactions that take place between the EM fields, exposure system, measuring instrumentation, the experimental subject, and any observed biological response may easily mislead one trained in a specific field of study. The spurious response may be an artifact from the exposure system or experimental protocol. Thus the outcome and significance of an investigation may often depend on the knowledge and experience of the investigator or investigators in the interdisciplinary team.

Because a single study seldom provides definitive evidence for or against a biological response, several independent studies are needed to arrive at a convincing answer to

the health effect question or at a statistically significant association. It should be noted that the scientific approach and acquisition of scientific knowledge are predicated on replication and confirmation by repeated experimentation.

Antennas as an essential component of exposure systems have played important roles in the research and development conducted to improve our understanding of RF and ELF interactions with biological systems. In addition to conventional antennas, a large variety of innovative approaches have been applied to the design and characterization of antennas and exposure systems used for biological research. It is emphasized that antennas are unique and integral parts of given exposure systems; they must be considered as a whole and characterized as such for the specific application or experimental scenario.

Following discussions of the salient features of *in vitro* and *in vivo* exposure systems, the antennas described and characterized in this chapter include ELF coils, TEM chambers, rectangular waveguides, radial transmission lines, wire patch cells, and coplanar waveguides for *in vitro* experiments under controlled laboratory conditions. Antennas for *in vivo* experiments can be divided into localized and partial-body exposures according to the protocols. Loop antennas and carousels are often used for experiments where only a portion of the body is subjected to EM field exposure. Antennas described for exposure of the whole body consist of ELF coils, radial waveguides, and TEM chambers for groups of mice and rats undergoing prolonged or lifelong exposures.

ACKNOWLEDGMENTS

The authors wish to thank Micaela Liberti and Alessandra Paffi for their valuable help in preparing the manuscript for this chapter.

REFERENCES

1. J. Schuderer, T. Samaras, W. Oesch, D. Spät, and N. Kuster, High peak SAR exposure unit with tight exposure and environmental control for *in vitro* experiments at 1800 MHz, *IEEE Trans. Microwave Theory Tech.*, Vol. 52, No. 8, pp. 2057–2066, 2004.
2. C. K. Chou, K. W. Chan, J. A. McDougall, and A. W. Guy, Development of a rat head exposure system for simulating human exposure to RF fields from handheld wireless telephones, *Bioelectromagnetics*, Vol. 20, pp. 75–92, 1999.
3. J. C. Lin, Dosimetric comparison between different possible quantities for limiting exposure in the RF band: rationale for the basic one and implications for guidelines, *Health Phys.*, Vol. 92, pp. 547–553, 2007.
4. N. Kuster and F. Schönborn, Recommended minimal requirements and development guidelines for exposure setups of bio-experiments addressing the health risk concern of wireless communications, *Bioelectromagnetics*, Vol. 21, pp. 508–514, 2000.
5. K. Courtney, J. C. Lin, A. W. Guy, and C. K. Chou, Microwave effect on rabbit superior cervical ganglion, *IEEE Trans. Microwave Theory Tech.*, Vol. 23, No. 10, pp. 809–813, 1975.
6. S. Arber and J. C. Lin, Microwave-induced changes in nerve cells: effects of modulation and temperature, *Bioelectromagnetics*, Vol. 6, pp. 257–270, 1985.
7. A. W. Guy, C. K. Chou, and J. A. McDougall, A quarter century of *in vitro* research: a new look at exposure methods, *Bioelectromagnetics*, Vol. 20, pp. 21–39, 1999.
8. F. Schönborn, K. Poković, M. Burkhardt, and N. Kuster, Basis for optimization of *in vitro* exposure apparatus for health hazard evaluations of mobile communications, *Bioelectromagnetics*, Vol. 22, pp. 547–559, 2001.

9. J. C. Lin, A. W. Guy, and C. C. Johnson, Power deposition in a spherical model of man exposed to 1–20 MHz EM fields, *IEEE Trans. Microwave Theory Tech.*, Vol. 21, No. 12, pp. 791–797, 1973.
10. J. Schuderer, D. Spät, T. Samaras, W. Oesch, and N. Kuster, *In vitro* exposure systems for RF exposures at 900 MHz, *IEEE Trans. Microwave Theory Tech*., Vol. 52, No. 8, pp. 2067–2075, 2004.
11. R. Merrit, C. Purcell, and G. Stroink, Uniform magnetic field produced by three, four, and five square coils, *Rev. Sci. Instrum.*, Vol. 54, No. 7, pp. 879–882, 1983.
12. K. Caputa and M. A. Stuchly, Computer controlled system for producing uniform magnetic fields and its application in biomedical research, *IEEE Trans. Instrum. Meas.*, Vol. 45, No. 3, pp. 701–709, 1996.
13. K. Yamazaki, H. Fujinami, T. Shigemitsu, and I. Nishimura, Low stray ELF magnetic field exposure system for *in vitro* study, *Bioelectromagnetics*, Vol. 21, pp. 75–83, 2000.
14. C. C. Davis, I. Barber, and M. Swicord, Food and drug administration low-level extremely-low-frequency magnetic field exposure facility, *Bioelectromagnetics*, Vol. 20, pp. 203–215, 1999.
15. G. Gottardi, P. Mesirca, C. Agostini, D. Remondini, and F. Bersani, A four coil exposure system (tetracoil) producing a highly uniform magnetic field, *Bioelectromagnetics*, Vol. 24, pp. 125–133, 2003.
16. J. Schuderer, W. Oesch, N. Felber, D. Spät, and N. Kuster, *In vitro* exposure apparatus for ELF magnetic fields, *Bioelectromagnetics*, Vol. 25, pp. 582–591, 2004.
17. M. Crawford, Generation of standard EM fields using TEM transmission cells, *IEEE Trans. Electromagn. Compat.*, Vol. 16, pp. 189–195, 1974.
18. M. Burkhardt, K. Poković, M. Gnos, T. Schmid, and N. Kuster, Numerical and experimental dosimetry of Petri dish exposure setups, *Bioelectromagnetics*, Vol. 17, pp. 483–493, 1996.
19. N. Nikoloski, J. Fröhlich, T. Samaras, J. Schuderer, and N. Kuster, Reevaluation and improved design of the TEM cell *in vitro* exposure unit for replication studies, *Bioelectromagnetics*, Vol. 26, pp. 215–224, 2005.
20. J. C. Lin, A new system for investigating nonthermal effect of microwaves on cells, in *Biological Effects of Electromagnetic Waves*, BRH/DHEW, Bethesda, MD, 1976, Vol. II, pp. 350–355.
21. M. J. Ottenbreit, J. C. Lin, S. Inoue, and W. D. Peterson, *In vitro* microwave effects on human neutrophil precursor cells, *Bioelectromagnetics*, Vol. 2, pp. 203–215, 1981.
22. P. Schönborn, K. Poković, A. M. Wobus, and N. Kuster, Design, optimization, realization, and analysis of an *in vitro* system for the exposure of embryonic stem cells at 1.71 GHz, *Bioelectromagnetics*, Vol. 21, pp. 372–384, 2000.
23. E. G. Moros, W. L. Straube, and W. F. Pickard, The radial transmission line as a broad-band shielded exposure system for microwave irradiation of large numbers of culture flasks, *Bioelectromagnetics*, Vol. 20, pp. 65–80, 1999.
24. Q. Balzano, C. K. Chou, R. Cicchetti, A. Faraone, and R. Y.S. Tay, An efficient RF exposure system with precise whole-body average SAR determination for *in vivo* animal studies at 900 MHz, *IEEE Trans. Microwave Theory Tech.*, Vol. 48, No. 11, pp. 2040–2049, 2000.
25. J. C. Lin and C. L. Wu, Scattering of microwaves by dielectric materials used in laboratory animal restrainers, *IEEE Trans. Microwave Theory Tech.*, Vol. 24, pp. 219–233, 1976.
26. J. C. Lin, H. J. Bassen, and C. L. Wu, Perturbation effects of animal restraining materials on microwave exposure, *IEEE Trans. Biomed. Eng.*, Vol. 24, pp. 80–83, 1977.
27. L. Laval, P. Leveque, and B. Jecko, A new *in vitro* exposure device for the mobile frequency of 900 MHz, *Bioelectromagnetics*, Vol. 21, pp. 255–263, 2000.

28. L. Ardoino, V. Lopresto, S. Mancini, R. Pinto, and G. A. Lovisolo, 1800 MHz *in vitro* exposure device for experimental studies on the effects of mobile communication systems, *Rad. Prot. Dos.*, Vol. 112, No. 3, pp. 419–428, 2004.
29. E. Neher and B. Sakman, *Single Channel Recording*, 2nd ed., Plenum Press, New York, 1998.
30. M. Liberti, F. Apollonio, A. Paffi, M. Pellegrino, and G. d'Inzeo, A coplanar-waveguide system for cell exposure during electrophysiological recordings, *IEEE Trans. Microwave Theory Tech.*, Vol. 52, No. 11, pp. 2521–2528, 2004.
31. K. C. Gupta, R. Garg, I. Bahl, and P. Barthia, *Microstrip Lines and Slotlines*, 2nd ed., Artech House, Boston, 1996.
32. C. C. Johnson and A. W. Guy, Nonionizing electromagnetic wave effects in biological materials and systems, *Proc. IEEE*, Vol. 60, pp. 692–718, 1972.
33. R. H. Lenox, O. P. Gandhi, J. L. Meyerhoff, and H. M. Grove, A microwave applicator for *in vivo* rapid inactivation of enzymes in the central nervous system, *IEEE Trans. Microwave Theory Tech.*, Vol. 24, pp. 58–61, 1976.
34. A. W. Guy and C. K. Chou, Effects of high-intensity microwave pulse exposure on rat brain, *Radio Sci.*, Vol. 17, pp. 169S–178S, 1982.
35. J. C. Lin and M. F. Lin, Microwave hyperthermia-induced blood–brain barrier alterations, *Radiation Res.*, Vol. 89, pp. 77–87, 1982.
36. A. P. Sanders and W. T. Joines, The effects of hyperthermia and hyperthermia plus microwaves on rat brain energy metabolism. *Bioelectromagnetics*, Vol. 5, pp. 63–70, 1984.
37. P. O. Kramar, A. F. Emery, A. W. Guy, and J. C. Lin, The ocular effects of microwaves on hypothermic rabbits: a study of microwave cataractogenic mechanisms, in Tyler PE editor. *Biologic Effects of Nonionizing Radiation*, P. E. Tyler (Ed.), New York Academy of Science, New York, 1975, Vol. 247, pp. 155–165.
38. P. O. Kramar, C. Harris, A. F. Emery, and A. W. Guy, Acute microwave irradiation and cataract formation in rabbits and monkeys, *J. Microwave Power*, Vol. 13, pp. 239–249, 1976.
39. H. A. Kues, L. W. Hirst, G. A. Lutty, S. A. D'Anna, and G. R. Dunkelberger, Effects of GHz microwaves on primate corneal endothelium, *Bioelectromagnetics*, Vol. 6, pp. 177–188, 1985.
40. P. Leveque, C. Dale, B. Veyret, and J. Wiart, Dosimetric analysis of a 900-MHz rat head exposure system, *IEEE Trans. Microwave Theory Tech.*, Vol. 52, No. 8, pp. 2076–2083, 2004.
41. E. G. Moros, W. L. Straube, and W. F. Pickard, A compact shielded exposure system for the simultaneous long-term UHF irradiation of forty small mammals: I. Electromagnetic and environmental design, *Bioelectromagnetics*, Vol. 19, pp. 459–468, 1998.
42. F. Schönborn, K. Poković, and N. Kuster, Dosimetric analysis of the carousel setup for the exposure of rats at 1.62 GHz, *Bioelectromagnetics*, Vol. 25, pp. 16–26, 2004.
43. P. Corona, J. Ladbury, and G. Latmiral, Reverberation-chamber research—then and now: a review of early work and comparison with current understanding, *IEEE Trans. Electromagn. Compat.*, Vol. 44, No. 1, pp. 87–94, 2002.
44. V. W. Hansen, A. K. Bitz, and J. R. Streckert, RF exposure of biological systems in radial waveguides, *IEEE Trans. Electromagn. Compat.*, Vol. 41, No. 4, pp. 487–493, 1999.
45. L. Ardoino, V. Lopresto, S. Mancini, C. Marino, R. Pinto, and G. A. Lovisolo, A radio-frequency system for *in vivo* pilot experiments aimed at the studies on biological effects of electromagnetic fields, *Phys. Med. Biol.*, Vol. 50, pp. 3643–3654, 2005.
46. W. R. Smythe, *Static and Dynamic Electricity*, 2nd ed., Mc Graw-Hill, New York, 1950.

PART VI

METHODS OF ANALYSIS, MODELING, AND SIMULATION

CHAPTER 29

Antenna Modeling Using Integral Equations and the Method of Moments

ANDREW F. PETERSON

29.1 INTRODUCTION

There are three dominant numerical approaches for antenna analysis. Two of these are the *finite-element frequency-domain method* and the *finite-difference time-domain method*, which are discussed in companion chapters and involve the numerical solution of differential equations. In this chapter we focus on integral equation methods in the frequency domain. Electromagnetics problems have been formulated in terms of integral equations since Pocklington analyzed wave propagation along wires in 1897, although the numerical solution of these equations in the modern sense did not begin until the 1960s. Integral equations offer the advantage that they incorporate a Green's function that exactly accounts for wave propagation over any distance. Thus there is no need to truncate the computational domain by approximate means, as with the finite-element and finite-difference methods.

We briefly review the formulation of several integral equations and the implementation of the most common numerical techniques, with an emphasis on the limits of these approaches. In general, our goal is to provide the reader with guidance on the choice of technique and help point the reader to the location of commercial software or freeware that might do the job. Readers seeking additional information are encouraged to consult a number of textbooks encompassing the scope of this discipline [1–9]; we do not attempt to provide the in-depth consideration of the necessary details here for the readers to develop their own computer codes.

29.2 INTEGRAL EQUATIONS

Material in this section and the following section has been adapted from Ref. 10.

Consider a perfect electric conducting body in an infinite homogeneous environment of permittivity ε and permeability μ. The body is illuminated by a sinusoidal steady-state source of electromagnetic radiation, having radian frequency ω, and all the time-varying

Modern Antenna Handbook. Edited by Constantine A. Balanis

quantities are represented by phasors with suppressed time dependence $e^{j\omega t}$. Although the structure represents the surface of an antenna, we pose the problem as though the antenna is scattering the field of the feed source. In the absence of the antenna, the source would produce an electric field $\vec{E}^{\text{inc}}(x, y, z)$ and a magnetic field $\vec{H}^{\text{inc}}(x, y, z)$ throughout the surrounding space. (These are denoted the *incident* fields.) In the presence of the conducting antenna, the fields are perturbed from these to the *total* fields $\vec{E}^{\text{tot}}(x, y, z)$ and $\vec{H}^{\text{tot}}(x, y, z)$. The perturbation can be accounted for by the presence of equivalent induced currents on the antenna surface.

The induced surface current density can be expressed as $\vec{J}(s, t)$, where s and t are parametric variables on the surface. (Henceforth, all references to the variable t are intended to denote position, not time.) The current density, if treated as a source function that exists in the absence of the conducting antenna, produces the *scattered* fields $\vec{E}^{s}(x, y, z)$ and $\vec{H}^{s}(x, y, z)$. The various fields are related by

$$\vec{E}^{\text{inc}} + \vec{E}^{s} = \vec{E}^{\text{tot}} \tag{29.1}$$

$$\vec{H}^{\text{inc}} + \vec{H}^{s} = \vec{H}^{\text{tot}} \tag{29.2}$$

Since the scattered fields are produced in infinite homogeneous space by $\vec{J}$, they are readily determined using any of the standard source-field relations, such as

$$\vec{E}^{s} = \frac{\nabla(\nabla \cdot \vec{A}) + k^2 \vec{A}}{j\omega\varepsilon} \tag{29.3}$$

$$\vec{H}^{s} = \nabla \times \vec{A} \tag{29.4}$$

(as in Ref. 7, Eq. (29.4) is employed as an alternative to the more traditional $\vec{B} = \nabla \times \vec{A}$), where the wavenumber of the medium is given by $k = \omega\sqrt{\mu\varepsilon}$, and the magnetic vector potential function is

$$\vec{A}(x, y, z) = \iint_{\text{surface}} \vec{J}(s', t') \, \frac{e^{-jkR}}{4\pi R} ds' \, dt' \tag{29.5}$$

In Eq. (29.5), R is the distance from a point (s', t') on the surface to the point (x, y, z) where the field is evaluated. Throughout this chapter, primed coordinates are used to describe the "source" of the electromagnetic field (the current density), while unprimed coordinates denote the "observer" location where that field is evaluated.

The total fields in Eqs. (29.1) and (29.2) must satisfy the electromagnetic boundary conditions on the surface of the perfect electric antenna:

$$\hat{n} \times \vec{E}^{\text{tot}}|_{\text{surface}} = 0 \tag{29.6}$$

$$\hat{n} \times \vec{H}^{\text{tot}}|_{\text{surface}} = \vec{J} \tag{29.7}$$

where $\hat{n}$ is the outward normal unit vector, and in Eq. (29.7) the surface is approached from the exterior. By combining Eqs. (29.1), (29.3), and (29.6), we obtain

$$\hat{n} \times \vec{E}^{\text{inc}}|_{\text{surface}} = -\hat{n} \times \left. \frac{\nabla(\nabla \cdot \vec{A}) + k^2 \vec{A}}{j\omega\varepsilon} \right|_{\text{surface}} \tag{29.8}$$

This equation is one form of the *electric field integral equation* (EFIE). (While it is in fact an integrodifferential equation, the term "integral equation" is commonly used for simplicity.) The equation can be solved in principle for the surface current density $\vec{J}$ appearing within the magnetic vector potential $\vec{A}$.

By combining Eqs. (29.2), (29.4), and (29.7), we obtain the *magnetic field integral equation* (MFIE)

$$\hat{n} \times \vec{H}^{\text{inc}}|_{\text{surface}} = \vec{J} - \hat{n} \times (\nabla \times \vec{A})|_{\text{surface}} \tag{29.9}$$

Equation (29.9) is enforced in the limiting case from the exterior of the surface of the body. This equation may also be solved for $\vec{J}$. Once $\vec{J}$ is determined by a solution of Eq. (29.8) or (29.9), the fields and other observable quantities associated with the antenna can be determined by direct calculation.

The MFIE in Eq. (29.9) is only applicable to closed conducting bodies where Eq. (29.7) is a valid boundary condition, but the EFIE can be used for thin shells as well as solid structures where Eq. (29.6) holds. This restriction on the MFIE is due to the fact that Eq. (29.7) has been specialized to the situation where the magnetic field is zero on the inner side of the surface. The form of the MFIE in Eq. (29.9) is not valid if the surface is open and the field is nonzero on both sides of the surface. Most antenna analysis employs the EFIE, and we focus on that equation in the following.

An alternative form of the EFIE is the so-called mixed potential integral equation (MPIE), in which the EFIE of Eq. (29.8) is recast into the form

$$\hat{n} \times \vec{E}^{\text{inc}}|_{\text{surface}} = \hat{n} \times \{j\omega\mu\vec{A} + \nabla\Phi_e\}|_{\text{surface}} \tag{29.10}$$

where the electric scalar potential function is

$$\Phi_e(x, y, z) = \frac{1}{-j\omega\varepsilon} \iint_{\text{surface}} \{\nabla'_s \cdot \vec{J}\} \frac{e^{-jkR}}{4\pi R} ds' dt' \tag{29.11}$$

The MPIE is sometimes a more convenient form for implementation because it offers one less derivative outside the integrals. Since the divergence of the current density is proportional to the surface charge density,[†] and the charge density is the dominant source of the near-zone electric field, the scalar potential term in Eq. (29.10) is likely to contribute fields of greater magnitude than the vector potential to the equation. The accuracy of a numerical solution is highly dependent on the accuracy of the near fields; hence the representation of the charge density and the scalar potential function are important, and the MPIE draws attention to those terms.

While we do not consider aperture antennas in this discussion, we note that those antennas can be analyzed by similar integral equations that usually employ the equivalent magnetic current density as the primary unknown. Appropriate formulations may be found in the literature [2, 5–7].

[†]Note that the divergence of the current density must be treated as a generalized function, incorporating any Dirac delta functions that might arise in the process.

29.3 THE METHOD OF MOMENTS

The vector integral equations (29.8), (29.9), and (29.10) have the form

$$L\{\vec{\mathrm{J}}\} = \vec{\mathrm{V}} \tag{29.12}$$

where L denotes a linear vector operator, $\vec{\mathrm{V}}$ is the given excitation function, and the equality holds for tangential vector components over the antenna surface. To solve Eq. (29.12) numerically, it must be projected from the continuous infinite-dimensional space to a finite-dimensional subspace (or in less mathematical terms, converted into a matrix equation). This process is known as *discretization*. One procedure for accomplishing this discretization is known as the *method of moments* (MoM) [1]. As applied here, the MoM is essentially equivalent to the *boundary element method*, the *weighted residual method*, and the *Rayleigh–Ritz procedure*.

Our implementation of the MoM procedure involves approximating the quantity to be determined, $\vec{\mathrm{J}}$, by an expansion in linearly independent vector basis functions

$$\vec{\mathrm{J}}(s,t) \cong \sum_{n=1}^{N} I_n \vec{\mathrm{B}}_n(s,t) \tag{29.13}$$

The N coefficients $\{I_n\}$ in Eq. (29.13) become the unknowns to be determined. To obtain N linearly independent equations, both sides of Eq. (29.12) are multiplied (using a scalar or dot product) with suitable vector testing functions $\{\vec{\mathrm{T}}_m(s,t)\}$ and integrated over the surface of the antenna to obtain

$$\iint_{\text{surface}} \vec{\mathrm{T}}_m \cdot L\{\vec{\mathrm{J}}\}\, ds\, dt = \iint_{\text{surface}} \vec{\mathrm{T}}_m \cdot \vec{\mathrm{V}}\, ds\, dt, \quad m = 1,\ 2, \ldots, N \tag{29.14}$$

By enforcing the equation in this manner, the boundary condition associated with the integral equation (29.12) is imposed on average (or in a weak sense) throughout the domain of the testing function.

The equations in (29.14) can be organized into a matrix of the form $\mathbf{ZI} = \mathbf{V}$, where the $N \times N$ matrix $\mathbf{Z}$ has entries

$$Z_{mn} = \iint_{\text{surface}} \vec{\mathrm{T}}_m \cdot L\{\vec{\mathrm{B}}_n\}\, ds\, dt \tag{29.15}$$

and the $N \times 1$ column vector $\mathbf{V}$ has entries

$$V_m = \iint_{\text{surface}} \vec{\mathrm{T}}_m \cdot \vec{\mathrm{V}}\, ds\, dt \tag{29.16}$$

The numerical solution of the matrix equation yields the coefficients $\{I_n\}$.

As expressed in Eqs. (29.13)–(29.16), the basis and testing functions may be spread over the entire surface, or defined with a more limited domain of support. The use of entire-domain basis functions, which are spread over the entire surface, is usually limited to codes developed for specific geometries. Because they allow flexibility in the geometry, codes based on subsectional functions are applicable to a broad class of problems, and most of the available computer codes for antenna analysis fall into this category.

As noted in the preceding section, the near-zone electric fields are tightly coupled to the surface charge density, which can be obtained from the current density through

$$\rho_s = \frac{1}{-j\omega}\nabla_s \cdot \vec{\mathrm{J}} \tag{29.17}$$

Although the current density may be viewed as the primary unknown in the EFIE, the representation for the charge density must be considered when constructing a numerical solution. It has been observed that inconsistent representations for charge density are associated with poor accuracy in numerical solutions [11].

29.4 THIN-WIRE ANTENNAS

The formulation and numerical solution process is illustrated using a conceptually simple antenna: a linear dipole (Figure 29.1). The dipole is a cylinder aligned with the z-axis, with radius a and overall length $2h$. For the moment, we neglect the details of the feed structure, except to suppose that the feed produces an electric field $\vec{\mathrm{E}}^{\mathrm{inc}}$ that illuminates the cylindrical arms of the dipole. More precisely, $\vec{\mathrm{E}}^{\mathrm{inc}}$ is the field of the feed as it would exist *in the absence* of the metal arms of the dipole. The surface equivalence principle is invoked to remove the dipole arms from the problem, with their effect accounted for by an equivalent surface current density $\vec{\mathrm{J}}$ radiating in free space.

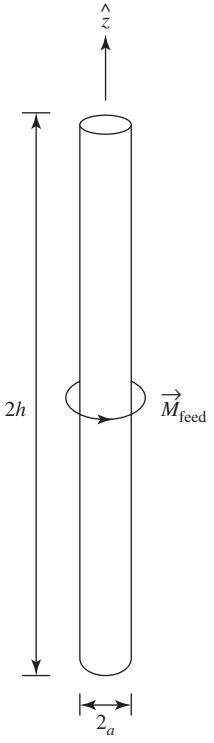


Figure 29.1 A linear dipole antenna.

Several simplifications are possible under the assumption that the cylinders are thin compared with the overall length ($a \ll 2h$) and the wavelength ($a \ll \lambda$). We also assume that the feed is ϕ-independent and produces no ϕ-component of electric field. These conditions imply that (1) there is no appreciable current flowing on the end caps of the cylinders and (2) the surface current density is ϕ-independent. (Equivalently, the first condition is also realized if the cylinders are hollow.) Thus the equivalent surface current density is only a function of z and can be expressed as

$$\vec{\mathrm{J}}(\phi, z) = \hat{\mathrm{z}}\ J_z(z) \tag{29.18}$$

The equivalent surface current is the primary unknown to be determined in the formulation. Once $\vec{\mathrm{J}}$ is determined, the radiation pattern, the directivity, and the input impedance of the antenna are readily found by secondary calculations.

The EFIE from Eq. (29.8) can be specialized to this situation to produce

$$\begin{aligned} E_z^{\text{inc}}(z)|_{\rho=a,\phi=0} &= \frac{-1}{j\omega\varepsilon}\hat{\mathrm{z}} \cdot \{\nabla(\nabla \cdot \vec{\mathrm{A}}) + k^2\vec{\mathrm{A}}\}|_{\rho=a,\phi=0,-h<z<h} \\ &= \frac{-1}{j\omega\varepsilon}\left\{\frac{\partial^2}{\partial z^2} + k^2\right\} A_z \\ &= \frac{-1}{j\omega\varepsilon}\left\{\frac{\partial^2}{\partial z^2} + k^2\right\} \int_{z'=-h}^{h} J_z(z') \int_{\phi'=0}^{2\pi} \frac{e^{-jkR}}{4\pi R} a\ d\phi'\ dz' \end{aligned} \tag{29.19}$$

where

$$R = \sqrt{(a - a\cos\phi')^2 + (-a\sin\phi')^2 + (z - z')^2} \tag{29.20}$$

After replacing the surface current density by the total current, using

$$I(z') = 2\pi a J_z(z') \tag{29.21}$$

Eq. (29.19) may be rewritten

$$E_z^{\text{inc}}(z) = \frac{-1}{j\omega\varepsilon}\left\{\frac{\partial^2}{\partial z^2} + k^2\right\} \int_{z'=-h}^{h} I(z')G(z - z')\ dz' \tag{29.22}$$

where

$$G(z - z') = \frac{1}{2\pi} \int_{\phi'=0}^{2\pi} \frac{e^{-jkR}}{4\pi R} d\phi' \tag{29.23}$$

and where R can be specialized to

$$R = \sqrt{4a^2\sin^2(\phi'/2) + (z - z')^2} \tag{29.24}$$

The equality in Eq. (29.22) holds over the original surface of the dipole ($\rho = a$, $-h < z < h$). Equation (29.22) has the form of Eq. (29.12) and can be solved in principle for $I(z)$.

The integral equation (29.22) can be generalized to bent or curved wires, wires with variable radius, wires loaded with lumped element resistances or reactances, and so on [3, 12–14]. The fundamental principles of the analysis of more general wire shapes do not differ substantially from that of the linear dipole, so the following discussion is limited to Eq. (29.22).

29.5 WIRE ANTENNA FEED MODELS

The field E_z^{inc} in Eq. (29.22) is the field of the antenna feed in the absence of the antenna itself. The most widely used feed models for wire antennas are the magnetic frill source [15] and the applied voltage source [12]. The frill feed, which consists of a magnetic current density associated with the TEM mode of a coaxial cable, is particularly well suited for modeling a monopole or other wire fed through a ground plane by a coaxial cable. For a ground plane at $z = 0$ and a coaxial aperture centered at the origin, that model produces an incident field of the form

$$E_z^{\text{inc}}(\rho, z) = \frac{V_0}{\ln(b/a)} \int_{\phi'=0}^{2\pi} \left\{ \frac{e^{-jkR_1}}{4\pi R_1} - \frac{e^{-jkR_2}}{4\pi R_2} \right\} d\phi' \tag{29.25}$$

where V_0 is the voltage across the coax aperture, a and b are the inner and outer dimensions of the aperture, respectively,

$$R_1 = \sqrt{\rho^2 + a^2 - 2\rho a \cos\phi' + z^2} \tag{29.26}$$

and

$$R_2 = \sqrt{\rho^2 + b^2 - 2\rho b \cos\phi' + z^2} \tag{29.27}$$

The applied voltage feed is not as physically valid as the frill feed model. It consists of the assumption that

$$\int T_m(z) E_z^{\text{inc}}(z)\, dz = \begin{cases} V_0, & \text{observer in feed region} \\ 0, & \text{otherwise} \end{cases} \tag{29.28}$$

where T_m is the testing function from Eq. (29.14) that is located in the feed region, and V_0 is the desired feed voltage. The various dipole feed models all produce highly localized fields near the feed point; hence Eq. (29.28) gives a reasonable first-order result without an explicit form for E_z^{inc}.

In general, a feed model that is closely tied to the actual feed structure is expected to produce more accurate results for input impedance. Because it is less physically valid, applied voltage feed results for input impedance are expected to be less accurate than those produced by a better feed model. However, far-field pattern results are relatively independent of feed models, and the applied voltage type of model is usually more than adequate for directivity calculations.

Other feed models have been proposed [3, 12, 13] but are not widely used. Practical feed structures such as a T-match feed [16] and various types of baluns [17] can sometimes be modeled by additional wire segments. In practice, there may be parasitic capacitance associated with antenna feed structures that is neglected in the typical feed model.

29.6 METHOD OF MOMENTS SOLUTION OF EQ. (29.22)

For generality, we limit our consideration to the use of subsectional basis functions, where each basis function is localized to a small region of support on the antenna surface.

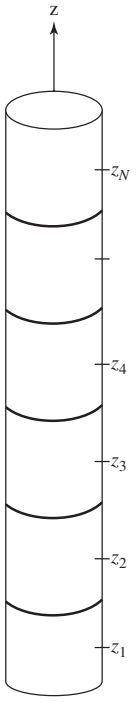


Figure 29.2 Division of the linear dipole into cells.

Suppose that the dipole in question is divided into cells along z (Figure 29.2). As in Eq. (29.13), the current $I(z)$ may be represented by a superposition of basis functions

$$I(z) = \sum_{n=1}^{N} I_n B_n(z) \tag{29.29}$$

After Eq. (29.29) is substituted into Eq. (29.22), the N coefficients $\{I_n\}$ become the unknowns to be determined.

While a large variety of subsectional basis function types have been studied, two widely used for wire antenna analysis are the *sinusoidal triangle* and the *three-term sinusoid*. Each sinusoidal triangle basis function straddles two cells and interpolates to the value of the current at the junction (Figure 29.3). If the cells are located in the intervals $-d < z < 0$ and $0 < z < d$, the basis function interpolating at the origin can be defined as

$$B_n(z) = \begin{cases} \dfrac{\sin(k[d - |z|])}{\sin(kd)}, & -d < z < d \\ 0, & \text{otherwise} \end{cases} \tag{29.30}$$

Figure 29.3, plotted for a cell dimension d equal to one-tenth the wavelength, illustrates that the sinusoidal triangle basis function is almost indistinguishable from an ordinary linear (polynomial or spline) triangle function. Additional basis functions with the functional dependence of Eq. (29.30) are centered at each of the cell junctions along the wire

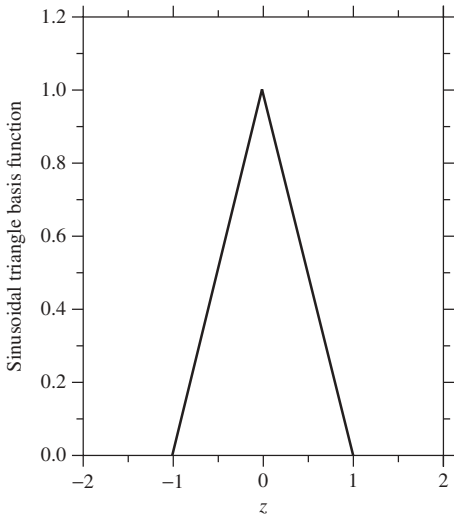


Figure 29.3 A single sinusoidal triangle basis function.

to provide the full representation. If the current is to vanish at the ends (for a hollow dipole or if the end effect is to be neglected), no basis function is located at the end of the extreme cells.

The sinusoidal triangle basis functions are usually used with sinusoidal triangle testing functions, identical to the basis functions, to produce an $N \times N$ system of equations.

The three-term sinusoid has the form

$$B_n(z) = \alpha_n + \beta_n \ \sin(kz) + \gamma_n \ \cos(kz) \tag{29.31}$$

where the coefficients $\{\alpha_n, \beta_n, \gamma_n\}$ are chosen to prescribe a specific behavior at cell junctions. In a region where the wire is smooth and of constant radius, these coefficients are usually selected to produce a sinusoidal spline behavior (Figure 29.4). Consider three adjacent cells, with a functional dependence consistent with Eq. (29.31) in each cell. Eight of the nine coefficients in these three cells can be obtained by enforcing continuity of the basis function and its first two derivatives at the two junctions, and the conditions that the basis function and its first derivative vanish at the outer edges of the three cells. For uniform cell sizes with cells located in the intervals $-\Delta < z < -\Delta/3$, $-\Delta/3 < z < \Delta/3$, and $\Delta/3 < z < \Delta$, the result is the spline function

$$S(z) = \begin{cases} S_0 \sin\left(\dfrac{k\Delta}{3}\right)\{1 - \cos(k[\Delta - |z|])\}, & \dfrac{\Delta}{3} < |z| < \Delta \\ S_0 \left\{\sin\left(\dfrac{k\Delta}{3}\right) - \sin(k\Delta) + \sin\left(\dfrac{2k\Delta}{3}\right)\cos(kz)\right\}, & |z| < \dfrac{\Delta}{3} \end{cases} \tag{29.32}$$

where, to normalize the function to unity at $z = 0$,

$$S_0 = \frac{1}{\sin\left(\dfrac{2k\Delta}{3}\right) + \sin\left(\dfrac{k\Delta}{3}\right) - \sin(k\Delta)} \tag{29.33}$$

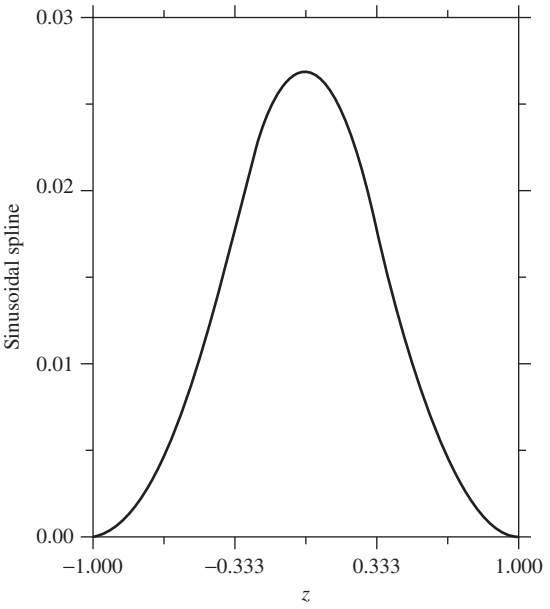


Figure 29.4 A sinusoidal spline basis function.

(Note that the normalization is not necessary, since the function defined in Eq. (29.32) does not interpolate to the value of the current at the middle of its domain, where adjacent basis functions also contribute a nonzero value.) The spline basis functions in Eq. (29.32) offer the advantage that the near fields they generate along the surface of the dipole are bounded, continuous functions. In contrast, near fields produced by the sinusoidal triangle basis functions, and many other types of basis functions, contain singularities at the cell junctions. Consequently, the smoother fields of the spline functions do a better job of satisfying the boundary condition in Eq. (29.6) inherent in the integral equation (except at the dipole ends).

Because the near fields are smooth, the testing functions usually employed in connection with sinusoidal spline basis functions are Dirac delta functions, located at the cell centers.

We note that at the ends of a wire, or at a junction between wires of different radii, the form of the spline in Eq. (29.32) is usually modified to incorporate the appropriate behavior for the current and charge densities [12]. In practice, the basis functions do not "hang off" the ends of the wire.

The dipole ends deserve special mention. The specific structure of the end caps determines the local behavior of the current and charge densities in that region. The dipole may be hollow, with a knife-edge geometry at the cylinder ends, or solid, with end caps that are flat, spherical, or some other shape. If end caps are to be modeled, then a special basis function can be incorporated that transitions from the expansion along the cylinder surface to an expansion of the current on the end cap (vanishing at the wire center). Such a basis function ordinarily incorporates the proper edge condition for the current and charge density if there is a discontinuity in the antenna surface at the transition

[18–20]. For dipoles, the proper treatment of the ends typically has little impact on the predicted input impedance for $a < 0.001\lambda$ but has greater impact as the radius increases.

A wide variety of other types of basis functions can be used and have been considered for wire antenna modeling, including ordinary polynomial functions such as Legendre, Lagrangian, Chebyshev, and Jacobi polynomials.

29.7 SINGULARITY IN THE GREEN'S FUNCTION

When the argument of the function G in Eq. (29.23) vanishes at $z = z'$, that function exhibits a logarithmic singularity. The integrand within G exhibits a $(1/R)$ singularity when $\phi' \to 0$ and $z \to z'$. The numerical evaluation of the MoM matrix entries must employ special procedures to properly handle these singularities, especially when high accuracy is desired. Discussions of general procedures for singular integrals are included in Ref. 1–9 and 21–23 and are briefly summarized in one of the following sections. A detailed consideration of their implementation is beyond the scope of this chapter.

In the distant past, it was suggested that Eq. (29.22) could be simplified by considering the current to be a filament located on the axis of the dipole instead of a cylinder on the dipole surface.[†] This approximation modifies the distance R to

$$R \cong R_a = \sqrt{a^2 + (z - z')^2} \tag{29.34}$$

Since R_a does not depend on ϕ', it follows that G in Eq. (29.23) can be approximated by

$$G_a(z - z') = \frac{e^{-jkR_a}}{4\pi R_a} \tag{29.35}$$

The function G_a is known as the "thin wire" kernel or the "reduced" kernel. It offers two computational advantages over G: the ϕ' integration in Eq. (29.23) is eliminated, reducing the necessary computations, and the kernel G_a is bounded and there is no longer a singularity at $z = z'$. Furthermore, since G_a is not singular, the derivatives in Eq. (29.22) can be pushed under the integral sign, to simplify the equation. The resulting equation has been widely used and often gives numerical results for the current and the input impedance that exhibit good agreement with numerical results from Eq. (29.22). Despite the fact that the overall results often agree, the reduced kernel G_a in Eq. (29.35) is usually *not* a good approximation to the exact G in Eq. (29.23) when $|z - z'| < 10a$. The success of the reduced kernel appears to stem from the explanation in the preceding footnote, namely, that an equally valid formulation of the problem can be obtained for a solid dipole by evaluating the boundary condition that the electric field is zero on the axis of the dipole. Thus the two approaches produce similar results even though the equations are substantially different. Despite the fact that as typically implemented the use of G_a is approximate and prevents results from converging, many codes use the approximation $G \cong G_a$.

[†]An equivalent and more justifiable result for a solid dipole can be obtained by leaving the current on the cylinder surface and moving the observer (unprimed coordinates) to the dipole axis. The feed field in Eq. (29.22) should then be evaluated at radius $\rho = 0$. (This is actually the approach taken in many codes.) A self-consistent formulation requires currents on the end caps to ensure that the fields inside the dipole are actually zero!

29.8 STEPS WITHIN A TYPICAL ANALYSIS

Once a user is reasonably familiar with the operation of a particular electromagnetic analysis software package, the steps involved in modeling an antenna proceed as follows.

The first step in the process is to describe the antenna geometry to the code. At its most detailed and tedious extreme, this may involve preparing an input file according to a prescribed format with data such as the coordinates of wire ends, number of segments along each wire, segment radius, frequency of operation, and soon. For solid surface modeling (to be discussed later), the data set may be more extensive and involve a lengthy list of coordinates on the surface, coupled with pointer arrays that associate cells on the surface with nodes (coordinates) and with edges defined between the nodes. In that manner, the precise location of each cell, and each basis function, is specified. In practice, users of commercial software tools should expect that the software will incorporate a graphical user interface that facilitates the description of the antenna geometry, and may allow the user to define the overall structure from a set of primitives and relatively few coordinates. In principle, a sophisticated software package should be capable of removing the burden of defining specific basis function locations from the user and using instead some form of adaptive refinement to guide the model creation internally. However, integral equation analysis codes with this level of sophistication are not widespread.

The second step in an antenna analysis is to define the excitation or antenna feed. Here the user is usually able to choose from a set of available feed models, including those discussed previously, and may also specify multiple feeds in the case of antenna array applications.

Once the model and excitation have been described, the method-of-moments analysis can be carried out, yielding coefficients of the basis functions for the current density. Most antenna codes will automatically calculate input impedance or admittance at the various feed points, perform a calculation of the antenna radiation pattern over some specified increments of view angles, and determine the total radiated power. It also may be possible to obtain copolarized and cross-polarized radiation patterns.

The data set encompassing the model of the antenna structure and the various results are stored in files and can be utilized for later postprocessing tasks and additional graphical processing. As discussed in a later section of this chapter, substantial physical insight can be obtained by mining these data in creative ways.

29.9 ACCURACY AND CONVERGENCE OF NUMERICAL SOLUTIONS

Two obvious questions arise in connection with a method-of-moments solution to an equation such as (29.22): *Do the results converge?* and *Do they converge to the correct answer?* Considerable effort has been directed toward these questions throughout the past four decades. While the answers to both questions can be "yes" under controlled conditions, it is important to understand that most electromagnetic analysis codes are unlikely to produce results that can stand up to scrutiny in this regard. (This is not to imply that the results are useless; we simply warn the reader that there is a limit as to how far things can be pushed.)

For results to converge, any approximation involved in the implementation of the MoM procedure must be well controlled, and the numerical precision of all the computations must be adequate. For instance, the use of a nonphysical feed model such as the applied

voltage source in Eq. (29.28) prevents numerical results from converging as the number of basis functions is increased, because the feed model itself changes as the cell sizes change. The use of the approximation $G \cong G_a$ also prevents convergence in the absence of end cap currents. The required numerical precision must exceed the likely number of digits lost to rounding errors in the various computations; normally this can be estimated from the condition number of the system matrix. In practice, as the number of basis functions is pushed in an investigation of convergence, the associated matrix condition numbers can exceed 10^6. Under those conditions, if it is desired to see results for input impedance converge to 5 digits, it is necessary that the matrix entries be computed to at least 11 digits of accuracy and that those many digits be preserved in all computations. Ensuring 11 digits of accuracy in the integrals over the singular Green's functions over a wide range of parameters is not a trivial task. Furthermore, most computer codes developed for wire antenna problems were not created with that objective in mind.

To illustrate that numerical results can converge, Table 29.1 shows the input admittance for a linear dipole with flat end caps, obtained from a high-order MFIE formulation where numerical precision was tightly controlled [24].

To investigate whether or not the numerical results converge to the "correct" answer, one must first take into account the idealizations incorporated into Eq. (29.22), and how they differ from the physical realization of a manufactured antenna. These include the idealization of perfectly conducting metal, a perfect boundary shape, a symmetrical feed field that might ignore the presence of a balun and matching network, and the absence of other electromagnetic scatterers in the vicinity of the antenna. These issues have long been relevant for electromagnetic field analysis. The presence of corners (or any type of surface discontinuity) in the idealized model deserves special mention, since current and charge densities exhibit singular behavior at these locations that should be incorporated into the basis function representation. An inappropriate representation at a discontinuity may prevent the numerical results from converging to the correct answer.

Despite the fact that most computer codes in use for modeling wire antennas will not truly converge to the "correct" results due to the aforementioned issues, useful results can be obtained by employing a sufficient number of unknowns and following other user guidelines [12].

TABLE 29.1 Input Admittance of a Center-Fed Linear Dipole[a] with Length 0.5λ, Radius $a = 0.0625\lambda$, and Flat End Caps

Unknowns	Re{Y} (mS)	Im{Y} (mS)
120	9.9173	3.6594
240	9.9180	3.6596
360	9.9182	3.6597
480	9.9183	3.6597
600	9.9183	3.6597

[a] The feed is a magnetic frill having $b/a = 1.2$. MFIE formulation is solved with degree-11 polynomial basis functions and Dirac delta testing functions, with equal-sized cells along the cylinder. These data are used to illustrate the convergence of input admittance possible under controlled conditions.

Source: Adapted from Ref. 24.

29.10 EXAMPLE: MUTUAL IMPEDANCE MATRIX FOR THREE-ELEMENT ARRAY

To illustrate the use of numerical analysis tools, consider the design of a three-element endfire array. Suppose we have three center-fed half-wave dipole elements, each aligned parallel to the z-axis in the $y = 0$ plane, with the positions along the x-axis given by $x = 0$, $x = \lambda/3$, and $x = 2\lambda/3$, where λ is the wavelength. The dipoles have radius $a = 0.005\lambda$. To produce an ideal endfire pattern with the main beam along the $+x$ axis, the currents at the feed points should be $\{I_0 e^{j0^\circ}, I_0 e^{-j120^\circ}, I_0 e^{-j240^\circ}\}$. In a simple design that neglects the mutual coupling between the dipole elements, the desired currents would be obtained through the use of feed voltages of the same relative phases, namely, $\{V_0 e^{j0^\circ}, V_0 e^{-j120^\circ}, V_0 e^{-j240^\circ}\}$.

Figure 29.5 shows the pattern obtained using NEC-2 [12] to model the three-element array with feed voltages $\{1e^{j0^\circ}, 1e^{-j120^\circ}, 1e^{-j240^\circ}\}$. In this analysis, each dipole was represented with 21 equal-sized segments, with an applied voltage feed imposed on each center segment. However, because of mutual coupling between the elements, the dipole currents are perturbed from the desired values, and the pattern deviates considerably from the ideal endfire pattern. (The ideal pattern has nulls at 90°, 180°, and 270°.) The feed currents determined by the method-of-moments solution process, in milliamperes, are $I_1 = 10.7e^{-j37.4^\circ}$, $I_2 = 5.2e^{-j176.7^\circ}$, and $I_3 = 3.0e^{-j277.0^\circ}$. These exhibit a substantial magnitude variation and some phase deviation from the desired values.

To obtain the desired endfire pattern, the effect of mutual coupling must be taken into account. In other words, the feed voltages must be corrected to those that produce the proper feed currents. These voltages are those prescribed by the relation

$$\begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix} = \begin{bmatrix} Z_{11} & Z_{12} & Z_{13} \\ Z_{12} & Z_{22} & Z_{12} \\ Z_{13} & Z_{12} & Z_{11} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} \tag{29.36}$$

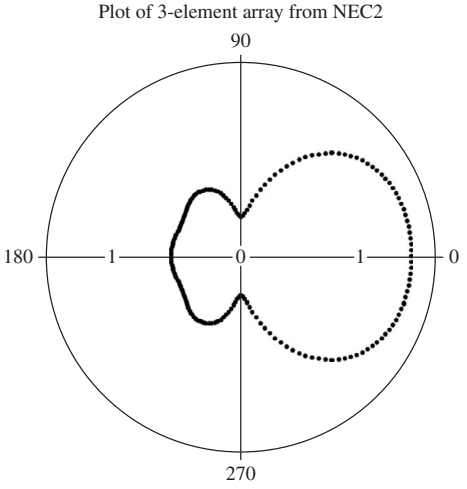


Figure 29.5 Pattern of three-element endfire array (from NEC-2), fed with ideal voltages that do not take mutual coupling into account.

where the currents are $\{I_0 e^{j0^\circ}, I_0 e^{-j120^\circ}, I_0 e^{-j240^\circ}\}$, and where the symmetry of the array has been taken into account to reduce the number of independent entries of the impedance matrix. The entries of the impedance matrix are not known *a priori* but may be determined using NEC-2.

A general entry in the impedance matrix, Z_{ij} in Eq. (29.36), is formally defined as "the voltage at feed i divided by the current at feed j, while dipole j is excited with all the other feeds open-circuited." In theory, a program such as NEC-2 can be used to determine the mutual impedances directly. This requires that the three-element array be driven by element j, while the other elements are open-circuited at their feed points. The difficulty with such an approach is the apparent need to model two open-circuited dipoles. NEC-2, like other wire antenna codes, is not well suited for treating open-circuited feeds, since the standard feed models typically involve a source imposed over a short-circuited wire. In addition, it is difficult to insert a "gap" between wire segments in such a manner so as to agree with the physical feed structure (which, in practice, almost never involves a gap between ideal cylinders). Finally, codes seldom have a provision for determining the voltage across such a gap. In actuality, analysis codes of this type are best suited for treating short-circuit feeds, not open-circuited feeds.

Instead of approaching the impedance matrix entries directly, we instead consider the inverse of Eq. (29.36),

$$\begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} = \begin{bmatrix} Y_{11} & Y_{12} & Y_{13} \\ Y_{12} & Y_{22} & Y_{12} \\ Y_{13} & Y_{12} & Y_{11} \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix} \tag{29.37}$$

and work directly with the admittance matrix. The formal definition of entry Y_{ij} in Eq. (29.37) is "the current at feed i divided by the voltage at feed j, with dipole j excited while all the other feeds are short-circuited." This definition is much easier to correlate with the result from an analysis code, since all that is required is that one element of the array be excited in the usual manner while the other two dipoles are modeled as continuous cylinders. Thus NEC-2 can be executed once with dipole #1 excited with voltage $V_1 = 1e^{j0^\circ}$, to determine Y_{11} (the current at center segment of dipole #1), Y_{12} (the current at the center segment of dipole #2), and Y_{13} (the current at the center segment of dipole #3). Figure 29.6 shows the input data file used with NEC-2. The results are obtained directly from the appropriate values of the current at each feed segment in the output data file. The code must be executed a second time with dipole #2 excited to determine Y_{22}. Finally, the admittance matrix in Eq. (29.37) may be inverted using standard linear algebra techniques to produce the impedance matrix in Eq. (29.36).

For the 21-segment dipole models described above, the numerical values obtained for the entries of the admittance and impedance matrix are given in Table 29.2. The

```
CM Example - array of 3 dipoles
CE Center-fed half-wave dipole with two parasitic dipoles
GW 1 21 0. 0. -0.25 0. 0. 0.25 0.005
GW 2 21 0.333 0. -0.25 0.333 0. 0.25 0.005
GW 3 21 0.667 0. -0.25 0.667 0. 0.25 0.005
GE
EX 0 1 11 0 1.0 0.0
FR 0 1 0 0 299.8
RP 0 5 0 1000 0. 0. 45. 0.
EN
```

Figure 29.6 Input data file used with NEC-2.

TABLE 29.2 Entries of the Mutual Admittance Matrix and Mutual Impedance Matrix for the Three-Element Endfire Array Example

Entry	Re{Y} (mS) or Re{Z} (Ω)	Im{Y} (mS) or Im{Z} (Ω)
Y_{11}	6.75	−2.98
Y_{12}	2.68	2.94
Y_{13}	−0.03	−0.60
Y_{22}	5.57	−1.43
Y_{23}	2.69	2.93
Z_{11}	94.2	52.1
Z_{12}	17.3	−53.7
Z_{13}	−37.2	6.1
Z_{22}	91.6	57.3
Z_{23}	17.1	−53.8

necessary feed voltages are obtained from Eq. (29.36) using those values, which yields $V_1 = 59.7e^{j28.8^\circ}$, $V_2 = 133.8e^{-j63.8^\circ}$, and $V_3 = 198.8e^{-j201.7^\circ}$. (These values can, of course, be scaled by any complex constant.) The overall antenna pattern obtained from NEC-2 using these feed voltages is shown in Figure 29.7 and exhibits good agreement with the ideal endfire pattern for an array of three point sources. The nulls in the pattern are not perfect, since, unlike ideal point sources, there is some phase variation in z along each dipole.

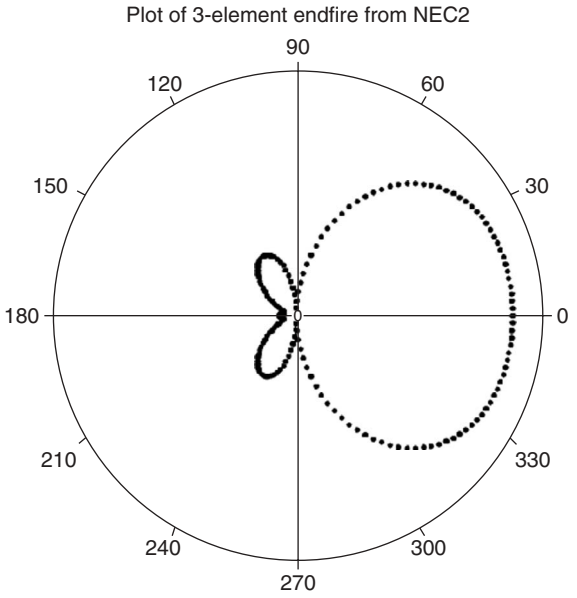


Figure 29.7 Pattern of three-element endfire array (from NEC-2), fed with voltages obtained by the mutual coupling analysis.

29.11 ANTENNAS CONFIGURED FROM ARBITRARY CONDUCTOR SHAPES

A number of antenna structures are constructed from conducting materials other than thin wires. Numerical techniques for treating these structures involve a representation for the surface in terms of flat or curved polygonal cell shapes (Figure 29.8). For illustration, we limit our consideration to flat-faceted triangular cell models. The surface current density is a vector quantity, and the most widely used basis functions defined on such surfaces are the vector Rao–Wilton–Glisson (RWG) triangular rooftop functions shown in Figure 29.9 [25]. Each function is centered on an edge in the surface model and straddles the two cells adjacent to that edge. That basis function provides a vector representation that interpolates to the surface current density flowing across the interior edge.

In the EFIE developed in Eqs. (29.1)–(29.8), the function $\vec{E}^{\text{inc}}$ is the field of the feed in the absence of the structure and is assumed known. The function $\vec{E}^{s}$ is the "scattered" electric field produced by the surface current $\vec{J}$, also in the absence of the structure, and

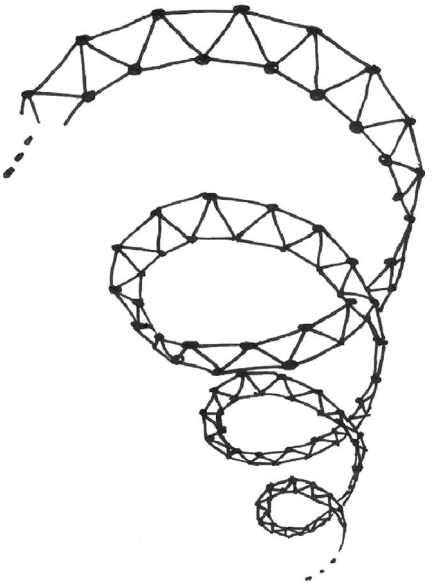


Figure 29.8 A portion of a conical spiral antenna modeled with flat triangular facets.

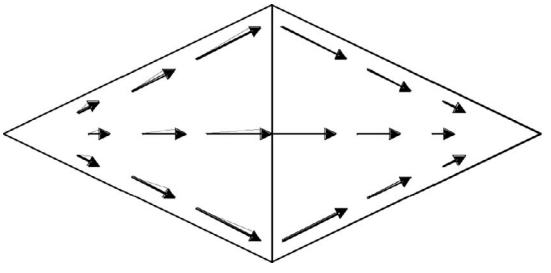


Figure 29.9 An RWG basis function straddling two triangular cells. The basis function interpolates to the component of current flowing across the central edge.

may be obtained by manipulating Eq. (29.3) into the form

$$\vec{\mathrm{E}}^s = -j\omega\mu \iint_{\text{surface}} \vec{\mathrm{J}}(s', t')G \; ds'dt' + \frac{1}{j\omega\varepsilon}\nabla \iint_{\text{surface}} \nabla' \cdot \vec{\mathrm{J}}(s', t')G \; ds' \; dt' \quad (29.38)$$

where G is now used to denote the free-space Green's function

$$G(R) = \frac{e^{-jkR}}{4\pi R} \quad (29.39)$$

$$R = \sqrt{[x(s,t) - x(s',t')]^2 + [y(s,t) - y(s',t')]^2 + [z(s,t) - z(s',t')]^2} \quad (29.40)$$

and the integration is performed over the antenna surface.

The surface current density is represented by the summation

$$\vec{\mathrm{J}}(s,t) = \sum_{n=1}^{N} I_n \vec{\mathrm{B}}_n(s,t) \quad (29.41)$$

where $\{\vec{\mathrm{B}}_n\}$ denote the RWG vector basis functions, and $\{I_n\}$ denote N complex-valued coefficients that henceforth are the unknowns to be determined.

The MoM procedure requires a testing function to provide a means for obtaining N linearly independent equations; the testing process converts the equation under consideration into a weak form, which can also be used to reduce the number of derivatives acting on the unknown function. If the general EFIE of Eq. (29.6) is multiplied (by scalar product) with a vector testing function $\vec{\mathrm{T}}(s, t)$, tangential to the surface, one obtains

$$\iint_{\text{surface}} \vec{\mathrm{T}} \cdot \vec{\mathrm{E}}^{\text{inc}} \; ds \; dt = -\iint_{\text{surface}} \vec{\mathrm{T}} \cdot \vec{\mathrm{E}}^s \; ds \; dt \quad (29.42)$$

With the EFIE, testing functions are often taken to be the same as the basis functions:

$$\vec{\mathrm{T}}_m(s,t) = \vec{\mathrm{B}}_m(s,t) \quad (29.43)$$

The equations can be organized into the form of a matrix equation

$$\mathbf{ZI} = \mathbf{V} \quad (29.44)$$

where $\mathbf{I}$ and $\mathbf{V}$ are N by 1 column vectors and $\mathbf{Z}$ is an N by N matrix. The entries of $\mathbf{I}$ are the coefficients $\{I_n\}$, while those of $\mathbf{V}$ are given by

$$V_m = \iint_{\text{surface}} \vec{\mathrm{T}}_m \cdot \vec{\mathrm{E}}^{\text{inc}} \; ds \; dt \quad (29.45)$$

The entries of $\mathbf{Z}$ can be manipulated into several mathematically equivalent forms, one of which is given by [10]

$$\begin{aligned} Z_{mn} = {} & j\omega\mu \iint_{s,t} \iint_{s',t'} \vec{\mathrm{T}}_m(s,t) \cdot \vec{\mathrm{B}}_n(s',t') \; G \; ds' \; dt' \; ds \; dt \\ & + \tfrac{1}{j\omega\varepsilon} \iint_{s,t} \nabla \cdot \vec{\mathrm{T}}_m \iint_{s',t'} \nabla' \cdot \vec{\mathrm{B}}_n \; G \; ds' \; dt' \; ds \; dt \end{aligned} \quad (29.46)$$

Equation (29.46) illustrates symmetry between the basis and testing functions and also highlights the obvious fact that in the absence of some other simplification the matrix entries involve four-dimensional integrals. Once these matrix entries have been computed, the matrix equation (29.44) may be solved to produce the coefficients $\{I_n\}$, after which any other quantity of interest may be obtained by integrating over the current density in Eq. (29.41).

When the testing and basis functions overlap a common cell in Eq. (29.46), there will generally be points where the function R within G vanishes. The resulting $1/R$ singularity complicates the evaluation of the integrals. There are several possible ways to evaluate the singular integral; for illustration we briefly review the use of a transformation due to Khayat and Wilton [23].

When the integrand is singular, the domain of the inner integral in Eq. (29.46) can be divided into three triangular subcells, each with the singularity at one corner. (The subdivision into three cells is not always optimal but illustrates the idea.) This approach reduces the integration to three integrals, each of the form

$$I = \int_{v=0}^{1} \int_{u=0}^{1-v} f(u, v) \frac{1}{\sqrt{u^2 + (1-v)^2}} du\ dv \tag{29.47}$$

where the function f incorporates the basis and testing functions, the numerator of the Green's functions, the dot product in Eq. (29.46), and any Jacobian arising from the cell transformation (or from a separate mapping associated with curved cells). The transformation of Khayat and Wilton [23] involves a change of variable from u to w, where

$$u = (1 - v) \sinh w \tag{29.48}$$

$$du = \sqrt{u^2 + (1-v)^2} dw \tag{29.49}$$

The limits of integration are modified, with the lower limit $u = 0$ replaced by $w = 0$, and the upper limit of $u = 1 - v$ replaced by $w = \text{arcsinh}(1)$. The transformation changes the triangular domain to a rectangular domain and produces

$$I = \int_{v=0}^{1} \int_{w=0}^{\text{arcsinh}(1)} f(u, v)\ dw\ dv \tag{29.50}$$

For flat cells, the factor in Eq. (29.49) exactly cancels the denominator of Eq. (29.47). While the denominator is not exactly canceled under more general conditions (curved cells or other mappings) the limiting behavior of the singularity *is* canceled. Equation (29.50) may be evaluated by adaptive quadrature.

Examples of surface-patch modeling of antennas abound. Makarov presents a tutorial-level discussion of modeling bow-tie, slot, and helical antennas using essentially the approach considered here [8]. The approach can be extended to more accurate analysis by incorporating curved cells [10] and higher-order vector basis and testing functions [26], although most available codes for conducting surfaces employ the RWG basis functions with flat faceted cell models.

29.12 PRINTED CIRCUIT ANTENNAS

The convenience of printed circuit manufacturing techniques motivates the use of microstrip patch, slot, and related printed circuit antennas. The analysis of these antennas is similar in most respects to that described in the preceding sections of this chapter, provided that the Green's function employed within the EFIE is modified to take into account the ground plane and substrate material.

The *free-space* Green's function used in Eqs. (29.19) and (29.39) is essentially the magnetic vector potential associated with a point source of current in infinite, homogeneous space. The magnetic vector potential of a tangentially directed source located on the surface of a grounded substrate of thickness h and constant relative permittivity ε_r may be obtained by appropriate analysis of that boundary value problem. Such an analysis yields [27]

$$\vec{\mathrm{A}}(x, y, h) = \iint \vec{\mathrm{J}}(x', y')\ G_A(x - x', y - y')\ dx'dy' \tag{29.51}$$

where the integration encompasses the support of the current density, and

$$G_A(x, y) = \frac{1}{2\pi} \int_0^\infty \frac{-jk_\rho J_0(k_\rho \rho)}{D_2} dk_\rho \tag{29.52}$$

If the mixed potential form of the EFIE in Eq. (29.10) is employed, the scalar potential function can be expressed as

$$\Phi_e(x, y, h) = \frac{1}{-j\omega\varepsilon} \iint \{\nabla_s' \cdot \vec{\mathrm{J}}\} G_\Phi(x - x', y - y')\ dx'dy' \tag{29.53}$$

where, for a grounded dielectric substrate,

$$G_\Phi(x, y) = \frac{1}{2\pi} \int_0^\infty \frac{-jk_\rho J_0(k_\rho \rho) D_3}{D_1 D_2} dk_\rho \tag{29.54}$$

In these expressions, $\rho = \sqrt{x^2 + y^2}$, $J_0(k_\rho \rho)$ denotes the zero-order Bessel function,

$$D_1 = \beta_d - j\beta\varepsilon_r \cot(\beta_d h) \tag{29.55}$$

$$D_2 = \beta - j\beta_d \cot(\beta_d h) \tag{29.56}$$

and

$$D_3 = \beta_d - j\beta \cot(\beta_d h) \tag{29.57}$$

where, if the various parameters k, k_d, and k_ρ are real valued,

$$\beta = \begin{cases} \sqrt{k^2 - k_\rho^2}, & k^2 > k_\rho^2 \\ -j\sqrt{k_\rho^2 - k^2}, & \text{otherwise} \end{cases} \tag{29.58}$$

$$\beta_d = \begin{cases} \sqrt{k_d^2 - k_\rho^2}, & k_d^2 > k_\rho^2 \\ -j\sqrt{k_\rho^2 - k_d^2}, & \text{otherwise} \end{cases} \tag{29.59}$$

The parameter k is the wavenumber in the region above the substrate, while k_d is the wavenumber in the substrate dielectric material.

The integrals in Eqs. (29.52) and (29.54) are a form of inverse Fourier transform often known by the name *Sommerfeld integrals*. They are usually evaluated by deforming the path of integration from the real k_ρ-axis in the complex k_ρ-plane to a path that is at least slightly above the real axis in order to avoid singularities at zeros of the denominators.[†] The integrals in Eqs. (29.52) and (29.54) are difficult to evaluate using numerical techniques due to the need to avoid zeros of the denominators and due to the growing oscillatory nature of the function $k_\rho J_0(k_\rho \rho)$ in the numerators! The complexity of their evaluation is the principal difficulty associated with numerical solutions of the EFIE for printed circuit antennas. A number of procedures have been proposed for their efficient evaluation [9, 27–33], and we refer the reader to the literature for additional details.

Several types of feeds are possible for planar antennas, including probe and aperture feeds through the ground plane, and microstrip feed lines on the surface of the substrate [33]. Feed models might consist of an impressed electric current source to represent a probe, or an impressed magnetic current source to represent an aperture or a voltage source. To excite a microstrip feed line, it is common practice to employ either an applied voltage feed (like that discussed earlier) or an impressed current feed, despite the fact that neither is closely related to a physical feed mechanism [34]. These feeds are typically located some distance away from the antenna, and a separate de-embedding procedure is used to remove the nonphysical effects if it is necessary to obtain an input impedance value [35].

The numerical solution of the EFIE or MPIE proceeds in accordance with the preceding examples. The current density on the metal surface of the printed circuit antenna is represented by vector basis functions as in Eq. (29.41). Testing functions are introduced to reduce the EFIE to a matrix equation of finite dimension. The solution of that system of equations yields the coefficients of the basis functions. Using slightly more general Green's functions than those in Eqs. (29.52) and (29.54), the fields radiated by the currents can be obtained at any point above the substrate or within the substrate, to produce the radiation pattern, input impedance, and other information [33].

Structures involving multiple substrate layers, superstrate layers, or an infinite dielectric half-space (perhaps representing antennas over earth) can be handled by extending the approach to incorporate Green's functions appropriate for the situation [31–33, 36, 37]. A rigorous implementation of Green's functions for these environments almost always involves Sommerfeld integrals.

29.13 EXAMPLE: MICROSTRIP PATCH ANALYSIS

To illustrate the numerical modeling process for a planar antenna, consider the design of a rectangular microstrip patch for use at 6.5 GHz. The patch will be located on a grounded

[†]The integration path in the complex plane may be further deformed and subdivided depending on the specific approach taken to evaluate the integrals. In that situation, k_ρ, β, and β_d become complex-valued quantities.

substrate of alumina ($\varepsilon_r = 9.8 - j0.004$) having thickness 0.4 mm. A pencil-and-paper design involves the determination of the resonant dimension, $d_x = \lambda/2$, and the nonresonant dimension d_y, often taken to be $d_y \cong 0.75\lambda$. Using the wavelength in alumina and neglecting fringing fields, we obtain $d_x \cong 7.4$ mm and $d_y \cong 11.0$ mm. When corrected for fringing fields, using standard expressions such as those in Appendix B of Ref. 33, d_x is reduced to $d_x = 7.19$ mm. Furthermore, by considering only the dominant (Eq. (29.10)) mode of the equivalent resonator, we estimate that a feed point located at a distance of 2.61 mm from the edge, along the centerline $d_y/2$, yields a reasonable match to a 50-Ω feed. This feed could be a coaxial cable penetrating the ground plane and substrate, or it could be a microstrip line inserted into the overall patch from the side. A 50-Ω microstrip feed line on this substrate is determined to have width $w = 0.39$ mm.

To verify and improve this preliminary design, we carry out a numerical analysis of the patch antenna using the program MultiSTRIP40 [38, 39]. This analysis program employs the EFIE with rectangular rooftop basis functions, which are similar to the RWG functions discussed earlier but straddle rectangular cells instead of triangular cells. The program requires as input the frequency range, constitutive parameters, the dimensions and geometry of the metallization pattern, and the cell dimensions over which the basis functions reside. The input data resides in an ASCII file easily edited by the user (Figure 29.10), which may also be modified through the program's user interface. Demonstration files provided with the software suggest the proper file format requirements.

For simplicity, we model the patch and the microstrip feedline with uniform, rectangular cells. (Improved accuracy for a given amount of computation might be obtained by the use of nonuniform cells, with smaller cells near the patch edges and the feed region.) To conveniently model the inserted feed line, the cell dimension along the nonresonant direction is selected to be $\Delta y = 0.39$ mm, or approximately $\Delta y = \lambda_d/38$, and 27 cells are therefore required to span the 11-mm dimension of the patch. (While Δy is smaller than necessary to model the relatively uniform current density along this dimension of the patch, it would actually be better to use multiple cells across the feed line to more accurately represent the feed current.) In the other direction, a cell dimension of $\Delta x = 0.218$ mm ($\Delta x = \lambda_d/68$) allows 33 uniform cells to span the resonant dimension of 7.19 mm, while 12 such cells approximate the estimated feed insertion distance of 2.61 mm. The resulting patch model, requiring the use of 1722 basis functions to represent the current density on the patch and feed line, is illustrated by the geometry definition in Figure 29.10.

A number of simulations were carried out using MultiSTRIP40, over a range of frequencies near 6.5 GHz. Plots of the current on the feed line and patch, as displayed by the MultiSTRIP40 user interface, were used to qualitatively assess the extent of the resonance. The most enhanced resonance appeared to occur at 6.74 GHz, a shift of 3.7% from the nominal design frequency. This suggests that the physical dimension of the patch be adjusted by a similar percentage, to $d_x = 7.45$ mm, to shift the resonance back to 6.5 GHz. Furthermore, the impedance match was poor over the entire range of frequencies, with the minimum reflection coefficient magnitude of 0.46 occurring at 6.74 GHz. (The poor impedance match is not unexpected, since the pencil-and-paper design neglected all modes except the dominant, which is a large approximation.) By adjusting the feed location to be closer to the edge, by a trial-and-error process, it was determined that an insertion distance of 1.31 mm (6 cells from the edge) produced a reasonable impedance match near the 6.74-GHz resonance. Near the resonant frequency,

```
+---------+---------+---------+---------+---------+---------+---------+
!                                                                     !
!             Microstrip antenna with an inserted feed                !
!                                                                     !
+---------+---------+---------+---------+---------+---------+---------+
!         !  START  !  STOP   ! INCREM. ! POINTS  ! (FREQUENCY IN GHZ)!
! FREQU.  ! 6.700000! 6.780000! 0.010000! 9.000000!---------+---------!
+---------+---------+---------+---------+---------+---------+---------+
!         !   NR.   ! DIELEC. ! LOSSES  ! HEIGHT  !  RE(ZS) ! IM(ZS)  !
!         !         !         ! (10E-3) ! (IN MM) !  (IN OHM * 10E-3) !
! LAYER   ! 1.000000! 9.800000!-4.000000! 0.400000! 0.000000! 0.000000!
! LAYER   ! 2.000000! 1.000000!-1.000000! 0.000000! 0.000000! 0.000000!
! LAYER   ! 3.000000! 1.000000!-1.000000! 0.000000! 0.000000! 0.000000!
! LAYER   ! 4.000000! 1.000000!-1.000000! 0.000000! 0.000000! 0.000000!
! LAYER   ! 5.000000! 1.000000!-1.000000! 0.000000! 0.000000! 0.000000!
+---------+---------+---------+---------+---------+---------+---------+
!         !   NR.   ! SX(MM)  ! SY(MM)  !  NO USE ! NO USE  ! NO USE  !
! SEGMENT !   1.    ! 0.218000! 0.390000!---------+---------+---------+
! GROUND  ! 0.000000! 0.000000!         !         !         !         !
! PRECIS  ! 1.000000!20.000000!         !         !         !         !
! ITERAT  ! 0.010000!   2.    !         !         !         !         !
+---------+---------+---------+---------+---------+---------+---------+
+---------+---------+---------+---------+---------+---------+---------+
!         ! NUMBER  ! NOT IN  ! NOT IN  ! NOT IN  ! NOT IN  ! NOT IN  !
!         ! OF STRU.! USE     ! USE     ! USE     ! USE     ! USE     !
! STRUKTU !    1.   !         !         !         !         !         !
+---------+---------+---------+---------+---------+---------+---------+
.............................................................
.............................................................
.......................#################################.....
.......................#################################.....
.......................#################################.....
.......................#################################.....
.......................#################################.....
.......................#################################.....
.......................#################################.....
.......................#################################.....
.......................#################################.....
.......................#################################.....
.......................#################################.....
.......................#################################.....
.............................###########################.....
.1######################################################.....
.............................###########################.....
.......................#################################.....
.......................#################################.....
.......................#################################.....
.......................#################################.....
.......................#################################.....
.......................#################################.....
.......................#################################.....
.......................#################################.....
.......................#################################.....
.......................#################################.....
.......................#################################.....
.......................#################################.....
.............................................................
.............................................................
+---------+---------+---------+---------+---------+---------+---------+
```

Figure 29.10 Input file used with MultiSTRIP40.

the reflection coefficients obtained from the simulation after the feed line was repositioned are as shown in Table 29.3.

The procedure outlined above should be repeated, after adjusting the physical dimensions, for a range of basis function densities and feed locations, in order to ensure that the computed results for this structure are relatively stable for the parameter ranges in use. A program such as MultiSTRIP40 makes it relatively easy to adjust parameters and repeatedly analyze a given antenna geometry.

TABLE 29.3 Reflection Coefficient Magnitude and Standing Wave Ratio (SWR) Obtained by Multi-STRIP40 for a Microstrip Patch[a]

f (GHz)	Magnitude of Reflection Coefficient	SWR
6.75	0.05	1.103
6.76	0.02	1.032
6.77	0.04	1.077

[a] The microstrip patch has dimensions of 7.19 mm by 11 mm, and is fed with an inset microstrip line terminated 1.31 mm past the patch edge. The patch model is described in Figure 29.10.

29.14 ANALYSIS OF LARGE ARRAYS AND PERIODIC STRUCTURES

Arrays of a few elements can be analyzed by the direct application of the preceding approaches, as illustrated by the earlier example of the three-element endfire array. For large arrays, or an antenna in the presence of a periodic structure (sometimes ground planes are periodically perforated; often radome coverings are constructed with periodic inclusions to perform filtering or reduce the scattering cross section of the antenna [40]), the direct application is usually prohibitive due to the large number of required unknowns. In those situations, approaches that incorporate periodic Green's functions into the integral equation might be applicable.

A periodic structure contains a lattice of identical cells. If illuminated by a uniform plane wave, the fields and currents along the structure are identical from cell to cell, with the exception of a phase shift [5, 7]. For example, the current density at some point on a structure with period a in the x direction and period b in the y direction can be expressed in terms of the current density within the principal unit cell as

$$\vec{\mathrm{J}}(x + ma, y + nb, z) = \vec{\mathrm{J}}(x, y, z)e^{-jk_x^{\mathrm{inc}}\, ma}e^{-jk_y^{\mathrm{inc}}\, nb} \tag{29.60}$$

where k_x^{inc} and k_y^{inc} are the phase shifts per cell imposed by the incident plane wave. Assuming that the structure consists entirely of conductors that are represented by the equivalent electric currents, the response of that structure is proportional to the convolution of the current on the principal unit cell with the periodic Green's function

$$G_p(x, y, z) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \frac{e^{-jkR_{mn}}}{4\pi R_{mn}} e^{-jk_x^{\mathrm{inc}}\, ma}e^{-jk_y^{\mathrm{inc}}\, nb} \tag{29.61}$$

where

$$R_{mn} = \sqrt{(x - ma)^2 + (y - nb)^2 + z^2} \tag{29.62}$$

When used with a periodic Green's function, the unknown current density is restricted to the principal unit cell of the lattice, thus constraining the total number of unknowns. The limitation imposed by the approach is primarily that the excitation must be a uniform plane wave; to treat a more general excitation it must first be decomposed into plane waves.

The application of the method of moments procedure to discretize the integral equation is essentially identical to the approaches described previously. However, when the entries of the moment-method matrix involve a periodic Green's function, such as Eq. (29.61), special procedures must be employed to accurately compute the slowly converging summations. Exponentially converging acceleration methods have been developed [7, 41] and are recommended over earlier methods that only provided algebraic convergence.

The numerical analysis of infinite phased arrays has been used to glean information about scan impedance and scan blindness [42, 43]. Similar techniques have been used to model ground planes with periodic perforations [44–46]. More general reviews of the analysis of periodic structures are also available [40, 47].

29.15 VISUALIZATION OF NUMERICAL RESULTS

The primary numerical data generated by an integral equation code are the coefficients of the basis functions. From that data, the code may automatically generate radiation pattern and input impedance information. Often, users limit their use of the results to those parameters. Sometimes additional graphical tools are used to illustrate certain behavior or help verify the likely correctness of a particular result. However, substantial insight into the operation of an antenna may be gained with additional processing and visualization tools.

As one example illustrating the benefit of additional processing, Nakano carried out detailed numerical modeling of the current on a number of axial-mode helices, using the method of moments for traditional thin-wire structures [48]. By studying the current distribution magnitude and phase, Nakano showed that within a few wavelengths of the feed point, the current on relatively long helical wires attenuates smoothly toward a minimum. Beyond this point, a slow wave with a fairly uniform amplitude is established along the structure. These results confirmed the presence of a slow wave along the helical structure, with a phase behavior appropriate for the helix pitch. Nakano also investigated tapering the ends of the helix to reduce the reflection from the end, in order to prevent a deterioration of the axial ratio [48]. In this manner, computational modeling followed by simple visualization was used to establish physical insight into the operation of axial-mode helix and provide guidance for helical antenna design.

The importance of visualization is stressed by Shaeffer, who states, "if all we do with our predictions is output data in the same form as measurements, then we are throwing away a treasure trove of potentially useful information" [49]. He likens computations to an iceberg, where the traditional results obtained from computations correspond to the tip of the iceberg and the additional, potentially useful data that is often ignored by the user corresponds to the massive submerged part. By analyzing the "submerged" part of the results, one is able to better understand specific physical mechanisms that underlie the performance of an antenna or other system and gain valuable insight into design trade-offs. For instance, inexpensive radiation images may be generated to graphically illustrate what portions of the antenna contribute to the far field in a given configuration [49]. In some cases, time-harmonic results can be converted into transient (time-domain) results for better visualization. Innovative plots of transient radiation from linear dipoles, as in Ref. 50, clearly enable a better understanding of the radiation mechanism. Graphical depictions of the Poynting vector in the vicinity of an antenna illustrate the precise flow of power and can also improve physical insight and motivate improved designs [51, 52].

29.16 ANTENNA DESIGN THROUGH NUMERICAL OPTIMIZATION

Computer platforms and numerical analysis software are now powerful enough that analysis tools can be incorporated within optimization algorithms that can be used to shape the geometry, change the material parameters, or otherwise vary the structure of an antenna to enhance its performance. Such an approach was postulated in the 1960s by Harrington [1], and occasionally used in antenna design in the 1970s [53, 54], but did not become commonplace for antenna design until the 1980s [51]. In recent years, optimization strategies based on local search techniques (gradient-based algorithms) have been replaced by global search approaches such as the genetic algorithm [55] and the particle swarm algorithm [56]. The former has been employed for a wide variety of antenna designs, including wire, slot, and patch elements, integrated printed circuit antennas, antennas loaded with parasitic resonators, and multibeam arrays. Many of these automated design procedures incorporate an integral equation analysis within the optimization loop.

Consider the design of a planar circuit antenna consisting of a particular metallization pattern on the substrate surface. If that surface is divided into small cells that are either to be covered by a conductor or not (the cells may overlap slightly to ensure a connection.), an approach such as the genetic algorithm may be used to determine the metallization pattern that maximizes gain in one or more frequency ranges, minimizes cross polarization, or otherwise optimizes some specification set. When the algorithm requires the evaluation of the specific objective function being minimized or maximized by a particular metallization pattern, an integral equation analysis is performed to yield the necessary parameters.

29.17 FAST NUMERICAL METHODS FOR LARGE OR COMPLEX PROBLEMS

Integral equation techniques produce a fully populated matrix equation that must be solved to obtain the antenna currents. The computational complexity associated with the generation of the matrix entries grows as N^2, where N is the number of unknowns, and the solution of such a system by standard LU factorization algorithms grows in proportion to N^3. For an individual antenna element in empty space, these computational requirements are usually not excessive. However, in many situations it is desired to model coupling between the antenna and its environment, and the inclusion of nearby structures may increase the required unknowns to a prohibitive number. Furthermore, modeling an electrically large antenna array or a reflector can be difficult.

During the past two decades, substantial progress has been made in "fast" methods that are applicable to integral equation problems when N grows large. Most of these methods employ iterative matrix solution algorithms as alternatives to direct methods of solution [57]. Iterative methods do not require that the system matrix be explicitly stored, and therefore various means of exploiting symmetry or otherwise compressing the storage can be implemented. During the iterative solution process, the system matrix must be repeatedly multiplied with various column vectors. It is possible to accelerate that matrix–vector multiplication, by reducing the computations associated with that operation from N^2 to something approaching $N \log N$. If the iterative algorithm converges quickly, the resulting computation can grow much slower than N^3. The most widely known of these acceleration procedures is the "fast multipole method" [58], although there are

several alternative techniques that offer comparable speed-ups [59, 60]. While those approaches are iterative in nature, progress has also been made in recent years on the acceleration of direct methods of solution for large problems [61].

29.18 A PERSPECTIVE ON NUMERICAL MODELING

There are obvious advantages of numerical analysis procedures: they can be used to simulate proposed designs over wide ranges of parameters, and they are usually far less expensive and offer a faster turnaround time than a program of experimental prototyping and measurement. If desired, numerical techniques allow trends in some performance metric or the sensitivity of some parameter to be studied as a function of any of the other parameters. When coupled with visualization tools, numerical solutions can be used to facilitate a better physical understanding of an antenna's operation. However, there is a principal drawback to numerical modeling: the results obtained by numerical simulation may be inaccurate or completely incorrect, and most of the simulation tools in use today provide little or no guidance that a user can rely on to determine the validity of a result.

Difficulties with accuracy arise for several reasons. First, there may be theoretical limitations to the underlying theory. The EFIE, for instance, is known to fail for closed bodies whose surfaces coincide with resonant cavities [7]. Some implementations of the EFIE will fail for objects that are electrically small [62]. Second, there are obvious approximations made to allow the physical structure under consideration to be modeled as a perfectly conducting metal, with a perfect shape, in a lossless environment, and so on. In many cases, the precise permittivity of a material (substrate, microwave absorber, etc.) may not be known and may not be uniform. Third, there may be numerical limitations to the calculations performed in specific cases (such as an unusually large amount of round-off error or high matrix condition number). The amount of such error may depend on user-supplied input data, such as the antenna geometric model. (A poor geometric model may lead to an ill-conditioned system matrix.) Fourth, there may be subtle bugs in the computer code that have not been discovered at the time an analysis is performed. Finally, there are limits on the size and complexity of a particular modeling task that may force a user to compromise on good modeling technique in order to obtain a result.

While the developers of an analysis code are expected to eradicate bugs, the user of such a tool should have independent methods to verify and validate specific numerical results. The user of a general-purpose code would be well advised to try that code on a problem for which there is a known solution, in order to ensure that the code is being used correctly, and gradually increase the complexity of structures being modeled while attempting to validate those results. Ideally, software should come with a variety of test cases "hard-wired" into the code so that the correct installation of the software can be verified, models of different scale can be attempted, and accuracy reference levels can be established [63]. Unfortunately, at the present time, such features, while possible, are not widely available within electromagnetic analysis codes.

An additional issue for consideration is that the environment around an antenna can have a substantial effect on the antenna radiation pattern and input impedance [64]. Most software packages provide an analysis of an antenna in infinite free space or over an idealized ground plane. The integration of the resulting antenna into its actual environment is usually left to the user, although there are codes available that incorporate scatterers [65].

29.19 SOME SOURCES OF SOFTWARE

We close this chapter by providing some links to references or websites that either provide software or indicate where shareware or commercial software packages may be obtained.

For wire antenna modeling, a widely used public-domain code is the Numerical Electromagnetics Code (NEC-2) [12]. A number of versions of this code (optimized for different processors, etc.), manuals, and ancillary software may be downloaded from the Internet [66]. There are numerous references that describe the use of NEC-2; for instance, Ref. 67 walks the reader through a number of modeling examples. An alternative to NEC-2 is the Antenna Scatterers Analysis Program (ASAP), also available through the Internet [68].

For arbitrary conducting structures, a number of software packages implement variations on the RWG basis solution of the EFIE discussed previously. One implementation is known as the Finite Element Radiation Model (FERM) [69]. Similar codes in the FORTRAN 77 or FORTRAN 90 languages have been developed under various U.S. government programs [70–72]. A MATLAB version of this type of code is also available [8].

Other approaches for the solution of the EFIE for wires and surfaces are possible. One is the Electromagnetic Surface Patch Code (ESP) developed at the Ohio State University [73, 74], which is designed to model thin wires, polygonal plates, and dielectric materials. The ESP uses basis functions of the sinusoidal triangle variety to represent the current densities. Another is the WIPL code [75], which employs polynomial functions as basis functions on surfaces and wires.

For the analysis of microstrip patch antennas or other antennas in the presence of layered media, there are several available commercial software packages, including EMPLAN [76]. The program MultiSTRIP used for the previous example is another [38, 39]. Some well established commercial packages include the Ansoft Designer [77], Agilent Momentum [78], and FEKO [79]. These products involve computational engines that incorporate techniques that are similar in most respects to those described earlier, coupled to fairly elaborate user interfaces for ease of use.

Reference 80 provides a fairly comprehensive website, maintained by the Electromagnetic Compatibility Laboratory of the University of Missouri at Rolla, with links to a variety of software tools, many of which are based on integral equation methods and suitable for antenna analysis. This site includes links to both commercial products and to shareware.

REFERENCES

1. R. F. Harrington, *Field Computation by Moment Methods*, IEEE Press, New York, 1993. (Reprint of the original 1968 edition.)
2. R. Mittra (Ed.), *Computer Techniques for Electromagnetics*. Hemisphere Publishing, New York, 1987. (Reprint of the original 1973 edition.)
3. J. Moore and R. Pizer (Eds.), *Moment Methods in Electromagnetics*, Research Studies Press (Wiley), Lechtworth, 1984.
4. N. Morita, N. Kumagai, and J. R. Mautz, *Integral Equation Methods for Electromagnetics*, Artech House, Boston, 1990. (Translation of the original 1987 edition.)
5. J. J. H. Wang, *Generalized Moment Methods in Electromagnetics—Formulation and Solution of Integral Equations*, John Wiley & Sons, Hoboken, NJ, 1991.

6. K. Umashankar and A. Taflove, *Computational Electromagnetics*, Artech House, Boston, 1993.
7. A. F. Peterson, S. L. Ray, and R. Mittra, *Computational Methods for Electromagnetics*, IEEE Press, New York, 1998.
8. S. N. Makarov, *Antenna and EM Modeling with MATLAB*, John Wiley & Sons, Hoboken, NJ, 2002.
9. D. B. Davidson, *Computational Electromagnetics for RF and Microwave Engineering*, Cambridge University Press, Cambridge, 2005.
10. A. F. Peterson, *Mapped Vector Basis Functions for Electromagnetic Integral Equations*, Morgan/Claypool, 2006.
11. L. Gurel, K. Sertel, and I. K. Sendur, On the choice of basis functions to model surface electric current densities in computational electromagnetics, *Radio Sci.*, Vol. 34, No. 6, pp. 1373–1387, November–December 1999.
12. G. J. Burke and A. J. Poggio, *Numerical Electromagnetics Code (NEC)—Method of Moments*, Technical Document 116, Naval Ocean System Center, San Diego, CA, January 1981.
13. B. D. Popovic, M. B. Dragovic, and A. R. Djordjevic, *Analysis and Synthesis of Wire Antennas*, John Wiley & Sons, Hoboken, NJ, 1982.
14. S. T. Li, J. C. Logan, and J. W. Rockway, *The MININEC System: Microcomputer Analysis of Wire Antennas*, Artech House, Boston, 1988.
15. L. L. Tsai, A numerical solution for the near and far fields of an annular ring of magnetic current, *IEEE Trans. Antennas Propag.*, Vol. 20, pp. 569–576, September 1972.
16. T. A. Milligan, *Modern Antenna Design*, John Wiley & Sons, Hoboken, NJ, 2005.
17. B. A. Munk, Baluns, etc., in *Antennas for All Applications*, J. D. Kraus and R. J. Marhefka (Eds.), McGraw-Hill, New York, 2002.
18. N. Stevens and L. Martens, An efficient method to calculate surface currents on a PEC cylinder with flat end caps, *Radio Sci.*, Vol. 38, No. 1, pp. 6.1–6.10, 2003.
19. M. M. Bibby, Impact of some discontinuities on the convergence of numerical methods in electromagnetics, *ACES J.*, Vol. 18, pp. 151–161, November 2003.
20. A. Heldring, E. Ubeta, and J. M. Rius, Efficient computation of the effect of wire ends in thin wire analysis, *IEEE Trans. Antennas Propag.*, Vol. 54, pp. 3034–3037, October 2006.
21. R. D. Graglia, On the numerical integration of the linear shape functions times the 3-D Green's function or its gradient on a plane triangle, *IEEE Trans. Antennas Propag.*, Vol. 41, pp. 1448–1455, October 1993.
22. S. Jarvenpaa, M. Taskinen, and P. Yla-Oijala, Singularity extraction technique for integral equation methods with higher order basis functions on plane triangles and tetrahedral, *Int. J. Numerical Methods Eng.*, Vol. 58, pp. 1149–1165, 2003.
23. M. A. Khayat and D. R. Wilton, Numerical evaluation of singular and near-singular potential integrals, *IEEE Trans. Antennas Propag.*, Vol. 53, pp. 3180–3190, October 2005.
24. A. F. Peterson and M. M. Bibby, High-order numerical solutions of the MFIE for the linear dipole, *IEEE Trans. Antennas Propag.*, Vol. 52, pp. 2684–2691, 2004.
25. S. M. Rao, D. R. Wilton, and A. W. Glisson, Electromagnetic scattering by surfaces of arbitrary shape, *IEEE Trans. Antennas Propag.*, Vol. 30, pp. 409–418, May 1982.
26. R. D. Graglia, D. R. Wilton, and A. F. Peterson, Higher order interpolatory vector bases for computational electromagnetics, *IEEE Trans. Antennas Propag.*, Vol. 45, No. 3, pp. 329–342, March 1997.
27. J. R. Mosig, Arbitrarily shaped microstrip structures and their analysis with a mixed potential integral equation, *IEEE Trans. Microwave Theory Tech.*, Vol. 36, pp. 314–323, February 1988.
28. F. E. Gardiol, Tips for evaluating Sommerfeld integrals, *IEEE Antennas Propag. Soc. Newslett.*, pp. 24–25, October 1986.

29. R. Janaswamy and F. E. Gardiol, Correction to "Tips for evaluating Sommerfeld integrals," *IEEE Antennas Propag. Soc. Newslett.*, pp. 18–19, February 1987.
30. J. R. Mosig, Integral equation technique, in *Numerical Techniques for Microwave and Millimeter-Wave Passive Structures*, T. Itoh (Ed.), John Wiley & Sons, Hoboken, NJ, 1989.
31. V. W. Hansen, *Numerical Solution of Antennas in Layered Media*, Research Studies Press, Taunton, 1989.
32. A. K. Bhattacharyya, *Electromagnetic Fields in Multilayered Structures*, Artech House, Boston, 1994.
33. R. Garg, P. Bhartia, I. Bahl, and A. Ittipiboon, *Microstrip Antenna Design Handbook*, Artech House, Boston, 2001.
34. G. V. Eleftheriades and J. R. Mosig, On the network characterization of planar passive circuits using the method of moments, *IEEE Trans. Microwave Theory Tech.*, Vol. 44, pp. 438–445, March 1996.
35. J. C. Rautio, A de-embedding algorithm for electromagnetics, *Int. J. Microwave Millimeter-Wave Computer-Aided Eng.*, Vol. 1, pp. 282–287, July 1991.
36. J. R. Wait, *Electromagnetic Waves in Stratified Media*, IEEE Press, New York, 1996. (Reprint of the original 1962 edition.)
37. J. R. Wait, Characteristics of antennas over lossy earth, in *Antenna Theory*, R. E. Collin and F. J. Zucker (Eds.), McGraw-Hill, New York, 1969.
38. G. Splitt, Improved numerical strategies for rigorous analysis of microstrip antennas, in *23rd European Microwave Conference*, Madrid, Spain, September 1993, pp. 354–356.
39. http://www.e-technik.fh-kiel.de/~splitt/html/mstrip.htm.
40. B. A. Munk, Frequency-selective surfaces and periodic structures in *Antennas for All Applications*, J. D. Kraus and R. J. Marhefka (Eds.), McGraw-Hill, New York, 2002.
41. K. E. Jordon, G. R. Richter, and P. Sheng, An efficient numerical evaluation of the Green's function for the Helmholtz operator on periodic structures, *J. Comp. Phys.*, Vol. 63, pp. 222–235, 1986.
42. D. M. Pozar and D. H Schaubert, Scan blindness in infinite phased arrays of printed dipoles, *IEEE Trans. Antennas Propag.*, Vol. 32, pp. 602–610, June 1984.
43. D. M. Pozar and D. H Schaubert, Analysis of an infinite array of rectangular microstrip patches with idealized probe feeds, *IEEE Trans. Antennas Propag.*, Vol. 32, pp. 1101–1107, October 1984.
44. B. Rubin and H. Bertoni, Reflection from a periodically perforated plane using a subsectional current approximation, *IEEE Trans. Antennas Propag.*, Vol. 31, pp. 829–836, November 1983.
45. G. Pan, X. Zhu, and B. Gilbert, Analysis of transmission lines of finite thickness above a periodically perforated ground plane at oblique orientations, *IEEE Trans. Microwave Theory Tech.*, Vol. 43, pp. 383–392, February 1995.
46. A. Mathis and A. F. Peterson, Efficient electromagnetic analysis of a doubly infinite array of rectangular apertures, *IEEE Trans. Microwave Theory Tech.*, Vol. 46, pp. 46–54, January 1998.
47. R. Mittra, C. H. Chan, and T. Cwik, Techniques for analyzing frequency selective surfaces—a review, *Proc. IEEE*, Vol. 76, pp. 1593–1615, December 1988.
48. H. Nakano, *Helical and Spiral Antennas—A Numerical Approach*, Research Studies Press, Letchworth, 1987.
49. J. Shaeffer, Electromagnetic visualization," *IEEE Potentials*, pp. 13–18, October/November 2000.
50. G. S. Smith, Teaching antenna radiation from a time-domain perspective, *Am. J. Phys.*, Vol. 69, pp. 288–300, March 2001.
51. F. M. Landstorfer and R. R. Sacher, *Optimization of Wire Antennas*, Research Studies Press, Letchworth, 1985.

52. E. K. Miller and F. J. Deadrick, Visualizing near-field energy flow and radiation, *IEEE Antennas Propag. Mag.*, Vol. 42, No. 6, pp. 46–54, December 2000.

53. B. J. Strait and K. Hirasawa, Constrained optimization of the gain of an array of thin wire antennas, *IEEE Trans. Antennas Propag.*, Vol. 20, pp. 665–666, September 1972.

54. T. K. Sarkar, An optimization program for linear arrays of parallel wires, *IEEE Trans. Antennas Propag.*, Vol. 22, pp. 631–632, July 1974.

55. Y. Rahmat-Samii and E. Michielssen, *Electromagnetic Optimization by Genetic Algorithms*, John Wiley & Sons, Hoboken, NJ, 1999.

56. J. Robinson and Y. Rahmat-Samii, Particle swarm optimization in electromagnetics, *IEEE Trans. Antennas Propag.*, Vol. 52, pp. 397–407, February 2004.

57. T. K. Sarkar (Ed.), *Application of the Conjugate Gradient Method to Electromagnetics and Signal Analysis*, Elsevier, New York, 1991.

58. R. Coifman, V. Rokhlin, and S. Wandzura, The fast multipole method for the wave equation: a pedestrian prescription, *IEEE Antennas Propag. Mag.*, Vol. 35, No. 3, pp. 7–12, June 1993.

59. W. C. Chew, J.-M. Jin, E. Michielssen, and J. Song, *Fast and Efficient Algorithms in Computational Electromagnetics*, Artech House, Boston, 2001.

60. E. Bleszynski, M. Bleszynski, and T. Jaroszewicz, AIM: adaptive integral method for solving large-scale electromagnetic scattering and radiation problems, *Radio Sci.*, Vol. 35, pp. 1225–1251, September–October 1996.

61. F. X. Canning and K. Rogovin, A universal matrix solver for integral-equation-based problems, *IEEE Antennas Propag. Mag.*, Vol. 45, No. 1, pp. 19–26, February 2003.

62. J.-S. Zhao and W. C. Chew, Integral equation solution of Maxwell's equations from zero frequency to microwave frequencies, *IEEE Trans. Antennas Propag.*, Vol. 48, pp. 1635–1645, October 2000.

63. P. J. Roache, Building PDE codes to be verifiable and validatable, *Computing Sci. Eng.*, Vol. 6, pp. 30–38, September/October 2004.

64. W. D. Burnside and R. J. Marhefka, Antennas on aircraft, ships, or any large, complex environment, in *Antenna Handbook: Theory, Applications, and Design*, Y. T. Lo and S. W. Lee (Eds.), Van Nostrand Reinhold, New York, 1988.

65. R. J. Marhefka, *Numerical Electromagnetic Code—Basic Scattering Code (NEC-BSC Version 4.2), User's Manual*, Technical Report, The Ohio State University, ElectroScience Laboratory, June 2000.

66. http://www.si-list.org/swindex2.html.

67. L. Sevgi, *Complex Electromagnetic Problems and Numerical Simulation Approaches*, John Wiley & Sons, Hoboken, NJ, 2003.

68. http://home.att.net/~ray.l.cross/asap/index.html.

69. S. Lee, D. A. Shnidman, and F. A. Lichauco, *Numerical Modeling of RCS and Antenna Problems*, Technical Report 705, Lincoln Laboratory/Massachusetts Institute of Technology, December 1987.

70. J. F. Shaeffer, *MOM3D Method of Moments Code—Theory Manual*, NASA Technical Report 189594, NAS1-18603, March 1992.

71. R. M. Sharpe, J. B. Grant, N. J. Champagne, W. A. Johnson, R. E. Jorgenson, D. R. Wilton, W. J. Brown, and J. W. Rockway, EIGER: electromagnetic interactions generalized, in *Digest of the 1997 IEEE Antennas and Propagation International Symposium*, Montreal, Quebec, July 1997, pp. 2366–2369.

72. J. M. Song, C. C. Lu, W. C. Chew, and S. W. Lee, Fast Illinois Solver Code (FISC), *IEEE Antennas Propag. Mag.*, Vol. 40, No. 3, pp. 27–34, June 1998.

73. E. H. Newman and D. M. Pozar, Electromagnetic modeling of composite wire and surface geometries, *IEEE Trans. Antennas Propag.*, Vol. 26, pp. 784–788, November 1978.

74. http://esl.eng.ohio-state.edu/codes/codes.php.
75. B. M. Kolundzija, J. S. Ognjanovic, and T. K. Sarkar, *WPIL-D Microwave—Software and User's Manual*, Artech House, Boston, 2005.
76. N. Kinayman and M. I. Askun, *EMPLAN: Electromagnetic Analysis of Printed Structures in Planarly Layered Media*, Software and User's Manual, Artech House, Boston, 2000.
77. http://www.ansoft.com/products.cfm.
78. http://eesof.tm.agilent.com/products/momentum_main.html.
79. http://www.feko.info/.
80. http://www.emclab.umr.edu/numer.html.

CHAPTER 30

Finite-Difference Time-Domain Method Applied to Antennas

GLENN S. SMITH and JAMES G. MALONEY

30.1 INTRODUCTION

The finite-difference time-domain (FDTD) method is a computational procedure for solving Maxwell's equations that is based on a clever algorithm first proposed by Kane S. Yee in 1966 [1]. When Yee proposed his algorithm, the method was computationally intensive in terms of both storage and run time, and only problems of very modest size could be solved using the best computational facilities (mainframe computers). Since then the power of computers has steadily increased, as has the popularity of the FDTD method. The first comprehensive analyses of practical antennas using the method were performed during the early 1990s, and today such computations are routinely performed on personal computers [2–6].

The purpose of this chapter is to introduce the reader to the rudiments of the FDTD method as applied to practical antennas. It is hoped that after delving into the chapter, the reader will understand the power and limitations of the method and be in a position to decide whether or not the FDTD method is suitable for analyzing his/her antenna problem. Because of the limited space, we cannot provide the details for implementing the method in a computer program. Readers interested in writing their own programs are referred to Ref. 7 for the details; others may wish to use one of the commercially available FDTD computer codes.

30.2 THE BASIC FDTD ALGORITHM

In the Yee algorithm, both space and time are discretized, with the increments in space for rectangular coordinates being Δx, Δy, Δz and the increment in time being Δt [8, 9].

Figure 30.1 is a schematic drawing showing a typical volume in which Maxwell's equations are to be solved. The volume is divided into $N = N_x N_y N_z$ unit cells each of

Modern Antenna Handbook. Edited by Constantine A. Balanis

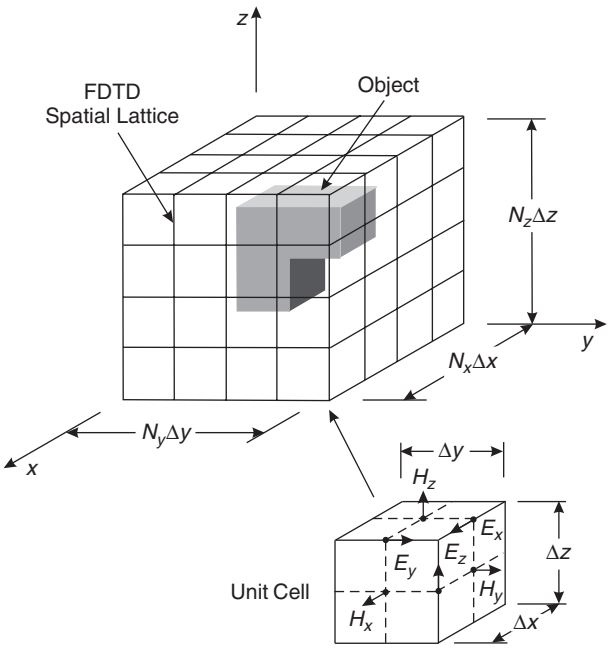


Figure 30.1 Schematic drawing showing the computational volume, FDTD spatial lattice, and unit cell.

volume $\Delta V = \Delta x \Delta y \Delta z$. The electromagnetic constitutive parameters ($\varepsilon = \varepsilon_r \varepsilon_0$, $\mu = \mu_r \mu_0$, σ) can vary from cell to cell, and they are used to define different objects within the volume. Here we use only simple materials with constant permittivity, permeability, and electrical conductivity. In the FDTD method there are techniques to handle more complicated materials, such as those with dispersive and anisotropic properties [9]. The six components of the electromagnetic field (E_x, E_y, E_z; H_x, H_y, H_z) are distributed over a unit cell (Yee cell) as shown in the inset. Note that all of the components are located at different points within the cell, and the components of **H** are displaced from those of **E** by one-half of a spatial increment, for example, $\Delta x/2$. Although not shown in the figure, the components of **H** are also evaluated at points displaced by one-half of a time increment, namely, $\Delta t/2$, from those of **E**.

The partial derivatives in Maxwell's equations are approximated by ratios of differences, for example,

$$\frac{\partial E_x}{\partial z} \approx \frac{\Delta E_x}{\Delta z}, \qquad \frac{\partial H_y}{\partial t} \approx \frac{\Delta H_y}{\Delta t} \tag{30.1}$$

For the spatial derivatives, the increment that occurs in the numerator is formed by differencing corresponding field components from adjacent unit cells, and for the temporal derivatives, it is formed by differencing field components from two adjacent time steps, for example, t and $t + \Delta t$. The discretized Maxwell's equations are arranged to form two sets of difference equations known collectively as "update equations." The first set of difference equations, which we call A, determines the change in the magnetic field, $\mathbf{H}(t + \Delta t/2) - \mathbf{H}(t - \Delta t/2)$, from the electric field at an intermediate time step, $\mathbf{E}(t)$, and

the second set of difference equations, which we call B, determines the change in the electric field, $\mathbf{E}(t + \Delta t) - \mathbf{E}(t)$, from the magnetic field at an intermediate time step, $\mathbf{H}(t + \Delta t/2)$.

At the start of the computation, we have the initial conditions: Throughout the computational volume, the electric field is known at time $t = 0$, and the magnetic field is known at the earlier time $t = -\Delta t/2$. The update equations A are then used with the initial conditions to obtain the magnetic field at time $t = \Delta t/2$. Next, the update equations B are used with the magnetic field that we just obtained at time $t = \Delta t/2$ and the electric field at time $t = 0$ to obtain the electric field at time $t = \Delta t$. This procedure of alternately applying update equations A and B to advance the solution in time is known as "marching-in-time" or "stepping-in-time." It is repeated until the electromagnetic field is known throughout the computational volume at the desired time $t = t_{\max} = N_t\ \Delta t$.

The choice for the increments of space and time (Δx, Δy, Δz, and Δt) is critical to the success of the algorithm, because their sizes determine how well the solution to the difference equations approximates the solution to Maxwell's equations. The spatial and temporal increments cannot be chosen independently; we can show that for convergence (as $\Delta x \to 0$, $\Delta t \to 0$, etc.) and stability of the algorithm the increments must satisfy the Courant–Friedrichs–Lewy condition, which for free space is

$$c\ \Delta t \sqrt{\frac{1}{\Delta x^2} + \frac{1}{\Delta y^2} + \frac{1}{\Delta z^2}} \leq 1 \tag{30.2}$$

For cubical cells, $\Delta x = \Delta y = \Delta z$, Eq. (30.2) becomes $S = c\Delta t/\Delta x \leq \sqrt{1/3}$, where S is referred to as the "Courant number," and a reasonable choice is $S = \frac{1}{2}$.

Additional restrictions for the spatial and temporal increments can only be obtained from knowledge of the variation of the field (the solution) in space and time. We basically have to make Δz and Δt in Eq. (30.1) small enough that the errors incurred by replacing the derivatives by the ratios of differences are acceptable. One obvious requirement is that the size of the spatial cells must be small enough to resolve all of the important structural features and the local field surrounding these features. Another requirement is that the error introduced by a phenomenon known as "numerical dispersion" must be negligible.

When there is numerical dispersion, a pulse that starts out with one shape ends up with a different shape after propagating through the FDTD lattice. Numerical dispersion is caused by the different frequency components of the pulse propagating through the lattice with different phase velocities. It can be quantified by considering a time-harmonic plane wave of angular frequency ω propagating in free space along one of the axes of the FDTD lattice, say, the x-axis. Assuming cubical cells, the numerical phase velocity, $\overline{v}_p$, for the wave, normalized to the speed of light in free space c, is

$$\frac{\overline{v}_p}{c} = \pi \left\{ N_\lambda \sin^{-1} \left[\frac{1}{S} \sin \left(\frac{\pi S}{N_\lambda} \right) \right] \right\}^{-1} \tag{30.3}$$

in which $N_\lambda = \lambda/\Delta x$ is the number of cells per wavelength [10]. Figure 30.2 is a graph of this equation showing the relative error in the phase velocity in percent (solid line) and a related quantity, the error in the phase per cell in degrees (dashed line). Note that the phase velocity is less than the speed of light, and that the error decreases monotonically with an increase in N_λ. For large N_λ (say, $N_\lambda > 10$), the error in the phase velocity is

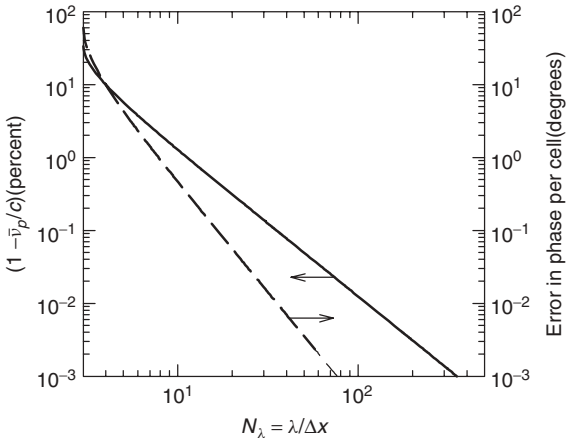


Figure 30.2 Numerical dispersion as a function of the number of cells per wavelength, N_λ, for a time-harmonic plane wave propagating along one of the axes of an FDTD lattice of cubical cells. Solid line, the relative error in the phase velocity in percent. Dashed line, the error in the phase per cell in degrees. $S = 0.5$.

approximately $(\pi^2/6)(1 - S^2)/N_\lambda^2$, so halving the cell size reduces the error by a factor of four. In theory, any desired accuracy can be obtained by increasing N_λ.

Ideally, given an electromagnetics problem, we would like to be able to estimate accurately the computational resources (computer memory and execution time) required to solve the problem using the FDTD method. Of course, this estimate is highly dependent on the problem and the computer being used. In practice, the estimate is usually made by comparing the requirements for the problem under consideration with those of a "benchmark problem" that has been run using a particular FDTD code on a particular computer. Even though the specific requirements for the algorithm are computer dependent, general rules for the scaling of the required memory and execution time with cell size are easily obtained.

Consider a computational volume that is a cube composed of cubical FDTD cells; then the total number of cells is $N = N_x^3$. Because only the most recent values of the electric and magnetic fields are needed at each step of the algorithm, the total storage required scales as N or N_x^3, that is, as the third power of the number of cells along the edge of the cubical volume. The simulation must be run for a time roughly proportional to that required for light to cross the volume, $t_{\max} \propto N_x\ \Delta x/c$. Thus the number of time steps required is $N_t = t_{\max}/\Delta t \propto N_x/\ S \propto N_x$. Now the execution time is proportional to the product of the number of cells with the number of times the cells must be updated: $N \times N_t \propto N_x^4$. The execution time scales as the fourth power of the number of cells along the edge of the cubical volume. Thus if we halve the dimensions of the cells, the storage will increase by a factor of 8, and the execution time will increase by a factor of 16.

30.3 FORMULATION OF THE ANTENNA PROBLEM IN THE FDTD METHOD

Antennas are customarily used in two states: transmission and reception. While the two states are related due to the reciprocity inherent in Maxwell's equations, not all quantities

for one state can be obtained from the other. Thus we must have two separate FDTD formulations for the antenna problem, one for the transmitting antenna and the other for the receiving antenna.

30.3.1 Transmitting Antenna

Figure 30.3a is a schematic drawing showing the basic elements involved in the FDTD analysis of a transmitting antenna. The figure is for a cross section through the computational volume, and the antenna is located near the center of the volume. The arrangement used to excite the antenna is shown in Figure 30.4a. The antenna is connected to the source by a transmission line (waveguide) of characteristic impedance R_0, and the source is matched to the characteristic impedance (there is no reflection for a wave entering the source). Throughout the chapter we assume that the characteristic impedance of a transmission line is real, a resistance. The specified excitation is the outward-propagating (incident) voltage wave $V_t^+(t)$ for a single mode at the reference plane in the line. Note that at this reference plane there is also a voltage $V_t^-(t)$ associated with an inward-propagating (reflected) wave.

The finite computational volume in Figure 30.3a is surrounded by an absorbing boundary. The objective for this boundary is to reproduce at its interior surface the same conditions for the electromagnetic field that would exist if the volume were infinite.

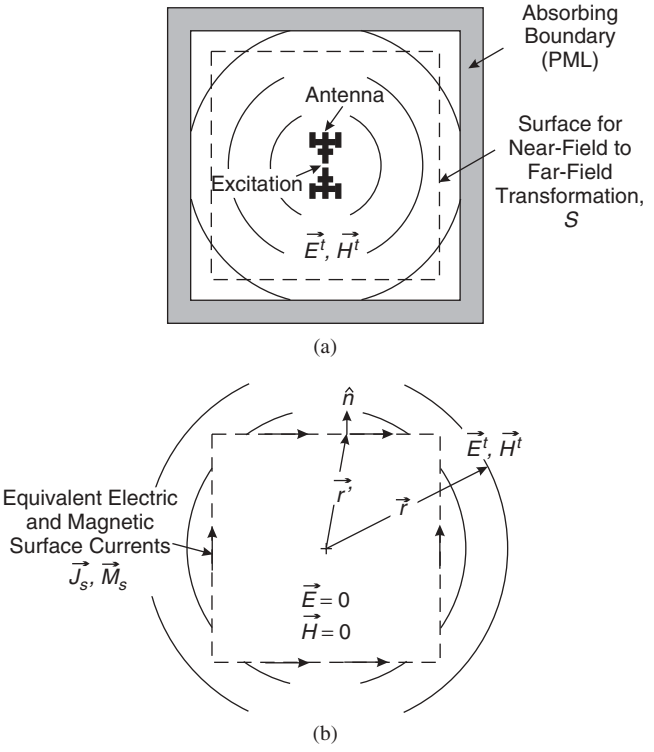


Figure 30.3 (a) Schematic drawing showing the basic elements involved in the FDTD analysis of a transmitting antenna. (b) Details for the near-field to far-field transformation.

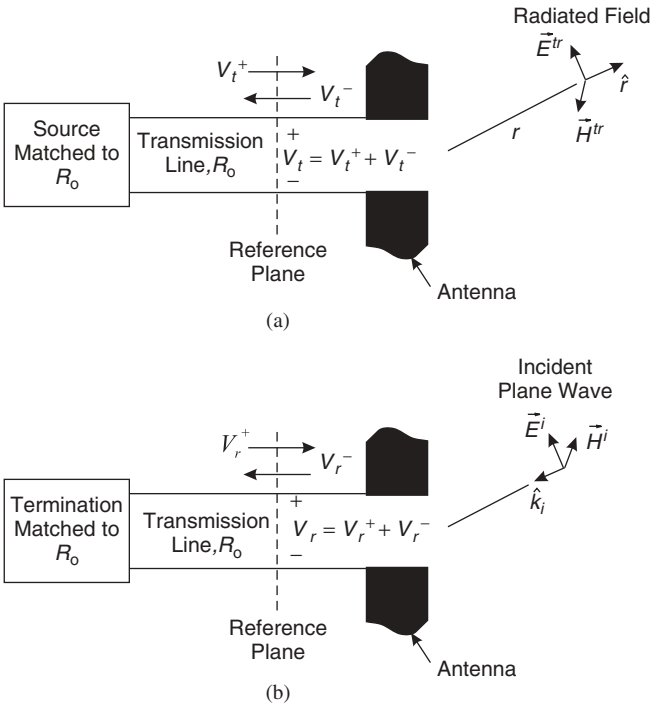


Figure 30.4 The details for the feed region of (a) the transmitting antenna and (b) the receiving antenna. The characteristic impedance of the transmission line is R_0, and the source and termination are matched to this impedance.

Stated differently, if we consider the electromagnetic field within the volume to be composed of a spectrum of plane waves, both outward propagating and evanescent, all of these waves should be absorbed without reflection by the boundary. At this time, the most effective absorbing boundaries are the perfectly matched layers (PMLs). Their implementation is discussed in the literature [11, 12].

The FDTD method provides the electromagnetic field for all lattice points within the finite computational volume. However, for many antenna applications, we would like to know the radiated or far-zone field, which is the field in the limit as the radial distance from the antenna becomes infinite ($r \to \infty$). This field can be obtained by applying what is known as a near-field to far-field (NFFF) transformation. For the implementation of this transformation, a closed surface S is placed around the antenna and inside the absorbing boundary. It is shown by the dashed line in Figure 30.3. The field ($\mathbf{E}^t$ and $\mathbf{H}^t$) on this surface is obtained for the time period of interest, and it is used to calculate the following electric and magnetic surface current densities:

$$\mathbf{J}_s(\mathbf{r}', t) = \hat{\mathbf{n}} \times \mathbf{H}^t(\mathbf{r}', t) \tag{30.4}$$

$$\mathbf{M}_s(\mathbf{r}', t) = -\hat{\mathbf{n}} \times \mathbf{E}^t(\mathbf{r}', t) \tag{30.5}$$

Here, as shown in Figure 30.3b, $\mathbf{r}'$ locates a point on the surface, and $\hat{\mathbf{n}}$ is the outward-pointing unit vector normal to the surface at that point. Outside the surface S,

these currents produce the same electromagnetic field as the transmitting antenna ($\mathbf{E}^t$, $\mathbf{H}^t$), and inside the surface they produce a null field ($\mathbf{E} = 0$, $\mathbf{H} = 0$).

At the position $\mathbf{r}$, the radiated or far-zone field (indicated by the additional superscript r) is obtained using these currents with a version of Huygens' principle for electromagnetic fields [8]:

$$\mathbf{E}^{tr}(\mathbf{r}, t) = \frac{\mu_0}{4\pi r} \oiint_S \left\{ \hat{\mathbf{r}} \times \hat{\mathbf{r}} \times \frac{\partial}{\partial t'} \left[\mathbf{J}_s(\mathbf{r}', t') \right] + \frac{1}{\zeta_0} \hat{\mathbf{r}} \times \frac{\partial}{\partial t'} \left[\mathbf{M}_s(\mathbf{r}', t') \right] \right\}_{t'=t_r} dS' \quad (30.6)$$

$$\mathbf{H}^{tr}(\mathbf{r}, t) = \frac{1}{\zeta_0} \hat{\mathbf{r}} \times \mathbf{E}^{tr}(\mathbf{r}, t) \quad (30.7)$$

in which the retarded time is

$$t_r = t - (r - \hat{\mathbf{r}} \cdot \mathbf{r}')/c \quad (30.8)$$

and $\zeta_0 = \sqrt{\mu_0/\varepsilon_0}$ is the wave impedance of free space.

In some situations, we may require the near field at points that are so far from the antenna that it is impractical to extend the computational volume to include these points. We can then use a near-field to near-field (NFNF) transformation to obtain these results: The FDTD analysis is performed for a volume, such as that shown in Figure 30.3a, and the field on the surface of the volume is transformed to obtain the near field outside the volume. Details for the NFNF transformation can be found in Refs. 13 and 14.

30.3.2 Receiving Antenna

Figure 30.5a is a schematic drawing showing the basic elements involved in the FDTD analysis of a receiving antenna. As for the transmitting antenna, the figure is for a cross section through the computational volume, and the finite computational volume is surrounded by an absorbing boundary. The excitation for the antenna is an incident, transverse electromagnetic (TEM) plane wave propagating in the direction $\hat{\mathbf{k}}_i$ with the field

$$\mathbf{E}^i(\mathbf{r}, t), \quad \mathbf{H}^i(\mathbf{r}, t) = \frac{1}{\zeta_0} \hat{\mathbf{k}}_i \times \mathbf{E}^i(\mathbf{r}, t) \quad (30.9)$$

Here, the vector $\mathbf{E}^i$ is transverse to $\hat{\mathbf{k}}_i$, namely, $\hat{\mathbf{k}}_i \cdot \mathbf{E}^i = 0$.

The closed surface S with outward-pointing unit normal vector $\hat{\mathbf{n}}$ is placed around the antenna and inside the absorbing boundary. And, as shown in Figure 30.5b, the following electric and magnetic surface current densities are placed on this surface to produce the incident field ($\mathbf{E}^i$, $\mathbf{H}^i$) inside the surface and a null field ($\mathbf{E} = 0$, $\mathbf{H} = 0$) outside the surface:

$$\mathbf{J}_s(\mathbf{r}, t) = -\hat{\mathbf{n}} \times \mathbf{H}^i(\mathbf{r}, t), \qquad \mathbf{M}_s(\mathbf{r}, t) = \hat{\mathbf{n}} \times \mathbf{E}^i(\mathbf{r}, t) \quad (30.10)$$

For the receiving antenna, we generally want to know the voltage produced in the antenna by the incident wave, and the arrangement used to accomplish this is shown in Figure 30.4b. The antenna is connected to the termination by a transmission line (waveguide) of characteristic impedance R_0, and the termination is matched to the characteristic impedance (there is no reflection for a wave entering the termination). The

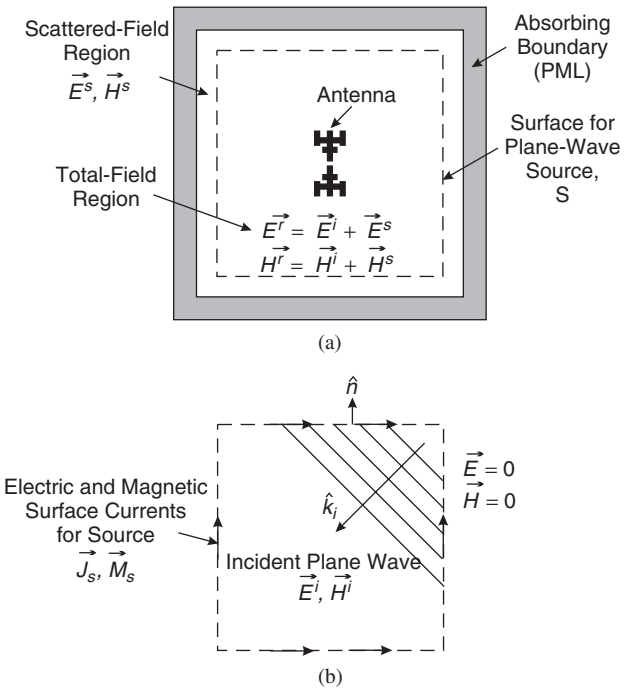


Figure 30.5 (a) Schematic drawing showing the basic elements involved in the FDTD analysis of a receiving antenna. (b) Details for the plane wave source.

desired response is the inward-propagating voltage wave $V_r^-(t)$ for a single mode at the reference plane in this line.

The scattered field is the field produced by the currents induced in the antenna by the incident field. Note from Figure 30.5a that the field inside the surface S is the total field, that is, the sum of the incident and scattered fields ($\mathbf{E}^r = \mathbf{E}^i + \mathbf{E}^s$, $\mathbf{H}^r = \mathbf{H}^i + \mathbf{H}^s$). However, the field outside the surface, in the region between the surface and the absorbing boundary, is only the scattered field ($\mathbf{E}^s$, $\mathbf{H}^s$). If we are interested in the scattering properties of the antenna, we can obtain them from knowledge of the field in this region. For example, the far-zone scattered field can be determined using a near-field to far-field transformation, as in the case for the transmitting antenna. The surface for the transformation must be placed between the surface for the plane wave source and the absorbing boundary.

30.3.3 Reciprocity

As mentioned earlier, some quantities for the states of transmission and reception are related through reciprocity. For example, when the arrangements shown in Figure 30.4 are used for the source and termination, the following relationship applies [15]:

$$V_t^+(t) * V_r^-(t) = \frac{2\pi R_0}{\zeta_0}\left[c\int_{t'=-\infty}^{t} \mathbf{E}^i(0,t')\,dt'\right] \cdot * \left[r\mathbf{E}^{tr}(-r\hat{\mathbf{k}}_i, t + r/c)\right] \qquad (30.11)$$

in which * indicates time convolution, and $\cdot$ * indicates the scalar product with time convolution. Here, the origin for the spherical coordinate system is centered on the antenna, as in Figure 30.4a, and the incident electric field $\mathbf{E}^i$ is evaluated at the origin ($r = 0$) of this system. The radiated electric field $\mathbf{E}^{tr}$ is evaluated at the radial distance r in the direction $(-\hat{\mathbf{k}}_i)$ from which the incident field arrives and at the time $t + r/c$. Sometimes this relationship can be used to eliminate the need for analyzing one of the two states (transmission or reception) when the other is known, or it can be used for verifying results from one state with results from the other.

30.3.4 Frequency Domain

The FDTD method is inherently a time-domain technique. When quantities are needed in the frequency domain (angular frequency ω), they are obtained using the Fourier transformation (discrete), which is indicated by $V(t) \leftrightarrow V(\omega)$. The quantities customarily used for evaluating the performance of an antenna in the frequency domain are determined from the transformed variables. For the transmitting antenna, the voltage reflection coefficient Γ_A and input impedance Z_A are

$$\Gamma_A(\omega) = \frac{V_t^-(\omega)}{V_t^+(\omega)} \tag{30.12}$$

$$Z_A(\omega) = R_0 \left[\frac{1 + \Gamma_A(\omega)}{1 - \Gamma_A(\omega)} \right] \tag{30.13}$$

and the realized gain G_{Rel} (gain including mismatch) and gain G in the direction $\hat{\mathbf{r}}$ are

$$G_{\text{Rel}}(\hat{\mathbf{r}}, \omega) = \frac{4\pi r^2 \hat{\mathbf{r}} \cdot \text{Re}\left[\mathbf{S}_c^{tr}(\mathbf{r}, \omega)\right]}{\text{Power available from source}} = \frac{4\pi R_0 r^2 |\mathbf{E}^{tr}(\mathbf{r}, \omega)|^2}{\zeta_0 |V_t^+(\omega)|^2} \tag{30.14}$$

$$G(\hat{\mathbf{r}}, \omega) = \frac{4\pi r^2 \hat{\mathbf{r}} \cdot \text{Re}\left[\mathbf{S}_c^{tr}(\mathbf{r}, \omega)\right]}{\text{Power accepted by antenna}} = \frac{1}{1 - |\Gamma_A(\omega)|^2} G_{\text{Rel}}(\hat{\mathbf{r}}, \omega) \tag{30.15}$$

in which $\mathbf{S}_c$ is the complex Poynting vector.

For the receiving antenna, the realized effective area $A_{\text{Rel}}(\hat{\mathbf{k}}_i, \omega)$ and the effective area $A_e(\hat{\mathbf{k}}_i, \omega)$ for an incident plane wave propagating in the direction $\hat{\mathbf{k}}_i$ are

$$A_{\text{Rel}}(\hat{\mathbf{k}}_i, \omega) = \frac{\text{Power accepted by termination}}{\hat{\mathbf{k}}_i \cdot \text{Re}\left[\mathbf{S}_c^i(\mathbf{r}, \omega)\right]} = \frac{\zeta_0}{R_0} \frac{|V_r^-(\omega)|^2}{|\mathbf{E}_i(\mathbf{r}, \omega)|^2} \tag{30.16}$$

and

$$A_e(\hat{\mathbf{k}}_i, \omega) = \frac{\text{Power available from antenna}}{\hat{\mathbf{k}}_i \cdot \text{Re}\left[\mathbf{S}_c^i(\mathbf{r}, \omega)\right]} = \frac{1}{1 - |\Gamma_A(\omega)|^2} A_{\text{Rel}}(\hat{\mathbf{k}}_i, \omega) \tag{30.17}$$

TABLE 30.1 Characteristics for Various Input Signals

Signal	Frequency for Spectrum 10% of Peak	Fractional Bandwidth at 10% points	Frequency for Spectrum 1% of Peak	Fractional Bandwidth at 1% points
Gaussian	$\omega = 2.15/\tau_p$		$\omega = 3.03/\tau_p$	
Differentiated Gaussian	$\omega = \begin{cases} 0.06/\tau_p \\ 2.76/\tau_p \end{cases}$	$\dfrac{\Delta\omega}{\omega_{\text{pk}}} = 2.70$	$\omega = \begin{cases} 0.01/\tau_p \\ 3.57/\tau_p \end{cases}$	$\dfrac{\Delta\omega}{\omega_{\text{pk}}} = 3.57$
Modulated sinusoid $\omega_0\ \tau_p \gg 1$	$\omega - \omega_0 = \pm\ 2.15/\tau_p$	$\dfrac{\Delta\omega}{\omega_0} = \dfrac{4.29}{\omega_0\tau_p}$	$\omega - \omega_0 = \pm\ 3.03/\tau_p$	$\dfrac{\Delta\omega}{\omega_0} = \dfrac{6.07}{\omega_0\tau_p}$

The gain and the effective area are related through reciprocity, Eq. (30.11); for a polarization match, that is, $|\mathbf{E}^{tr} \cdot \mathbf{E}^{i}| = |\mathbf{E}^{tr}||\mathbf{E}^{i}|$, we have[†]

$$A_{\text{Rel}}(\hat{\mathbf{k}}_i, \omega) = \frac{\lambda^2}{4\pi} G_{\text{Rel}}(\hat{\mathbf{r}} = -\hat{\mathbf{k}}_i, \omega), \qquad A_e(\hat{\mathbf{k}}_i, \omega) = \frac{\lambda^2}{4\pi} G(\hat{\mathbf{r}} = -\hat{\mathbf{k}}_i, \omega) \tag{30.18}$$

30.3.5 Input Signals

As we have seen, there is generally an input signal, $f(t)$, required for the antenna problem. For the transmitting antenna it is the incident voltage in the feeding transmission line, and for a receiving antenna it is the incident plane wave. The temporal behavior for some typical input signals used with the FDTD method is shown in Figure 30.6, and some of the characteristics for these signals are given in Table 30.1.

The choice for the input signal will depend on the particular application. For example, when we are determining the locations on an antenna at which reflections or radiation originate, a narrow pulse, such as the Gaussian pulse shown as a solid line in Figure 30.6a, may be an appropriate choice:

$$f(t) = \exp\left[-(t/\tau_p)^2/2\right], \qquad F(\omega) = \sqrt{2\pi}\,\tau_p \exp\left[-(\omega\tau_p)^2/2\right] \tag{30.19}$$

in which τ_p is the characteristic time. However, the spectrum for the Gaussian pulse contains significant low-frequency content (including dc), and this usually is not radiated by the antenna (the dc component never is). Thus the field near the antenna may take an unacceptably long time to settle when a Gaussian pulse is used.

When we are interested in the performance of an antenna over a band of frequencies, a pulsed input signal with zero mean is useful, followed by the Fourier transform to obtain the desired frequency-domain response. For this case, the appropriate choice for

[†]For a polarization match, the state of polarization for the incident plane wave in a particular direction (reception) is matched to the state of polarization for the radiated field in the same direction (transmission). For example, if the radiated electric field is linearly polarized, the electric field of the incident plane wave is linearly polarized and points in the same direction. If the radiated electric field is right-handed circularly polarized, the electric field of the incident plane wave is right-handed circularly polarized.

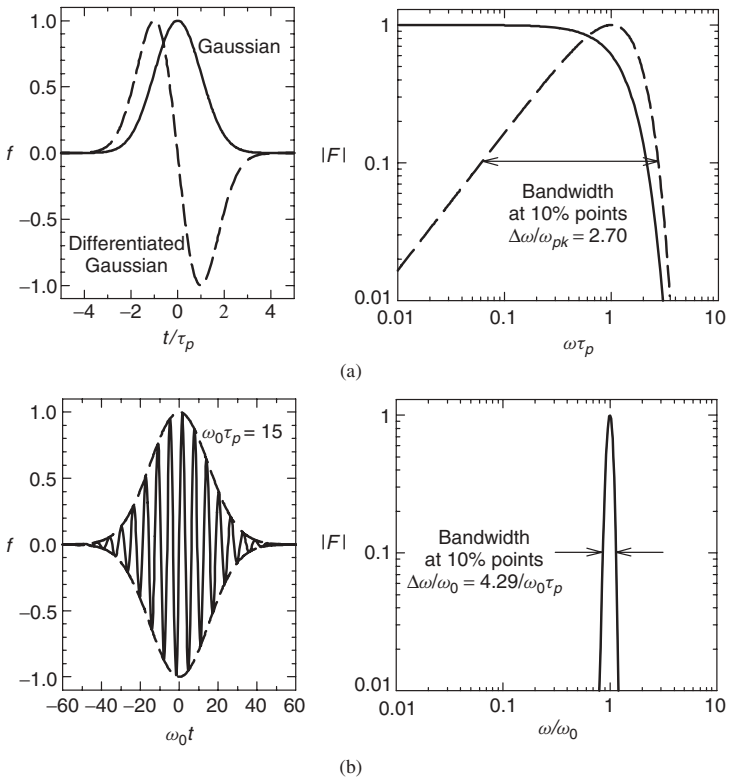


Figure 30.6 (a) The Gaussian pulse (solid line) and the differentiated Gaussian pulse (dashed line) and the magnitude of their Fourier transforms. (b) The sinusoid of frequency ω_0 amplitude modulated by a Gaussian pulse and the magnitude of its Fourier transform, $\omega_0\tau_p = 15$. All waveforms are normalized to have a maximum value of 1.0.

the input signal might be the differentiated Gaussian pulse shown as a dashed line in Figure 30.6a,

$$\begin{aligned} f(t) &= -\left(\frac{t}{\tau_p}\right)\exp\left\{-\left[(t/\tau_p)^2 - 1\right]/2\right\} \\ F(\omega) &= j\sqrt{2\pi}\omega\tau_p^2\exp\left\{-\left[(\omega\tau_p)^2 - 1\right]/2\right\} \end{aligned} \tag{30.20}$$

or the sinusoid of frequency ω_0 amplitude modulated by a Gaussian pulse shown in Figure 30.6b,

$$\begin{aligned} f(t) &= \exp\left[-(t/\tau_p)^2/2\right]\sin(\omega_0 t) \\ F(\omega) &= -j\sqrt{\pi/2}\tau_p\left(\exp\left\{-\left[(\omega-\omega_0)\tau_p\right]^2/2\right\} - \exp\left\{-\left[(\omega+\omega_0)\tau_p\right]^2/2\right\}\right) \end{aligned} \tag{30.21}$$

The differentiated Gaussian pulse has a rather large fractional bandwidth that is fixed; for example, the bandwidth associated with the points at which the spectrum is 10%

(−20 dB) of the maximum is $\Delta\ \omega/\omega_{\rm pk} \approx 2.70$, where $\omega_{\rm pk} = 1/\tau_p$ is the frequency at the peak. The modulated sinusoid has a variable fractional bandwidth that is controlled by the relative width of the modulating pulse, $\omega_0\tau_p$; for example, the bandwidth associated with the points at which the spectrum is 10% of the maximum is $\Delta\ \omega/\omega_0 \approx 4.29/\omega_0\tau_p$ (when $\omega_0\tau_p \gg 1$). For the case shown in Figure 30.6b, $\omega_0\tau_p = 15$, so the fractional bandwidth is $\Delta\ \omega/\omega_0 \approx 0.29$, which is much narrower than the fractional bandwidth for the differentiated Gaussian pulse shown in Figure 30.6a.

30.4 EXAMPLES OF THE USE OF THE METHOD FOR ANTENNA ANALYSIS

In the previous sections, we presented the rudiments of the FDTD method and described in general how the method is used to analyze an antenna for both transmission and reception. In this section, we show results obtained by applying the method to analyze particular antennas. These examples were chosen to illustrate specific issues that arise and must be dealt with when applying the method.

30.4.1 Cylindrical Monopole: Theoretical Model Versus Experimental Model

The ultimate test for any physical theory is how well its predictions agree with experimental measurements, and, of course, this is the case for electromagnetic theory when applied to antennas. One of the most important factors that affect the agreement is how closely the theoretical model for the antenna agrees with the experimental model. To examine this issue we consider the FDTD analysis of the cylindrical monopole, the image equivalent of the cylindrical dipole, which is arguably the most fundamental antenna.

The monopole antenna, shown in Figure 30.7a, is formed by extending the metallic center conductor of a coaxial line the distance h above an infinite metallic image plane [2, 8]. The dimensions of the transmission line, inner conductor radius a and outer conductor radius b, are chosen so that only the TEM mode propagates in the line for the signals of interest. The FDTD model for the transmitting monopole is shown in Figure 30.7b. All of the conductors in the model are perfect (perfect electric conductors, PECs), and the structure is surrounded by a PML, not shown in the figure [16]. Because of the rotational symmetry of the structure and the excitation, a two-dimensional cylindrical lattice (ρ, z) with the spatial increments $\Delta\rho$ and Δz is used in the FDTD analysis. A "one-way source" excites the coaxial line. This consists of the electric and magnetic surface currents

$$\mathbf{J}_s(\rho, t) = -\frac{V_t^+(t)}{2\pi R_0 \rho}\hat{\boldsymbol{\rho}}, \qquad \mathbf{M}_s(\rho, t) = -\frac{\zeta_0 V_t^+(t)}{2\pi R_0 \rho}\hat{\boldsymbol{\phi}} \tag{30.22}$$

on the plane $z = -l$ that produce the incident TEM voltage wave, V_t^+, above the source and a null field below the source. An absorbing boundary is placed at the bottom of the line. With this configuration, only the reflected TEM voltage wave, V_t^-, appears below the source, so it is easily determined. Note the similarity of this arrangement to the plane wave source used with the receiving/scattering antenna in Figure 30.5.

Figure 30.8 is a comparison of results from the FDTD simulation (solid line) with measurements (dots) made on an experimental model corresponding to the geometry

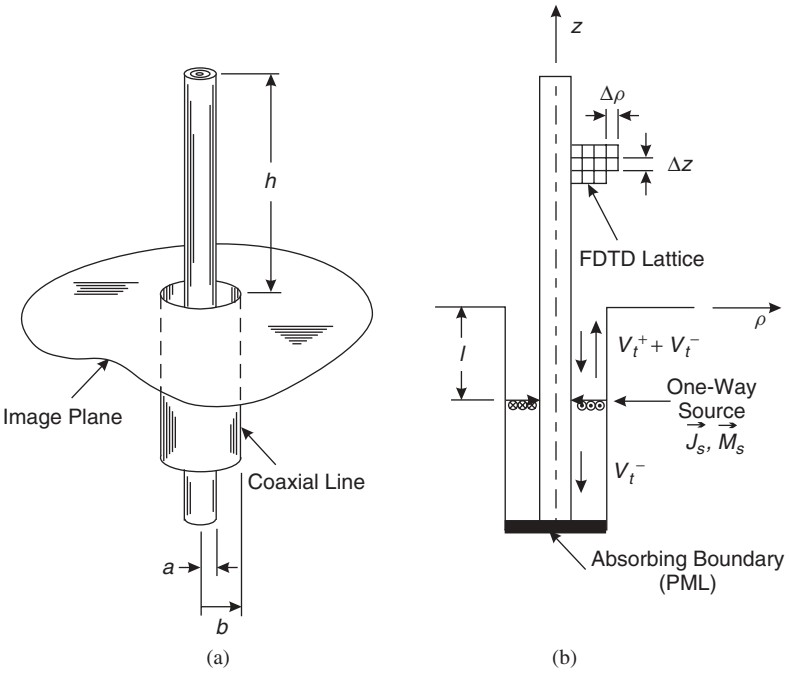


Figure 30.7 (a) Cylindrical monopole antenna fed through an image plane from a coaxial transmission line. (b) FDTD model for the cylindrical monopole antenna. The PML that surrounds the computational space is not shown.

in Figure 30.7a. The height of the monopole is $h = 5.0$ cm, and the dimensions of the coaxial line (precision line with APC-7 connector) are $a = 1.52$ mm and $b = 3.5$ mm, which gives a characteristic impedance of $R_0 = (\zeta_0/2\pi)\ln(b/a) = 50\ \Omega$. The excitation V_t^+ is a unit-amplitude Gaussian pulse in time, Eq. (30.19), with the characteristic time $\tau_p = 1.61 \times 10^{-1}\tau_a$, where $\tau_a = h/c$ is the time for light to travel the length of the monopole. In the experiment the reflections from the edges of the finite-sized image plane were windowed out in time.

Figure 30.8a is for the reflected voltage, V_t^-, in the transmission line, and Figure 30.8b is for the electric field on the image plane at the radial distance $\rho/h = 12.7$; both are shown as a function of the normalized time t/τ_a. In Figure 30.8a, we see the initial reflection of the incident pulse from the drive point (A), followed by its initial reflection from the open end of the monopole (B). As expected, these events are separated by roughly the time for light to make a round trip on the monopole, $(t_B - t_A)/\tau_a \approx 2$. Additional reflections of decreased amplitude occur each time the pulse encounters the drive point and the open end. In Figure 30.8b, we see that radiation occurs each time the pulse encounters the drive point or the open end of the monopole. As expected, the initial radiation from the drive point (A) is separated from the initial radiation from the open end of the monopole (B) by roughly the time for light to travel the length of the monopole, $(t_B - t_A)/\tau_a \approx 1$. The agreement of the theoretical and measured results is seen to be very good.

The FDTD method inherently provides information about the electromagnetic field within the computational volume over the entire period of the simulation. Only a small fraction of this information is used when investigating conventional antenna parameters,

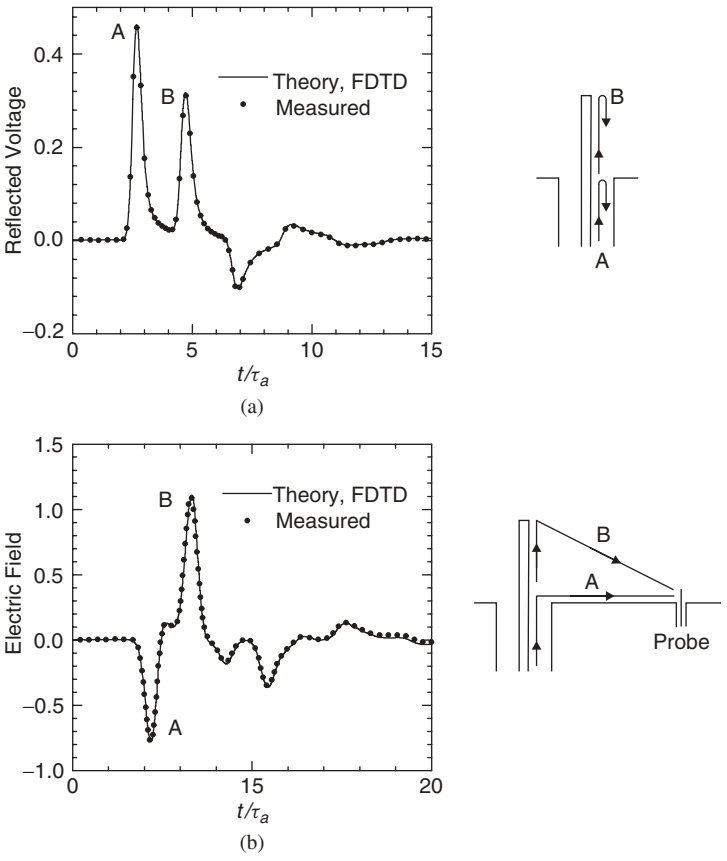


Figure 30.8 Comparision of theoretical and measured results for the cylindrical monopole antenna. (a) Reflected voltage in the coaxial line. (b) Electric field on the image plane at $\rho/h = 12.7$.

such as the results shown in Figure 30.8. Sometimes this additional information can be used to perform "numerical experiments" that improve our understanding of the radiation process for the antenna. This is illustrated in Figure 30.9, where we show the instantaneous Poynting vector in the region surrounding the monopole [17]. On the right-hand side of these figures, the logarithm of the magnitude of the Poynting vector, $|\mathbf{S}|$, is plotted on a gray scale. The intensity of the field increases as the shade goes from black to white, and the range for the values of $|\mathbf{S}|$ displayed is 10^4:1. On the left-hand side of these figures, the arrows indicate the direction of the Poynting vector, and the length of an arrow is proportional to the logarithm of $|\mathbf{S}|$. The excitation is a Gaussian voltage pulse with $\tau_p = 5.37 \times 10^{-2}\tau_a$. For this value of τ_p, about three nonoverlapping pulses fit along the length of the monopole, so the reflections associated with different points are separated and easily identified.

In Figure 30.9a, the pulse has just left the drive point and is traveling up the monopole. A spherical wavefront W_1 centered on the drive point has formed, and it is attached to the outward propagating pulses of charge/current on the monopole and image plane. In Figure 30.9b, the pulse has encountered the open end of the monopole, and it is traveling back down the monopole. A second spherical wavefront W_2 centered on the open end has

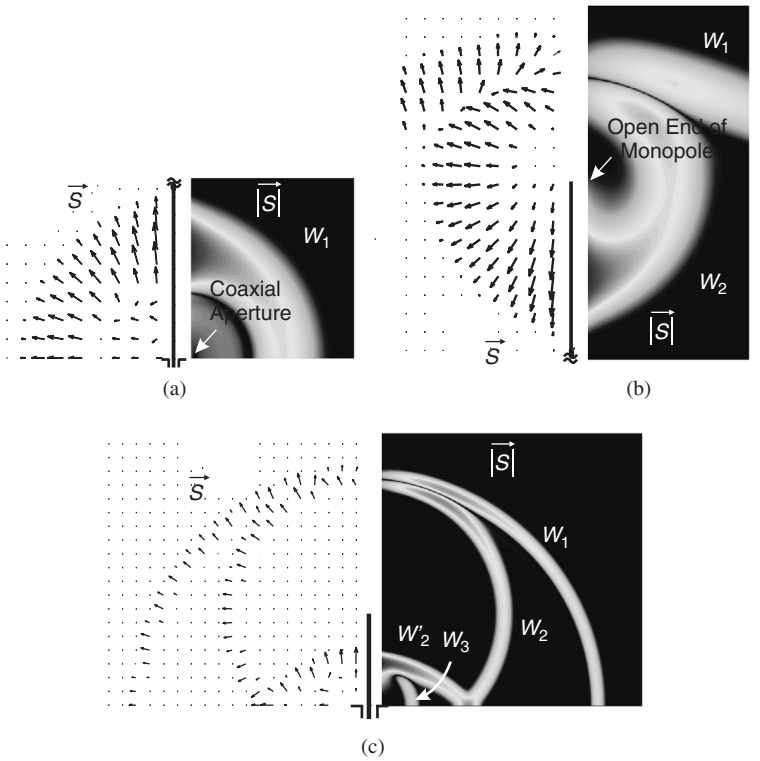


Figure 30.9 Three snapshots in time showing the magnitude (right) and direction (left) of the Poynting vector surrounding the cylindrical monopole antenna: (a) near the coaxial aperture at time $t/\tau_a = 0.3$, (b) near the open end of the monopole at time $t/\tau_a = 1.3$, and (c) around the whole structure at time $t/\tau_a = 2.5$. Logarithmic scaling is used for both plots. Note that (a) and (b) only show a portion of the monopole. (Adapted from Ref. 17; copyright © 2001 IEEE.)

formed, and it connects the inward propagating pulse of charge/current on the monopole with the wavefront W_1. Additional wavefronts, W_2', W_3, ..., shown in Figure 30.9c, are produced each time the pulse encounters the drive point and the open end. All of these spherical wavefronts travel outward at the speed of light. The Poynting vectors are seen to be predominantly normal to the wavefronts, which indicates that energy is being transported away from both the drive point and the open end.

The input impedance $Z_A(\omega)$ or admittance $Y_A(\omega) = 1/Z_A(\omega) = G_A(\omega) + jB_A(\omega)$ of the monopole antenna is a useful parameter for practical applications, and it is also a sensitive measure of the accuracy of any theoretical model. It is easily calculated from the FDTD time-domain results using Eqs. (30.12) and (30.13). In Figure 30.10 the input admittance is graphed as a function of frequency for a monopole with the same dimensions as used for Figure 30.8 [18]. FDTD results (lines) for three different levels of discretization (A, B, C) are compared with measurements (dots). The parameters for the three levels of discretization are given in Table 30.2. In this graph we observe the convergence of the FDTD method. Consider the input susceptance, B_A; the result for discretization A is slightly displaced from the measured values, while the results for discretizations B and C are essentially the same as the measured values. Hence we can conclude that,

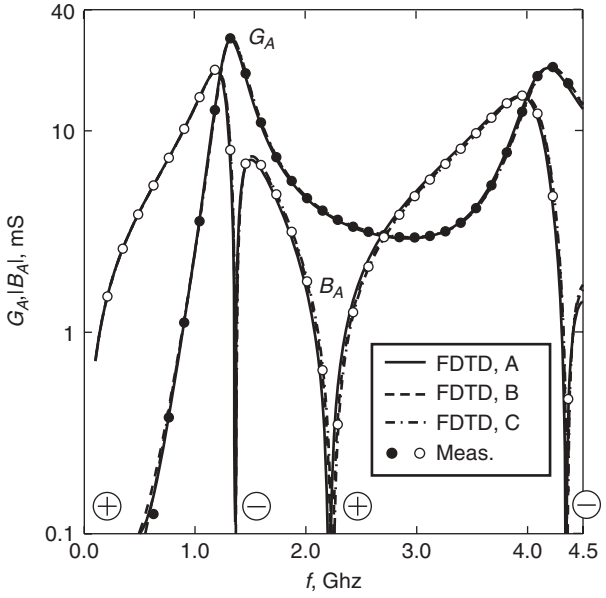


Figure 30.10 Comparison of theoretical and measured results for the input admittance of the cylindrical monopole antenna. Results are shown for three levels of discretization (A, B, C) in the FDTD method. (Adapted from Ref. 18; copyright © 2003 IEEE.)

TABLE 30.2 Parameters for FDTD Analysis of the Monopole Antenna

	Level of Discretization		
Parameter	A	B	C
$(b-a)/\Delta\ \rho$	2	4	12
$h/\Delta\ \rho$	50	101	303
Min N_λ ($f = 4.5$ GHz)	67	135	303
$1 - \overline{v}_p/c$ ($S = 0.5$)	2.75×10^{-4}	6.77×10^{-5}	7.56×10^{-6}

for practical purposes, the FDTD results for the input admittance have converged to the measured values at discretization B, which is for four FDTD cells across the gap in the coaxial line or 101 cells along the length of the monopole. We should mention that the dimensions of the FDTD cell for this example had to be chosen so that an integral number of cells fit along the dimensions of the antenna, so the cells are not perfectly square. Note that discretization B corresponds to 135 cells per wavelength at the highest frequency ($f = 4.5$ GHz) and a relative error in the phase velocity (Eq. (30.3) and Figure 30.2 for $S = 0.5$) of only 6.77×10^{-3} %. For this example, it is not the error in the phase velocity that determines the accuracy of the solution. The fine details of the structure must be accurately modeled, and this requires cells that are much smaller than needed for a small error in the phase velocity.

The very good agreement of the theoretical results with the measurements evident in Figures 30.8 and 30.10 is a consequence of the close match of the theoretical model for

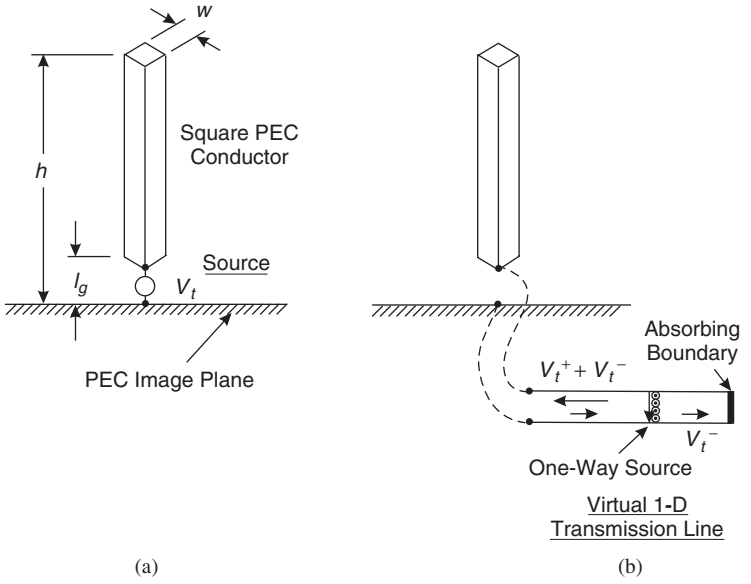


Figure 30.11 Simplified models for the cylindrical monopole antenna. (a) Model incorporating a "hard source." (b) Model incorporating a virtual one-dimensional (1D) transmission line. The monopole conductor has a square cross section in both models.

the monopole, Figure 30.7b, to the experimental model, Figure 30.7a. In some cases, additional constraints on the analysis require a reduction in the fidelity of the FDTD model, and such good agreement cannot be expected. To illustrate the effect a reduction in the fidelity of the model can have on the accuracy of the results, we examine some common simplifications used for the FDTD model of the monopole.

For the models shown in Figure 30.11, the cylindrical conductor of the monopole has been replaced by an equivalent square conductor of side $w = 1.69a$ [19]. Thus the monopole now can be analyzed using the conventional three-dimensional rectangular FDTD lattice rather than the two-dimensional cylindrical lattice of Figure 30.7b.

The excitation for the monopole has also been changed from that in Figure 30.7b. For the model in Figure 30.11a, a "hard source" is used. This specifies the total voltage $V_t = V_t^+ + V_t^-$ across the gap of length l_g at the base of the monopole. And for the model in Figure 30.11b, a virtual one-dimensional transmission line is connected across the gap at the base of the monopole [20]. This transmission line contains the same elements as the transmission line in Figure 30.7b, in particular, a one-way source that specifies the incident voltage V_t^+. We refer to this line as virtual because it does not appear in the FDTD lattice surrounding the monopole. It is in a different location and coupled to the monopole through the voltage and current at its terminals. The hard source, while simple to implement, suffers from two drawbacks not present with the transmission line feed. There is no damping in the hard source, unless resistance is added, so the current on the antenna can ring for a long period of time. And the total voltage is specified, so the reflected voltage, a quantity often of interest in time-domain simulations, is not readily available.

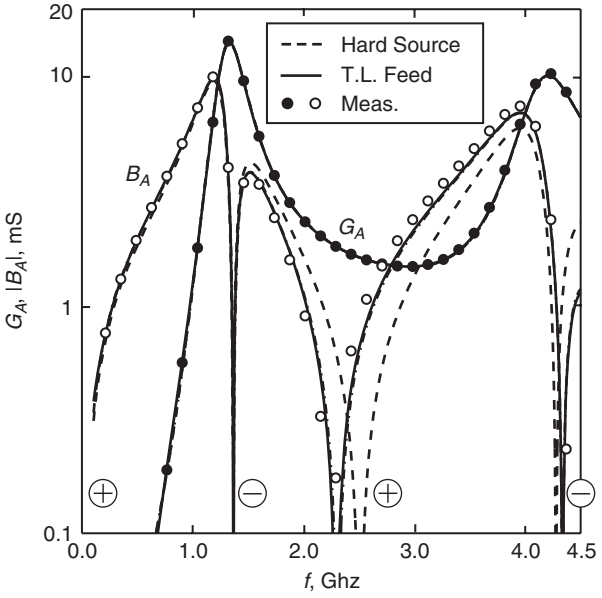


Figure 30.12 Comparison of theoretical and measured results for the input admittance of the cylindrical monopole antenna. Results are shown for the two simplified FDTD models. (Adapted from Ref. 18; copyright © 2003 IEEE.)

In Figure 30.12, FDTD results for the input admittance for both models in Figure 30.11 are compared with measurements made with the configuration shown in Figure 30.7b [18]. The level of discretization used is such that the simulations have converged for practical purposes. The theoretical results for the input conductance, G_A, for both models are in very good agreement with the measurements; however, those for the input susceptance, B_A, differ from the measurements, particularly for the hard source (dashed line). The difference in susceptance is a consequence of the geometry for the simplified models not accurately representing the experimental model, Figure 30.7a, in the vicinity of the drive point (the aperture of the coaxial line). The susceptance for the simplified models can be brought into better agreement with the measured results by adding a small capacitance in parallel with the terminals of the monopole [18].

30.4.2 Metallic Horns and Spirals: Stair-Stepped Surfaces

For the monopole antennas discussed in the previous section, the boundaries of the FDTD cells as well as the boundaries of all material regions (PECs) coincided with surfaces of constant coordinate. Thus the boundaries of material regions never passed obliquely through an FDTD cell. This is a special case that is not encountered for most antennas.

Figure 30.13a illustrates the more general case. It shows the cross section of a PEC object with the rectangular FDTD lattice superimposed. The curved surface of the object does not coincide with any of the lattice lines. Note, that for the computation we only need to know the field in the FDTD cells that are exterior to the PEC, because both **E** and **H** are zero inside the PEC. There are different approaches that can be used for this

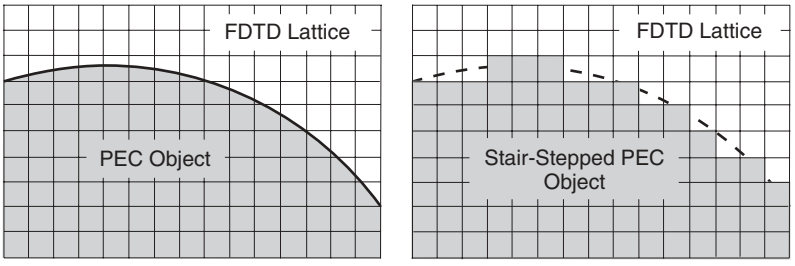


Figure 30.13 (a) Rectangular FDTD lattice superimposed on the cross section of an object that is a perfect electric conductor (PEC). (b) The surface of the object has been deformed to conform to the rectangular lattice; the surface of the object has been replaced by a stair-stepped approximation.

case. One approach is to introduce nonrectangular FDTD cells that conform to the surface of the object; these cells could be used throughout the computational volume or just adjacent to the object [21–23]. Another much simpler approach, shown in Figure 30.13b, is to deform the curved surface of the object so that it conforms to the rectangular FDTD lattice. The surface of the object is said to be replaced by a "stair-stepped" or "staircase" approximation. The stair-stepped approximation will introduce an error, and often the error can be made negligible by simply choosing the size of the staircase to be small compared to the physical dimensions of the object [24, 25]. The stair-stepped approximation is commonly used, and it is the only approach we consider in this introductory treatment.

We now consider two practical antennas for which the stair-stepped approximation was used in modeling the structure in the FDTD analysis. As these examples will show, when properly used, the approximation can yield results that are in good agreement with experimental measurements. The first example is the metallic, pyramidal horn shown in Figure 30.14 (Flann Microwave Instruments Ltd. Model 1624-20). Antennas like this are used in many microwave applications, and sometimes they serve as gain standards (standard gain horns). The small drawings at the bottom of the figure show the lengths and angles that describe this particular horn antenna: $a = 10.95$ cm, $b = 7.85$ cm, $D = 2.284$ cm, $l_w = 5.08$ cm, $\alpha = 10.74^\circ$, and $\beta = 8.508^\circ$. The waveguide feeding the horn is type WR-90 (X-band, with the operational bandwidth 8.2–12.4 GHz).

In the FDTD model, the cubical cells have the side length $\Delta x = 0.635$ mm, and the perfectly conducting walls are plates two cells thick. The inset shows the faces of the individual cells that model the bottom wall of the horn; the cells are shown seven times actual size. The slanted sides of the horn are stair stepped, as indicated in the figure, with a "tread length-to-rise" of approximately six cells to one. The horn is fed by a probe inserted into the section of rectangular waveguide, and the incident and reflected voltages in a one-dimensional transmission line ($R_0 = 50\ \Omega$) connected to the probe are used in the analysis.

The structure is symmetrical about the x-z plane, and this symmetry was used in the analysis to reduce the size of the computational volume, which was $519 \times 116 \times 183$ cells. The sides of the antenna were 20 cells from the PML absorbing boundary (10 cells thick), except the front side (radiating aperture), which was 40 cells from the absorbing boundary.

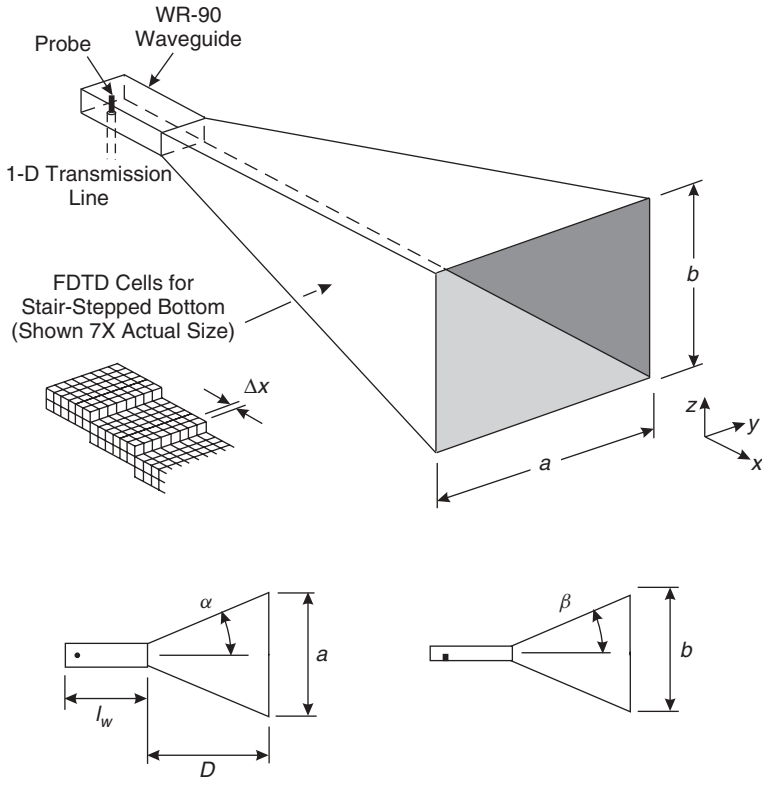


Figure 30.14 Schematic drawing for the pyramidal horn antenna. The inset shows the FDTD cells used to model the bottom of the horn.

The pyramidal horn was first analyzed as a transmitting antenna. The excitation in the transmission line, $V_t^+(t)$, was a differentiated Gaussian pulse, Eq. (30.20), with the characteristic time $\tau_p = 1.59 \times 10^{-11}$ s. This pulse has significant energy over the operational bandwidth of the horn: 8.2–12.4 Ghz. The peak of the spectrum for the pulse is at 10.0 GHz, and the spectrum drops to 10% of the peak at 600 MHz and 27.6 GHz.

At the highest frequency (shortest wavelength) within the operational bandwidth of the horn we have $\Delta x = 0.026\lambda$, which corresponds roughly to 38 cells per wavelength. From this result, we can estimate the numerical dispersion using Figure 30.2 or Eq. (30.3). The relative error in the phase velocity is about 0.1%, which is equivalent to 8.1×10^{-3} degrees of phase error per cell, or a total error of 4.2 degrees of phase error for propagation across the longest side of the computational volume.

Figure 30.15 is a comparison of the FDTD results (solid line) for this antenna with measurements (dots). The measured data were kindly supplied by Dr. David G. Gentle of the National Physical Laboratory, Teddington, Middlesex, U.K. Figures 30.15a and 30.15b show the E- and H-plane field patterns at the frequency 10 GHz, and Figure 30.15c shows the gain on boresite as a function of frequency. The results from the FDTD calculations are seen to be in very good agreement with the measurements. The small differences that do exist in the H-plane field pattern are for angles at which the

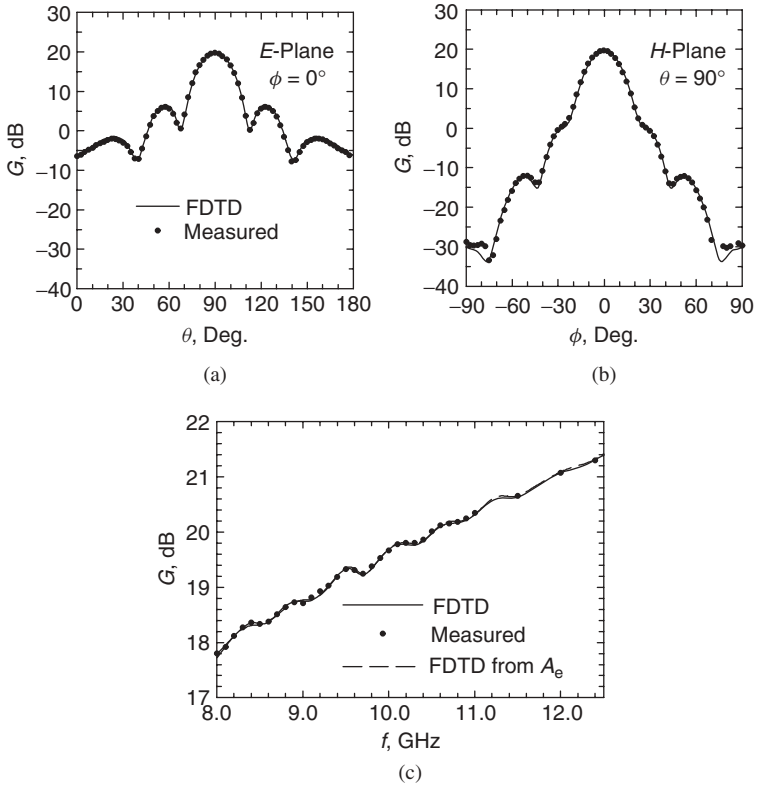


Figure 30.15 Comparison of theoretical and measured results for the pyramidal horn antenna. (a) E-plane pattern and (b) H-plane pattern at 10 GHz. (c) Boresite gain versus frequency.

field is very weak, 50 dB below the peak. We note that the precise details of the probe feeding the waveguide in the FDTD model do not affect the calculation of the gain, Eq. (30.15), of the horn. This would not be the case if the realized gain, Eq. (30.14) (gain including mismatch), were determined.

The pyramidal horn was also analyzed as a receiving antenna. For this case, a plane wave was incident from the boresite direction ($\hat{\mathbf{k}}_i = -\hat{\mathbf{x}}$) with the electric field pointing in the z direction. The incident electric field was a differentiated Gaussian pulse in time, Eq. (30.20), with the same characteristic time as used for the transmitting case, $\tau_p = 1.59 \times 10^{-11}$ s. The effective area obtained from the receiving analysis was converted to a gain using Eq. (30.18), and the result is shown as a dashed line in Figure 30.15c. As expected from reciprocity, the results from the two FDTD calculations (transmitting and receiving) are nearly identical.

The FDTD method provides the field throughout the computational volume, and it can be used to construct graphical results that illustrate the process of radiation for the transmitting horn antenna. For such illustrations, we want an excitation whose spectrum lies within the operational bandwidth of the antenna. Frequencies outside this band will either be cut off in the waveguide or overmode the waveguide. A good choice for the voltage $V_t^+(t)$ is the sinusoid of frequency ω_0 amplitude modulated by a Gaussian pulse, that is Eq. (30.21) shown in Figure 30.6b. With $f_0 = \omega_0/2\pi = 10.0$ GHz, and

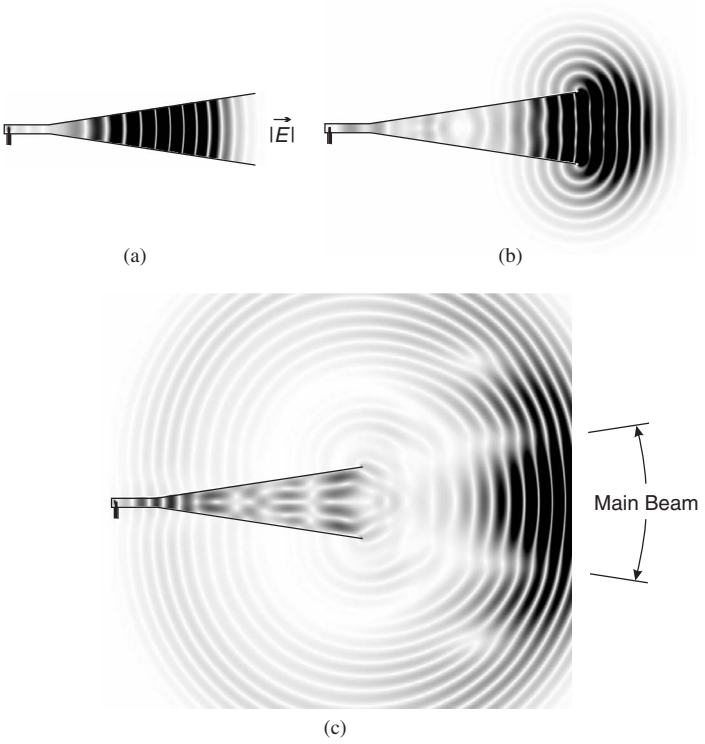


Figure 30.16 Gray scale plots for the magnitude of the electric field on the vertical symmetry plane of the transmitting horn antenna. The excitation is a sinusoid amplitude modulated by a Gaussian pulse.

$\tau_p = 7.96 \times 10^{-11}$ s, the spectrum for this signal is 10% of its peak at $f = 5.7$ GHz and $f = 14.3$ GHz.

Figure 30.16 shows three gray scale plots for the magnitude of the electric field on the x-z plane of the transmitting antenna. In Figure 30.16a the pulse has entered the horn from the waveguide, but it has not reached the aperture. The spacing between the white lines (nulls) roughly corresponds to one-half of a guide wavelength. Note that this spacing decreases on going from the throat of the horn toward the aperture. In the rectangular waveguide, the guide wavelength is about 1.3 times the free-space wavelength, whereas at the aperture of the horn it is closer to the free-space wavelength. Figure 30.16b is for a time when the pulse has reached the aperture. Note that the white lines in the horn near the aperture are distorted; there is a small segment that is concave to the right. This is caused by the reflection from the aperture that is traveling back toward the throat of the horn. Directly in front of the aperture, the radiated wave is roughly planar. In Figure 30.16c, the field has propagated away from the horn, and a spherical wavefront has formed that is approximately centered on the aperture. The change in the shade of gray in going around the antenna (dark in front to light in back) clearly shows a large "front-to-back ratio" for the horn. In the forward direction, minima appear along the wavefront, and these minima define the main beam in the far zone. Back in the horn,

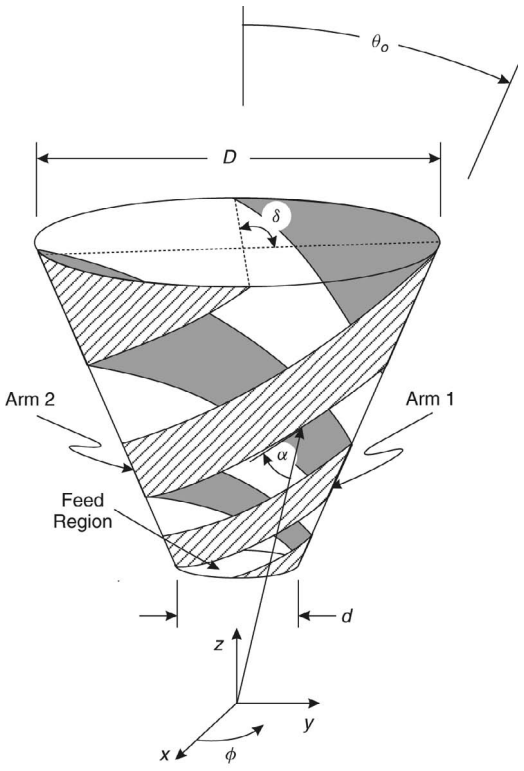


Figure 30.17 Geometry for the two-arm conical spiral antenna. (Adapted from Ref. 26; copyright © 2002 IEEE.)

the field has several minima and maxima across its width. They indicate the presence of higher order modes that were excited when the initial pulse encountered the aperture.

The second example we consider is the two-arm, conical spiral antenna shown in Figure 30.17 [26]. It is used in applications that require an antenna to radiate circular polarization over a broad bandwidth. This antenna is formed by winding two metallic strips around the surface of a truncated cone. The angles and dimensions for the particular antenna we are considering are $d = 1.9$ cm, $D = 15.2$ cm, $\theta_0 = 7.5^\circ$, $\alpha = 75^\circ$, and $\delta = 90^\circ$. It is designed to have constant gain and input impedance ($Z_A \approx 100\ \Omega$) over an operational bandwidth extending from $f_{\min} = 0.5$ GHz to $f_{\max} = 3.3$ GHz.

In the FDTD model, the arms of the spiral are formed by making selected faces of the cubical cells ($\Delta x = 0.8$ mm) PEC. The result is the stair-stepped approximation in Figure 30.18. For clarity, only the lower 10% of the antenna is shown in the figure. The spiral is fed by a one-dimensional (1D) transmission line ($R_0 = 100\ \Omega$) connected at the bottom of the antenna; the same arrangement as used with the monopole antenna in Figure 30.11b. The excitation in the transmission line, $V_t^+(t)$, is a differentiated Gaussian pulse, Eq. (30.20), whose spectrum is centered on the operational bandwidth of the antenna.

The computational volume was $691 \times 240 \times 240$ cells, with the sides of the antenna 15 cells from the PML absorbing boundary (10 cells thick), except the bottom side (main

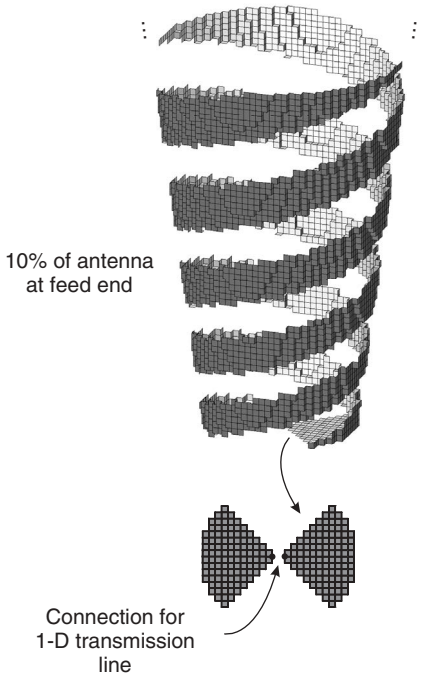


Figure 30.18 Schematic drawing showing the arrangement of FDTD cells used to model the conical spiral antenna. For clarity, only the lower 10% of the antenna is shown. (Adapted from Ref. 26; copyright © 2002 IEEE.)

direction for radiation), which was 30 cells from the absorbing boundary. At the highest frequency (shortest wavelength) within the operational bandwidth of the horn we have $\Delta x = 0.0093\lambda$, which corresponds roughly to 107 cells per wavelength. From this result, we can estimate the numerical dispersion using Figure 30.2 or Eq. (30.3). The relative error in the phase velocity is about 0.01%, which is equivalent to 3.6×10^{-4} degrees of phase error per cell, or a total error of 0.25 degrees of phase error for propagation across the longest side of the computational volume. As with the earlier case of the monopole antenna, it is not the error in the phase velocity that determines the accuracy of the solution but the degree to which the fine details of the structure are modeled.

Figure 30.19 is a comparison of the FDTD results (solid line) for this antenna with measurements (dashed line). Figure 30.19a shows the magnitude of the reflection coefficient at the terminals of the antenna, and Figure 30.19b shows the realized gain, Eq. (30.14), at boresite ($-\hat{\mathbf{z}}$ direction) as a function of frequency. The results from the FDTD calculations are seen to be in fairly good agreement with the measurements. The differences that do exist are most likely caused by elements in the experimental model that were not included in the theoretical model. In the experimental model, the metallic arms were on a very thin dielectric substrate (Kapton, thickness 0.051 mm), which was not included in the theoretical model. In addition, the terminal measurements were made through a balun, and the imperfections in the balun were not taken into account.

The FDTD method provides detailed information about the electromagnetic field surrounding the spiral, and it can be used to graphically illustrate how energy is radiated

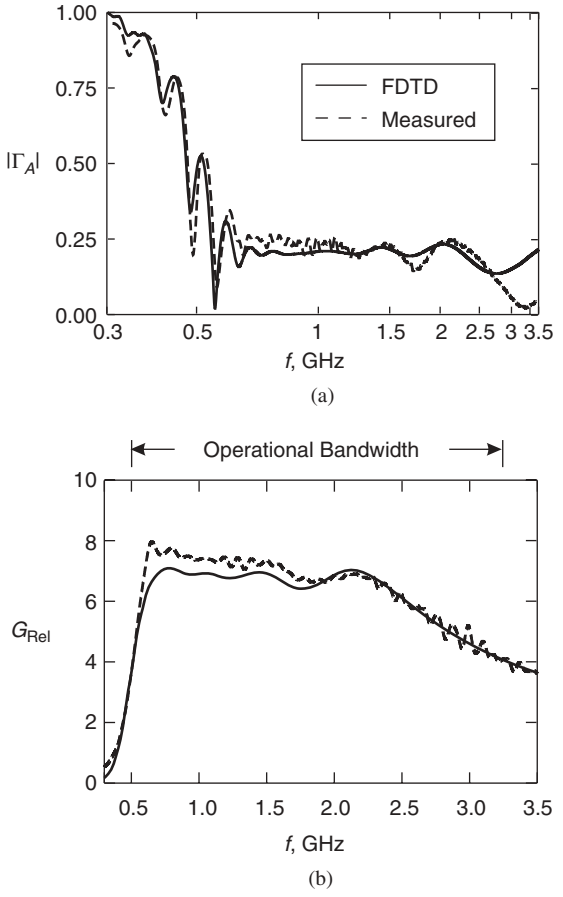


Figure 30.19 Comparison of theoretical and measured results for the conical spiral antenna. (a) Magnitude of reflection coefficient versus frequency. (b) Realized gain in the boresite direction versus frequency. (Adapted from Ref. 26; © 2002 IEEE.)

from this structure [27]. Figure 30.20 shows three gray scale plots of the magnitude of the x component of the electric field on the x-z plane. Each plot is for a different normalized time t/τ_L, were τ_L is the time for light to travel the length of the spiral arm. We can see that the radiation is roughly periodic with the spacing between the nulls (white lines) being $\lambda/2$. The frequency corresponding to this wavelength is indicated on each plot. These plots clearly show that the region from which radiation leaves the antenna changes with the wavelength, moving from the small end (diameter d) for the shortest wavelengths (highest frequencies) to the large end (diameter D) for the longest wavelengths (lowest frequencies). This is in keeping with the "active-region concept," which states that the radiation originates at the cross section of the spiral that is approximately one wavelength in circumference [28].

In the previous two examples, the stair-stepped approximations used for the geometry of the antennas in the FDTD models were adequate for obtaining theoretical results that were in good agreement with the measurements. This is a consequence of choosing the

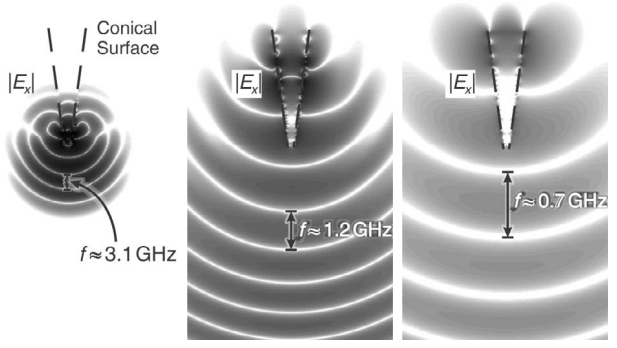


Figure 30.20 Gray scale plots for the magnitude of the electric field near the conical spiral antenna for three instants in time: (a) $t/\tau_L = 0.1$, (b) $t\ /\ \tau_L = 0.6$, and (c) $t\ /\ \tau_L = 1.1$, where τ_L is the time for light to travel the length of the spiral arm. (Adapted from Ref. 27; copyright © 2003 IEEE.)

size of the steps to be small compared to the dimensions defining the geometry of the antennas. For example, for the pyramidal horn, the height of the stair step is only about 10% of the smallest dimension of the antenna (the height of the rectangular waveguide). We now consider a case in which the stair-stepped approximation leads to significant errors in the calculated results.

The transverse electromagnetic (TEM) horn is a simple antenna used for applications that require broad bandwidth. The FDTD model for the monopole version of this antenna is shown in Figure 30.21a. It is formed from a PEC plate that is an isosceles triangle of side length s and angle at the apex α. The plate is inclined at the angle $\beta/2$ to the PEC image plane, and the antenna is fed by a transmission line connected between the apex of the plate and the image plane. The plate–image plane forms a TEM transmission line, and for the example to be discussed ($\alpha = 25.4^\circ$, $\beta = 11.2^\circ$), the characteristic impedance of this line is $R_0 \approx 50\ \Omega$ [29–31]. The transmission line feeding the antenna has the same characteristic impedance.

The plate for this antenna is stair stepped in the FDTD model in the manner shown in Figure 30.21b. Two different sizes for the staircase are examined: case A for which the rise is $\Delta z = 1$ mm and the tread length is $\Delta s = 1$ cm, and case B for which $\Delta z = 2$ mm and $\Delta s = 2$ cm. Note that the level of discretization for case B is twice as coarse as that for case A. The smallest dimensions for the horn are at the drive point, where the initial tread for both cases is 4 mm above the image plane. So for case A, the rise of the staircase, Δz, is about 25% of the smallest dimension of the horn; whereas for case B it is about 50% of the smallest dimension of the horn.

Figure 30.22a shows the reflected voltage, $V_t^-(t)$, in the feeding transmission line of the horn when the incident voltage, $V_t^+(t)$, is a unit-amplitude, differentiated Gaussian pulse, Eq. (30.20), with the characteristic time $\tau_p = 5.31 \times 10^{-11}$ s. The peak of the spectrum for the pulse is at 3.0 GHz. The solid line is for case A and the dashed line is for case B. The initial reflection from the drive point is evident and is similar for both cases, and the reflection from the open end of the horn has been windowed out. There is a pronounced ripple in the result for the coarser staircase, case B. The ripple is clearly due to the staircase, because its period roughly corresponds to the round-trip time on a

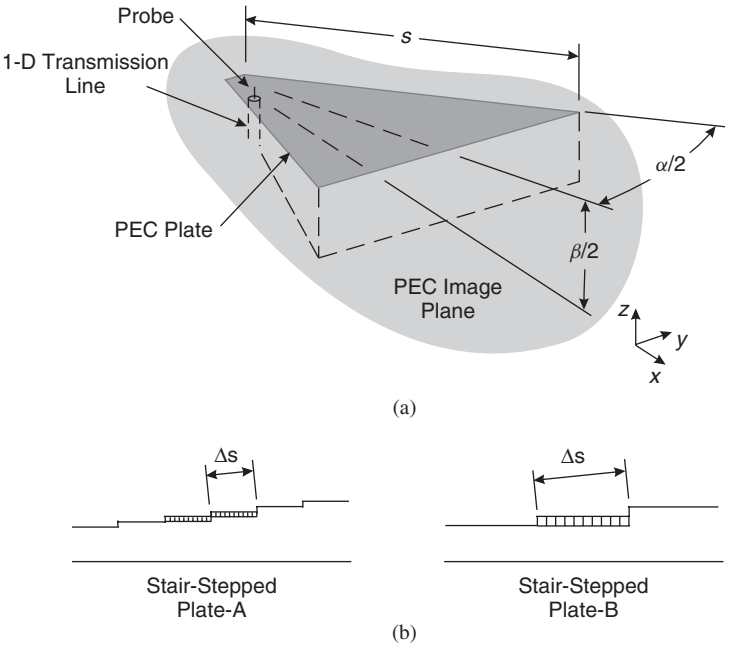


Figure 30.21 (a) Schematic drawing for the TEM horn antenna (monopole configuration). (b) Cross sections showing the stair-stepped approximation to the plate for two different cases, A and B.

tread, which is $\Delta t = 2\Delta s_B/c \approx 2.5\tau_p$. Note that the amplitude of the ripple decreases with time. This is because the reflections that occur later in time are from stair steps further out along the antenna, where the rise of the staircase, Δz, is a smaller fraction of the separation between the plate and the image plane.

Figure 30.22b shows the magnitude of the Fourier transform (spectrum) of the reflection coefficient for the antenna. Note that the results for the two cases, A and B, are quite different. Specifically, for case B there is a distinct dip in the reflection coefficient near $2f\ \Delta s_B/c = 1$ ($f = 7.5$ GHz). At this frequency, $\Delta s_B/\lambda = \frac{1}{2}$, so the small reflections from all of the steps in the staircase add in phase.

To avoid the problem described above, we must use a finer staircase, such as in case A. For TEM horns with low characteristic impedance (generally small β), this can require a very fine level of discretization. A similar problem is encountered with bow-tie antennas with low characteristic impedance [32].

30.4.3 Microstrip Patches: Excessive Ringing for Narrowband Antennas

The antennas we examined in the previous section, a conical spiral and horns, are fairly wideband antennas. Now we consider the other extreme, namely, narrowband antennas. For our example, we use the basic, rectangular microstrip patch antenna shown in Figure 30.23.

In the mid-1980s, Chang et al. [33] made extensive measurements of this antenna, and we first compare our FDTD results with their measurements. The dimensions for a patch designed for frequencies around $f = 7.0$ GHz are $s = 1.1$ cm, $w = 1.7$ cm, and

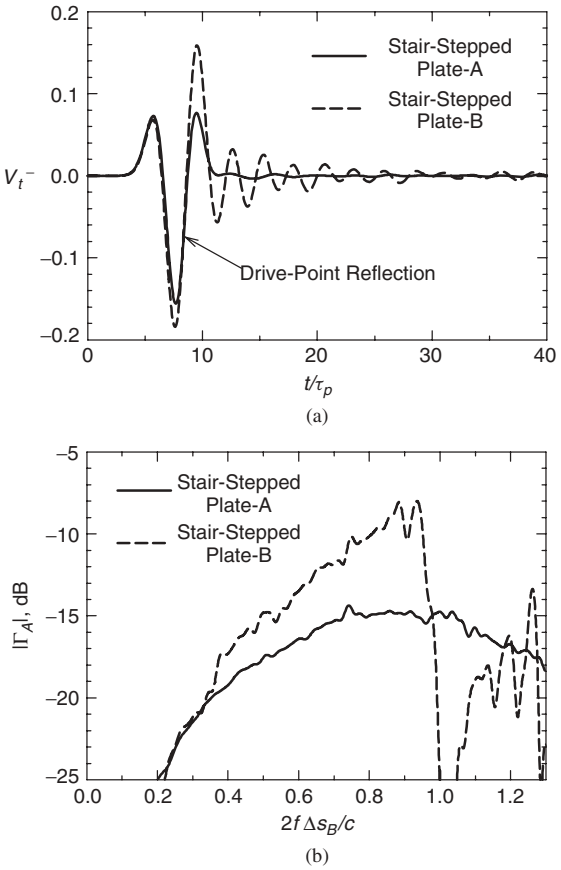


Figure 30.22 Results for two different stair-stepped approximations (A and B) applied to the TEM horn antenna. (a) The reflected voltage in the feeding transmission line; the reflection from the open end of the horn has been windowed out. (b) The magnitude of the Fourier transform of the reflection coefficient for the antenna.

$h = 3.175$ mm. As shown in the figure, the probe of the feeding coaxial line ($R_0 = 50\ \Omega$) is displaced from the broad side of the patch by $l_p = 1.5$ mm. In the model, the dielectric substrate is 10 cm $\times$ 10 cm with the electrical properties $\varepsilon_r = 2.33$ and $\sigma = 2.1 \times 10^{-3}$ S/m, and the ground plane is infinite. The incident voltage, $V_t^+(t)$, in the feeding transmission line is a unit-amplitude, differentiated Gaussian pulse, Eq. (30.20), with the characteristic time $\tau_p = 2.65 \times 10^{-11}$ s. The peak of the spectrum for this pulse is at 6.0 GHz.

The dimensions of the FDTD rectangular cells ($\Delta x = 0.529$ mm, $\Delta y = 0.500$ mm, $\Delta z = 0.500$ mm) were chosen so that all of the details of the coaxial feed line could be included in the model, and the time step was $\Delta t = 9.44 \times 10^{-13}$ s. The number of time steps, N_t, required for the simulation was determined by observing the magnitude of the reflected voltage $|V_t^-|$ in the feeding transmission line versus the normalized time $t/\Delta t$; this is shown in Figure 30.24a. Note that the vertical scale is logarithmic. When $t/\Delta t = 3000$, the reflected voltage has dropped by six orders of magnitude from its peak,

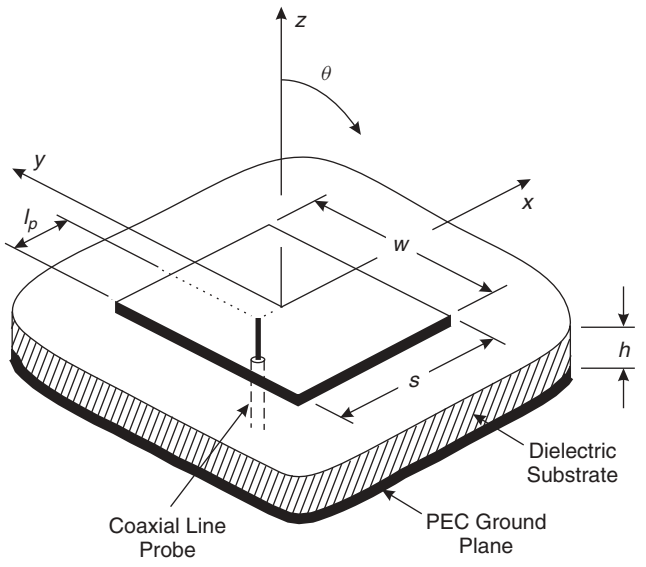


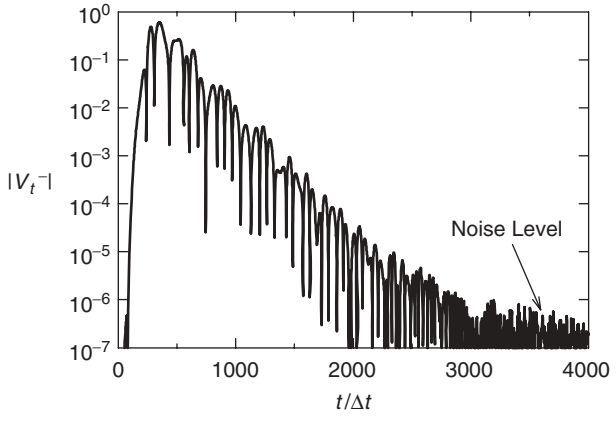
Figure 30.23 Rectangular microstrip patch antenna fed by a coaxial line probe.

and it is at the noise level for the computation. So any number of time steps greater than 3000 was deemed adequate for the simulation ($N_t = 4000$ was actually used).

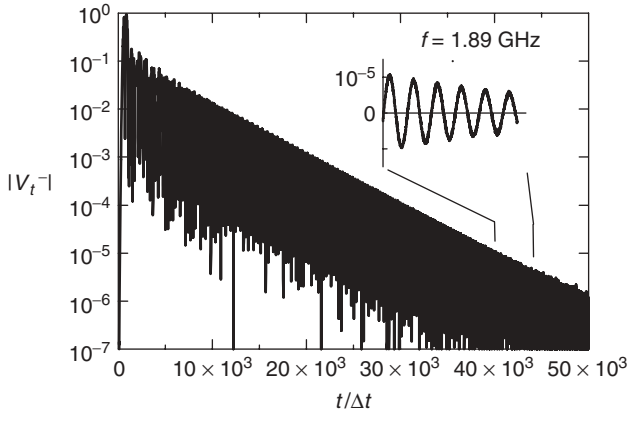
Figure 30.25 is a comparison of the FDTD theoretical results with the measurements. The graph in Figure 30.25a shows the magnitude of the reflection coefficient versus frequency: theory (solid line) and measurement (dots). The agreement is reasonably good, particularly when we consider that some of the geometrical details for the measurement, such as the precise geometry at the feed, were not known for use in the FDTD model.

The field patterns were measured with the 10-cm × 10-cm substrate mounted at the center of a circular aluminum image plane of diameter 1 m. We chose not to model this configuration with the same fine resolution used for the FDTD calculation of the reflection coefficient, because of the large amount of memory that would be required. Instead, larger cells were used with the dimensions $\Delta x = 1.59$ mm, $\Delta y = 1.42$ mm, and $\Delta z = 1.57$ mm. The use of the larger cells causes little error in the far-zone field patterns. The FDTD and measured field patterns for the frequency $f = 6.8$ GHz are compared in Figure 30.25b. These plots show the gain, Eq. (30.15), versus the angle θ, normalized to 0 dB at the peak. Results are given for both the E-plane (x-z plane, solid line and dots) and the H-plane (y-z plane, dashed line and triangles). Again the agreement is reasonably good.

For our second example, we chose a rectangular microstrip patch antenna designed to operate around $f = 1.9$ GHz that is similar to one reported in the literature [34]. The dimensions for the patch are $s = 5.12$ cm, $w = 6.0$ cm, and $h = 1.575$ mm, and the probe of the feeding coaxial line ($R_0 = 50\ \Omega$) is displaced from the broad side of the patch by $l_p = 1.64$ cm. The dielectric substrate ($\epsilon_r = 2.2$ and $\sigma = 1.1 \times 10^{-3}$ S/m) and the ground plane are the same size: 11.5 cm × 11.5 cm. The incident voltage, $V_t^+(t)$, in the feeding transmission line is a unit-amplitude, differentiated Gaussian pulse, Eq. (30.20), with the characteristic time $\tau_p = 1.061 \times 10^{-10}$ s, and the peak of the spectrum for this pulse is at 1.5 GHz. Again, the parameters for the FDTD simulation allow complete modeling



(a)



(b)

Figure 30.24 The magnitude of the reflected voltage in the feeding coaxial line versus the normalized time: (a) rectangular microstrip patch and (b) narrowband, rectangular microstrip patch.

of the details of the coaxial feed line ($\Delta x = 0.529$ mm, $\Delta y = 0.500$ mm, $\Delta z = 0.500$ mm, $\Delta t = 9.91 \times 10^{-13}$ s). The electrical thickness of the substrate for this example is about one-eighth of that for the previous example, $h/\lambda = 0.010$ (for $f = 1.9$ GHz) versus $h/\lambda = 0.077$ (for $f = 7.3$ GHz), so we expect this antenna to have a significantly narrower bandwidth [35].

Figure 30.24b shows the magnitude of the reflected voltage $|V_t^-|$ in the feeding transmission line (logarithmic scale) versus the normalized time $t/\Delta t$. As a consequence of the narrower bandwidth, the reflected voltage decreases much more slowly with increasing $t/\Delta t$ than in the previous example, Figure 30.24a. The reflected voltage has dropped by six orders of magnitude from its peak and is approaching the noise level for the computation when $t/\Delta t = 50{,}000$. So about 50,000 time steps ($N_t = 50{,}000$) are required for the simulation, as compared to 3000 for the previous example! The inset in Figure 30.24b shows the reflected voltage, plotted on a linear scale, for times around

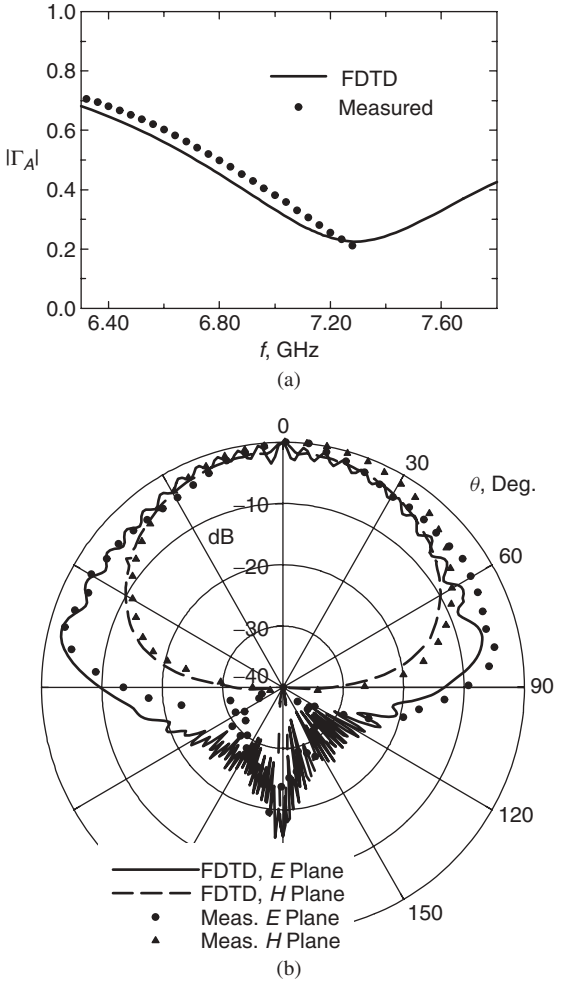


Figure 30.25 Comparison of theoretical and measured results for the rectangular microstrip patch antenna: (a) magnitude of reflection coefficient versus frequency and (b) field patterns for E- and H-planes at the frequency $f = 6.8$ GHz. Measured results are from Ref. 33.

$t/\Delta t = 40{,}000$. The voltage is seen to be a slowly decaying sinusoid at the frequency $f \approx 1.89$ Ghz.

In Figure 30.26 we show the magnitude of the reflection coefficient versus frequency for simulations with different numbers of time steps: $N_t = 8000$, $N_t = 24{,}000$, and $N_t = 50{,}000$. For each case, a Hanning window is applied in time to eliminate truncation artifacts. The antenna is seen to be matched at the frequency $f = 1.89$ GHz, and the "apparent" bandwidth for the match is seen to depend on the number of time steps used for the simulation. Thus if one were to underestimate the number of time steps required for the simulation to converge, one would think that the antenna had a much wider bandwidth for the reflection coefficient than it actually has. With the detailed analysis presented above, this point may appear to be obvious. However, sometimes, particularly

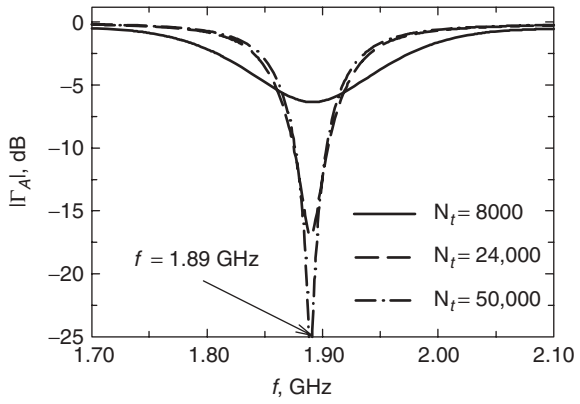


Figure 30.26 Narrowband, rectangular microstrip patch antenna: magnitude of reflection coefficient versus frequency for three different numbers of time steps.

when a computation is automated, this degree of analysis may not be performed every time a parameter for the antenna, such as the thickness of the substrate, is changed.

In some cases, special techniques can be applied to shorten the computation for a narrow bandwidth antenna. For example, because of the well defined, decaying sinusoidal waveform in the reflection coefficient for this antenna, a shorter computation time, say, $N_t = 20{,}000$, could be used with an extrapolation for the remainder of the waveform. Such techniques are discussed in the literature [36].

30.5 SUMMARY

In this chapter, we have presented an introduction to the finite-difference time-domain method, aimed at individuals who have little or no experience with the method. Thus we have limited the presentation to the basics of the method, and we have avoided mention of many refinements that are generally restricted to particular applications. To give the reader a sense of the breadth of application allowed by these refinements, we present a partial list below.

- Techniques for handling materials with dispersive properties (properties that are a function of the frequency), anisotropic properties (properties that depend on the direction of the field components), and nonlinear properties.
- Methods for incorporating impedance boundary conditions.
- Subcell methods for treating material sheets that are thinner than a FDTD cell.
- Methods for incorporating periodic boundary conditions, which are useful in treating antenna arrays.
- Higher order FDTD schemes that have lower error (numerical dispersion) than the conventional Yee algorithm.
- Techniques for incorporating nonuniform and nonorthogonal grids.
- Special procedures for handling objects that are bodies of revolution.

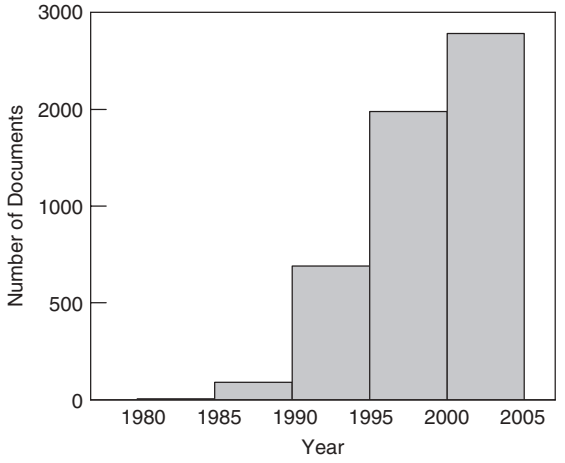


Figure 30.27 Number of documents published over the last twenty-five years that include the words "finite-difference time-domain" or "FDTD" in the title. Each bar shows the total number of documents published during a five-year period.

The brevity of the contribution precluded the derivation of the mathematical formulas associated with the method, for example, FDTD update equations and equations for the perfectly matched layer. These formulas can be found in the in-depth treatment of the method contained in the book edited by A. Taflove and S. C. Hagness [9].

To assess the popularity of the FDTD method, a search was done with INSPEC for documents that included either "finite-difference time-domain" or "FDTD" in the title. (A few of these documents apply the finite-difference time-domain method to problems other than electromagnetic, such as acoustic problems.) The results of the search, presented in Figure 30.27, clearly show the rapid growth in the popularity of the method over the last twenty-five years.

In keeping with the subject of this handbook, the emphasis throughout the chapter has been on the application of the FDTD method to the analysis of antennas. After brief discussions of the special formulations associated with transmitting and receiving antennas, the details for the analysis of a few different types of antennas were presented. Again, because of the brevity of the contribution, no attempt was made to mention all of the different antennas that have been analyzed with the method. Many individuals have used the method to treat antennas; as an indication of the number, the INSPEC search mentioned above listed over 500 documents with FDTD and antenna(s) in the title.

All of the numerical results presented in the examples were obtained by us or our students. Thus we have very detailed knowledge for each example and can make fairly accurate statements about the results. These examples were chosen not only to show the power of the FDTD method, particularly the good agreement with experimental measurements, but also to show that the method has some limitations; albeit, the limitations are sometimes due to the crudeness of the theoretical model for the antenna or the choice of the parameters for the simulation.

The refinement of the FDTD method and its application to practical problems is an ongoing story. Undoubtedly, there will be some exciting accomplishments made in the future. One area with great promise is the use of the method for antenna synthesis. Here

we do not mean the conventional approach in which the method is coupled with an optimization routine and used to choose the parameters for a standard antenna (dipole, horn, etc.) so that certain criteria for the performance are met. What we have in mind for antenna synthesis is a quite different, a more modern approach. In this approach, the structure of the antenna is not completely predetermined with only a few parameters to be chosen, but the structure of the antenna is actually developed as part of the synthesis! The FDTD method is well suited for use in such schemes; because of the flexibility of the method, a new structure can easily be introduced. The antenna structure is changed by simply changing the electromagnetic constitutive parameters associated with individual cells. The particular technique we have employed for such synthesis is based on the "fragmented aperture" concept and is discussed in Refs. [37–40].

ACKNOWLEDGMENTS

The senior author (G.S.) would like to thank his former and current Ph.D. students at the Georgia Institute of Technology who have contributed to his understanding of the FDTD method for antenna analysis: they include J. M. Bourgeois, I. R. Capoglu, T. H. Hertel, R. T. Lee, J. G. Maloney, T. P. Montoya, D. R. Reid, W. R. Scott, Jr., and K. L. Shlager.

REFERENCES

1. K. S. Yee, Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media, *IEEE Trans. Antennas Propag.*, Vol. 14, pp. 302–307, May 1966.
2. J. G. Maloney, G. S. Smith, and W. R. Scott, Jr., Accurate computation of the radiation from simple antennas using the finite-difference time-domain method, *IEEE Trans. Antennas Propag.*, Vol. 38, pp. 1059–1068, July 1990.
3. J. J. Boonzaaier and C. W. Pistorius, Thin wire dipoles—a Finite-difference time-domain approach, *Electron. Lett.*, Vol. 26, pp. 1891–1892, 25 October 1990.
4. D. S. Katz, M. J. Picket-May, A. Taflove, and K. R. Umashankar, FDTD analysis of electromagnetic wave radiation from systems containing horn antennas, *IEEE Trans. Antennas Propag.*, Vol. 39, pp. 1203–1212, August 1991.
5. P. A. Tirkus and C. A. Balanis, Finite-difference time-domain method for antenna radiation, *IEEE Trans. Antennas Propag.*, Vol. 40, pp. 334–340, March 1992.
6. R. J. Luebbers and J. Beggs, FDTD calculation of wide-band antenna gain and efficiency, *IEEE Trans. Antennas Propag.*, Vol. 40, pp. 1403–1407, November 1992.
7. J. G. Maloney and G. S. Smith, Modeling of Antennas, in *Advances in Computational Electrodynamics, The Finite-Difference Time-Domain Method*, A. Taflove (Ed.), Artech House, Boston, 1998, chap 7, pp. 409–460. Also, J. G. Maloney, G. S. Smith, E. Thiele, O. Ghandi, N. Chavannes, and S. Hagness, in *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, 3rd ed., A. Taflove, and S. Hagness, (Eds.), Artech House, Boston, 2005, chap. 14, pp. 607–676.
8. G. S. Smith, *An Introduction to Classical Electromagnetic Radiation*, Cambridge University Press, Cambridge, UK, 1997.
9. A. Taflove and S. C. Hagness (Eds.), *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, Artech House, Boston, 2005.

10. J. B. Schneider and C. L. Wagner, FDTD dispersion revisited: faster-than-light propagation, *IEEE Microwave Guided Wave Lett.*, Vol. 9, pp. 54–56, February 1999.
11. S. Gedney, An anisotropic perfectly matched layer-absorbing medium for the truncation of FDTD lattices, *IEEE Trans. Antennas Propag.*, Vol. 44, pp. 1630–1639, December 1996.
12. S. Gedney, Perfectly matched layer absorbing boundary conditions, in *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, 3rd ed., A. Taflove and S. C. Hagness (Eds.), Artech House, Boston, 2005, chap 7, pp. 273–328.
13. K. L. Shlager and G. S. Smith, Near-field to near-field transformation for use with FDTD method and its application to pulsed antenna problems, *Electron. Lett.*, Vol. 30, pp. 1262–1264, 4, August 1994.
14. K. L. Shlager and G. S. Smith, Comparison of two near-field to near-field transformations applied to pulsed antenna problems, *Electron. Lett.*, Vol. 31, pp. 936–938, 8, June 1995.
15. G. S. Smith, A direct derivation of a single-antenna reciprocity relation for the time domain, *IEEE Trans. Antennas Propag.*, Vol. 52, pp. 1568–1577, June 2004.
16. J. G. Maloney, M. P. Kesler, and G. S. Smith, *Generalization of PML to cylindrical geometries*, in *13th Annual Review of Progress in Applied Computational Electromagnetics*, Monterey, CA, March 1997, pp. 900–908.
17. G. S. Smith and T. W. Hertel, On the transient radiation of energy from simple current distributions and linear antennas, *IEEE Antennas Propag. Mag.*, Vol. 43, pp. 49–62, June 2001.
18. T. W. Hertel and G. S. Smith, On the convergence of common FDTD feed models for antennas, *IEEE Trans. Antennas Propag.*, Vol. 51, pp. 1771–1779, August 2003.
19. R. W. P. King, *The Theory of Linear Antennas*, Harvard University Press, Cambridge, MA, 1956, p. 20.
20. J. G. Maloney, K. L. Shlager, and G. S. Smith, A Simple FDTD model for transient excitation of antennas by transmission lines, *IEEE Trans. Antennas Propag.*, Vol. 42, pp. 289–292, February 1994.
21. S. Dey and R. Mittra, A locally conformal finite-difference time-domain (FDTD) algorithm for modeling three-dimensional perfectly conducting objects, *IEEE Microwave Guided Wave Lett.*, Vol. 7, pp. 273–275, September 1997.
22. A. Taflove, M. Celuch-Marcysiak, and S. Hagness, Local subcell models of fine geometrical features, in *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, 3rd ed., A. Taflove and S. C. Hagness (Eds.), Artech House, Boston, 2005, Chap. 10, pp. 407–462.
23. S. Gedney, F. Lansing, and N. Chavannes, Nonuniform grids, nonorthogonal grids, unstructured grids, and subgrids, in *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, 3rd ed., A. Taflove and S. C. Hagness (Eds.), Artech House, Boston, 2005, Chap. 11, pp. 463–516.
24. A. C. Cangellaris and D. B. Wright, Analysis of the numerical error caused by the stair-stepped approximation of a conducting boundary in FDTD simulations of electromagnetic phenomena, *IEEE Trans. Antennas Propag.*, Vol. 39, pp. 1518–1525, October 1991.
25. R. Holland, Pitfalls of staircase meshing, *IEEE Trans. Electromagnetic Compatibility*, Vol. 35, pp. 434–439, November 1993.
26. T. W. Hertel and G. S. Smith, Analysis and design of two-arm conical spiral antennas, *IEEE Trans. Electromagnetic Compatibility*, Vol. 44, pp. 25–37, February 2002.
27. T. W. Hertel and G. S. Smith, On the dispersive properties of the conical spiral antenna and its use for pulsed radiation, *IEEE Trans. Antennas Propag.*, Vol. 51, pp. 1426–1433, July 2003.
28. J. D. Dyson, The characteristics and design of the conical log-spiral antenna, *IEEE Trans. Antennas Propag.*, Vol. 13, pp. 488–499, July 1965.

29. K. L. Shlager, G. S. Smith, and J. G. Maloney, Accurate analysis of TEM horn antennas for pulse radiation, *IEEE Trans. Electromagnetic Compatibility*, Vol. 38, pp. 414–423, August 1996.

30. R. T. Lee and G. S. Smith, On the characteristic impedance of the TEM horn antenna, *IEEE Trans. Antennas Propag.*, Vol. 52, pp. 315–318, January 2004.

31. R. T. Lee and G. S. Smith, A design study for the basic TEM horn antenna, *IEEE Antennas Propag. Mag.*, Vol. 46, pp. 86–92, February 2004.

32. K. L. Shlager, G. S. Smith, and J. G. Maloney, Optimization of bow-tie antennas for pulse radiation, *IEEE Trans. Antennas Propag.*, Vol. 42, pp. 975–982, July 1994.

33. E. Chang, S. A. Long, and W. F. Richards, An experimental investigation of electrically thick rectangular microstrip antennas, *IEEE Trans. Antennas Propag.*, Vol. 34, pp. 767–772, June 1986.

34. H. Abdallah, W. Wasylkiwskyj, K. Parikh, and A. Zaghloul, Comparison of return loss calculations with measurements of narrow-band microstrip patch antennas, *ACES J.*, Vol. 19, pp. 184–186, November 2004.

35. D. R. Jackson and N. G. Alexopoulos, Simple approximate formulas for the input resistance, bandwidth, and efficiency of a resonant rectangular patch, *IEEE Trans. Antennas Propag.*, Vol. 39, pp. 407–410, March 1991.

36. S. Chebolu, R. Mittra, and W. D. Becker, The analysis of microwave antennas using the FDTD method, *Microwave J.*, Vol. 39, pp. 134–150, January 1996.

37. J. G. Maloney, P. H. Harms, M. P. Kesler, T. L. Fountain, and G. S. Smith, Novel, planar antennas designed using the genetic algorithm, in *1999 USNC/URSI Radio Science Meeting*, Orlando, FL, July 1999, p. 237.

38. J. G. Maloney, M. P. Kesler, P. H. Harms, T. L. Fountain, and G. S. Smith, The fragmented aperture antenna: FDTD analysis and measurement, in *Millennium Conference on Antennas and Propagation (AP 2000)*, Davos, Switzerland, April 2000.

39. J. G. Maloney, M. P. Kesler, P. H. Harms, and G. S. Smith, Fragmented aperture antennas and broadband ground planes, U.S. Patent No. 6,323,809 B1, 27 November 2001.

40. L. N. Pringle, P. H. Harms, S. P. Blalock, G. N. Kiesel, E. J. Kuster, P. G. Friederich, R. J. Prado, J. M. Morris, and G. S. Smith, A reconfigurable aperture antenna based on switched links between electrically small metallic patches, *IEEE Trans. Antennas Propag.*, Vol. 52, pp. 1434–1445, June 2004.

CHAPTER 31

Finite-Element Analysis and Modeling of Antennas

JIAN-MING JIN, ZHENG LOU, NORMA RILEY, and DOUGLAS RILEY

31.1 INTRODUCTION

The finite-element method (FEM) is a numerical procedure to obtain approximate solutions to boundary-value problems of mathematical physics with the aid of an electronic computer. The method was first proposed by Courant in 1943 to solve variational problems in potential theory [1]. Thereafter, the method has been developed and applied extensively to problems of structural analysis and increasingly to problems in other fields. Today, the FEM is recognized as a general preeminent method applicable to a wide variety of engineering and mathematical problems, including those in antenna and microwave engineering.

The first application of the FEM to microwave engineering and electromagnetics appeared in 1969 when Silvester employed it to analyze wave propagation in a hollow waveguide [2]. The importance of the method was quickly recognized and successful applications were achieved for the analysis of electrostatic, magnetostatic, and dielectric-loaded waveguide problems. In 1974, Mei developed a technique that combined the FEM with eigenfunction expansion to deal with open-region electromagnetics problems such as antenna and scattering analysis [3]. In 1982, Marin developed an alternative method to deal with open-region scattering problems, which combined the FEM and a boundary integral equation [4].

An important breakthrough in the finite-element analysis of vector electromagnetic field problems occurred in the 1980s with the development of edge-based vector elements [5–7]. These new elements accurately model the nature of the electric and magnetic fields and eliminate many of the challenges associated with traditional node-based scalar elements. Since the development of the vector elements, the FEM has become a very powerful numerical technique in computational electromagnetics. Today, the method is used as a major design tool for antennas and microwave devices. Its basic principle and a variety of applications have been described in many books, such as those by Silvester and Ferrari [8], Jin [9], and Volakis, Chatterjee, and Kempel [10].

The application of the FEM for the analysis and design of various antennas dates back to the 1970s when Mei developed the first accurate approach that enabled the FEM to

Modern Antenna Handbook. Edited by Constantine A. Balanis

deal with unbounded open-region problems [3]. The method was applied to axisymmetric antennas. For many years, the FEM was limited to simplified two-dimensional models of antennas because of the difficulty of using the node-based elements to model vector electromagnetic fields [11, 12]. The first full-wave three-dimensional finite-element analysis of antennas appeared in the early 1990s [13], thanks to the development of edge-based elements. Thereafter, a variety of the FEM-based numerical techniques have been developed for the analysis and simulation of various antennas and antenna arrays. Most notably, the FEM-based numerical techniques have been developed to analyze infinitely periodic array antennas [14–17], finite array antennas [18–21], complex horn antennas [22–26], conformal antennas [27–29], dielectric lens antennas [30, 31], antennas residing on complex materials [32, 33], and antennas mounted on a finite platform [34–38]. Most of these techniques were developed for analysis in the frequency domain. To perform a frequency sweep analysis, a model-order reduction technique has also been proposed [39, 40]. Recently, the FEM has been developed for antenna analysis directly in the time domain [41–44]. Such a time-domain analysis is highly efficient for the characterization of broadband responses and is capable of modeling nonlinear materials and devices. Novel numerical schemes have been developed for the analysis of both finite and infinitely periodic antenna arrays using time-domain finite-element formulations [45–47].

31.2 FINITE-ELEMENT FORMULATION IN THE FREQUENCY DOMAIN

We first formulate the finite-element analysis in the frequency domain. Consider a generic antenna, sketched in Figure 31.1. The antenna is excited by a current source with electric current density denoted by $\mathbf{J}_{\text{imp}}$. This current radiates an electromagnetic field, which is modified by the antenna. The main objective of an antenna analysis is to predict the performance characteristics of the antenna, which include the input impedance and radiation patterns. To this end, one has to solve Maxwell's equations

$$\nabla \times \mathbf{E} = -j\omega\mu\mathbf{H} - \mathbf{M}_{\text{imp}} \tag{31.1}$$

$$\nabla \times \mathbf{H} = j\omega\varepsilon\mathbf{E} + \mathbf{J}_{\text{imp}} \tag{31.2}$$

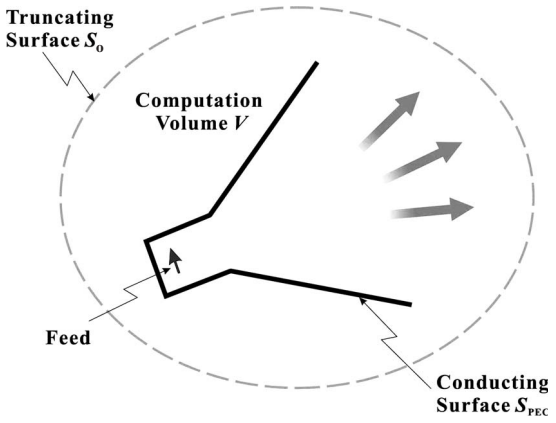


Figure 31.1 Computational domain for the finite-element analysis of an antenna. An artificial surface truncates the computational domain into a finite volume.

$$\nabla \cdot (\varepsilon \mathbf{E}) = -\frac{1}{j\omega} \nabla \cdot \mathbf{J}_{\text{imp}} \tag{31.3}$$

$$\nabla \cdot (\mu \mathbf{H}) = -\frac{1}{j\omega} \nabla \cdot \mathbf{M}_{\text{imp}} \tag{31.4}$$

together with the boundary condition

$$\hat{\mathbf{n}} \times \mathbf{E} = 0 \quad \text{on } S_{\text{PEC}} \tag{31.5}$$

where S_{PEC} denotes the perfect electrically conducting (PEC) surface of the antenna. In Eq. (31.1), $\mathbf{M}_{\text{imp}}$ denotes the magnetic current density of an impressed magnetic current. Although a magnetic current does not exist in reality, it is a useful quantity to model certain antenna feeds, such as a magnetic frill generator. In addition to Eq. (31.5), the electric and magnetic fields have to satisfy the Sommerfeld radiation condition at infinity:

$$\lim_{r \to \infty} r \left[\nabla \times \begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix} + jk_0 \hat{\mathbf{r}} \times \begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix} \right] = 0 \tag{31.6}$$

where k_0 is the free-space wavenumber. The electromagnetic problem defined by Eqs. (31.1)–(31.6) can be solved analytically only for a very few cases where S_{PEC} has a very simple shape. For most practical problems, one has to resort to a numerical method such as the FEM. Since the antenna radiates an electromagnetic field to infinity, the problem has an unbounded solution space. To use the FEM, this unbounded space must be truncated into a finite space. This can be accomplished by introducing a fictitious surface to enclose the antenna. This surface is denoted as S_{o} in Figure 31.1. To uniquely define the electromagnetic problem bounded by S_{o}, one has to specify a boundary condition on S_{o}. This boundary condition should make S_{o} as transparent as possible to the radiated field. The ideal boundary condition is to make S_{o} completely transparent, such that the radiated field can pass through it without any distortion or reflection. This is, unfortunately, not possible in practice. Therefore we instead employ an approximate boundary condition

$$\hat{\mathbf{n}} \times \nabla \times \begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix} + jk_0 \hat{\mathbf{n}} \times \hat{\mathbf{n}} \times \begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix} \approx 0, \quad \text{on } S_{\text{o}} \tag{31.7}$$

which is similar to the Sommerfeld radiation condition in Eq. (31.6). This approximate radiation condition is also called the first-order absorbing boundary condition (ABC). For this ABC to be reasonably accurate, S_{o} must be placed some distance away from the antenna. More accurate treatment of the truncation surface S_{o} is given in Section 31.4.

The electromagnetic problem defined by Eqs. (31.1)–(31.5) and (31.7) can be solved in terms of either the electric field $\mathbf{E}$ or the magnetic field $\mathbf{H}$. Here, we describe the solution procedure for the electric field; the magnetic field can be solved in a similar manner. By eliminating $\mathbf{H}$ in Eqs. (31.1)–(31.4), we can derive the vector wave equation for $\mathbf{E}$ as

$$\nabla \times \left(\frac{1}{\mu_r} \nabla \times \mathbf{E} \right) - k_0^2 \varepsilon_r \mathbf{E} = -jk_0 Z_0 \mathbf{J}_{\text{imp}} - \nabla \times \left(\frac{\mathbf{M}_{\text{imp}}}{\mu_r} \right), \quad \text{in } V \tag{31.8}$$

where $\mu_r = \mu/\mu_0$ and $\varepsilon_r = \varepsilon/\varepsilon_0$ are the relative permeability and permittivity, respectively, $k_0 = \omega\sqrt{\mu_0\varepsilon_0}$ and $Z_0 = \sqrt{\mu_0/\varepsilon_0}$ are the free-space wavenumber and intrinsic impedance, respectively, and V denotes the volume enclosed by S_o. To solve Eq. (31.8) for $\mathbf{E}$, we can multiply Eq. (31.8) by an appropriate testing function $\mathbf{T}$ and integrate over V to find

$$\iiint_V \mathbf{T} \cdot \left[\nabla \times \left(\frac{1}{\mu_r}\nabla \times \mathbf{E}\right) - k_0^2 \varepsilon_r \mathbf{E}\right] dV = -\iiint_V \mathbf{T} \cdot \left[jk_0 Z_0 \mathbf{J}_{\text{imp}} + \nabla \times \left(\frac{\mathbf{M}_{\text{imp}}}{\mu_r}\right)\right] dV \tag{31.9}$$

By invoking the vector identity

$$\mathbf{T} \cdot \left[\nabla \times \left(\frac{1}{\mu_r}\nabla \times \mathbf{E}\right)\right] = \frac{1}{\mu_r}(\nabla \times \mathbf{T}) \cdot (\nabla \times \mathbf{E}) - \nabla \cdot \left[\mathbf{T} \times \left(\frac{1}{\mu_r}\nabla \times \mathbf{E}\right)\right] \tag{31.10}$$

and Gauss's theorem

$$\iiint_V \nabla \cdot \left[\mathbf{T} \times \left(\frac{1}{\mu_r}\nabla \times \mathbf{E}\right)\right] dV = \oiint_{S_o} \left[\mathbf{T} \times \left(\frac{1}{\mu_r}\nabla \times \mathbf{E}\right)\right] \cdot d\mathbf{S} \tag{31.11}$$

we obtain

$$\iiint_V \left[\frac{1}{\mu_r}(\nabla \times \mathbf{T}) \cdot (\nabla \times \mathbf{E}) - k_0^2 \varepsilon_r \mathbf{T} \cdot \mathbf{E}\right] dV = \iint_{S_{\text{PEC}}} \frac{1}{\mu_r}(\hat{\mathbf{n}} \times \mathbf{T}) \cdot (\nabla \times \mathbf{E})\, dS - jk_0 \oiint_{S_o} (\hat{\mathbf{n}} \times \mathbf{T}) \cdot (\hat{\mathbf{n}} \times \mathbf{E})\, dS - \iiint_V \mathbf{T} \cdot \left[jk_0 Z_0 \mathbf{J}_{\text{imp}} + \nabla \times \left(\frac{\mathbf{M}_{\text{imp}}}{\mu_r}\right)\right] dV \tag{31.12}$$

after the application of the first-order ABC in Eq. (31.7).

To find a numerical solution of Eq. (31.12) using the FEM, the entire volume V is first divided into small finite elements, such as hexahedral, tetrahedral, prism, and/or pyramidal cells. Within each small finite-element volume, $\mathbf{E}$ can be interpolated using a set of discrete values. One approach is to assign $\mathbf{E}$ at a few points on the element and then interpolate $\mathbf{E}$ elsewhere using a set of scalar interpolation functions. This approach turns out to be very problematic because of the difficulty to apply correct boundary conditions to the interpolated $\mathbf{E}$-field. A better approach is to assign the tangential component of $\mathbf{E}$ at each edge of the element and then interpolate $\mathbf{E}$ elsewhere using a set of vector basis functions. For example, the field in a tetrahedral element can be interpolated as

$$\mathbf{E}^e(x, y, z) = \sum_{i=1}^{6} \mathbf{N}_i^e(x, y, z) E_i^e \tag{31.13}$$

where E_i^e denotes the tangential component of $\mathbf{E}$ at edge i of element e, and $\mathbf{N}_i^e$ is the corresponding interpolation or basis function. Denoting the simplex coordinates of a

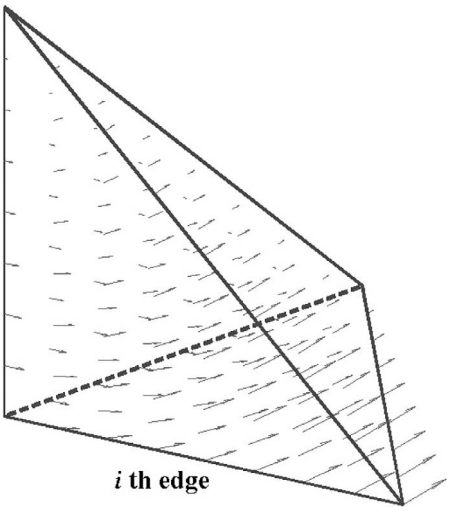


Figure 31.2 Vector basis function for a tetrahedral element. There are six such first-order vector basis functions for a tetrahedral element associated with the six edges.

tetrahedron as λ_l ($l = 1, 2, 3, 4$) [8], the vector basis function associated with the edge that connects nodes α and β can be written as

$$\mathbf{N}^e_{\alpha\beta}(\mathbf{r}) = \lambda_l \nabla\lambda_k - \lambda_k \nabla\lambda_l \tag{31.14}$$

where (α, β, k, l) are the six permutations of (1, 2, 3, 4) such that $\alpha < \beta$. Figure 31.2 shows one of the six first-order vector basis functions for a tetrahedral element. Clearly, such basis functions have a tangential component only along the associated edge, and as such they ensure the tangential continuity of the interpolated field while allowing the normal component to be discontinuous at a material discontinuity. Hence they accurately model the nature of the vector field **E**. Higher-order vector basis functions can also be constructed to achieve better interpolation accuracy.

When the field **E** is interpolated in each element using its tangential values at the edges of the element, the field **E** in the entire volume V can be expressed as

$$\mathbf{E} = \sum_{i=1}^{N_{\text{edge}}} \mathbf{N}_i E_i \tag{31.15}$$

where N_{edge} denotes the total number of edges excluding those on S_{PEC}, E_i denotes the tangential component of **E** at the ith edge, and $\mathbf{N}_i$ is the corresponding basis function. Obviously, for an edge inside V, $\mathbf{N}_i$ spans over several neighboring elements that share the common edge. Also note that by excluding the edges on S_{PEC} in Eq. (31.15), the interpolated field satisfies the required boundary condition in Eq. (31.5).

By substituting Eq. (31.15) into Eq. (31.12) and using the same $\mathbf{N}_i$ as the weighting function **T**, we obtain

$$\sum_{j=1}^{N_{\text{edge}}} K_{ij} E_j = b_i, \quad i = 1, 2, \ldots, N_{\text{edge}} \tag{31.16}$$

where

$$K_{ij} = \iiint_V \left[\frac{1}{\mu_r} (\nabla \times \mathbf{N}_i) \cdot (\nabla \times \mathbf{N}_j) - k_0^2 \varepsilon_r \mathbf{N}_i \cdot \mathbf{N}_j \right] dV + jk_0 \oiint_{S_o} (\hat{\mathbf{n}} \times \mathbf{N}_i) \cdot (\hat{\mathbf{n}} \times \mathbf{N}_j) \, dS \tag{31.17}$$

$$b_i = - \iiint_V \mathbf{N}_i \cdot \left[jk_0 Z_0 \mathbf{J}_{\text{imp}} + \nabla \times \left(\frac{\mathbf{M}_{\text{imp}}}{\mu_r} \right) \right] dV \tag{31.18}$$

Note that the integral over S_{PEC} in Eq. (31.12) vanishes here since $\hat{\mathbf{n}} \times \mathbf{N}_i = 0$ on S_{PEC}. Equation (31.16) can be written compactly as

$$[K]\{E\} = \{b\} \tag{31.19}$$

which can be solved for $\{E\}$. Because the elemental interactions in Eq. (31.17) are local in nature, $[K]$ is a sparse and symmetric matrix that can be solved efficiently using a sparse matrix solver. Once $\{E\}$ is obtained, the field everywhere in V can be calculated using Eq. (31.15), from which other parameters such as the input impedance and radiation patterns can be computed.

The formulation can easily be extended to the case where μ_r and ε_r are tensors that represent an anisotropic material. As a frequency-domain based formulation, it can easily handle dispersive materials since the simulation is performed only at a single frequency.

31.3 FINITE-ELEMENT FORMULATION IN THE TIME DOMAIN

The finite-element formulation described in Section 31.2 operates in the frequency domain. It takes the frequency as an input parameter and solves for the electric field at that specific frequency. However, for most antenna analyses, one is interested in the characteristics of an antenna over a frequency band. In such a case, one often has to repeat a frequency-dependent analysis at many frequencies, which can be quite time consuming for both narrowband and broadband antennas. For narrowband antennas, the antenna characteristics change rapidly with frequency due to resonance. A very small frequency step has to be used to accurately capture impedance variations near resonance. For broadband antennas, the frequency step can be made larger; however, the entire frequency band may be very wide, thus requiring many frequency samples. Furthermore, it is difficult to accurately model nonlinear devices/media by using a frequency-domain based method. These two challenges can be alleviated by using the FEM formulated in the time domain.

In the time domain, the first two of Maxwell's equations (31.1) and (31.2) become

$$\nabla \times \mathbf{E}(t) = -\mu \frac{\partial \mathbf{H}(t)}{\partial t} - \mathbf{M}_{\text{imp}}(t) \tag{31.20}$$

$$\nabla \times \mathbf{H}(t) = \varepsilon \frac{\partial \mathbf{E}(t)}{\partial t} + \sigma \mathbf{E}(t) + \mathbf{J}_{\text{imp}}(t) \tag{31.21}$$

where σ denotes the conductivity. The boundary condition on the perfectly conducting surface remains the same as in Eq. (31.5) and the boundary condition corresponding to Eq. (31.7) becomes

$$\hat{\mathbf{n}} \times \left(\frac{1}{\mu_0} \nabla \times \mathbf{E} \right) + Y_0 \hat{\mathbf{n}} \times \left(\hat{\mathbf{n}} \times \frac{\partial}{\partial t} \mathbf{E} \right) \approx 0, \quad \text{on } S_o \tag{31.22}$$

where $Y_0 = 1/Z_0$. By eliminating the magnetic field from Eqs. (31.20) and (31.21), we obtain the vector wave equation for the electric field

$$\nabla \times \left[\frac{1}{\mu} \nabla \times \mathbf{E}(t) \right] + \varepsilon \frac{\partial^2 \mathbf{E}(t)}{\partial t^2} + \sigma \frac{\partial \mathbf{E}(t)}{\partial t} = -\frac{\partial \mathbf{J}_{\text{imp}}}{\partial t} - \nabla \times \left(\frac{\mathbf{M}_{\text{imp}}}{\mu} \right) \tag{31.23}$$

and its weak-form representation

$$\begin{aligned} &\iiint_V \left[\frac{1}{\mu} (\nabla \times \mathbf{T}) \cdot (\nabla \times \mathbf{E}) + \varepsilon \mathbf{T} \cdot \frac{\partial^2 \mathbf{E}}{\partial t^2} + \sigma \mathbf{T} \cdot \frac{\partial \mathbf{E}}{\partial t} \right] dV + Y_0 \oiint_{S_o} (\hat{\mathbf{n}} \times \mathbf{T}) \cdot \left(\hat{\mathbf{n}} \times \frac{\partial \mathbf{E}}{\partial t} \right) dS \\ &\quad = -\iiint_V \mathbf{T} \cdot \left[\frac{\partial \mathbf{J}_{\text{imp}}}{\partial t} + \nabla \times \left(\frac{\mathbf{M}_{\text{imp}}}{\mu} \right) \right] dV \end{aligned} \tag{31.24}$$

The derivation follows that in the frequency domain. Note that the vector testing function is assumed to satisfy the boundary condition $\hat{\mathbf{n}} \times \mathbf{T} = 0$ on S_{PEC}, since Eq. (31.5) is enforced in the solution of Eq. (31.24).

To seek the finite-element solution of Eq. (31.24), we first perform the spatial discretization in exactly the same manner as in the frequency domain. To be more specific, by subdividing the solution volume into small finite elements and expanding the electric field within each element using the vector basis functions, we can express the electric field as

$$\mathbf{E}(\mathbf{r}, t) = \sum_{i=1}^{N_{\text{edge}}} \mathbf{N}_i(\mathbf{r}) E_i(t) \tag{31.25}$$

Substituting this into Eq. (31.24) yields a second-order ordinary differential equation

$$[T] \frac{d^2\{E\}}{dt^2} + [R] \frac{d\{E\}}{dt} + [S]\{E\} = \{f\} \tag{31.26}$$

where $[T]$, $[R]$, and $[S]$ represent sparse, symmetric matrices whose elements are given by

$$T_{ij} = \iiint_V \varepsilon \mathbf{N}_i \cdot \mathbf{N}_j dV \tag{31.27}$$

$$R_{ij} = \iiint_V \sigma \mathbf{N}_i \cdot \mathbf{N}_j dV + Y_0 \oiint_{S_o} (\hat{\mathbf{n}} \times \mathbf{N}_i) \cdot (\hat{\mathbf{n}} \times \mathbf{N}_j) \, dS \tag{31.28}$$

$$S_{ij} = \iiint_V \frac{1}{\mu} (\nabla \times \mathbf{N}_i) \cdot (\nabla \times \mathbf{N}_j) dV \tag{31.29}$$

Furthermore, $\{E\} = [E_1, E_2, \ldots, E_{N_{\text{edge}}}]^T$ and the elements of the excitation vector $\{f\}$ are given by

$$f_i = -\iiint_V \mathbf{N}_i \cdot \left[\frac{\partial \mathbf{J}_{\text{imp}}}{\partial t} + \nabla \times \left(\frac{\mathbf{M}_{\text{imp}}}{\mu}\right)\right] dV \tag{31.30}$$

Equation (31.26) can be solved by using direct integration or the finite-difference method [9]. In the finite-difference method, the time variable t is uniformly discretized such that it is represented by $t = n\ \Delta t$ $(n = 0,1,\ldots)$, where Δt is denoted as the time step. The continuous differentiation is then approximated by a finite difference, which yields an equation that allows the calculation of the unknown vector $\{E\}$ based on its previous values. This process is called time marching. As shown in Jin [9], the use of forward differencing will result in an unstable time-marching equation. The use of backward differencing will result in an *unconditionally* stable (the time step size Δt is not constrained by the spatial discretization) time-marching equation, which unfortunately is only first-order accurate—the accuracy of the solution is proportional to $O(\Delta t)$. The use of central differencing will yield a second-order accurate time-marching equation, which is *conditionally* stable—the time marching is stable only when Δt is smaller than a certain value dictated by the spatial discretization. For Eq. (31.26), the best choice is to use a differencing formula derived from the Newmark-beta integration method [48–50], which is equivalent to using the central differencing for the first- and second-order differentiations

$$\frac{d\{E\}}{dt} \approx \frac{\{E\}^{(n+1)} - \{E\}^{(n-1)}}{2\Delta t} \tag{31.31}$$

$$\frac{d^2\{E\}}{dt^2} \approx \frac{\{E\}^{(n+1)} - 2\{E\}^{(n)} + \{E\}^{(n-1)}}{(\Delta t)^2} \tag{31.32}$$

and using a weighted average for the undifferentiated quantities

$$\{E\} \approx \beta\{E\}^{(n+1)} + (1 - 2\beta)\{E\}^{(n)} + \beta\{E\}^{(n-1)} \tag{31.33}$$

$$\{f\} \approx \beta\{f\}^{(n+1)} + (1 - 2\beta)\{f\}^{(n)} + \beta\{f\}^{(n-1)} \tag{31.34}$$

where β is a parameter that takes a value between 0 and 1. In Eqs. (31.31)–(31.34), the superscript denotes the time at which the associated quantity is evaluated; for example, $\{E\}^{(n)}$ denotes the vector $\{E\}$ evaluated at $t = n\Delta t$. Substituting these into Eq. (31.26), we obtain

$$\begin{aligned}\left\{\frac{1}{(\Delta t)^2}[T] + \frac{1}{2\Delta t}[R] + \beta[S]\right\}\{E\}^{(n+1)} \\ = \left\{\frac{2}{(\Delta t)^2}[T] - (1 - 2\beta)[S]\right\}\{E\}^{(n)} - \left\{\frac{1}{(\Delta t)^2}[T] - \frac{1}{2\Delta t}[R] + \beta[S]\right\}\{E\}^{(n-1)} \\ + \beta\{f\}^{(n+1)} + (1 - 2\beta)\{f\}^{(n)} + \beta\{f\}^{(n-1)}\end{aligned} \tag{31.35}$$

Obviously, when $\beta = 0$, this equation reduces to the one obtained using the central differencing. However, it can be shown that when $\beta \geq \frac{1}{4}$, this equation is unconditionally stable [49] while maintaining second-order accuracy. Given the initial values of $\{E\}$, specifically, $\{E\}^{(0)}$ and $\{E\}^{(1)}$, and the values of the excitation vector $\{f\}$, Eq. (31.35)

can be employed to compute all the subsequent values for $\{E\}$. To compute each new $\{E\}$, one has to solve a matrix equation at each time step.

The formulation described above assumes that the media involved are dispersion free. In other words, both the permeability μ and the permittivity ε are invariant with respect to frequency. This formulation can be modified to deal with dispersive media [51, 52] and to demonstrate this, we consider the case where the permittivity ε is a function of frequency while the permeability μ remains frequency independent (electrically dispersive material). In this case, the electric flux density $\mathbf{D}(t)$ is related to the electric field intensity $\mathbf{E}(t)$ by the constitutive relation

$$\begin{aligned}\mathbf{D}(t) &= \varepsilon_0\varepsilon_\infty \mathbf{E}(t) + \varepsilon_0 \chi_e(t) \otimes \mathbf{E}(t) \\ &= \varepsilon_0\varepsilon_\infty \mathbf{E}(t) + \varepsilon_0 \int_0^t \chi_e(t-\tau)\mathbf{E}(\tau)d\tau\end{aligned} \tag{31.36}$$

where ε_∞ denotes the high-frequency limiting value for the relative permittivity, $\chi_e(t)$ represents the electrical susceptibility tensor, and $\otimes$ denotes the time convolution. From Maxwell's equations, we obtain

$$\nabla \times \left[\frac{1}{\mu}\nabla \times \mathbf{E}(t)\right] + \varepsilon_0\varepsilon_\infty \frac{\partial^2 \mathbf{E}(t)}{\partial t^2} + \varepsilon_0 \frac{\partial^2}{\partial t^2}\left[\chi_e(t) \otimes \mathbf{E}(t)\right] = -\frac{\partial \mathbf{J}_{\text{imp}}}{\partial t} - \nabla \times \left(\frac{\mathbf{M}_{\text{imp}}}{\mu}\right) \tag{31.37}$$

Note that

$$\frac{\partial}{\partial t}\left[\chi_e(t) \otimes \mathbf{E}(t)\right] = \chi_e(0)\mathbf{E}(t) + \int_0^t \frac{d}{dt}\chi_e(t-\tau)\mathbf{E}(\tau)d\tau \tag{31.38}$$

and

$$\begin{aligned}\frac{\partial^2}{\partial t^2}\left[\chi_e(t) \otimes \mathbf{E}(t)\right] &= \chi_e(0)\frac{\partial \mathbf{E}(t)}{\partial t} + \left[\frac{d}{dt}\chi_e(t)\right]_{t=0}\mathbf{E}(t) + \int_0^t \frac{d^2}{dt^2}\chi_e(t-\tau)\mathbf{E}(\tau)d\tau \\ &= \chi_e(0)\frac{\partial \mathbf{E}(t)}{\partial t} + \dot{\chi}_e(0)\mathbf{E}(t) + \ddot{\chi}_e(t) \otimes \mathbf{E}(t)\end{aligned} \tag{31.39}$$

By using Eq. (31.39), Eq. (31.37) can be written as

$$\begin{aligned}\nabla \times \left[\frac{1}{\mu}\nabla \times \mathbf{E}(t)\right] + \varepsilon_0\varepsilon_\infty \frac{\partial^2 \mathbf{E}(t)}{\partial t^2} + \varepsilon_0\chi_e(0)\frac{\partial \mathbf{E}(t)}{\partial t} + \varepsilon_0\dot{\chi}_e(0)\mathbf{E}(t) \\ + \varepsilon_0\ddot{\chi}_e(t) \otimes \mathbf{E}(t) = -\frac{\partial \mathbf{J}_{\text{imp}}}{\partial t} - \nabla \times \left(\frac{\mathbf{M}_{\text{imp}}}{\mu}\right)\end{aligned} \tag{31.40}$$

The weak-form representation of Eq. (31.40) becomes

$$\begin{aligned}\iiint_V \left[\frac{1}{\mu}(\nabla \times \mathbf{T}) \cdot (\nabla \times \mathbf{E}) + \varepsilon_0\varepsilon_\infty \mathbf{T} \cdot \frac{\partial^2 \mathbf{E}}{\partial t^2} + \varepsilon_0\chi_e(0)\mathbf{T} \cdot \frac{\partial \mathbf{E}}{\partial t} + \varepsilon_0\dot{\chi}_e(0)\mathbf{T} \cdot \mathbf{E}\right. \\ \left. + \varepsilon_0 \mathbf{T} \cdot \ddot{\chi}_e(t) \otimes \mathbf{E}\right] dV + Y_0 \iint_{S_0} (\hat{\mathbf{n}} \times \mathbf{T}) \cdot \left(\hat{\mathbf{n}} \times \frac{\partial \mathbf{E}}{\partial t}\right) dS \\ = -\iiint_V \mathbf{T} \cdot \left[\frac{\partial \mathbf{J}_{\text{imp}}}{\partial t} + \nabla \times \left(\frac{\mathbf{M}_{\text{imp}}}{\mu}\right)\right] dV\end{aligned} \tag{31.41}$$

The finite-element discretization of Eq. (31.41) follows the same procedure as described earlier. For plasma-, Debye-, and Lorentz-type materials, the finite-element discretization of the convolution term yields $[Z]\{\psi\}$, where $[Z]$ is a constant matrix and $\{\psi\}$ is a vector whose entries are given by

$$\psi_i(t) = \varphi(t) \otimes E_i(t) \tag{31.42}$$

In this expression, $E_i\ (t)$ is the finite-element expansion coefficient for $\mathbf{E}(t)$ and $\varphi(t)$ can be written as

$$\varphi(t) = \mathrm{Re}\left[ae^{-bt}u(t)\right] \tag{31.43}$$

where $u(t)$ represents the unit step function and a and b are parameters related to specific materials. As a result, the convolution in Eq. (31.42) can be evaluated recursively as

$$\psi_i^{(n+1)} = \mathrm{Re}\left[\tilde{\psi}_i^{(n+1)}\right] \tag{31.44}$$

$$\tilde{\psi}_i^{(n+1)} = e^{-b\Delta t}\tilde{\psi}_i^{(n)} + ae^{-b(n+1)\Delta t}\int_{n\Delta t}^{(n+1)\Delta t} e^{b\tau}E_i(\tau)\,d\tau \tag{31.45}$$

Instead of assuming $E_i(t)$ to be constant in the time interval $[n\Delta\mathrm{t},\ (n+1)\Delta t]$, we can employ a linear interpolation in the evaluation of Eq. (31.45) to obtain second-order accuracy, yielding

$$\tilde{\psi}_i^{(n+1)} = e^{-b\,\Delta t}\tilde{\psi}_i^{(n)} + \frac{a\ \Delta t}{2}\left(E_i^{(n+1)} + e^{-b\,\Delta t}E_i^{(n)}\right) \tag{31.46}$$

The use of this recursive convolution significantly saves computation time and memory.

A similar idea can be employed to extend the formulation above to handle magnetically lossy and dispersive materials [53]. The treatment of anisotropic materials is as straightforward in the time domain as in the frequency domain (cf. [9] for the frequency-domain formulation).

31.4 FINITE-ELEMENT MESH TRUNCATION

As illustrated in Section 31.2, one of the major challenges in the finite-element analysis of antenna problems is the truncation of infinite space into a finite computational domain. This truncation can be accomplished by introducing an artificial surface to enclose the antenna. However, to emulate the original free-space environment, the artificial truncation surface should absorb as much of the radiated field as possible in order to reduce any artificially reflected fields. This situation is similar to what occurs in the antenna measurement in an anechoic chamber, where absorbers are used to cover the walls of the chamber so that reflections from the walls do not interfere with the measurement of antenna characteristics.

Although in principle we can use the FEM to simulate an anechoic chamber for an antenna analysis, this approach is not practical because an anechoic chamber is electrically very large, and hence its numerical simulation is extremely time consuming. Fortunately, there are many other approaches to reducing the reflection of an artificial surface. These include the use of a mathematical boundary condition, the use of fictitious absorbing material layers, and the use of a surface integral equation.

31.4.1 Absorbing Boundary Conditions

Among the three approaches for mesh truncation, the use of a mathematical boundary condition is the simplest. It is well known that when a plane wave is incident normally on a planar impedance surface, the impedance surface can absorb all the incident power without any reflection if the surface impedance matches the intrinsic impedance of free space. It can easily be verified that on the surface the electric field $\mathbf{E}$ satisfies the relation

$$\hat{\mathbf{n}} \times (\nabla \times \mathbf{E}) = -jk_0 \hat{\mathbf{n}} \times (\hat{\mathbf{n}} \times \mathbf{E}) \tag{31.47}$$

where $\hat{\mathbf{n}}$ is the normal unit vector of the surface and k_0 is the wavenumber of the plane wave. Equation (31.47) can also be considered as a boundary condition, because it provides a relationship between the tangential components of the electric and magnetic fields. When this boundary condition is applied to the artificial truncation surface, the surface will absorb any normally incident plane wave. Hence this boundary condition is often referred to as an absorbing boundary condition (ABC). Note that this condition is identical to the approximate Sommerfeld radiation condition, Eq. (31.7). In reality, the radiated field at a point on the truncation surface is usually a summation of many plane waves propagating in different directions. For plane waves incident at a large angle, a significant reflection may be caused by the boundary. However, when placed sufficiently far away from all sources of excitation, most waves would be incident on the truncation boundary only at small angles from normal and hence would be absorbed. The truncation distance, measured from the antenna to the boundary, required for good absorptive performance depends on the nature of the radiation. Typically, a minimum distance of one-half wavelength is necessary for obtaining useful results.

It can be shown that the boundary condition, Eq. (31.47), is only the first-order approximation to the true Sommerfeld radiation condition, Eq. (31.6). More terms can be included in the expansion, resulting in higher-order ABCs [54, 55]. These higher-order ABCs provide improved absorbing performances at the expense of more complicated formulations. Unfortunately, none are able to provide good absorption at near-grazing incident angles.

The major advantage of the ABC, apart from its simplicity, is that it leads to a localized relation between boundary fields and consequently preserves the highly sparse and banded pattern of the finite-element system matrices. In addition, the boundary condition can easily be incorporated into a weak-form wave equation such as Eq. (31.12). Thus the implementation of the first-order ABC in the FEM is straightforward, as demonstrated in Section 31.2.

31.4.2 Perfectly Matched Layers

Instead of using a mathematical boundary condition, the FEM mesh may also be terminated by using absorptive materials. Whereas it is extremely costly to model the absorbers used in anechoic chambers due to their electrically large thickness and their wedged and pyramidal shapes, we can instead design thin layers of artificial absorbers solely for simulation purposes.

Specifically, by adjusting the number and thickness of layers, as well as their permittivity, permeability, and conductivity (all of which can be made tensors), these artificial absorbing layers can be designed to provide negligible reflection and sufficient attenuation to the transmitted field such that the field vanishes as it propagates into the layers. An early attempt to use this method in the FEM was made in the frequency domain

[56], which provided good absorbing performance at a specific frequency. An alternative and popular approach was proposed by Berenger within the finite-difference time-domain (FDTD) setting [57] and is known as perfectly matched layers (PMLs).

A PML is an artificial material that is theoretically defined to create no reflections regardless of the frequency, polarization, and angle of incidence of a plane wave. The frequency independence is especially important because it enables broadband simulation with a time-domain method. In its original form, the PML was formulated with the aid of nonphysical "split" fields. Later, it was found that the PML could be derived alternatively from a modified form of Maxwell's equations based on stretched coordinates [58]. In addition, a PML can also be derived based on an artificial anisotropic medium model [59, 60]. Among these different formulations, the anisotropic medium model is most suitable for an FEM implementation. In this model, the PML is considered as a uniaxial anisotropic medium with diagonal permittivity and permeability tensors given by

$$\overline{\boldsymbol{\varepsilon}} = \begin{bmatrix} \varepsilon_{xx} & 0 & 0 \\ 0 & \varepsilon_{yy} & 0 \\ 0 & 0 & \varepsilon_{zz} \end{bmatrix}, \qquad \overline{\boldsymbol{\mu}} = \begin{bmatrix} \mu_{xx} & 0 & 0 \\ 0 & \mu_{yy} & 0 \\ 0 & 0 & \mu_{zz} \end{bmatrix} \tag{31.48}$$

It can be shown that the medium will be reflectionless to any impinging plane wave if $\overline{\boldsymbol{\varepsilon}}$ and $\overline{\boldsymbol{\mu}}$ are chosen according to

$$\overline{\boldsymbol{\varepsilon}} = \varepsilon\overline{\boldsymbol{\Lambda}}, \qquad \overline{\boldsymbol{\mu}} = \mu\overline{\boldsymbol{\Lambda}} \tag{31.49}$$

where

$$\overline{\boldsymbol{\Lambda}} = \begin{bmatrix} \dfrac{s_y s_z}{s_x} & 0 & 0 \\ 0 & \dfrac{s_z s_x}{s_y} & 0 \\ 0 & 0 & \dfrac{s_x s_y}{s_z} \end{bmatrix} \tag{31.50}$$

This holds true for any frequency, polarization, and incident angle; hence the name perfectly matched layers. In Eq. (31.50), the parameters s_x, s_y, and s_z are chosen based on the PML orientation. If the PML is perpendicular to the x-axis, then s_y and s_z are constant, whereas s_x can be arbitrary. If s_x is chosen to have a negative imaginary part, then the wave propagating inside the PML will be attenuated in the x-direction.

Using the property of the PML described above, we can set up a simulation domain such as the one shown in Figure 31.3. Note that for numerical simulation, the outer boundary of the PML must still be truncated, and this is typically done with a conducting surface or an ABC [61, 62]. Therefore it is necessary to make sure that the PML provides sufficient attenuation such that when the field is reflected by the truncation boundary, it becomes negligible when it reenters into the physical solution domain. This can be done relatively easily since the PML attenuation can be estimated with a simple formula. However, since the PML primarily attenuates waves propagating normal to the PML orientation, the truncated PML will provide less attenuation for obliquely incident waves; consequently, a significant nonphysical reflection can occur for waves incident at a near-grazing angle. For this reason, the PML has to be placed some distance (typically a quarter to a half of a wavelength) away from the antenna, similar to the case of a simple ABC grid termination.

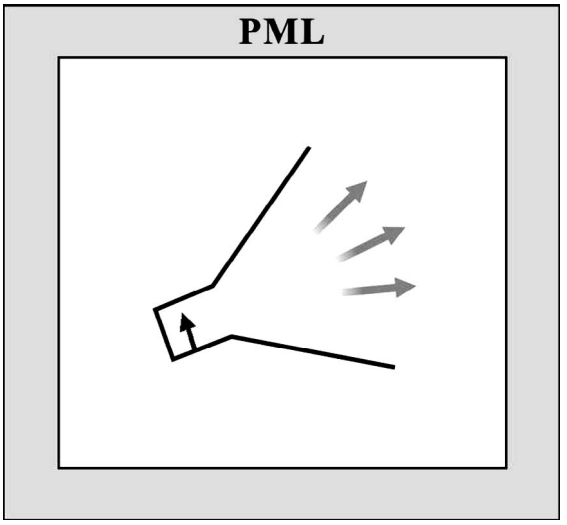


Figure 31.3 PML setup for the finite-element analysis of an antenna. Each PML sidewall is assigned parameters based on its orientation, and the outer surface is terminated by a PEC or ABC.

Although theoretically the PML interface is reflectionless, this may not hold true in numerical simulations. When an abrupt material change occurs and the FEM discretization inside the PML is not sufficiently dense to resolve the change, undesirable numerical reflections may occur [63]. One approach to avoiding this problem is to vary the material parameters smoothly within the PML. For example, for a PML perpendicular to the x-direction and interfacing with air, the PML parameters can be defined as the following mth-order polynomial:

$$s_x = 1 + \frac{\sigma_{\max}}{j\omega\varepsilon_0}\left|\frac{x - x_0}{L}\right|^m \qquad (m = 0, 1, 2, \ldots) \tag{31.51}$$

where x_0 is the x-coordinate of the PML–air interface, L is the thickness of the PML, and $\sigma_{\max}$ is the maximum conductivity inside the PML. It is found that $m = 2$ is generally a good choice. For a given thickness L, the absorbing performance of the PML can be improved by increasing the maximum conductivity and the discretization level.

Apart from the above numerical constraints, the PML is also known to have poor absorption for evanescent waves. Although the effect of evanescent waves in free-space propagation can usually be neglected, this may not be the case in waveguide and antenna simulations. Hence in these applications the PML must be placed sufficiently far away from waveguide discontinuities or the antenna. A modified version of the PML, referred to as the complex-frequency shifted (CFS) PML, has been proposed in the FDTD method to provide better absorption of evanescent waves [64, 65]. However, the CFS PML cannot absorb low frequency propagating waves. This challenge has recently been alleviated by using the second-order PML, which absorbs both the evanescent and propagating waves for all frequencies [66].

By adopting the anisotropic medium model, the implementation of the PML in the frequency-domain FEM is straightforward. The weak-form wave equation for the PML

is the same as Eq. (31.12) except that the scalar material parameters ε_r and μ_r are replaced with their tensor counterparts given in Eq. (31.49). The implementation of the PML in the time-domain FEM, however, is much more involved since the permittivity and permeability tensors are frequency dependent. Therefore the PML has to be modeled as an anisotropic, dispersive medium [67, 68]. Toward this end, the general procedure for the time-domain FEM modeling of dispersive media described in Section 31.3 can be applied.

A major advantage of the PML over the ABC approach is that the absorbing performance of the PML can be improved systematically by simply increasing the number of PML layers or equivalently by increasing the conductivity and discretization inside the PML. Despite the fact that the PML has been used extensively and successfully in the FDTD method, its application in the FEM has not achieved a similar popularity. In the frequency-domain FEM, it has been observed that the system matrix becomes poorly conditioned when PML absorbing layers are present [69, 70]. As a result, the number of iterations required by an iterative solver increases substantially. In the time-domain FEM, the PML implementation becomes much more complicated due to the need to model anisotropic dispersive media. Moreover, it has been found that an improper use of high PML conductivity values may cause undesirable instability problems. A better approach is to utilize the well established implementation of the PML in the FDTD method by combining the time-domain FEM and FDTD. This approach is discussed in Section 31.4.4.

31.4.3 Boundary Integral Equations

Both the ABC and PML approaches have a distinct advantage that they produce a highly sparse matrix, which can be generated and solved very efficiently. However, both share a common disadvantage that they are approximate, or in other words, they are not reflectionless for obliquely incident waves. Although the performance of the PML can be improved by increasing the number of layers, one still has to place it some distance (typically a quarter to a half of a wavelength) away from the antenna—thereby increasing the size of the computational volume. The third approach, to be discussed here, is the one based on boundary integral equations, which can provide a truly perfect boundary condition for mesh truncation at a higher computational cost.

It is well known that the integral-equation based methods are particularly suited for analyzing open-region radiation/scattering problems because they accurately model the wave propagation into free space via the use of appropriate Green's functions. However, these methods can encounter challenges when handling geometries with complicated materials present, typical of complex antenna and feed region designs. The FEM, on the other hand, is most suitable for modeling geometries that include complicated material compositions. The individual successes of the two methods in their own realms of applications have led to the development of a hybrid technique, known as the finite element–boundary integral (FE-BI) method [71–77], which combines the advantages of both methods and permits accurate analysis of highly complicated electromagnetic problems.

The FE-BI method employs an arbitrary boundary to terminate the FEM mesh. Interior to the boundary, the finite-element discretization is applied. Exterior to the boundary, the fields are represented by boundary integral equations. The fields in the two regions are then coupled at the boundary via field continuity conditions, leading to a coupled system

from which the interior and boundary fields can be determined. The interior problem can be formulated by rewriting the weak-form wave equation (31.9) as

$$\iiint_V \left[\frac{1}{\mu_r}(\nabla \times \mathbf{T}) \cdot (\nabla \times \mathbf{E}) - k_0^2 \varepsilon_r \mathbf{T} \cdot \mathbf{E}\right] dV + jk_0 \oiint_{S_o} \hat{\mathbf{n}} \cdot (\mathbf{T} \times \overline{\mathbf{H}})\, dS$$
$$= -\iiint_V \mathbf{T} \cdot \left[jk_0 Z_0 \mathbf{J}_{\text{imp}} + \nabla \times \left(\frac{\mathbf{M}_{\text{imp}}}{\mu_r}\right)\right] dV \tag{31.52}$$

where $\overline{\mathbf{H}} = Z_0 \mathbf{H}$. After expanding the electric field $\mathbf{E}$ according to Eq. (31.15) and a similar expansion of the surface magnetic field, Eq. (31.52) can be converted into the matrix equation

$$\begin{bmatrix} K_{II} & K_{IS} & 0 \\ K_{SI} & K_{SS} & H_{SS} \end{bmatrix} \begin{Bmatrix} E_I \\ E_S \\ \overline{H}_S \end{Bmatrix} = \begin{Bmatrix} b^{\text{imp}} \\ 0 \end{Bmatrix} \tag{31.53}$$

assuming that $\mathbf{J}_{\text{imp}}$ and $\mathbf{M}_{\text{imp}}$ are strictly inside S_o. In Eq. (31.53), $\{E_I\}$ represents the discrete electric field inside S_o, $\{E_S\}$ the discrete tangential electric field on S_o, and $\{\overline{H}_S\}$ the discrete tangential magnetic field on S_o. The matrix $[K]$ comprising the four submatrices is a sparse symmetric matrix, and the matrix $[H_{SS}]$ is also a sparse matrix.

Equation (31.53) is incomplete for solving for $\{E_I\}$, $\{E_S\}$, and $\{H_S\}$. It has to be complemented with another relation between $\{E_S\}$ and $\{H_S\}$, which comes from the formulation of the exterior field. By using the vector Green's theorem and the free-space Green's function, we can derive the following electric field integral equation (EFIE):

$$\mathbf{E} = -L(\overline{\mathbf{J}}_S) + K(\mathbf{M}_S) \tag{31.54}$$

where

$$L(\mathbf{X}) = jk_0 \oiint_{S_o} \left[\mathbf{X}(\mathbf{r}')G_0(\mathbf{r}, \mathbf{r}') + \frac{1}{k_0^2}\nabla' \cdot \mathbf{X}(\mathbf{r}')\nabla G_0(\mathbf{r}, \mathbf{r}')\right] dS' \tag{31.55}$$

$$K(\mathbf{X}) = \oiint_{S_o} \mathbf{X}(\mathbf{r}') \times \nabla G_0(\mathbf{r}, \mathbf{r}')\, dS' \tag{31.56}$$

and

$$\overline{\mathbf{J}}_S = \hat{\mathbf{n}} \times \overline{\mathbf{H}}_S, \qquad \mathbf{M}_S = \mathbf{E}_S \times \hat{\mathbf{n}} \tag{31.57}$$

Similarly, we can derive the magnetic field integral equation (MFIE) as

$$\overline{\mathbf{H}} = -K(\overline{\mathbf{J}}_S) - L(\mathbf{M}_S) \tag{31.58}$$

Either Eq. (31.54) or (31.58) can be discretized into a desired matrix equation to be solved together with Eq. (31.53). However, when used independently for a closed surface S_o, both of the EFIE and MFIE can support singular frequencies (nonphysical interior resonances), which are notorious for corrupting the numerical solutions. To effectively

eliminate this problem, we may combine the EFIE and MFIE to form the combined field integral equation (CFIE)

$$\hat{\mathbf{n}} \times \text{EFIE} + \hat{\mathbf{n}} \times \hat{\mathbf{n}} \times \text{MFIE} \tag{31.59}$$

Using the same FEM expansions for $\mathbf{E}_S$ and $\overline{\mathbf{H}}_S$, we can convert Eq. (31.59) into the matrix equation

$$[P]\{E_S\} + [Q]\{\overline{H}_S\} = 0 \tag{31.60}$$

which together with Eq. (31.53) forms a complete, coupled system

$$\begin{bmatrix} K_{II} & K_{IS} & 0 \\ K_{SI} & K_{SS} & H_{SS} \\ 0 & P & Q \end{bmatrix} \begin{Bmatrix} E_I \\ E_S \\ \overline{H}_S \end{Bmatrix} = \begin{Bmatrix} b^{\text{imp}} \\ 0 \\ 0 \end{Bmatrix} \tag{31.61}$$

for the solution of $\{E_I\}$, $\{E_S\}$, and $\{H_S\}$. An iterative solution of Eq. (31.61) can be greatly accelerated using a robust preconditioning technique [78].

Unlike the ABCs, the boundary conditions derived from the boundary integral equations are exact. This allows the truncation boundary to be placed conformal to the antenna structure to be analyzed. However, this accuracy is achieved at the cost of an increased computational burden because the boundary integrals produce full matrices for the boundary unknowns whose computation and solution are much more costly than that required for sparse matrices. Fortunately, fast algorithms such as the fast multipole method (FMM) [79, 80] can be employed to speed up the computation.

The hybridization of the FEM with the integral equation method can be achieved in the time domain as well. Recent developments in the time-domain FEM and time-domain integral equation methods have made such hybridization possible. The time-domain FE-BI method can be formulated in one of two different approaches. The first approach directly transforms the above frequency-domain FE-BI formulations into the time domain, via the Laplace transform. Thus the spatial discretization and coupling schemes remain the same as in the frequency-domain method. The second approach employs two artificial boundaries [81]. On the exterior boundary, a first-order mixed boundary condition is postulated. Given the field calculated by the FEM at a specific time step, the time-domain boundary integral equations are used to compute the mixed boundary condition on the exterior boundary at the next time step from the field data on the interior boundary. This mixed boundary condition is then used for the FEM calculation of the internal field at the next time step. The process continues for all time steps. Again, in this case, the calculation of the time-domain boundary integrals can be accelerated using a fast algorithm, as demonstrated in Jiao et al. [81].

31.4.4 FDTD Interface

The finite-difference time-domain (FDTD) method has been widely used for the analysis and design of antennas [82]. The basic FDTD technique is highly efficient since it utilizes both a structured grid and explicit time marching. A three-dimensional *structured* grid can be referenced by three integer indices, typically *I-J-K*, and as such storage of the full mesh is not required. In contrast, finite-element grids based on tetrahedral elements are typically *unstructured*, and in this case the entire mesh generally needs to be stored and a simple integer cell-referencing scheme is generally not possible. In addition, because the standard

FDTD algorithm uses *explicit* time marching, inversion of the governing system matrix is simple because this matrix is diagonal. The FDTD method can also be parallelized in a distributed sense in a very straightforward manner [83], and the PML grid-truncation technology is mature and numerically stable [60]. These features make the FDTD method a powerful tool; however, because the FDTD technique is primarily based on a rectilinear grid, the accurate modeling of complex antenna feed regions and/or complex antenna geometry may require an exceedingly fine level of spatial discretization along with large computational resources. In addition, because the formulation is conditionally stable, which is a requirement for explicit time marching, the Courant limit dictates that a fine spatial discretization also requires an increase in the number of time steps to obtain an accurate solution.

An alternative approach to the global application of FDTD is to utilize a hybridization of the time-domain FEM with the FDTD method [84–91]. In this way, the antenna details can be characterized by the unconditionally stable and conformal aspects of the FEM, while the efficiency of FDTD can be utilized in the external regions of the problem domain. The advantage of this approach is that fine discretization, common in antenna feed regions such as spirals and other antennas, does not require an increase in the number of time steps. At the same time, the efficiency of FDTD is advantageous in the air surrounding the antenna. An example of such an interface is shown in Figure 31.4. This hybrid FDTD–FEM concept is provably numerically stable up to the Courant limit of the FDTD region that supports the finite-element domains [89, 91]. In other words, the grid associated with the internal FEM regions can expand and contract as needed to

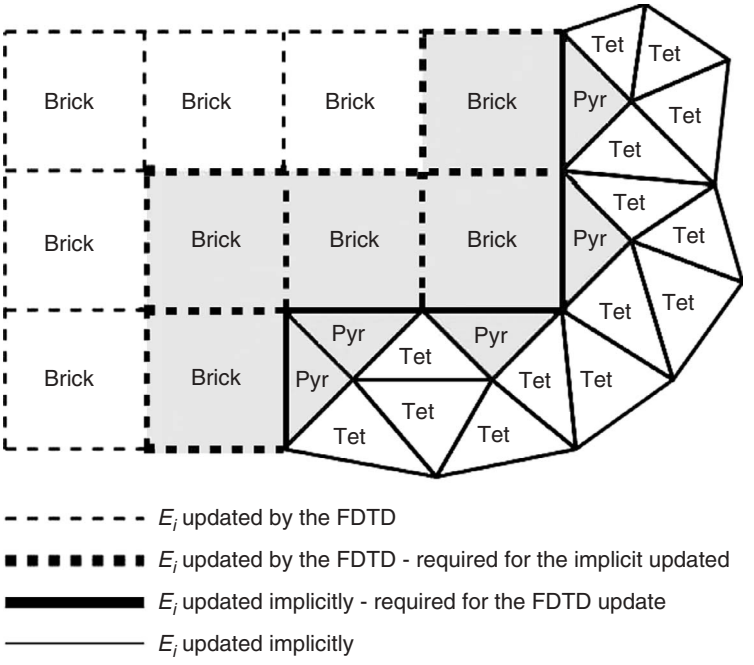


Figure 31.4 Interface between time-domain FEM and FDTD grid regions. "Brick" denotes rectangular hexahedral (FDTD) elements, "Pyr" denotes pyramidal transition elements, and "Tet" denotes tetrahedral elements. (Adapted from Ref. 91; Copyright © 2005 Artech House.)

characterize the local geometry without imposing further restrictions on the global time step with regard to numerical stability.

The success of this hybridization is fundamentally dependent on an underlying equivalence that exists between the FDTD and the FEM formulations that are based on vector edge basis functions. Specifically, when trapezoidal integration is applied in the FEM to rectangular hexahedral cells, the FDTD algorithm can be recovered [91, 92]. Consequently, a symmetric interface between these two methods can be constructed, which is typically a necessary condition for proving numerical stability [91].

31.5 ANTENNA PARAMETER MODELING

To characterize an antenna, the principal measured parameters of interest are the input impedance and radiation patterns, from which other parameters, such as antenna gain, can be derived. For example, the reflection coefficient and the voltage standing wave ratio (VSWR) observed at the feed can be calculated from the input impedance, and the antenna directivity and gain can be calculated from the radiation pattern. The calculation of the antenna input impedance depends on the feed model employed in the numerical simulation; hence an appropriate feed model is often essential to obtaining correlation of measured and predicted input impedance. Simplified feed models, such as voltage gap sources, are frequently used to predict accurate far-field patterns and in initial antenna design studies. By using Huygens' principle [9], the antenna radiation pattern can be computed from a near-field surface integral based on equivalent currents located on the surfaces within the FEM mesh. Because integration tends to average out small errors in the currents, antenna radiation patterns are much less sensitive to the fidelity of the feed model than are impedance predictions.

In this section, the numerical modeling of the input impedance and far-field patterns are discussed in some detail. In addition to these two important antenna characteristics, the near-field patterns may also reveal useful information about the underlying physics of the antenna performance. Therefore the visualization for the near field is also briefly described at the end of the section.

31.5.1 Feed Modeling

A proper antenna feed model not only provides the desired excitation to the antenna but also provides a means by which the input impedance and other parameters can be calculated. Despite the various types of feeds an antenna may actually employ, the numerical models for an antenna feed can be classified into two categories. The first type of feed model incorporates the precise feed geometry into the numerical simulation. Although numerically accurate, these models increase the burden of mesh generation. More importantly, challenges may arise when the antenna feed is small in dimension as compared to the wavelength. In these situations, a much denser spatial discretization has to be employed in the vicinity of the feed to resolve the feed geometry, resulting in a highly imbalanced mesh over the entire computation domain. When handled inappropriately, this may cause a significant increase in run time due to an increase in the condition number of the matrix.

The second type of feed model employs a simplified version (usually the low frequency approximation) of the actual antenna feed. Such models are numerically convenient and

efficient and do not require especially dense spatial discretization in the vicinity of the feed. Moreover, they enable the separation of antenna properties from the effect of feeding networks. This may be helpful during the design process in which the objective is to understand the behavior of the antenna itself. However, the accuracy of these models may depend on the frequency, the antenna type, and the parameters of interest.

31.5.1.1 Current Probe Simplified feed models are widely used in the numerical simulation of antennas. Although a voltage source, such as a delta-gap source, is commonly used in the method of moments (MoM) simulations, it is more convenient to model a current source in the FEM simulations. The simplest current source model is a short and infinitesimally thin current probe inserted at the feed point of the antenna, as illustrated in Figure 31.5a. In this simple model, the excitation current density $\mathbf{J}_{\text{imp}}$ can be modeled as a delta function in three-dimensional space; for example, a short current probe oriented in the z-direction can be modeled as

$$\mathbf{J}(x, y, z) = \hat{\mathbf{z}} I_o \delta(x, y), \quad z \in L \tag{31.62}$$

where I_{o} is the lumped current flowing into the antenna and L is the length of the probe. Such a current source can also be considered as an infinitesimal dipole with a constant current distribution. Using Eq. (31.62) as excitation in Eq. (31.8) together with proper boundary conditions, the electric field everywhere inside the computational domain can be uniquely determined. Once the electric field is determined, the voltage across the probe can be computed according to

$$V = \int_L \mathbf{E} \cdot \hat{\mathbf{l}} \, dl \tag{31.63}$$

and the input impedance of the antenna, by its definition, can be calculated as

$$Z_{\text{in}} = V/I \tag{31.64}$$

The use of an electrically small input current source corresponds to a Norton equivalent network for the source. As such, if one desired a voltage source with amplitude V_o and internal resistance R_o, then the definition of I_o in Eq. (31.62) would be V_o/R_o. In addition,

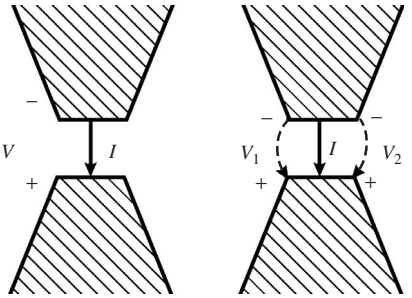


Figure 31.5 Electric current probe excitation: (a) simple model and (b) improved model that calculates the voltage along offset lines. (Adapted from Ref. 43; copyright © 2005 IEEE.)

the input current I in Eq. (31.64) can be represented by $I = I_o - V/R_o$. This modeling strategy can be extended to model commonly used circuit elements in the FEM [93].

The current probe model can be used to approximate a variety of actual antenna feeds, such as coaxial cable and microstrip lines. The length of the probe has to be small compared to the shortest wavelength in the frequency band of interest in order for the distributed effect to be negligible. The electric current probe model is usually accurate in predicting the far-field behaviors such as radiation pattern and polarization characteristics. For predicting parameters such as the input impedance and VSWR, the accuracy may not be sufficient. Unfortunately, the accuracy cannot be improved by employing a finer discretization. One reason for the inaccuracy of the current probe model is that the simplified numerical model no longer corresponds exactly to the original geometry of the antenna feed. For example, in order to insert the current probe, the inner conductor of the feed is often broken to create a gap. Another intrinsic drawback in this model is that since a radius is not specified for the probe, the electromagnetic field is singular in the vicinity of the probe, even though this singularity is dampened in the numerical solution. Numerical errors due to this drawback will be most apparent in parameters such as input impedance because their evaluation depends heavily on the localized field solution.

To avoid the challenge caused by the singular field at the probe, we can calculate the voltage using the field on a cylindrical surface centered at the probe and having a radius equal to that of the physical feed (the inner conductor in the case of a coaxial feed). Since the field on this cylindrical surface is quite axisymmetric, it is sufficient to calculate the field along one or two lines (which are called the observation probes here) on the surface, as illustrated in Figure 31.5b. When two observation probes are used, the terminal voltage is then calculated as the averaged voltage on the two probes. Numerical experiments show that the input impedance calculated by this improved probe feed model provides improved agreement with precise feed modeling. On certain occasions, there may still exist a small amount of discrepancy in the imaginary part of the input impedance due to the additional self-susceptance associated with the artificial gap introduced by the probe model. However, such discrepancy can easily be compensated in the numerical simulation if a reasonably good numerical model for the self-susceptance of the feed can be established [43].

31.5.1.2 Distributed Voltage Source (Gap Generator) As previously noted, the concept of the delta-gap voltage generator is widely used for the analysis of linear antennas based on the MoM technique [94]. Although the current probe model discussed in the preceding section is commonly used in FEM analyses, the delta-gap voltage generator is also straightforward to apply within the FEM. The delta-gap source assumes that the impressed voltage induces an incident electric field that exists only within the region of the gap. A simple dipole antenna is shown in Figure 31.6 and the feed region is represented by the surface S_g, which corresponds to a "belt" of width d ($\ll \lambda$) around the circumference of the $\hat{\mathbf{z}}$-directed linear antenna with radius a. The potential difference across the belt is defined to be V_o. Consequently, the electric field on S_g can be defined as

$$\mathbf{E}(\mathbf{r}) = -\frac{V_o}{d}\hat{\mathbf{z}}, \quad \mathbf{r} \in S_g \tag{31.65}$$

In the finite-element setting, this expression can conveniently be incorporated as a Dirichlet boundary condition by projecting the defined electric field onto the finite-element edges located on S_g. Specifically, the coefficient for the gap electric field associated with the ith finite-element edge on S_g is defined by

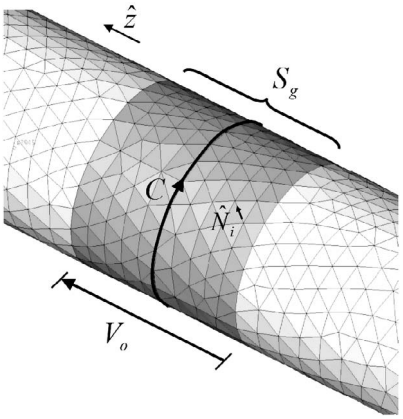


Figure 31.6 Linear dipole antenna showing local spatial grid region for impressing a distributed voltage source for modeling a delta-gap voltage generator.

$$E_i(t) = -\frac{V_o}{d}\hat{\mathbf{z}} \cdot \hat{\mathbf{N}}_i \tag{31.66}$$

The input impedance is subsequently obtained by evaluating

$$Z_{\text{in}} = \frac{V_o}{\oint_C \mathbf{H} \cdot d\mathbf{l}} \tag{31.67}$$

where the integration contour C is illustrated in Figure 31.6.

31.5.1.3 Waveguide Feed Use of a waveguide port often minimizes the problem size when modeling structures excited by a waveguide. For instance, in the simulation of a horn antenna fed by a rectangular waveguide, it is possible to employ a waveguide feed model to account for the modal distribution inside the waveguide. This approach eliminates the need to model the details of the probe excited the waveguide itself, which may be some distance down the waveguide.

Antennas are commonly fed by waveguide structures, such as coaxial cables, striplines, and rectangular and circular waveguides. An accurate FEM analysis of waveguide feeds requires that an appropriate boundary condition be imposed on the waveguide port that is locally used to truncate the mesh. The desired boundary condition should be able to launch an incident wave into the waveguide and, at the same time, absorb the reflected wave from the waveguide discontinuity without any spurious reflection. Moreover, it is desirable to place the port boundary as close to the discontinuity as possible to reduce the computation domain. One approach to terminating a waveguide port is to use the PML. However, the original PML was not designed to absorb the evanescent waves that may be present in the waveguide, although a modified formulation of the PML has been proposed recently to improve its ability to absorb evanescent waves without compromising its ability to absorb propagating waves [66].

A more accurate approach is to impose a mixed boundary condition, specifically designed for the waveguide, on the waveguide port [9]. The exact boundary condition can be derived based on waveguide modal expansions. All waveguide modes, including both

propagation and evanescent modes, can be perfectly absorbed by the boundary condition. This allows the waveguide to be truncated relatively close to the waveguide discontinuity. The boundary condition, referred to as the waveguide port boundary condition (WPBC), and its implementation in the FEM are described in the following.

It is well known that the electric field in a homogeneous waveguide can be expressed as the superposition of orthogonal TEM (if it exists), TE, and TM waveguide modes as

$$\begin{aligned}\mathbf{E} &= \mathbf{E}^{\text{inc}} + \mathbf{E}^{\text{ref}} \\ &= \mathbf{E}^{\text{inc}} + a_0\mathbf{e}_0^{\text{TEM}} + \sum_{m=1}^{\infty} a_m\mathbf{e}_m^{\text{TE}} + \sum_{m=1}^{\infty} b_m\mathbf{e}_m^{\text{TM}}\end{aligned} \tag{31.68}$$

where m is the modal index, a_m and b_m are the modal amplitudes, and $\mathbf{e}_0^{\text{TEM}}$, $\mathbf{e}_m^{\text{TE}}$, and $\mathbf{e}_m^{\text{TM}}$ are the modal functions for the TEM, TE, and TM modes, respectively. Although analytical modal functions can be found for certain waveguide geometries, a numerical method has to be employed in order to compute the modal functions for an arbitrary waveguide cross section. Applying a curl operator to both sides of Eq. (31.68) yields

$$\begin{aligned}\hat{\boldsymbol{n}} \times (\nabla \times \mathbf{E}) = \hat{\mathbf{n}} \times (\nabla \times \mathbf{E}^{\text{inc}}) + jka_0\mathbf{e}_0^{\text{TEM}} \\ + \sum_{m=1}^{\infty} \sqrt{k_{cm}^2 - k^2}\, a_m\mathbf{e}_m^{\text{TE}} + \sum_{m=1}^{\infty} \frac{-k^2}{\sqrt{k_{cm}^2 - k^2}} b_m\mathbf{e}_{tm}^{\text{TM}}\end{aligned} \tag{31.69}$$

where k_{cm} is the cutoff wavenumber of the mth mode and $\mathbf{e}_{tm}^{\text{TM}}$ denotes the transverse component of $\mathbf{e}_m^{\text{TM}}$. By modal orthogonality, the modal amplitudes a_m and b_m can be found as

$$a_0 = \iint_S \mathbf{e}_0^{\text{TEM}} \cdot [\mathbf{E} - \mathbf{E}^{\text{inc}}]dS \tag{31.70}$$

$$a_m = \iint_S \mathbf{e}_m^{\text{TE}} \cdot [\mathbf{E} - \mathbf{E}^{\text{inc}}]dS, \quad m = 1, 2, \ldots \tag{31.71}$$

$$b_m = \iint_S \mathbf{e}_{tm}^{\text{TM}} \cdot [\mathbf{E} - \mathbf{E}^{\text{inc}}]dS, \quad m = 1, 2, \ldots \tag{31.72}$$

Substituting Eqs. (31.70)–(31.72) into Eq. (31.69), Eq. (31.69) can be written in a compact form as

$$\hat{\mathbf{n}} \times (\nabla \times \mathbf{E}) + P(\mathbf{E}) = \mathbf{U}^{\text{inc}} \tag{31.73}$$

where

$$\begin{aligned}P(\mathbf{E}) = -jk\mathbf{e}_0^{\text{TEM}} \iint_S \mathbf{e}_0^{\text{TEM}} \cdot \mathbf{E}\, dS - \sum_{m=1}^{\infty} \sqrt{k_{cm}^2 - k^2}\mathbf{e}_m^{\text{TE}} \iint_S \mathbf{e}_m^{\text{TE}} \cdot \mathbf{E} dS \\ - \sum_{m=1}^{\infty} \frac{-k^2}{\sqrt{k_{cm}^2 - k^2}} \mathbf{e}_{tm}^{\text{TM}} \iint_S \mathbf{e}_{tm}^{\text{TM}} \cdot \mathbf{E} dS\end{aligned} \tag{31.74}$$

$$\mathbf{U}^{\text{inc}} = \hat{\mathbf{n}} \times (\nabla \times \mathbf{E}^{\text{inc}}) - jk\mathbf{e}_0^{\text{TEM}} \iint_S \mathbf{e}_0^{\text{TEM}} \cdot \mathbf{E}^{\text{inc}} dS - \sum_{m=1}^{\infty} \sqrt{k_{cm}^2 - k^2}\mathbf{e}_{tm}^{\text{TE}} \iint_S \mathbf{e}_{tm}^{\text{TE}} \cdot \mathbf{E}^{\text{inc}}\, dS - \sum_{m=1}^{\infty} \frac{-k^2}{\sqrt{k_{cm}^2 - k^2}}\mathbf{e}_{tm}^{\text{TM}} \iint_S \mathbf{e}_{tm}^{\text{TM}} \cdot \mathbf{E}^{\text{inc}}\, dS \tag{31.75}$$

Equation (31.73) is referred to as the WPBC [95], which can readily be incorporated into the frequency-domain weak-form wave equation (31.12), yielding

$$\iiint_V \left[\frac{1}{\mu_r}(\nabla \times \mathbf{T}) \cdot (\nabla \times \mathbf{E}) - k_0^2 \varepsilon_r \mathbf{T} \cdot \mathbf{E}\right] dV + jk_0 \oiint_{S_o} (\hat{\mathbf{n}} \times \mathbf{T}) \cdot (\hat{\mathbf{n}} \times \mathbf{E})\, dS - \iint_{S_p} \frac{1}{\mu_r}\mathbf{T} \cdot P(\mathbf{E})\, dS = -\iint_{S_p} \frac{1}{\mu_r}\mathbf{T} \cdot \mathbf{U}^{\text{inc}} dS - jk_0 Z_0 \iiint_V \mathbf{T} \cdot \mathbf{J}_{\text{imp}}\, dV \tag{31.76}$$

where S_p denotes the waveguide port surface. Following the same spatial discretization procedures outlined in Section 31.2, a linear system of equations can be derived from Eq. (31.76) for the unique determination of the electric field everywhere inside the waveguide. The procedure described above can be extended to inhomogeneously filled waveguides.

The above frequency-domain WPBC can be converted to the time domain in a straightforward manner. The Laplace transform of Eqs. (31.74) and (31.75) yields

$$P(\mathbf{E}) = -\mathbf{e}_0^{\text{TEM}} \iint_S \mathbf{e}_0^{\text{TEM}} \cdot \frac{1}{c}\frac{\partial}{\partial t}\mathbf{E} dS - \sum_{m=1}^{\infty} \mathbf{e}_m^{\text{TE}} \iint_S \mathbf{e}_m^{\text{TE}} \cdot \left(\frac{1}{c}\frac{\partial}{\partial t}\mathbf{E} + h_m \otimes \mathbf{E}\right) dS - \sum_{m=1}^{\infty} \mathbf{e}_{tm}^{\text{TM}} \iint_S \mathbf{e}_{tm}^{\text{TM}} \cdot \left(\frac{1}{c}\frac{\partial}{\partial t}\mathbf{E} + g_m \otimes \mathbf{E}\right) dS \tag{31.77}$$

$$\mathbf{U}^{\text{inc}} = \hat{\mathbf{n}} \times (\nabla \times \mathbf{E}^{\text{inc}}) - \mathbf{e}_0^{\text{TEM}} \iint_S \mathbf{e}_0^{\text{TEM}} \cdot \frac{1}{c}\frac{\partial}{\partial t}\mathbf{E}^{\text{inc}} dS - \sum_{m=1}^{\infty} \mathbf{e}_m^{\text{TE}} \iint_S \mathbf{e}_m^{\text{TE}} \cdot \left(\frac{1}{c}\frac{\partial}{\partial t}\mathbf{E}^{\text{inc}} + h_m \otimes \mathbf{E}^{\text{inc}}\right) dS - \sum_{m=1}^{\infty} \mathbf{e}_{tm}^{\text{TM}} \iint_S \mathbf{e}_{tm}^{\text{TM}} \cdot \left(\frac{1}{c}\frac{\partial}{\partial t}\mathbf{E}^{\text{inc}} + g_m \otimes \mathbf{E}^{\text{inc}}\right) dS \tag{31.78}$$

where $\otimes$ again stands for time-domain convolution and h_m and g_m are time-domain functions with expressions that are found analytically as

$$h_m(t) = \frac{k_{cm}}{t} J_1(k_{cm}ct)u(t) \tag{31.79}$$

$$g_m(t) = \frac{k_{cm}}{t} J_1(k_{cm}ct)u(t) - k_{cm}^2 c J_0(k_{cm}ct)u(t) \tag{31.80}$$

where $u(t)$ denotes a step function and J_0 and J_1 denote the first- and second-order Bessel functions, respectively. Equations (31.73), (31.77), and (31.78) can easily be incorporated into the time-domain weak-form wave equation (31.24), which after spatial and temporal discretization can be converted into a time-marching system [95].

In both of its frequency- and time-domain versions, the boundary condition (31.73) is expressed as a summation of an infinite number of waveguide modes. When the series are not truncated, the expression is the exact boundary condition satisfied on the waveguide port. In a numerical implementation, however, the infinite summation is always truncated and hence the boundary condition is only approximate. Nevertheless, since the magnitudes of the higher-order evanescent modes decay quickly, it is usually sufficient to include only a small number of modes in the boundary condition.

31.5.2 Input Impedance Computation

In many applications, the feed point of an antenna can be modeled as a pair of terminals which are connected to a voltage or current generator. The input impedance of the antenna is then defined as the ratio of the voltage across the terminals to the current flowing through the terminals. The computation of the antenna input impedance depends on the specific feed model that is employed in the numerical simulation. If a current probe source is employed, then the input impedance is simply calculated using Eq. (31.64), where the voltage V is calculated from the electric field via Eq. (31.63).

If a waveguide model is employed to model the antenna feed, the reflection coefficient (or S_{11}) at the waveguide port can be extracted from the FEM solutions according to

$$R = \frac{\iint\limits_{S_p} (\mathbf{E} - \mathbf{E}^{\text{inc}}) \cdot \mathbf{E}^{\text{inc}} dS}{\iint\limits_{S_p} |\mathbf{E}^{\text{inc}}|^2 dS} \tag{31.81}$$

Then the input impedance can be calculated from the reflection coefficient R according to

$$Z_{\text{in}} = \frac{1+R}{1-R} Z_0 \tag{31.82}$$

where Z_0 is the characteristic impedance of the waveguide.

31.5.3 Far-Field Pattern Computation

The far field radiated by an antenna can be calculated numerically from the near field on an arbitrary surface completely enclosing the antenna structure, denoted here as S_{NTF}. This procedure is referred to as the near-to-far-field transformation. Given the near-zone electromagnetic field on S_{NTF}, it can be shown that the electric field in the far zone can be expressed as

$$E_\theta(\mathbf{r}) = -\frac{jk_0 e^{-jk_0 r}}{4\pi r}(L_\phi + Z_0 N_\theta) \tag{31.83}$$

$$E_\phi(\mathbf{r}) = \frac{jk_0 e^{-jk_0 r}}{4\pi r}(L_\theta - Z_0 N_\phi) \tag{31.84}$$

where Z_0 is the free-space impedance and

$$\mathbf{N}(\hat{\mathbf{r}}) = \iint\limits_{S_{\text{NTF}}} \mathbf{J}(\mathbf{r}')e^{jk_0\mathbf{r}'\cdot\hat{\mathbf{r}}}\,dS' \tag{31.85}$$

$$\mathbf{L}(\hat{\mathbf{r}}) = \iint\limits_{S_{\text{NTF}}} \mathbf{M}(\mathbf{r}')e^{jk_0\mathbf{r}'\cdot\hat{\mathbf{r}}}\,dS' \tag{31.86}$$

where $\mathbf{J} = \hat{\mathbf{n}} \times \mathbf{H}$ and $\mathbf{M} = -\hat{\mathbf{n}} \times \mathbf{E}$ are equivalent surface electric and magnetic currents, respectively. The near-to-far-field transformation surface S_{NTF} must be a closed surface containing all the sources inside. If an infinite ground plane is present (e.g., in the case of monopole antennas), the near-to-far-field transformation surface S_{NTF} and its image S'_{NTF} together comprise a closed surface. The equivalent surface currents on S'_{NTF} can easily be obtained by invoking image theory.

In time-domain simulations, the far field can be calculated in either the frequency or the time domain. In the frequency-domain approach, the time-domain surface currents $\mathbf{J}$ and $\mathbf{M}$ are transformed to the frequency domain via FFT, or alternatively their frequency-domain representation at specific frequencies can be constructed concurrently with the time-stepping process via a DFT. Next, the frequency-domain far-field pattern is calculated using Eqs. (31.83)–(31.86). In the time-domain approach, the far-field pattern is directly calculated from the time-domain surface currents $\mathbf{J}$ and $\mathbf{M}$ according to

$$E_\theta(\mathbf{r}, t) = -\frac{1}{4\pi r c_0}\left[\frac{\partial}{\partial t}L_\phi\left(\hat{\mathbf{r}}, t - \frac{\mathbf{r}\cdot\hat{\mathbf{k}}}{c_0}\right) + Z_0\frac{\partial}{\partial t}N_\theta\left(\hat{\mathbf{r}}, t - \frac{\mathbf{r}\cdot\hat{\mathbf{k}}}{c_0}\right)\right] \tag{31.87}$$

$$E_\phi(\mathbf{r}, t) = \frac{1}{4\pi r c_0}\left[\frac{\partial}{\partial t}L_\theta\left(\hat{\mathbf{r}}, t - \frac{\mathbf{r}\cdot\hat{\mathbf{k}}}{c_0}\right) - Z_0\frac{\partial}{\partial t}N_\phi\left(\hat{\mathbf{r}}, t - \frac{\mathbf{r}\cdot\hat{\mathbf{k}}}{c_0}\right)\right] \tag{31.88}$$

where

$$\mathbf{N}(\hat{\mathbf{r}}, \tau) = \iint\limits_{S_{\text{NTF}}} \mathbf{J}\left(\mathbf{r}', \tau + \frac{\mathbf{r}'\cdot\hat{\mathbf{r}}}{c_0}\right)dS' \tag{31.89}$$

$$\mathbf{L}(\hat{\mathbf{r}}, \tau) = \iint\limits_{S_{\text{NTF}}} \mathbf{M}\left(\mathbf{r}', \tau + \frac{\mathbf{r}'\cdot\hat{\mathbf{r}}}{c_0}\right)dS' \tag{31.90}$$

In terms of efficiency, the frequency-domain near-to-far-field transformation is most suitable for computing far-field patterns with dense angular sampling at only a few frequencies, while the time-domain near-to-far-field transformation is most suitable for computing far-field patterns over a wide frequency range at only a few observation angles.

31.5.4 Near-Field Visualization

Antenna measurements typically provide knowledge of impedance, radiation pattern, and gain. Because measurements frequently reveal that improvements to a particular antenna's performance are necessary, it becomes important to understand the physical

cause of the undesired performance characteristics. Visualization of near fields provides insight that is not typically available through measurement. In this way, analysis augments measurement, allowing antenna designers to perform analytical experiments.

Visualization can be used in numerous useful ways, each providing insight into the physics driving antenna performance. Each antenna application will drive the specific use of field visualizations. While it is impossible to discuss all possible uses here, a few examples are given where field visualizations have proved their usefulness to practicing antenna engineers.

Visualization of the electric and magnetic vector fields, as well as the Poynting vector, often provides valuable physical insight. For instance, strong magnetic field lines might unintentionally couple the horizontal and vertical ports of an antenna, causing degraded cross polarization. Alternatively, when designing antennas having anisotropic substrates, visualization of the fields provides physical insight into how the various field components are affected by the anisotropy. As a final example, visualization of the Poynting vector can be extremely useful toward providing printed-antenna designers insight into the physical mechanisms related to substrate losses, including surface wave phenomena.

31.6 PHASED ARRAY MODELING

Phased array antennas are of great importance in modern radar and communication systems. Accurate prediction of array parameters using numerical methods not only reduces the development cost and design period but also provides invaluable information to design engineers. Increasing demands on array performances may require nontraditional designs using anisotropic dispersive materials or complex radiating elements. Furthermore, a successful antenna array design may require a complicated feeding and matching network, which by itself may impose certain numerical challenges. Among the various numerical methods, the FEM appears to be well suited to perform such analyses due to its versatility in geometry and material modeling.

In this section, two general scenarios in the array analysis are considered. In the first scenario, the array is considered to be infinitely periodic, usually in a two-dimensional xy-plane. Although not a realistic configuration, the infinite array model provides a good approximation to the performance of the interior elements in a large finite array. The numerical analysis of an infinite array is relatively easy because the computation domain can be confined to a single array element (unit cell) due to the periodicity in the electromagnetic fields. In the second scenario, arrays with a finite size are faithfully modeled. Since the electromagnetic fields no longer satisfy the periodic relations, the computation domain has to be extended to host the entire array. However, the repetition in periodic array geometries can often be exploited in the numerical simulation, which helps to lessen the tremendous computational burden typically associated with a finite array analysis.

31.6.1 Infinite Arrays

An infinitely periodic structure with a uniform progressively phased excitation enables the numerical analysis to be confined to a single array element, or a unit cell. In the interior region of the unit cell, the standard FEM discretization is applied. On the boundaries of the unit cell, proper boundary conditions are imposed to correctly model the field

behavior on both the periodic and radiation boundaries. Such boundary conditions have been developed in both the frequency and time domains.

31.6.1.1 Frequency-Domain Modeling When a phased array is operating with harmonic time dependence (frequency domain), each antenna element is excited by an individual source with a specific phase difference relative to the other sources, which results in a main beam being radiated in a desired direction. In order to steer the beam, the relative phase shift between the elements is changed such that the beam radiates in a new direction. An infinite antenna array can be analyzed using a single unit cell, as shown in Figure 31.7, where the field at one periodic surface of a unit cell is related to the field at the opposite parallel surface of the unit cell through a simple phase shift. This fact makes it straightforward to utilize a unit cell to numerically analyze infinitely periodic phased array antenna structures in the frequency domain.

Consider a planar array extending infinitely in the xy-plane as shown in Figure 31.7. According to the Floquet theorem [15], the electric field satisfies the following periodic relation:

$$\mathbf{E}(x + T_x, y + T_y) = \mathbf{E}(x, y)e^{-j(k_x^s T_x + k_y^s T_y)} \tag{31.91}$$

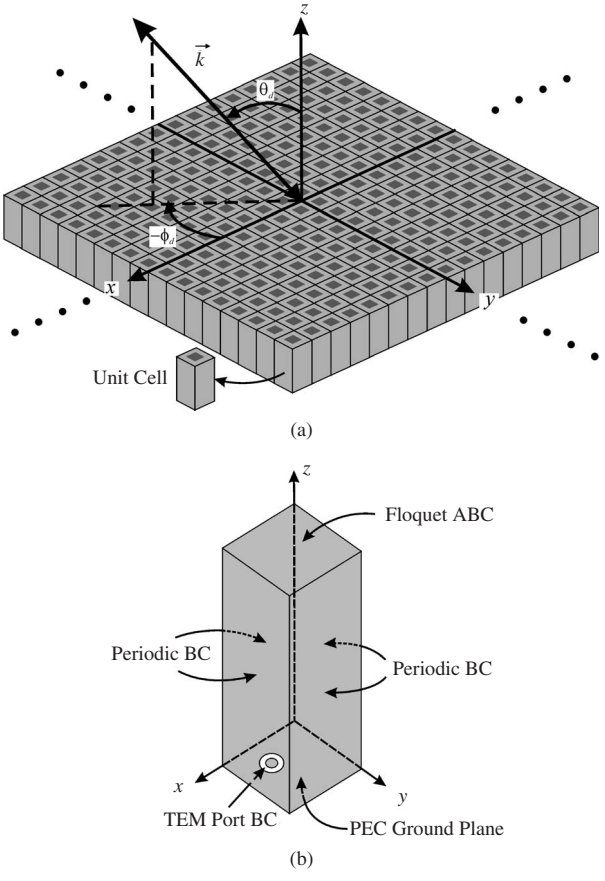


Figure 31.7 (a) Infinite periodic array and (b) unit cell assuming a ground plane and waveguide port excitation. (Adapted from Ref. 47; copyright © 2006 IEEE.)

where T_x and T_y are the periodicities in the x- and y-directions, $k_x^s = k_0 \sin\theta_s \cos\phi_s$ and $k_y^s = k_0 \sin\theta_s \sin\phi_s$ with (θ_s, ϕ_s) being the scan angle of the array. Applying Eq. (31.91) to the four sidewalls of the unit cell results in a periodic boundary condition that relates the electric field on the opposite sidewalls with a phase shift term determined from the scan angles of the array.

Since an infinite ground plane has been assumed on the lower surface of the array, a radiation boundary condition is required only on the top surface. Although approximate radiation boundary conditions, such as the first-order ABC, can be applied on this surface, a rigorous, periodic version of the boundary integral equation is presented in this section. Due to the periodicity in the x- and y-directions, the integration in the spatial domain can be carried out as a double summation in the spectral domain using the Fourier transform, and the equation can be written as [17]

$$\mathbf{H}(\mathbf{r}) = -2jk_0 Y_0 \sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} \tilde{\overline{\mathbf{G}}}(k_{xp}, k_{yq}) \cdot \tilde{\mathbf{M}}_{pq} e^{-j(k_{xp}x + k_{yq}y)} \tag{31.92}$$

where $\tilde{\mathbf{M}}_{pq}$ is the two-dimensional (2D) Fourier transform of the magnetic current on the top surface S_T

$$\tilde{\mathbf{M}}_{pq} = \frac{1}{T_x T_y} \iint_{S_T} \mathbf{M}(x, y) e^{j(k_{xp}x + k_{yq}y)} dS \tag{31.93}$$

and $\tilde{\overline{\mathbf{G}}}(k_{xp}, k_{yq})$ is the dyadic Green's function in the spectral domain, which can be expressed explicitly as

$$\tilde{\overline{\mathbf{G}}}(k_{xp}, k_{yq}) = \frac{1}{2jk_z k_0^2} \begin{bmatrix} k_0^2 - k_{xp}^2 & -k_{xp}k_{yq} \\ -k_{xp}k_{yq} & k_0^2 - k_{yq}^2 \end{bmatrix} \tag{31.94}$$

where

$$k_{xp} = \frac{2\pi}{T_x} p + k_x^s, \qquad k_{yq} = \frac{2\pi}{T_y} q + k_y^s, \qquad k_z = \sqrt{k_0^2 - k_{xp}^2 - k_{yq}^2} \tag{31.95}$$

The integral equation (31.92) enforces an exact radiation boundary condition on S_T in the presence of the infinite array. It is therefore referred to as the periodic radiation boundary condition. Numerically, the infinite summations in Eq. (31.92) are always truncated with $-M \le p \le M$ and $-N \le q \le N$, where M and N are the highest order of the free-space propagating Floquet modes included in the summations. Larger array periods permit more propagating free-space Floquet modes; hence more terms should be included in the summation. Meanwhile, it is also important that the spatial discretization on S_T should be dense enough to resolve the spatial variation of the highest-order free-space Floquet mode.

31.6.1.2 Time-Domain Modeling As illustrated above, when analyzing an infinite periodic antenna array using a unit cell, it is necessary to impose periodic boundary conditions that reflect a uniform phase shift between the antenna elements. This is straightforward in the frequency domain. It is, however, more challenging in the time

domain when the antenna elements are excited by a pulse in time with an appropriate time shift to steer the main beam into a particular direction. The difficulty of enforcing the periodic boundary conditions for this case is due to the fact that the simple phase shift in the frequency domain translates into a time shift in the time domain. This time shift of the field at two parallel periodic surfaces requires future knowledge of the field at one of the two surfaces in order to enforce periodic boundary conditions. Clearly, this is a major obstacle for a time-domain numerical method to overcome. There have been various methods suggested in the literature to tackle the problem of periodic boundary conditions in the time domain for FDTD analysis [96, 97] and for integral equation analysis [98]. One of the techniques originally suggested for periodic FDTD analysis has recently been adopted for time-domain FEM analysis [46, 47]. This method is based on transforming Maxwell's equations into a new set of equations involving transformed field variables instead of the regular field variables. By doing this, a transformed field variable at one periodic surface can be made equal to the transformed field variable at the parallel periodic surface for all time, so that the periodic boundary conditions can be easily implemented.

To be more specific, we introduce the transformed field variable $\mathbf{P}$, which is related to the electric field vector by

$$\mathbf{E}(x, y, z; \omega) = \mathbf{P}(x, y, z; \omega) e^{-j(k_x^s x + k_y^s y)} \tag{31.96}$$

This is done to extract out the phase variation associated with the periodic boundary conditions. By substituting Eq. (31.96) into Eq. (31.91), we obtain the periodic boundary condition for $\mathbf{P}$ as

$$\mathbf{P}(x, y, z; \omega) = \begin{cases} \mathbf{P}(x + T_x, y, z; \omega) \\ \mathbf{P}(x, y + T_y, z; \omega) \end{cases} \tag{31.97}$$

Note that Eq. (31.97) remains the same in the time domain, which makes the periodic boundary conditions straightforward to enforce because there is no time delay between the transformed fields at two parallel surfaces of the unit cell. By substituting the transformed field variable defined in Eq. (31.96) into the wave equation (31.8), and transforming the resulting equation into the time domain, we obtain the following modified wave equation (neglecting the source terms for simplicity):

$$\begin{aligned} &\nabla \times \left(\frac{1}{\mu_r} \nabla \times \mathbf{P} \right) + \frac{1}{c^2} \left(\varepsilon_r - \frac{\sin^2 \theta_s}{\mu_r} \right) \frac{\partial^2 \mathbf{P}}{\partial t^2} + \mu_0 \sigma \frac{\partial \mathbf{P}}{\partial t} - \frac{1}{c} \nabla \times \left(\frac{\mathbf{k}_t^s}{\mu_r} \times \frac{\partial \mathbf{P}}{\partial t} \right) \\ &\quad - \frac{1}{c} \frac{\mathbf{k}_t^s}{\mu_r} \times \nabla \times \frac{\partial \mathbf{P}}{\partial t} + \frac{1}{c^2} \mathbf{k}_t^s \left(\frac{\mathbf{k}_t^s}{\mu_r} \cdot \frac{\partial^2 \mathbf{P}}{\partial t^2} \right) = 0 \end{aligned} \tag{31.98}$$

where the transverse component, $\mathbf{k}_t^s$, of the unit vector of the scanning vector wavenumber is given by

$$\mathbf{k}_t^s = \sin \theta_s \cos \phi_s \hat{\mathbf{x}} + \sin \theta_s \sin \phi_s \hat{\mathbf{y}} \tag{31.99}$$

Equation (31.98) is to be solved over a unit cell with the associated truncation boundaries shown in Figure 31.7. With the coordinate system shown for the unit cell in Figure 31.7b,

we have the following periodic boundary conditions:

$$\begin{aligned}\mathbf{P}(x=0, y, z; t) &= \mathbf{P}(x=T_x, y, z; t)\\ \mathbf{P}(x, y=0, z; t) &= \mathbf{P}(x, y=T_y, z; t)\end{aligned} \tag{31.100}$$

and for simplicity we implement a first-order ABC,

$$\hat{\mathbf{z}} \times \left(\nabla \times \mathbf{P} - \frac{1}{c}\hat{\mathbf{k}}_t^s \times \frac{\partial \mathbf{P}}{\partial t}\right) = \frac{1}{c\cos\theta_s}\hat{\mathbf{k}}_t^s \left(\hat{\mathbf{k}}_t^s \cdot \frac{\partial \mathbf{P}}{\partial t}\right) - \frac{1}{c}\cos\theta_s \hat{\mathbf{z}} \times \left(\hat{\mathbf{z}} \times \frac{\partial \mathbf{P}}{\partial t}\right) \tag{31.101}$$

that is capable of perfectly absorbing the main beam, which corresponds to the fundamental Floquet mode. More sophisticated boundary conditions for absorbing higher-order Floquet modes can also be derived by taking the Laplace transform of Eq. (31.92). Its implementation, however, is much more involved [47].

It is important to briefly note the modeling of coaxial waveguide-fed phased array antennas, where the antenna elements reside on top of a perforated ground plane and are fed through the apertures in the ground plane by coaxial waveguides. For such a case, the proper port boundary condition (considering only the TEM mode) is

$$\hat{\mathbf{n}} \times \left(\frac{1}{\mu_r}\nabla \times \mathbf{P}\right)_{z=z_{\text{port}}} = -\frac{(\varepsilon_r/\mu_r)^{1/2}}{c}\hat{\mathbf{n}} \times \hat{\mathbf{n}} \times \left(\frac{\partial \mathbf{P}}{\partial t} - 2\frac{\partial \mathbf{P}^{\text{inc}}}{\partial t}\right)_{z=z_{\text{port}}} \tag{31.102}$$

where $\mathbf{P}^{\text{inc}}$ is the transformed incident field.

31.6.2 Finite Arrays

The analysis and design of a large finite phased array antenna, consisting of hundreds or even thousands of array elements, is technically very challenging because of its large electrical size. Although planar finite-sized arrays can be approximated by an infinite array (unit cell) model in conjunction with array theory [99], and arrays with one dimension much larger than the other can be approximated by a finite-by-infinite array model [100], such an analysis neglects the complicated edge truncation effects, which often are dealt with by expensive and time-consuming measurements. This is particularly true with regard to understanding the variation in the input impedance seen by the outer elements of the array as a function of their distance from the edge of the array. However, it is important to note that, for the radiation *patterns* associated with planar finite periodic arrays of arbitrary dimension, it is often the case that simple array theory based on a unit-cell analysis may be completely sufficient and requires negligible computational cost beyond the unit-cell simulation.

For applications that require a rigorous analysis of the finite array, a conventional FEM formulation can be applied; however, this often results in linear systems with many millions of unknowns, requiring huge amounts of memory and floating-point operations. In the attempts to fully model large finite periodic arrays more efficiently, various fast algorithms have been developed, including the array decomposition method [18], the characteristic basis function method [101], and the domain decomposition methods [19, 21]. Domain decomposition methods for both frequency- and time-domain models are discussed next.

31.6.2.1 Frequency-Domain Modeling Among the various numerical schemes proposed for large-scale FEM simulations, the domain decomposition method is probably the most efficient and versatile. Here we describe one such domain decomposition method that is especially well suited for simulating large, finite array antennas. This method is called a dual primal finite-element tearing and interconnecting method for electromagnetic analysis (FETI-DPEM) [21]. It is an extension of the FETI-DP method for solving the Helmholtz equation in computational acoustics [102–104]. The basic ideas of FETI and FETI-DP have also been adapted for solving electromagnetic problems [19, 70]. The general principle of the FETI-DPEM is first to divide the entire computational domain into nonoverlapping subdomains, where an incomplete solution of the field is first evaluated using a direct solver for the FEM equations (a *direct* solver, in contrast to an *iterative* solver, provides a complete factorization of the governing FEM system matrix in the subdomains). Next, tangential continuity is enforced at the subdomain interfaces by using the Lagrange multipliers. This yields a reduced-order interface problem, which can be solved using an iterative algorithm. The solution to the interface problem can then serve as the boundary condition for individual subdomain problems to evaluate the field inside the subdomains.

For an array-type geometry, the natural domain decomposition is to divide the entire domain such that each array element becomes a subdomain. For an array that consists of $M \times N$ array elements, such a decomposition yields $M \times N$ subdomains in addition to subdomains external to the array, as sketched in Figure 31.8. Assuming all the array elements are identical (which is not absolutely necessary), there are nine different subdomains in the entire computational domain; all other subdomains can then be identified with one of these nine basic subdomains (due to the enforcement of exterior boundary conditions such as the ABC and PML, the subdomains at the edges and corners are different from those in the interior). At the interface between two subdomains, say, subdomains

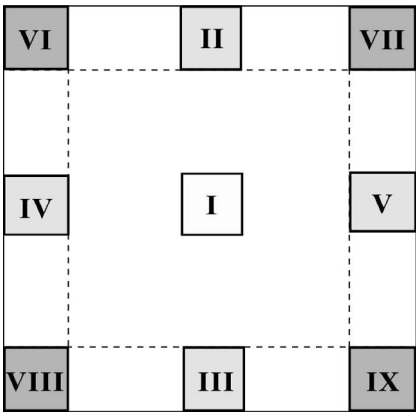


Figure 31.8 Nine different elements in a rectangular array. Due to the exterior boundary condition, the FEM equations for these nine elements are different; however, all other elements can be identified with one of these elements.

i and j, the electric and magnetic fields satisfy the tangential continuity conditions

$$\begin{aligned} \hat{\mathbf{n}}_i \times (\hat{\mathbf{n}}_i \times \mathbf{E}_i) &= \hat{\mathbf{n}}_j \times (\hat{\mathbf{n}}_j \times \mathbf{E}_j) \\ \hat{\mathbf{n}}_i \times \left(\frac{1}{\mu_{ri}} \nabla \times \mathbf{E}_i \right) &= -\hat{\mathbf{n}}_j \times \left(\frac{1}{\mu_{rj}} \nabla \times \mathbf{E}_j \right) = \mathbf{\Lambda}_{ij} \end{aligned} \tag{31.103}$$

where $\hat{\mathbf{n}}_i$ and $\hat{\mathbf{n}}_j$ are the unit normal vectors at the interface pointing to the exterior of subdomains i and j, respectively. Note that $\mathbf{\Lambda}_{ij}$ at the interface is actually unknown. However, by using the FEM, we can formulate the system of equations for each subdomain with both the electric field $\mathbf{E}$ and $\mathbf{\Lambda}$ as the unknowns. For identical subdomains, the FEM matrices will be identical. All the subdomain's systems of equations can then be assembled to form a large, global system of equations based on Eq. (31.103). However, we can eliminate all the unknowns associated with the electric field in the global system to obtain a much smaller system that contains only the unknowns associated with $\mathbf{\Lambda}$. Since $\mathbf{\Lambda}$ is defined only on the interfaces between the subdomains, this smaller system is called the *interface problem*. When formulated carefully, the elimination of the electric field unknowns in the global system requires only the inversion of the FEM matrix for each subdomain. Since this matrix is relatively small and sparse, it can be factorized very efficiently using a sparse direct solver as previously noted. Furthermore, since there are nine different subdomains, only nine such FEM matrices have to be factorized in general. This significantly speeds up the simulation and hence makes the solution of large finite arrays tractable.

The much smaller interface problem is solved using an iterative algorithm such as GMRES with the aid of a preconditioner. After the interface problem is solved, the electric field inside each subdomain can be evaluated individually by using the known Neumann boundary condition [the second equation in Eq. (31.103)] at the interface for each subdomain. It has been shown that the number of iterations for solving the interface problem is independent of the number of subdomains, which means that an n-times larger problem can be solved in parallel in a constant time by using n-times the number of processors. This makes the FETI-DPEM especially attractive for parallel computation. Moreover, because of its efficient reuse of the subdomain FEM matrices, the method can efficiently handle large array problems even on a single computer.

31.6.2.2 Time-Domain Modeling Developing an efficient time-domain formulation for a large finite array is not as straightforward as the analogous frequency-domain formulation described previously. Whereas the system matrix is solved only once in a frequency-domain simulation, it is solved at each time step in a time-domain simulation. Hence the matrix solving is more critical in a time-domain simulation and a tremendous amount of research effort has been directed into this aspect. In order to minimize the total CPU time, solving the linear system using a direct solver is preferable in the time domain since the factorization can be reused at each time step. For large finite arrays, however, the memory storage required by a direct solver becomes impractically large, even for the factorization of a sparse matrix resulting from a global FEM discretization. This is because the inverse of a sparse matrix is not generally sparse. As a result, the time-domain FEM often resorts to an iterative solver for large problems to avoid excessive memory usage. However, the convergence of an iterative solver heavily depends on the properties of the system matrix and the preconditioner it employs.

One strategy to solve large-scale problems more efficiently in the time domain is again to employ the domain decomposition method by dividing the original global computation domain into several smaller subdomains. With a reduced size, each smaller subdomain problem can then be factored and solved using a sparse direct solver such that the overall computational complexity can be reduced as compared to the original single-domain (global) problem. To further reduce the computation time, the subdomain problems can be distributed onto a massively parallel computing system and solved in parallel. To this end, we can use the FETI-DPEM described earlier. A drawback of this method is that the global interface problem, which is usually solved by an iterative solver, needs to be solved at each time step. A more efficient time-domain domain decomposition scheme that does not require solving a global problem has recently been proposed [44, 45]. This method, referred to as the dual-field domain decomposition (DFDD) time-domain FEM, solves the dual-field second-order vector wave equations in each subdomain and relates the adjacent subdomains explicitly using the equivalent surface currents on the subdomain interfaces. Since adjacent subdomains are explicitly related at each time step, a global interface problem does not need to be formulated and solved. At every time step, each subdomain problem is solved efficiently using its local prefactorized matrix.

The DFDD method can be further accelerated for treating large finite array problems by exploiting the geometrical repetition in the array geometry. As discussed earlier, when the computational domain is divided such that each array element becomes a subdomain, there are nine different subdomains. For each of the nine basic subdomains, an FEM matrix is assembled, factored, and stored in memory at the beginning of the simulation. During the time marching, each of the array elements is updated using the appropriate factorized matrix according to its position in the array. Thus the DFDD requires the storage of only nine factorized matrices regardless of the actual size of the array. Considering that matrix factorization typically dominates the memory consumption, a significant reduction in the memory requirement can be achieved by using this method for large arrays. The same argument can be made for the factorization time. However, different from the frequency-domain FEM, the total amount of computation time required by the time-marching process unfortunately is not improved. The reason for this is that even though the system matrices are the same for the subdomains with identical geometry, the electric fields throughout these subdomains are different at every time iteration.

31.7 NUMERICAL AND PRACTICAL CONSIDERATIONS

Over the past several decades, many electromagnetic simulation techniques have been developed. In this section, we discuss various numerical and practical considerations regarding the choice of simulation techniques, the applicability of frequency- or time-domain formulations, the benefit of using parallel computers, and the issue of numerical convergence.

31.7.1 Choice of Solver Technology

Computational electromagnetics tools support a wide variety of engineering applications, such as analytically designing antennas and predicting the impact of platforms on antenna performance. In addition, they address more complex applications including calibration of antenna systems supported by analytical predictions, estimating cosite interference

of multiple antenna systems on a platform, and scattering from low observable antenna installations. Frequency-domain FEM and time-domain FEM span this class of applications, although generally for a subset of the frequency bands of interest to antenna engineers. The antenna engineer's expectation of analysis tools often differs from one application to another. For instance, when predicting impedance and radiation patterns of an existing antenna design, physical insight may not be needed; whereas if a novel class of antenna is being developed, physical insight is of great importance.

Every validated analysis approach has a place in the antenna engineer's suite of tools. A wide variety of analysis tools are available to antenna engineers including integral and differential equation solvers implemented in both frequency and time domains, as well as high frequency solvers that are useful when the physical size of the geometry is electrically very large. The remainder of this section is intended to support antenna design by identifying specifically when frequency- and time-domain finite-element formulations are an appropriate choice of analytical tool. Due to the many competing considerations, it is difficult to provide hard-and-fast rules; however, general guidance is provided.

The attraction of FEM solvers is their generality. With a good mesh generator, the FEM can model arbitrarily complex geometries that have very general material parameters. As an example, a wideband sinuous antenna, etched on a thick conical dielectric shell with coaxial feed lines, would be a good candidate for the time-domain FEM, or possibly the frequency-domain FEM, while challenging many other solution methods. The reasons for this are the presence of the material regions along with a very wide range of element sizes, as well as the absence of a known Green's function that describes the conical substrate. The price one pays for the generality with finite-element approaches is the increased time required to build volumetric meshes as compared to the surface meshes generally used by moment-method integral equation formulations.

When selecting an analysis tool, it is important to understand both the capabilities and limitations of the techniques and codes that are available. Capabilities and limitations may depend on either a code's specific implementation or on the fundamental underlying theory. For instance, although theory exists to allow thin wires, thin slots, and thin materials to be accurately included in finite-element analysis tools without directly meshing these geometrical details, a particular computer code may not have implemented that theory. Moreover, although it may, in principle, be possible to mesh a particular antenna geometry, the specific mesh generator integrated into a software package may be challenged by an especially complex geometry. In practice, limitations are often experienced as excessively long setup and/or run times.

Although any of a variety of analysis tools may be able to provide accurate solutions, it is important to choose the tool that provides the necessary accuracy within reasonable computer execution times. For instance, although predicting coupling between two monopoles on a large structure may be possible using either frequency- or time-domain FEM, run times for high frequency methods may be orders of magnitude faster and should be used when acceptable accuracy can be demonstrated. However, it is important to keep in mind that high frequency methods may not provide sufficient accuracy when one antenna is deep in the shadow region, and in such a case, finite-element (or integral equation) methods might be required to predict coupling to antennas in regions of the structure where there is no line-of-sight.

Finally, it is important to realize that the appropriate choice of tool, for a specific problem, slowly changes as new capabilities are formulated within the various classes of solvers. For instance, wide angle scanning of phased arrays to nearly 90° has recently

become possible in the time-domain FEM, making it an excellent choice for the analysis of wideband phased arrays. Another example is the acceleration of integral equation solvers using the FMM. This acceleration now makes integral equation solvers a practical choice for applications that are moderately electrically large. In addition, research currently being performed to add general materials to the FMM will likely further increase the class of problems to which integral equation solvers will apply in the future. Finally, advancements in domain decomposition techniques for the FEM now permit these methods to efficiently address electrically large applications such as large finite arrays.

31.7.2 Frequency-Domain Versus Time-Domain Simulations

It is important to begin a simulation with the desired goal in mind. Specifically, what information is needed from the predictions? Does one need simply to predict only impedance and radiation patterns of an existing antenna? Or is the goal of the analysis to explain why a particular antenna has specific measured characteristics? For instance, what is causing the measured cross polarization to be so high?

If the goal is to predict measurable performance parameters of an existing antenna accurately and efficiently, a time-domain technique is a good choice when many frequencies are needed. Alternatively, a frequency-domain method that utilizes a direct solver is a good choice when many excitations are needed, such as in the case when many different plane waves are incident from numerous angles.

When it is necessary to fully understand the underlying physics, both the frequency- and time-domain responses of the antenna provide useful and different insights. A frequency-domain simulation directly provides visualizations of how field lines and Poynting vectors in the near field spatially vary at a given frequency. For instance, a visualization of strong field lines coupling the vertical polarization port of a printed antenna to the horizontal polarization port might guide design improvements in cross-polarization performance of that antenna. Alternatively, time-domain predictions provide additional insight into where physical phenomena occur within an antenna model, in the same way that time-domain reflectometry measurements provide knowledge of the locations from where signals are reflected by discontinuities along a transmission line. For instance, edge reflections from finite arrays can be clearly observed, as reflected waves move across an array and enter individual radiator feeds. Distances to reflections and other physical phenomena that cause the impedance of wideband antennas to vary with frequency are easily determined. Knowledge of those distances guides the antenna designer by providing valuable insight into the locations of reflections, as well as the coupling to adjacent sources, where both affect wideband impedance.

A good example of solution applicability is for the case of electrically small antennas. Frequency-domain FEM is well suited to analyzing electrically small antennas and does not suffer numerical challenges due to the small electrical size until perhaps the dc limit is approached. On the other hand, time-domain approaches may not be well suited to electrically small antennas because these structures often require exceedingly long run times to reach a steady state.

As previously noted, wideband phased array analysis is an excellent application of time-domain FEM because of its efficiency for this type of application. Wideband predictions are obtained with a single run, often in less time than is required to sample the frequency band using a frequency-domain FEM. In addition, due to its unconditional stability, the time-domain FEM provides reasonable run times even at wide scan angles.

Because phased arrays are designed to eliminate grating lobes, a well-designed infinite phased array radiates a single plane wave within its intended band of operation. Improved ABCs for both the time-domain FEM and frequency-domain FEM, which are specialized to absorb energy nearly perfectly when a plane wave is incident from a known scan angle up to nearly 90°, have recently been formulated to significantly improve the prediction of scan impedance particularly over wide bandwidths. Furthermore, the fine frequency resolution available in a time-domain simulation ensures that sharp resonant anomalies, sometimes found in flared notch arrays and log-periodic antennas, are captured.

In both frequency- and time-domain formulations, various interpolation and extrapolation strategies can be useful. For example, frequency interpolation has been implemented in some frequency-domain FEM solvers to increase efficiency when many frequencies are of interest [39, 40]. Similarly, a postprocessing extrapolation of the time-domain response into the late time may be possible [105]. This extrapolation improves the efficiency of time-domain methods for highly resonant antennas, which would otherwise require long run times to obtain accurate antenna performance characteristics. While such methods extend the usefulness of a particular code, added care is required to ensure that these techniques are applicable and properly used. In addition, visualizations of all field quantities may not be available for the interpolated frequencies or time steps. However, by adopting various interpolation and extrapolation strategies, it is often possible to extend the utility of a given method so that it can address a design problem that is slightly outside the primary strengths of the method.

31.7.3 Domain Decomposition and Parallel Computing

As discussed in Section 31.6, a highly effective approach to solving large-size antenna and array problems is to divide the original computation domain into several smaller subdomains. With a reduced size, the subdomain problems can then be distributed onto a massively parallel computing system and solved in parallel. Therefore a robust and efficient domain decomposition method is of great interest for the FEM simulation of large-scale electromagnetic problems.

For an efficient solution on a parallel computer that provides good scaling with an increasing number of processors, it is important to have equally distributed work loads. As previously discussed, because of the repetitive nature of the large finite array problem, this type of application can easily be decomposed such that the computational work is nearly ideally balanced across a very large number of processors.

For more general three-dimensional applications, there are at least two fundamental challenges associated with very large-scale FEM simulations. First, the direct creation of unstructured grids that consist of tens or even hundreds of millions of elements can be challenging. Grids of a large magnitude are often constructed in parallel and then "stitched" together to form the global mesh. Because grids of realistic geometry will typically have elements with widely varying dimensions, this leads to the second challenge of logically decomposing this global mesh such that the computational work levels are uniformly balanced across the distributed computing platform. A variety of techniques are currently available to accomplish this decomposition [106].

31.7.4 Numerical Convergence

It is important to be careful to use a given numerical method only within its range of applicability. When a numerical method is applied outside of its bounds, one often obtains

poor results, and then concludes the numerical simulation is useless because the results lack sufficient accuracy. In light of the enormous amount of physical insight that can be gained from modern numerical methods, and the range of the design space that can be explored only by simulation, this is an unfortunate conclusion. Therefore, it is very important to use a numerical method within its range of applicability. One of the most probable causes for obtaining poor results with the FEM is the use of a finite element mesh that is not sufficiently fine to resolve the field variation, which yields unconverged results. Such unconverged results are particularly troubling because they often appear to be physically plausible but are in fact inaccurate. The temptation to mesh a geometry with a coarser mesh than is necessary to generate a converged solution is especially high when a user is running large numbers of design iterations using an optimizer. For problems where the field variation is mostly due to wave propagation, a general rule of thumb for FEM is to use 20 points per wavelength to obtain a reasonably accurate finite-element solution. However, this rule has to be used with care because it does not guarantee any specific accuracy—different problems often have different accuracy in their FEM solutions. For instance, to obtain acceptable grid dispersion errors using FEM for problems requiring field propagation over many tens of wavelengths, greater than 20 points per wavelength may be necessary.

To understand this further, it is necessary to briefly address the issue related to the accuracy of a finite-element solution. In the FEM presented in this chapter, the field is spatially interpolated within each finite element using a set of basis functions. Since the basis functions have a finite interpolation order, such as the first-order basis in (31.14), this interpolation often cannot represent an arbitrary field exactly. As a result, there is always an interpolation error in a finite-element solution. For static problems, this is usually the major source of error. For wave related problems, the imperfect interpolation causes waves to propagate on a finite-element mesh at a slightly different speed than the true speed, which, in the frequency domain, leads to an error in the phase of the complex field. This error is often referred to as grid dispersion error, which is particularly detrimental to the solution of wave-related problems. This is because phase errors can accumulate as waves propagate over a finite-element mesh, and the larger the finite-element computational domain in terms of wavelength, the larger the final accumulated phase error, which can eventually render a completely inaccurate solution. As a result, even with a fixed mesh density, the accuracy of a finite-element solution will be different for different problems; typically, the accuracy decreases for a given mesh as the size of the computational domain (in terms of wavelength) increases. Theoretical, as well as numerical, investigations [9] show that for the first-order basis functions defined in (31.14), the grid dispersion error is proportional to n^{-2}, where n denotes the number of points per wavelength. As a result, one can always reduce the error in a finite-element solution by increasing the mesh density. For example, when the mesh density is doubled (say, from ten points per wavelength to 20 points per wavelength), the error will be reduced by a factor of four.

A much more effective approach to reducing the grid dispersion error is to employ higher-order basis functions in the finite-element formulation. It has been shown [9] that the grid dispersion error is proportional to n^{-2p}, where p denotes the order of the basis functions employed[†]. Therefore, for a mesh density of ten points per wavelength, if the

[†]Note that the order of vector basis functions here is defined differently from that in [9], which causes a different expression for the grid dispersion error. The basis functions defined in (2.14) are referred to as the first order here, instead of the zeroth order as in [9].

order of basis functions is increased from the first to second (which will yield a similar number of unknowns to a mesh with 20 points per wavelength), the dispersion error will be reduced by a factor of 100, and also results in much faster convergence. Of course, this can only be done with codes that have implemented higher-order basis functions.

Finally, we note that besides the field variation caused by wave propagation, geometry singularities such as edges, corners, and sharp tips and material discontinuities can also cause fields to vary rapidly. These issues should also be considered in the finite-element discretization by using finer meshes around these singularities and discontinuities. Of course, all numerical solution methods are subject to these considerations and are therefore not unique to finite-element formulations.

In summary, it is recommended that numerical simulations be run with either multiple different discretizations/mesh densities, or an increased order of the basis functions, to check for a change in physical quantities of interest so that a converged solution is ensured. Note that different physical quantities may require different mesh densities; for example, a low mesh density may be sufficient for calculating antenna radiation patterns, whereas a much higher mesh density, or a higher-order basis function, is typically required for near-field quantities such as the input impedance.

31.8 APPLICATION EXAMPLES

To demonstrate the capability and versatility of the FEM, we consider its application to the simulation of a variety of antennas, antenna arrays, and the interaction of the antennas with the mounting platforms.

31.8.1 Narrowband Antennas

Narrowband antennas are often based on resonant configurations. Because of resonance, the electromagnetic field in a narrowband antenna decays very slowly in time and varies drastically as a function of frequency. A small frequency step has to be employed in the frequency-domain simulation to capture the drastic frequency variation, and a very long time has to be simulated in the time domain for the fields to reach a steady state. Two typical narrowband antennas are considered here.

31.8.1.1 Patch Antenna This example provides the input impedance for a microstrip patch antenna geometry consisting of a 5.0-cm $\times$ 3.4-cm rectangular conducting patch residing on a dielectric substrate having thickness $d = 0.08770$ cm, relative permittivity $\varepsilon_r = 2.17$, and conductivity $\sigma = 0.362$ mS/m. The substrate is housed in a 7.5-cm $\times$ 5.1-cm rectangular cavity recessed in a ground plane (Figure 31.9). The patch is excited by a current probe applied at $x_f = 1.22$ cm and $y_f = 0.85$ cm. A 50-Ω impedance load is placed at $x_L = -2.2$ cm and $y_L = -1.5$ cm. Figure 31.10 shows the input impedance of the antenna as a function of frequency from 1 to 4 GHz. The results are calculated using the frequency-domain FE-BI and the time-domain FEM with a PML truncation. Both solutions agree with each other quite well. Although not shown, they also agree very well with the experimental data [9].

31.8.1.2 Coaxial-Fed Monopole Monopole antennas have a typical bandwidth of approximately 10% and are widely used in communication systems. A coaxial-fed

monopole antenna mounted on a small cylinder and an infinite ground plane is shown in Figure 31.11a. The height h of the antenna is 0.206756 m and the radius a is 1.6205 mm, which lead to the antenna thickness factor, $2 \ln (2h/a) = 11$, corresponding to a moderately thin monopole. Both the radius and height of the small mounting cylinder are 24.308 mm. All metal surfaces were assumed to be perfect electrical conductors, and the coaxial feed model was based on the techniques described in Section 31.5.1.3. Measurements [107] and predictions for the input admittance are shown in Figure 31.11b, where the predictions were based on the hybrid time-domain FEM and FDTD algorithm discussed in Section 31.4.4.

31.8.2 Broadband Antennas

Broadband antennas are often more complicated and their geometrical modeling more challenging than that for narrowband antennas. We consider four broadband antennas to demonstrate their FEM modeling and simulation.

31.8.2.1 Ridged Horn Ridged horns are broadband antennas that are widely used for experimental measurements. A 4-in. × 14-in. single-polarization design is shown in Figure 31.12a. The antenna is fed by a 50-Ω coaxial transmission line. Predictions and measurements for S_{11} versus frequency are shown in Figure 31.12b. The hybrid time-domain FEM with FDTD algorithm described in Section 31.4.4 was used to generate the predicted performance. The antenna and coaxial feed were modeled with the time-domain FEM algorithm, while the region external to the antenna used the FDTD algorithm with a PML termination. The coaxial feed model was based on the techniques described in Section 31.5.1.3. The global time step was based on the cell size in the FDTD region, and even though the maximum to minimum element size in the finite-element region was approximately 100:1, no further reduction in the time step was required in the finite-element region due to its local unconditional stability.

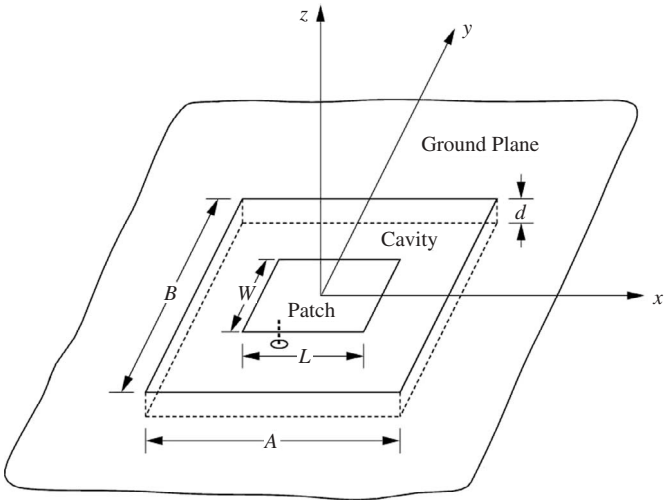
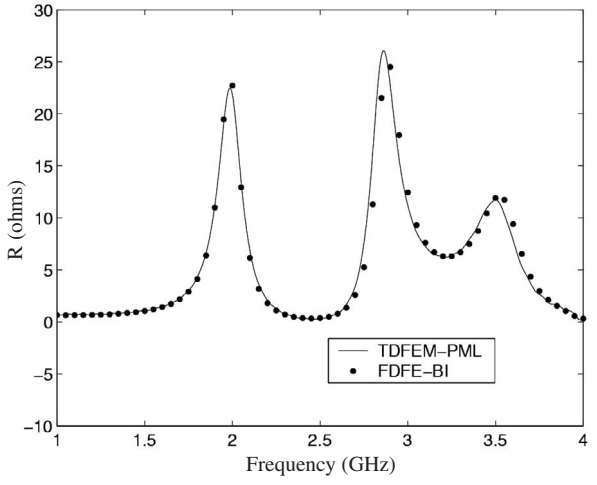
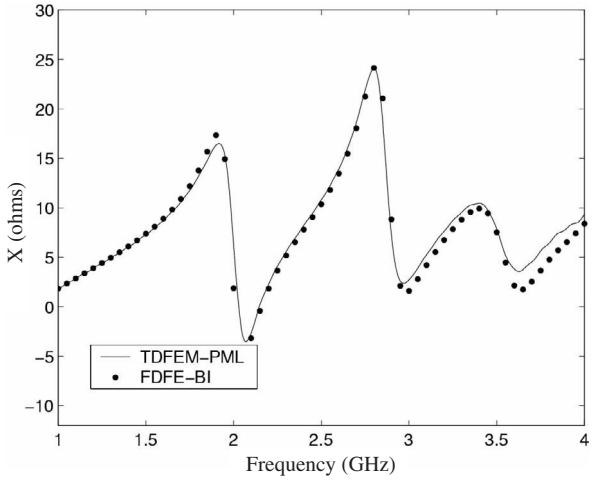


Figure 31.9 Microstrip patch antenna recessed in a ground plane. The antenna is fed by a coaxial line, which is modeled as an electric current probe.



(a)



(b)

Figure 31.10 Input impedance of a loaded microstrip patch antenna: (a) resistance and (b) reactance. (Adapted from Ref. 41.)

31.8.2.2 Spiral In this example, a logarithmic spiral antenna consisting of two free-standing conducting arms is simulated. The front view of the antenna is shown in Figure 31.13a. The centerline of the arms is prescribed by $r = r_0\ \tau^{\phi/2\pi}$ in the polar coordinate system, where r_0 is a constant and the scaling factor τ is chosen to be 1.588. The inner and outer radii of the spiral are 0.22 and 3.5 cm, respectively. In the simulation, the antenna is fed by the probe feed model described in Section 31.5.1. A detailed view of the feed region is shown in Figure 31.13b. A 0.56-mm long current probe is placed at the center to excite the antenna, and the voltage across the feed is measured along the two observation probes 0.28 mm away from the current probe. The calculated input resistance (solid line) and reactance (dashed line) versus frequency are shown in Figure 31.13c. The curves exhibit a transition from the resonant region to the broadband

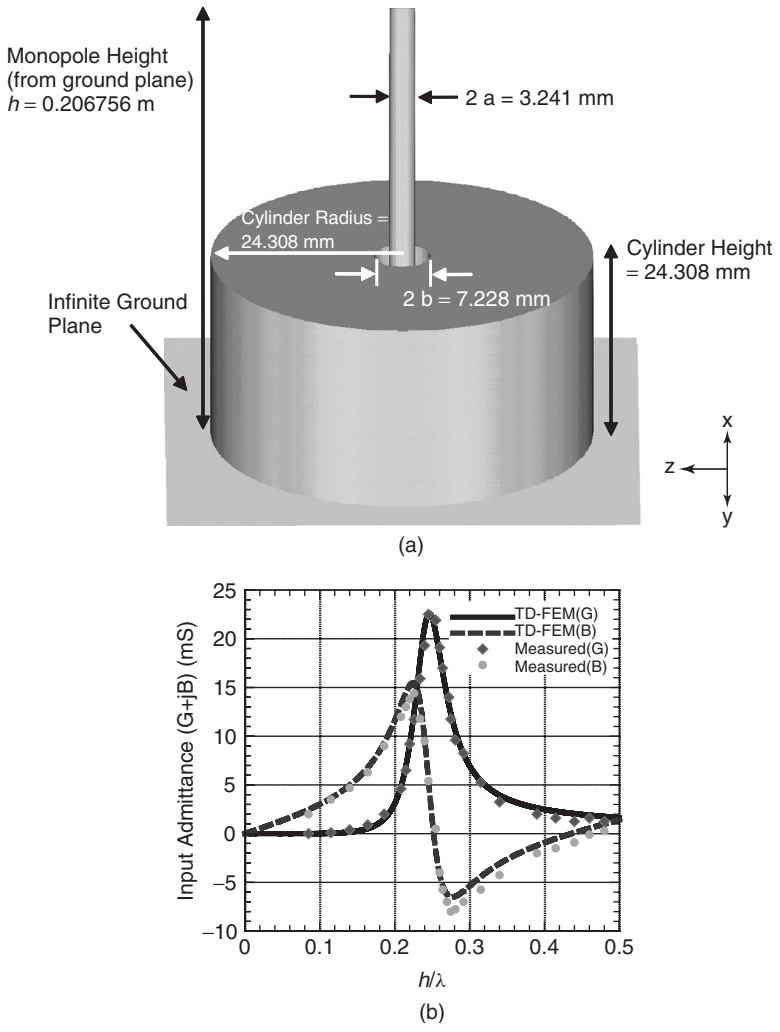


Figure 31.11 Coaxially fed monopole antenna: (a) geometry and (b) input admittance. Measured data is by Liu and Grimes [107].

region as the frequency increases and a nice broadband behavior for frequencies beyond 8 GHz. We also observe that the calculated input impedance converges to the theoretical value (188.5 Ω, arm-to-arm, for the two-arm self-complementary antenna) [108] at the higher frequency end.

31.8.2.3 Sinuous Because the sinuous antenna has a very wide range in length scales from the outer boundary of the antenna to the feed region, it presents significant challenges to any predictive method. The variation of the edge lengths required to mesh this type of antenna can be greater than 100:1 depending on the spatial extent of the feed region (which controls the high frequency performance of the antenna). Sinuous antennas are typically driven to obtain right- or left-hand circular polarization. Coaxial feeds are usually applied to each arm and are generally phased to obtain different performance

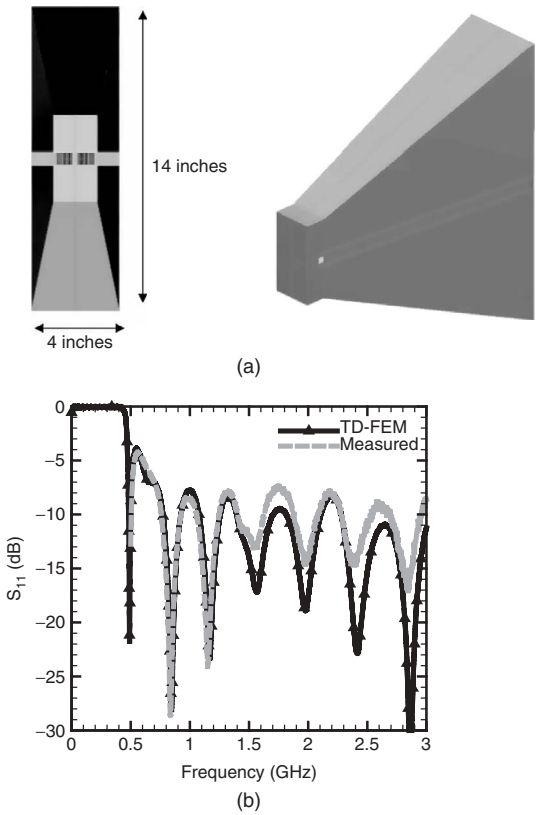


Figure 31.12 Ridged horn: (a) geometry and (b) S_{11} versus frequency.

characteristics. A four-arm design is shown in Figure 31.14a. The hybrid time-domain FEM and FDTD technique described in Section 31.4.4 was applied to this geometry. The use of a local, unconditionally stable solution method is particularly beneficial for this antenna because the time step does not need to be reduced for numerical stability as the element size becomes progressively smaller into the feed region.

The arm-to-common input impedance for the four-arm sinuous antenna over a bandwidth of 4–8 GHz is shown in Figure 31.14b. The arm-to-common input impedance for an N-arm ideal self-complementary antenna in free space is given by the expression $30\ \pi/\sin\ (\pi\ \ M/N)$, where $M = 1$ and $M = N - 1$ correspond to the "sum mode" and $M = 2$ and $M = N - 2$ correspond to the "difference mode." Note that the arm-to-common input impedance is one-half the opposing arm-to-arm impedance. The sum mode, arm-to-common, active input impedance of the ideal four-arm sinuous antenna in free space is 133.3 Ω. As shown in Figure 31.14b, the presence of the thin dielectric substrate significantly affects the input impedance compared to the free-space theory, which is also confirmed by measurements.

31.8.2.4 Inverted Conical Spiral Figure 31.15a shows a two-arm conical spiral antenna that has been inverted and fed against a ground plane to support wide bandwidth antenna performance. An inset of the finite-element mesh of the coaxial feed region is

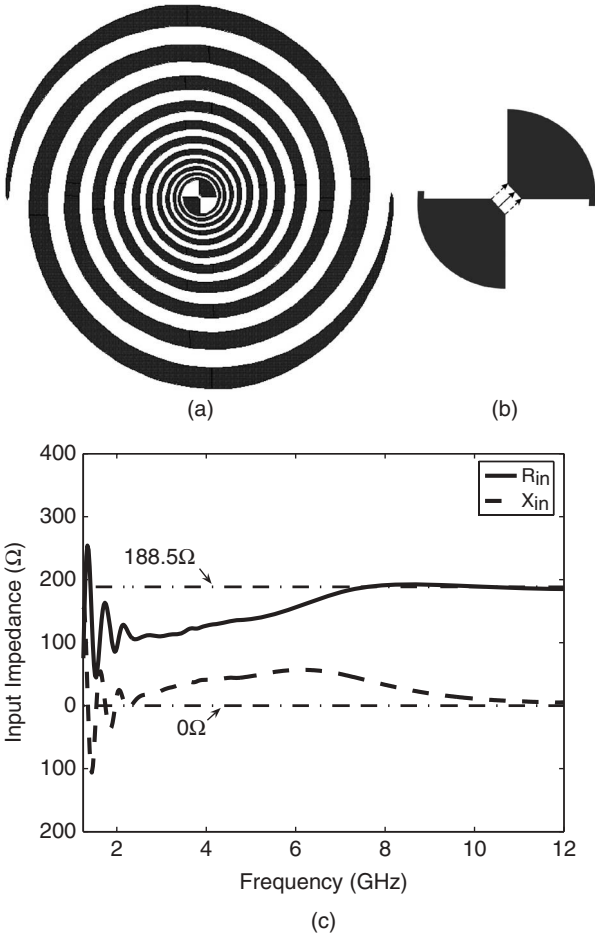


Figure 31.13 Two-arm logarithmic spiral antenna: (a) geometry of the arms, (b) enlarged feed region, and (c) input impedance. (Adapted from Ref. 43; copyright © 2005 IEEE.)

shown in Figure 31.15b. As frequency decreases, significant current density moves out along the arms, thereby increasing the distance of the peak current from the feed region and ground plane. Preserving this electrical distance above the ground plane is important in obtaining broadband performance of conical multiarm antennas in the presence of a ground plane.

While this early model is etched on the outer surface of a solid dielectric cone, it is also possible to use a thin material approximation inside the FEM to analyze the spiral etched on a thin dielectric shell. In practice, multiarm printed antennas such as spiral, sinuous, helical, or multiarmed log-periodic antennas can be etched on a thin dielectric shell having three-dimensional shapes that include not only cones but also truncated cones or hemispheres, in a similar way.

Figure 31.15c shows the input impedance of this inverted two-arm conical spiral for both a free-space and a solid dielectric cone of relative permittivity equal to 4.

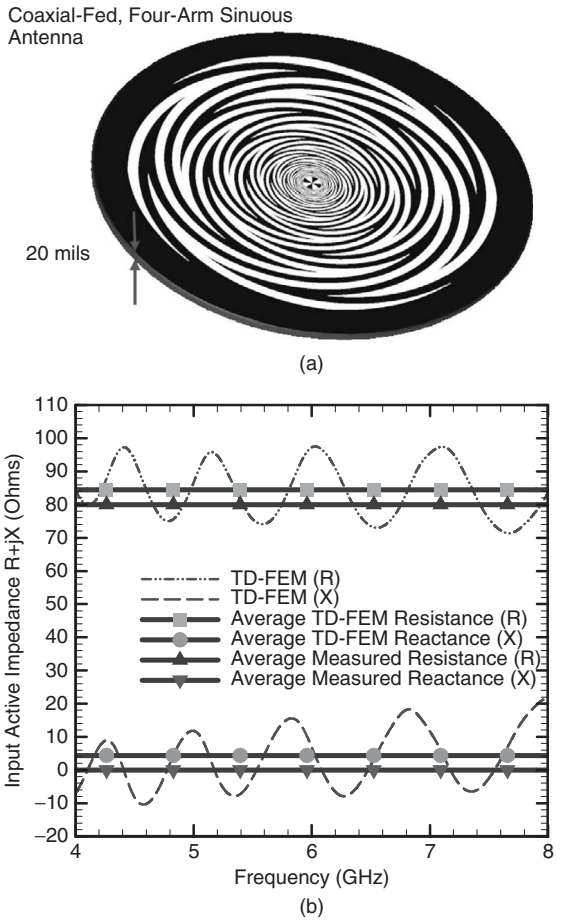


Figure 31.14 Four-arm sinuous antenna etched on a 20-mil dielectric substrate: (a) geometry and (b) predicted arm-to-common active input impedance and average measured impedance. (Adapted from Ref. 91; copyright © 2005 Artech House.)

31.8.3 Infinite Periodic Array

Three examples are presented to demonstrate the analysis of infinite periodic arrays in both frequency and time domains.

31.8.3.1 Frequency Domain Here we present the frequency-domain FEM simulation of an infinite array of microstrip patch antennas. The microstrip patches, shown in Figure 31.16, are placed on a grounded dielectric substrate with height h and permittivity ε_r. The patches are fed with coaxial cables that enter the substrate from below the ground plane. In the simulation, the computation domain, or unit cell, contains only a single patch and proper periodic boundary conditions are imposed on its boundaries. For a phased array, a critical performance parameter is the active reflection coefficient. One

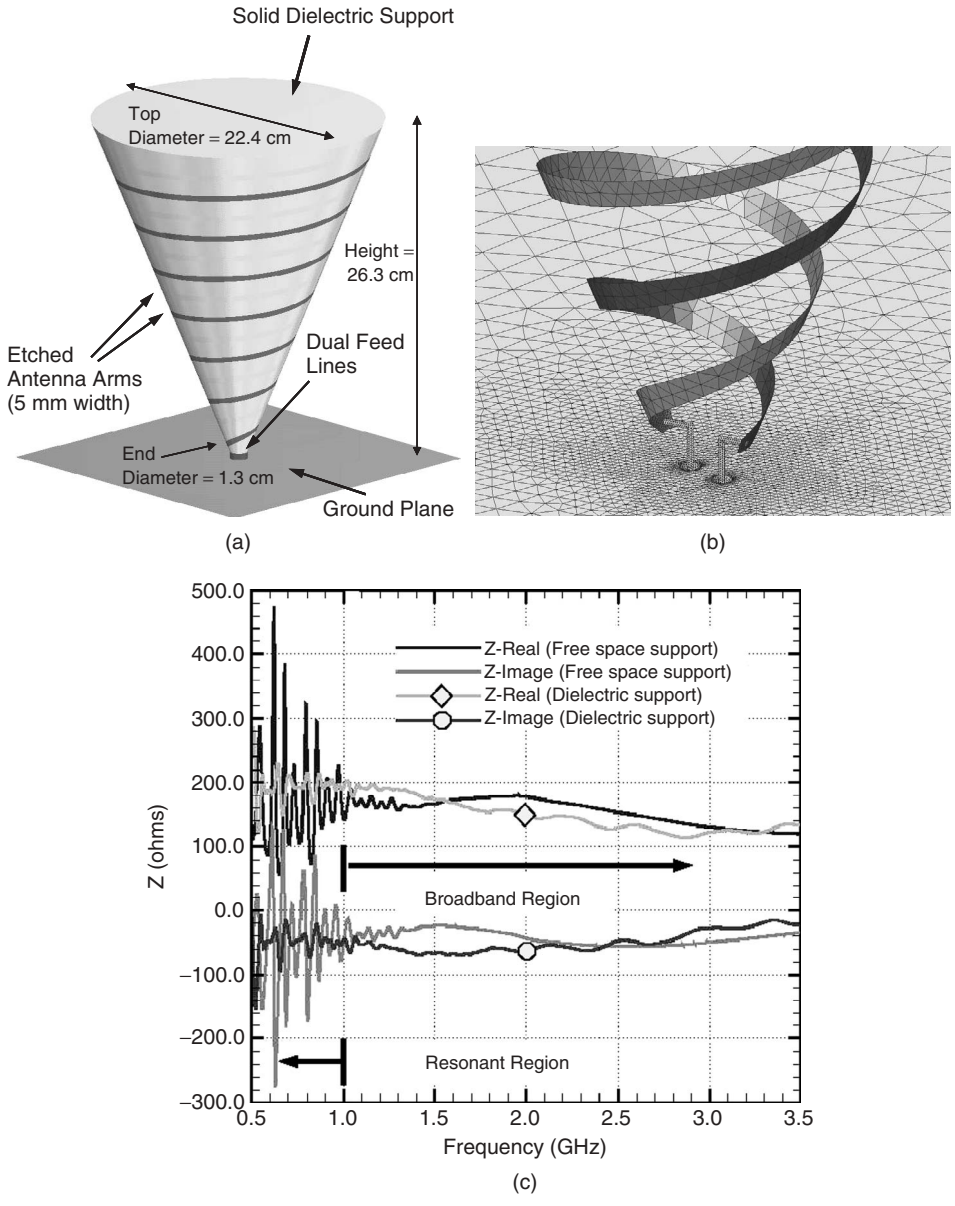


Figure 31.15 Inverted two-arm conical spiral over a ground plane: (a) geometry, (b) finite-element surface mesh local to the inverted conical spiral feed region, and (c) input impedance for a free-space and a solid dielectric ($\varepsilon_r = 4$) cone.

definition for this is given by

$$R(\theta_s, \phi_s) = \frac{Z_{\text{in}}(\theta_s, \phi_s) - Z_{\text{in}}(0, 0)}{Z_{\text{in}}(\theta_s, \phi_s) + Z^*_{\text{in}}(0, 0)} \tag{31.104}$$

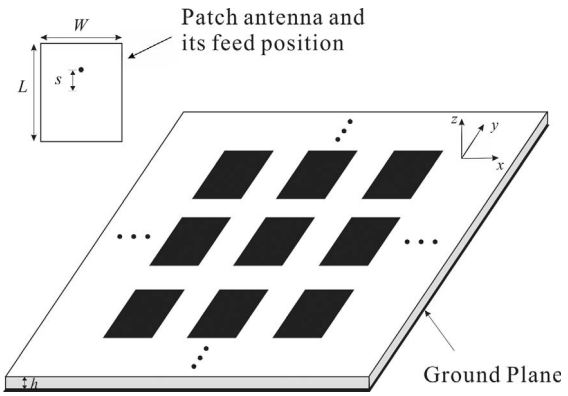


Figure 31.16 Infinite periodic microstrip patch array on a substrate. Each patch is fed by a coaxial line.

where Z_{in} is the input impedance of the antenna. The use of the conjugate operator in Eq. (31.104) discounts the imaginary part of the input impedance and more clearly reveals the scan properties of the array. The array under consideration has a period of $0.5\lambda_0$ in both x- and y-directions. The patches are $0.3\lambda_0$ by $0.3\lambda_0$ in size and the permittivity of the substrate is 2.55. The E-plane active reflection coefficients are relative calculated for two different substrate heights: $h = 0.02\lambda_0$ and $h = 0.06\lambda_0$. The results calculated using the simplified and precise coaxial feed models are plotted in Figure 31.17. A MoM analysis by Pozar and Schaubert [109] is used here as reference results. It is noted that all three results agree with one another very well. All results predict a scan blindness at $\theta = 68.8^\circ$ for $h = 0.06\lambda_0$, as a result of surface wave excitation. Moreover, for this particular array, the agreement between the simplified and precise feed models suggests that scan performances such as active reflection coefficients are insensitive to the specific feed scheme, as indicated in Ref. 109.

31.8.3.2 Time Domain Consider a phased array antenna based on a Vivaldi element shown in Figure 31.18, which is related to a design discussed in Refs. 15 and 17. A dielectric substrate ($\varepsilon_r = 6$) with an etched PEC surface on one side is positioned upright on the ground plane such that it partially covers the aperture of a coaxial waveguide. The side of the substrate containing the etched PEC surface bisects the aperture of the coaxial waveguide, and the PEC surface is etched such that it forms a short coplanar waveguide starting at the ground plane with the dimension of the two slots being equal to the width of the aperture of the coaxial waveguide. One of the slots of the coplanar waveguide transitions into a flared slotline, and the other slot is terminated by an open load formed by an etched circle of radius 2.5 mm. The dimensions of the unit cell are $T_x = 36$ mm and $T_y = 34$ mm, and the height of the substrate is $d = 33.3$ mm with a thickness $h = 1.27$ mm. The coaxial waveguide is empty with the radii of the inner and outer conductors equal to 0.375 and 0.875 mm, respectively. The halfwidth of the flared slotline is given by $0.25e^{0.123z}$ mm, which gives a halfwidth of 15 mm at the open mouth. The etched circle forming the open load of the other slot intersects the slot at a distance 3.5 mm from the ground plane. Note that the xz-plane that coincides with the center of the flared slotline divides the unit cell into two equal sized parts. Figure 31.19 shows the simulated results for the magnitude of S_{11} as a function of frequency for

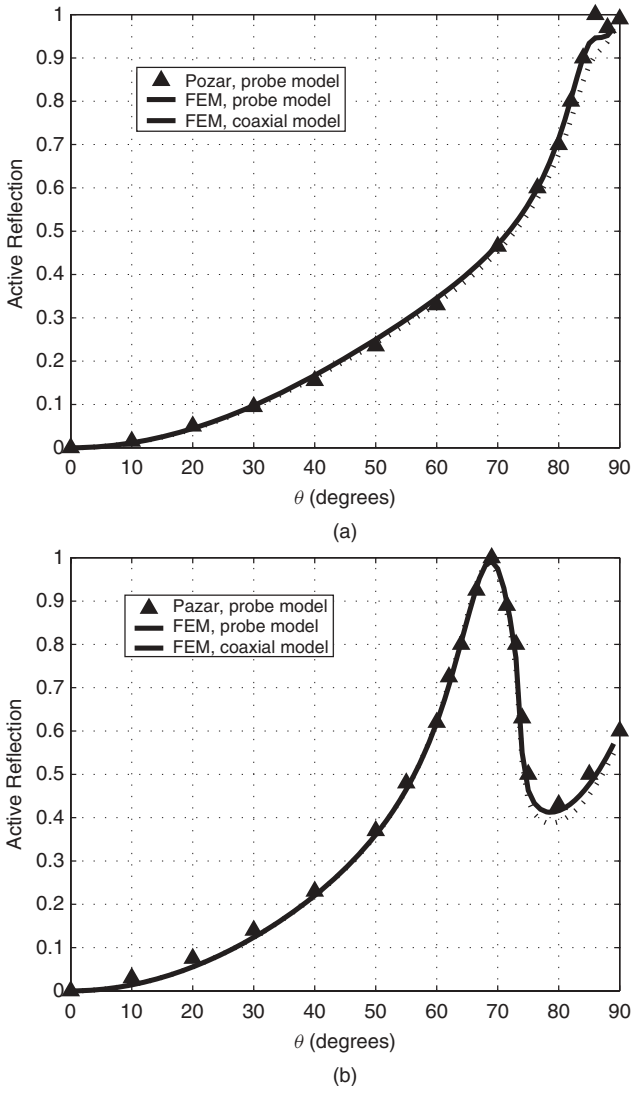


Figure 31.17 Active reflection coefficients for the microstrip patch array: (a) $h = 0.02\lambda_0$ and $s = 0.075\lambda_0$; and (b) $h = 0.06\lambda_0$ and $s = 0.14\lambda_0$. (Adapted from Ref. 17.)

broadside scanning $(\theta_s, \phi_s) = (0^\circ, 0^\circ)$ and oblique scanning $(\theta_s, \phi_s) = (45^\circ, 135^\circ)$. The results are computed using the time-domain FEM [47] and are compared with another solution obtained using the commercial software suite HFSS, which is based on the frequency-domain FEM.

Scanning of periodic arrays to a very wide scan angle is often problematic for time-domain methods such as FDTD because a significant reduction in the time step is required [82]. The periodic time-domain FEM method described in Section 31.6.1.2 imposes no such restriction on the time step due to its unconditional stability. This is demonstrated with the coaxially fed, doubly periodic monopole array

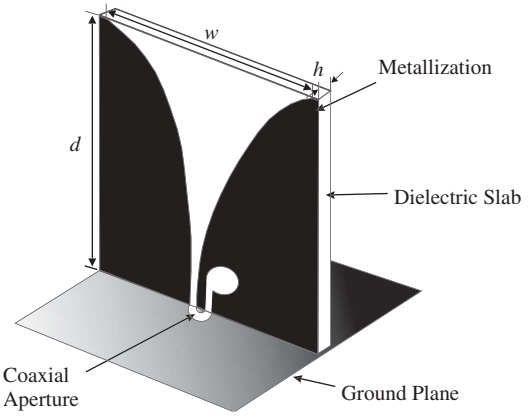


Figure 31.18 Unit cell of an infinite phased array of Vivaldi antennas.

shown in Figure 31.20a. Scan performance over $0^\circ \leq \theta_s < 90^\circ$ in both the principal (Figure 31.20b) and diagonal (Figure 31.20c) planes is shown. Comparison data are based on an approximate analytical solution that assumed a sinusoidal current distribution on the monopoles [110].

31.8.4 Finite Periodic Array

Two examples are presented to demonstrate the analysis of finite periodic arrays in both frequency and time domains.

31.8.4.1 Frequency Domain The example considered here is a broadband single-polarized Vivaldi antenna array having 31×31 array elements, which are identical to the one shown in Figure 31.18. Figure 31.21a shows the magnitude of the active reflection coefficient (defined by using the characteristic impedance of the feeding coaxial line as the reference impedance) as a function of frequency for the center element. The result, obtained using the technique (FETI-DPEM) described in Section 31.6.2.1, is compared with those for the corresponding infinite array from the time-domain FEM and HFSS [47]. Good agreement can be observed for the three different methods over a wide frequency band.

To demonstrate the edge effects of the finite array as compared with the corresponding infinite array, the active reflection coefficient for the elements from the central row of the array is computed at 3 GHz and the result is shown in Figure 31.21b with respect to the element position. As expected, it is evident that the infinite array approximation is accurate only for the inner elements, and the edge effects are prominent for the outer elements. Finally, the radiation patterns of the arrays are shown in Figure 31.22. For this simulation, nearly 40 million unknowns were used and the computation required 541 MB of memory and 12 minutes of computing time on a 1.5-GHz Itanium II processor using the FETI-DPEM method.

31.8.4.2 Time Domain As an example of time-domain FEM simulation of broadband phased array antennas, a 10×10 Vivaldi array shown in Figure 31.23a is analyzed here. The unit cell configuration is shown in Figure 31.23b. This configuration is similar

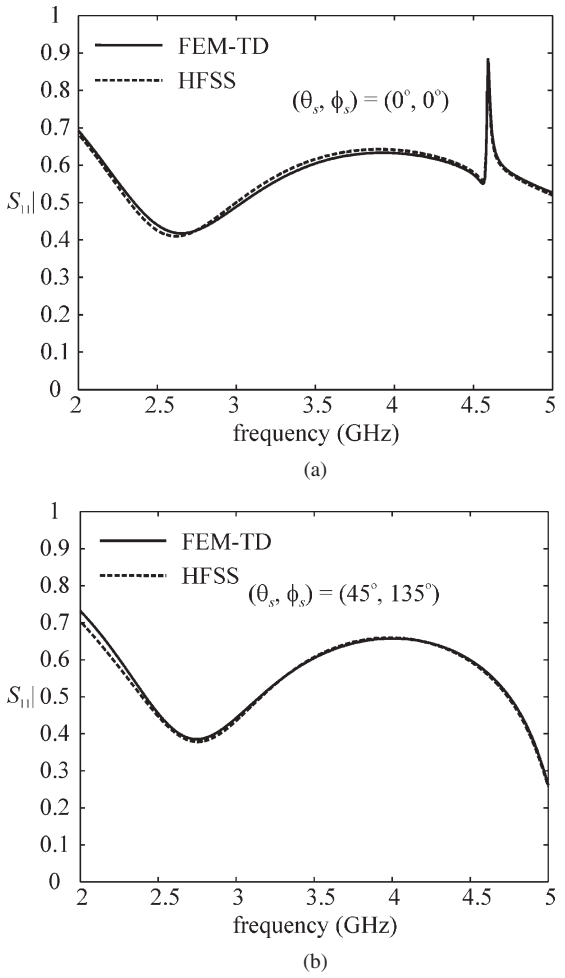
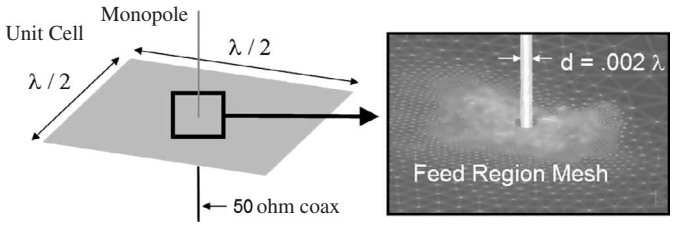
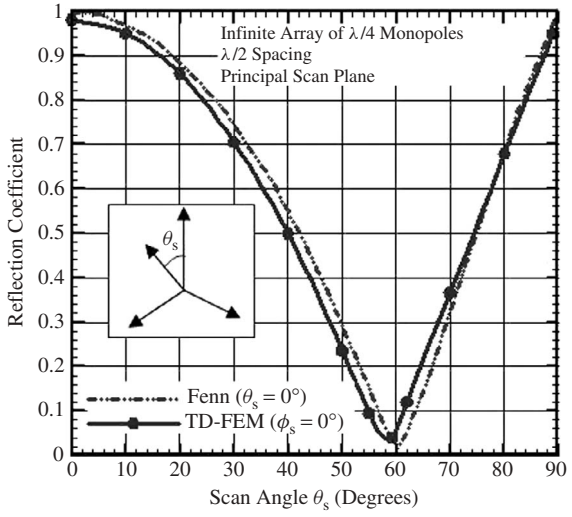


Figure 31.19 Reflection coefficient of a coaxial waveguide connected to an element in an infinite phased array of Vivaldi antennas. (Adapted from Ref. 47; copyright © 2006 IEEE.)

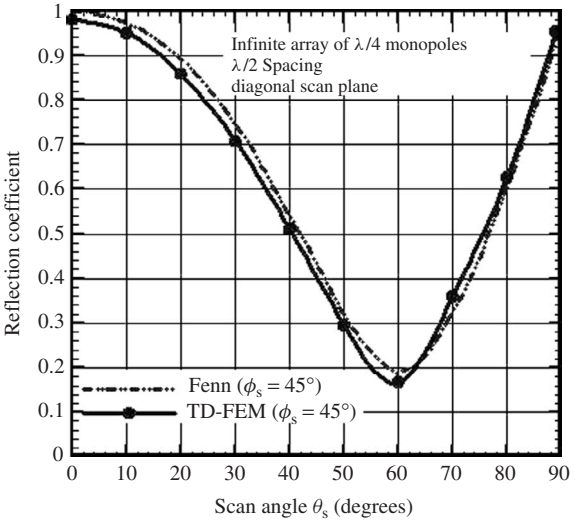
to the one considered in Sections 31.8.3.2 and 31.8.4.1, but now having different dimensions and the feed is a stripline from below the ground plane. The spacing between array elements is 40 mm in both the x- and y-directions. The separation corresponds to one-half of a wavelength at 3.75 GHz. An infinite ground plane is assumed and the stripline feed is modeled as a TEM port. The simulation employs 144 subdomains and the total number of unknowns is approximately 3.3 million. The VSWR parameter calculated at the feed port of the center element is shown in Figure 31.23c. A prominent feature of the VSWR curve is the two sharp peaks occurring at 3.92 and 5.32 GHz when the array is configured for broadside radiation. Such phenomena, referred to as the impedance anomalies, have been observed in the frequency analysis of the infinite array of Vivaldi antennas [17, 111]. The anomalies are explained by the excitation of certain resonant modes in the unit cell cavities, which are formed by the ground plane, the electric conducting surfaces on the sidewalls, and a magnetic conducting surface on the open aperture. The resonant



(a)



(b)



(c)

Figure 31.20 Quarter-wavelength monopole array: (a) unit cell geometry and the finite-element surface mesh local to the feed region, (b) scan performance in the principal ($\phi_s = 0^\circ$) plane, and (c) scan performance in the diagonal ($\phi_s = 45^\circ$) plane.

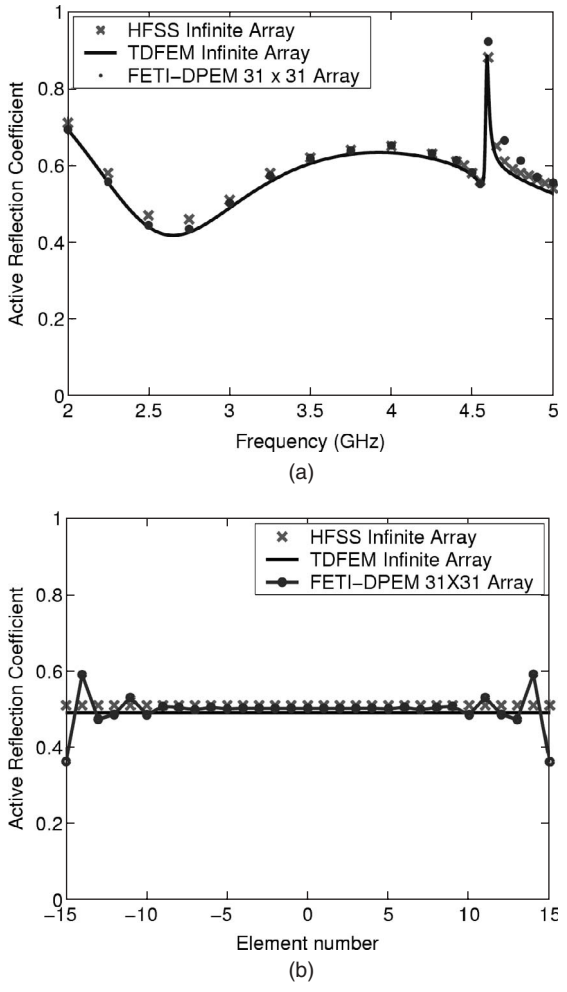


Figure 31.21 Active reflection coefficient for a 31×31 Vivaldi antenna array: (a) for the center element as a function of frequency and (b) for a midrow of elements at 3 GHz. (Adapted from Ref. 21; copyright © 2006 IEEE.)

frequencies predicted by the cavity model are $f_{10} = 3.99$ GHz and $f_{11} = 5.55$ GHz, which are similar to the resonant frequencies obtained by the numerical simulation. Note that the cavity model does not take into account the thin dielectric substrate. If the effect of the dielectrics is considered, the predicted resonant frequencies would become slightly lower.

31.8.5 Platform Interaction

In practice, antennas are generally integrated into platforms, such as airplanes, missiles, satellites, and ships. Mounting antennas on a platform inevitably introduces distortion in their radiation patterns and causes coupling between the antenna and the platform. The distortion in the radiation patterns may reduce the desired coverage for effective

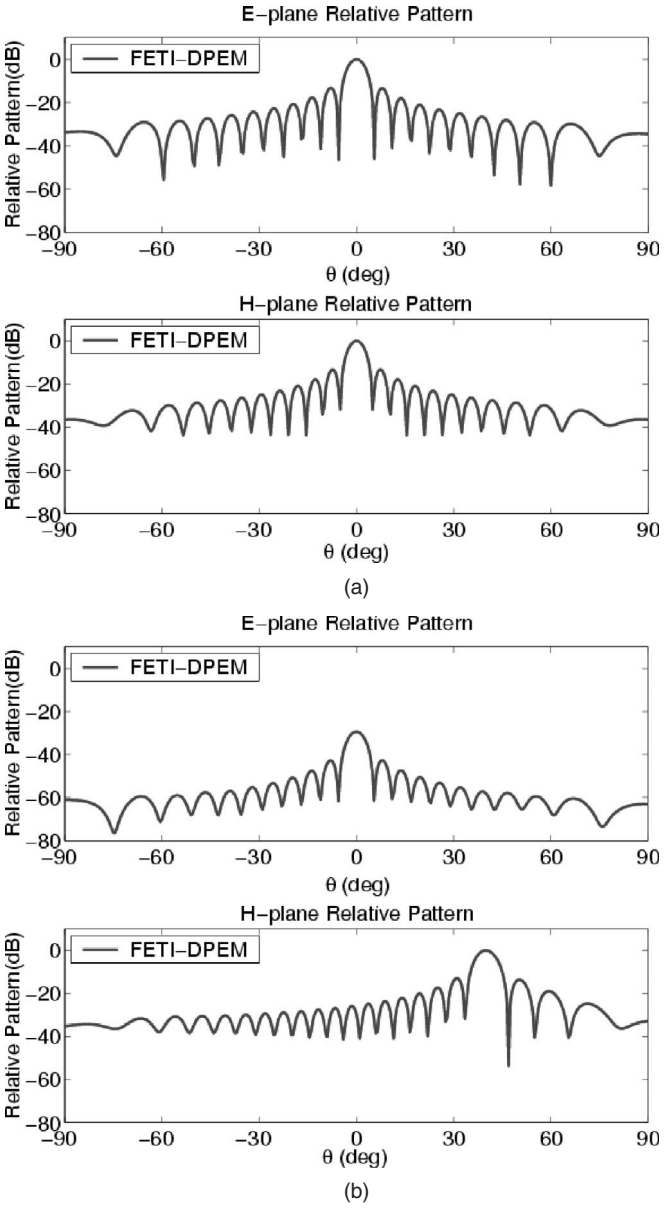


Figure 31.22 Radiation patterns of a 31×31 Vivaldi antenna array at 3 GHz: (a) broadside and (b) $\theta_s = 40^\circ$ and $\phi_s = 0^\circ$. (Adapted from Ref. 21; copyright © 2006 IEEE.)

communications or compromise the accuracy for isolating and locating targets. The existence of mutual coupling, caused by space waves, surface waves, and scattering by the platform, can increase the electromagnetic coupling between multiple onboard antennas and consequently make it difficult to operate the antennas simultaneously. In addition, antenna systems utilizing amplitude and phase received at various antenna ports often experience degraded performance. Analytical predictions of anticipated amplitude and

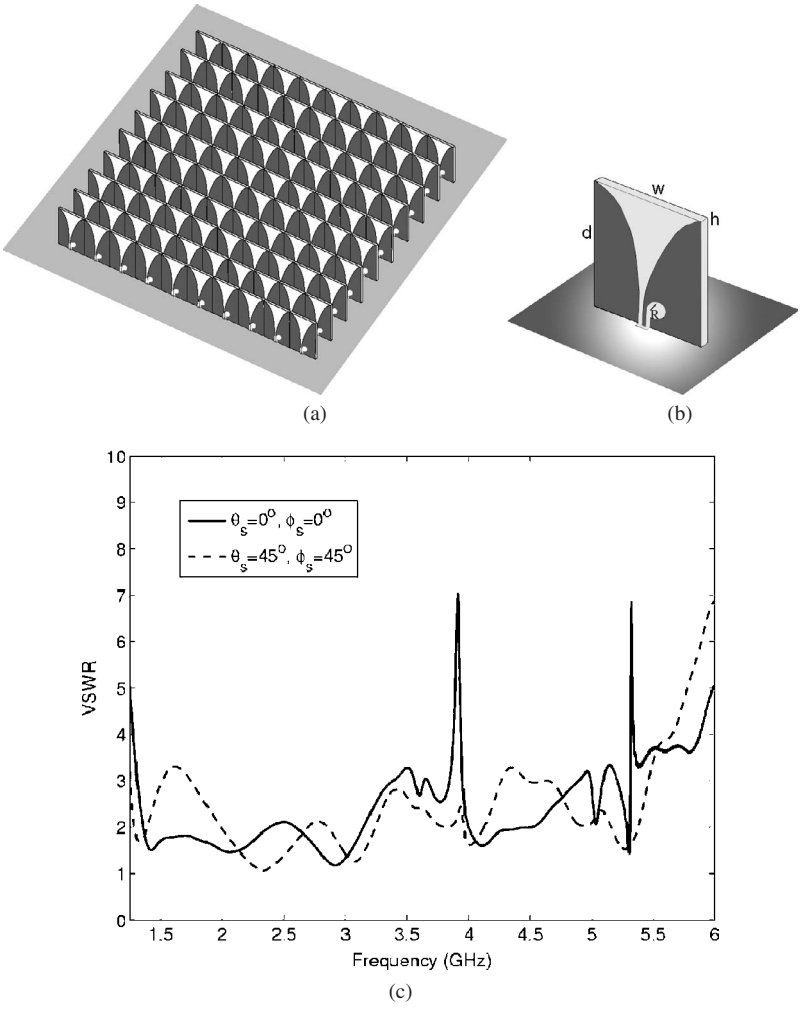


Figure 31.23 10×10 Vivaldi array: (a) geometry of the array and the unit cell and (b) VSWR calculated at the feed port of the center element. (Adapted from Ref. 45; copyright © 2007 Elsevier.)

phase received at the antenna ports, typically as a function of the incident angle of the received signal, can often be used to improve the performance of the antenna system. Therefore it is important to develop accurate numerical prediction tools to characterize the radiation patterns and mutual coupling of the antennas mounted on a complex, often large, platform.

Because most platforms are electrically very large, analyzing antennas mounted on a platform is technically very challenging. Various approximate numerical schemes have been developed for this analysis. The simplest is first to analyze the antenna by assuming that the antenna is either residing in free space or on a ground plane (depending on specific problems) to obtain the current distribution on the antenna or the radiated near field, and then compute the radiation pattern (far field) by letting the current radiate in the presence of the platform or by letting the near field propagate in the presence of the

platform [35, 36]. The second step can be carried out efficiently using an asymptotic technique such as physical optics or the geometrical theory of diffraction. This approach incorporates the effect of the platform on the radiation pattern but ignores the effect of the platform on the antenna itself. An improvement to this approach is to incorporate the first-order effect of the platform on the antenna by considering the reflected field by the platform as an incident field on the antenna. There are two ways to accomplish this. If the analysis employs the FE-BI method, this can be accomplished by modifying the Green's function in the boundary integral equation to include the presence of the platform [34]. Alternatively, we can first let the antenna radiate in free space and then compute the field reflected back to the antenna by the platform using an asymptotic technique. The analysis is then repeated by considering the reflected field as the excitation to the antenna. This process can be repeated until a converged solution is obtained. In practice, a single iteration is often sufficient unless the antenna is placed in a highly resonant platform.

Since asymptotic (ray optics) techniques are not capable of modeling surface waves and near-field scattering accurately on complex platforms, the previously described approach cannot accurately simulate the mutual coupling of the antennas when surface wave and near-field physics are important. In this situation, it is necessary to use a first principle method, such as the FEM or its various hybridizations, to carry out the analysis. In the hybrid FE-BI method, the FEM is used to model the details of the antennas and the boundary integral equation is employed to model the effect of the platform. The time-consuming evaluation of boundary integrals over the platform can be accelerated by using a fast algorithm such as the FMM. One such formulation is described by Liu and Jin [38], which included a numerical example to demonstrate its capability. The example consists of a microstrip patch antenna housed in a cavity that resides on a platform consisting of a conducting circular cylinder and a conducting plate (wing). The microstrip patch antenna is designed to operate at 3.3 GHz and its long edge is aligned with the cylinder's axis. The entire structure is shown in Figure 31.24a and the detailed information about the patch antenna is given in Figure 31.24b,c. The normalized radiation pattern in the H-plane is shown in Figure 31.25 for the case in which the patch antenna is placed 45° from the wing. It is seen that the numerical results agree well with the measurement for both copolarization and cross polarization. The effect of the platform is significant in both cases.

An additional example based on time-domain FEM hybridized with FDTD is shown in Figure 31.26. This geometry consists of two, surface conformal, dual-frequency circularly polarized stacked patch antennas mounted on a platform that is approximately 40 wavelengths in length and 20 wavelengths in circumference. The two antennas are located on opposite sides of the platform. A domain decomposition approach that isolated the patch antennas from the platform yet retained full coupling was used. The time-domain FEM was used to model the antennas and a thin volumetric layer around the perimeter of the platform. The FEM region transitioned to an FDTD grid that was terminated with a PML using the technique described in Section 31.4.4. Predictions and measurements for the right-hand circularly polarized (RHCP) gain of the antennas on the platform are shown in Figure 31.26b. The cut plane is through the antennas and along the axis of the platform. Good agreement is seen for all angles with the exception of the tail section. The positioner that was used to support and rotate the physically large platform adversely affected the measurements for observation angles in this region.

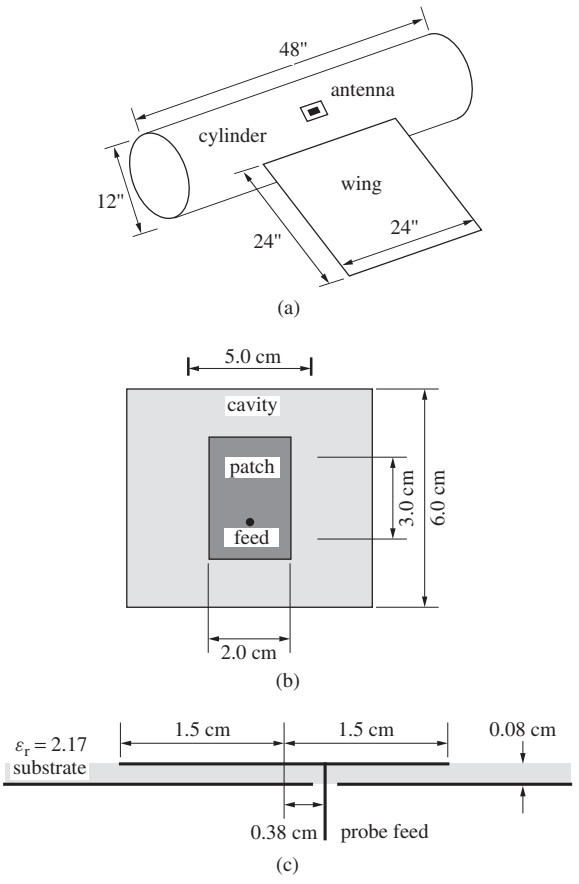


Figure 31.24 Microstrip patch antenna placed on a cylinder with a wing: (a) entire structure, (b) top view of the microstrip patch antenna, and (c) side view of the microstrip patch antenna. (Adapted from Ref. 38.)

31.9 SUMMARY

In this chapter, we described the finite-element analysis of complex antenna problems. First, we described the formulations of the FEM in both the frequency and time domains using vector basis functions, which are also known as edge elements. We then discussed two important issues for the FEM simulation of antennas. One was the mesh truncation and the approaches covered included absorbing boundary conditions, perfectly matched layers, boundary integral equations, and the FDTD interface. The other issue was the modeling of antenna feeds and the extraction of antenna parameters such as the input impedance and radiation patterns. The antenna feed models discussed included the current probe, the gap generator, and the waveguide feed such as coaxial lines. This was followed by the description of numerical schemes for modeling infinite periodic

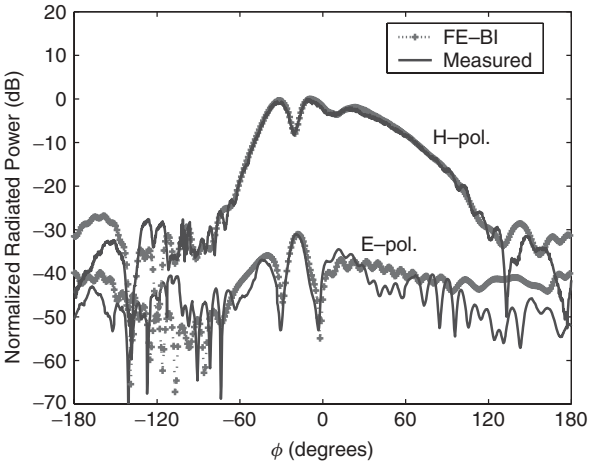


Figure 31.25 Normalized radiation pattern in the H-plane for a microstrip patch antenna placed on a cylinder with a wing. (Adapted from Ref. 38.)

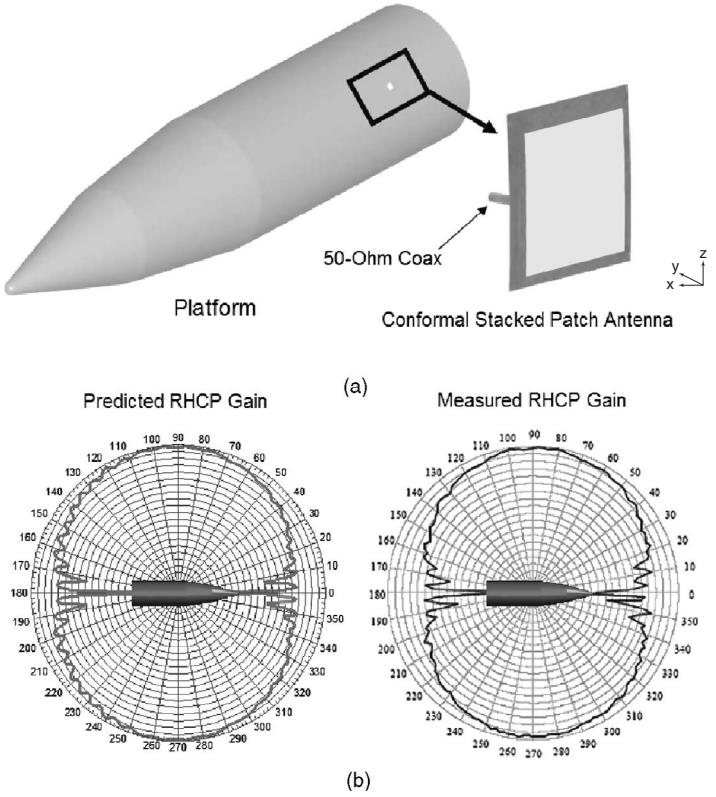


Figure 31.26 Two dual-frequency, coaxially fed, stacked patch antennas mounted on electrically large platform: (a) geometry and (b) predicted and measured RHCP gain for a cut plane taken axially through the antennas. Equivalent dB scales with 2-dB increments. Platform positioner impacted measured data for observation angles in the tail section.

phased arrays as well as finite arrays in both the frequency and time domains. In addition, practical guidelines were provided for the choice of solution technology based on either frequency- or time-domain methodologies. Numerical application examples were presented to demonstrate the FEM analysis of a variety of antennas. These included narrowband antennas such as microstrip patch and monopole antennas, broadband antennas such as a ridged horn, a spiral, a sinuous, and an inverted conical spiral antenna, and both infinite and finite periodic array antennas. Finally, we briefly addressed the FEM modeling of antenna–platform interaction. The application examples clearly demonstrated that the FEM is a powerful numerical simulation tool for the analysis and design of highly complicated antennas, and that it holds great potential for future antenna developments.

REFERENCES

1. R. L. Courant, Variational methods for the solution of problems of equilibrium and vibration, *Bull. Am. Math. Soc.*, Vol. 49, pp. 1–23, 1943.
2. P. P. Silvester, Finite element solution of homogeneous waveguide problems, *Alta Freq.*, Vol. 38, pp. 313–317, May 1969.
3. K. K. Mei, Unimoment method of solving antenna and scattering problems, *IEEE Trans. Antennas Propag.*, Vol. 22, pp. 760–766, November 1974.
4. S. P. Marin, Computing scattering amplitudes for arbitrary cylinders under incident plane waves, *IEEE Trans. Antennas Propag.*, Vol. 30, pp. 1045–1049, November 1982.
5. J. C. Nedelec, Mixed finite elements in R^3, *Numerische Math.*, Vol. 35, pp. 315–341, 1980.
6. A. Bossavit and J. C. Verite, A mixed FEM-BIEM method to solve 3-D eddy current problems, *IEEE Trans. Magn.*, Vol. 18, pp. 431–435, March 1982.
7. M. L. Barton and Z. J. Cendes, New vector finite elements for three-dimensional magnetic field computation, *J. Appl. Phys.*, Vol. 61, pp. 3919–3921, April 1987.
8. P. P. Silvester and R. L. Ferrari, *Finite Elements for Electrical Engineers*, 3rd ed., Cambridge University Press, Cambridge, UK, 1996.
9. J.-M. Jin, *The Finite Element Method in Electromagnetics*, 2nd ed., John Wiley & Sons, Hoboken, NJ, 2002.
10. J. L. Volakis, A. Chatterjee, and L. C. Kempel, *Finite Element Method for Electromagnetics: Antennas, Microwave Circuits and Scattering Applications*, IEEE Press, New York, 1998.
11. T. Orikasa, S. Washisu, T. Honma, and I. Fukai, Finite element method for unbounded field problems and application to two-dimensional taper, *Int. J. Num. Meth. Eng.*, Vol. 19, pp. 157–168, 1983.
12. J. D'Angelo and I. D. Mayergoyz, Finite element methods for the solution of RF radiation and scattering problems, *Electromagnetics*, Vol. 10, pp. 177–199, 1990.
13. J. M. Jin and J. L. Volakis, A hybrid finite element method for scattering and radiation by microstrip patch antennas and arrays residing in a cavity, *IEEE Trans. Antennas Propag.*, Vol. 39, pp. 1598–1604, November 1991.
14. J. M. Jin and J. L. Volakis, Scattering and radiation analysis of three-dimensional cavity arrays via a hybrid finite element method, *IEEE Trans. Antennas Propag.*, Vol. 41, pp. 1580–1586, November 1993.
15. D. T. McGrath and V. P. Pyati, Phased array antenna analysis with the hybrid finite element method, *IEEE Trans. Antennas Propag.*, Vol. 42, pp. 1625–1630, December 1994.
16. E. W. Lucas and T. P. Fontana, A 3-D hybrid finite element/boundary element method for the unified radiation and scattering analysis of general infinite periodic arrays, *IEEE Trans. Antennas Propag.*, Vol. 43, No. 2, pp. 145–153, February 1995.

17. Z. Lou and J. M. Jin, Finite element analysis of phased array antennas, *Microwave Optical Technol. Lett.*, Vol. 40, No. 6, pp. 490–496, March 2004.
18. R. Kindt, K. Sertel, E. Topsakal, and J. Volakis, Array decomposition method for the accurate analysis of finite arrays, *IEEE Trans. Antennas Propag.*, Vol. 51, pp. 1364–1372, June 2003.
19. S.-C. Lee, M. N. Vouvakis, and J.-F. Lee, A non-overlapping domain decomposition method with non-matching grids for modeling large finite antenna arrays, *J. Comput. Phys.*, Vol. 203, pp. 1–21, February 2005.
20. J. Rubio, M. A. Gonzalez, and J. Zapata, Generalized-scattering-matrix analysis of a class of finite arrays of coupled antennas by using 3-D FEM and spherical mode expansion, *IEEE Trans. Antennas Propag.*, Vol. 53, No. 3, pp. 1133–1144, March 2005.
21. Y. Li and J. M. Jin, A vector dual-primal finite element tearing and interconnecting method for solving 3-D large-scale electromagnetic problems, *IEEE Trans. Antennas Propag.*, Vol. 54, pp. 3000–3009, October 2006.
22. G. C. Chinn, L. W. Epp, and D. J. Hoppe, A hybrid finite-element method for axisymmetric waveguide-fed horns, *IEEE Trans. Antennas Propag.*, Vol. 44, No. 3, pp. 280–285, March 1996.
23. C. Zuffada, T. Cwik, and V. Jamnejad, Modeling radiation with an efficient hybrid finite-element integral-equation waveguide mode-matching technique, *IEEE Trans. Antennas Propag.*, Vol. 45, No. 1, pp. 34–39, January 1997.
24. D. T. McGrath and C. E. Baum, Scanning and impedance properties of TEM horn arrays for transient radiation, *IEEE Trans. Antennas Propag.*, Vol. 47, No. 3, pp. 469–473, March 1999.
25. A. D. Greenwood and J. M. Jin, Finite element analysis of complex axisymmetric radiating structures, *IEEE Trans. Antennas Propag.*, Vol. 47, No. 8, pp. 1260–1266, August 1999.
26. J. M. Gil, J. Monge, J. Rubio, and J. Zapata, A CAD-oriented method to analyze and design radiating structures based on bodies of revolution by using finite elements and generalized scattering matrix, *IEEE Trans. Antennas Propag.*, Vol. 54, No. 3, pp. 899–907, March 2006.
27. G. E. Antilla and N. G. Alexopoulos, Radiation and scattering from complex 3D curvilinear geometries using a hybrid finite element–integral equation method, *IEEE AP-S Int. Symp. Dig.*, pp. 1758–1761, 1992.
28. M. A. Gonzalez de Aza, J. A. Encinar, J. Zapata, and M. Lambea, Full-wave analysis of cavity-backed and probe-fed microstrip patch arrays by a hybrid mode-matching generalized scattering matrix and finite-element method, *IEEE Trans. Antennas Propag.*, Vol. 46, No. 2, pp. 234–242, February 1998.
29. C. A. Macon, L. C. Kempel, S. W. Schneider, and K. D. Trott, Modeling conformal antennas on metallic prolate spheroid surfaces using a hybrid finite element method, *IEEE Trans. Antennas Propag.*, Vol. 52, No. 3, pp. 750–758, March 2004.
30. A. D. Greenwood and J. M. Jin, A field picture of wave propagation in inhomogeneous dielectric lenses, *IEEE Antennas Propag. Mag.*, Vol. 41, No. 5, pp. 9–18, October 1999.
31. C. S. Liang, D. A. Streater, J. M. Jin, E. Dunn, and T. Rozendal, A quantitative study of Luneberg lens reflectors, *IEEE Antennas Propag. Mag.*, Vol. 47, No. 2, pp. 30–42, April 2005.
32. M. N. Vouvakis, C. A. Balanis, C. R. Birtcher, and A. C. Polycarpou, Ferrite-loaded cavity-backed antennas including nonuniform and nonlinear magnetization effects, *IEEE Trans. Antennas Propag.*, Vol. 51, No. 5, pp. 1000–1010, May 2003.
33. F. Bilotti, A. Toscano, and L. Vegni, FEM-BEM formulation for the analysis of cavity-backed patch antennas on chiral substrates, *IEEE Trans. Antennas Propag.*, Vol. 51, No. 2, pp. 306–311, February 2003.
34. C. J. Reddy, M. D. Deshpande, C. R. Cockrell, and F. B. Beck, Radiation characteristics of cavity backed aperture antennas in finite ground plane using the hybrid FEM/MoM technique

and geometrical theory of diffraction, *IEEE Trans. Antennas Propag.*, Vol. 44, No. 10, pp. 1327–1333, October 1996.

35. A. D. Greenwood, S. S. Ni, J. M. Jin, and S. W. Lee, Hybrid FEM/SBR method to compute the radiation pattern from a microstrip patch antenna in a complex geometry, *Microwave Optical Technol. Lett.*, Vol. 13, No. 2, pp. 84–87, October 1996.
36. J. M. Jin, J. A. Berrie, R. Kipp, and S. W. Lee, Calculation of radiation patterns of microstrip antennas on cylindrical bodies of arbitrary cross section, *IEEE Trans. Antennas Propag.*, Vol. 45, No. 1, pp. 126–132, January 1997.
37. M. Alaydrus, V. Hansen, and T. F. Eibert, Hybrid^2: combining the three-dimensional hybrid finite element–boundary integral technique for planar multilayered media with the uniform geometrical theory of diffraction, *IEEE Trans. Antennas Propag.*, Vol. 50, No. 1, pp. 67–74, January 2002.
38. J. Liu and J. M. Jin, Analysis of conformal antennas on a complex platform, *Microwave Optical Technol. Lett.*, Vol. 36, No. 2, pp. 139–142, January 2003.
39. D. Jiao and J. M. Jin, Fast frequency-sweep analysis of cavity-backed microstrip patch antennas, *Microwave Optical Technol. Lett.*, Vol. 22, No. 6, pp. 389–393, September 1999.
40. D. Jiao and J. M. Jin, Fast frequency-sweep analysis of microstrip antennas on a dispersive substrate, *Electron. Lett.*, Vol. 35, No. 14, pp. 1122–1123, July 1999.
41. D. Jiao and J. M. Jin, Time-domain finite element simulation of cavity-backed microstrip patch antennas, *Microwave Optical Technol. Lett.*, Vol. 32, No. 4, pp. 251–254, February 2002.
42. F. Edelvik, G. Ledfelt, P. Lötstedt, and D. J. Riley, An unconditionally stable subcell model for arbitrarily oriented thin wires in the FETD method, *IEEE Trans. Antennas Propag.*, Vol. 51, No. 8, pp. 1797–1805, August 2003.
43. Z. Lou and J. M. Jin, Modeling and simulation of broadband antennas using the time-domain finite element method, *IEEE Trans. Antennas Propag.*, Vol. 53, No. 12, pp. 4099–4110, December 2005.
44. Z. Lou and J. M. Jin, A novel dual-field time-domain finite-element domain-decomposition method for computational electromagnetics, *IEEE Trans. Antennas Propag.*, Vol. 54, pp. 1850–1862, June 2006.
45. Z. Lou and J. M. Jin, A dual-field domain-decomposition method for time-domain finite-element analysis of large finite arrays, *J. Comput. Phys.*, Vol. 222, pp. 408–427, March 2006.
46. L. E. R. Petersson and J. M. Jin, A three-dimensional time-domain finite element formulation for periodic structures, *IEEE Trans. Antennas Propag.*, Vol. 54, No. 1, pp. 12–19, January 2006.
47. L. E. R. Petersson and J. M. Jin, Analysis of periodic structures via a time-domain finite element formulation with a Floquet ABC, *IEEE Trans. Antennas Propag.*, Vol. 54, No. 3, pp. 933–944, March 2006.
48. N. M. Newmark, A method of computation for structural dynamics, *J. Eng. Mech. Div. Proc. Am. Soc. Civil Eng.*, Vol. 85, pp. 67–94, July 1959.
49. J.-F. Lee and Z. Sacks, Whitney elements time domain (WETD) methods, *IEEE Trans. Magn.*, Vol. 31, No. 3, pp. 1325–1329, May 1995.
50. S. D. Gedney and U. Navsariwala, Unconditionally stable finite element time-domain solution of the vector wave equation, *IEEE Microwave Guided Wave Lett.*, Vol. 5, pp. 332–334, 1995.
51. D. Jiao and J. M. Jin, Time-domain finite-element modeling of dispersive media, *IEEE Microwave Wireless Components Lett.*, Vol. 11, No. 5, pp. 220–222, May 2001.
52. F. Maradei, A frequency-dependent WETD formulation for dispersive materials, *IEEE Trans. Magn.*, Vol. 37, No. 5, Pt. 1, pp. 3303–3306, September 2001.

53. D. Riley and J. M. Jin, Modeling of magnetic loss in the finite-element time-domain method, *Microwave Optical Technol. Lett.*, Vol. 46, No. 2, pp. 165–168, July 2005.
54. A. F. Peterson, Absorbing boundary conditions for the vector wave equation, *Microwave Optical Technol. Lett.*, Vol. 1, pp. 62–64, April 1988.
55. J. P. Webb and V. N. Kanellopoulos, Absorbing boundary conditions for the finite element solution of the vector wave equation, *Microwave Optical Technol. Lett.*, Vol. 2, pp. 370–372, October 1989.
56. J. M. Jin, J. L. Volakis, and V. V. Liepa, A fictitious absorber for truncating finite element meshes in scattering, *Proc. Inst. Elec. Eng., Part H*, Vol. 139, No. 5, pp. 472–476, October 1992.
57. J.-P. Berenger, A perfectly matched layer for the absorption of electromagnetic waves, *J. Comput. Phys.*, Vol. 114, No. 2, pp. 185–200, 1994.
58. W. C. Chew and W. H. Weedon, A 3D perfectly matched medium from modified Maxwell's equations with stretched coordinates, *Microwave Optical Technol. Lett.*, Vol. 7, pp. 599–604, September 1994.
59. Z. S. Sacks, D. M. Kingsland, R. Lee, and J.-F. Lee, A perfectly matched anisotropic absorber for use as an absorbing boundary condition, *IEEE Trans. Antennas Propag.*, Vol. 43, pp. 1460–1463, December 1995.
60. S. D. Gedney, An anisotropic perfectly matched layer absorbing medium for the truncation of FDTD lattices, *IEEE Trans. Antennas Propag.*, Vol. 44, pp. 1630–1639, December 1996.
61. J. M. Jin and W. C. Chew, Combining PML and ABC for finite element analysis of scattering problems, *Microwave Optical Technol. Lett.*, Vol. 12, No. 4, pp. 192–197, July 1996.
62. J. M. Jin, X. Q. Sheng, and W. C. Chew, Complementary perfectly matched layers to reduce reflection errors, *Microwave Optical Technol. Lett.*, Vol. 14, No. 5, pp. 284–287, April 1997.
63. W. C. Chew and J. M. Jin, Perfectly matched layers in the discretized space: an analysis and optimization, *Electromagnetics*, Vol. 16, No. 4, pp. 325–340, July 1996.
64. M. Kuzuoglu and R. Mittra, Frequency dependence of the constitutive parameters of causal perfectly anisotropic absorbers, *IEEE Microwave Guided Lett.*, Vol. 6, No. 12, pp. 447–449, December 1996.
65. J. A. Roden and S. D. Gedney, Convolution PML (CPML): an efficient FDTD implementation of the CFS-PML for arbitrary media, *Microwave Optical Technol. Lett.*, Vol. 27, No. 5 pp. 334–339, December 2000.
66. D. Correia and J. M. Jin, On the development of a higher-order PML, *IEEE Trans. Antennas Propag.*, Vol. 53, No. 12, pp. 4157–4163, December 2005.
67. D. Jiao, J. M. Jin, E. Michielssen, and D. Riley, Time-domain finite-element simulation of three-dimensional scattering and radiation problems using perfectly matched layers, *IEEE Trans. Antennas Propagat.*, Vol. 51, No. 2, pp. 296–305, February 2003.
68. T. Rylander and J. M. Jin, Perfectly matched layers in three dimensions for the time-domain finite element method applied to radiation problems, *IEEE Trans. Antennas Propag.*, Vol. 53, No. 4, pp. 1489–1499, April 2005.
69. J.-Y. Wu, D. M. Kingsland, J.-F. Lee, and R. Lee, A comparison of anisotropic PML to Berenger's PML and its application to the finite-element method for EM scattering, *IEEE Trans. Antennas Propag.*, Vol. 45, No. 1, pp. 40–50, January 1997.
70. C. T. Wolfe, U. Navsariwala, and S. D. Gedney, A parallel finite-element tearing and interconnecting algorithm for solution of the vector wave equation with PML absorbing medium, *IEEE Trans. Antennas Propag.*, Vol. 48, No. 2, pp. 278–284, February 2000.
71. J. M. Jin and V. V. Liepa, Application of a hybrid finite element method to electromagnetic scattering from coated cylinders, *IEEE Trans. Antennas Propag.*, Vol. 36, No. 1, pp. 50–54, January 1988.

72. X. Yuan, Three-dimensional electromagnetic scattering from inhomogeneous objects by the hybrid moment and finite element method, *IEEE Trans. Microwave Theory Tech.*, Vol. 38, pp. 1053–1058, August 1990.

73. T. Eibert and V. Hansen, Calculation of unbounded field problems in free space by a 3D FEM/BEM-hybrid approach, *J. Electromagn. Waves Appl.*, Vol. 10, No. 1, pp. 61–78, 1996.

74. X. Q. Sheng, J. M. Jin, J. M. Song, C. C. Lu, and W. C. Chew, On the formulation of hybrid finite-element and boundary-integral method for 3D scattering, *IEEE Trans. Antennas Propag.*, Vol. 46, No. 3, pp. 303–311, March 1998.

75. J. Liu and J. M. Jin, A novel hybridization of higher order finite element and boundary integral methods for electromagnetic scattering and radiation problems, *IEEE Trans. Antennas Propag.*, Vol. 49, No. 12, pp. 1794–1806, December 2001.

76. M. N. Vouvakis, S.-C. Lee, K. Zhao, and J.-F. Lee, A symmetric FEM-IE formulation with a single-level IE-QR algorithm for solving electromagnetic radiation and scattering problems, *IEEE Trans. Antennas Propag.*, Vol. 52, No. 11, pp. 3060–3070, November 2004.

77. M. M. Botha and J. M. Jin, On the variational formulation of hybrid finite element–boundary integral techniques for electromagnetic analysis, *IEEE Trans. Antennas Propag.*, Vol. 52, No. 11, pp. 3037–3047, November 2004.

78. J. Liu and J. M. Jin, A highly effective preconditioner for solving the finite element–boundary integral matrix equation of 3-D scattering, *IEEE Trans. Antennas Propag.*, Vol. 50, No. 9, pp. 1212–1221, September 2002.

79. R. Coifman, V. Rokhlin, and S. Wandzura, The fast multipole method for the wave equation: a pedestrian prescription, *IEEE Antennas Propag. Mag.*, Vol. 35, pp. 7–12, June 1993.

80. W. C. Chew, J. M. Jin, E. Michielssen, and J. M. Song (Eds.), *Fast and Efficient Algorithms in Computational Electromagnetics*, Artech House, Norwood, MA, 2001.

81. D. Jiao, A. Ergin, B. Shanker, E. Michielssen, and J. M. Jin, A fast time-domain higher-order finite element–boundary integral method for 3-D electromagnetic scattering analysis, *IEEE Trans. Antennas Propag.*, Vol. 50, No. 9, pp. 1192–1202, September 2002.

82. J. Maloney, G. Smith, E. Thiele, O. Gandhi, N. Chavannes, and S. Hagness, Modeling of antennas, in *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, 3rd ed., A. Taflove and S. C. Hagness (Eds.), Artech House, Norwood, MA, 2005, pp. 607–676.

83. S. K. Maxumdar, J. E. Lumpp, and S. D. Gedney, Performance modeling of the finite-difference time-domain method on high performance parallel systems, *Appl. Comput. Electromagn. J.*, Vol. 13, No. 2, pp. 147–159, 1998.

84. D. J. Riley and C. D. Turner, Interfacing unstructured tetrahedron grids to structured-grid FDTD, *IEEE Microwave Guided Wave Lett.*, Vol. 5, No. 9, pp. 284–286, September 1995.

85. D. J. Riley and C. D. Turner, VOLMAX: a solid-model-based, transient, volumetric Maxwell solver using hybrid grids, *IEEE Antennas Propag. Mag.*, Vol. 39, No. 1, pp. 20–33, February. 1997.

86. M. Feliziani and F. Maradei, Mixed finite-difference/Whitney-elements time domain (FD/WE-TD) method, *IEEE Trans. Magn.*, Vol. 34, No. 5, Pt. 1, pp. 3222–3227, September 1998.

87. T. Rylander and A. Bondeson, Stable FDTD-FEM hybrid method for Maxwell's equations, *Comput. Phys. Comm.*, Vol. 125, pp. 75–82, 2000.

88. T. Rylander and A. Bondeson, Stability of explicit–implicit hybrid time-stepping schemes for Maxwell's equations, *J. Comput. Phys.*, Vol. 179, pp. 426–438, 2002.

89. T. Rylander, Stable FDTD-FEM hybrid method for Maxwell's equations, Ph.D. thesis, Department of Electromagnetics, Chalmers University of Technology, Gothenburg, Sweden, 2002.

90. F. Edelvik, Hybrid solvers for the Maxwell equations in time-domain, Ph.D. thesis, Department of Information Technology, Scientific Computing, Uppsala University, Uppsala, Sweden, 2002.
91. T. Rylander, F. Edelvik, A. Bondeson, and D. Riley, Advances in hybrid FDTD-FE techniques, in *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, 3rd ed., A. Taflove and S. C. Hagness (Eds.), Artech House, Norwood, MA, 2005, pp. 907–953.
92. G. Cohen and P. Monk, Gauss point mass lumping schemes for Maxwell's equations, *Numer. Methods Partial Differential Eq.*, Vol. 14, pp. 63–88, 1998.
93. M. Feliziani and F. Maradei, Modeling of electromagnetic fields and electrical circuits with lumped and distributed elements by the WETD method, *IEEE Trans. Magn.*, Vol. 35, No. 3, Pt. 1, pp. 1666–1669, May 1999.
94. W. L. Stutzman and G. A. Thiele, *Antenna Theory and Design*, John Wiley & Sons, Hoboken, NJ, 1981.
95. Z. Lou and J. M. Jin, An accurate waveguide port boundary condition for the time-domain finite element method, *IEEE Trans. Microwave Theory Tech.*, Vol. 53, No. 9, pp. 3014–3023, September 2005.
96. M. E. Veysoglu, R. T. Shin, and J. A. Kong, A finite-difference time-domain analysis of wave scattering from periodic structures: oblique incidence case, *J. Electromagn. Waves Appl.*, Vol. 7, pp. 1595–1607, December 1993.
97. J. Ren, O. P. Gandhi, L. R. Walker, J. Fraschilla, and C. R. Boerman, Floquet-based FDTD analysis of two-dimensional phased array antennas, *IEEE Microwave Guided Wave Lett.*, Vol. 4, pp. 109–111, April 1994.
98. N.-W. Chen, B. Shanker, and E. Michielssen, Integral-equation-based analysis of transient scattering from periodic perfectly conducting structures, *IEE Proc. Microwave Antennas Propag.*, Vol. 150, pp. 120–124, April 2003.
99. C. A. Balanis, *Antenna Theory*, Harper & Row, New York, 1982.
100. Z. Lou and J. M. Jin, Higher-order finite element analysis of finite-by-infinite arrays, *Electromagnetics*, Vol. 24, No. 7, pp. 497–514, October 2004.
101. J. Yeo, R. Mittra, and V. Prakash, Efficient analysis of a class of microstrip antennas using the characteristic basis function method (CBFM), *Microwave Optical Technol. Lett.*, Vol. 39, pp. 456–464, December 2003.
102. C. Farhat, A. Macedo, and M. Lesoinne, A two-level domain decomposition method for the iterative solution of high frequency exterior Helmholtz problems, *Numerische Math.*, Vol. 85, pp. 283–308, 2000.
103. C. Farhat, M. Lesoinne, P. Le Tallec, K. Pierson, and D. Rixen, FETI-DP: a dual-primal unified FETI method—part I: a faster alternative to the two-level FETI method, *Int. J. Numer. Meth. Eng.*, Vol. 50, pp. 1523–1544, 2001.
104. C. Farhat, P. Avery, R. Tezaur, and J. Li, FETI-DPH: a dual-primal domain decomposition method for acoustic scattering, *J. Computat. Acoustics*, Vol. 13, No. 3, pp. 499–524, September 2005.
105. W. L. Ko and R. Mittra, A combination of FDTD and Prony's methods for analyzing microwave integrated circuits, *IEEE Trans. Microwave Theory Tech.*, Vol. 39, pp. 2176–2181, December 1991.
106. W. Gropp, E. Lusk, and T. Sterling, *Beowulf Cluster Computing with Linux*, 2nd ed., MIT Press, Cambridge, MA, 2003.
107. G. Liu and C. A. Grimes, A quasistatic FDTD source model for coaxially driven monopole antennas, *Microwave Optical Technol. Lett.*, Vol. 26, pp. 30–34, July 2000.
108. G. Deschamps, Impedance properties of complementary multiterminal planar structures, *IRE Trans. Antennas Propag.*, Vol. 7, No. 8, pp. S371–S379, December 1959.

109. D. M. Pozar and D. H. Schaubert, Analysis of an infinite array of rectangular microstrip patches with idealized probe feeds, *IEEE Trans. Antennas Propagat.*, Vol. 32, pp. 1101–1107, October 1984.

110. A. J. Fenn, Theoretical and experimental study of monopole phased array antennas, *IEEE Trans. Antennas Propag.*, Vol. 34, No. 10, pp. 1118–1126, October 1985.

111. H. Holter, T.-H. Chio, and D. H. Schaubert, Elimination of impedance anomalies in single- and dual-polarized endfire tapered slot phased arrays, *IEEE Trans. Antennas Propag.*, Vol. 48, pp. 122–124, February 2000.

CHAPTER 32

Genetic Algorithms for Antennas

RANDY L. HAUPT

32.1 ANTENNA OPTIMIZATION

Optimization is the process of minimizing or maximizing an objective function. In this chapter we equate optimization to minimization and call the objective function the cost function. The idea is to find the set of variable values that minimizes the output of a cost function. Algorithms that start with a random guess of the variable values and move to the closest minimum are called local search algorithms. Algorithms that can leap out of the valley of the closest minima in search of the best or global minimum are called global search algorithms.

Optimizing antennas to closely approximate desired far-field responses or desired impedance characteristics over a bandwidth is an important part of antenna design today. Traditional optimization techniques are local optimization algorithms that often require computing derivatives of the cost function. Methods that use derivatives quickly converge to a local optimum solution that is close to the starting point. In addition, they require derivative calculations, work on only continuous variables, and are limited to optimizing a few variables. Random search methods do not require gradients and are less prone to getting stuck in a local minimum but are very slow. Neither class of algorithm is capable of optimizing many current antenna design problems. A number of interesting random search algorithms that are guided by laws of nature have emerged, including simulated annealing [1, 2], particle swarm optimization [3], and genetic algorithms (GAs) [4, 5]. These algorithms have similar characteristics and can find solutions to optimization problems that have befuddled traditional numerical optimization algorithms. GAs have emerged as the primary global optimization technique for antenna applications. This chapter introduces the GA and some examples of antenna design using the GA.

A GA has several advantages over traditional numerical optimization approaches:

1. It optimizes with either continuous or discrete variables.
2. It does not require calculating derivatives.
3. It simultaneously searches diverse areas of the cost surface.
4. It works with many variables.

Modern Antenna Handbook. Edited by Constantine A. Balanis

5. It is ideal for parallel computers.
6. It finds the optimum of multimodal cost surfaces.
7. It finds a list of optimum solutions rather than a single solution.
8. It works with numerically generated data, experimental data, or analytical functions.

These advantages have been put to good use in the design of antennas. This chapter starts with an introduction to GAs, then moves on to present a sampling of antenna design examples.

32.1.1 Introduction to Genetic Algorithms

A GA is a global random search that is guided by the principles of genetics and natural selection. Figure 32.1 is a flowchart of the GA that optimizes a cost function with five input variables. A single guess of the optimum input to the cost function is placed in a row vector called a chromosome. The GA works with many guesses at once, so a matrix with rows that are chromosomes is formed. At the start of the algorithm, all the chromosomes in the initial population are random. This matrix is passed to the cost function, and the output from the cost function (the cost) is placed in a column vector (see Figure 32.1). Most variations of a GA require a binary encoding of the input variables of the cost function. The cost function translates the binary values into integers or quantized real values.

The next step in the algorithm is natural selection. Chromosomes with low costs survive, while chromosomes with high costs are discarded. This step either keeps a certain percentage of the population or discards members with costs that exceed a certain level. Surviving chromosomes are known as the mating pool. In Figure 32.1, the best four chromosomes are retained while the worst four are discarded.

Discarded chromosomes from the population are replaced by new chromosomes called offspring. In order to create the offspring, parents must be found. The two most common approaches to finding parents are called roulette wheel selection and tournament selection. Roulette wheel selection assigns a probability of selection to each chromosome based on its cost or standing in the population. Chromosomes with low costs have a

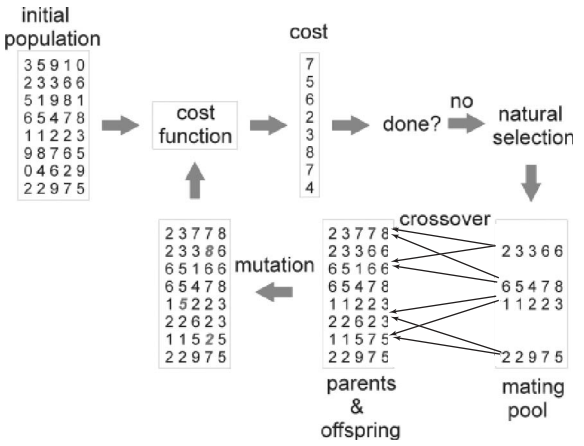


Figure 32.1 Flowchart of the GA.

higher probability of selection than chromosomes with high costs. Another approach is to randomly select a few chromosomes into two subpopulations. The chromosomes with the lowest costs in both subpopulations become parents. This approach is called tournament selection. Both methods mimic nature where the strongest in a population have the greatest odds of finding a mate and reproducing. In general, two parents produce two discarded chromosomes.

The next step is mating or creation of the offspring. Two parents produce two offspring by exchanging and/or combining values between two chromosomes. For two binary chromosomes, the generic procedure is called uniform crossover. A uniform random row vector that is the same length as a chromosome serves as a mask. If the random mask value at position m is less than one, then the value at position m of parent 1 becomes the value of offspring 1 at position m and the value at position m of parent 2 becomes the value of offspring 2 at position m. Alternatively, if the random mask value at position m is greater than one, then the value at position m of parent 2 becomes the value of offspring 1 at position m and the value at position m of parent 1 becomes the value of offspring 1 at position m. An example of uniform crossover with binary chromosomes is demonstrated in Eq. (32.1).

1	0	1	1	0	0	0	1	1	0	First parent
1	**1**	**1**	**0**	**0**	**1**	**0**	**1**	**0**	**1**	Second parent
0.93	0.51	0.77	0.13	0.45	0.26	0.33	0.64	0.95	0.17	Mask
1	0	1	**0**	**0**	**1**	**0**	1	1	**1**	First offspring
1	**1**	**1**	1	0	0	0	**1**	**0**	0	Second offspring

(32.1)

The offspring exhibit traits of both parents. Other common alternatives are single-point and double-point crossover.

At this point, mutations occur in the full population matrix. Mutations alter randomly selected values within the population. A binary mutation changes a one to zero or zero to one. A continuous value mutation replaces a value with another random value or a random perturbation to the value. Normally, the best chromosome receives no mutations. Figure 32.1 shows three mutations that occurred in the population (bold, italic numbers).

Costs are calculated for all the new chromosomes that are offspring and/or mutated. From here, the process repeats starting with natural selection. The algorithm stops when it finds a good enough solution or after a set number of generations or cost function calls. GAs do not really have a proof of convergence. When possible, it is usually a good idea to run a local optimization algorithm when the GA is finished and using the best chromosome as a starting point. Combining a GA with a local optimizer is called a hybrid GA.

As an example, consider the two dipole antennas that are L long and separated by a distance x in Figure 32.2. First, a cost function must be defined. In this case, the inputs

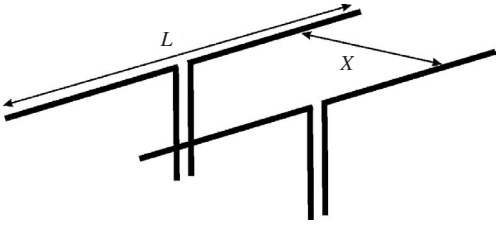


Figure 32.2 Diagram of the two-dipole model with L and x as variables.

to the cost function are (x, L). A method of moments program is used to calculate the currents and fields associated with the dipoles. One possible cost is the gain of the two dipoles when they have equal voltages applied. Since we would want to maximize gain, the cost would be the negative of the gain. The cost surface associated with the gain is shown in Figure 32.3 for input values $0.3 \leq x \leq 1.5$ and $0.3 \leq L \leq 2.0$. There are two very distinct minima within the variable constraints. The one at (x, L) = (0.63, 1.22) is the global minimum with a gain of 10.7 dB. The local minimum occurs at (x, L) = (1.65, 1.23) with a gain of 9.13 dB. A completely different optimization results when the reflection coefficient is the cost. Figure 32.4 is a plot of the magnitude of the reflection coefficient as a function of x and L over the same ranges as before. The transmission line is assumed to be 70 ohms, and the reflection coefficient, s_{11}, is found

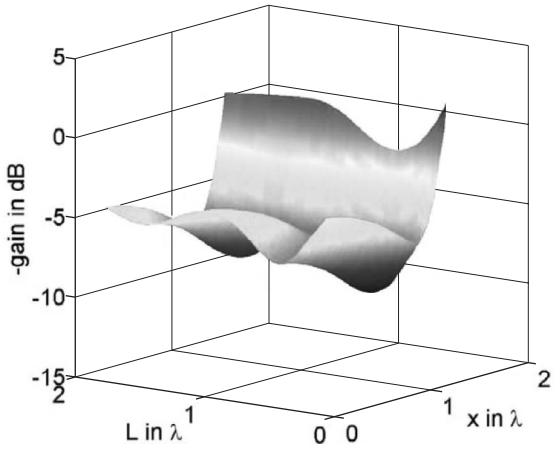


Figure 32.3 Cost surface when the cost is the negative of the array gain as a function of dipole spacing and length.

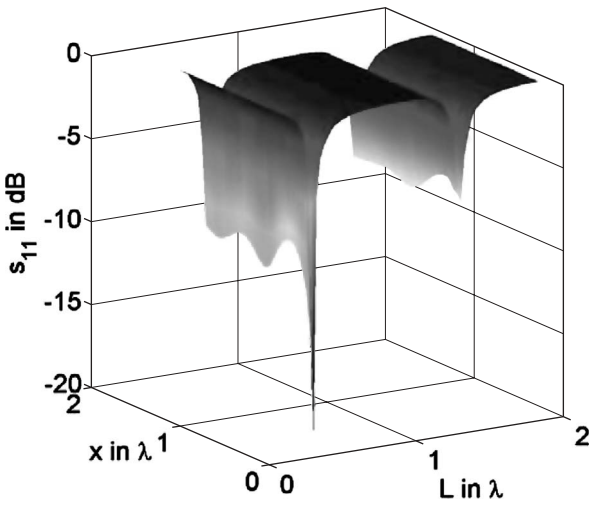


Figure 32.4 Cost surface when the cost is the magnitude of the reflection coefficient as a function of dipole spacing and length.

TABLE 32.1 Highest Gain for the Two Dipoles Found by the Nelder Meade Algorithm

Starting Point	Ending Point	Cost
(0.5, 0.5)	(0.63, 1.22)	−10.07
(1.5, 2.0)	(0.75, 3.25)	−7.41
(1.3, 1.7)	(1.65, 1.23)	−9.13

using the calculated antenna input impedance. This time the cost function has two deep troughs with many minima in each trough. In this case, the cost function has a global minimum at $(x, L) = (0.67, 0.48)$ with a $s_{11} = -18.4$ dB.

To demonstrate the limitations associated with traditional local optimization techniques, the Nelder Mead downhill simplex method is used to find the minimum of the cost surface in Figure 32.3. The minimum found depends on the starting point. Table 32.1 lists three starting points for the Nelder Meade algorithm and their corresponding ending points and cost of the optimal solution found. In this case, the first starting point found the global minimum. Not only did the second starting point end up in a local minimum, but that local minimum was out of the range of the variables. It is possible to do a transformation of variables to keep the optimization within bounds, but that would complicate the optimization process. The third starting point found the local minimum within the variable constraints. This local minimum might be good enough for the antenna design, but it is almost 1 dB less than the global minimum.

Can a GA do better than the Nelder Meade algorithm? Since the GA is a random search, five independent runs were made to optimize the gain of the dipole pair. The GA has a population size of 8, has a mutation rate of 10%, used tournament selection, and used 50% replacement. Figure 32.5 shows the convergence of all five runs. Each time, the GA converged to the global minimum. Sometimes a GA may start narrowing in on the local minimum, then either crossover or mutation pushes a chromosome into the

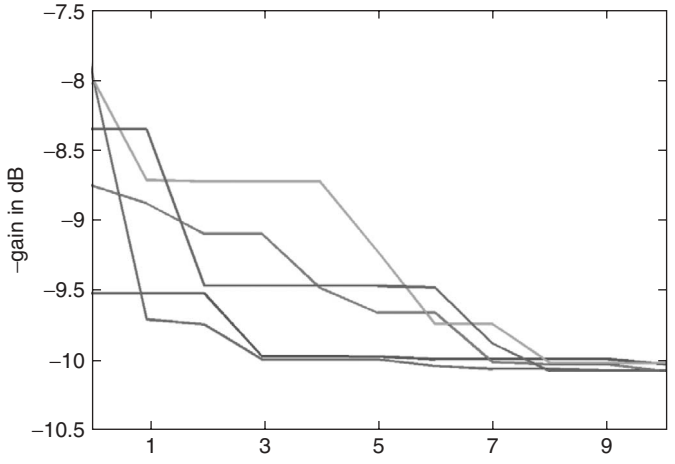


Figure 32.5 GA convergence over five independent runs.

region of the global minimum. Determining the variable constraints is important, because the GA will not exceed these constraints like a typical local optimizer.

32.1.2 Multiple Objective Optimization

Oftentimes, there is more than one objective in the optimization. For instance, an antenna design may require the highest gain and the lowest sidelobe levels possible.

$$\begin{aligned} f_1(\mathbf{v}) &= \max\{\text{sidelobe level}\} \\ f_2(\mathbf{v}) &= -gain \end{aligned} \tag{32.2}$$

where $\mathbf{v}$ are the input variables, such as element weights. Placing a negative sign in front of the gain switches the function so that a maximum gain results in a minimum cost. Low sidelobes and high gain are conflicting objectives when designing an antenna array having a set number of elements and element spacing. Thus there are many optima depending on the trade-off between gain and sidelobe levels.

Equation (32.2) implies there is no one best solution with respect to all the objectives. Consider the case of minimizing the sidelobe level of a 21-element array by an optimized phase taper. For this case, a plot of f_1 versus f_2 for 1000 points is shown in Figure 32.6 for $0 \leq f_1, f_2 \leq 1$. The minimum point of each independent function (min $\{f_1\}$, min $\{f_2\}$) occurs outside the feasible region (set of all possible points that satisfy the constraints) in the plot. The set of points bounding the bottom and left side of the feasible region is the Pareto front. A dashed line connects the points that approximate the Pareto front in Figure 32.6. Two possible approaches to multiple objective optimization presented here are (1) weighting and summing the cost functions and (2) finding the Pareto front.

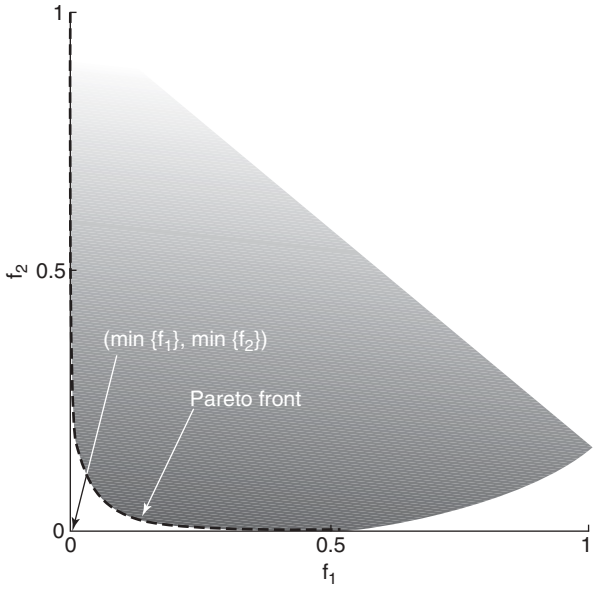


Figure 32.6 There are 10,000 points points plotted with the Pareto front indicated by the dashed line.

The most straightforward approach to multiple objective optimization is to weight the output of each objective function and add them together, so that

$$F = \sum_{n=1}^{N} w_n f_n \tag{32.3}$$

where

$$f_n = \text{cost function } n \text{ and } 0 \leq f_n \leq 1$$

$$w_n = \text{weighting factor and } \sum_{n=1}^{N} w_n = 1$$

Finding the best values for w_n is challenging, because different weights result in different values of F for the same f_n. Weighting and summing the f_n provides the one value that the GA minimizes. For the low sidelobe array

$$F = wf_1 + (1 - w) f_2 \tag{32.4}$$

Since this approach results in a single cost value, no modification to the GA is required.

Another approach to optimizing multiple cost functions is the Pareto GA. Pareto optimal means that reallocation of resources cannot minimize the output of one objective without raising the output of another objective. Lacking additional information, all solutions on the Pareto front are equally satisfactory. Thus a selection of alternatives are available that yield acceptable outcomes.

The population of a Pareto GA consists of chromosomes that are the current guesses for some points on the Pareto front. A chromosomes position within the population is ranked according to dominance. Chromosome n dominates chromosome m if chromosome n has a lower output than chromosome m for at least one of the objective functions and is not worse with respect to all other objective functions. In other words, c dominates d if

$$f_1(c) < f_1(d) \quad \text{and} \quad f_2(c) \leq f_2(d) \tag{32.5}$$

or

$$f_1(c) \leq f_1(d) \quad \text{and} \quad f_2(c) < f_2(d) \tag{32.6}$$

A chromosome is Pareto optimal if no other chromosome dominates that chromosome with respect to the cost functions. A chromosome is nondominated if no chromosome can be found that dominates it. Once this set of chromosomes is found, then the user can select a single chromosome based on various postoptimization trade-offs that equate to different weightings of the component costs. One way of finding the Pareto front is to run a GA for many different combinations of the cost function weights in Eq. (32.3). Each optimal chromosome found from a particular weighting combination is on the Pareto front. This approach requires too many runs for all the different weighting combinations to estimate the Pareto set.

The first attempt at a multiobjective GA, called the vector evaluated GA (VEGA), tended to converge to a single solution [6] rather than finding a Pareto front. Newer algorithms were developed to overcome this problem. Some examples include a niching

strategy called sharing [7], the multiobjective GA (MOGA) [8], and a nondominated sorting GA (NSGA) [9].

A Pareto GA needs a large population size in order to define the Pareto front. The GA used to find the Pareto front in Figure 32.7 had $N_{\text{pop}} = 500$, discards 50% of the population, has a mutation rate of 10%, and uses tournament selection. Those chromosomes in the final population on the Pareto front are connected by a dashed line. The Pareto front allows an easy way to trade off between the two cost functions.

Pareto GAs have been used in electromagnetics for looking at trade-offs involved with absorber and antenna design. An NSGA was used to optimize the broadband behavior of multilayer absorptive coatings [10]. The NSGA proved to perform better than two other implementations of a Pareto GA. Results were comparable to those of a GA optimization with weighted objective functions, except the NSGA provides a large number of design alternatives. A demonstration of the effectiveness of a Pareto GA in performing a trade-off analysis between beamwidth and sidelobe levels for a linear array appeared in Ref. 11. Pareto GAs require a large number of objective function evaluations. Neural network interpolation and special zooming strategies were combined with a Pareto GA to reduce the number of function evaluations in the design of an electromagnet [12]. A Pareto GA was compared to single objective optimization in the design of Yagi–Uda antennas [13]. Antenna size, bandwidth, and efficiency were the objectives in the design of an electrically small planar antenna using an inductively coupled feed using a Pareto GA [14].

32.1.3 Genetic Algorithm Parameters

The mutation rate and population size for a GA are the major contributions to the convergence speed of a GA. Other operators and parameters affect GA convergence but to a lesser extent. De Jong did the first study of GA parameters versus. performance [15].

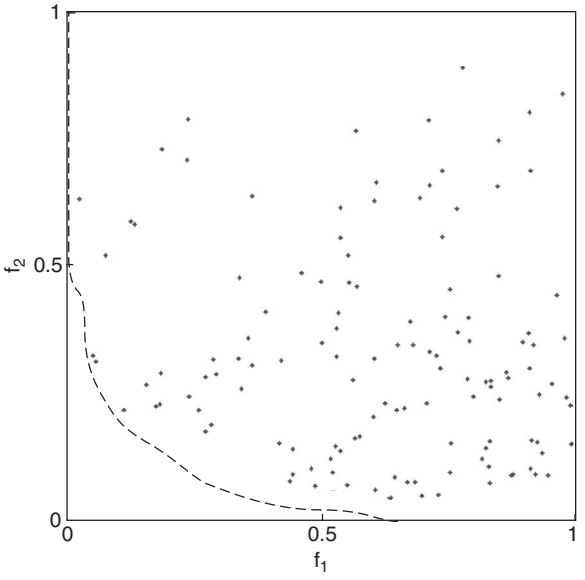


Figure 32.7 The 500 members of the GA are plotted with the dashed line drawn through the estimated Pareto front.

He tested his GAs on the five cost functions for different mutation rates, population sizes, crossover rates, and replacement percentages.

Grefenstette used a GA to optimize parameters for GAs [16] that found the minimum of the De Jong test functions. Another study added five more cost functions and used 8400 possible combinations of parameter settings [17]. Thomas Back postulated that the optimal mutation rate is $1/\ell$, where ℓ is the length of the chromosome [18]. He also found that convergence improved when starting the GA with large mutation rates of up to 50% and gradually decreasing to $1/\ell$. Gao showed that the larger the probability of mutation and the smaller the population, the faster the GA should converge for short-term performance [19]. Iterative approaches where mutation rate varies over the course of a run such as done by Back [20, 21] and Davis [22] are probably best, but there is no acceptable strategy as to how to adaptively change the parameters. A GA parameter study was done on antenna array factors [23], and in these experiments, the best mutation rate for GAs was found to be between 5% and 20% while the population size should be less than 16. Parameter values are highly dependent on the cost function and the particular implementation of the GA. A more lengthy discussion of GA parameters can be found in Ref. 24.

32.2 GA OPTIMIZATION OF ANTENNA ARRAYS

The cost function for most phased array optimization is based on the array factor. An arbitrary array in three-dimensional space has an array factor given by

$$AF = \sum_{n=1}^{N} w_n e^{jk[x_n \sin\theta \cos\phi + y_n \sin\theta \sin\phi + z_n \cos\theta]} \tag{32.7}$$

where

$$\begin{aligned} w_n &= a_n e^{j\delta_n} = \text{complex weight at element } n \\ a_n &= \text{amplitude weight at element } n \\ \delta_n &= \text{phase weight at element } n \\ k &= 2\pi/\lambda \\ \lambda &= \text{wavelength} \\ (x_n, y_n, z_n) &= \text{location of element } n \\ \theta &= \text{elevation angle} \\ \phi &= \text{azimuth angle} \\ N &= \text{number of array elements} \end{aligned}$$

For a given frequency, the potential optimization variables are the amplitude weights, the phase weights, and/or the element locations. Possible costs include sidelobe levels, beamwidth, nulls, and bandwidth. The examples in this section all try to minimize the maximum sidelobe level in the array factor of the linear array architecture in Figure 32.8.

GAs have found the most use in optimizing antenna arrays. There are far too many papers to list all of them in this chapter. The reader is referred to Ref. 25, which gives an overview of the various applications of the GA in antenna arrays and provides an extensive list of references by category. A few of the journal articles that cover various

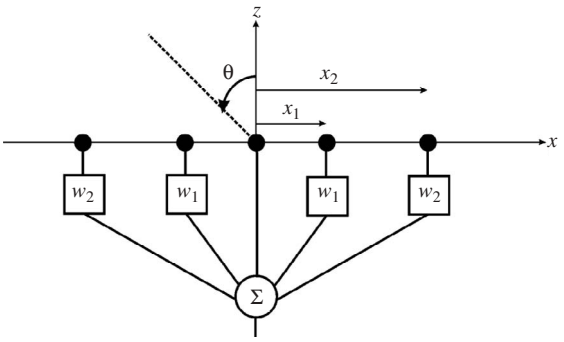


Figure 32.8 Diagram of the linear array to be optimized.

array GA optimization topics are given in Refs. 26–37. The examples that follow find the lowest maximum sidelobe level of a nonuniformly spaced array, a phase tapered array, and a subarray tapered array.

32.2.1 Nonuniform Spacing

The first example has fixed weights at the elements but the element spacing is variable. Nonuniformly spacing the elements of an array effectively places an amplitude taper across the array: the amplitude is highest where the element density is highest. A GA is ideal for finding the spacings because the cost function is multimodal and the variables are inside the cosine function [38]. As an example, a GA minimizes the maximum sidelobe level of a 31-element nonuniformly spaced linear array lying along the x-axis. The array factor cost function for an odd number of elements having symmetry about the center of the array is given by

$$x_n = x_{n-1} + d_0 + \Delta_n$$

$$sll = \max\left\{1 + 2\sum_{n=1}^{(N-1)/2} \cos(kx_n \sin\theta)\right\} \quad \text{for } \theta > \theta_{\text{MB}} \tag{32.8}$$

where

$$\begin{aligned} x_0 &= 0 \\ d_0 &= \lambda/2 \\ \Delta_n &= \text{variable spacing between 0 and } \lambda/2 \\ \theta_{\text{MB}} &= \text{first null next to the main beam} \end{aligned}$$

Elements have a minimum spacing of d_0 and a maximum spacing of 1λ. Since the center element is at $x = 0$ and the left side of the array has the same spacing as the right hand of the array, there are 15 input variables to the cost function. Figure 32.9 is the optimized array factor with a peak sidelobe of the array factor at -22.13 dB below the peak of the main beam. The relative spacing of the elements appears at the top of the plot. It is hard to determine if this result is the global optimum. Since peaks of several of the sidelobes are at the maximum sidelobe level, this result is probably close to optimal.

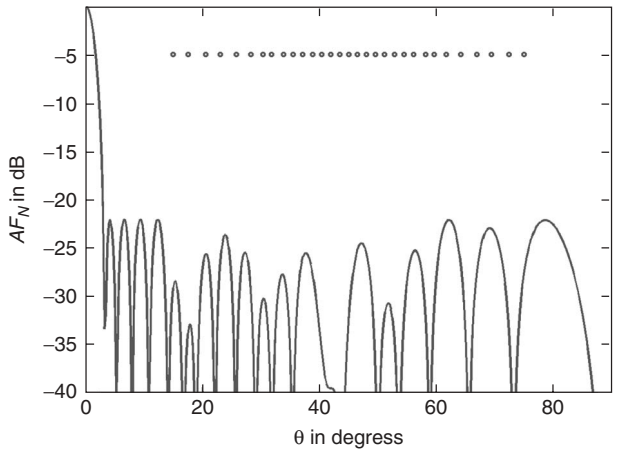


Figure 32.9 Optimized array factor for nonuniform spacing of a 31-element array. The small circles at the top of the figure are the relative element spacings.

32.2.2 Phase Tapering

The second example finds the phase weights that generate low sidelobes in the array factor [39]. The cost function returns the maximum sidelobe level.

$$sll = \max\left\{1 + 2\sum_{n=1}^{(N-1)/2} e^{j\delta_n}\cos[k(n-1)d_0\sin\theta]\right\} \quad \text{for } \theta > \theta_{\text{MB}} \tag{32.9}$$

Figure 32.10 is the optimized array factor with a maximum sidelobe of -17.0 dB below the peak of the main beam. The peak of the main beam is normalized to that of the array with zero phase shifts. Gain loss is the trade-off for getting low sidelobes. All the sidelobes are at nearly the same height, so the solution should be nearly optimal.

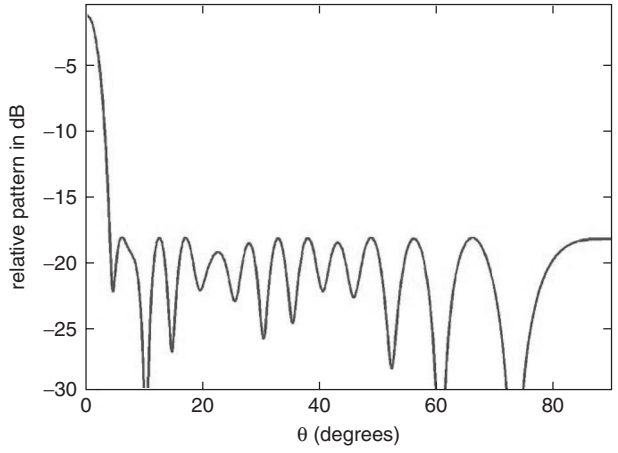


Figure 32.10 Optimized array factor for phase tapering a 31-element array.

32.2.3 Subarrays

Feed networks for large low sidelobe phased arrays usually use subarrays to simplify design and manufacture of the array. One way to reduce the cost of a large array is to place low sidelobe amplitude weights at the subarray output ports rather than at each element. Each element in the subarray is uniformly weighted, so all subarrays have identical feed networks. When the subarrays are identical, then the design, testing, and maintenance of the array become cheaper than having all the subarrays with different feed networks. Unfortunately, amplitude weighting at the subarray ports creates grating lobes due to the effective quantization of the array amplitude taper.

A trade-off exists between low sidelobe performance and simplicity of design. A genetic algorithm (GA) has been used to design optimum amplitude tapers for both the elements in the subarrays and the subarray outputs for linear and planar arrays [40].

The cost function is based on the linear subarray model in Figure 32.11. Element amplitude tapers are identical in each subarray. Subarray weights multiply the element weights to produce an effective weight at each element. The weights are symmetric about the center of the array so the array factor is given by

$$AF = 2\sum_{q=1}^{N_s} b_q \sum_{n=1}^{N_e} a_n \cos\{kd_0 \sin\theta[n - 0.5 + (q-1)N_e]\} \tag{32.10}$$

where

a_n = element amplitude weights
b_q = subarray amplitude weights
$2N_s$ = number of subarrays
N_e = number of elements in a subarray

Consider a 128-element array that is broken into 16 subarrays of 8 elements each. Assume all element tapers within each subarray are identical. The element closest to the center of the array has an amplitude of one or $a_1 = 1$. Subarrays are symmetric about the

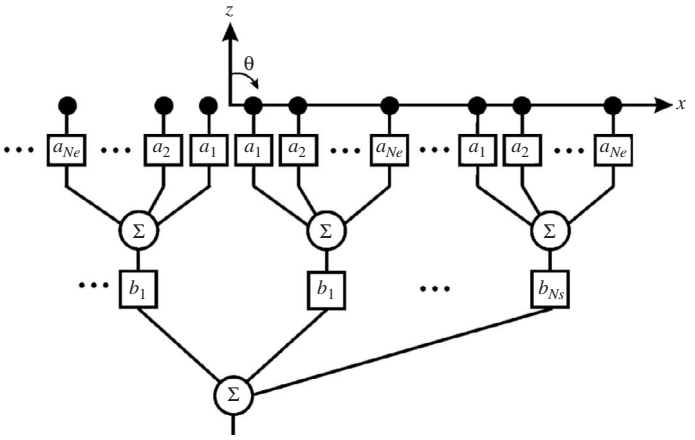


Figure 32.11 Linear array with subarrays.

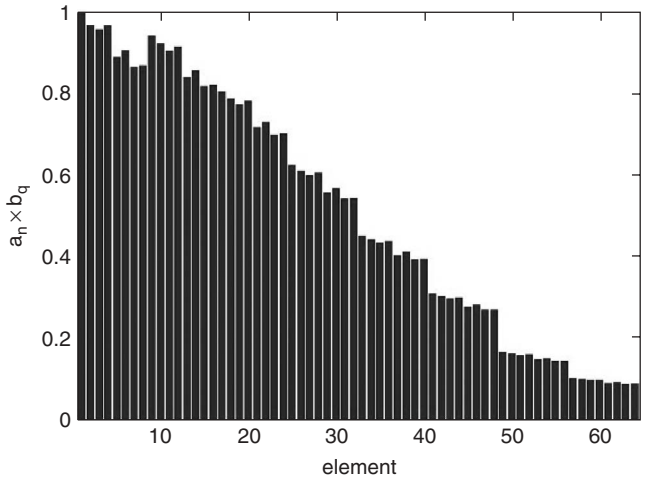


Figure 32.12 Graph of the subarray weight times the element weight for each element.

center of the array and the subarray closest to the center has an amplitude of one or $b_1 = 1$. As a result, there are a total of $7 + 7 = 14$ unknowns for this problem. Figure 32.12 is a plot of the subarray weight times the element weight at each element in the array. The element-to-element variation is relatively small. The corresponding array factor appears in Figure 32.13. Sidelobes remain at least 38.1 dB below the peak of the main beam. Remnants of the "split" grating lobes appear in the predicted locations [41] of

$$\theta_m = \sin^{-1}\left(\frac{m\lambda}{N_e d_0}\right) \tag{32.11}$$

or $\theta_m = 14.5^\circ$, 30°, 48.6°, and 90° for Figure 32.13. This array has very low sidelobes while at the same time having a simplified feed network.

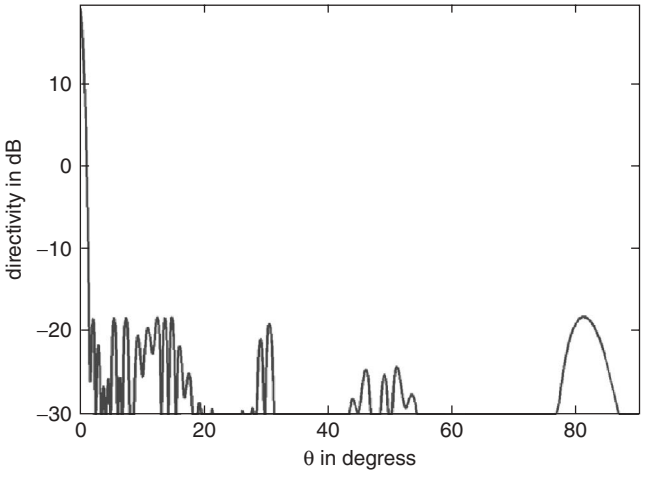


Figure 32.13 Optimized element and subarray amplitude weights produce this array factor.

32.3 GA OPTIMIZATION OF APERTURE ANTENNAS

The GA has been extensively applied to optimize microstrip patches. Horns and reflectors have received much less attention, perhaps because the modeling of horns and reflectors is more computationally expensive. This section starts with horn antennas, moves on to reflector antennas, and ends with microstrip antennas.

32.3.1 Horn Antennas

The GA has been applied to optimize the design of several different types of horn antennas. Some examples include an ultrawideband tapered resistive TEM horn [42], the optimal choke distribution for a horn at X-band [43], and corrugated horns [44–47].

If the horn is used to feed a reflector antenna, then the phase center of the horn is placed at the focal point of the reflector. The phase center of the horn moves with frequency and is not generally located at the same point for the E-and H-planes. Also, the phase center is a function of angle. Consequently, the phase center is relevant over the field of view of the reflector surface.

The example presented here applies a GA to the design of a horn antenna having a common phase center in the E-and H-planes over a limited angular range. Calculations are made using a method of moments model of a horn at 10.3 GHz [48]. A diagram of the X-band horn appears in Figure 32.14. The waveguide has dimensions of $a = 2.29$ cm and $b = 1.02$ cm and the feed probe is $\lambda/4$ from the back wall.

The cost function is based on the standard deviation of the phase in the E-and H-planes when $-45^\circ \leq \phi \leq 45^\circ$ and $45^\circ \leq \theta \leq 135^\circ$

$$f(x_c, L_x, A, B) = \max\{\sigma_\theta, \sigma_\phi\} \tag{32.12}$$

where the input variables are defined by

x_c = distance of phase center from horn aperture, $0 \leq x_c \leq \ell + 2$ cm
L_x = distance from waveguide/horn junction and horn aperture, $0.5\lambda \leq L_x \leq 4.5\lambda$
A = horn dimension in y-direction, 2.29 cm $\leq A \leq$ 11.43 cm
B = horn dimension in z-direction, 1.02 cm $\leq B \leq$ 7.11 cm

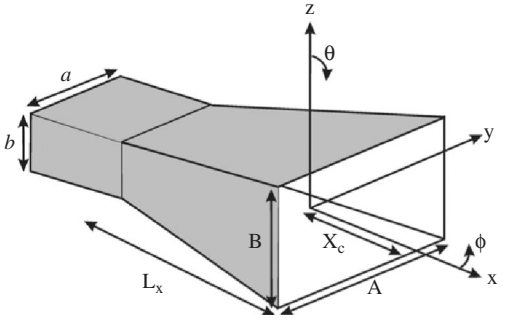


Figure 32.14 Diagram of a horn antenna.

and the costs are given by the mean square error of the field phase in the E-and H-planes, or

$$\sigma_\theta = \sqrt{\frac{1}{N}[\angle E_\theta(\theta_n, \phi = 0^\circ) - \overline{\angle E_\theta(\theta, \phi = 0^\circ)}]^2} \tag{32.13}$$

$$\sigma_\phi = \sqrt{\frac{1}{N}[\angle E_\theta(\theta = 90^\circ, \phi_n) - \overline{\angle E_\theta(\theta = 90^\circ, \phi_n)}]^2} \tag{32.14}$$

where

N = number of angles
$\angle E$ = phase of E
$\overline{E}$ = average of E
N = number of field points

If $\sigma_\theta = 0$ and $\sigma_\phi = 0$, then the phase center in the E-and H-planes are at the origin of the coordinate system.

A hybrid GA consisting of a continuous GA with a population size of 8 and a mutation rate of 20% coupled with a Nelder Mead downhill simplex algorithm found the minimum of $f = 0.57^\circ$ at $p_c = 0.28$ cm, $\ell = 4.31$ cm, $A = 4.22$ cm, and $B = 2.19$ cm. The optimized horn has small phase variations in the E-and H-plane cuts over $\pm 45^\circ$ from the peak of the main beam. Figure 32.15 is a plot of the phase of E_θ in the E-plane, and Figure 32.16 is a plot of the phase of E_θ in the H-plane. The vertical dotted lines bound the region where the phase is to be constant. In these regions, the phase is smooth and has little variation.

32.3.2 Reflector Antennas

Reflector antennas may have phased array feeds in order to beam steer, place nulls, shape the radiation pattern, or create multiple beams [49]. As the size of the array increases to provide more control over the antenna pattern, the blockage also increases. Blockage

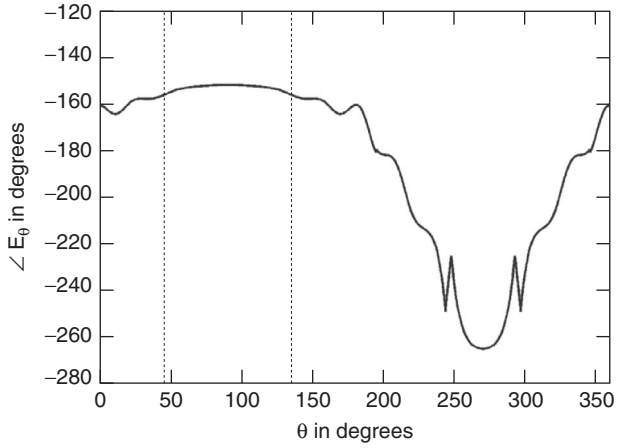


Figure 32.15 Plot of E_θ in the E-plane.

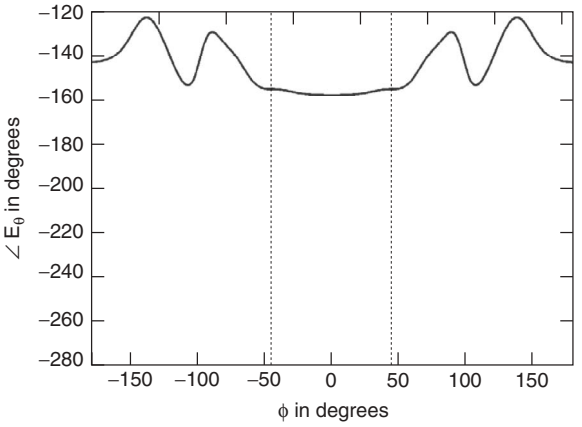


Figure 32.16 Plot of E_θ in the H-plane.

is undesirable, because it reduces the antenna gain and increases the sidelobe levels. A reflector antenna can be calibrated for maximum gain in the presence of blockage by adjusting the amplitude and phase of the array elements [49–52]. A GA can be used to find the element weights that maximize the output power from the reflector.

The two-dimensional offset reflector in Figure 32.17 has an aperture size of 20λ in the y-direction and an $F/D = 0.4$. Its feed has seven elements spaced $\lambda/2$ apart. A perfect conducting strip that is 4λ wide is placed $\lambda/4$ behind the array. The center element has an amplitude of one and a phase of zero. All other elements can take on continuous amplitude values between zero and one and continuous phase values between zero and 2π. If the array feed is uniformly weighted, then the antenna pattern is shown by the dashed line in Figure 32.18. The main beam has a shallow null at boresight. To get rid of the null, the reflector antenna is calibrated using a GA. The GA found the 12 amplitude and phase weights that maximize the output power at $\phi = 0^\circ$. After optimization, the main beam increases by 4.3 dB. The calibrated pattern also has better main beam resolution and overall lower sidelobes than the uncalibrated pattern.

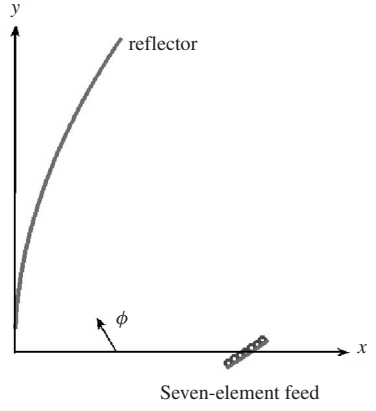


Figure 32.17 Diagram of the offset reflector with seven-element phased array feed.

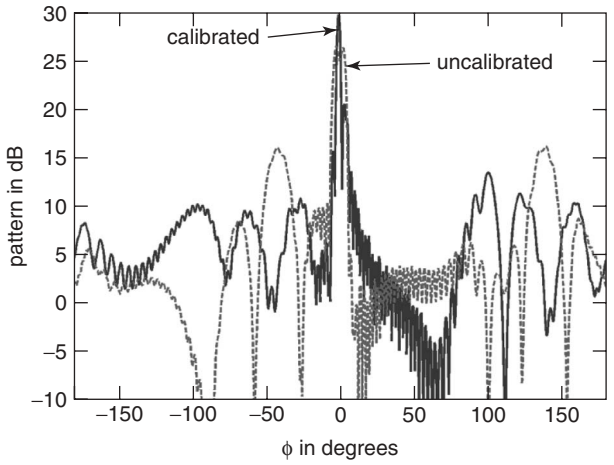


Figure 32.18 The calibrated reflector pattern (solid line) is superimposed on the uncalibrated reflector pattern (dashed line).

32.3.3 Microstrip Antennas

GAs have been used extensively to optimize microstrip patch antennas. Some of the initial work on microstrip optimization with GAs was reported in Refs. [53–55]. These designs were narrowband and involved optimizing the shape of a patch. More recent work concentrates on the design of broadband patch antennas [56–58]. Some of the designs result in interesting shapes that are nonintuitive.

The example here is to design a rectangular patch for circular polarization (either right-hand or left-hand) at 10 GHz. A diagram of the patch antenna appears in Figure 32.19. Input variables for the cost function are

$$
\begin{aligned}
(F_x, F_y) &= \text{position of probe feed} \\
L_x, L_y &= \text{patch length in } x\text{-and } y\text{-directions} \\
h &= \text{substrate thickness} \\
\varepsilon_r &= \text{relative dielectric constant of the substrate}
\end{aligned}
$$

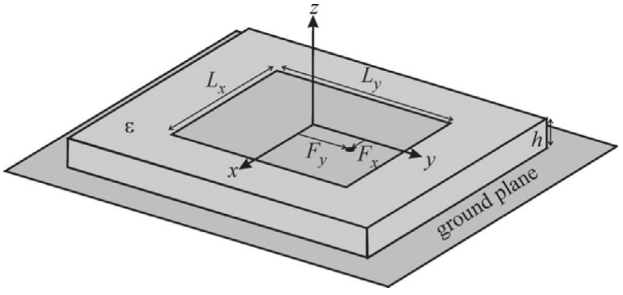


Figure 32.19 Diagram of probe-fed microstrip patch.

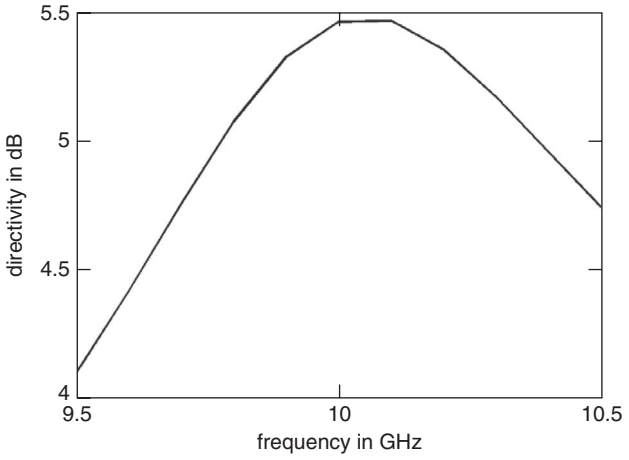


Figure 32.20 Plot of the directivity of the optimized patch as a function of frequency.

The cost function is calculated using the method of moments and returns the following value:

$$cost = \max\left\{\frac{||E_\theta| - |E_\varphi||}{|E_\theta| + |E_\varphi|}, \left||\angle E_\theta - \angle E_\varphi| - \frac{\pi}{2}\right|, |s_{11}|\right\} \tag{32.15}$$

An infinite ground plane and dielectric were assumed in the model. The first two terms in Eq. (32.15) are zero for circular polarization, and $|s_{11}|$ is zero for a perfect match at 50 Ω. When the patch is circularly polarized and perfectly matched, then cost = 0.

A hybrid GA, in which a continuous variable GA was combined with a Nelder Mead downhill simplex algorithm, was used to optimize the patch. The first run used a coarse model of the patch having a triangular mesh with sides that were λ/15 long. Next, the best result from this optimization run was then used as a starting point for the Nelder Mead downhill simplex algorithm that optimized a model with triangular mesh with sides that were λ/25 long. The final cost was 0.0317 with corresponding optimal variables given by

$$\begin{aligned}(F_x, F_y) &= (3.60, 1.21) \text{ mm}\\ L_x, L_y &= 10.35, 8.935 \text{ mm}\\ h &= 1.47 \text{ mm}\\ \varepsilon_r &= 2.04\end{aligned}$$

Figure 32.20 is a plot of the optimized patch directivity for left-hand circular polarization over a frequency range from 9.5 to 10.5 GHz. Figures 32.21 and 32.22 are graphs of s_{11} and the inverse axial ratio over the same frequency range. These graphs show that the design at 10 GHz was met with success.

32.4 GA OPTIMIZATION OF WIRE ANTENNAS

Many different types of wire antennas have been optimized by GAs. Altshuler and Linden invented the "genetic antenna" by using a GA to bend a monopole into strange shapes

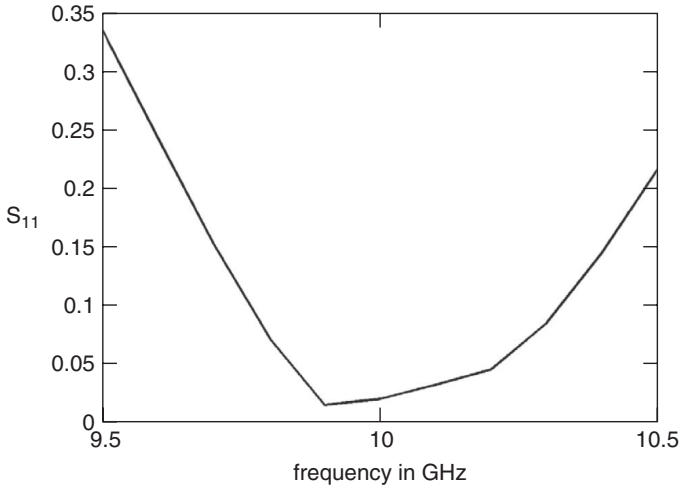


Figure 32.21 Plot of s_{11} of the optimized patch as a function of frequency.

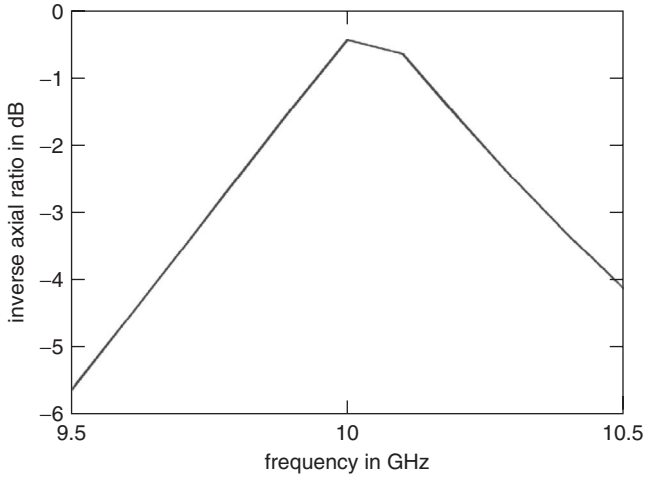


Figure 32.22 Plot of the inverse axial ratio of the optimized patch as a function of frequency.

in order to get the desired performance characteristics [59]. They used a method of moments model and supported the results with experimental measurements. Figure 32.23 is an example of how the "crooked antenna" is constructed from straight wire segments. The number of wire segments, the segment lengths, and the volume that contains the antenna are constraints. A voltage source feeds the antenna at $(x,y,z) = (0,0,0)$ and the antenna is situated above an infinite ground plane. Chromosomes contain the end points of each wire segment.

A binary GA was used to do the optimization with each dimension represented by five bits, making a chromosome 105 bits long [60]. Each side of the volume defining the antenna is 0.5λ long and each of the seven segments are long. The goal was to have

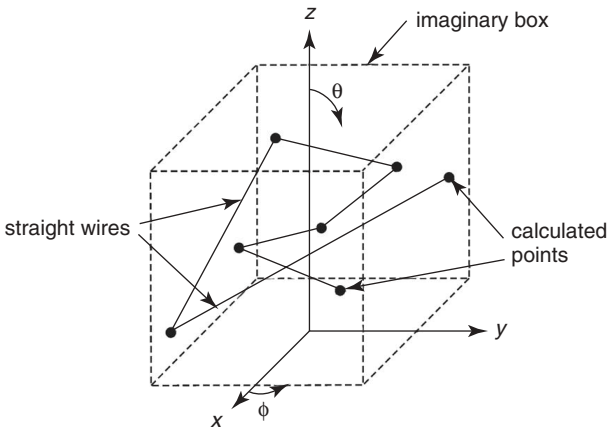


Figure 32.23 Model of crooked wire antenna used for cost function.

circular polarization from $-80^\circ \leq \theta \leq 80^\circ$ for all values of ϕ. At the design frequency of 1600 MHz, the cost was formulated as

$$cost = \sum [gain(\theta, \phi) - \overline{gain}]^2 \tag{32.16}$$

where the overbar represents average. When the GA finished, it found the wire segment end points given in Table 32.2 [60]. A diagram of this crooked wire antenna appears in Figure 32.24. The end points from Table 32.2 were entered into a method of moments program, then the gain patterns were calculated. Gain pattern results for three ϕ cuts are shown in Figure 32.25. This genetic antenna has a relatively flat gain response for circular polarization across the pattern cuts shown.

There are many different wire antennas that can be optimized by a GA. For instance, Yagi arrays have been optimized in Refs. 60–64. A combination of Yagi arrays and log-periodic antennas were optimized in Ref. 65. Finally, variations of log-periodic antennas were optimized in Ref. 66. Inputs to these antenna cost functions include length of wires and spacing between wires. Costs include gain, sidelobe level, and VSWR. The GA has created new designs that would not have been thought of from traditional design approaches.

TABLE 32.2 Optimized Locations of the Wire Segment End Points

x (cm)	y (cm)	z (cm)
0.00	0.00	0.00
−1.66	0.45	7.14
−3.18	−1.66	1.70
−3.18	−2.87	7.75
−3.18	4.39	1.40
−3.18	0.45	6.24
−1.06	3.78	8.66
−1.06	2.57	2.30

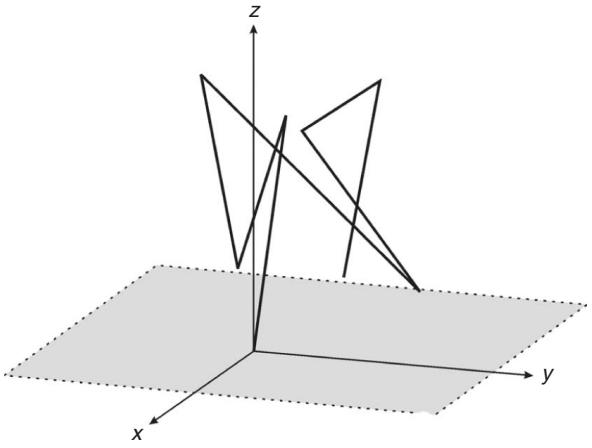


Figure 32.24 Optimum crooked wire antenna design.

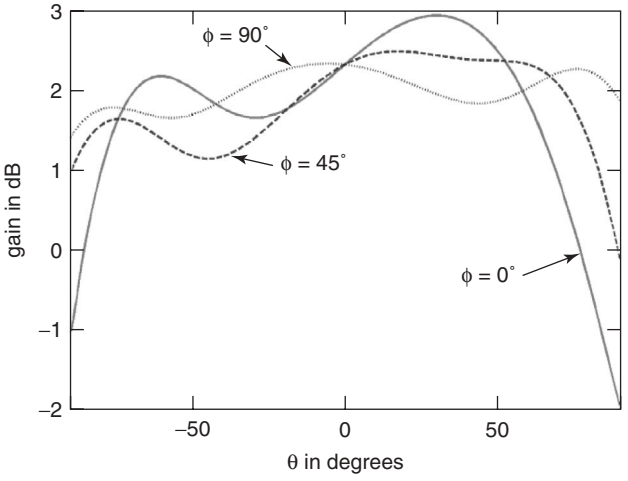


Figure 32.25 Gain patterns at three different pattern cuts for left-hand circular polarization.

32.5 SMART ANTENNAS

Adaptive antennas place nulls in the antenna pattern in the directions of the interference while minimally perturbing the main beam. A sidelobe canceler [67] adjusts the output of an omnidirectional antenna until its output cancels the interference entering a sidelobe of the high gain antenna. Since the omnidirectional antenna has a low gain compared to the main beam of the high gain antenna, the main beam was not affected much. This concept was generalized to arrays by adjusting the weights of the elements in the array to receive the desired signal and cancel the sidelobe interference [68]. Most adaptive algorithms need a receiver at each element in order to detect the amplitude and phase of the signals at each element in order to form the covariance matrix. Receivers

are expensive and require frequent calibration, so adaptive nulling is somewhat of a luxury.

A totally different approach to adaptive nulling has the goal of minimizing the total output power of the array [69]. Since the array output is the sum of the desired signal and interference signals, some insurance is needed to null only the sidelobes and not the main beam. One way to constrain the nulling is to use only the least significant bits (LSBs) of the amplitude and phase weights or only a fraction of the elements in the array. In this way, the signal strength associated with the LSBs is too small to create a null in the main beam. The concept of using LSBs is similar to that of the omnidirectional antenna in the sidelobe canceler. A binary GA is ideal, because most adjustable amplitude and phase weights have binary controls. Results have previously been reported for a computer model of phase-only adaptive nulling with linear arrays [70] and experimental results for amplitude and phase adaptive nulling with an eight-element conformal array [71].

The adaptive array consists of elements lying along the x-axis (Figure 32.26) and has a corresponding array factor given by

$$AF = \sum_{n=1}^{N} w_n e^{jk(n-1)d_0 \cos\phi} \tag{32.17}$$

Controlling the weights modifies the main beam peak and nulls. Usually, the array weights are digital and have a finite number of settings. The desired signal has an amplitude of one and is incident at boresight.

The problem with this cost function formulation is that the desired signal and the interfering signals are mixed together. Minimizing the output power decreases the desired signal in addition to the interfering signals unless we assume the desired signal enters the main beam and the adaptive weights are constrained to small values that cannot place a null in the main beam. Each binary chromosome in the population contains a few of the LSBs of the phase and/or amplitude settings at each element in the array. The goal

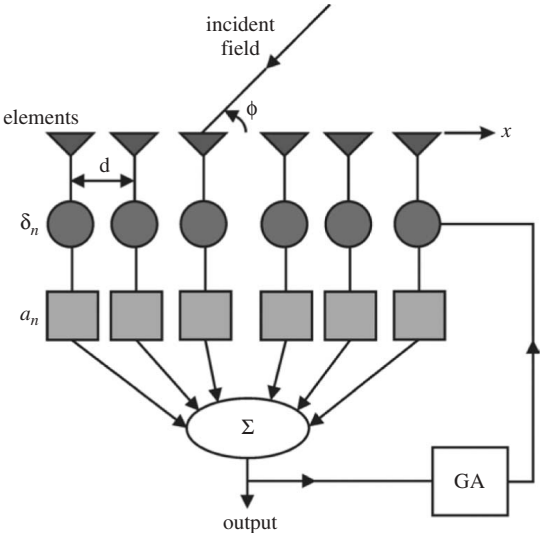


Figure 32.26 Diagram of an adaptive array controlled by a GA.

of the GA is to minimize the total output power of the antenna by adjusting these array settings. Since the algorithm must be fast and a global minimum is not necessary, the GA uses a small population size.

If a six element uniform array has six-bit phase shifters, then the bits correspond to the six phase settings in Table 32.3. The LSBs have small phase values that produce small changes to the main beam. Assume that the six-element array is symmetric about its center and the center two elements have zero phase. The phase on the right side of the array is the negative phase of the phase on the left side of the array. Figure 32.27 shows the array output from the main beam ($\phi = 90^\circ$) and a sidelobe ($\phi = 45^\circ$) as a function of phase shifter settings when up to four out of six bits are used. When one LSB is used, then the middle and edge elements each have one bit to vary for a total of four different combinations. When those same elements have two bits, then there are 16 total combinations. Figure 32.27 shows from one to four LSBs, so at four LSBs, there are 256 total combinations possible. At least four bits are needed to null the sidelobe. If the array had low sidelobes, then fewer bits would be needed to place a null in the sidelobe. Using up to four LSBs will not significantly perturb the main beam. On the other hand, five or six bits can place a null in the main beam.

All the adaptive nulling examples in this section are demonstrated on a 20-element array of isotropic point sources spaced 0.5λ apart. The elements have six-bit amplitude and phase weights. The array has a 20-dB, $\overline{n} = 3$ Taylor amplitude taper. The 0-dB desired signal is incident on the peak of the main beam at $\phi = 90^\circ$.

In the first example, the array uses two LSBs of the amplitude weights and three of the phase weights to deal with two 30-dB interference sources at 111° and 117°. The GA has a population size of 8, uses roulette wheel selection with a 50% selection rate, uses uniform

TABLE 32.3 Phase Associated with Each Bit of a Six-Bit Phase Shifter

Bit	1	2	3	4	5	6
Phase in radians	$\frac{\pi}{32}$	$\frac{\pi}{16}$	$\frac{\pi}{8}$	$\frac{\pi}{4}$	$\frac{\pi}{2}$	π

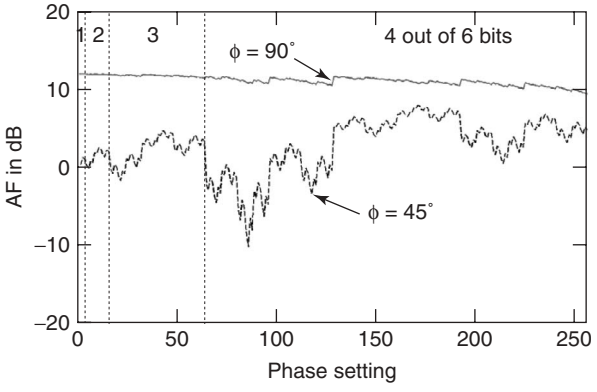


Figure 32.27 Array factor at the main beam ($\phi = 90^\circ$) and a sidelobe ($\phi = 45^\circ$) when all combinations of one through four LSBs out of a total of six bits in the phase shifters are tried.

crossover, and mutates 10% of the chromosomes. The adapted pattern is the solid line and the quiescent pattern is the dashed line in Figure 32.28. The nulls were successfully placed in just a short time (Figure 32.29) with little perturbation to the antenna pattern. Convergence took place in only $8 + 17 \times 7 = 127$ power measurements. Even though the output power decreases monotonically, the individual interference source received power varies. The power received from one interference source will change relative to the power received from the second interference source. The desired signal does not change much, because the main beam remains virtually unperturbed.

A second example of amplitude and phase nulling has two 30-dB interference signals at 50° and 130°. In this case, four amplitude and three phase LSBs were used (more bits were needed to place these nulls). Figure 32.30 shows the adapted pattern. This time, sidelobe distortion is more pronounced but the main beam shows little change.

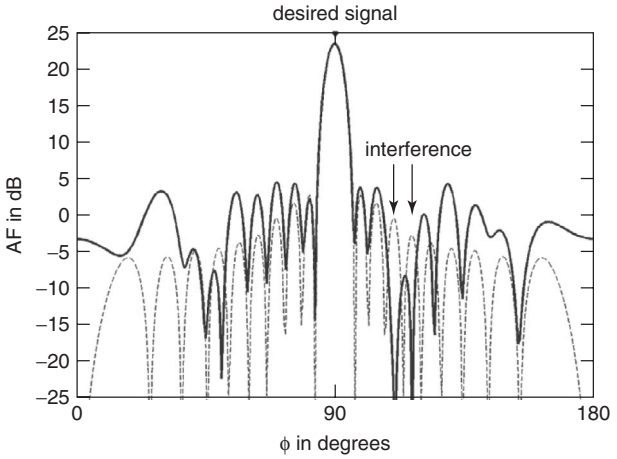


Figure 32.28 The solid line is the adapted array factor and the dashed line is the quiescent array factor.

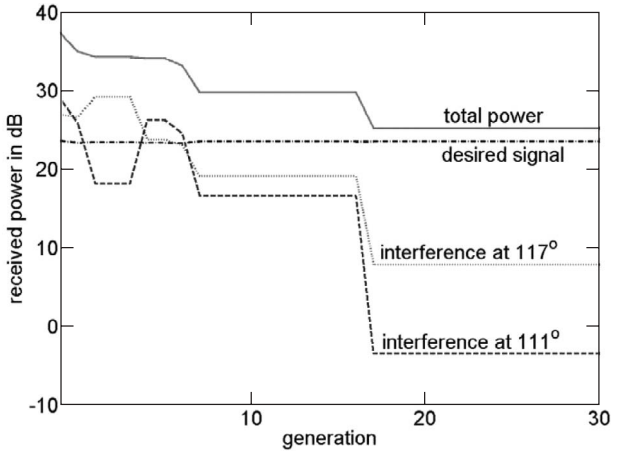


Figure 32.29 GA convergence for amplitude and phase adaptive nulling.

The previous two scenarios were repeated using only the phase weights. All the amplitude weights are equal to one. Three LSBs from the six-bit phase shifters were used to place nulls at 111° and 117°. Figure 32.31 shows the adapted array factor with little main beam perturbation but a significant increase in some sidelobes. The phase-only convergence in Figure 32.32 is not as good as with the amplitude and phase adaptive nulling. After only 18 generations, the algorithm found two deep nulls in the sidelobes.

Phase-only adaptive nulling with interference sources at symmetric angles about the main beam is more difficult than for amplitude and phase nulling. Only one null can be placed by using three least significant phase bits. A fourth bit was added to produce the two nulls shown in Figure 32.33. The nulls were placed at the cost of reducing the main

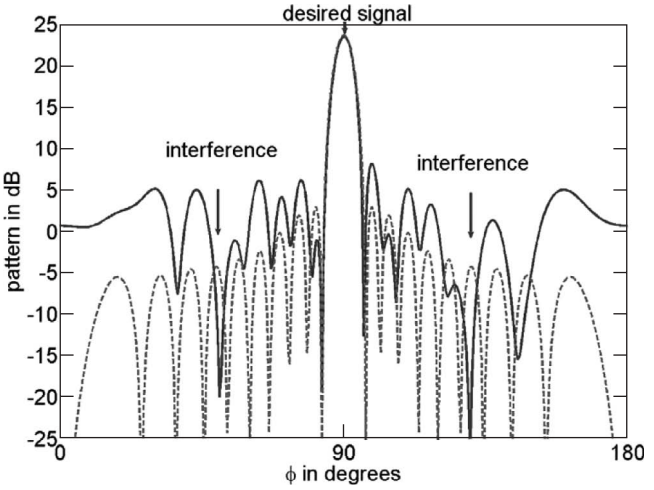


Figure 32.30 Adapted pattern when there are two symmetric interference sources.

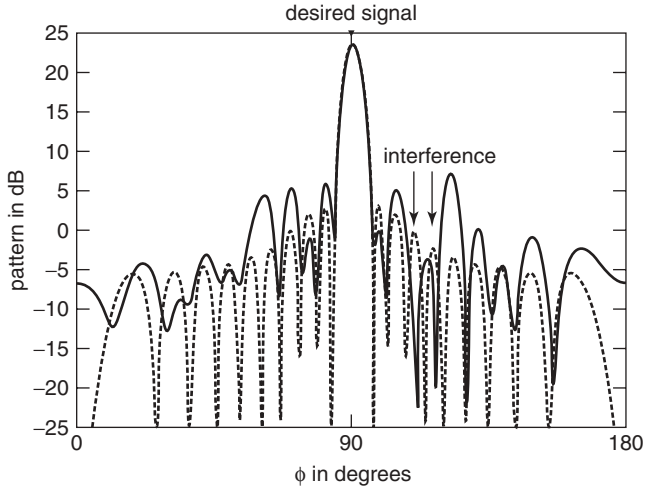


Figure 32.31 The solid line is the adapted array factor and the dashed line is the quiescent array factor.

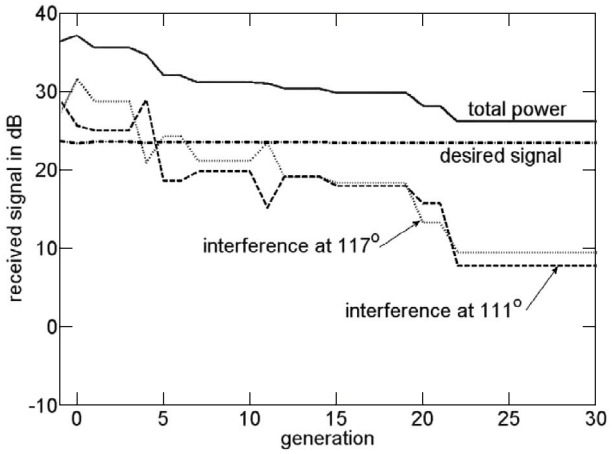


Figure 32.32 GA convergence for phase-only adaptive nulling.

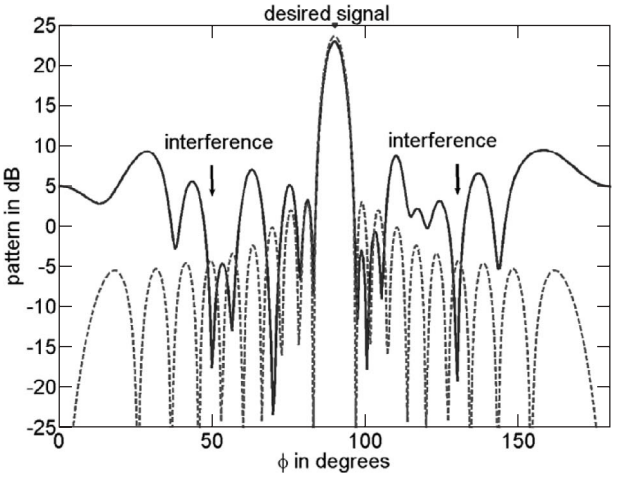


Figure 32.33 Adapted pattern for phase-only nulling when there are two symmetric interference sources.

beam gain. Nulling symmetric interference sources is a general problem with phase-only nulling that can only be rectified with large phase shifts.

32.6 OTHER GA ANTENNA DESIGNS

This chapter gave a brief overview of antenna optimization using a GA. Many topics were skipped due to space constraints. Some good references for other antenna topics include Refs. 72 and 73.

Future applications of GAs to antennas are highly dependent on antenna modeling tools and advances in computing power. GAs are a powerful optimization tool that can

create new designs, but computers and software models are too slow for the cost function evaluations. There are three ways to get the GA to converge faster:

1. *Use Faster GAs:* Finding the best GA parameters such as population size and mutation rate can make orders of magnitude difference in the number of function calls needed to find an acceptable solution [24]. Adapting the parameters as the GA progresses has shown some success. Hybrid approaches that combine the GA with local optimizers or other global techniques can combine the advantages of different algorithms. Human interaction with the GA and cost function might produce better and faster results. Subjective cost functions have been used in art and music, why not antenna design [74]?
2. *Use Faster Cost Functions:* Starting with fast, approximate function evaluations in early generations and converting to slower, increasingly accurate function evaluations in later generations has some promise. Efficient hybrid methods can be used in the cost function. Eliminating variables that have little effect on the outcome encourages the optimization algorithms to converge faster. Adaptively adjusting the grid size can result in significant time savings for many problems.
3. *Use Faster Computers:* Not only are CPU time and memory important, but computer architectures are also important. GAs are highly parallel in nature, so they can be fast on multiprocessor machines.

REFERENCES

1. N. Metropolis et al., Equation of state calculations by fast computing machines, *J. Chem. Phys.*, Vol. 21, pp. 1087–1092, 1953.
2. S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, Optimization by simulated annealing, *Science*, Vol. 220, pp. 671–680, 13 May 1983.
3. J. Kennedy and R. C. Eberhart, Particle swarm optimization, in *Proceedings IEEE International Conference on Neural Networks, IV*, IEEE Service Center, Piscataway, NJ, 1995, pp. 1942–1948.
4. J. H. Holland, *Adaptation in Natural and Artificial Systems*, The University of Michigan Press, Ann Arbor, 1975.
5. D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, Boston, 1989.
6. J. D. Schaffer, Some experiments in machine learning using vector evaluated genetic algorithms, Ph.D. dissertation, Vanderbilt University, Nashville, TN, 1984.
7. D. E. Goldberg and J. Richardson, Genetic algorithms with sharing for multi-modal function oprimixation, in *Genetic Algorithms and Their Applications: Proceedings of the 2nd Intternational Conference on Genetic Algorithms*, J. J. Grefenstette (Ed.), Lawrence Erlbaum, Hillsdale, NJ, 1987, pp. 41–49.
8. C. M. Foneseca and P. J. Flemming, Genetic algorithms for multi-objective optimization: formulation, discussion, and generalization, in *Proceedings of the 5th International Conference on Genetic Algorithms*, Morgan Kaufmann, San Francisco, CA, 1993, pp. 416–423.
9. K. Deb et.al, A fast and elitist multiobjective genetic algorithm: NSGA-II, *IEEE Trans. Evol. Comput.*, Vol. 6, No. 2, pp. 182–197, 2002.
10. D. S. Weile, E. Michielssen, and D. E. Goldberg, Genetic algorithm design of Pareto optimal broadband microwave absorbers, *IEEE AP-S Trans.*, Vol. 38, No. 3, p. 525, August 1996.

11. D. S. Weile and E. Michielssen, Integer coded Pareto genetic algorithm design of constrained antenna arrays, *Electron. Lett.*, Vol. 32, No. 19, pp. 1744–1745, 12 September 1996.
12. M. Farina, A. Bramanti, and P. Di Barba, GRS method for Pareto-optimal front identification in electromagnetic synthesis, *IEE Proc. Sci. Meas. Technol.*, Vol. 149, No. 5, pp. 207–213, September 2002.
13. H. Choo and H. Ling, Design of electrically small planar antennas using inductively coupled feed, *Electron. Lett.*, Vol. 39, No. 22, pp. 1563–1564, 30 October 2003.
14. N. V. Venkatarayalu and T. Ray, Optimum design of Yagi–Uda antennas using computational intelligence, *IEEE AP-S Trans.*, Vol. 52, No. 7, pp. 1811–1818, July 2004.
15. K. A. De Jong, Analysis of the behavior of a class of genetic adaptive systems, Ph.D. dissertation, The University of Michigan, Ann Arbor, 1975.
16. J. J. Grefenstette, Optimization of control parameters for genetic algorithms, *IEEE Trans. Systems Man and Cybernetics*, Vol. 16, p. 128, January/February 1986.
17. J. D. Schaffer, et al., A study of control parameters affecting online performance of genetic algorithms for function optimization, in *Proceedings of the Third International Conference on Genetic Algorithms*, J. D. Schaffer (Ed.), Morgan Kaufmann, San Francisco, 1989, pp. 51–60.
18. T. Back, Optimal mutation rates in genetic search, in *Proceedings of the 5th International Conference on Genetic Algorithms*, S. Forrest (Ed.), Morgan Kaufmann, San Francisco, 1993, pp. 2–9.
19. Y. Gao, An upper bound on the convergence rates of canonical genetic algorithms, *Complexity Int.*, Vol. 5, 1998.
20. T. Back, and M. Schutz, Intelligent mutation rate control in canonical genetic algorithms, in *Foundations of Intelligent Systems 9th International Symposium*, Z. W. Ras and M. Michalewicz (Eds.), Springer Verlag, Berlin, 1996, pp. 158–167.
21. T. Back, Evolution strategies: an alternative evolutionary algorithm, in *Artificial Evolution*, J. M. Alliot et al. (Eds.), Springer Verlag, Berlin, 1996, pp. 3–20.
22. L. Davis, Parameterizing a genetic algorithm, *Handbook of Genetic Algorithms*, in L. Davis (Ed.), Van Nostrand Reinhold, New York, 1991.
23. R. L. Haupt and S. E. Haupt, Optimum population size and mutation rate for a simple real genetic algorithm that optimizes array factors, *Appl. Computat. Electromagn. Soc. J.*, Vol. 15, No. 2, pp. 94–102, July 2000.
24. R. L. Haupt and S. E. Haupt, *Practical Genetic Algorithms*, 2nd ed, John Wiley & Sons, Hoboken, NJ, 2004.
25. R. L. Haupt, Genetic algorithm applications for phased arrays, *Appl. Computat. Electromagn. Soc. J.*, Vol. 21, No. 2, July 2006.
26. R. L. Haupt, Thinned arrays using genetic algorithms, *IEEE AP-S Trans.*, Vol. 42, pp. 993–999, July 1994.
27. F. Ares et al., Application of genetic algorithms and simulated annealing technique in optimising the aperture distributions of antenna array patterns, *Electron. Lett.*, Vol. 32, pp. 148–149, 1 February 1996.
28. R. J. Mitchell, B. Chambers, and A. P. Anderson, Array pattern synthesis in the complex plane optimised by a genetic algorithm, *Electron. Lett.*, Vol. 32, pp. 1843–1845, 26 September 1996.
29. Y. Keen-Keong and L. Yilong, Sidelobe reduction in array-pattern synthesis using genetic algorithm, *IEEE AP-S Trans.*, Vol. 45, pp. 1117–1122, July 1997.
30. J. M. Johnson and Y. Rahmat-Samii, Genetic algorithms in engineering electromagnetics, *IEEE AP-S Mag.*, Vol. 39, pp. 7–21, August 1997.
31. F. J. Ares-Pena, J. A. Rodriguez-Gonzalez, E. Villanueva-Lopez, and S. R. Rengarajan, Genetic algorithms in the design and optimization of antenna array patterns, *IEEE AP-S Trans.*, Vol. 47, pp. 506–510, March 1999.

32. B. J. Barbisch D. H. Werner, and P. L. Werner, A genetic algorithm optimization procedure for the design of uniformly excited and nonuniformly spaced broadband low sidelobe arrays, *Appl. Computat. Electromagn. Soc. J.*, Vol. 15, No. 2, pp. 34–42, July 2000.
33. S. Mummareddy, D. H. Werner, and P. L. Werner, Genetic optimization of fractal dipole antenna arrays for compact size and improved impedance performance over scan angle, *IEEE AP-S Int. Symp.*, pp. 98–101, June 2002.
34. M. G. Bray et al., Optimization of thinned aperiodic linear phased arrays using genetic algorithms to reduce grating lobes during scanning, *IEEE AP-S Trans.*, Vol. 50, pp. 1732–1742, December 2002.
35. R. J. Allard, D. H. Werner, and P. L. Werner, Radiation pattern synthesis for arrays of conformal antennas mounted on arbitrarily-shaped three-dimensional platforms using genetic algorithms, *IEEE AP-S Trans.*, Vol. 51, pp. 1054–1062, May 2003.
36. R. Haupt, Generating a plane wave in the near field with a planar array antenna, *Microwave J.* August 2003.
37. R. L. Haupt, Interleaved thinned linear arrays, *IEEE AP-S Trans.*, Vol. 53, pp. 2858–2864, September 2005.
38. R. L. Haupt, An introduction to genetic algorithms for electromagnetics, *IEEE AP-S Mag*., Vol. 37, pp. 7–15, April 1995.
39. R. L. Haupt, Optimum quantised low sidelobe phase tapers for arrays, *Electron. Lett.*, Vol. 31, pp. 1117–1118, 6 July 1995.
40. R. Haupt, Optimization of subarray amplitude tapers, *IEEE AP-S Int. Symp.*, pp. 1830–1833, June 1995.
41. R. J. Mailloux, *Phased Array Antenna Handbook*, Artech House, Boston, 2005.
42. L. Chung, T. Chang, and W. D. Burnside, An ultrawide-bandwidth tapered resistive TEM horn antenna, *IEEE AP-S Trans.*, Vol. 52, No. 2, pp. 1848–1857, February 2004.
43. P. L. Garcia-Muller, Optimisation of compact horn with broad sectoral radiation pattern, *Electron. Lett.*, Vol. 37, No. 6, pp. 337–338, 15 March 2001.
44. D. Yang, Y. C. Chung, and R. L. Haupt, Genetic algorithm optimization of a corrugated conical horn antenna, *IEEE AP-S Symp.*, Vol. 1, pp. 342–345, June 2002.
45. L. Lucci, R. Nesti, G. Pelosi and S. Selleri, Optimization of profiled corrugated circular horns with parallel genetic algorithms, *IEEE AP-S Symp.*, Vol. 2, pp. 338–341, June 2002.
46. L. Lucci, R. Nesti, G. Pelosi, and S. Selleri, NURBS profiled corrugated circular horns, *IEEE AP-S Symp.*, Vol. 1, pp. 161–164, June 2003.
47. S. Sinton, J. Robinson, and Y. Rahmat-Samii, Standard and micro genetic algorithm optimization of profiled corrugated horn antennas, *Microwave Optical Technol. Lett.*, Vol. 35, No. 6, pp. 449–453, 20 December 2002.
48. R. Haupt, A horn-fed reflector optimized with a genetic algorithm, *2005 ACES Conference*, Honolulu, HI, April 2005.
49. M. Vall-llossera, J. M. Rius, N. Duffo, and A. Cardama, Design of single-shaped reflector antennas for the synthesis of shaped contour beams using genetic algorithms, *Microwave Optical Technol. Lett.*, Vol. 27, No. 5, pp. 358–361, 5 December 2000.
50. Y. Lu, X. Cai, and Z. Gao, Optimal design of special corner reflector antennas by the real-coded genetic algorithm, in *Asia-Pacific Microwave Conference*, December 2000, pp. 1457–1460.
51. K. Barkeshli, F. Mazlumi and R. Azadegan, The synthesis of offset dual reflector antennas by genetic algorithms, *IEEE AP-S Symp.*, Vol. 1, pp. 670–673, June 2002.
52. S. L. Avila, W. P. Carpes, and J. A. Vasconcelos, Optimization of an offset reflector antenna using genetic algorithms, *IEEE Trans. Magnetics*, Vol. 40, No. 2, pp. 1256–1259, March 2004.
53. J. M. Johnson and Y. Rahmat-Samii, Genetic algorithms in engineering electromagnetics, *IEEE AP-S Mag.*, Vol. 39, No. 4, pp. 7–21, August 1997.

54. C. Delabie, M. Villegas, and O. Picon, Creation of new shapes for resonant microstrip structures by means of genetic algorithms, *Electron. Lett.*, Vol. 33, No. 18, pp. 1509–1510, August 1997.
55. J. M. Johnson, and Y. Rahmat-Samii, Genetic algorithms and method of moments (GA/MOM) for the design of integrated antennas, *IEEE Trans. Antennas Propag.*, Vol. 47, No. 10, pp. 1606–1614, October 1999.
56. A. Mitchell, M. Lech, D. M. Kokotoff, and R. B. Waterhouse, Search for high-performance probe-fed stacked patches using optimization, *IEEE Trans. Antennas Propag.*, Vol. 51, No. 2, pp. 249–255, February 2003.
57. H. Choo and H. Ling, Design of broadband and dual-band microstrip antennas on a high-dielectric substrate using a genetic algorithm, *IEE Proc Microwaves Antennas Propag.*, Vol. 150, No. 3, pp. 137–142, June 2003.
58. F. J. Villegas, T. Cwik, Y. Rahmat-Samii, and M. Manteghi, A parallel electromagnetic genetic-algorithm optimization (EGO) application for patch antenna design, *IEEE Trans. Antennas Propag.*, Vol. 52, No. 9, pp. 2424–2435, September 2004.
59. E. E. Altshuler and D. S. Linden, A process for the design of antennas using genetic algorithms, U.S. Patent No. 5,719,794, 17 February 1998.
60. E. Altshuler and D. Linden, Wire-antenna designs using genetic algorithms, *IEEE Antennas Propag. Mag.*, Vol. 39, No. 2, pp. 33–43, April 1997.
61. E. A. Jones and W. T. Joines, Design of Yagi–Uda antennas using genetic algorithms, *IEEE AP Trans.*, Vol. 45, No. 9, pp. 1386–1392, September 1997.
62. N. V. Venkatarayalu and T. Ray, Optimum design of Yagi–Uda antennas using computational intelligence, *IEEE AP Trans.*, Vol. 52, No. 7, pp. 1811–1818, July 2004.
63. Y. Kuwahara, Multiobjective optimization design of Yagi–Uda antenna, *IEEE AP Trans.*, Vol. 53, No. 6, pp. 1984–1992, June 2005.
64. P. K. Varlamos, P. J. Papakanellos, S. C. Panagiotou, and C. N. Capsalis, Multi-objective genetic optimization of Yagi–Uda arrays with additional parasitic elements, *IEEE AP-S Mag.*, Vol. 47, No. 4, pp. 92–97, August 2005.
65. E. A. Jones and W. T. Joines, Genetic design of linear antenna arrays, *IEEE AP-S Mag.*, Vol. 42, No. 3, pp. 92–100, June 2000.
66. Y. C. Chung and R. L. Haupt, Log-periodic dipole array optimization, *J. Electromagn. Waves Applications*, Vol. 15, No. 9, pp. 1269–1280, September 2001.
67. S. P. Applebaum, Adaptive arrays, Syracuse University Research Corporation Report SPL TR 66–1, August 1966.
68. B. Widrow et al., Adaptive antenna systems, *IEEE Proc.*, Vol. 55, No. 12, pp. 2143–2159, December 1967.
69. C. A. Baird and G. G. Rassweiler, Adaptive sidelobe nulling using digitally controlled phase-shifters, *IEEE AP Trans.*, Vol. 24, No. 5, pp. 638–649, September 1976.
70. R. L. Haupt, Phase-only adaptive nulling with genetic algorithms, *IEEE AP-S Trans.*, Vol. 45, No. 5, pp. 1009–1015, June 1997.
71. R. L. Haupt and H. L. Southall, Experimental adaptive nulling with a genetic algorithm, *Microwave J.*, Vol. 42, No. 1, pp. 78–89, January 1999.
72. Y. Rahmat-Samii and E. Michielssen (Eds.), *Electromagnetic Optimization by Genetic Algorithms*, John Wiley & Sons, Hoboken, NJ, 1999.
73. R. L. Haupt and D. H. Werner, *Genetic Algorithms in Electromagnetics*, John Wiley & Sons, Hoboken, NJ, 2007.
74. R. L Haupt and S. E. Haupt, The creative use of genetic algorithms, *IEEE Potentials*, Vol. 19, No. 2, pp. 26–29, April 2000.

CHAPTER 33

Neural Networks for Antennas

CHRISTOS G. CHRISTODOULOU and AMALENDU PATNAIK

33.1 INTRODUCTION

In the last twenty years neural computing has experienced an unbelievable growth, both in terms of novel neural network (NN) architectures that have appeared in the literature, and new applications where neural networks have been used successfully. Most of the neural network architectures that have been introduced in the last twenty years share the common advantages of neural network computing: that is, model-free design, generalization capabilities, inherent parallelism, graceful performance degradation in the presence of noise, portability, and so on.

Neural networks, also called artificial neural networks (ANNs) are considered as grossly simplified models of the human brain. An NN is a computational tool that learns from experience (training), generalizes from previous examples to new ones, and abstracts essential characteristics from input containing irrelevant data. Although ANNs can serve to further the understanding of brain functions, engineers are interested in ANNs for problem solving. Because of their massively parallel nature, ANNs can perform computations at high speed and because of their adaptive nature, ANNs can adapt to changes in data and learn the characteristics of input signals.

Furthermore, due to their nonlinear nature, they can perform functional approximation and signal filtering operations, which are beyond optimal linear techniques. So any practical problem that can be put either in nonlinear mapping formation form, function approximation form, or data classification form can be implemented using ANNs.

The art of nonlinear black-box system modeling in engineering using neural network structures is gradually gaining momentum. In the last two decades, ANN technology has leaped forward and is now widely applied to a variety of practical problems such as signal processing, speech processing, and controls systems, analysis and optimization of microwave circuits, and intermodulation and power analysis. The distinguished characteristics of ANNs such as learning from data, to generalizing patterns in data, and to modeling nonlinear relationships makes them a good candidate for many antenna engineering applications. From a user's viewpoint, the key problem is to find a suitable model structure, within which a good model is to be found. Fitting a model within a given structure is then, in most cases, a lesser problem.

Modern Antenna Handbook. Edited by Constantine A. Balanis

In this chapter we present the basics of neural networks and how they can be applied to antenna problems.

33.2 NEURAL NETWORK CONCEPTS

We describe neural network (NN) structural issues to better understand what NNs are and why they have the ability to represent antenna behaviors. We study NNs from an external input–output point of view, and also from an internal neuron information processing point of view. The most popularly used NN structure—the multilayer perceptron (MLP)—is described in detail. The effects of structural issues on modeling accuracy are discussed.

33.2.1 Basic Components

A typical NN structure has two types of basic components, namely, the processing elements and the interconnections between them. The processing elements are called *neurons* and the connections between the neurons are known as *links* or *synapses*, as shown in Figure 33.1. Every link has a corresponding *weight* parameter associated with it. Each neuron receives stimulus from other neurons connected to it, processes the information, and produces an output. Neurons that receive stimuli from outside the network are called *input neurons*, while neurons whose outputs are externally used are called *output neurons* (see Figure 33.2). Neurons that receive stimuli from other neurons and whose outputs are stimuli for other neurons in the network are known as *hidden neurons*. Different NN structures can be constructed by using different types of neurons and by connecting them differently.

33.2.2 Concept of a Neural Network Model

Let n and m represent the number of input and output neurons of an NN as shown in Figure 33.2. Let $\mathbf{x}$ be an n-vector containing the external inputs to the NN, let $\mathbf{y}$ be an m-vector containing the outputs from the output neurons, and let $\mathbf{w}$ be a vector containing all the weight parameters representing various interconnections in the NN. The definition

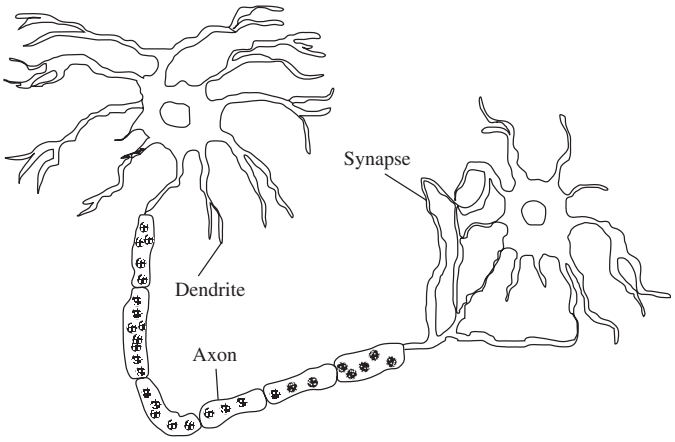


Figure 33.1 A simple neuron and its connection.

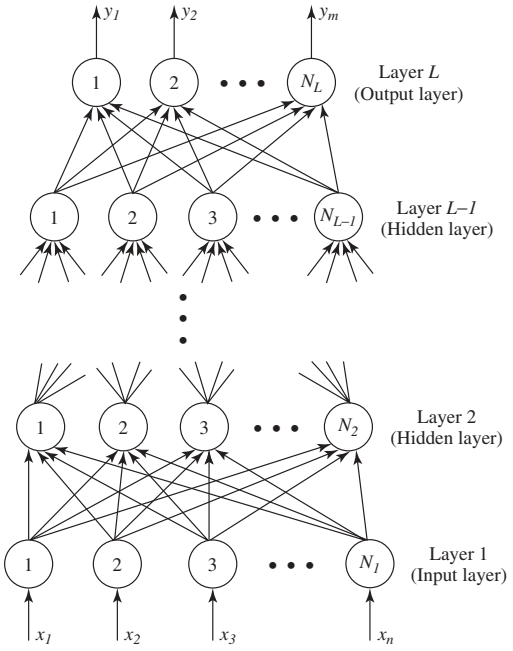


Figure 33.2 MLP neural network structure. Typically, an MLP network consists of an input layer, one or more hidden layers, and an output layer. (From Ref. 8. Reprinted with permission, copyright © 2003 IEEE.)

of **w** and the manner in which **y** is computed from **x** and **w** determine the structure of the NN.

Let us consider a circular microstrip antenna for an example. The dimensions and the construction parameters of the microstrip antenna are variables and any change in the values of these parameters affects the resonant frequency of the microstrip antenna. Assume that there is a need to develop a neural model that can represent such an input–output relationship. Inputs and outputs of the corresponding microstrip antenna neural model are given by

$$\mathbf{x} = [\varepsilon_r, h, r]^T \tag{33.1}$$

$$\mathbf{y} = [f_r]^T \tag{33.2}$$

where f_r is frequency, and ε_r, h, and r represent the dielectric constant, height of the substrate and radius of the radiating patch, respectively. The superscript T indicates the transpose of a vector or matrix. So we can express the original electromagnetics (EM) based microstrip antenna modeling problem as

$$\mathbf{y} = f(\mathbf{x}) \tag{33.3}$$

where f represents a detailed EM-based input–output relationship. The NN model (mapping problem) for the microstrip antenna is given by

$$\mathbf{y} = f(\mathbf{x}, \mathbf{w}) \tag{33.4}$$

Several ($\mathbf{x}$, $\mathbf{y}$) samples called *training data* need to be generated either from the microstrip antenna EM simulator or from actual measured data. The objective of training is to adjust the NN weights $\mathbf{w}$ such that the neural model outputs best match the training data outputs (i.e., mapping the input to the output data). A trained neural model can be used during the antenna design process to provide instant answers to the task it has learned. In the microstrip antenna case, the neural model can be used to provide fast estimation of the resonant frequency against the antenna's dimensions and construction parameter values.

33.2.3 Neural Network Versus Conventional Modeling

Next, the NN approach is compared with conventional approaches for a better understanding. The first approach is the detailed modeling approach, that is, the electromagnetic, full-wave based modeling, where the model is defined by a well established theory. The detailed models are accurate but could be computationally time consuming. The second approach is an approximate modeling approach, which uses either empirical or equivalent-circuit-based models for antennas. These models are developed using a mixture of simplified component theory, heuristic interpretation and representation, and/or fitting of experimental data. Evaluation of approximate models is much faster than that of the full-wave EM models. However, the models are limited in terms of accuracy and input parameter range over which they can be accurate. The NN approach is a new type of modeling approach where the model can be developed by learning from detailed (accurate) data of the antenna. After training, the NN becomes a fast and accurate model representing the original component behaviors. The following are some of the situations where use of the NN technique is advisable:

- When closed form solutions do not exist and trial-and-error methods are the only approaches to solving the problem at hand. By training an NN one can use it to predict solutions to the problem.
- When the application requires real-time performance. This is an important feature because NNs can be trained off-line and then implemented in hardware and embedded on any device.
- When faster convergence rates and smaller errors are required in the optimization of large systems.
- When enough measured data exist to train an NN for prediction purposes, especially when no analytical tools exist.

33.2.4 MLP Neural Network

33.2.4.1 Structure and Notation MLP is the simplest and most widely used NN structure in electromagnetics. In the MLP neural network, the neurons are grouped into layers. The first and last layers are called input and output layers, respectively, and the remaining layers are called hidden layers. Typically, an MLP neural network consists of an input layer, one or more hidden layers, and an output layer, as shown in Figure 33.2. For example, an MLP neural network with an input layer, one hidden layer, and an output layer is referred to as a three-layer MLP.

Suppose the total number of layers is L. The first layer is the input layer, the Lth layer is the output layer, and layers 2 to $L-1$ are hidden layers. Let the number of neurons in the lth layer be N_l, $l = 1, 2, \ldots, L$. Let w^l_{ij} represent the weight of the link between the

jth neuron of the $(l-1)$th layer and the ith neuron of the lth layer. Let x_i represent the ith external input to the MLP and $z^l{}_i$ be the output of the ith neuron of the lth layer. There is an additional weight parameter for each neuron (w^l_{i0}) representing the bias for the ith neuron of the lth layer. As such, the vector $\mathbf{w}$ of the MLP includes w^l_{ij}, $j = 0, 1, \ldots, N_{l-1}$, $i = 1, 2, \ldots, N_l$, and $l = 2, 3, \ldots, L$; that is $\mathbf{w} = [w^2_{10} w^2_{11} w^2_{12} \cdots w^L_{N_L N_{L-1}}]^T$. The parameters in the weight vector are real numbers, which are initialized before the MLP training. During training, they are changed (updated) iteratively in a systematic manner [1, 2]. Once the NN training is completed, the vector $\mathbf{w}$ remains fixed throughout the usage of the NN as a model.

33.2.4.2 Anatomy of Neurons In the MLP network, each neuron processes the stimuli (inputs) received from other neurons. The process is done through a function called the activation function in the neuron, and the processed information becomes the output of the neuron. For example, every neuron in the lth layer receives stimuli from the neurons of the $(l-1)$th layer, that is, $z^{l-1}_1, z^{l-1}_2, \ldots, z^{l-1}_{N_{l-1}}$. A typical ith neuron in the lth layer processes this information in two steps. First, each of the inputs is multiplied by the corresponding weight parameter and the products are added to produce a weighted sum γ^l_i:

$$\gamma^l_i = \sum_{j=0}^{N_{l-1}} w^l_{ij} z^{l-1}_j \tag{33.5}$$

In order to create the effect of the bias parameter w^l_{i0}, we assume a fictitious neuron in the $(l-1)$th layer whose output is $z^{l-1}_0 = 1$. Second, the weighted sum in Eq. (33.5) is used to activate the neuron's activation function $\sigma(\cdot)$ to produce the final output of the neuron $z^l = \sigma(\gamma^l_i)$. In turn, this output can become the stimulus to neurons in the $(l+1)$th layer. The most commonly used hidden neuron activation function is the sigmoid function [2] given by

$$\sigma(\gamma) = \frac{1}{(1 + e^{-\gamma})} \tag{33.6}$$

Other functions that can also be used are the arc-tangent function and hyperbolic-tangent function [2]. All these are smooth switch functions that are bounded, continuous, monotonic, and continuously differentiable. Input neurons use a relay activation function and simply relay the external stimuli to the hidden layer neuron; that is $z^l_i = x_i$, $i = 1, 2, \ldots, n$. In the case of neural networks for antenna design, where the purpose is to model continuous parameters, a linear activation function can be used for output neurons. The output neuron computation is given by

$$\sigma(\gamma^l_i) = \gamma^l_i = \sum w^L_{ij} z^{L-1}_j \tag{33.7}$$

33.2.4.3 Feedforward Computation Given the input vector $\mathbf{x} = [x_1\ x_2\ \ldots\ x_n]^T$ and the weight vector $\mathbf{w}$, the NN feedforward computation is a process used to compute the output vector $\mathbf{y} = [y_1\ y_2\ \ldots\ y_n]^T$. Feedforward computation is useful not only during NN training, but also during the usage of the trained neural model. The external inputs are first fed to the input neurons (i.e., first layer) and the outputs from the input neurons are fed to the hidden neurons of the second layer. Continuing in this way, the outputs of

the $(L-1)$th layer neurons are fed to the output layer neurons (i.e., the Lth layer). During feedforward computation, NN weights $\mathbf{w}$ remain fixed. The computation is given by

$$z_i^1 = x_i,\ i = 1, 2, \ldots, N_1,\ n = N_1 \tag{33.8}$$

$$z_i^l = \sigma\left(\sum w_{ij}^l z_j^{l-1}\right),\ i = 1, 2, \ldots, N_l,\ l = 2, 3, \ldots, L \tag{33.9}$$

$$y_i = z_i^L, i = 1, 2, \ldots, N_L,\ m = N_L \tag{33.10}$$

33.2.4.4 Important Features It may be noted that the simple formulas (33.8)–(33.10) are now intended for used as antenna models. It is evident that these formulas are much easier to compute than numerically solving theoretical EM equations. This is the reason why NN models are much faster than detailed numerical models of antennas. For the microstrip antenna modeling example described earlier, Eqs. (33.8)–(33.10) represent the model of resonant frequency as functions of dielectric constant, height of the substrate, and radiating patch radius. The question of why such simple formulas in the NN can represent complicated microstrip antenna (or, in general, EM) behavior can be answered by the universal approximation theorem.

The universal approximation theorem [3] states that there always exists a three-layer MLP neural network that can approximate any arbitrary nonlinear continuous multidimensional function to any desired accuracy. This forms a theoretical basis for employing NNs to approximate antenna behaviors, which can be functions of dimensional/construction parameters. MLP neural networks are distributed models; that is no single neuron can produce the overall **x**-**y** relationship. For a given **x**, some neurons are switched on, some are off, and others are in transition. It is this combination of neuron switching states that enables the MLP to represent a given nonlinear input–output mapping. During the training process, the MLP's weight parameters are adjusted and, at the end of training, they encode the component information from the corresponding **x**-**y** training data.

33.2.5 Network Size and Layers

For the NN to be an accurate model of the problem to be learned, a suitable number of hidden neurons are required. The number of hidden neurons depends on the degree of nonlinearity of **f** and dimensionality of **x** and **y** (i.e., values of n and m). Highly nonlinear components need more neurons and smoother items need fewer neurons. However, the universal approximation theorem does not specify the size of the MLP network. The precise number of hidden neurons required for a given modeling task remains an open question. Users can employ either experience or a trial-and-error process to judge the number of hidden neurons. The appropriate number of neurons can also be determined through adaptive processes, which add/delete neurons during training [4, 5]. The number of layers in the MLP can reflect the degree of hierarchical information in the original modeling problem. In general, the MLPs with one or two hidden layers [6] (i.e., three- or four-layer MLPs) are commonly used for antenna applications.

33.2.6 Other Neural Network Configurations

In addition to the MLP, there are other ANN structures [7] used for antennas, for example, radial basis function (RBF) networks, ARTMAP, and self-organizing map (SOM) networks [2]. In order to select a neural network structure for a given application, one starts by identifying the nature of the **x**-**y** relationship.

33.3 NEURAL NETWORK MODEL DEVELOPMENT

The NN does not represent any antenna unless we train it with antenna data. To develop an NN model, we need to identify input and output parameters of the component in order to generate and preprocess data, and then use this data to carry out NN training.

33.3.1 Problem Formulation and Data Processing

33.3.1.1 ANN Inputs and Outputs The first step toward developing an NN is the identification of inputs (**x**) and outputs (**y**). The output parameters are determined based on the purpose of the NN model. Other factors influencing the choice of outputs are (1) ease of data generation and (2) ease of incorporation of the neural model into circuit simulators. Neural model input parameters are those antenna parameters that affect the output parameter values.

33.3.1.2 Data Range and Sample Distribution The next step is to define the range of data to be used in ANN model development and the distribution of **x**-**y** samples within that range. Suppose the range of input space (i.e., **x**-space) in which the neural model would be used after training (during design) is $[\mathbf{x}_{\min}, \mathbf{x}_{\max}]$. Training data is sampled slightly beyond the model utilization range, that is, $[\mathbf{x}_{\min} - \Delta, \mathbf{x}_{\max} - \Delta]$, in order to ensure reliability of the neural model at the boundaries of model utilization range. Test data is generated in the range $[\mathbf{x}_{\min}, \mathbf{x}_{\max}]$.

Once the range of input parameters is finalized, a sampling distribution needs to be chosen. Commonly used sample distributions include the uniform grid distribution, the nonuniform grid distribution, the design of experiments (DoE) methodology, the star distribution, and random distributions. In the uniform grid distribution, each input parameter x_i is sampled at equal intervals. In the nonuniform grid distribution, each input parameter is sampled at unequal intervals. This is useful when the problem behavior is highly nonlinear in certain subregions of the **x**-space and dense sampling is needed in such subregions. Sample distributions based on DoE and star distribution are used in situations where training data generation is expensive.

33.3.1.3 Data Generation In this step, **x**-**y** sample pairs are generated using either simulation data or measured data. The generated data could be used for training the NN and testing the resulting NN model. In practice, both simulations and measurements could suffer from errors that will affect the accuracy of the NN predictions. Considering this, we introduce a vector **d** to represent the outputs from simulation/measurement corresponding to an input **x**. Data generation is then defined as the use of simulation/measurement data to obtain sample pairs $(\mathbf{x}_k, \mathbf{d}_k)$, $k = 1, 2, \ldots, P$. The total number of samples P is chosen such that the developed neural model best represents the given problem **f**. A general guideline is to generate a larger number of samples for a nonlinear high-dimensional problem and fewer samples for a relatively smooth low-dimensional problem.

33.3.1.4 Data Organization The generated ($\mathbf{x}$, $\mathbf{d}$) sample pairs could be divided into three sets, namely, training data, validation data, and test data. Let T_r, V, T_e, and D represent index sets of training data, validation data, test data, and generated (available) data, respectively. *Training* data is utilized to guide the training process, that is, to update the NN weight parameters during training. *Validation* data is used to monitor the quality of the NN model during the training and to determine termination criteria for the training process. Test data is used to independently examine the final quality of the trained neural model in terms of accuracy and generalization capability.

Ideally, each of the data sets T_r, V, and T_e should adequately represent the original component behavior $\mathbf{y} = \mathbf{f}(\mathbf{x})$. In practice, available data D can be split depending on its quantity. When D is sufficiently large, it can be split into three mutually disjoint sets. When D is limited due to expensive simulation or measurement, it can be split into just two sets. One of the sets is used for training and validation ($T_r, = V$) and the other is used for testing (T_e) or, alternatively, one of the sets, is used for training (T_r) and the other for validation and testing ($V = T_e$).

33.3.1.5 Data Preprocessing Contrary to binary data (0's and 1's), the orders of magnitude of various input ($\mathbf{x}$) and output ($\mathbf{d}$) parameter values in antenna applications can be very different from one another. As such, a systematic *preprocessing* of training data called *scaling* is desirable for efficient NN training. At the end of this step, the scaled data is ready to be used for training.

33.3.2 Neural Network Training

33.3.2.1 Weight Parameter Initialization In this step, we prepare the NN for training. The NN weight parameters ($\mathbf{w}$) are initialized so as to provide a good starting point for training (optimization). The widely used strategy for MLP weight initialization is to initialize the weights with small random values (e.g., in the range $[-0.5, 0.5]$.

33.3.2.2 Formulation of Training Process The most important step in the development of an NN model is the training. The training data consists of sample pairs $\{(\mathbf{x}_k, \mathbf{d}_k),\ k \in T_r\}$, where $\mathbf{x}_k$ and $\mathbf{d}_k$ are n- and the m-vectors representing the inputs and desired outputs of the NN. We define the NN training error as

$$E_{T_r}(\mathbf{w}) = \frac{1}{2} \sum_{k \in T_r} \sum_{j=1}^{m} \left| y_j(\mathbf{x}_k, \mathbf{w}) - d_{jk} \right|^2 \tag{33.11}$$

where d_{jk} is the jth element of $\mathbf{d}_k$ and $y_j(\mathbf{x}_k, \mathbf{w})$ is the jth NN output for input $\mathbf{x}_k$.

The purpose of NN training, in basic terms, is to adjust $\mathbf{w}$ such that the error function $E_{T_r}(\mathbf{w})$ is minimized. Since $E_{T_r}(\mathbf{w})$ is a nonlinear function of the adjustable (i.e., trainable) weight parameter $\mathbf{w}$, iterative algorithms are often used to explore the $\mathbf{w}$-space efficiently. One begins with an initialized value of $\mathbf{w}$ and iteratively updates it. Gradient-based iterative training techniques update $\mathbf{w}$ based on error information $E_{T_r}(\mathbf{w})$ and the error derivative information $\partial E_{T_r} / \partial \mathbf{w}$. The subsequent point in $\mathbf{w}$-space denoted as $\mathbf{w}_{\text{next}}$ is determined by a step down from the current point $\mathbf{w}_{\text{now}}$ along a direction vector $\mathbf{h}$; that is, $\mathbf{w}_{\text{next}} = \mathbf{w}_{\text{now}} + \eta \mathbf{h}$. Here, $\Delta \mathbf{w} = \eta \mathbf{h}$ is called the weight update and η is a positive step size known as the *learning rate*. For example, the backpropagation (BP) training algorithm [1, 2] updates $\mathbf{w}$ along the negative direction of the gradient of training error as $\mathbf{w} = \mathbf{w} - \eta(\partial E_{T_r} / \partial \mathbf{w})$.

33.3.2.3 Error Derivative Computation As mentioned earlier, gradient-based training techniques require error derivative computation, that is $\partial E_{T_r}/\partial \mathbf{w}$. For the MLP neural network, these derivatives are compared using a standard approach often referred to as error backpropagation, which is described here. We define a per-sample error function E_k given by

$$E_k = \frac{1}{2}\sum_{j=1}^{m}(y_j(\mathbf{x}_k, \mathbf{w}) - d_{jk})^2 \tag{33.12}$$

for the kth data sample $k \in T_r$. Let δ_i^L represent the error between the ith NN output and the ith output in the training data; that is,

$$\delta_i^L = y_i(\mathbf{x}_k, \mathbf{w}) - d_{ik} \tag{33.13}$$

Starting from the output layer, this error can be backpropagated to the hidden layers as

$$\delta_i^l = \left(\sum_{j=1}^{N_{l+1}} \delta_j^{l+1} w_{ji}^{l+1}\right) z_i^l(1 - z_i^l), \quad l = L-1, L-2, \ldots, 3, 2 \tag{33.14}$$

where δ_i^l represents the local error at the ith neuron in the lth layer. The derivative of the per-sample error in Eq. (33.12) with respect to a given NN weight parameter w_{ij}^l is given by

$$\frac{\partial E_k}{\partial w_{ij}^l} = \delta_i^l z_j^{l-1}, \quad l = L, L-1, \ldots, 2 \tag{33.15}$$

Finally, the derivative of the training error in Eq. (33.11) with respect to w_{ij}^l can be computed as

$$\frac{\partial E_{Tr}}{\partial w_{ij}^l} = \sum_{k \in T_r} \frac{\partial E_k}{\partial w_{ij}^l} \tag{33.16}$$

Using error backpropagation, $(\partial E_{T_r}/\partial \mathbf{w})$ can be evaluated systematically for the MLP neural network structure and can be provided to gradient-based training algorithms for the determination of weight update $\Delta \mathbf{w}$.

A flowchart summarizing major steps in NN training and testing is shown in Figure 33.3 [8].

33.4 OTHER NEURAL NETWORK MODELS USED FOR ANTENNA MODELING

33.4.1 Radial Basis Function (RBF) Neural Network

The RBF network [7] is another powerful supervised trained network that can be used for pattern classification and function approximation. It has a simpler structure (one hidden layer). One of the special basis functions (i.e., the nonlinearities) that are commonly used is a Gaussian kernel function. The basis functions in the hidden layer of the network produce a significant nonzero response when the input to them falls within a relatively

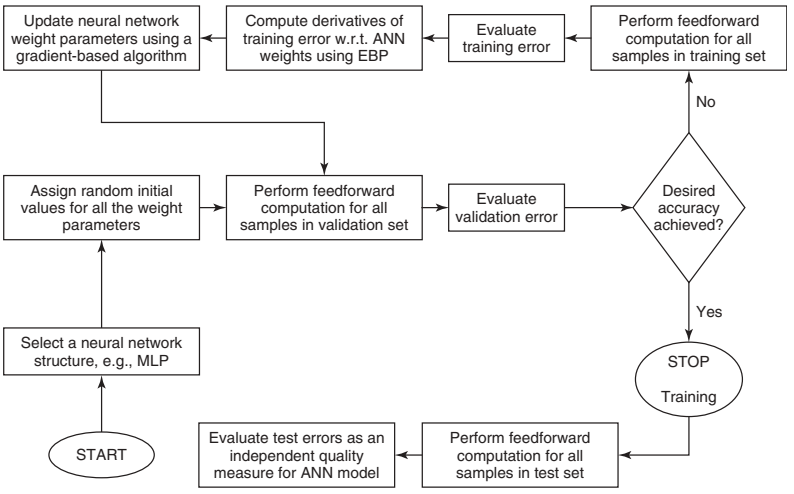


Figure 33.3 Flowchart demonstrating NN training, neural model testing, and use of training, validation, and test data sets in ANN modeling approach. (From Ref. 8. Reprinted with permission, copyright © 2003 IEEE.)

small, localized region of the input space. For the RBF networks, the input patterns of the training data are often used as the candidate of the centers of the radial basis functions, but the strategy to choose the radius (or scale) parameter of the radial basis functions is not clearly defined. In many cases, RBF networks will train much more quickly than feedforward MLPs trained by backpropagation. The RBF network was used previously with antenna arrays for direction of arrival (DoA) determination and beam steering [2].

33.4.2 Self-Organizing Map (SOM) Neural Network

There are many different types of self-organizing neural networks; however, they all share a common characteristic. This is the ability to assess the input patterns presented to the network, organize itself to learn on its own the similarities among the collective set of inputs, and categorize (or cluster) them into groups of similar patterns. Therefore these types of NNs [7] learn without a "teacher," that is, through unsupervised learning.

In general, self-organized learning (or unsupervised learning) involves the frequent modification of the network's synaptic weights in response to a set of input patterns. The weight modifications are carried out in accordance with a set of learning rules. After repeated applications of these patterns to the network, a configuration emerges that is of some significance. Basically, from numerous originally random local interactions within a network, in response to the input patterns, there emerges global order. This global order can ultimately lead to some form of congruous behavior.

The SOM NN consists of an input layer of nodes, where the inputs to the NN are applied, and an output layer of nodes, where the categorization (grouping/clustering) of the inputs are formed. The nodes in the output layer, most often, are organized in a two-dimensional array. Training is performed in an unsupervised way using the Kohonen learning algorithm [9]. The Kohonen SOM is the only unsupervised trained network found in antenna literature. The training, in this case, can be viewed as a procedure that learns to group input patterns in clusters in a way inherent to the data. To train the SOM NN, continuous valued input vectors are presented in random sequence to the network. The

mapping from the external input patterns to the network's activity patterns is realized by correlating the input patterns with the connection weights. After enough input patterns have been presented, weights converging to output nodes of the SOM NN specify cluster centers that represent the input patterns.

The adaptive training process for SOM is described in Ref. 7. In this, the weights of a two-dimensional SOM network is updated according to

$$\mathbf{w}_j(t+1) = \mathbf{w}_j(t) + \eta(t) h_{j,i(x)}(t)(\mathbf{x} - \mathbf{w}_j(t)) \tag{33.17}$$

where $\eta(t)$ is the time-varying learning parameter given by $\eta(t) = \eta_0 \exp(-t/\tau_2)$, $t = 0, 1, 2, \ldots$; η_0 (≈ 0.1) and τ_2 (≈ 1000) are constants. $h_{j,i(x)}(t)$ is the neighborhood function given by

$$h_{j,i(x)}(t) = \exp\left(\frac{-d_{j,i}^2}{2\sigma^2(t)}\right), \quad t = 0, 1, 2, \ldots \tag{33.18}$$

where σ is the "effective width" of the topological neighborhood given by $\sigma(t) = \sigma_0 \exp(-t/\tau_1)$, $t = 0, 1, 2, \ldots$; $\tau_1 = (1000/\log \sigma_0)$ and σ_0 (≈ 1) is a constant. $d_{j,i}$ is the lateral distance between winning neuron i and excited neuron j.

33.5 A TYPICAL EXAMPLE

In order to apply ANN techniques for a particular antenna problem, the first task is to bring the problem to a form that is suitable for an ANN application: that is, the problem has to be formulated into one of the categories of mapping, function approximation, or classification. Data has to be generated to create the training environment for the ANN. This can be done using experiments or simulation packages or with rigorous theoretical derivations. Depending on the type of problem and looking to other factors, like number of data patterns available, number of input/output parameters, or complexity of the problem, a training algorithm has to be chosen. Standard training algorithms [7] are available. The aim of the training is to get a specified set of weights that can be used later to determine the outputs of the ANN or the system as a whole. Whether the obtained network is a properly trained one or not can be verified using the test data set, a data set that is usually complementary to the training data set.

Before describing the use of NNs for different types of antenna applications, here a simple example is described, to get an idea of approaching a specific problem for NN application. The example describes the calculation of resonant frequency for an equilateral triangular microstrip antenna, shown in Figure 33.4, with the use of NNs [10].

Problem

Calculate the resonant frequencies for an equilateral triangular microstrip antenna using NNs.

Why NNs for This Problem?

1. Determination of the resonant frequency of a triangular microstrip patch antenna for the TM_{mn} mode requires the computation of the effective side length and the effective permittivity constant. Finding these is complicated and time consuming.

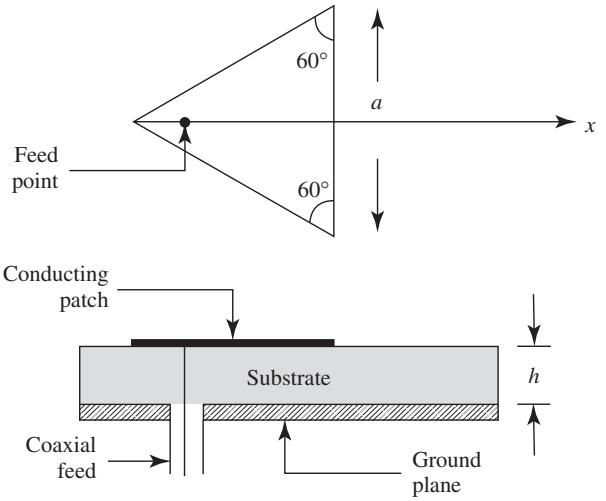


Figure 33.4 Geometry of equilateral triangular microstrip antenna. (From Ref. 10. Reprinted with permission from John Wiley & Sons, Inc.)

On the other hand, when the relationship between the parameters of the microstrip antenna and the measured resonant frequency results are modeled in the NN, it neither requires a formula nor the calculations of the effective side length and effective permittivity constant. The NN model only requires the side length of the patch, the permittivity of the substrate, and its thickness.

2. The theoretical resonant frequency results obtained from the trained NN are in very good agreement with the experimental results and are far better than results obtained using empirical formulas.

Implementation

Step 1. Generation of Data: Experimentally measured resonant frequency data available in the literature were used for training of the NN. Fifteen sets of such data values were collected for different dimensions and modes of the antenna, out of which 3 sets were used as a test set and the rest as a training set, because of the limited experimental data available. It should be emphasized here that better and more robust results can be obtained from the NN models if more input data set values are available for training.

Step 2. Choosing the Training Algorithm and the Network Structure: A backpropagation training algorithm was used in this case because this is the most widely used training algorithm for function implementation and at the same time it is easy to use.

The task of choosing the number of appropriate neurons in the input and output layers is problem dependent. In this particular example the intention is to find the resonant frequency, which is a function of the antenna parameters: side length (a), height of the substrate (h), substrate permittivity (ε_r), and the mode numbers (m and n). Therefore five neurons in the input layer and one neuron in the output layer are required. There is no specific rule to determine the number of neurons in the hidden layer(s). Moreover, as mentioned earlier, theoretically it has been

proved that a multilayer NN with at least one hidden layer can model arbitrarily complex nonlinear input–output relationships. In this example, after several trials, it was found that the two-layer network achieved the task with high accuracy. So the network configuration chosen in this case was $5 \times 5 \times 3 \times 1$. This means that the number of neurons was five for the first hidden layer and three for the second hidden layer.

Step 3. Choosing the Training Parameters: After fixing the network structure, the efficiency of training depends on many training parameters such as learning rate, momentum, and training tolerance. The optimum values of these parameters are problem dependent and found by trial and error or experience. This is more of an art than science. The values of the different training parameters for this example are given in Table 33.1.

Step 4. Testing of the Developed Network: In order to ascertain that the developed network is properly trained, the trained NN has to be tested for validation with the test data set. If the network is providing proper results within the desired tolerance, then the developed network will be treated as a properly trained one. A typical set of results from the tested NN, in this example, is given in Table 33.2.

Benefits of Using an NN for the Present Problem

1. Because the neural model developed in this example has high accuracy and requires no complicated mathematical functions, it can be very useful for the development of fast CAD algorithms.
2. The very good agreement between the measured values and the computed resonant frequency values supports the validity of the neural model even with the limited data set. This CAD model, capable of accurately predicting the resonant frequencies of triangular microstrip antennas, is also useful to antenna engineers. Using this model, one can calculate accurately the resonant frequency of triangular patch antennas, without any background knowledge of microstrip antennas.
3. It takes only a few microseconds to produce the resonant frequencies on a PC. Even if the training time takes less than 10 minutes, after training, the calculation time is less than 100 μs in real-time calculations. Thus the neural model is very fast after training.

TABLE 33.1 Training Parameter Values

Parameter	Value
Number of input layer neuron(s)	5
Number of output layer neuron(s)	1
Number of hidden layers	2
Number of first hidden layer neuron(s)	5
Number of second hidden layer neurons	3
Learning rate	0.3
Momentum	0.5
Number of epochs	25,000

Source: Ref. 10.

TABLE 33.2 Comparison of Measured and Calculated Resonant Frequencies of the First Five Modes of an Equilateral Triangular Microstrip Antenna

Mode	f_{me} (MHz)	f_{mn} Present Method (MHz)
TM_{10}	1519	1526
TM_{11}	2637	2637
TM_{20}	2995	2995
TM_{21}	3973	3973
TM_{30}	4439	4439
TM_{10}	1489	1478
TM_{11}	2596	2596
TM_{20}	2969	2969
TM_{21}	3968	3968
TM_{30}	4443	4443
TM_{10}	1280	1280
TM_{11}	2242	2242
TM_{20}	2550	2550
TM_{21}	3400	3400
TM_{30}	3824	3829

Source: Ref. 10.

[a] f_{me}, measured values [11, 12]; f_{mn}, NN calculated values.

33.6 ANTENNA APPLICATIONS

A list of applications of NNs for antennas and arrays is given in Table 33.3, along with the type of network structure used and the purpose of application. In this chapter, we are not describing the applications for the arrays; rather, we are confining ourselves to the detailed descriptions for single antennas of various types.

33.6.1 Horn Antenna Analysis and Synthesis

33.6.1.1 Necessity of NN Use In many reflector and lens antennas, corrugated circular horns constitute one of the best feed solutions due to their polarization purity and small size. However, they require very accurate procedures for their modeling, design, and fabrication in order to avoid a poor yield. Characterization of such antennas using full-wave software simulators, based on mode matching and on the combined field integral equation (CFIE) techniques [38]. This is computationally time consuming, on the order of several minutes for each frequency point. Conventional optimization techniques rely on a very large number of such analyses, so automatic horn design is an exceedingly demanding task even for powerful computers. In order to overcome such difficulty, an unconventional approach, based on the use of ANNs, can be utilized [13].

33.6.1.2 Problem Formulation and Implementation The paradigm of analysis and synthesis for the horn antenna problem is shown in Figures 33.5 and 33.6, respectively. In the analysis problem, where an ANN is used as an alternative to the electromagnetic simulator, the scheme of Figure 33.5 has to be followed for the ANN learning process. The input parameters of the electromagnetic analyzer constitute the inputs of the ANN, while the results obtained by the analyzer are the desired outputs.

TABLE 33.3 Some of the Applications of NNs for Antennas/Arrays

Applications	Purpose/Advantage	Type of Network Used	Reference
Horn antenna analysis and synthesis	To reduce the computation time that is required by the full-wave software simulators	Multilayer feedforward network	13
Aperture antenna shape prediction	To significantly reduce the time of production associated with this relatively broad class of antennas To provide a methodology that can be used in the control of active aperture antennas	Multilayer feedforward network	14
Design and analysis of microstrip antenna/CPW patch antenna	To bypass the repeated use of complex iterative processes To avoid CPU intensive simulation procedures	Multilayer feedforward network, radial basis function network, and knowledge-based NN	15–25
Design and analysis of reconfigurable antenna	To reduce the heavy computational resources required for analyzing the multiscale nature of the reconfigurable antenna	Multilayer feedforward network, self-organizing map NN	26, 27
Wideband mobile antenna design	To avoid lengthy full-wave EM analysis with a faster method	Multilayer feedforward and knowledge-based NNs	28
Direction of arrival (DoA) estimation	To reduce the computational complexities of the previously available methods of DoA estimation	Radial basis function network	29–35
Adaptive beam forming	Real-time implementation of beam former to respond to the time-varying environment	Hopfield network and radial basis function network	36, 37

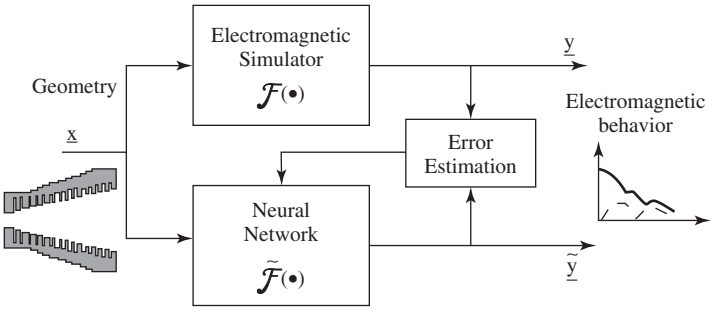


Figure 33.5 ANN learning scheme for the analysis problem. (From Ref. 13. Reprinted with permission, copyright © 2001 IEEE.)

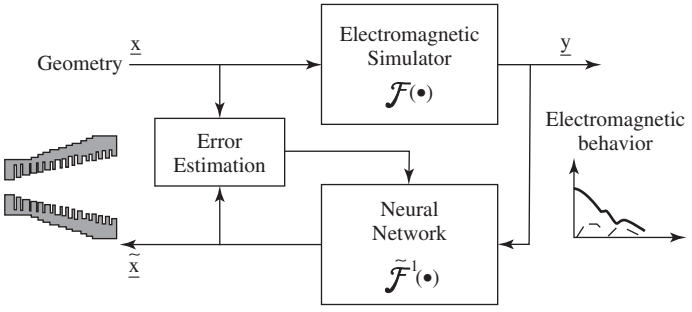


Figure 33.6 ANN learning scheme for the synthesis problem. (From Ref. 13. Reprinted with permission, copyright © 2001 IEEE.)

On the other hand, for the synthesis problem, the approach of Figure 33.6 has to be chosen. In this case the inputs of the full-wave analysis are the desired outputs of the ANN, while the results given by the electromagnetic simulator are the ANN inputs. Here it is important to point out that, for the synthesis case, the ANN learning sets have to be single valued; that is, the same output values may not correspond to more than one single input value. If this situation occurs, the ANN learning process may not converge to a solution. Hence a suitable selection of the learning patterns (training data) set produced by the EM analyzer has to be determined, in order to avoid this situation. A supervised feedforward-type ANN trained in the backpropagation mode is used in this work.

Figure 33.7 depicts the basic geometry of the horn. Half a longitudinal section is shown. Geometrical parameters are the throat radius r, the aperture radius R, the horn length L, and the corrugation parameters s, w, b, and t, which are the corrugation depth, period, and the void/metal occupation in each period, respectively. Alternatively, the quantity $n = \lambda/w$, which gives the number of corrugations per wavelength and the metal–void ratio t/b can be used. The horn has a profiled longitudinal section according to the law

$$r(z) = r + (R - r)\left[\frac{z}{L}(1 - A) + A\sin^2\left(\frac{\pi}{2}\frac{z}{L}\right)\right] \tag{33.19}$$

where r, R, and L have the aforementioned meanings, z is the coordinate along the horn axis, and A determines the amount of profiling. $A = 0$ gives a standard linear horn, while $A = 1$ gives a sine squared profile. In Ref. 13, a value $A = 0.7$ has been chosen, which

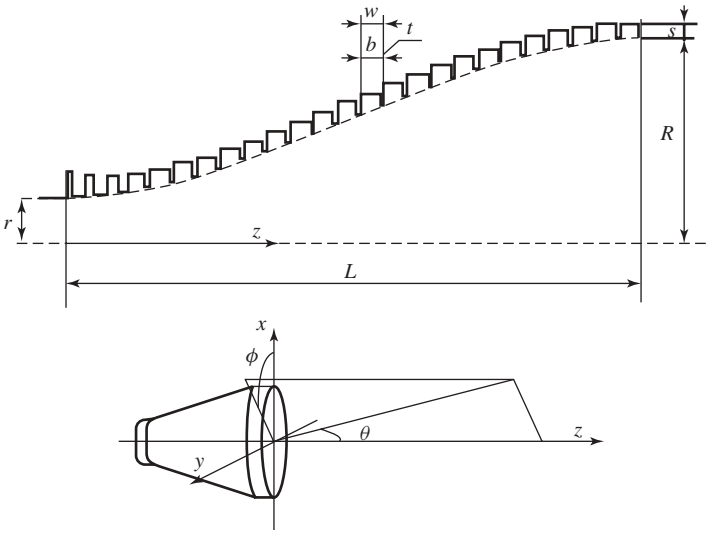


Figure 33.7 Geometry of a profiled corrugated circular horn and reference system for radiation patterns. (From Ref. 13. Reprinted with permission, copyright © 2001 IEEE.)

allows for shorter horns having the same aperture R and a comparable return loss Γ with respect to a linear horn.

For the analysis task the geometrical parameters R, L, s, t/b, and n are considered as ANN inputs. The outputs are the samples of the copolar and cross-polar patterns over a suitable set of directions and the return loss Γ. This number of inputs is so high that some reduction in the ANN learning set is necessary because an exhaustive search in the input space would require an enormous amount of examples. Therefore some suitable method has to be adopted for the determination of a particular learning set. In order to simplify the analysis problem and speed up the learning process, several ANNs are required rather than only one. This is always an approach that can be used when the training of the ANN becomes difficult due to the many possible scenarios that can exist for training. Each ANN is related to the task of mapping a subspace of the input space to a subspace of the output space. Figure 33.8 describes this approach where a subspace selection logic unit (SSLU) is used. Then each ANN of the selected set computes its portion of output values.

In the work described, the radiation pattern output has been limited to the $\theta \in [0^\circ, 40^\circ]$ range with a 2° step, which is the most significant one. The network set consists of four ANNs for the copolar and four ANNs for the cross-polar pattern at each desired ϕ cut (only the $\phi = 45^\circ$ cut is taken here, the most relevant for the cross-polar component) and one ANN for the Γ. So a total of 90 ANNs have been trained, with the input space divided into ten subspaces.

Figure 33.9 shows the copolar and cross-polar patterns for a horn with $R = 2.45\lambda$, $L = 8.25\lambda$, $s = 0.27\lambda$, $t/b = 0.35$, and $n = 5$. The analysis has been performed with the conventional mode matching–CFIE technique and with the proposed ANN approach.

The synthesis procedure could be carried out by training an ANN to directly approximate the function mapping the electromagnetic characteristics into the geometrical parameters. Hence the ANN inputs are the typical design requirements for a feed: the main

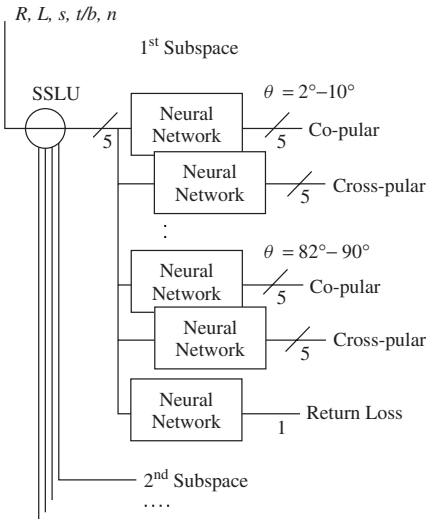


Figure 33.8 ANN setup for horn analysis with input decision logic and output subspaces. (From Ref. 13. Reprinted with permission, copyright © 2001 IEEE.)

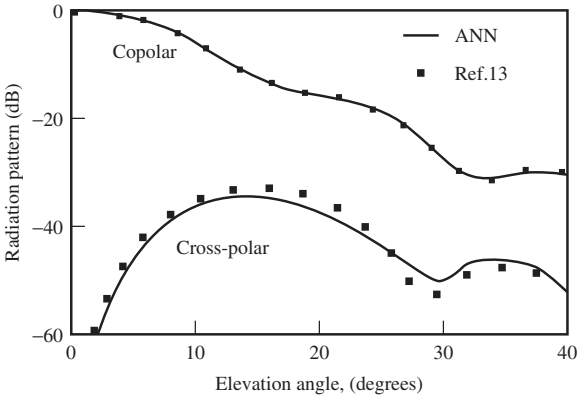


Figure 33.9 Comparision between the ANN and the full-wave simulation. (From Ref. 13. Reprinted with permission, copyright © 2001 IEEE.)

beamwidth, the maximum level of the cross-polar level $X_{\max}$, and the return loss Γ. The former is described, instead of as the 3-dB angle, in a way characteristic of feeds for reflector antennas, that is, the angle θ_{ET} for which the radiated field level is ET dB below the maximum ($ET =$ edge taper). The outputs are the geometrical parameters *R, L, A, s, n*, and *t/b*.

To reduce this number of outputs, some fixed values are assumed: $A = 0.7$, $n = 4$, $s = \lambda/4$, and $t/g = 0.2$. This assures good performance and at the same time a low production cost, since horns with a small number of relatively large corrugations are considerably cheaper. Then Γ is not considered among the ANN inputs. This assumption can be taken since profiled corrugated circular horns easily exhibit very low Γ levels. Furthermore, by

training the network with data relative to horns with good performance—that is, with a low Γ—it is expected that the geometries generated by the ANN also exhibit low return loss, and this proved to be the case.

In designing the ANN, care must be exercised when there are many groups of geometrical parameters leading to the same horn performance. When this occurs, the function $f(R, L) \rightarrow [\theta_{ET}, ET, X_{\max}]$ cannot be inverted. Since the ANN should invert such a function to perform the horn design, the function needs to be made single valued by considering only those geometries with the lowest value of R for a given θ_{ET} and ET pair.

The developed ANN has been used to design a horn with $\theta_{ET} = 9.5^{\circ}$, $ET = 15$ dB, and minimum $X_{\max}$. The obtained feed has $R = 3.92\lambda$, $L = 29.45\lambda$, and the usual values for the corrugation parameters. The return loss Γ is equal to 46.1 dB. A conventional design has also been used for comparison, that is, a heuristic trial-and-error technique, based on the full-wave electromagnetic analysis. The latter required many weeks of work. Figure 33.9 shows both the full-wave analysis of the ANN designed horn (curve a) and the heuristically designed horn (curve b) . The horn obtained by using the ANN approach has been measured, only for the copolar pattern, and the data are also reported in Figure 33.10.

The obtained ANN algorithms, both for analysis and synthesis, are easily embeddable in existing CAD tools for antenna design, which will take full advantage of such an accurate method with low memory and CPU time requirements. This ANN approach for the horn analysis can be inserted into a standard optimization procedure. Alternatively, the ANN designed horn can be used as a starting point for a full-wave simulator-based optimization.

33.6.1.3 Points to Note from this Application It is not always feasible to develop a single trained network for the entire range of problem input/output space. In that case, the input space can be split into a number of subspaces and individual trained NNs can be developed for each subspace. Here provisions should be made to select the required subspace depending on the input data range.

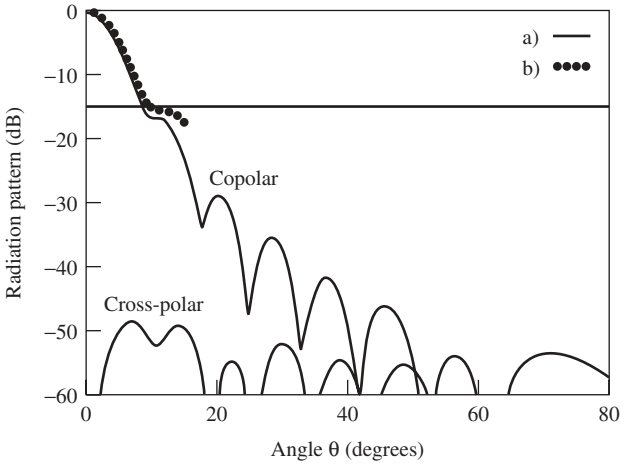


Figure 33.10 Full-wave analysis for a horn synthesized via the ANN approach: (a) measured pattern and (b) prototype built accordingly to the ANN design. (From Ref. 13. Reprinted with permission, copyright © 2001 IEEE.)

33.6.2 Aperture Antenna Shape Prediction

33.6.2.1 Necessity of NN Use With the advances in aperture antenna fabrication, the design of nonuniform, semiparaboloidal shaped apertures has emerged. This class of antennas can produce any desired far-field pattern. Calculating the antenna far-field beam pattern as a function of aperture shape is relatively straightforward. The inverse problem, where one calculates the shape necessary to obtain a particular ground pattern, is not as simple. The existing techniques require trial-and-error approaches that are always timeconsuming. Washington [14] has shown that a feedforward NN can be used to establish a methodology in designing this class of antennas, once a "desired" pattern is defined. Once the NN is trained, for a given desired pattern, it can predict the aperture shape that generates this radiation pattern in real time.

33.6.2.2 Problem Formulation and Implementation The paradigm of application of NN for aperture antenna shape prediction is shown in Figure 33.11. In this particular application, a multilayer perceptron NN trained in the backpropagation mode was used to perform a nonlinear mapping between the far-field radiation pattern (NN input) of reflector antenna and the numerical contribution of each mode of the reflector (NN output), calculated using a mechanical engineering problem, considering a cylindrical reflector surface made of two flexible shells connected at the center. The numerical contribution of each mode in turn can be used to calculate the displacement of the reflector.

A three-layer NN with 41 inputs (that means the pattern was sampled at 41 points), 30 hidden layer-1 neurons, 30 hidden layer-2 neurons, and two outputs (for the first two reflector modes) was utilized. The first two reflector modes can adequately describe beam steering and shaping, so two output neurons were taken. If more modes were taken into account, it would be possible to reconstruct the surface more accurately from a linear combination of modes.

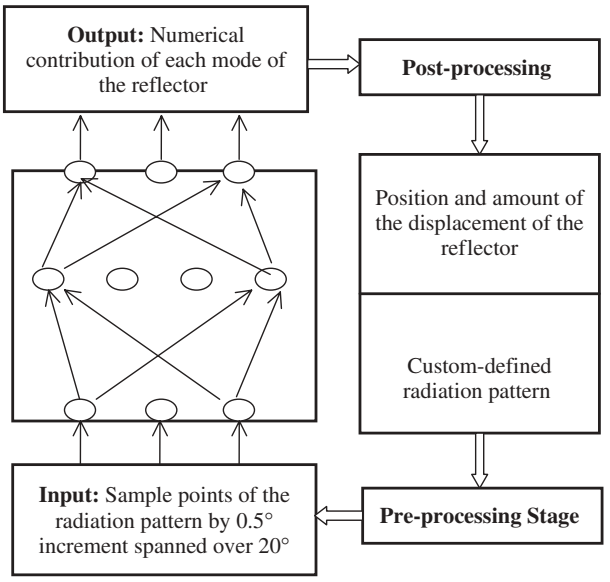


Figure 33.11 Paradigm of NN application for aperture antenna shape prediction.

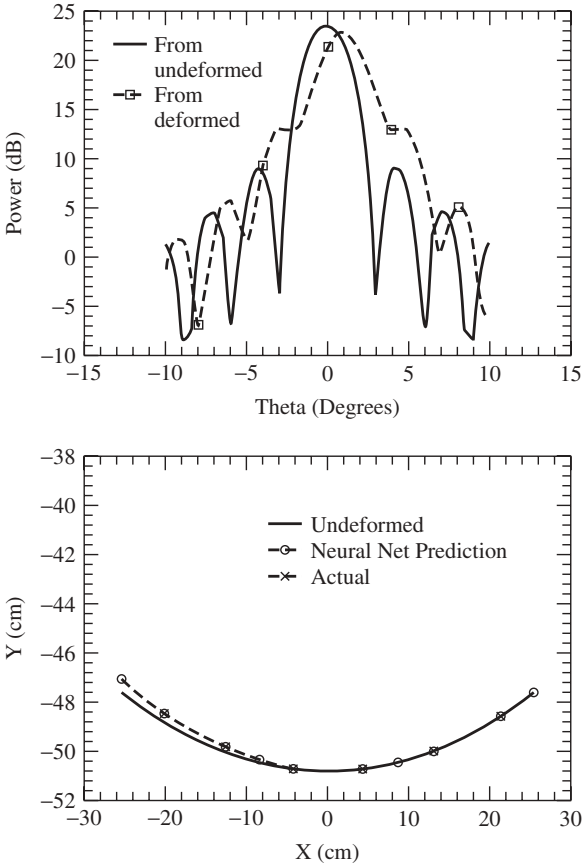


Figure 33.12 Predicted and actual shape. (From Ref. 14. Reprinted with permission, copyright © 1997 IEEE.)

A typical beam pattern used as the input to the trained NN and the predicted and actual shape are shown in Figure 33.12. The result reveals that NNs can be utilized to accurately determine reflector shapes given the associated far-field power pattern. Since the mathematical function represented by the NN is just an input–output mapping, it can easily be employed in real time. Although the training time is large, the final input–output mapping takes seconds to run and can be used to predict the shapes of many different antennas. Once the network is designed, it can be used over and over again, thereby saving the designer time.

33.6.2.3 Points to Note from This Application It has been theoretically proved that a multilayer NN with at least one hidden layer can model arbitrarily complex nonlinear input–output relationships. Even then, if the problem at hand is very complex, a larger number of hidden layers with the required number of neurons can be taken, in order to ease the task.

Additionally, in some cases, raw data is not suitable to apply directly to the NN. In that case some preprocessing technique has to be applied in order to make the data fit for training and testing and also reduce the number of neurons at the input or output layers.

33.6.3 Microstrip Antenna Analysis and Design

33.6.3.1 Necessity of NN Use Although the structure of the microstrip antenna looks somewhat simple, its analysis is very involved. The full-wave analysis methods used for microstrip antennas are computationally complex and hence take an appreciable amount of time to find a single analysis parameter, whereas hundreds of simulations are required for getting an optimized result. Methods based on empirical formulas are efficient from the computation time point of view but vary in terms of their applicability range. On the other hand, the neurocomputational models are fast and at the same time robust from an accuracy point of view. So the models are suitable for CAD applications.

33.6.3.2 Problem Formulation and Implementation A lot of applications of NNs for microstrip antennas can be found in the literature, both for analysis and design [15–25]. On the analysis side, it includes the calculation of resonant frequency of electrically thin and thick rectangular and circular microstrip antennas [16, 18–20], bandwidth calculation of rectangular microstrip antennas [17], and input resistance of circular microstrip antennas [23]. The example described in the previous section—calculating the resonant frequency of equilateral triangular microstrip antenna [10]—is one such application.

In all these applications, the job of the network is to act as a simple novel alternative for the computationally involved procedure to find the analysis parameter. The data generation in these works is either from simulations based on full-wave analysis or from some experiment, and the trained network was developed to map the different design parameters of the antenna with the output analysis parameter. These applications are individual networks for individual parameters. However, in Ref. 22 a generalized network has been developed, in which a single NN can be used to calculate the resonant frequencies of rectangular, circular, and triangular microstrip antennas.

On the design side, NNs were used to find the patch size of some regular shape microstrip antennas for a specific operating frequency [24, 25]. The paradigm of application for design of a square patch antenna is shown in Figure 33.13. In the design of

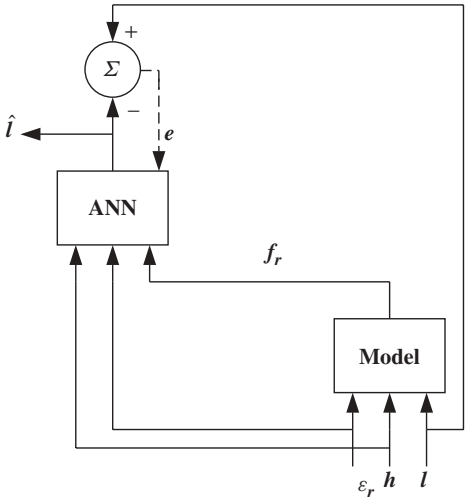


Figure 33.13 Calculation paradigm of design of square patch antenna.

square patch antennas, the resonant frequency in the dominant mode is given, ε_r and h are supplied, and l is calculated.

Here, the Wolf–Knoppik formula [25, 39] is used to find the resonant frequency of the square patch antenna for different input parameters (ε_r, h, and l) in specified ranges. The generated data were then arranged in two matrices. A three-column matrix containing the values of ε_r, h, and f_r is used as the input training data for the network. The other one-column matrix containing the corresponding values of l is used as the output training data for the network.

A distinct advantage of neurocomputing is that, after proper training, an NN completely bypasses the repeated use of complex iterative processes for new cases presented to it. The developed model in the present example produces the length of the side of a square patch almost instantaneously for the other three specified parameters of the square patch antenna. Moreover, this single network structure can predict the results for a frequency range up to 15 GHz provided the values of ε_r and h are in the domain of training values.

The same procedure can be applied for designing any regular shape microstrip antenna [24].

33.6.3.3 Points to Note from this Application The accuracy of a trained network depends on the accuracy of the data used for its training. Since obtaining data can be expensive, care must be taken while generating the training data set. If experimental facilities are available, additional data can be used.

33.6.4 Reconfigurable Antenna Analysis and Design

33.6.4.1 Necessity of NN Use In response to the ever-increasing needs of antenna bandwidth, a considerable amount of effort is currently underway to develop multiband antennas. Mostly planar designs are preferred for these structures due to their added advantage of small size, low manufacturing cost, and conformability. The multiband property can be achieved by making the structure reconfigurable, in which the different radiating elements of an array are connected together using switches to form groups of elements that resonate at different frequency bands [40–45]. Reconfigurable multiband antennas are attractive for many military and commercial applications where it is desirable to have a single antenna that can be dynamically reconfigured to transmit and/or receive on multiple frequency bands. Such antennas find applications in space-based radar, unmanned aerial vehicles, communication satellites, electronic intelligence aircraft, and many other communications and sensing applications. The technology of design and fabrication of microelectromechanical systems (MEMSs) for RF circuits has had a major positive impact on reconfigurable antennas [46–48].

For frequency reconfigurable antennas, the challenging tasks are (1) finding an analytical procedure to locate the frequency bands of operation (analysis problem) and (2) determining how to connect the radiating elements together such that the resulting module will have desired frequency bands (design problem), or in other words, determining which switches to turn ON so that a specific set of elements will be active to make the structure operate in the desired frequency bands.

Due to the multiscale nature of reconfigurable antennas, a single analytical method cannot characterize the whole structure. On the other hand, the use of different analytical methods for a single structure makes it a computationally intensive task, leading to the use

of heavy computational resources. So there is a need to search for an analysis procedure for reconfigurable antennas that can characterize the antenna accurately.

Neural network application for the analysis and design of a multiband reconfigurable microstrip antenna has been studied by Patnaik et al. [26]. Two different neural architectures have been used for analysis and design of a reconfigurable antenna. In the analysis phase, NNs are used to locate the operational frequency bands for different combination of switches. This is treated as a mapping formation problem and is accomplished by an MLP trained in the backpropagation mode. In the design phase, the job of the NN is to determine the switches that are to be ON for the structure to resonate at specific bands. This task is handled as a classification type of problem and is accomplished by a self-organizing map (SOM) neural network.

33.6.4.2 Problem Formulation and Implementation The proposed neurocomputational technique was investigated for a laboratory prototype antenna. The structure under investigation is shown in Figure 33.14.

The antenna was fabricated on a Duroid substrate ($\varepsilon_r = 2.2$) with no radiating element touching the adjacent elements. In the absence of actual MEMS switches, their electromagnetic performance was considered ideal and their placement was accomplished by small physical connections of the antenna's adjacent conducting parts.

Neurocomputational Technique for Analysis Through different combinations of switches, it has been observed that the bands at which the antenna resonates depend on the switch positions that are in the "ON" state. In order to implement this nonlinear mapping function between the ON switch positions and the resonance pattern of the reconfigurable antenna, we have used an MLP neural network trained in the backpropagation mode.

The parameters to be mapped, for which data generation is required, are the reconfigurable antenna structure and the corresponding frequency response. Experimentally measured S_{11} plots were sampled for various combination of structures. Depending on the ON and OFF switch positions in the reconfigurable structure, bit strings of 1's and 0's were formed to make the input data set and the corresponding sampled values of the S_{11} plot form the output data set. The trained network was then tested for the frequency response of the reconfigurable structure for different combinations of switches. The response of some of the typical combinations is shown in Figure 33.15. The responses are also compared with the measured values. The developed network can be used to find the operational frequency bands of the multiband antenna. The advantage of using NNs

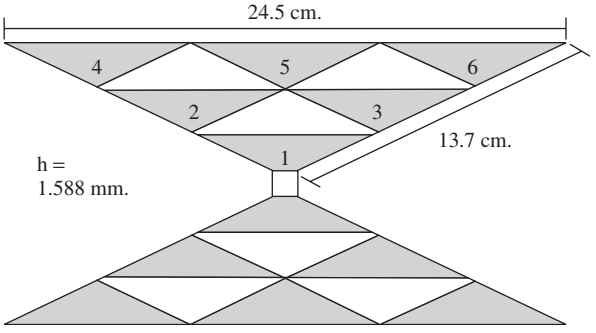


Figure 33.14 Multiband antenna under investigation.

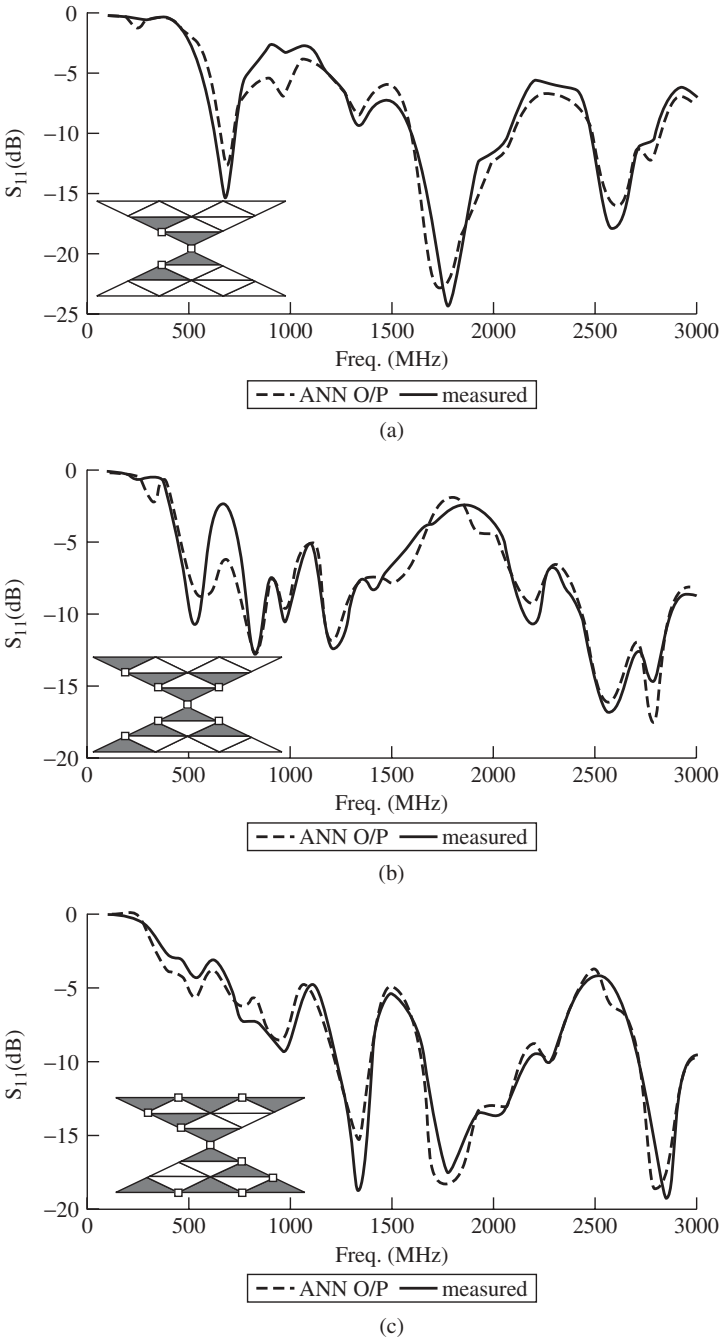


Figure 33.15 Comparision of NN output of some typical reconfigurable structures with the measured value. (From Ref. 26. Reprinted with permission, copyright © 2005 IEEE.)

is that it avoids the computational complexity involved in the numerical modeling of the antenna. Furthermore, the response time is very fast for NNs.

Neurocomputational Technique for Design The purpose in the design process is to correlate any frequency response, within the operational range of the antenna, with a reconfigurable structure as close as possible. The design procedure was approached as a clustering problem and used a SOM neural network [9] for classification of the frequency response plots. The task of the SOM is to map a continuous input space of activation patterns onto a discrete output space of neurons by a process of competition among the neurons in the network. Based on the shape of the frequency response plots, the SOM NN classifies the responses into different clusters (Figure 33.16). Each cluster has some similarity among the frequency responses in the operational range and their depths. The clusters so formed are then related with their corresponding antenna structure (Figure 33.17) Therefore, for each cluster, a set of typical reconfigurable structures was formed. Given a frequency response plot, the corresponding approximate reconfigurable structure can be traced out from these sets of typical structures, depending on to which cluster the frequency response is adapting, as determined by the SOM. The paradigm of the neurocomputational design process is shown in Figure 33.18.

The developed SOM network can be used for design purposes. The steps of approaching the design process are as follows:

Step 1. Give the network the desired frequency response (input).

Step 2. The SOM NN matches the frequency response to the closest cluster.

Step 3. The antenna configuration can be chosen from the set of structures corresponding to that cluster, starting with a simple structure with minimum number of switches.

Step 4. Depending on the requirement, more elements can be excited over the original structure.

33.6.4.3 Points to Note from This Application Numerical data is required for training NNs. So some suitable method has to be adopted to obtain numerical data for the problem, if initially it is not expressed explicitly in numerical form. Also, depending on the nature of the problem, whether classification or regression type, a corresponding suitable neural architecture has to be chosen.

33.7 SOME ISSUES OF USING NEURAL NETWORKS FOR ANTENNA PROBLEMS

1. For using NNs, the identified problem in hand first has to be checked for its suitability for NN implementation. That means, for simple linear functions or for problems that can be implemented through a direct closed form formula, it is advisable not to resort to NN techniques.
2. After specifying the problem, either it can be implemented in total using NN or the whole problem can be divided into parts and NNs can be used to implement a part. By implementing the problem in its totality, the NN acts as a black-box (Figure 33.19) and does not disclose the physics behind it to the end user. On the other hand, part implementation with an NN preserves the background phenomena

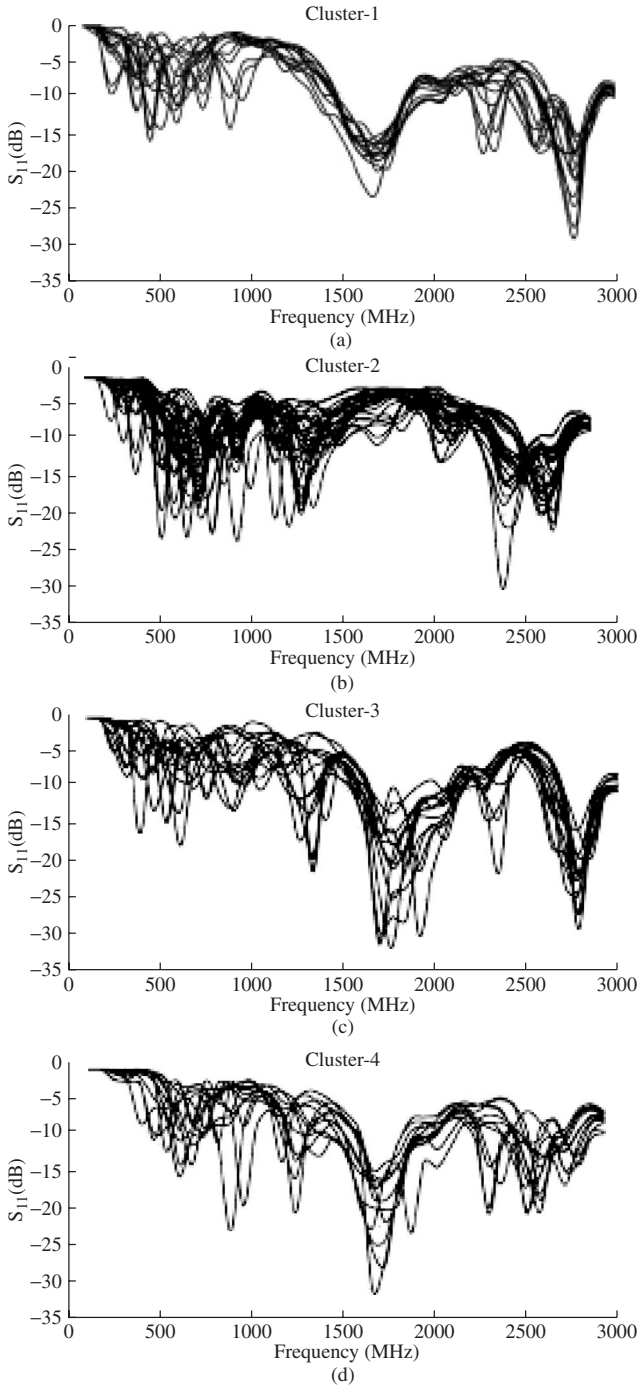


Figure 33.16 Clusters of the frequency responses of the reconfigurable structures as made by the SOM network. (From Ref. 26. Reprinted with permission, copyright © 2005 IEEE.)

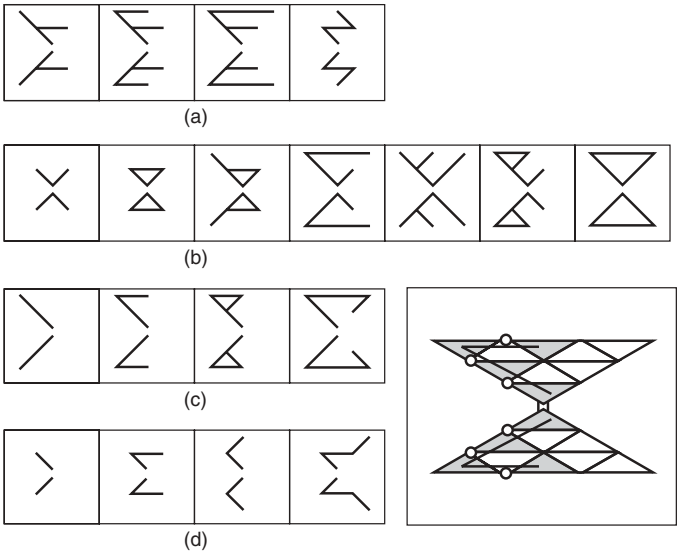


Figure 33.17 Reconfigurable structures corresponding to clusters 1–4 as shown in (a)–(d), respectively. The inset picture shows the formation of paths for a typical configuration. Structures corresponding to these or their variants can be used to get the desired frequency response of the reconfigurable antenna. (From Ref. 26. Reprinted with permission, copyright © 2005 IEEE.)

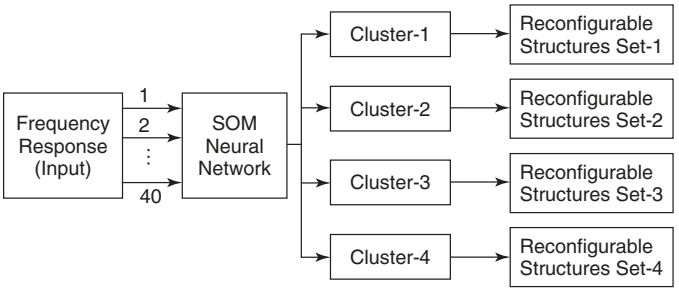


Figure 33.18 Paradigm of reconfigurale structure design.

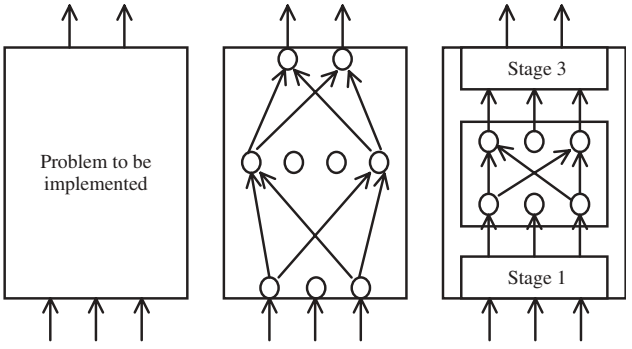


Figure 33.19 The problem in hand can be modeled using NNs as a whole, where it works as a black-box, or the NN can be used to model a part of the whole problem.

of the problem to some extent. The knowledge-based NN also preserves the background physics of the problem as well. In some cases the existing prior knowledge is used to train the network [28].

3. The accuracy of a properly trained network depends on the accuracy of the data used to train the network. Therefore care should be taken while generating training data, whether it is simulated or experimentally generated data. Sometimes preprocessing of input and output data reduces the training time of the network to a large extent. Effective data representation is another step in this direction.
4. With the increase in network size, the number of training patterns required for proper generalization also increases. Because the generation of data in RF/microwave problems is very expensive, it is desirable to develop the network with the minimum number of neurons in the hidden layer(s) (the number of input and output layer neurons are problem dependent and fixed), at the same time avoiding overtraining and undertraining. For this, generated data may be divided into training and test sets for observing the error behavior with the progress of training. As a rule of thumb, the minimum number of hidden layer neurons required to avoid overtraining and to obtain good accuracy for testing data is $\sqrt{[\text{Number of input layer neurons}) \times (\text{Number of output layer neurons})]}$.

33.8 DISADVANTAGES OF USING NEURAL NETWORKS

The conventional approach to ANN-based design described previously, which is also known as the "black-box" approach, has the following three main disadvantages:

1. Time required to generate sufficient training, validation, and testing samples.
2. Unreliability of the optimal solution when it lies outside the training region (due to the poor extrapolation performance of ANNs).
3. "Curse of dimensionality," which refers to the fact that the number of learning samples needed to approximate a function grows exponentially with the ratio between the dimensionality and its degree of smoothness.

The conventional neural optimization approach is indeed very suitable when the device's physics is not fully understood (i.e., when there is no empirical model available for the device), but the device's outputs for specified inputs are available either from measurement or from accurate simulations. On the other hand, an important advantage of the conventional neural optimization approach is its adequacy for full automation.

33.9 NEW TRENDS

ANNs have also been employed in improving the efficiency of computational methods used in the design of complex antenna structures. Since NNs can be implemented on high performance computers or fast hardware such as ASICS or FPGAS, they can yield more computationally efficient electromagnetic modeling tools. ANNs have been used in conjunction with finite-element methods (FEMs) [49–54] to solve several problems in electromagnetics and wave propagation. ANNs have also been used very successfully

with the finite-difference time-domain method [55–57] to solve EM scattering and EMC problems. Finally, ANNs have also been used in conjunction with the method of moments [58, 59] to solve for the radiation patterns of arbitrarily shaped microstrip antennas.

33.10 SUMMARY

Neural networks are a computing paradigm that involves several, simple processing elements that can communicate with each other, originally developed to mimic the action of neurons in the human brain. These connections form a massively parallel network used to perform tasks that would otherwise be very difficult. Each connection is assigned a numerical weight that is modified through training to meet the particular task, such as finding the resonant frequency of the antenna or the size of the antenna that will yield a specific resonance or radiation pattern. Given certain information about the antenna through either simulation or measured data, the neural network can be used to design the antenna for a desired outcome.

Once the network is trained off-line, it can then "generalize" other cases in real time. That means, given a certain desired antenna radiation pattern, resonance, or input impedance, for example, the neural network can determine the appropriate type of aperture or size of antenna. In array antennas, one can give a certain radiation pattern and the neural network can define the number of elements and their complex excitation required to yield such radiation pattern.

Several antenna examples that can be designed using neural networks have been presented and discussed.

REFERENCES

1. F. Wang, V. K. Devabhaktuni, C. Xi, and Q. J. Zhang, Neural network structures and training algorithms for microwave applications, *Int. J. RF Microwave Computer-Aided Eng.*, Vol. 9, pp. 216–240, 1999.
2. C. G. Christodoulou and M. Georgiopoulos, *Applications of Neural Networks in Electromagnetics*, Artech House, Boston, 2001.
3. K. Hornik, M. Stinchcombe, and H. White, Multilayer feedforward networks are universal approximators, *Neural Networks*, Vol. 2, pp. 359–366, 1989.
4. V. K. Devabhaktuni, M. Yagoub, and Q. J. Zhang, A robust algorithm for automatic development of neural network models for microwave applications, *IEEE Trans. Microwave Theory Tech.*, Vol. 49, pp. 2282–2291, December 2001.
5. T. Y. Kwok and D. Y. Yeung, Constructive algorithms for structure learning in feedforward neural networks for regression problems, *IEEE Trans. Neural Networks*, Vol. 8, pp. 630–645, May 1997.
6. J. de Villiers and E. Barnard, Backpropagation neural nets with one and two hidden layers, *IEEE Trans. Neural Networks*, Vol. 4, pp. 136–141, January 1992.
7. S. Haykins, *Neural Networks: A Comprehensive Foundation*, IEEE Press/IEEE Computer Society Press, New York, 1994.
8. Q. J. Zhang, K. C. Gupta, and V. K. Devabhaktuni, Artificial neural networks for RF and microwave design—from theory to practice, *IEEE Trans. Microwave Theory Tech.*, Vol. 51, No. 4, pp. 1339–1350, April 2003.
9. T. Kohonen, The self-organizing map, *Proc. IEEE*, Vol. 78, No. 9, pp. 1464–1480, 1990.

10. S. Sagiroglu and K. Guney, Calculation of resonant frequency for an equilateral triangular microstrip antenna with the use of artificial neural networks, *Microwave Optical Technol. Lett.*, Vol. 14, No. 2, pp. 89–93, February 1997.
11. J. S. Dahele and K. F. Lee, On the resonant frequencies of the triangular microstrip patch antennas, *IEEE Trans. Antennas Propag.*, Vol. 35, No. 1, pp. 100–101, 1987.
12. W. Chen, K. F. Lee, and J. S. Dahele, Theoretical and experimental studies of the resonant frequencies of the equilateral triangular microstrip antenna, *IEEE Trans. Antennas Propag.*, Vol. 40, No. 10, pp. 1253–1256, 1992.
13. G. Fedi, S. Manetti, G. Pelosi, and S. Selleri, Profile corrugated circular horns analysis and synthesis via an artificial neural network, *IEEE Trans. Antennas Propag.*, Vol. 49, No. 11, pp. 1597–1602, November 2001.
14. G. Washington, Aperture antenna shape prediction by feedforward neural networks, *IEEE Trans. Antennas Propag.*, Vol. 45, No. 4, pp. 683–688, April 1997.
15. N. P. Somasiri, X. Chen, and A. A. Rezazadeh, Neural network modeler for design optimization of multilayer patch antennas, *Proc. IEE—Microwaves Antennas Propag.*, Vol. 151, No. 6, pp. 514–518, December 2004. On-Line No. 20040862.
16. S. Sagiroglu, K. Guney, and M. Erler, Resonant frequency calculation for circular microstrip antennas using artificial neural networks, *Int. J. RF Microwave Computer-Aided Eng.*, Vol. 8, pp. 270–277, 1998.
17. S. Sagiroglu, K. Guney, and M. Erler, Calculation of bandwidth for electrically thin and thick rectangular microstrip antennas with the use of multilayered perceptrons, *Int. J. RF Microwave Computer-Aided Eng.*, Vol. 9, pp. 277–286, 1999.
18. D. Karaboga, K. Guney, S. Sagiroglu, and M. Erler, Neural computation of resonant frequency of electrically thin and thick rectangular microstrip antennas, *Proc. IEE—Microwaves Antennas Propag.*, Vol. 146, No. 2, pp. 155–159, April 1999.
19. K. Guney, M. Erler, and S. Sagiroglu, Artificial neural networks for the resonant resistance calculation of electrically thin and thick rectangular microstrip antennas, *Electromagnetics*, Vol. 20, pp. 387–400, 2000.
20. K. Guney, S. Sagiroglu, and M. Erler, Comparison of neural networks for resonant frequency computation of electrically thin and thick rectangular microstrip antennas, *J. Electromagn. Waves Applications*, Vol. 15, pp. 1121–1145, 2001.
21. K. Guney, S. Sagiroglu, and M. Erler, Design of rectangular microstrip antenna with the use of artificial neural networks, *Neural Network World*, Vol. 4, pp. 361–370, 2002.
22. K. Guney, S. Sagiroglu, and M. Erler, Generalized neural method to determine resonant frequencies of various microstrip antennas, *Int. J. RF Microwave Computer-Aided Eng.*, Vol. 12, pp. 131–139, 2002.
23. K. Guney and N. Sarikaya, Artificial neural networks for calculating the input resistance of circular microstrip antennas, *Microwave Optical Technol. Lett.*, Vol. 37, pp. 107–111, 2003.
24. R. K. Mishra and A. Patnaik, Design of circular microstrip antenna using neural network, *IETE J. of Res.*, Vol. 44, No. 1 & 2, pp. 35–39, 1998.
25. R. K. Mishra and A. Patnaik, Neural network based CAD model for design of square patch antenna, *IEEE Trans. Antennas Propag.*, Vol. 46, No. 12, pp. 1890–1891, 1998.
26. A. Patnaik, D. Anagnostou, C. G. Christodoulou, and J. C. Lyke, Neurocomputational analysis of a multiband reconfigurable planar antenna, *IEEE Trans. Antennas Propag.*, Vol. 53, No. 11, pp. 3453–3458, November 2005.
27. A. Patnaik, D. Anagnostou, R. K. Mishra, C. G. Christodoulou, and J. C. Lyke, Applications of neural networks in wireless communications, *IEEE Antennas Propag. Mag.*, Vol. 46, No. 3, pp. 130–137, June 2004.

28. P. M. Watson, G. L. Creech, and K. C. Gupta, Knowledge based EM-ANN models for the design of wide bandwidth CPW patch/slot antennas, *IEEE Int. Symp. Antennas Propag. Dig.*, Vol. 4, pp. 2588–2594, 1999.
29. H. L. Southall, J. A. Simmers, and T. H. O'Donnell, Direction finding in phased arrays with a neural beamformer, *IEEE Trans. Antennas Propog.*, Vol. 43, No. 12, pp. 1369–1374, 1995.
30. A. H. El Zooghby, C. G. Christodoulou, and M. Georgiopoulos, Performance of radial-basis function networks for direction of arrival estimation with antenna arrays, *IEEE Trans. Antennas Propag.*, Vol. 45, No. 11, pp. 1611–1615, 1997.
31. E. Charpentier and J.-J. Laurin, An implementation of a direction-finding antenna for mobile communications using a neural network, *IEEE Trans. Antennas Propag.*, Vol. 47, No. 7, pp. 1152–1159, July 1999.
32. A. H. El Zooghby, C. G. Christodoulou, and M. Georgiopoulos, A neural network-based smart antenna for multiple source tracking, *IEEE Trans. Antennas Propag.*, Vol. 48, No. 5, pp. 768–776, May 2000.
33. C.-S. Shieh and C.-T. Lin, Direction of arrival estimation based on phase differences using neural fuzzy network, *IEEE Trans. Antennas Propag.*, Vol. 48, No. 7, pp. 1115–1124, July 2000.
34. S. Jha and T. S. Durrani, Direction of arrival estimation using artificial neural networks, *IEEE Trans. Syst. Man Cybern.*, Vol. 21, No. 5, pp. 1192–1201, September/October 1991.
35. T. Lo, L. Henry, and L. John, Radial basis function neural network for direction of arrival estimation, *IEEE Signal Processing Lett.*, Vol. 1, pp. 45–47, February 1994.
36. P. R. Chang, W. H. Yang, and K. K. Chan, A neural network approach to MVDR beamforming problem, *IEEE Trans. Antennas Propag.*, Vol. 40, No. 3, pp. 313–322, 1992.
37. A. H. El Zooghby, C. G. Christodoulou, and M. Georgiopoulos, Neural network-based adaptive beamforming for one- and two-dimensional antenna arrays, *IEEE Trans. Antennas Propag.*, Vol. 46, No. 12, pp. 1891–1893, 1998.
38. E. Kuhn and V. Hombach, Computer-aided analysis of corrugated horns with axial or ring-loaded radial slots, in *Institution Electrical Engineers ICAP*, 1983, pp. 127–131.
39. I. Wolf and N. Knoppik, Rectangular and circular microstrip disk capacitors and resonators, *IEEE Trans. Microwave Theory Tech.*, Vol. 22, pp. 857–864, October 1974.
40. K. C. Gupta, J. Li, R. Ramadoss, C. Wang, Y. C. Lee, and V. M. Bright, Design of frequency-reconfigurable rectangular slot ring antennas, *Proc. IEEE Antennas Propagat. Int. Symp.*, Vol. 1, p. 326, July 2000.
41. M. A. Ali and P. Wahid, A reconfigurable Yagi array for wireless applications, *Proc. IEEE Antennas Propag. Int. Symp.*, pp. 466–468, June 2002.
42. J. C. Veihl, R. E. Hodges, D. McGrath, and C. Monson, Reconfigurable aperture decade bandwidth array, *Proc. IEEE Antennas Propag. Int. Symp.*, Vol. 1, pp. 314–317, July 2000.
43. J. Hazen, R. Clark, P. Mayes, and J. T. Bernhard, Stacked reconfigurable antennas for space-based radar applications, *Proc. Antenna Applications Symp.*, pp. 59–69, September 2001.
44. G. H. Huff, J. Feng, S. Zhang, and J. T. Bernhard, A novel radiation pattern and frequency reconfigurable single turn square spiral microstrip antenna, *IEEE Microwave Wireless Propag. Lett.*, Vol. 13, No. 2, pp. 57–59, February 2003.
45. L. N. Pringle, P. H. Harms, S. P. Blalock, G. N. Kiesel, E. J. Kuster, P. G. Friederich, R. J. Prado, J. M. Morris, and G. S. Smith, A reconfigurable aperture antenna based on switched links between electrically small metallic patches, *IEEE Trans. Antennas Propag.*, Vol. 52, No. 6, pp. 1434–1444, 2004.
46. E. R. Brown, RF-MEMS switches for reconfigurable integrated circuits, *IEEE Trans. Microwave Theory Tech.*, Vol. 46, No. 11, pp. 1868–1880, 1998.

47. W. H. Weedon, W. J. Payne, and G. M. Rebeiz, MEMS-switched reconfigurable antennas, *Proc. IEEE Antennas Propag. Int. Symp.*, Vol. 3, pp. 654–657, July 2001.

48. J. H. Schaffner, R. Y. Loo, D. F. Sevenpiper, F. A. Dolezal, G. L. Tangonan, J. S. Colburn, J. J. Lynch, J. J. Lee, S. W. Livingston, R. J. Broas, and M. Wu, Reconfigurable aperture using RF MEMS switches for multi-octave tenability and beam steering, *Proc. IEEE Antennas Propag. Int. Symp.*, Vol. 1, pp. 321–324, July 2000.

49. I. E. Lagaris, A. L. Likas, and D. Papageorgiou, Neural-network methods for boundary value problems with irregular boundaries, *IEEE Trans. Neural Networks*, Vol. 11, No. 5, pp. 1041–1049, 2000.

50. J. Takeuchi and Y. Kosugi, Neural network representation of finite element method, *Neural Networks*, Vol. 7, No. 2, pp. 389–395, 1994.

51. L. Ziemianski, Hybrid neural network/finite element modeling of wave propagation in infinite domains, *Computers Structures*, Vol. 81, pp. 1099–1109, 2003.

52. H. Yamashita, N. Kowata, V. Cingoski, and K. Kaneda, Direct solution method for finite element analysis using a Hopfield neural network, *IEEE Trans. Magn.*, Vol. 31, pp. 1964–1967, 1995.

53. F. Guo, P. Zhang, F. Wang, X. Ma, and G. Qiu, Finite element analysis based on Hopfield neural network model for solving nonlinear electromagnetic field problems, *IEEE Int. Joint Conf. Neural Networks*, Vol. 6, pp. 4399–4403, 1999.

54. D. D. Dyck and D. A. Lowther, Determining an approximate finite element mesh density using neural network techniques, *IEEE Trans. Magn.*, Vol. 28, No. 2, pp. 1767–1770, 1992.

55. J. Litva, C. Wu, and E. Navarro, On incorporating signal processing and neural networks techniques with the FDTD method for solving electromagnetic problems, *Int. Symp. Signals Systems Electron. (ISSE'95)*, pp. 327–330, 1995.

56. C. Wu, A. Navarro, and J. Litva, Combination of finite impulse response neural network technique with FDTD method for simulation of electromagnetic problems, *Electron. Lett.*, Vol. 32, No. 12, pp. 1112–1113, 1996.

57. S. Goasguen, S. M. Hammadi, and S. M. El-Ghazali, A global modeling approach using artificial neural networks, *IEEE MTT-S Dig.*, Vol. 1, pp. 153–156, 1999.

58. L. Vegni and A. Toscano, Analysis of microstrip antennas using neural networks, *IEEE Trans. Magn.*, Vol. 33, No. 2, pp. 1414–1417, 1997.

59. P. M. Piel et al., Fast hybrid integral equation–neural network method for the modeling of multiconductor transmission lines, *IEEE MTT-S*, pp. 11673–1676, 1999.

INDEX

Modern Antenna Handbook. Edited by Constantine A. Balanis