

RF Systems, Components, and Circuits Handbook

Second Edition

For a listing of recent titles in the *Artech House Microwave Library*,
turn to the back of this book.

RF Systems, Components, and Circuits Handbook

Second Edition

Ferril A. Losee



**ARTECH
HOUSE**

BOSTON | LONDON
artechhouse.com

Library of Congress Cataloging-in-Publication Data

Losee, Ferril.

RF systems, components, and circuits handbook/Ferril A. Losee.—2nd ed.

p. cm.—(Artech House microwave library)

Includes bibliographical references and index.

ISBN 1-59693-010-1 (alk. paper)

1. Radio circuits—Design and construction—Handbooks, manuals, etc. 2. Radio frequency integrated circuits—Design and construction—Handbooks, manuals, etc.

I. Title. II. Series

TK6560.L65 2005

621.384—dc22

2005048170

British Library Cataloguing in Publication Data

Losee, Ferril

RF systems, components, and circuits handbook.—2nd ed.—(Artech House microwave library)

1. Radio circuits—Design and construction—Handbooks, manuals, etc.

I. Title

621.3'84

ISBN-10 1-59693-010-1

Cover design by Igor Valdman

© 2005 ARTECH HOUSE, INC.

685 Canton Street

Norwood, MA 02062

All rights reserved.

Printed and bound in the United States of America. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.

All terms mentioned in this book that are known to be trademarks or service marks have been appropriately capitalized. Artech House cannot attest to the accuracy of this information. Use of a term in this book should not be regarded as affecting the validity of any trademark or service mark.

International Standard Book Number: 1-59693-010-1

10 9 8 7 6 5 4 3 2 1

Contents

Preface	<i>xvii</i>
PART I	
RF Systems	1
CHAPTER 1	
Telephone Systems	3
1.1 Introduction	3
1.2 Telephone Lines	4
1.2.1 Open-Wire Lines	4
1.2.2 Twisted-Pair Lines	4
1.2.3 Coaxial Cable	4
1.2.4 Cordless Telephones	5
1.2.5 2.4-GHz Cordless Telephones	6
1.2.6 5.8-GHz Digital Spread Spectrum (DSS) Cordless Telephones	6
1.3 Telephone Relay Systems	7
1.3.1 Microwave Relay Systems	7
1.3.2 Fiber Optics Cable Telephone Relay Systems	8
1.3.3 Submarine Cable Relay Systems	10
1.3.4 Communication Satellite Relay Systems	11
1.4 Cellular Telephone Systems	12
1.4.1 Introduction to Cellular Telephones	12
1.4.2 The Concept of Spatial Frequency Reuse	13
1.4.3 Propagation Characteristics of Cellular Telephone Systems	14
1.4.4 Types of Cellular Telephone Systems	14
1.4.5 How a Cellular Telephone Call Is Made	18
1.5 Modern Commercially Available Cellular Telephone Systems	19
References	19
CHAPTER 2	
Wireless Communication Systems	21
2.1 Introduction	21
2.2 Background Information	21
2.2.1 Units and Conversion Information	21
2.2.2 Frequency Bands for Communication Systems	22

2.2.3	The Use of Decibels	22
2.3	Frequency Allocation and FCC Regulations	23
2.4	Types of Communication Services as Defined by the ITU	24
2.4.1	Aeronautical Mobile Service	24
2.4.2	Aeronautical Mobile Satellite Service	25
2.4.3	Amateur Service	25
2.4.4	Broadcasting Service	26
2.4.5	Broadcasting Satellite Service	27
2.4.6	Citizen Band Radio	27
2.4.7	VHF and UHF FM Business and Personal Two-Way Radio	28
2.4.8	Pager Systems	28
2.4.9	Mobile Service	29
2.4.10	Remote Control Systems	29
2.5	WLANs, IEEE 802.11 and Bluetooth	29
2.5.1	WLANs	29
2.5.2	IEEE 802.11 Standard	30
2.5.3	IEEE 802.11b Standard	30
2.5.4	IEEE 802.11g Standard	31
2.5.5	Bluetooth	31
	References	31

CHAPTER 3

	Radionavigation and Global Positioning Systems (GPS)	33
3.1	Older Radionavigation Systems	33
3.1.1	Omega System	33
3.1.2	Loran System	34
3.1.3	Radio Beacons and Airborne Direction Finders	35
3.1.4	VOR Systems	36
3.1.5	Distance-Measuring Equipment (DME)	37
3.1.6	Instrument Landing System (ILS)	37
3.1.7	Tactical Air Navigation	38
3.1.8	Microwave Landing System	39
3.1.9	Air Traffic Control Radar Beacon System	39
3.1.10	Transit System	39
3.2	GPS Navigation System	40
3.2.1	System Architecture	40
3.2.2	Applications for GPS	50
	References	51

CHAPTER 4

	Radar Systems	53
4.1	Basic Radar Concepts	53
4.2	Radars Frequencies	54
4.3	Types of Radars Based on Frequency	54
4.3.1	MF Radar (0.3–3.0 MHz)	54
4.3.2	HF Radar (3.0–30 MHz)	55
4.3.3	VHF Radar (30–300 MHz)	55

4.3.4	UHF Radar (300–1,000 MHz)	55
4.3.5	L-Band Radar (1.0–2.0 GHz)	55
4.3.6	S-Band Radar (2.0–4.0 GHz)	56
4.3.7	C-Band Radar (4.0–8.0 GHz)	56
4.3.8	X-Band Radar (8.0–12.0 GHz)	56
4.3.9	K_u -Band Radar (12.0–18.0 GHz)	56
4.3.10	K-Band Radar (18.0–27.0 GHz)	56
4.3.11	K_a -Band Radar (26.5–40.0 GHz)	57
4.3.12	V-Band Radar (40.0–75.0 GHz)	57
4.3.13	W-Band Radar (75–110 GHz)	57
4.3.14	Millimeter-Wave Radar (110–300 GHz)	57
4.4	Types of Radar	57
4.4.1	Surveillance Radars	57
4.4.2	CW Speed Measurement Radar	58
4.4.3	Airborne Weather-Avoidance Radar	58
4.4.4	Radar Altimeters	59
4.4.5	Airborne Doppler Navigation Radar	59
4.4.6	Ship-Based Search and Surveillance Radar	60
4.4.7	Shore-Based Search and Surveillance Radar	60
4.4.8	Space Applications of Radar	60
4.4.9	Ground-Based Instrumentation Radars for Locating and Tracking Missiles and Satellites	60
4.4.10	Airborne Military Multiple Function Radars	61
4.4.11	Airborne Terrain-Following Radar	61
4.4.12	Airborne Side-Looking Radar	61
4.4.13	Ground-Based Military Radar for Locating and Tracking Aircraft and Missiles	62
4.4.14	Ground-Based Military Radars for Ground-Based Targets	62
4.4.15	Ship-Based Military Radar	62
4.4.16	ICBM Defense Radars	62
4.5	Radar Measurement	63
4.5.1	Radar Cross-Section of Targets	63
4.5.2	Radar Clutter	65
4.5.3	Range Measurements	68
4.5.4	Velocity Measurement Using CW Radar	70
4.5.5	Velocity Measurements Using FMCW Radar	72
4.5.6	Velocity Measurements Using a Pulse-Type Radar	72
4.5.7	Angle Measurements for Radars	73
4.6	Moving-Target Indicator (MTI) and Pulse Doppler Radars	74
4.6.1	MTI Radar Types	74
4.6.2	Signal Processing for MTI Radars	76
4.6.3	MTI from a Moving Platform	79
4.7	Tracking Radars	80
4.7.1	Monopulse Tracking Radars	80
4.7.2	Tracking in Range Using Sequential Gating	81
	References	82

CHAPTER 5

Radio Frequency Propagation	83
5.1 Antennas	83
5.1.1 Transmit Antennas	83
5.1.2 Receive Antenna Gain and Capture Area	84
5.2 Electromagnetic Waves, Fields, and Power Density	85
5.3 RF Electrical Field Waveforms and Vector Addition	86
5.4 Free-Space Path Loss	87
5.5 Excess Path Loss and Atmospheric Attenuation	87
5.5.1 Atmospheric Absorption	87
5.5.2 Attenuation Produced by Rain, Snow, and Fog	88
5.6 Atmospheric Refraction	89
5.7 Diffraction of Radio Waves	90
5.8 Multipath	92
5.9 Ionospheric Propagation	93
5.10 Ground-Wave Propagation	95
5.11 Scatter Propagation	96
5.12 Fiber Optic Cable Propagation	97
5.13 Radar Cross-Section of Targets	97
5.14 Equations for Calculating Propagation Performance for Communication Systems	99
5.14.1 Example 1: HF Ionospheric Reflection Communication System	99
5.14.2 Example 2: VHF Base Station to Mobile Unit Communication System	100
5.14.3 Example 3: Microwave Uplink to Satellite Relay Located at Geostationary Orbit	100
5.15 Equations for Calculating Propagation Performance for Radar Systems	100
5.15.1 Example 4: L-Band Aircraft Surveillance Radar	101
5.15.2 Example 5: X-Band Airborne Multiple-Function Radar	102
References	102

CHAPTER 6

RF Noise and Link Analysis	103
6.1 Concepts of RF Noise and Signal-to-Noise Ratio	103
6.2 Noise Power, Noise Temperature, and Noise Figure	104
6.3 Multiple-Stage Systems with Noise	106
6.4 Types of Noise	107
6.4.1 Atmospheric Noise	107
6.4.2 Galactic Noise	108
6.4.3 Solar Noise	110
6.4.4 Ground Noise	111
6.4.5 Man-Made Noise and Interference	111
6.5 Signal-to-Noise Improvement by Use of Integration	112
6.6 Signal-to-Noise Ratio	113
6.7 Communication System Link Analysis	115

6.8	Radar System Link Analysis	116
6.9	Performance Calculations for Radar Systems with Electronic Countermeasures	117
	References	119
CHAPTER 7		
	Modulation Techniques	121
7.1	Pulsed Continuous-Wave Signals	121
7.2	Conventional Amplitude Modulation	122
7.3	Double Sideband Suppressed Carrier Modulation	123
7.4	Vestigial Sideband Modulation	125
7.5	Single-Sideband Modulation	125
7.6	Standard Frequency Modulation	126
7.7	Modulation for Telemetry	129
7.8	Combination Communication and Range-Measurement Systems	130
7.9	Modulation for Radar	130
	7.9.1 Pulsed CW Modulation	130
	7.9.2 High-Power Impulse Generators and Ultra-Wideband, High-Power Microwave Generators	132
	7.9.3 Chirp Pulse Modulation	133
	7.9.4 Phase Code Modulated Pulse Modulation	133
	7.9.5 Continuous-Wave Modulation	134
	7.9.6 Frequency-Modulated CW Modulation	134
7.10	Single-Channel Transmitter System	134
7.11	Frequency Division Multiplex Transmitter System	136
7.12	Sample Circuits and Analog-to-Digital Converter Concepts	137
7.13	Time Division Multiplex Transmitter System with Pulse Code Modulation	137
7.14	Two-State Modulation Types for Binary Signals	139
	7.14.1 On-Off or Two-State Amplitude Keying	139
	7.14.2 Frequency Shift Keying	140
	7.14.3 Binary Phase-Shift Keying	140
7.15	Four-State and Eight-State Phase-Shift Keying	141
7.16	Sixteen Phase-State Keying (16-PSK)	141
7.17	Sixteen Amplitude-Phase Keying	142
7.18	Direct Sequence Spread Spectrum Modulation	142
	7.18.1 Pseudorandom Noise (PN) Generators	142
	7.18.2 Direct Sequence Spread Spectrum Systems	144
	7.18.3 Applications for Direct Sequence Spread Spectrum	147
	References	148
	Selected Bibliography	148
CHAPTER 8		
	RF Amplifiers, Oscillators, Frequency Multipliers, and Mixers	149
8.1	Amplifiers	149
	8.1.1 Front-End Low-Noise RF Amplifiers for Receivers	149
	8.1.2 IF Amplifiers	155

8.1.3	Audio and Other LF Amplifiers	157
8.1.4	Transmitter RF Amplifier Chains	157
8.1.5	Transmitter RF Power Amplifiers for Communication Systems	158
8.1.6	RF Power Amplifiers and Oscillators for Radars, Navigation, and Electronic Countermeasure Applications	161
8.2	Oscillators and Frequency Synthesizers	163
8.2.1	Transistor Feedback Oscillators	163
8.2.2	Negative Resistance Two-Terminal Oscillators	165
8.2.3	Frequency Synthesizers	165
8.3	Frequency Multipliers	166
8.3.1	Varactor Diode Frequency Multipliers	166
8.3.2	Step-Recovery Diode Frequency Multipliers	167
8.3.3	Transistor Multipliers	168
8.4	Mixers	169
8.4.1	Diode Mixers	169
8.4.2	Transistor Mixers	174
	References	176
	Selected Bibliography	176

CHAPTER 9

	Modulators and Demodulators	177
9.1	Modulators	177
9.1.1	Modulators for Conventional Amplitude Modulation	177
9.1.2	Modulators for Double-Sideband Modulation	179
9.1.3	Vestigial-Sideband Modulators	181
9.1.4	Modulators for Single-Sideband Modulation	181
9.1.5	Modulators for Frequency-Division Multiplex	184
9.1.6	Modulators for Standard Frequency Modulation	184
9.1.7	Modulators for Frequency-Shift Keying	187
9.1.8	Modulators for Phase-Shift Keying	187
9.1.9	Modulators for Pulse Code Modulation Time-Division Multiplex Modulation	190
9.1.10	Time-Division Multiple Access	191
9.2	Demodulators or Detectors	191
9.2.1	Amplitude Modulation Detectors	192
9.2.2	Product Detectors	192
9.2.3	Frequency Modulation Detector Concepts	194
9.2.4	Phase Detectors	197
	References	198

CHAPTER 10

	Example Communication System Block Diagrams	201
10.1	HF Communication System Using Single-Sideband Modulation	201
10.2	VHF or UHF Ground-to-Air Communication System Using Either Amplitude Modulation or Narrowband FM	203
10.3	Frequency Modulation Broadcast Systems	205
10.4	Microwave Relay Systems	208

10.5	Satellite Relay Communication Systems	211
10.6	Satellite Relay Earth Stations	215
	References	216

PART II

	RF Components and Circuits	217
--	----------------------------	-----

CHAPTER 11

	Transmission Lines and Transmission Line Devices	219
11.1	Two-Wire Transmission Lines	219
11.2	Coaxial Transmission Lines	220
11.3	Coaxial Cable Connectors	223
11.4	Microstrip Transmission Lines	224
11.5	Stripline Transmission Lines	225
11.6	Characteristics of Transmission Lines	227
	11.6.1 Wave Velocity on Transmission Lines	227
	11.6.2 Reflection Coefficients	227
	11.6.3 Standing-Wave Ratio	228
11.7	The Smith Chart	230
	11.7.1 Impedance and Admittance Coordinates	230
	11.7.2 Voltage Standing-Wave Ratio Circles	232
	11.7.3 Reflection Coefficients	232
	11.7.4 Examples Using the Smith Chart	232
11.8	Impedance Matching Using the Smith Chart	236
	11.8.1 Impedance Matching with a Quarter-Wave Transformer	236
	11.8.2 Impedance Matching with a Short-Circuited Stub	240
11.9	Coaxial Terminations	242
11.10	Coaxial Directional Couplers	242
11.11	Baluns	243
11.12	Two-Wire Transmission Line Impedance Transformer	244
11.13	Stripline and Microstrip Circuits	245
	11.13.1 Shunt Stub DC Returns	245
	11.13.2 Branch Line 90-Degree Hybrid Couplers	245
	11.13.3 Stripline or Microstrip Rat Race Hybrid Coupler	246
	11.13.4 Split Inline Hybrid Dividers and Combiners	247
	11.13.5 Quarter-Wave Coupled-Line Directional Couplers	248
	11.13.6 90-Degree Coupled-Line Hybrid Coupler	249
	11.13.7 Stripline Lowpass Filters	249
	11.13.8 Stripline Highpass Filters	250
	11.13.9 Stripline Bandpass Filters	251
11.14	Ferrite Circulators and Isolators	252
11.15	Coaxial Electromechanical Switches	254
11.16	PIN Diode Switches	256
11.17	Sparkgap Switches for Lightning Protection	257
	References	257
	Selected Bibliography	257

CHAPTER 12

Waveguides and Waveguide-Related Components	259
12.1 Introduction to Waveguides	259
12.2 Rectangular Waveguides	260
12.3 Higher-Order Modes in Rectangular Waveguides	266
12.4 Launching the TE_{10} Mode Using a Coaxial Line Input	268
12.5 Characteristic Wave Impedance for Waveguides	268
12.6 Other Types of Waveguides	270
12.6.1 Ridged Waveguides	270
12.6.2 Circular Waveguides	270
12.7 Waveguide Hardware	271
12.7.1 Waveguide Flanges	271
12.7.2 Rotary Joints	272
12.7.3 Tapered Transition Sections of Waveguides	272
12.7.4 Flexible Waveguides	272
12.7.5 Waveguide Accessories	273
12.8 Waveguide Hybrid Junctions	273
12.9 Waveguide Impedance Matching	275
12.10 Waveguide Resistive Loads and Attenuators	276
12.11 Waveguide Directional Couplers	276
12.12 Waveguide Ferrite Isolators, Circulators, and Switches	277
12.13 Waveguide Detectors and Mixers	279
12.14 Gas-Tube Switches	280
12.15 Duplexers	280
12.16 Cavity Resonators	283
References	285
Selected Bibliography	285

CHAPTER 13

Antennas	287
13.1 Monopole Antennas	289
13.1.1 Thin-Wire Monopole Antennas	289
13.1.2 Wideband Monopoles	290
13.1.3 Impedance of Monopole Antennas	290
13.1.4 Large-Size Monopole Antennas	291
13.1.5 Electrically Small Monopole Antennas	291
13.2 Dipole Antennas	292
13.2.1 Thin-Wire Dipole Antennas	292
13.2.2 Other Types of Dipole Antennas	294
13.2.3 Dipole Impedance	294
13.2.4 Dipole Current Distribution and Antenna Patterns for Different L/λ Ratios	296
13.2.5 Turnstile Antenna	297
13.3 Yagi-Uda Antennas	299
13.4 Sleeve Antennas	301
13.4.1 Sleeve Monopoles	301
13.4.2 Sleeve Dipoles	302

13.5	Loop Antennas	302
13.5.1	Air-Core Loop Antennas	302
13.5.2	Ferrite-Core Loop Antennas	304
13.6	Helical Antennas	304
13.7	Spiral Antennas	306
13.7.1	Equiangular Spiral Antennas	306
13.7.2	Archimedean Spiral Antennas	307
13.7.3	Conical Spiral Antennas	308
13.8	Log-Periodic Antennas	308
13.8.1	Log-Periodic Dipole Array	308
13.8.2	Trapezoidal-Toothed Log-Periodic Antennas	311
13.8.3	Triangular-Toothed Log-Periodic Antennas	313
13.9	Slot Antennas	313
13.9.1	Open-Slot Antennas	313
13.9.2	Cavity-Backed Rectangular-Slot Antennas	315
13.9.3	Waveguide-Fed Slot Antennas	315
13.10	Notch Antennas	318
13.11	Horn Antennas	318
13.12	Lens Antennas	320
13.12.1	Dielectric Lens Antenna	321
13.12.2	Luneburg Lens Antenna	321
13.12.3	Metallic-Plate Lens Antenna	322
13.13	Antenna Arrays	322
13.13.1	End-Fire Line Antenna Arrays	322
13.13.2	Broadside Line Antenna Arrays	324
13.14	Planar Arrays	326
13.15	Scanning Methods	326
13.15.1	Mechanically Scanned Arrays	326
13.15.2	Arrays with Space Feeds	329
13.16	Flat-Plate Reflector Type Antennas	330
13.16.1	Half-Wave Dipole Antennas with Reflectors	330
13.16.2	Corner Reflector Antennas	330
13.17	Parabolic Reflector Antennas	330
13.18	Patch Antennas	333
13.18.1	Introduction	333
13.18.2	Types of Patch Antennas	334
13.18.3	Antenna Feeds	335
13.18.4	Broadbanding	336
13.18.5	SSFIP Antennas	336
13.18.6	Dual Polarized and Circularly Polarized Patch Antennas	337
13.18.7	Patch Antenna Arrays	338
	References	339
CHAPTER 14		
	Lumped Constant Components and Circuits	341
14.1	Conductors and Skin Effect	341
14.2	RF Resistors	342

14.3	Inductors and Inductive Reactance	342
14.4	RF Chokes	346
14.5	Capacitors and Capacitive Reactance	346
14.6	Series Resonant RLC Circuits	349
14.7	Parallel Resonant RLC Circuits	350
14.8	Complex Resonant Circuits	352
14.9	The Use of the Smith Chart for Circuit Analysis	355
14.10	S-Parameters	357
14.11	Impedance Matching Using LC Circuits	358
14.12	Impedance-Matching Design Using the Smith Chart	360
14.13	LC Filters	365
	References	368
CHAPTER 15		
	RF Transformer Devices and Circuits	369
15.1	Conventional Transformers	369
15.2	Magnetic Core Material for RF Transformers	371
15.3	Tuned Transformers	373
15.4	High-Frequency Wideband Conventional RF Transformers	374
15.5	Transmission-Line Transformers	378
15.6	Power Combiners and Splitters	381
	References	382
CHAPTER 16		
	Piezoelectric, Ferrimagnetic, and Acoustic Devices and Circuits	383
16.1	Quartz Crystal Resonators and Oscillators	383
16.2	Monolithic Crystal Filters	386
16.3	Ceramic Filters	387
16.4	Dielectric Resonant Oscillators	388
	16.4.1 Dielectric Resonator Description and Parameters	388
	16.4.2 Coupling Between a Dielectric Resonator and a Microstrip Line	390
	16.4.3 Mechanical and Electrical Tuning of Dielectric Resonators	391
	16.4.4 Examples of Dielectrically Stabilized Oscillators	391
16.5	YIG Resonators and Filters	391
	16.5.1 Ferrimagnetic Resonance in Yttrium Iron Garnet Crystals	391
	16.5.2 YIG Bandpass Filters	393
	16.5.3 YIG-Tuned Oscillators	393
16.6	Surface Acoustic Wave Delay Lines	393
	16.6.1 Nondispersive Delay Lines	393
	16.6.2 Tapped Delay Lines	395
	16.6.3 Dispersive Delay Lines	396
16.7	Surface Acoustic Wave Delay Line Oscillators	399
16.8	Bulk Acoustic Wave Delay Lines	400
	References	402

CHAPTER 17

Semiconductor Diodes and Their Circuits	403
17.1 Semiconductor Materials	403
17.2 “Ordinary” Junction Diodes	403
17.3 Zener Diodes	404
17.4 Schottky-Barrier Diodes	404
17.5 PIN Diodes	405
17.6 Varactor Diodes	406
17.7 Step-Recovery Diodes	407
17.8 Microwave Tunnel Diodes and Circuits	408
17.9 Microwave Gunn Diodes and Circuits	408
17.10 Microwave IMPATT Diodes	410
17.11 Semiconductor IR Laser Diodes	412
17.12 Light-Emitting Diodes	413
17.13 IR Photodiodes	413
References	414

CHAPTER 18

Bipolar and Field-Effect Transistors and Their Circuits	415
18.1 Bipolar Junction Transistors	415
18.2 BJT Amplifier Configurations	417
18.2.1 Common-Emitter Amplifier	417
18.2.2 Common-Base Amplifier	422
18.2.3 Common-Collector Amplifier	424
18.3 Field Effect Transistors and Circuits	426
18.3.1 Junction Field-Effect Transistors	426
18.3.2 Metal-Semiconductor Field-Effect Transistors	428
18.3.3 Metal-Oxide Semiconductor Field Effect Transistors	429
18.4 Comparison of FET and BJT Amplifiers	431
18.5 High Electron Mobility Transistors and Heterojunction Bipolar Transistors	433
18.5.1 High Electron Mobility Transistors	433
18.5.2 Heterojunction Bipolar Transistors	434
18.6 DC Bias Circuits for BJT Amplifiers	434
18.7 Bias Circuits for FET Amplifiers	436
18.8 Stability with BJT and FET Amplifiers	436
18.9 Impedance Matching	437
18.10 Design Methods with S-Parameters	438
18.11 Manufacturers’ Data Sheets For Transistors	440
References	443

CHAPTER 19

High-Power Vacuum Tube Amplifiers and Oscillators	445
19.1 Grid Tubes	445
19.1.1 Triode, Tetrode, and Pentode Vacuum Tubes	445
19.1.2 Grid-Type Vacuum Tube Amplifiers and Modulated Amplifiers	447

19.1.3	The Use of Cavities as Resonators for High-Power, High-Frequency Triodes and Tetrodes	450
19.2	Microwave Tubes and Circuits	451
19.2.1	Introduction to Microwave Tubes	451
19.2.2	Multiple-Cavity Klystron Amplifiers	453
19.2.3	Helix-Type Traveling-Wave Tube Amplifiers	455
19.2.4	Coupled-Cavity Traveling-Wave Tube Amplifiers	458
19.2.5	Conventional Magnetrons	458
19.2.6	Coaxial Cavity Magnetron	459
19.2.7	Amplitrons	461
19.2.8	Gyrotron Oscillators and Amplifiers	462
19.2.9	Circuit Configurations for Microwave Tubes	462
	References	464
	About the Author	465
	Index	467

Preface

There have been many changes in the RF field in recent years. Many of our most important RF systems, components, and circuits did not even exist 20 years ago when many of the potential readers of this book received their educations. For many who need this information, their training was in other fields such as computers or business. In many other cases, they have simply forgotten things that they learned about RF. For many who are now in school, there is both a desire and a need to learn about this important subject as completely and as easily as possible. For these and other reasons, I felt the need to provide a second edition to my first book *RF Systems, Components, and Circuits Handbook*, which was published in 1997. My goal has been to keep it simple and easy to read. The subject is interesting and the book should demonstrate this fact. My goal was to add at least 30% of new material and to be as up-to-date as possible. The coverage should remain comprehensive, including all of the important subject material of RF at least at an introductory level. I have assumed that in many cases the reader would have a fairly limited technical background and I wrote the text accordingly.

This handbook is in two parts. Part I, which comprises Chapters 1 through 10, covers RF systems. Important topics include telephone systems, wireless communication systems, global positioning systems (GPS), radar systems, radio frequency signal propagation, RF noise, signal modulation techniques, RF subsystems, modulators and demodulators, and example communication system block diagrams.

Part II, which comprises Chapters 11 through 19, covers RF components and circuits. Important topics include transmission lines and transmission line devices; waveguides and waveguide-related systems; antennas; lumped constant components and circuits; RF transformer devices and circuits; piezoelectric, ferromagnetic, and acoustic wave devices and circuits; semiconductor diodes and their circuits; bipolar and field-effect transistors and their circuits; vacuum tubes, and microwave tubes. These are a large number of topics, and the book is fairly comprehensive in its coverage.

One of the ways that I have tried to make this book interesting and easy to understand is to use many figures or illustrations. In this second edition there are 163 figures in Part I (RF Systems), and 196 figures in Part II (RF Components and Circuits), for a total of 359 figures for the book.

Another way that I have tried to make the book easy to read and understand by readers with limited technical background is to use only limited mathematical discussions. This is done without serious loss of needed exactness and completeness.

In this second edition, I have made an effort to better consolidate subject material. For example, nearly all of the material on telephone systems is found in

Chapter 1 rather than in a number of chapters. Nearly all of the material on radar is found in Chapter 4 rather than in a number of chapters. Chapter 3 presents new material on global positioning satellite (GPS) systems. Chapter 7 includes new material on direct-sequence spread spectrum systems, and Chapter 13 includes new material on microstrip patch antennas.

PART I

RF Systems

Telephone Systems

1.1 Introduction

This author believes that the most important communication service we have is the telephone. This is especially true since the development of cellular telephone systems. Conventional telephones are used not only for voice communication, but also for computer communications via modems, Internet access, fax transmissions, and other services. Additional discussion of cellular telephone systems is provided in Section 1.4. Much of the information presented in this chapter is quoted or adapted from [1–10].

Figure 1.1 shows the main elements of a telephone system. On the left are the customer location and a number of telephone sets. The sets are connected by wire lines to a protector block. A single twisted pair of small-gauge wire known as the *local loop* connects each customer telephone system to a central office. Local loops from customers are combined in a cable. A typical cable has 1,400 pairs of copper wires carried in a single plastic-sheathed cable of about 3 inches in diameter.

The central offices or central switching center in different communities are connected to each other by so-called trunk lines. These lines may be telephone wire lines, or they may be some other RF communication system such as microwave communication links, microwave relay systems, satellite relay systems, and cable systems.

Toll networks are provided for long-distance telephone, which is provided by such companies as AT&T, MCI, and Sprint rather than by the local telephone company. Links provided by these companies include microwave relay systems, satellite relay systems, land-based coax cable systems, undersea cable systems, and fiber optic cable systems. (Each of these types of links is discussed later.)

It is common to send many signals over each channel so that the total number of voice channels is much greater than the number of wires or other signal channels. This is done by a process known as *multiplexing*. Two main approaches have been used in the past for multiplexing telephone signals. These are frequency-division multiplexing and time-division multiplexing. The earlier multiplexing systems used for telephone were mainly frequency multiplexing. The author believes that the age of frequency-division multiplexing is now over. Digital or time-division multiplexing has proved superior in nearly all respects and is now quite widespread.

The centralized switching systems or telephone exchanges use computers for interconnections. Computers are also used to keep track of the calls made by each subscriber and to provide subscriber billing information.

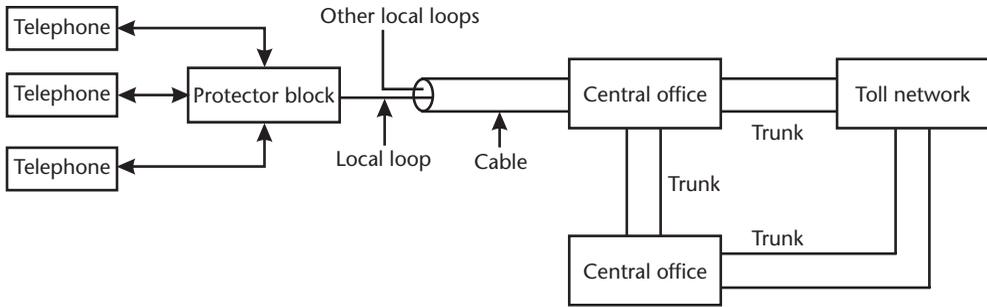


Figure 1.1 Main elements of a telephone system. (After: [2].)

1.2 Telephone Lines

1.2.1 Open-Wire Lines

The early telephone systems often used open-wire lines. These are still found in some rural areas. These open-wire lines consist of uninsulated bare copper wires strung on poles with insulators and separated by about 12 inches. In the early systems direct current was supplied to the telephone set through the two wires entering the set from the central office or switching center. A dc voltage of about 45V was normally used for communication with 75V dc used for ringing.

1.2.2 Twisted-Pair Lines

Twisted pair lines are now widely used in modern telephone systems. They consist of two individually insulated copper wires twisted together with a full twist about every 2 to 6 inches. Twisting the wires reduces interference to the lines from outside sources. Small gauge wires, such as 26 gauge (0.016-inch diameter) to 19 gauge (0.036-inch diameter), are typically used. Many twisted pairs may be combined into a single cable. These cables can be strung on poles, buried underground, or installed in a conduit. Since the telephonic speech signal has a required passband extending from approximately 300 Hz to 3300 Hz, the bandwidth requirement for a twisted pair is small.

1.2.3 Coaxial Cable

A large number of one-way voice circuits can be multiplexed together on a single coax cable. A number of 3/8 inch or smaller diameter cables typically are combined into a larger cable with an outside plastic, vinyl, or rubber coating. For example, the cable used with the AT&T L5 system has 11 pairs of coaxial cables enclosed in a single outer jacket. The characteristics and performance of the L5 are reported to be as follows:

Service date: 1978;

Technology: Integrated circuits;

Repeater spacing: 1.0 mile;

Capacity per 3/8 inch coax pair: 13,200 voice circuits;

Capacity per group (10 working pairs): 132,000 voice circuits.

1.2.4 Cordless Telephones

Figure 1.2 shows a cordless telephone system. The system includes a base unit connected to a two-wire telephone line and a handset that the user can carry several hundred feet from the base unit. The base unit is designed to hold the handset in a charging position when it is not in use. Figure 1.3 is a simplified system block diagram of one type of 49-MHz cordless telephone system. The base unit includes a small whip antenna, a low-powered (10 mW) FM transmitter, and an FM receiver. Full duplex operation is used so the transmitter and receiver are on slightly different frequencies in the 49-MHz band. A duplex system uses different frequencies for transmitting and receiving so it is possible to transmit and receive at the same time.

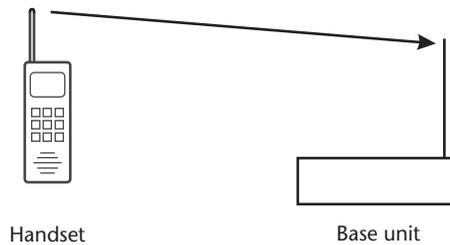


Figure 1.2 Cordless telephone system.

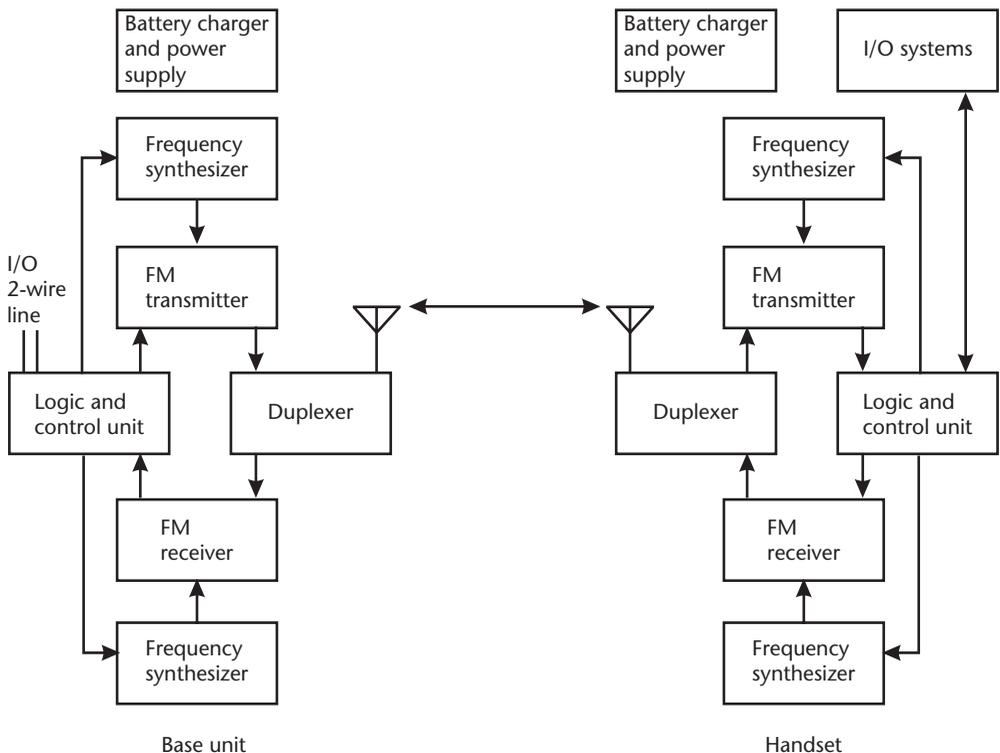


Figure 1.3 Simplified system block diagram of a cordless telephone system.

A simplex system uses the same frequency for transmit and receive, so it cannot transmit and receive simultaneously.

The base unit for the cordless telephone also includes a power supply, a battery charger for the handset, and a logic and control unit. The handset includes a rechargeable battery, a low (10 mW) FM transmitter, an FM receiver, a dialing unit, a speaker, a microphone and a logic and control unit.

The above example is provided in part to show the reader the marvelous degree of miniaturization that is possible with modern communication equipment through the use of digital techniques and large-scale integrated circuits. This is found even to a greater degree in cell phones and GPS systems that are discussed later.

When the first edition of this book was published in 1997, 49-MHz cordless telephones were frequently used, and 900-MHz cordless telephones were considered to be the top-of-the-line cordless telephone systems. Now such stores as RadioShack do not even sell 49-MHz cordless telephones, and 900-MHz systems are not frequently sold. The top-of-the-line cordless telephones are now 2.4-GHz and 5.8-GHz systems. Examples of these types of systems are discussed in the following sections.

1.2.5 2.4-GHz Cordless Telephones

There are two main types of 2.4 GHz cordless telephone system in use today. These are 2.4-GHz analog cordless phones and 2.4-GHz spread spectrum (DSS) cordless phones. The analog phones are the most popular cordless phones today, not only because of price, but also because of good quality and features that are included.

The digital spread spectrum (DSS) cordless phones provide excellent range, privacy, and clarity. The signal between the base and handset is digitally encrypted to ensure all conversations are secure. These phones also ensure incredible clarity by using multiple frequencies to transmit conversations across a very large bandwidth. Virtually all interference is eliminated.

1.2.6 5.8-GHz Digital Spread Spectrum (DSS) Cordless Telephones

The 5.8-GHz DSS cordless phones provide the greatest output power available in cordless telephone. It provides the maximum in range, privacy, and clarity available. It constantly searches for the clearest channel possible for the ultimate in security. It does not interfere with 802.11b, making it wireless network-friendly.

An example of a 5.8-GHz cordless telephone is the Panasonic model KX-TG5240M. The RadioShack listed cost for this system is \$149.99. The advertised product features are listed as follows:

- Wireless network friendly—will not interfere with wireless computer networks using 2.4-GHz technology;
- Add a phone to any room, even without a phone jack, expandable up to four handsets;
- 5.8-GHz technology for excellent clarity, range and privacy;
- Talking caller ID with 50-number memory and call waiting ID—talking caller ID announces the caller's name and number between rings using text-to-speech technology;

- Caller IQ compatible;
- All-digital answering system with 16 minutes of recording time;
- Duplex speakerphone on base and handsets;
- Lighted antenna visual ringer lets you see a call coming in, even if the ringer is turned off; also flashes slowly if you have a new message;
- Three-mailbox digital answering system;
- Voice enhancer technology helps clarify and improve sound reception, creating a natural-sounding voice that is easy to hear and understand;
- Battery provides up to 5 hours of talk time and up to 11 days of standby;
- Backlit four-line LCD on handset base for easier viewing of caller ID information;
- Two-digit LCD message counter on base;
- 50-station phonebook and dialer stores up to 32 digits per station so that you can add a long-distance access code;
- Use as desk phone or mount on wall;
- Headset jack;
- Base ringer;
- Voice scramble for security;
- Flash memory message backup;
- Pause and mute functions;
- Any-key answer;
- Selectable ringtones (three tones, four melodies) plus ringer melody download;
- Handset volume control;
- Ringer volume control;
- Lighted handset keypad for easy dialing even in the dark;
- Dual keypads—one on base and one on handset;
- Five-number memory redial.

One cannot help but be impressed with the large list of product features listed here. It shows that, in addition to great RF engineering in the development of this product, there has been great digital and mechanical engineering.

The information presented in this section is quoted or adapted from [9].

1.3 Telephone Relay Systems

1.3.1 Microwave Relay Systems

Microwave relay was once the backbone of the long-distance network and still is a fairly important technology. This type of system is illustrated in Figure 1.4. It uses radio towers for antennas spaced about 26 miles apart along the route on flat ground. Where possible, radio towers are located on hills to increase the distance for free space propagation without ground interference. A minimum system includes two antennas, one aimed in each of the two directions, with waveguide connected between the antennas and relay circuits. The most common types of

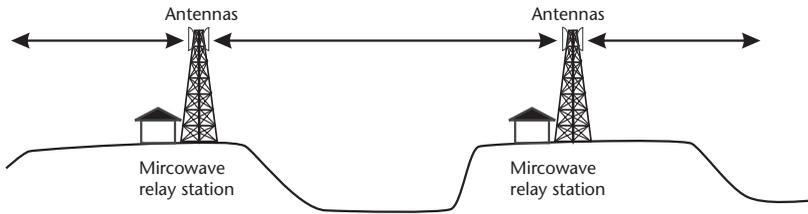


Figure 1.4 Two sections of a long-range microwave relay communication system.

antennas used are either horn antennas or parabolic dish reflector antennas, which are wideband antennas that may be dual polarized, permitting operation in each of two polarizations. That permits the capacity of the system to be doubled compared to a system with a single polarization.

Two microwave bands currently in use are 3.7–4.2 GHz (called the 4-GHz band) and 5.925–6.425 GHz (termed the 6-GHz band). Notice that the width of each band is 500 MHz. Each band is further subdivided into a number of channels. Channel widths are 20 MHz for the 4-GHz band and 30 MHz for the 6-GHz band.

As an indication of the performance capability of a microwave relay system for telephone service using frequency multiplex, consider the AR6A radio system. This system, introduced in 1981 by AT&T, used single-sideband, suppressed-carrier, microwave transmission, for the 6-GHz band. Six thousand voice channels could be placed in each radio channel. Thus, the seven active channels permitted 42,000 simultaneous two-way voice circuits. The AR6A was used in addition to the existing TD system, which used the 4-GHz band. The TD system had a capacity of 19,800 circuits. Combining the two systems, the total capacity, using the same antennas was 61,800 two-way voice circuits.

In the early 1980s, AT&T introduced digital radio systems that used time division multiplexing. Systems introduced included the DR6-30 system, the DR 11-40 system, the DR6-30-135 system, and the DR4-20-90 system. The DR6-30 operates in the 6-GHz band with channels that have 30-MHz bandwidth. The total capacity is 9,408 two-way digital voice circuits. The DR11-40 operates in the 11-GHz band with 40-MHz channels. The total capacity is 13,444 two-way digital voice channels. The DR6-135 system operates in the 6-GHz band and uses 64 quadrature amplitude modulation (64-QAM). It has a capacity of 14,112 two-way digital voice circuits. The DR11-40-135 system operates in the 11-GHz band and has a total system capacity of 20,160 two-way digital voice circuits.

The time-division multiplexed signal is transmitted using quadrature amplitude modulation (QAM), a technique that is a combination of phase and amplitude modulation. A 16-QAM has 16 possible states and represents 4 bits per baud. That involves 4 phase states and 4 amplitude states. A 64-QAM has 64 possible states and represents 6 bits per baud. That involves 8 amplitude states and 8 phase states.

1.3.2 Fiber Optics Cable Telephone Relay Systems

Fiber optic communication has become very important in recent years. Such systems may be used wherever coaxial cable or other transmission lines are used and may

replace other means of communications such as microwave relay systems. The main advantages for fiber optics cables include lower weight, lower cost, better security, and larger bandwidth. Infrared-emitting diodes are used as transmitters for fiber optic systems. Attenuation as a function of wavelength for silicon fibers is shown in Figure 1.5. This figure shows that for silicon fibers the lowest attenuation is just above and just below 1.4 microns, being only about 0.25 dB/km. The attenuation peak at 1.4 microns is a hydroxyl (OH) absorption band. Above about 1.6- μm wavelength, the attenuation for silicon fibers increases appreciably, being about 1 dB/km at 1.7 microns. Fibers based on fluoride and chalcogenide may exhibit ultra-low loss in the 2–5 μm wavelength range, possibly allowing repeaterless trans-oceanic fiber systems.

One type of optical fiber is a small-diameter cylindrical section of extremely pure glass with typical core diameter in the range of 2 to 10 μm . The optical fiber is covered with an outside cladding of glass of a slightly different chemical composition and a different refractive index. The outside cladding of glass is subsequently covered with a protective covering. This type of fiber is known as a *step index fiber*. It may be used for multimode or single-mode propagation. Single-mode propagation yields the best performance and the lowest attenuation. It is thus considered to be superior to the earlier graded-index fibers.

Figure 1.6 shows a block diagram of a fiber optic communication system. A diode laser is used to produce a modulated IR signal representing the input signal. This modulated beam is used as the input signal to an optical fiber. If the desired transmit distance is large, one or more power boosters or repeaters are spaced along the path at appropriate intervals to compensate for optic fiber cable loss. At the end of the line, an optical receiver is used to convert the IR signals to electrical signals, which are the output of the system.

The increase in the capacity of commercially available optical fiber communication systems is astonishing. AT&T's first fiber system, FT3, was introduced in 1979

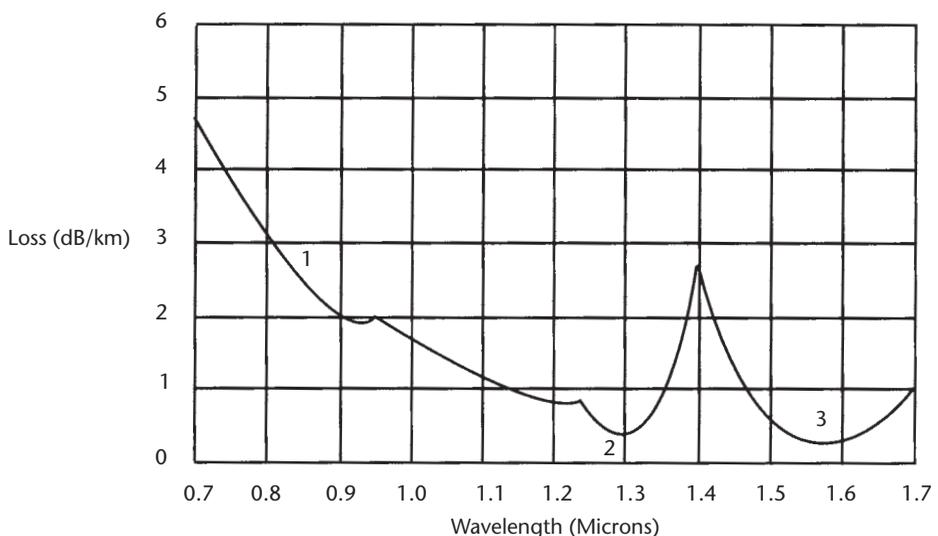


Figure 1.5 Attenuation as a function of wavelength for silicon fibers.

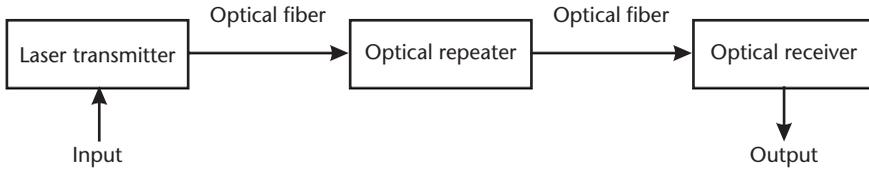


Figure 1.6 Fiber optics cable communication system.

and used graded-index fiber at 45 Mbps per fiber. Seventy-two fiber pairs were placed together in a single cable and 66 were active. The total route capacity of the system was 44,352 two-way digital voice circuits. The FT3C system was introduced soon afterward, with each fiber operating at 100 Mbps and carrying 1,344 digital voice circuits. Using cable with 66 active pairs, the system had a maximum route capacity of 88,704 two-way digital voice circuits. Repeaters were located every 4 miles along the route.

In 1984, AT&T introduced the FTX-180 system with each single-mode fiber operating at 180 Mbps and carrying 2,688 digital voice circuits. A 400-Mbps system introduced in 1968 had each single-mode fiber carrying 6,048 digital voice circuits and repeaters spaced every 20 miles.

Fiber optic systems operating at 2,000 Mbps are now standard. The capacities of today's fiber optics systems are mind-boggling. A 2,000-Mbps system has 31,250 voice circuits or 40 television signals per single-mode fiber.

1.3.3 Submarine Cable Relay Systems

Submarine cable lines are sometimes used for transoceanic communication. Such systems use power boosters spaced along the path at appropriate intervals to compensate for line losses experienced by the electrical signals. A system of this type is shown in Figure 1.7. Submarine cable systems have been around for a long time and have greatly improved in capacity with time. The first installed system had a design capacity of only 36 two-way voice circuits. That system, which was named TAT-1, had a service date of 1958. It used vacuum tube technology with repeaters spaced at intervals of 44 miles. The first system that used transistor technology was the

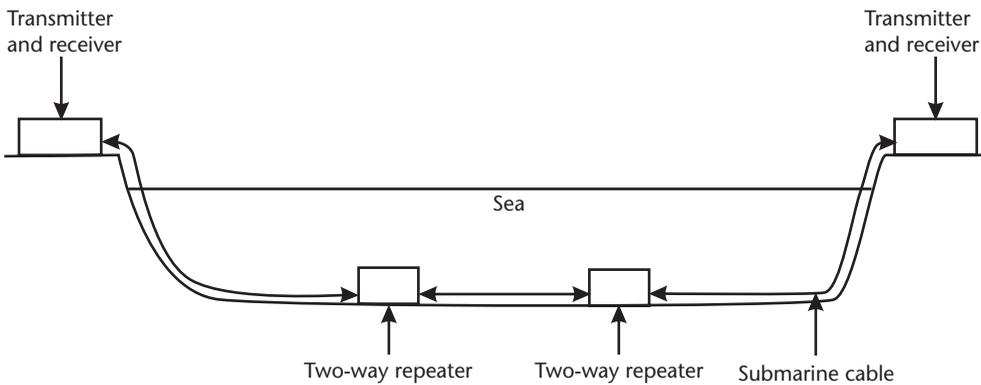


Figure 1.7 Submarine cable communication system.

TAT-5, having an in-service date of 1970. It had a design capacity of 845 two-way voice circuits. Repeaters for the TAT-5 system were spaced 12 miles apart. The TAT-7, with a service date of 1983, also used transistor technology. It had a design capacity of 4,200 two-way voice circuits. Repeaters for that system were spaced only 6 miles apart.

The first transatlantic submarine cable-system using fiber optics was the TAT-8, with a service date of 1988. It used three fiber pairs, diode lasers, and integrated circuit technology. It has a design capacity of 8,000 two-way digital voice circuits. That capacity can be extended using a technique called *time assignment speech interpolation* (TASI). TASI uses the silent intervals in speech conversation to carry signals from other speech conversations. The maximum capacity for the TAT-8 using TASI was 40,000 voice circuits. Repeaters for the TAT-8 system are spaced 41 miles apart.

The TAT-9 system had a service date of 1991. It is a fiber optic system with a maximum capacity of 80,000 two-way voice circuits. Repeaters are spaced every 75 miles. The large capacity and reliability of fiber optics make it the transmission medium of choice. The cost of international telephone calls will continue to decrease as capacity continues to increase through the use of fiber optics.

1.3.4 Communication Satellite Relay Systems

Satellite relay systems are used extensively for relaying telephone signals for long-distance telephone systems. Most of the fixed satellite services use geostationary orbit satellites with orbit altitudes of about 35,800 km and equatorial orbit planes. At that altitude, the orbital period is 24 hours, so the satellite appears stationary to observers on the surface of the earth. Satellite communication systems of this type typically use high-gain, parabolic dish antennas at the ground stations and high power transmitters for the ground-based transmitter.

The earliest frequency bands used for satellite relay were the same 4-GHz and 6-GHz bands used for terrestrial microwave transmission. The 4-GHz band (3.7–4.2 GHz) was used for the downlink, and the 6-GHz band (5.925 to 6.425 GHz) was used for the uplink. Those two bands taken together are called C-band (as in C-band radar). Some newer communication satellites use the radar K_u -band operating at 11 GHz (10.95 to 11.2 GHz plus 11.45 to 11.7 GHz) for the downlink and 14 GHz (14.0 to 14.5 GHz) for the uplink. International agreements also have authorized the use of a third band, the K_a -band, with operation at 17 and 30 GHz.

The typical spectrum width of the radio channel served by one transponder is 36 MHz. The modulation type is FM. A single transponder can be used for one color television signal, 1,200 voice circuits, or digital data at a rate of 50 Mbps. The total width of each half of the C-band is 500 MHz. Horizontal and vertical polarization of the radio signal are used to double the capacity. The result is that 24 channels are available for use, which provide for 12 two-way transponder pairs. Many satellites can be used with separations adequate to permit coverage of a single satellite per antenna beam.

1.4 Cellular Telephone Systems

1.4.1 Introduction to Cellular Telephones

Cellular telephone service is one of the newer types of communication service that has had remarkable growth and acceptance by the public worldwide. It offers many advantages not offered by the older type conventional telephone systems. These include the ability to operate in almost any environment including inside buildings, inside aircraft, inside automobiles, buses, and other vehicles, and the ability to operate at locations well removed from fixed telephone base stations. In many ways, the cell phones act like super cordless telephones but with many features added. Modern cell phones have a remarkable degree of miniaturization resulting in very small size and weight for handheld units. The author has a cell phone that is 0.75 inch deep by 1.75 inches wide by 4.5 inches long. Many add-on features are available for cell phones including digital cameras and display capability. It has been reported on national TV that, as of January 2005, there are more cellular telephones in use than there are conventional telephone sets. One of the reasons for this great popularity is the safety feature provided by this unit. If a person is driving and has car trouble, he or she can call for help using the cell phone in the car. If an older person falls or otherwise has a health problem, he or she can call for help using the cell phone which he or she carries at all times. Many more examples of safety advantages could be cited. The cell phone is of great utility for workers working away from home or office. Young people love to talk with their friends on cell phones. One of the great joys for the author is to go to BYU and see the many students with cell phones at their ears.

One interesting and important application for cell phones is in telematics. Telematics is the transmission of data communication between systems and devices. One example of this is OnStar. OnStar's in-vehicle safety, security, and information services use Global Positioning System (GPS) satellite and cellular technology to link the vehicle and driver to the OnStar center. At the OnStar center, advisors offer real-time, personalized help 24 hours a day, 365 days a year.

One example service is the GM Advanced Automatic Crash Notification System (AACN). This system uses front and side sensors as well as the sensing and diagnostic module (SDM) itself. The accelerometer located within the SDM measures the crash severity. The SDM transmits crash information to the vehicle's OnStar module. Within seconds of a moderate to severe crash, the OnStar module will send a message to the OnStar call center (OCC) through a cellular connection, informing the advisor that a crash has occurred. A voice connection between the advisor and the vehicle occupant is established. The advisor then can conference in 911 dispatch, or a public safety answering point which determines if emergency services are necessary. If there is no response from the occupants, the advisor can provide the emergency dispatcher with the crash information from the SDM that reveals the severity of the crash. The dispatcher can then identify what emergency services may be appropriate. Using the GPS satellite, OnStar advisors are able to tell emergency workers the location of the vehicle. The number and location of sensors and SDM may vary depending on vehicle model.

AACN is not available on every OnStar-equipped vehicle. AACN is available on the model year 2004 Chevy Malibu and is being rolled out on more OnStar-equipped GM vehicles starting in model year 2005. Model year 2005 vehicles with

AACN include the Buick Rainier, Cadillac STS, Chevrolet Cobalt, Chevrolet Malibu, Chevrolet Malibu MAXX, Chevrolet TrailBlazer, Chevrolet TrailBlazer EXT, GMC Envoy, GMC Envoy XL, GMC Envoy XUV, Isuzu Ascender, Isuzu Ascender EXT, Pontiac G6, Pontiac Pursuit (sold only in Canada), and SAAB 9-7X.

There are many other OnStar services that can be provided using the vehicle-located cellular telephone systems. Other types of telematic applications for cell phones are clearly possible and should increase in the future.

1.4.2 The Concept of Spatial Frequency Reuse

A key problem faced by the developers of cellular telephones was the problem of insufficient available bandwidth for the number of users. A partial solution to the problem was spatial frequency reuse. The concept of spatial frequency reuse can be illustrated as follows. Assume an analog FM system with 70 channels and a separate service area 100 miles on a side. The number of telephone calls that can be handled in parallel by this system is equal to the number of channels (70). Now assume that the 100×100 mile service area is divided into 10×10 -mile cells and the transmitter power is reduced. The 100×100 -mile service area would then contain 100 cells. Then assume there is one base station in each cell and let each base station use 10 of the 70 channels. Adjacent cells would use different channels to avoid interference.

The spacing of cells of the same type and the reduced transmitter power provide the needed attenuation to prevent interference between cells using the same frequencies. The same 70 channels can now handle 1,000 telephone calls in parallel. If the cell size is reduced to 2×2 -mile cells and the transmitter power is further reduced, there would be 2,500 cells each with base stations that use 10 channels. The result would be the ability to service 25,000 calls. As the cells become even smaller, the power levels used by the base stations and the portable units must be decreased to avoid interference with other systems operating on the same frequencies.

These concepts have been implemented in both analog and digital cellular telephone systems. The cells usually are hexagonal in shape, to better approximate circles of constant radius. Clusters of cells are used. This arrangement of cells and clusters is illustrated in Figure 1.8. In this figure, the base station is located at the center of the cell site is identified by a letter that indicates the cell cluster and by a number that indicates the cell site within the cluster. For example, cell site C4 is located in cell 4 in cell cluster C.

A second approach to base-station location is to have each cell site at the edge of a cell where three cells meet. A single antenna mast can be used with three directional antennas. Each antenna covers a 120-degree sector and each antenna covers a cell. This approach reduces the number of cell sites needed by a factor of 3.

Seven cells are used for each cluster in Figure 1.8. Other cluster sizes sometimes used are 4, 12, and 14 cells per cluster. In each case there is a large separation of cells of the same type. That results in large attenuation of signals of the same frequency from an adjacent cluster. A requirement is that the interference from other cells be at least 18 dB below the desired signal strength.

Not all cells in a service area are the same size. Cells in heavily populated areas are made smaller so more users can be accommodated. As the population density

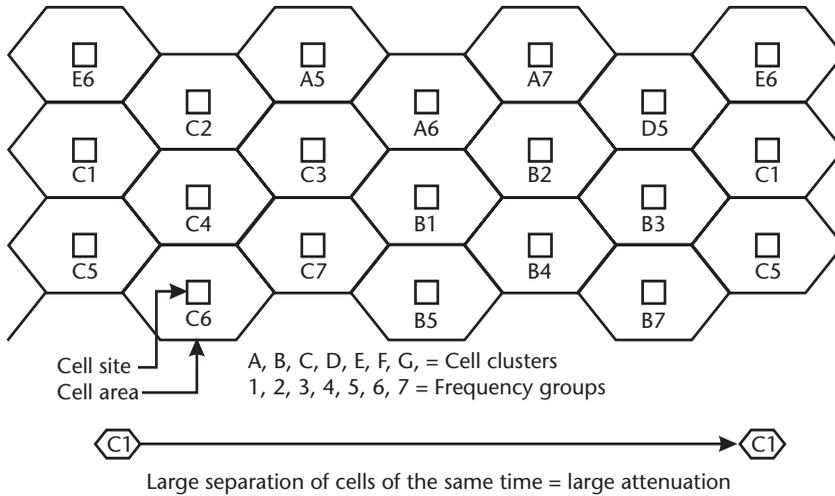


Figure 1.8 Cells and cell clusters for spatial frequency reuse. (After: [2].)

decreases, the cell size can be made larger. In rural areas, the cell size can be quite large (up to 50 km in diameter).

1.4.3 Propagation Characteristics of Cellular Telephone Systems

At the frequencies used by cellular telephones and the signal paths involved (ground wave and multipath) the attenuation tends to increase approximately as the fourth power of range. In contrast, free-space propagation attenuation (involving no interaction with the ground) increases as the second power of range. Thus, power received at the 4-mile range from the cellular transmitter would be a factor of 4 to the fourth power, or 256 (24 dB), below the power received at the 1-mile range. Similarly, the power received at the 10-mile range would be a factor of 10,000 (40 dB) less than the power received at the 1-mile range. The rapid rate of attenuation increase with distance is one of the key features that makes spatial frequency reuse possible [4].

An important propagation problem for operation in the cellular environment is the problem of multipath. Usually a number of propagation paths are present between the transmitter and the receiver. The strongest path will be the direct path. Other signals received are reflected signals both from the ground and from other reflecting objects. The resulting signal is the vector addition of all those signals. At some ranges, the reflected signals may add in phase, and the resulting signal will be greater than the direct signal. At other ranges the reflected signals may add 180 degrees out of phase from the direct signal, and the resulting signal will be less than the direct signal. With a moving mobile unit, the signal will fade in and out as a function of time.

1.4.4 Types of Cellular Telephone Systems

Table 1.1 shows characteristics of five main types of cellular telephone systems. Each of these systems are briefly discussed next.

Table 1.1 Characteristics of Five Main Types of Cellular Telephone Systems

<i>Characteristics</i>	<i>AMPS North America</i>	<i>IS-54 North America</i>	<i>IS-95 North America</i>	<i>GSM Europe</i>	<i>PDC Japan</i>
<i>Frequency</i>					
<i>RX (MHz)</i>	869–894	869–894	869–894	935–960	940–956 1,477–1,501
<i>TX (MHz)</i>	824–849	824–849	824–849	890–915	810–826 1,429–1,453
<i>Access method</i>	FDMA	TDMA/FDM	CDMA/FDM	TDMA/FDM	TDMA/FDM
<i>Duplex method</i>	FDD	FDD	FDD	FDD	FDD
<i>Channel spacing (kHz)</i>	30	30	1,250	200	25
<i>Number of channels per user</i>	832	832	10	124	1,600
<i>Users per channel</i>	1	3	118	8	3
<i>Bit rate (Kbps)</i>		48.6	1,228.8	270.83	42
<i>Modulation</i>	FM	$\pi/4$ DQPSK	BPSK/DQPSK	GMSK (0.3 Gaussian)	$\pi/4$ DQPSK

Source: [6].

1.4.4.1 Advanced Mobile Phone Service

The development of the Advanced Mobile Phone Service (AMPS) cellular telephone system was started in 1983. This is considered to be the first generation cellular telephone system. Table 1.1 shows parameters of the AMPS system. Figure 1.9 shows a block diagram of a cellular analog radio telephone (R/T). One is impressed with the large amount of electronic equipment that must fit into the small hand-held unit. Figure 1.10 illustrates this small unit with the cell site in the distance.

Figure 1.11 is a simplified system block diagram of a total analog cellular telephone system. The figure shows only a few of many cell sites (base stations) and a few of many possible mobile or portable subscriber units. Two-way radio links are provided between the cell sites and the handheld portable units. Separate links are provided for voice and data (command and control).

Two-way voice trunks and data links also are provided between the mobile telecommunication switching office (MTSO) and the cell sites. The MTSO is also connected to direct distance dialing networks, which connect to the party. The trunks may be microwave relay links, wire lines, or fiber optic links.

1.4.4.2 Digital Cellular Telephone Systems

Digital cellular telephone systems have been developed to greatly increase the number of subscribers that can be served with the limited bandwidth available for cellular telephone use. Three of the main types of digital telephone systems in use are the American digital cellular (ADC) system (IS-136), the Japanese digital cellular system (JDS), and the European digital cellular system known as the *Global System for Mobile Communications* (GSM). These systems all use TDMA modulation. While the three systems are similar, there are a few important differences.

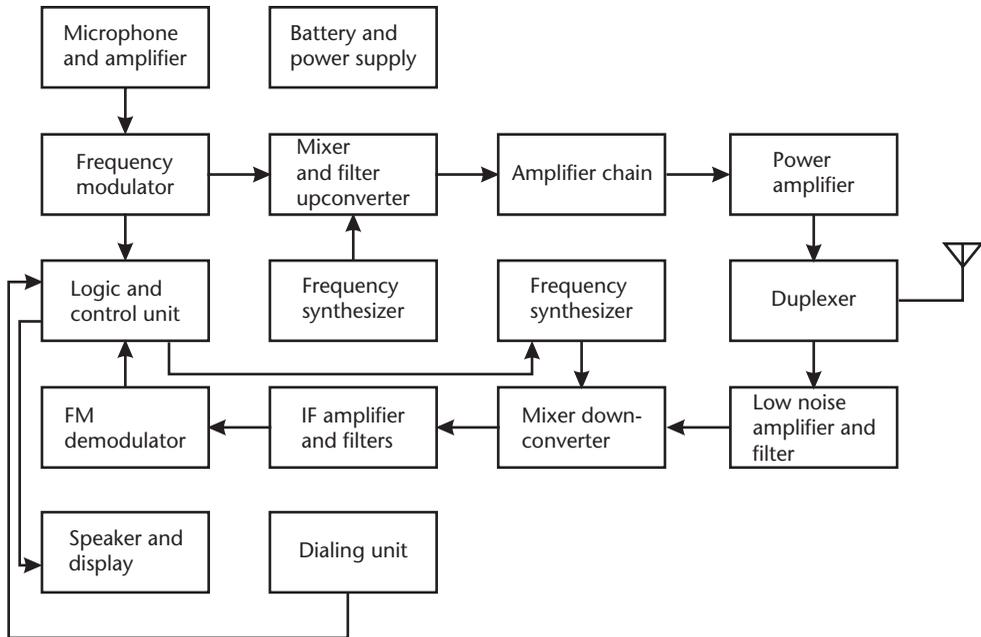


Figure 1.9 Block diagram of a cellular analog R/T.

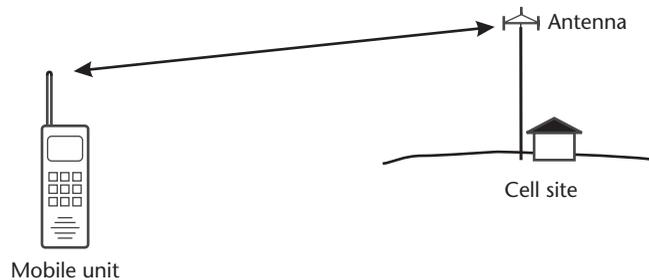


Figure 1.10 A cellular telephone.

The IS-136 digital cellular telephone system which replaced the IS-54 system uses the same frequency bands as are used for AMPS, the same cell structure, and the same overall control system design. One difference is the type of modulation used for voice channels. The AMPS system uses analog FM for voice, while the IS-136 digital system uses DQPSK modulation. The IS-136 provides three users per channel, while AMPS provides only one. The AMPS systems have been gradually phased out because of the increased capability of digital cellular systems.

The IS-95 digital cellular telephone system uses CDMA modulation. It has the potential for improving the efficiency of spectrum utilization over the older analog AMPS system by a factor of up to about 20. This system, which was developed by Qualcomm, is often called a third generation cellular telephone system. This system uses pseudorandom codes and 1.25-MHz channel bandwidth. One important advantage of the CDMA technology is that the systems can get by with lower signal-to-interference levels than the conventional narrowband FM techniques. This

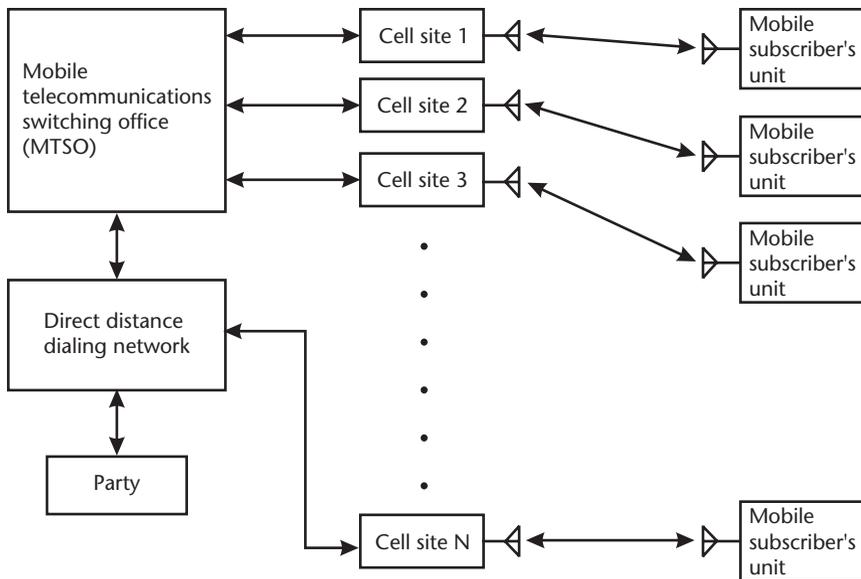


Figure 1.11 Simplified system block diagram of total cellular telephone system. (After: [7].)

allows CDMA systems to use the same set of frequencies in every cell, which provides a large improvement in capacity. Unlike other digital cellular systems, the Qualcomm system uses a variable rate vocoder with voice activity detection, which considerably reduces the required data rate and also the battery drain by the mobile transmitter.

In the early 1990s, a new specialized mobile radio service (SMA) was developed to compete with U.S. cellular carriers. Nextel and Motorola formed an extended SMR (E-SMR) network in the 800-MHz band that provides capacity and services similar to cellular. In 1995 Motorola's integrated radio system (MIRS) was replaced with the integrated digital enhanced network (iDen).

Personal Communication Service (PCS) licenses in the 1,800/1,900-MHz band were auctioned by the U.S. government to wireless providers in early 1995 and these have spawned new wireless services that complement, as well as compete with, cellular and SMR. Two examples are DCS-1900 (GSM), with a frequency band of 1.85–1.99 GHz, and PACS with the same frequency band. In the United States, the PACS standard, developed by Bellcore and Motorola, is likely to be used inside office buildings as a wireless voice and data telephone system or radio local loop.

In Europe earlier cellular systems such as ETACS, NMT-450 and NMT-900 are now being replaced by the pan-European digital cellular standard GSM (Global System for Mobile) which was first deployed in 1990 in a new 900-MHz band which all of Europe dedicated for cellular telephone service. The GSM standard has gained worldwide acceptance as the first universal digital cellular system with modern network features extended to each mobile user, and the leading digital air interface for PCS services above 1,800 MHz throughout the world.

In Japan, the Pacific Digital Cellular (PDC) standard with frequency band 810–1,501 MHz provides digital cellular coverage using a system similar to North America's USDC.

Some of the foregoing information is quoted or adapted from [10].

1.4.5 How a Cellular Telephone Call Is Made

When a cellular phone is turned on, but is not yet engaged in a call, it first scans the group of forward control channels to determine the one with the strongest signal, and then monitors that control channel until the signal drops below a usable level. At this point, it again scans the control channel in search of the strongest base station signal. The control channels are defined and standardized over the entire geographic area covered and typically make up about 5% of the total number of channels available in the system (the other 95% are dedicated to voice and data traffic for the end-users). Since the control channels are standardized and are identical throughout different markets within the country or continent, every phone scans the same channels while idle. When a telephone call is placed to a mobile user, the MSC dispatches the request to all base stations in the cellular system. The mobile identification number (MIN), which is the subscriber's telephone number, is then broadcast as a paging message over all of the forward control channels throughout the cellular system. The mobile receives the paging message sent by the base station, which it monitors, and responds by identifying itself over the reverse control channel. The base station relays the acknowledgment sent by the mobile and informs the MSC of the handshake. Then, the MSC instructs the base station to move the call to an unused voice channel within the cell (typically, between 10 to 60 voice channels and just one control channel are used in each cell's base station). At this point, the base station signals the mobile to change frequency to an unused forward and reverse voice channel pair, at which point another data message (called an *alert*) is transmitted over the forward voice channel to instruct the mobile to ring, thereby instructing the mobile user to answer the phone.

Once a call is in progress, the MSC adjusts the transmitter power of the mobile and changes the channel of the mobile unit and base stations in order to maintain call quality as the subscriber moves in and out of range of each station. This is called *handoff*. Special control signaling is applied to the voice channels so that the mobile unit may be controlled by the base station and the MSC while a call is in progress.

When a mobile originates a call, a call initiation request is sent on the reverse control channel. With this request the mobile unit transmits its telephone number (MIN), electronic serial number (ESN), and the telephone number of the called party. The mobile also transmits a station class mark (SCM), which indicates what the maximum transmitter power level is for the particular user. The cell base station receives this data and sends it to the MSC. The MSC validates the request, makes a connection to the called party through the PSTN, and instructs the base station and mobile user to move to an unused forward-and-reverse channel pair to allow the conversation to begin.

All cellular systems provide a service called *roaming*. This allows subscribers to operate in service areas other than the one from which service is subscribed. When a mobile enters a geographic area that is different from its home service area, it is registered as a roamer in the new area. Every few minutes, the MSC issues a global command over each FCC in the system asking for all mobiles, which are previously unregistered to report their MIN and ESN over the RCC. The MSC uses this data to request billing status from the home location register. Once registered, roaming mobiles are allowed to receive and place calls from the area, and billing is routed automatically to the subscriber's home service provider.

1.5 Modern Commercially Available Cellular Telephone Systems

Key features for modern commercially available cellular telephone systems include a very high degree of miniaturization, and the optional use of a number of add-on features such as digital cameras for camera-phone operation and GPS receivers. In the future, it will be possible to include many other add-on features.

The cost of cell phones continues to drop as the capability improves. We are now in the age when there are multiple cell phones per family. In fact, in Finland, where Nokia is headquartered, there are more cell phones than people.

References

- [1] Keiser, B. E., and E. Strange, *Digital Telephony and Network Integration*, 2nd ed., New York: Van Nostrand Reinhold, 1995, Appendix C.
- [2] Noll, A. M., *Introduction to Telephone and Telephone Systems*, 2nd ed., Norwood, MA: Artech House, 1991, Chap. 1, 3, 6.
- [3] RadioShack, *1997 Catalog*, pp. 6–59.
- [4] Macario, R. C. V., *Cellular Radio Principles and Design*, New York: McGraw-Hill, 1993, pp. 47–53.
- [5] Mehrotra, A., *Cellular Radio Analog and Digital Systems*, Norwood, MA: Artech House, 1994, pp. 245–257.
- [6] Williams, M., and D. Ong, “PCS and RF Components,” *Applied Microwave & Wireless*, Spring 1995.
- [7] Lee, W. C. Y., “Cellular Telecommunication Systems,” Chap. 46 in *Reference Data for Engineers*, 8th ed., M. E. Van Valkenberg, (ed.), SAMS, Carmel, IN: Prentice Hall Computer Publishing, 1993.
- [8] Keiser, B. E., and E. Strange, *Digital Telephony and Network Integration*, 2nd ed., New York: Van Nostrand Reinhold, 1995, p. 257.
- [9] RadioShack, *2005 Catalog*, <http://www.radioshack.com>.
- [10] Rappaport, T. S., *Wireless Communications Principles and Practices*, 2nd ed., Upper Saddle River, NJ: Prentice Hall, 2002, Chap. 1.

Wireless Communication Systems

2.1 Introduction

This chapter discusses a number of important wireless communication systems. It also presents some useful tutorial background information that may be helpful to readers of this book. This includes discussion of units and conversion information, frequency band information, and a discussion of the use of decibels in RF systems. It also discusses briefly frequency allocations and FCC regulations.

2.2 Background Information

2.2.1 Units and Conversion Information

Table 2.1 shows unit and conversion information for frequency and wavelength.

A wavelength is the distance an electromagnetic wave travels in one RF cycle. That distance depends on the speed of the wave in the medium involved. In the case of free space, the speed of the RF electromagnetic wave is the speed of light, c , which is approximately $3 \cdot 10^8$ meters per second. The equation for wavelength with free space propagation is

$$\lambda = \frac{c}{f} \quad (2.1)$$

where:

λ = wavelength in meters

c = speed of light in meters per second = 3×10^8 meter/second

f = frequency in hertz (1 Hz = 1 cycle per second)

In media other than free space, the RF energy travels at a speed less than the speed of light, and the wavelength at a given frequency is smaller. The ratio of the wavelength in free space to the wavelength in a medium other than free space is equal to the square root of the relative dielectric constant, permittivity, of the medium.

Table 2.1 Common Frequency and Wavelength Units and Conversion

<i>Frequency</i>		
1 hertz	= 1 Hz	= 1 cycle per second
1 kilohertz	= 1 kHz	= 10^3 Hz
1 megahertz	= 1 MHz	= 10^6 Hz
1 gigahertz	= 1 GHz	= 10^9 Hz
1 terahertz	= 1 THz	= 10^{12} Hz
<i>Wavelength</i>		
1 centimeter	= 1 cm	= 10^{-2} m
1 millimeter	= 1 mm	= 10^{-3} m
1 micron (or micrometer)	= 1μ	= 10^{-6} m
1 angstrom	= 1 A	= 10^{-10} m

2.2.2 Frequency Bands for Communication Systems

Table 2.2 shows frequency band names and frequency coverage information for communication systems. Corresponding free-space wavelengths are also shown.

2.2.3 The Use of Decibels

Decibels are frequently used to indicate ratios of powers in a logarithmic fashion. As such, a ratio expressed in decibels is simply 10 times the logarithm to the base 10 of the numeric power ratio. The following are examples.

2.2.3.1 Filter Loss

If the output power from a filter is one-half the input power, the insertion loss for the filter is $10 \log_{10} 2 = 3$ dB.

2.2.3.2 Amplifier Gain

If the power output of an amplifier is 50 times the input power, the gain of the amplifier is $10 \log_{10} 50 = 17$ dB.

2.2.3.3 Antenna Gain

If the power density in a given direction is 40 times what it would be if the antenna had been an isotropic radiator (uniform in all directions), the gain of the antenna with respect to isotropic would be $10 \log_{10} 40 = 16$ dBi, where “i” indicates isotropic. Sometimes antenna gain is given with respect to that of a dipole antenna. In that case, if the power density in a given direction is two times what it would be if the antenna had been a dipole antenna, the gain of the antenna would be $10 \log_{10} 2 = 3$ dBd (where “d” indicates dipole). That would be about 5 dBi, since the gain of a dipole with respect to isotropic is about 1.8 dBi.

Table 2.2 Frequency Bands for Communication Systems

<i>Frequency Bands for Communication Systems</i>		
<i>Frequency</i>	<i>Wavelength</i>	<i>Frequency Band</i>
3–30 kHz	10^5 – 10^4 m	VLF (very low frequency)
30–300 kHz	10^4 – 10^3 m	LF (low frequency)
0.3–3 MHz	10^3 – 10^2 m	MF (medium frequency)
3–30 MHz	10^2 –10 m	HF (high frequency)
30–300 MHz	10–1 m	VHF (very high frequency)
0.3–3 GHz	1–0.1 m	UHF (ultra high frequency)
3–30 GHz	10–1 cm	SHF (super high frequency)
30–300 GHz	1–0.1 cm	EHF (extremely high frequency)
0.3–3 THz	1–0.1 mm	Band 12
1–417 THz	300–0.72 mm	Infrared
417–789 THz	0.72–0.38 mm	Visible light
789 to 5×10^6 THz	0.38 to 6×10^{-5} mm	Ultraviolet
3×10^4 to 3×10^8 THz	100 to 1×10^{-2} Å	X-rays
$>3 \times 10^7$ THz	< 0.1 Å	Gamma rays

2.2.3.4 Power Level

If the transmitted power level is 300W, the power level expressed in decibel format would be $10 \log_{10} 300 = 25$ dBW, where W stands for watts. Equivalently, the power can be expressed as $10 \log_{10} (30,000 \text{ mW}) = 55$ dBm, where m stands for milliwatts. It is emphasized that dBm and dBW are ratios relative to 1 mW and 1 W, respectively.

If the receive power is $8 \cdot 10^{-9}$ W, or 8 nW (nanowatts), the received power expressed in dBW is $10 \log_{10}(8 \cdot 10^{-9}) = 10 \log_{10}(8) + 10 \log_{10}(10^{-9}) = 10(0.9) + 10(-9) = 9 - 90 = -81$ dBW.

2.2.3.5 Voltage Ratio

When expressing a voltage ratio in decibels, we express the ratio as a power ratio. Power is voltage squared divided by resistance. Thus, the voltage ratio 100/10 expressed in decibels is $20 \log_{10}(100/10) = 20$ dB.

2.3 Frequency Allocation and FCC Regulations

The frequency allocations for the different communication services are provided on a worldwide basis by the International Telecommunication Union (ITU). The allocations for Region 2, which includes North and South America, have been adopted by the U.S. Federal Communications Commission (FCC), which has responsibility for RF spectrum management within the United States. Individual frequency assignments or authorizations must be requested by the user and subsequently approved by the FCC before the user may transmit at the requested frequencies. Other FCC regulations also must be followed, such as the maximum transmitted power level, out-of-band harmonics, and spurious signal levels. A

complete listing of FCC frequency allocations as of 2002 is given in [1]. More recent allocations are available from other sources that were not available to the author at the time of this writing.

2.4 Types of Communication Services as Defined by the ITU

The following are some of the more important communication services as defined by the ITU. Each is discussed briefly in the following sections.

2.4.1 Aeronautical Mobile Service

Figure 2.1 shows illustrations of two types of aeronautical mobile service. These are ground to aircraft communication using HF ionospheric refraction propagation (sometimes referred to as sky wave propagation) for long distance coverage, and ground to aircraft communication using VHF or UHF free-space propagation for line-of-sight short-range communication. Allocated HF frequencies are in the frequency range of 2.85 MHz to 23.35 MHz. Different frequencies are used for different times of the day, seasons of the year, and other factors. HF communication is used mainly for communication between land-based stations and aircraft flying over the Atlantic or Pacific Oceans.

Most of the aircraft that fly only over land use only VHF or UHF frequencies. VHF frequencies used are in the range of 118 MHz to 237 MHz. That band normally is used for amplitude modulation (AM) signals. UHF frequencies used are in the range of 225 MHz to 400 MHz. That band normally is used for frequency modulation (FM) signals. The propagation mode in each case is free-space propagation, sometimes called *space wave*. The space wave represents the energy that travels

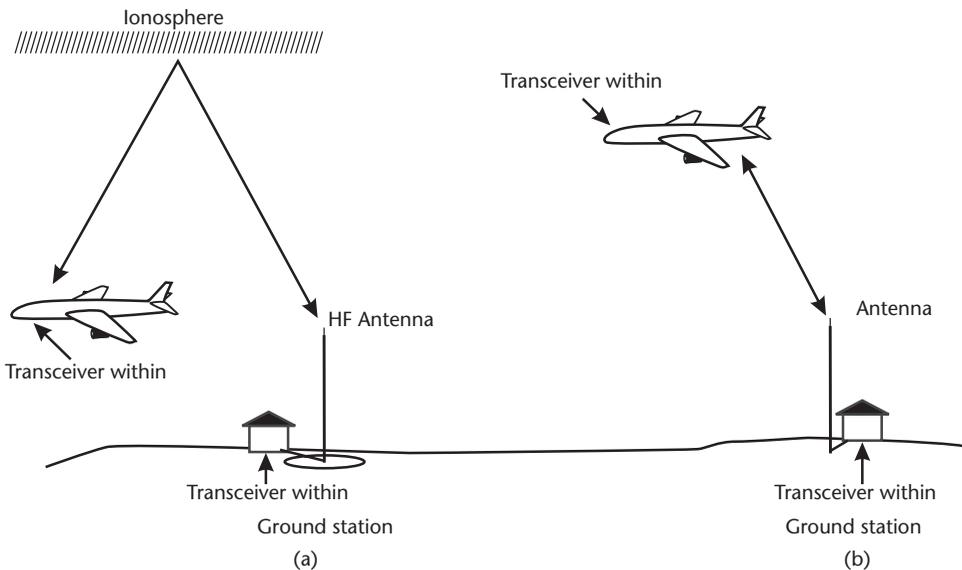


Figure 2.1 Types of aeronautical mobile service (a) aeronautical mobile service using HF ionospheric refraction propagation and (b) aeronautical mobile service using VHF or UHF free-space propagation.

from the transmitter to the receiver antenna in the Earth's troposphere, that is, the portion of the Earth's atmosphere in the first 10 miles adjacent to the Earth's surface. The range capability for free-space propagation is radio horizon limited and for airborne systems is typically less than 200 miles, depending on aircraft altitude and the surrounding topology.

2.4.2 Aeronautical Mobile Satellite Service

Figure 2.2 illustrates aeronautical mobile satellite service. This drawing shows two-way communication between a ground station and an aircraft perhaps flying over the Atlantic Ocean or the Pacific Ocean at a very long range from the ground station. A stationary orbit (geosynchronous) communication satellite serves as a relay station for the signals. The frequency allocations for the service are 1,545–1,555 MHz for the space-to-Earth link (downlink) and two bands at 1,646.5–1,656.5 MHz and 1,660–1,660.5 MHz for the Earth-to-space link (uplink). There are a number of important advantages for this type of system over the HF system of 2.4.1. The HF system has problems of crowded spectrum and possible radio blackout due to ionospheric disturbances.

2.4.3 Amateur Service

Amateur service is a radio-communication service for the purpose of self training, enjoyment, and technical investigations carried out by amateurs. Amateurs are defined as duly authorized persons interested in radio techniques solely with a personal aim and without pecuniary interest.

Much of the radio amateur activity is in the HF band where very long-range communication is possible using ionospheric refraction or sky wave propagation. Frequency allocations for amateurs in the HF band are 3.5–3.75 MHz, 7.0–7.3 MHz, 14.0–14.25 MHz, 18.069–18.168 MHz, 21.0–21.450 MHz, 24.89–24.99 MHz, and 28.0–28.45 MHz. Frequency allocations are also provided for amateurs in the MF, VHF, UHF, SHF, and EHF bands.

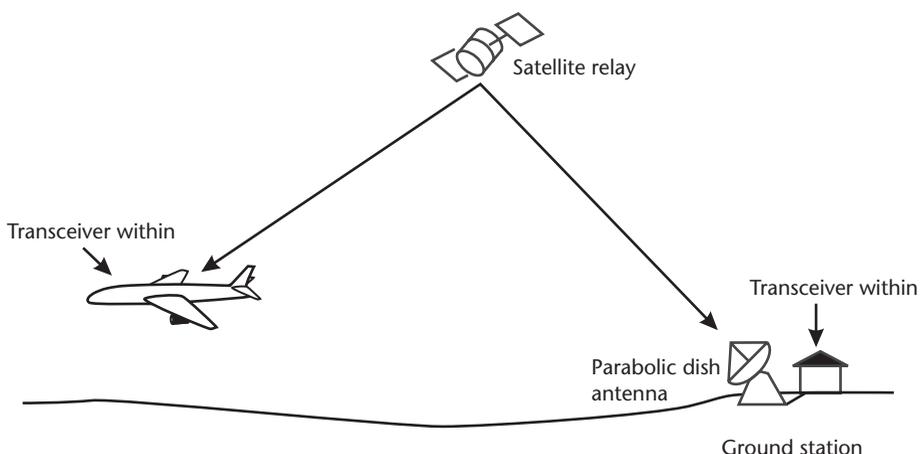


Figure 2.2 Aeronautical mobile satellite service.

2.4.4 Broadcasting Service

Figure 2.3 shows an illustration of VHF and UHF broadcast service. Broadcasting service is a radio communication service in which the transmissions are intended for direct reception by the general public. This service includes sound, television, and other types of transmission. Frequency allocations for this service include MF, HF, VHF, UHF and higher frequencies.

The MF frequencies are used for standard AM broadcasting. The mode of propagation in this case is normally ground wave during the daytime. Covering range in this case is usually less than 100 miles. Ground wave is sometimes called *surface wave*. It can exist when the transmitting and receiving antennas are close to the surface of the Earth and are vertically polarized. Ground wave is supported at its lower edge by the presence of the ground. Charges move in the Earth and constitute a current with loss as the wave moves. It is effective only at the lower frequencies because losses become too great at the higher frequencies.

At night it is sometimes possible to have longer-range capability at MF frequencies by means of E-layer ionospheric refraction propagation. This mode, however, is not depended on for MF broadcast service.

The HF frequencies are used for broadcasting over very long range using ionospheric propagation. Only narrowband operation is possible.

VHF frequencies are used for FM radio broadcasting and television broadcasting. FM radio broadcasting uses the 88–108-MHz band. VHF television broadcasting uses the 54–72-MHz band, the 76–88-MHz band, and the 174–216-MHz band. The UHF television broadcasting uses the 470–608-MHz band and the 614–890-MHz band.

FM radio and television broadcasting are illustrated in Figure 2.3. Here we see that a single high-power transmitter with a wide-angle antenna can provide signals to a large number of receivers. Coverage range can be 50 miles or more, depending on antenna height, transmitter power, and surrounding topology.

The normal mode of propagation for VHF and UHF broadcasting is free-space propagation. Diffraction and reflection also are often involved where the

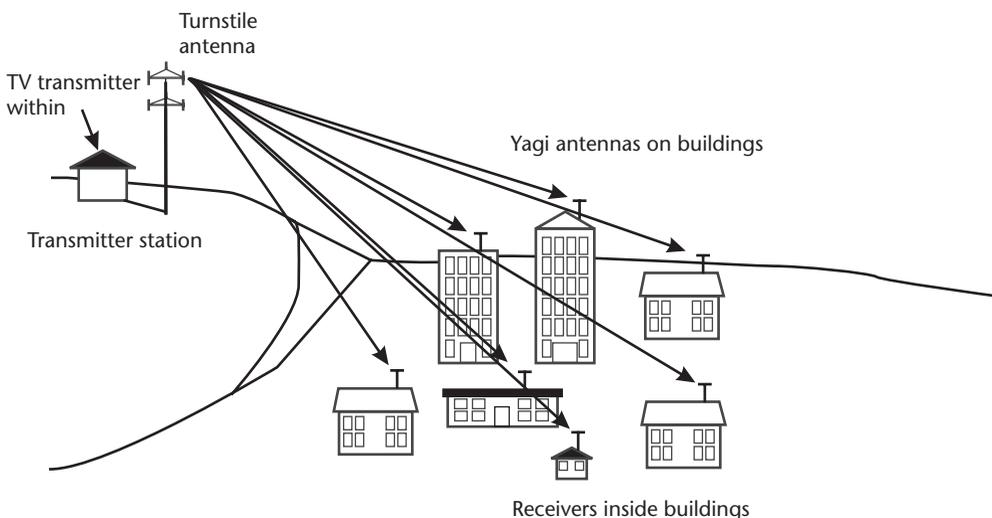


Figure 2.3 Illustration of VHF and UHF broadcast service.

line-of-sight path is blocked by trees, buildings, or hills. There is attenuation of the signal as the waves pass through trees and other natural and unnatural structures. These propagation modes are discussed in Chapter 5.

2.4.5 Broadcasting Satellite Service

Figure 2.4 shows the main concepts of broadcasting satellite service. Signals are relayed by a satellite for direct reception by the general public. The assumption is that stationary orbit satellites are used.

Frequency allocations for broadcasting satellite service include 2,520 to 2,670 MHz, 3.7 to 4.2 GHz, 12.2 to 12.7 GHz, and 17.5 to 17.7 GHz.

An example of broadcasting satellite service is direct satellite-to-home TV service where the home has its own dish antenna. Services such as DIRECTV®, PrimeStar, and Dish Network are examples of this type of service. Satellite telephone is also an example.

2.4.6 Citizen Band Radio

The 26.96–27.23-MHz frequency band is allocated for citizens who are not licensed amateurs but who desire to use radio transmitters and receivers for business or pleasure. These systems often are used by truck drivers and others who travel the nation’s highways. CB radio may also include ground stations for ground-to-mobile or ground-to-ground communication. The range capability for CB radio is about 10 miles.

An example of a modern CB transceiver made by RadioShack is a 40-channel AM mobile with digital signal processing. The signal processing virtually eliminates some types of noise and interference. The CB radio provides RF gain, squelch, and tone controls. By squelch is meant the ability to turn the sound on only when the desired signal is above a predetermined level in the receiver. The size of this unit is $1.75 \times 5.5 \times 8$ inches.

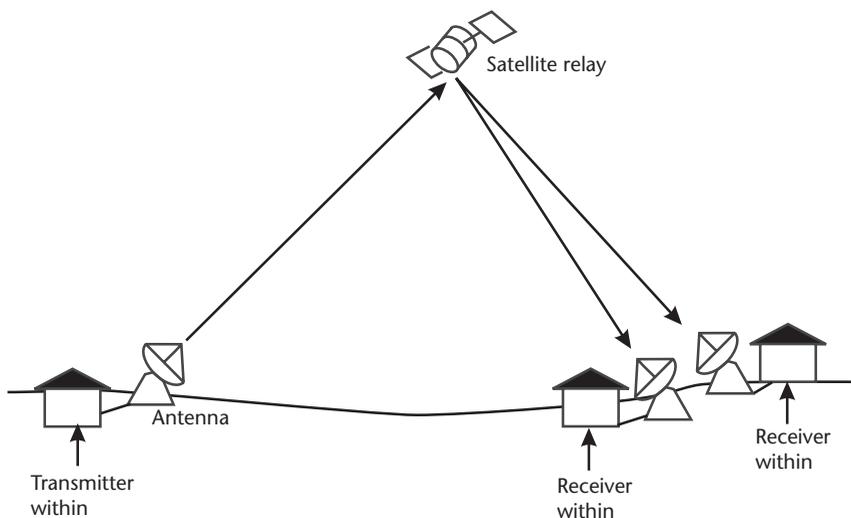


Figure 2.4 Broadcasting satellite service.

The newer RadioShack CB radio model TRC-485 provides both AM and single-sideband (SSB) capability. This unit provides improved range capability. It features dual watch, so the user can monitor one channel while listening to another. It features automatic modulation control and adjustable RF gain. The liquid crystal display (LCD) shows the exact frequency being used as well as the channel number.

2.4.7 VHF and UHF FM Business and Personal Two-Way Radio

Handheld VHF-FM and UHF-FM business two-way radios are available for ranges of up to several miles. These units require an FCC license to operate. This license is easy to get, and virtually any business, school, hospital, clinic, church, or organization can qualify. No test is required.

An example of a VHF system is one made by Motorola that provides 1W of output power. A typical frequency set at the factory is 154.60 MHz, but other assigned operating frequencies are easily set. A 2-W UHF-FM unit has a frequency of 464.550 MHz. The size of each unit is $6\text{-}7/8 \times 2\text{-}9/16 \times 1\text{-}3/8$ inches. Both the VHF and the UHF versions use PLL frequency synthesizers, rechargeable Ni-Cd batteries, and flexible quarter-wave antennas.

2.4.8 Pager Systems

Over 10 million people in the United States subscribe to paging services. Most of the earlier pager systems were fairly bulky and were worn on the users belt. Newer pagers are so small they can be carried in a shirt pocket or on the wrist like a watch. Older pagers simply beeped and the user had to telephone the paging office. Today pagers not only beep, but some also vibrate, for noiseless paging. Newer pagers present information in the form of LCDs, including the telephone number of the paging party, and in some systems, short messages. Other pagers can provide up to 8 seconds of voice message.

In some pager systems, the radio signal sent to the pager is carried on a subcarrier of a local FM radio station. The transmit frequency in that case is in the 88–108-MHz band. In other systems, the radio signal is sent from a base station with operation at a frequency near 900 MHz.

For example, in Utah, where the author lives, AT&T Wireless Services uses 35 base stations to cover the main population centers and the major highways in the state. Their operating frequency is 931.2125 MHz. Because of the availability of elevated transmit sites, coverage of about a 40-mile radius is possible for each station.

Long-distance paging is also possible. For example, AT&T Wireless Services uses communication satellites to carry paging requests across the continent to metropolitan areas, where local paging signals are then sent by radio.

There are many manufacturers of paging equipment, and the designs and cost of each type of system differ. Generally, the smaller the device, the higher the cost. Also, the more costly units provide the greatest information capability.

A pager package includes an antenna, a receiver, a digital logic circuit, an LCD circuit, a sound or speaker circuit, a vibrator unit, and a battery. All that equipment fits into a package that is only a few cubic inches in volume. A wristwatch-like system takes up less than 1 cubic inch. This is a remarkable degree of miniaturization for an RF system.

An example of a numeric-only pager is one made by RadioShack that costs about \$60. It is a dependable, one-button-operation pager with silent vibrator alert. Its flexible memory holds eleven 10-number messages or eight 20-number messages. An elapsed-time clock shows how long it has been since each page came in. It reminds the user for up to 12 hours that a new page has been received. Memory retention keeps the pages in memory when power is off. It has low-battery alert and uses one AA battery. It is only about 2 cubic inches in size.

2.4.9 Mobile Service

Mobile service is a radio communication service between mobile and land stations or between mobile stations. This is the type of communication service used by public safety systems, industrial systems, and land transportation systems. Public safety systems include police, fire, highway, forestry, ambulance, and emergency services. Industrial systems include power, petroleum, pipeline, forest products, factories, builders, ranchers, motion picture, press relay, and radio-controlled appliance repair service. Land transportation systems include taxis, trucks, buses, and railroads.

Most of these examples of communication systems operate in either the VHF or low UHF frequency bands. Typical operating frequencies are less than 470 MHz. Most of the systems use narrowband FM modulation. The mobile units use fairly small transmitter power and small antennas. The ground stations use somewhat higher power and elevated antennas.

2.4.10 Remote Control Systems

There are many types of wireless remote control systems. Examples of simple remote control systems are the remote garage door opener and the remote car door opener. By pushing a button the Garage door opener transmitter sends out an RF coded signal that is recognized by the receiver located within the garage. This in turn turns on the motor that causes the door to open. In a similar way by pushing a button on a small transmitter often connected to the car key, the remote car door opener sends out a coded signal that is recognized by a receiver located in the car. This in turn causes the selected door or the trunk to be unlocked. Other functions may also be provided.

A somewhat more complicated remote control unit using IR signals is used to operate a television set. It typically has many different controls such as power on-off, channel selection, volume control, brightness control, color control, and many others.

Many other types of remote control systems are used. Some of the information presented in Section 2.4 is quoted or adapted from [1–3].

2.5 WLANs, IEEE 802.11 and Bluetooth

2.5.1 WLANs

WLAN is an acronym for wireless local-area network. This is a type of local-area network that uses microwave radio rather than wires or cables to communicate

between elements of a computer network. Most WLANs are confined to a single building or a group of closely related buildings.

However one LAN can be connected to other LANs over any distance via telephone lines or radio links. A system of this kind is called a wide area network (WAN).

The primary reason for building a wireless network is for increased mobility so users of the network can move around from room to room or place to place without being tethered to a computer or network jack. It also provides a way to easily network computers together without having to snake wires and cables through walls. With WLANs many users of the network can share expensive equipment, such as laser printers, as well as data. Network users can use the LAN to communicate with each other by sending email or engaging in chat sessions.

There are many applications for WLANs. These include medical applications in hospitals, factory applications, warehouse applications, technical and business meetings. Many other applications could be listed.

2.5.2 IEEE 802.11 Standard

In the IEEE 802.11 standard for communication for wireless LANs there are two different ways to configure a network. These are ad hoc and infrastructure. In the ad hoc network, computers are brought together to form a network “on the fly”. There is no structure to the network; there are no fixed points, and usually every node is able to communicate with every other node. An example of this is a meeting where employees bring lap top computers together to communicate and share designs or financial information.

A second type of network used in wireless LANs is the infrastructure. This architecture uses fixed network access points with which mobile nodes can communicate. These network access points are sometimes connected to land lines to widen the LAN’s capability by bridging wireless nodes to other wireless nodes.

IEEE 802.11 provides for data rates up to 2 Mbps, with operation in the 2.4 to 2.4835 GHz frequency band. The modulation scheme used is either direct sequence spread spectrum, frequency hopping spread spectrum, or infrared (IR) pulse position modulation. It provides for 8 channels of data. This system is a relatively short-range system with range capability being less than 300 feet.

Security is a very important requirement for WLANs. With IEEE 802.11 security is provided by Wired Equivalent Privacy (WEP) and MAC address filters. WEP is a system for encrypting data to keep it private from unauthorized users. TKIP is a higher performing security system that can be added on top of WEP for better security. It offers new encryption algorithms and constantly changes the encryption key needed.

2.5.3 IEEE 802.11b Standard

IEEE 802.11b is a higher performance communication standard for LANs. It provides for up to 11 Mbps in the 2.4 GHz frequency band. The modulation scheme is direct sequence spread spectrum. It offers high-speed access to data at up to 300 feet from the base station. A total of 11 channels can be used. Security is provided by WEP.

2.5.4 IEEE 802.11g Standard

IEEE 802.11g provides for up to 54 Mbps in the 2.4 frequency band. The modulation scheme is DSSS below 20 Mbps and OFDM above 20 Mbps. Security is by WEP. A total of 11 channels can be used.

2.5.5 Bluetooth

Bluetooth is a standard and specification for small-form factor, low cost, short-range, radio links between mobile PCs, mobile phones and other portable devices. The technology allows users to form wireless connections between various communication devices, in order to transmit real-time voice and data communications. The Bluetooth radio is built into a small microchip and operates in the 2.4 GHz band. It uses frequency hopping spread spectrum modulation, which changes its frequency 1600 times per second.

The Bluetooth is intended to replace the cables connecting portable and/or fixed devices. Its key features are robustness, low complexity, low power, and low cost.

Some of the foregoing material is quoted or adapted from references [4] and [5].

References

- [1] Van Valkenburg, M. E., (ed.), *Reference Data for Engineers*, 9th ed, Boston, MA: Butterworth-Heinemann, 2002, Chap. 1, pp. 1-6-1-93.
- [2] Van Valkenburg, M. E., (ed.), *Reference Data for Engineers*, 8th ed., SAMS, Carmel, IN: Prentice Hall Computer Publishing, 1993.
- [3] Kennedy, G., *Electronic Communication Systems*, 3rd ed., New York: McGraw-Hill, 1985.
- [4] Lough, Daniel L., T. Blankenship, and Keven J. Kvizman, "A Short Tutorial on Wireless LANs and IEEE 802.11," <http://www.computer.org>.
- [5] Bluetooth Tutorial, <http://www.wave-report.com>.

Radionavigation and Global Positioning Systems (GPS)

3.1 Older Radionavigation Systems

A number of different types of radionavigation systems have been used in the past and some of these are still in use. These are briefly discussed in this section before moving into GPS. Some of this information is quoted or adapted from [1, 2].

3.1.1 Omega System

Omega was the first worldwide, continuously available radionavigation system. The system became operational in the 1970s. It was decommissioned in 1997.

Omega was realized with eight ground-based transmitters. Each station operated in a CW mode at five frequencies in the VLF band: 10.2 kHz, 11.05 kHz, 11.333 kHz, 13.6 kHz, and a unique frequency to identify the station. The radiated power of the transmitting stations was about 10 kW. Signals propagated around the globe in sky-wave mode.

Omega was a hyperbolic navigation aid with one subtle difference: Rather than measuring the time difference in the arrival of two signals, an Omega receiver measured the phase difference between the sinusoidal signals. A line of constant phase difference is a hyperbola, but not a unique one. Lines of constant phase (LOP) difference recur over the coverage area. Figure 3.1 shows hyperbolic LOPs corresponding to zero degrees phase difference. The region between the lines of zero phase difference is called a lane. In order for a phase measurement to provide an unambiguous position estimate, the user has to determine the correct lane first. For 10.2-kHz signals, the width of a lane along the line joining the two transmitters is about 15 km (one-half of a wavelength). The lane width can be made larger by using other frequencies in combination with 10.2 kHz. Phase measurements at 10.2 and 11.33 kHz can be used to give a lane width of 130 km.

Given an initial user position, the receiver can keep track of the lanes crossed in the course of a voyage and produce subsequent position estimates without any ambiguity.

The largest source of error for the Omega system was propagation model errors. Accuracies of 2–4 nautical miles were achieved in most of the coverage areas by this system.

The operation of an Omega system is shown in Figure 3.2.

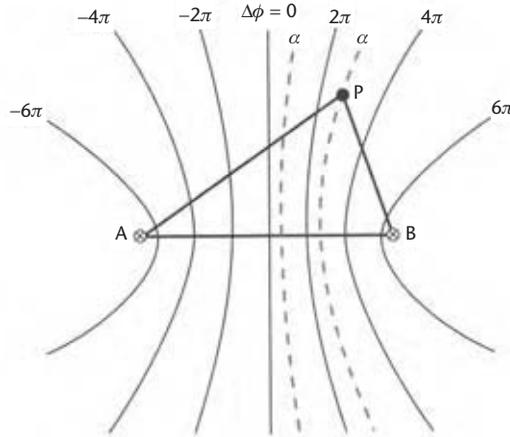


Figure 3.1 Omega lane boundaries and lane ambiguity. (After: [2].)

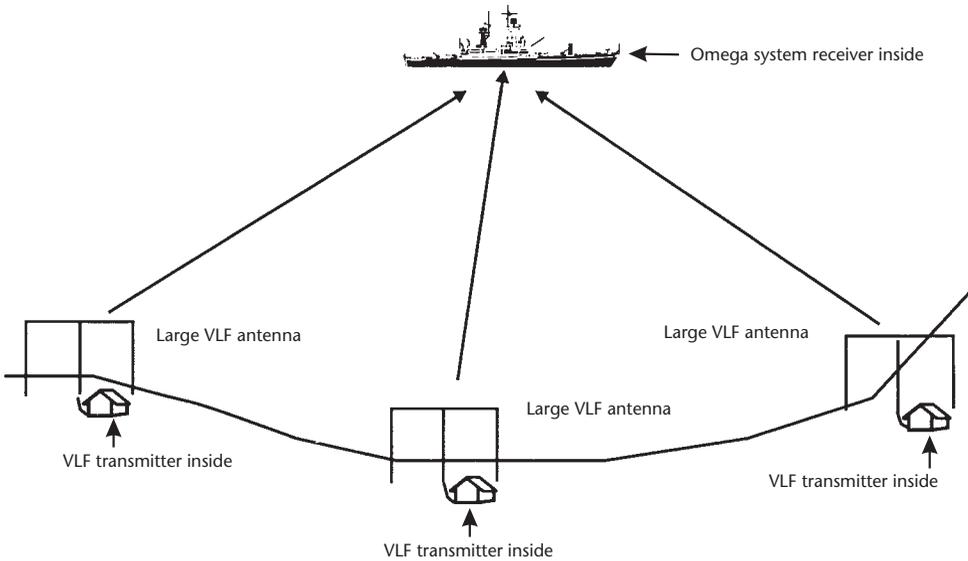


Figure 3.2 Omega system navigation aid.

3.1.2 Loran System

The radio navigation era began in earnest during World War II with the development of Gee in Great Britain to guide aircraft and Loran (Long-range Navigation System) in the United States primarily for guiding ships. Both systems were designed as high-power hyperbolic systems transmitting pulse signals. Loran-C, which was to have been decommissioned in 2000, but this decision was reversed. It has been upgraded and is likely to be around beyond 2008.

Loran-C is illustrated in Figure 3.3. It is a hyperbolic system operating in the LF-band at 90–110 kHz. The system is comprised of a set of chains of high-power transmitters. A typical chain consists of a master ground station and two slave or secondary ground stations, each separated from the master by about 1,000 km. Each station transmits about 1 MW of peak power. The U.S. Coast Guard operates 29 transmitting stations composing 13 chains covering the U.S. coastal waters, the contiguous 48 states, the Aleutian Islands, and the Bering Sea. Loran-C never became a global system, though it had the capability.

Loran-C has a 2-D rms positioning accuracy of about 250m. The synchronized transmitters of a chain radiate pulses of RF energy. A receiver onboard the ship or aircraft measures the time difference (TD) between arrival of pulses from the master and slave or secondary stations. Each measured TD defines a hyperbolic LOP for the user. Intersection of two LOPs defines the user in 2-D. The LF-band Loran signal has both ground-wave and sky-wave components. The ground-wave is generally stable and predictable. The sky-wave component is not suited for precise positioning because of uncertainties associated with the ionosphere.

3.1.3 Radio Beacons and Airborne Direction Finders

Radio beacons are nondirectional ground-based transmitters that operate in the LF and MF bands. Frequencies used are 190–415 kHz and 510–535 kHz. A radio beacon system is illustrated in Figure 3.4. A radio direction finder is used to measure the relative bearing to the transmitter with respect to the heading of an aircraft or marine vessel. The direction finder is sometimes referred to as an airborne direction finder (ADF). Angular accuracy is in the range of ± 3 to ± 10 degrees. The beacons transmit either a coded or a modulated CW signal for station identification. The radio direction finder may use a rotating loop antenna.

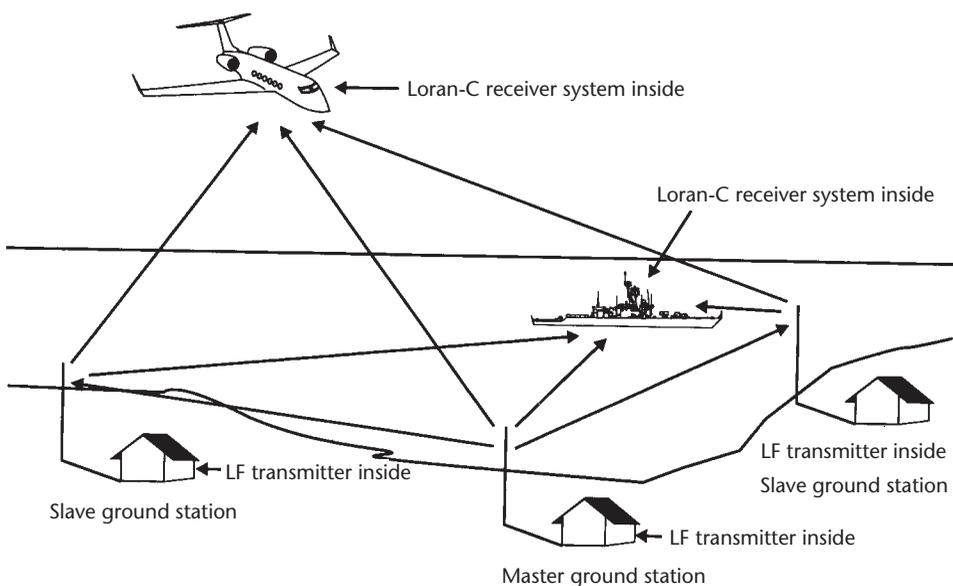


Figure 3.3 Loran-C system navigation aid.

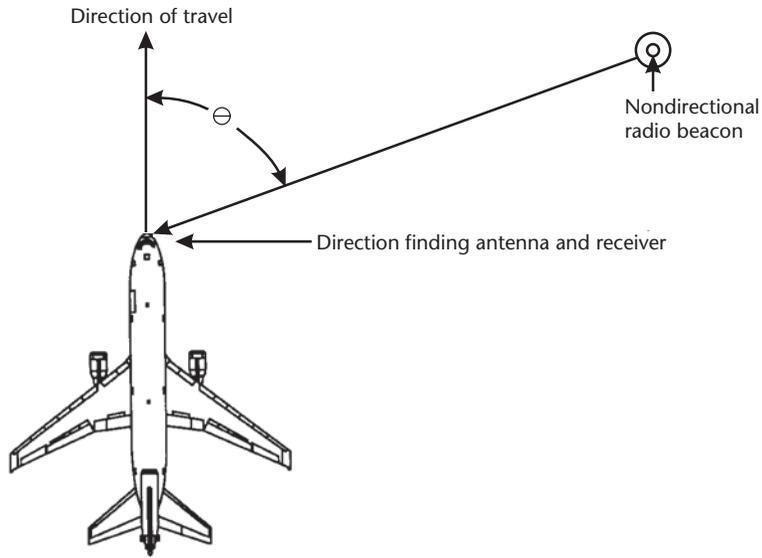


Figure 3.4 Radio beacon and ADF navigation aids.

3.1.4 VOR Systems

A VHF omnidirectional range (VOR) navigational aid transmits CW signals on one of 20 assigned channels in the 108–112-MHz band and 60 channels in the 112–118-MHz band, with 100 kHz channel separation. A 30-Hz reference signal is transmitted by an omnidirectional antenna. A frequency modulation of ± 480 Hz on a 9,960-Hz subcarrier is transmitted along with a carrier from a rotating antenna with a horizontal cardioid pattern. The cardioid antenna pattern rotates at a 30-Hz rate, allowing the airborne receiver to determine its bearing from the station as a function of phase between the reference and the rotating signal. A VOR navigation system is illustrated in Figure 3.5.

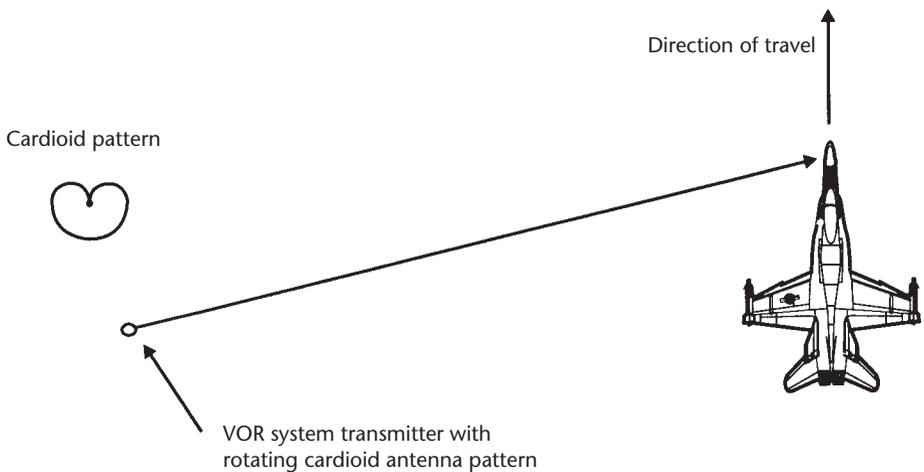


Figure 3.5 VOR system navigational aid.

3.1.5 Distance-Measuring Equipment (DME)

A typical DME system is illustrated in Figure 3.6. The airborne equipment (interrogator) generates a pulse signal that is recognized by the ground equipment (transponder), which then transmits a reply that is identified by the tracking circuit in the interrogator. The distance is computed by measuring the total round-trip time of the interrogation, the reply, and the fixed delay introduced by the ground transponder.

The airborne transponder transmits about 30 pulse pairs per second on one of 126 allocated channels between 1,025 and 1,150 MHz. The ground transponder replies on one of the paired channels in the 962–1,024-MHz band or in the 1,151–1,213-MHz band. A DME and a colocated VOR constitute an effective navigation system.

3.1.6 Instrument Landing System (ILS)

A typical instrument landing system (ILS) navigational aid is illustrated in Figure 3.7. Currently the ILS operating in the 108–112 MHz frequency band is the primary worldwide, ICAO-approved, precision landing system.

An ILS normally consists of two or three marker beacons, a localizer and a glide slope indicator to provide both vertical and horizontal guidance information. The localizer, operating in the 108–112-MHz band, normally is located 1,000 feet beyond the stop end of the runway. The glide slope normally is positioned 1,000 feet after the approach end of the runway and operates in the 328.6–335.4-MHz band. Marker beacons operating along the extension of the runway centerline at 75 MHz are used to indicate decision height points for the approach or distances to the threshold of the runway.

Azimuth guidance provided by the localizer is accomplished by use of a 90-Hz-modulated left-hand antenna pattern and a 150-Hz-modulated right-hand

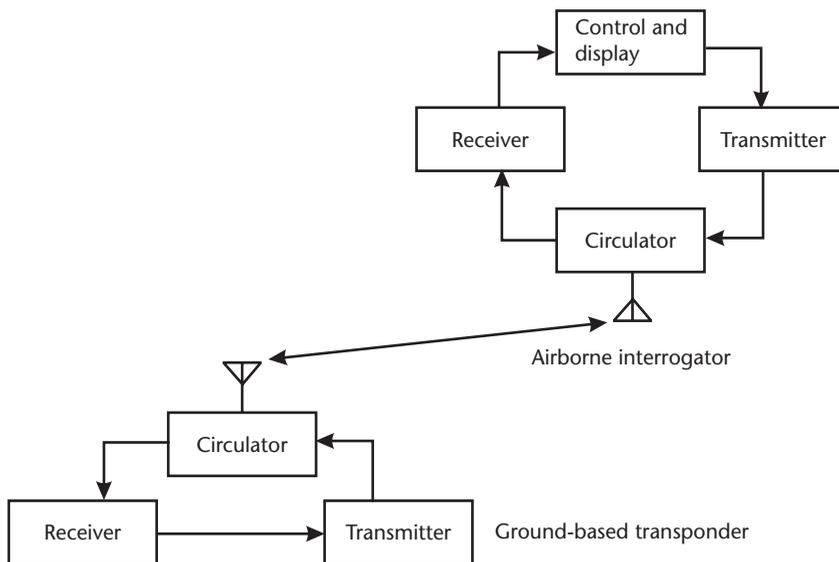


Figure 3.6 DME navigational aid.

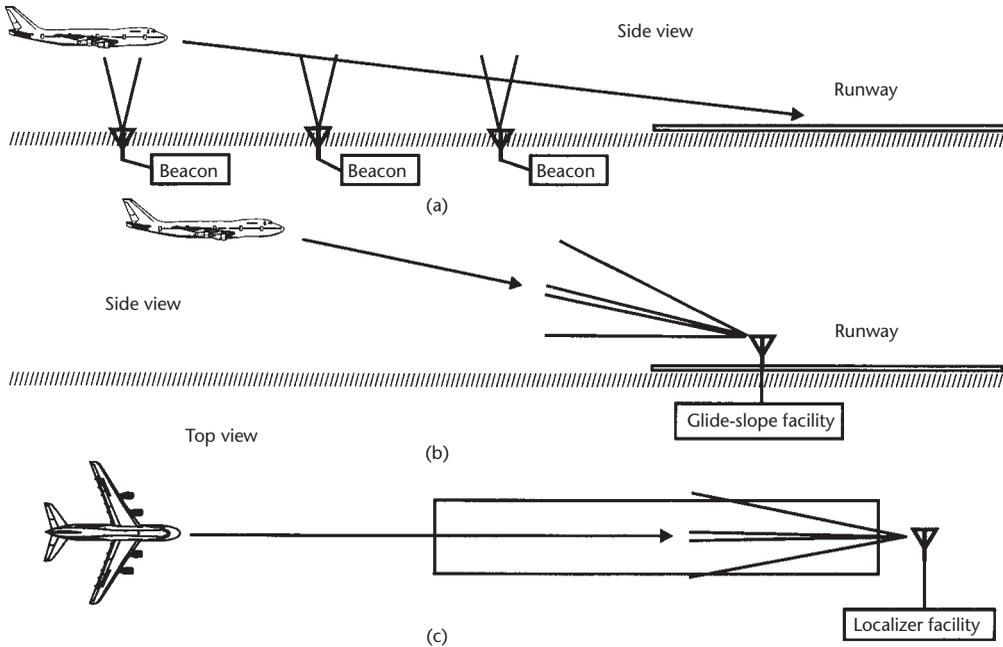


Figure 3.7 ILS: (a) fan-shaped markers along course; (b) glide-slope indicator; and (c) azimuth guidance.

antenna pattern as viewed from the aircraft on an approach. A 90-Hz signal detected by the aircraft receiver causes the course deviation indicator (CDI) to deviate to the right. A 150-Hz signal drives the CDI vertical to the left when the aircraft is right of the centerline course. When the aircraft is on the centerline, the CDI vertical needle is centered. The ILS localizer system provides a total of 40 channels. Each channel is paired with a possible glide-slope channel.

Vertical guidance is provided by the glide-slope facility, which is normally located to the side of the approach end of the runway. A total of 40 channels are provided in the 328.6–335.4-MHz band. The carrier radiated in the antenna pattern below the glide slope is amplitude modulated with a 150-Hz signal. The antenna pattern above the glide slope produces a signal with the 90-Hz amplitude modulation. When the approaching aircraft is on the glide slope, the CDI horizontal glide slope needle is centered. If the approaching aircraft is either too high or too low, that is indicated by the CDI glide slope needle.

The marker-beacon facilities along the course provide vertical fan markers to mark the key locations along the approach. The inner marker is normally at the runway threshold. The middle marker is about 3,500 feet from the runway threshold. The outer marker usually is about 5 miles from the runway threshold.

3.1.7 Tactical Air Navigation

The TACAN system provides both omnidirectional bearing and distance-measuring capability. The antenna system provides a rotating cardioid antenna pattern plus a rotating nine-lobed pattern. The antenna system rotates at a 15-Hz rate producing a 15-Hz course bearing and a 135-Hz (9×15 Hz) fine bearing in the aircraft.

Reference signals are transmitted by coded pulse trains to provide the phase reference. Bearing is obtained by the airborne receiver by comparing the 15-Hz and the 135-Hz sine waves with the reference pulse groups.

The TACAN system operates in the 960–1,215-MHz band with 1-MHz channel separations. The higher frequency used by TACAN permits the use of smaller antennas than are used by VOR. The multilobe principle improves the accuracy.

3.1.8 Microwave Landing System

The microwave landing system (MLS) is the ICAO-approved replacement for the current ILS system. The system is designed to meet the full range of user operational requirements for 2000 and beyond. The MLS system is based on the time-referenced scanning beams, referenced to the runway, that enable the airborne unit to determine the precise azimuth angle and elevation angle. Azimuth and elevation angle functions are provided by 200 channels in the 5,000–5,250-MHz band. Range information for the MLS system is provided by DMEs operating in the 960–1,215-MHz band. An option is included in the signal format to permit a special-purpose system operating in the 15,400–15,700-MHz band.

3.1.9 Air Traffic Control Radar Beacon System

The air traffic control radar beacon system (ATCRBS), is a ground-based radar system that operates with a beacon in the aircraft. The ground-based interrogator transmits at 1,030 MHz. The system uses a fan-shaped antenna beam that is wide in elevation but narrow in azimuth. The rotation or scan rate in azimuth is 5 Hz en route and 2.5 Hz for terminal areas. The interrogator transmits approximately 400 pulse pairs per second and receives replies from aircraft transponders that are within the beam of the antenna pattern.

The airborne transponder replies at 1,090 MHz with one of the 4,096 pulse codes available. The decoded replies are displayed on the surveillance radar PPI along with primary radar returns. An omnidirectional pulse pattern is also radiated from the ground to suppress unwanted sidelobe replies. This system is often referred to as a secondary surveillance radar.

3.1.10 Transit System

Transit was a satellite system operated by the U.S. Navy. It is no longer in use, having been decommissioned in 1996. It consisted of four or more satellites in approximately 600 nautical mile polar orbits. The satellites broadcasted ephemeris information continuously at 150 MHz and 400 MHz. A ship-based receiver measured successive Doppler shifts of the signal as the satellite approached or passed the user. The geographical position of the receiver was then calculated from the satellite position information and the Doppler measurements. The 2-D positioning accuracy for Transit was about 25m for a stationary user.

Coverage was worldwide, but not continuous. The period between updates of navigation information was as short as 1 hour but as long as 8 hours, depending on latitude. The Transit system has been replaced by GPS.

3.2 GPS Navigation System

Global Positioning System (GPS) is a worldwide satellite navigation system developed by the U.S. Department of Defense. Development was started in the early 1970s. Initial operational capability was completed in 1993. GPS was developed as a military system; however, it was later made available to civilian users and is now a dual-use system that can be accessed by both military and civilian users. Much of the following information is quoted or adapted from [2–4].

3.2.1 System Architecture

The GPS system is made up of three segments. These are the space segment, the control segment, and the user segment. Each segment is discussed in the following sections.

3.2.1.1 Space Segment

The space segment consists of a 24-satellite constellation with four satellites in each of six orbits. Orbits are nearly circular with a radius of 26,560 km, and a period of approximately 12 hours. Orbital planes are inclined at 55 degrees to the equatorial plane. This constellation is illustrated in Figure 3.8.

With this baseline constellation, almost all users worldwide would see a minimum of four satellites if not impeded by tall buildings in a large city such as New York. It is more likely that six to eight satellites would be in view. Each GPS satellite transmits a signal, which has a number of components: two sine waves or carriers, two digital codes, and a navigation message. The codes and the navigation message are added to the carriers as binary biphase modulation. The carriers and the codes are used mainly to determine the distance from the user's receiver to the GPS satellites. The navigation message contains, along with other information, the coordinates (the location) of the satellites as a function of time. The transmitted signals are controlled by highly accurate atomic clocks onboard the satellites.

3.2.1.2 GPS Satellite Generations

GPS satellite constellation build-up started with a series of 11 satellites known as Block I satellites. The first satellite in this series was launched on February 22, 1978. The last was launched on October 9, 1985. Block I satellites were built mainly for

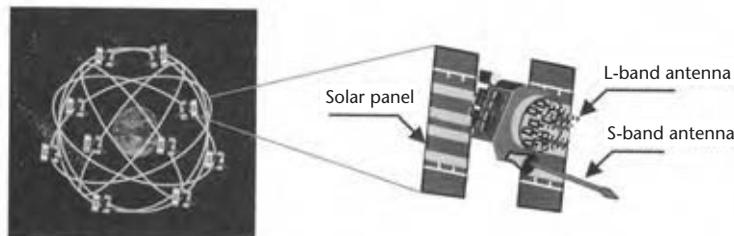


Figure 3.8 GPS constellation. (After: [2].)

experimental purposes. Although the design lifetime of Block I satellites was 4.5 years, some remained in service for more than 10 years. The last Block I satellite was taken out of service on November 18, 1995.

Detail about later GPS satellite systems are shown in Table 3.1.

The second generation of the GPS satellites is known as Block II/IIA. Block IIA is an advanced version of Block II, with an increase in the navigation message data storage capability from 14 days for Block II to 180 days for Block IIA. This means that these satellites can function continuously, without ground support for periods of 14 to 180 days. A total of 28 Block II/IIA satellites were launched during the period from February 1989 to November 1997. Of these 23 are currently still in service.

A new generation of GPS satellites is currently being launched. These are the Block IIR satellites. These will be followed by another system called Block IIF. It is clear from Table 3.1 that with each new generation the performance gets better and yet the unit cost gets lower.

3.2.1.3 Description of Satellite

Figure 3.9 is an illustration of a GPS satellite. Solar panels are used for prime power. An S-band receiver and antenna are used to receive information from the control segment. An L-band transmitter is used to send range measurement signals and navigation data to the user segment. The output power level for this transmitter is about 50W (17 dBW). The transmitter type is probably a traveling-wave tube (TWT) because this type has the best efficiency for the required power output. The antenna used with this transmitter is probably a helical antenna array. The reported antenna gain is about 14.7 dBi. The effective radiated power is reported as 487.63W or 26.8 dBW. The antenna polarization is circular, providing an important way to reduce the harmful effects of multipath. If the satellite antenna provides right-hand circular polarization signals, the user segment receiver would also use right-hand circular polarization antennas for no polarization mismatch loss. With multipath, the signals would be reflected as left-hand circular polarized signals and there would be large polarization loss, thus permitting rejection of the unwanted multipath signals.

The GPS satellite transmits a microwave signal composed of two carrier frequencies (or sine waves) modulated by two digital codes and a navigation message. The two carrier frequencies as illustrated in Figure 3.10(a) are generated at 1,575.42 MHz (referred to as the L1 carrier) and 1,227.60 MHz (referred to as the

Table 3.1 GPS Satellite Generations

	<i>Block II/IIA</i>	<i>Block IIR</i>	<i>Block IIF</i>
<i>Number</i>	28	21	12
<i>First launch</i>	1989	1997	2005*
<i>Satellite weight (kg)</i>	900	1,100	1,700*
<i>Power/solar panels (W)</i>	1,100	1,700	2,900*
<i>Design life (yrs)</i>	7.5	10	15
<i>Unit cost</i>	\$43 million	\$30 million	\$28 million*

*Estimates

Source: [2].

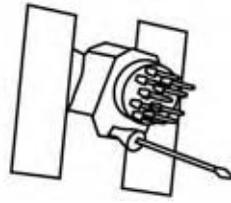


Figure 3.9 Illustration of GPS satellite. (After: [2].)

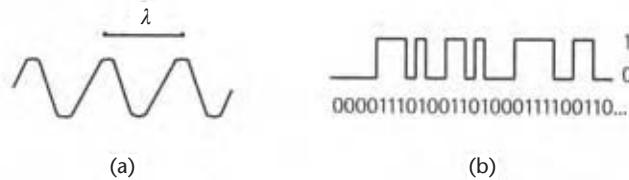


Figure 3.10 (a) Sinusoidal wave; and (b) digital code (After: [2].)

L2 carrier). The availability of two carrier frequencies allows for correcting a major GPS error known as ionospheric delay. The code modulation is illustrated in Figure 3.10(b). It is different for each satellite. This significantly minimizes the signal interference.

The two GPS codes are called course acquisition (C/A-code) and precision (P-code). Each code consists of a stream of binary digits, ones and zeros, known as bits or chips. The codes are commonly known as pseudorandom noise codes (PRN) because they look like random signals. The PRN codes are generated using digital circuits similar to that shown in Figure 3.11.

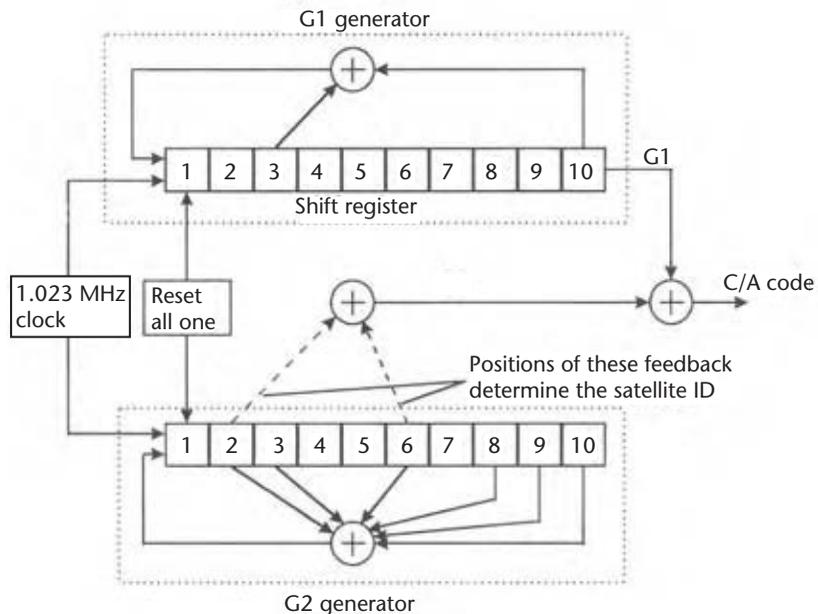


Figure 3.11 C/A code generator. (After: [3].)

Presently, the C/A-code is modulated onto the L1 carrier only, while the P-code is modulated onto both the L1 and the L2 carriers. The modulation type is biphasic modulation (0 and 180 degrees).

The C/A-code is a stream of 1,023 binary digits (i.e., 1,023 zeros and ones) that repeats itself every millisecond. Each satellite is assigned a unique C/A-code, which enables the GPS receivers to identify which satellite is transmitting a particular code.

The P-code is a very long sequence of binary digits that repeats itself after 266 days. It is 10 times faster than the C/A-code. The 266-day-long code is divided into 38 segments; each is one week long. Of these, 32 segments are assigned to the various GPS satellites.

Each GPS Block II and IIA satellite contains four atomic clocks, two cesium and two rubidium. Block IIR satellites carry rubidium clocks only. One of the clocks is selected to provide the frequency and the timing requirements for generating the GPS signals. The others are backups. The accuracy of atomic clocks is very high. The satellite clock error is only about 8.64 to 17.28 ns per day. One nanosecond error is equivalent to a range error of about 30 cm. These atomic clocks cost a few thousand dollars for rubidium clocks and about \$20,000 for cesium clocks. GPS receivers, in contrast, use inexpensive crystal clocks, which are much less accurate than satellite clocks.

3.2.1.4 Control Segment

The control segment of GPS consists of a master control station (MCS), a worldwide network of monitor stations, and ground control stations. The MCS located at Colorado Springs, Colorado, is the central processing facility of the control segment and is manned at all times.

There are five monitor stations, located in Colorado Springs (with the MCS), Hawaii, Kwajalein, Diego Garcia, and Ascension Island. The coordinates of these monitor stations are known very precisely.

Each monitor station is equipped with high quality GPS receivers and a cesium clock or oscillator for the purpose of continuous tracking all the GPS satellites in view. Three of the monitor stations (Kwajalein, Diego Garcia, and Ascension Island) are also equipped with ground antennas and transmitters for uploading information to the GPS satellites. All of the monitor stations and the ground control stations are unmanned and operated remotely from the MCS.

3.2.1.5 User Segment

The remarkable success of GPS for civil use is attributable almost entirely to the revolution in integrated circuits, which has made the GPS receivers compact, lightweight, and low cost. There are now hundreds of GPS receiver models on the market. Good quality handheld receivers are available for as little as \$200. It is estimated that more than 1 million receivers have been produced each year since 1997.

The received power level at the GPS receiver is very low as shown in Figure 3.12. At zenith and horizon the powers are at -160 dBw. The maximum power level is -158 dBw, which occurs at about 40 degrees. If the receiving antenna is taken into consideration, the received power will be modified by its antenna pattern.

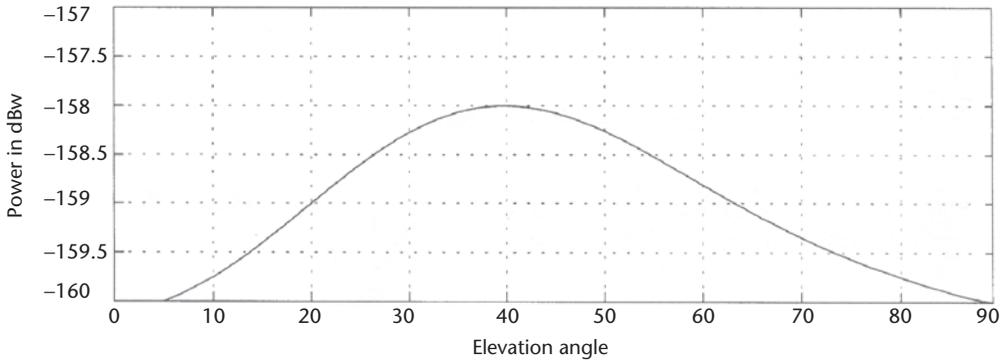


Figure 3.12 Receiver input power level versus elevation angle. (After: [4].)

The frequency spectrum of the C/A-code has a null-to-null bandwidth of 2.045 MHz. The frequency spectrum for the P-code has a null-to-null bandwidth of 20.46 MHz.

3.2.1.6 Receiver Hardware

The GPS receiver antenna should cover a wide spatial angle to receive the maximum number of signals. The common requirement is to receive signals from all satellites that are 5 degrees or more above the horizon. If an antenna is used to receive both the L1 (1575.42 MHz) and the L2 (1227.6 MHz) frequencies, the antenna can either have a wide bandwidth to cover the entire frequency range or have two narrow bands covering the desired frequency range. An antenna with two narrow bands can avoid interference from the signals between the two bands.

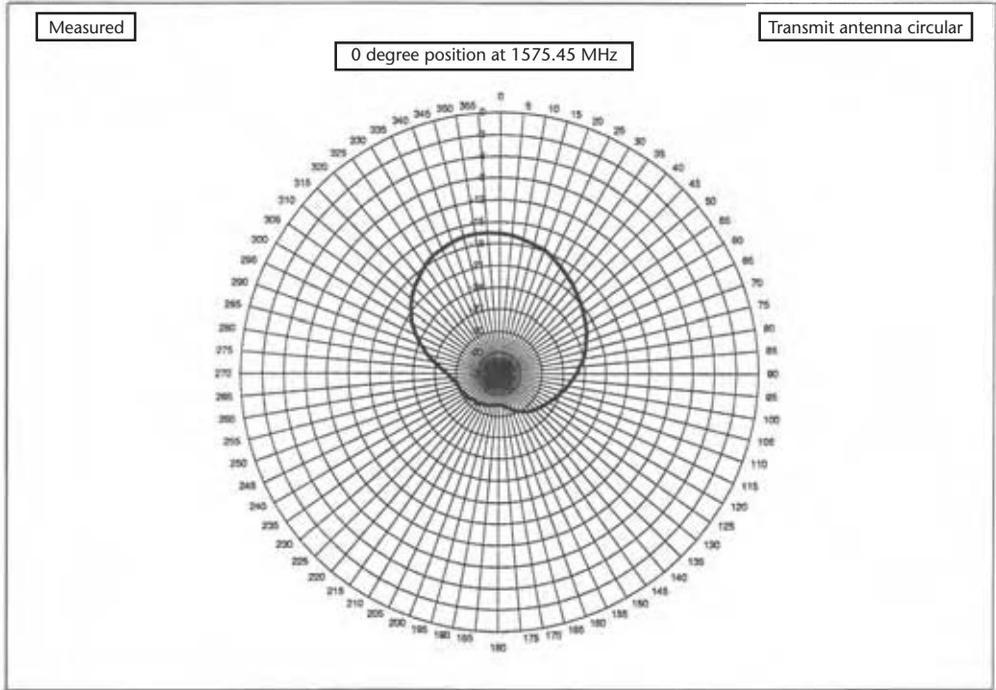
As pointed out earlier, the receiver antennas should use circular polarization to minimize the harmful effects of multipath. It is also desirable to use ground planes to minimize the multipath from the ground below the antenna. It is also desirable to use small antennas.

One common antenna design to receive a circular polarized signal is a spiral antenna, which inherently has wide bandwidth. Another type of popular design is a microstrip antenna called a *patch antenna*. If the shape is properly designed and the feed point is properly selected, a patch antenna can produce a circular polarized wave. The advantage of the patch antenna is its simplicity and small size.

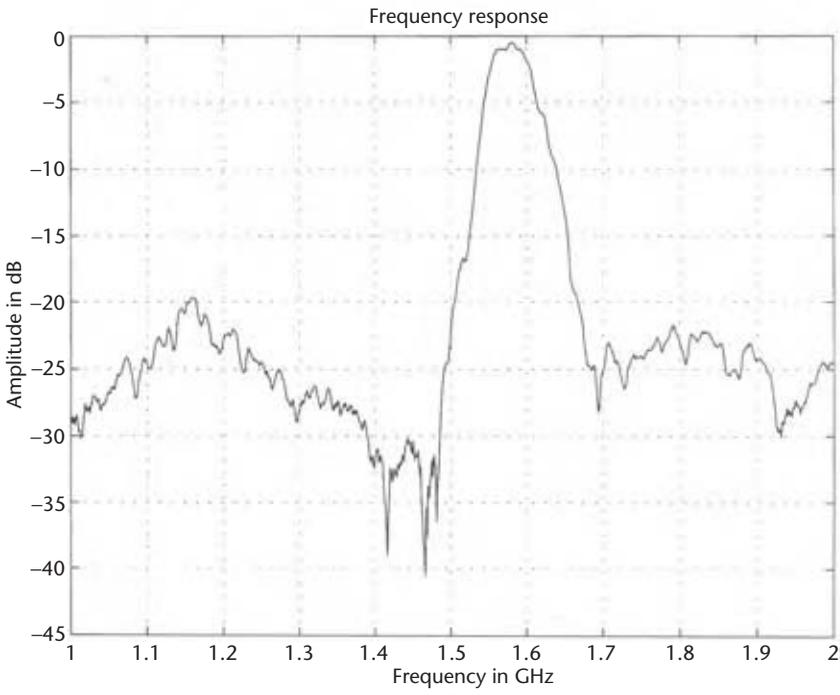
In some commercial GPS receivers the antenna is an integral part of the receiver unit. Other antennas are integrated with an amplifier. These antennas can be connected to the receiver through a long cable because the amplifier gain can compensate for the cable loss. A patch antenna (M/A COM ANP-C 114-5) with an integrated amplifier is discussed as follows. The internal amplifier has a gain of 26 dB with a noise figure of 2.5 dB. The antenna including the amplifier has a diameter of 3 inches and a thickness of 0.75 inch. The antenna pattern and frequency response are shown in Figure 3.13.

The estimated noise floor for a GPS receiver is about -111 dBm. If the GPS signal is at -130 dBm, the signal is 19 dB below the noise floor. Signal processing is used to greatly improve this signal-to-noise ratio. With this processing, the signal-to-noise ratio is generally adequate.

0_NR_CIR



(a)



(b)

Figure 3.13 (a) Measured patch antenna pattern; and (b) measured patch antenna frequency response. (After: [3].)

Figure 3.14 shows two arrangements for data collection or conversion from RF to digitized signals. In Figure 3.14(a), a low noise amplifier follows the antenna. A mixer is used to downconvert from 1,575.42 MHz to 21.25 MHz. Amplifiers 2, 3, and 4 are IF amplifiers which provide most of the gain. The overall gain of the four amplifiers and associated filters is about 100 dB. In Figure 3.14(b), most of the gain is in RF stages prior to the downconversion mixer. It is also possible to directly digitize at RF without downconverting.

The circuit of Figure 3.14(a) is usually preferred when pseudorange measurements using the C/A-codes or the P-codes are used because it is easier to build adequate bandpass filters at IF frequencies than it is at RF frequencies. When using carrier-phase measurements, it is best to use direct digitizing because of the increased accuracy of measurements.

In the circuits of Figure 3.14, the output stage is an analog-to-digital converter (ADC). This circuit provides the input to the digital architecture, which follows but is not shown. These digital circuits provide the functions of detecting and tracking the code signals, measuring the distance to the satellites, and computing the position of the receiver. The details of these systems are considered beyond the scope of this handbook. The interested reader is referred to [2–4].

The basic idea of GPS positioning is illustrated in Figure 3.15. Theoretically, only three distances to three simultaneously tracked satellites are needed to determine the position of the receiver. In this case, the receiver would be located at the

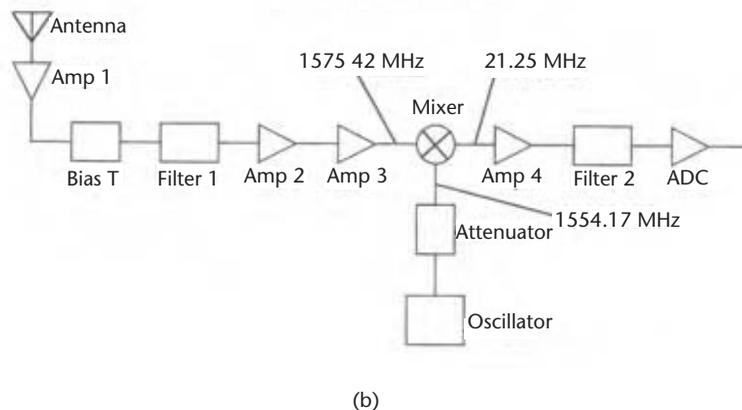
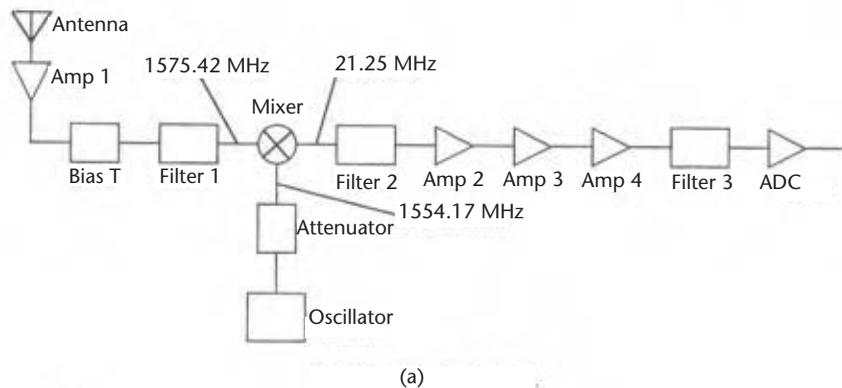


Figure 3.14 (a, b) Two receiver architectures for data collection. (After: [4].)

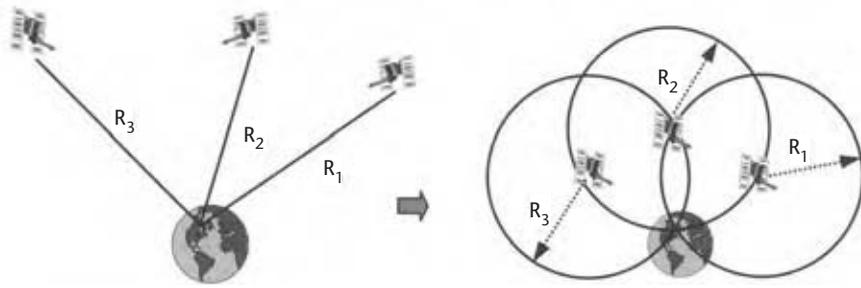


Figure 3.15 Basic idea of GPS positioning. (After: [3].)

intersection of three spheres; each has a radius of one receiver-satellite distance and is centered on that particular satellite. From the practical point of view, however, a fourth satellite is needed to account for receiver clock offset.

3.2.1.7 Pseudorange Measurements

The pseudorange is a measure of the range, or distance, between the GPS receiver's antenna and the GPS satellite's antenna. Either the C/A-code or the P-code can be used for measuring the pseudorange.

The procedure of the GPS range determination can be described as follows. Let us assume for a moment that both the satellite and the receiver clocks, which control the signal generation, are perfectly synchronized with each other. When the PRN code is transmitted from the satellite, the receiver generates an exact replica of that code. After some time, equivalent to the signal travel time in space, the transmitted code will be picked up by the receiver. The code generated at the receiver is shifted in time until an exact correlation exists between the two codes. This is equivalent to a matched filter, which provides pulse compression and maximum S/N integration gain. By measuring the time delay needed to bring the two codes into exact time correlation, the receiver can compute the signal travel time. Multiplying travel time by the speed of light (about 3×10^8 m/s) gives the range between the satellite and the receiver (see Figure 3.16).

Unfortunately, the assumption that the receiver and the satellite clocks are exactly synchronized is not entirely true. In fact, the measured range is contaminated, along with other errors and biases, by the synchronization error between the

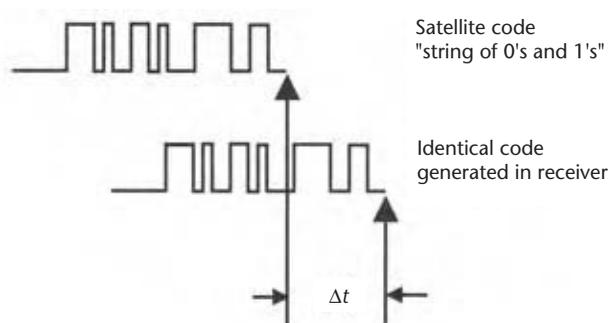


Figure 3.16 Pseudorange measurements. (After: [3].)

satellite and receiver clocks. For this reason, this quantity is referred to as the *pseudorange*, not the range.

GPS was designed so that the range determined by the civilian C/A-code would be less precise than that of the military P-code. This is based on the fact that the resolution of the C/A-code, 300m, is 10 times lower than the P-code. Surprisingly, due to the improvements in receiver technology, the obtained accuracy is almost the same from both codes.

3.2.1.8 Carrier-Phase Measurements

Another way to measure range to the satellites is to use carrier phase. The range would simply be the sum of the total number of full cycles plus fractional cycles at the receiver and the satellite multiplied by the carrier wavelength. This concept is illustrated in Figure 3.17.

The carriers are just pure sine waves and all cycles look the same. The receiver cannot determine the total number of complete cycles between the satellite and the receiver. It can only measure a fraction of a cycle very accurately (less than 2 mm), while the initial number of complete cycles remains unknown or ambiguous. Fortunately, the receiver has the capability to keep track of the phase changes after being switched on. That means that the initial cycle ambiguity remains unchanged over time, as long as no signal loss occurs.

It is clear that if the initial cycle ambiguity parameters are resolved, accurate range measurements can be obtained, which lead to accurate position determination. This high accuracy positioning can be achieved through the so-called relative positioning techniques either in real time or in the postprocessing mode. This requires two GPS receivers simultaneously tracking the same satellites in view.

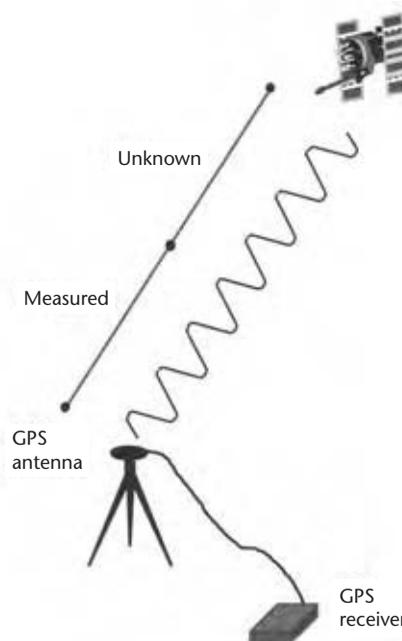


Figure 3.17 Carrier-phase measurements. (After: [3].)

The expected horizontal positioning accuracy with point positioning (using the PRN code for the civilian C/A-code receivers) is about 22m. GPS point positioning is used mainly when a relatively low accuracy is required. This includes recreation applications and low accuracy navigation.

The available position accuracy using GPS relative positioning is of the order of a subcentimeter to a few meters. Typical applications are mapping and precise navigation.

3.2.1.9 Differential GPS (D-GPS)

The following material is quoted or adapted from [5].

The differential GPS enhances standalone GPS accuracy and removes correlated errors from two or more receivers viewing the same satellites. In the basic form of DGPS, one of these receivers is called the monitoring or reference receiver and is surveyed in; that is, its precise position is known. The other receivers are called rovers or users and are in line of sight of the reference station. The reference station makes code-based GPS pseudorange measurements, just as any standard GPS receiver, but, because the monitoring station knows its precise position, it can determine the biases in the measurements. For each satellite in view of the monitoring station, these biases are computed by differencing the pseudorange measurement and the satellite-to-reference station geometric range. These biases contain errors incurred in the pseudorange measurement process. For real-time applications, the reference station transmits these biases, which are called *differential corrections*, to all users in the coverage area. The users incorporate these corrections to improve the accuracy of their position solution. Position errors less than 10m are typically realized.

To provide even greater submeter accuracy, DGPS techniques that utilize phase information of the satellite carrier frequencies have been developed. Extremely high accuracies (20 cm in dynamic applications and millimeter level for static applications) can routinely be achieved.

3.2.1.10 Other Satellite Navigation Systems

Other satellite navigation systems under development or in use include the GLONASS satellite system being developed by Russia, the Beidou Navigation System being developed by China, and the GALILEO navigation system being developed by Europe. Each of these systems is briefly discussed next. Some of the material presented is quoted or adapted from [2].

GLONASS Satellite System

The GLONASS satellite system has much in common with the GPS system. The nominal constellation of the GLONASS system consists of 21 operational satellites plus 3 spares at a nominal altitude of 19,100 km. Eight satellites are arranged in each of three orbital planes. The orbits are approximately circular, with an orbital period of 11 hours and 15 minutes and an inclination of 64.8 degrees. Similar to GPS, each GLONASS satellite transmits a signal that has a number of components: two L-band carriers, C/A-code on L1, P-code on both L1 and L2, and a navigation

message. However, unlike GPS, each pair of satellites transmits on its own carrier frequency in the bands 1,598.0625–1,604.25 MHz and 1,242.9375–1,247.75 MHz. Satellite pairs are placed on the opposite sides of the Earth so the user cannot see them simultaneously. GLONASS codes are the same for all satellites. As such, GLONASS receivers use the frequency channel rather than the code to distinguish the satellite. The chipping rates for the P-code and the C/A CODE are 5.11 and 0.511 Mbps, respectively. The navigation message is a 50-bps data stream. More details about this systems can be obtained from [3, 5].

Beidou System

The Chinese regional satellite navigation system (Beidou) is under development. The satellites are placed in geostationary orbits at an altitude of approximately 36,000 km above the Earth's surface. The primary use of the system is in land and marine transportation. China is also planning to build its second generation satellite positioning and navigation system, which will have more satellites and more coverage. The information about these systems is from [3].

Future European Global Satellite Navigation System (GALILEO System)

GALILEO is a satellite-based global-navigation system proposed by Europe. The selected system is a constellation of 30 medium Earth orbit satellites that will be evenly distributed over three orbital planes at an altitude of about 23,000 km. The signal characteristics of the GALILEO system were to be developed sometime in 2001 but are not known to this author at this time.

Galileo will be compatible at the user level with the existing GPS and GLONASS systems. However, unlike GPS and GLONASS, GALILEO will provide two levels of service: a basic, free-of-direct-charge service and a chargeable service that that offers additional features.

The GALILEO development plan will be divided into three different phases:

1. The definition phase was concluded at the end of 2000.
2. The development and validation phase began in 2001 and has been extended for a period of 4 years. This phase comprises a more detailed definition of the GALILEO system (e.g., frequency allocation). It also includes the construction of the various segments of the system (space, ground, and receiver). Some prototype satellites were launched in 2004, along with the establishment of a minimal ground infrastructure, to validate the system.
3. The constellation deployment phase is scheduled to begin in 2006 and extend until 2007. With the experience gained during the system validation phase, operational satellites will be gradually launched during this phase. In addition, ground infrastructure will be completed. The target date for gradual introduction of GALILEO operational service is 2008 or shortly thereafter.

3.2.2 Applications for GPS

There are many military and civil applications for GPS. Military applications include aircraft navigation, missile and other weapon guidance, land navigation, and ship navigation. Civil applications include GPS for the utilities industry, for

forestry and natural resources, for precision farming, for civil engineering applications, for monitoring structural deformations, for open pit mining, for land seismic surveying, and for marine seismic surveying. Other applications include airborne mapping, seafloor mapping, navigation for automobiles and other land vehicles, navigation for aircraft, navigation for ships, GPS for the retail industry, GPS for surveying, and GPS for recreation. Of these, automobile navigation and recreation probably have the greatest number of users. A good discussion of most if these applications is found in [3, 5]. As time goes on, undoubtedly there will be many other applications for GPS.

References

- [1] Robinson, E. A., "Radio Navigation Aids" Chap. 37 in *Reference Data for Engineers*, 8th ed., M. E. Van Valkenberg, (ed.), SAMS, Carmel, IN: Prentice Hall Computer Publishing, 1993.
- [2] Misra, P., and P. Enge, *Global Positioning System: Signals, Measurements, and Performance*, Lincoln, MA: Ganga-Jamuna Press, 2001.
- [3] El-Rabbany, A., *Introduction to GPS: The Global Positioning System*, Norwood, MA: Artech House, 2002.
- [4] Tsui, J., *Fundamentals of Global Positioning System Receivers: A Software Approach*, New York: John Wiley & Sons, 2000.
- [5] Kaplan, E. D., editors, "Understanding GPS Principles and Applications," Norwood, MA: Artech House, 1996.

Radar Systems

4.1 Basic Radar Concepts

The term radar refers to any type of radio frequency system in which the main objective is to gain information regarding remote objects or targets by means of reflections of transmitted signals. Radar is an acronym for radio detection and ranging. Types of information that can be obtained by radar include range or distance to targets, azimuth and elevation angles to targets, radar cross-section (RCS) of targets, and radial velocity of targets with respect to the radar. Some radars measure all those parameters, while others measure only some of them.

Radars are of two general types: pulse radar and continuous wave (CW) radar. With pulse radar, the range to the target is measured by the time delay between the short transmitted RF pulses and the received (reflected) RF pulses.

Angles are measured by using highly directional antennas and target-scanning methods. Radar cross-section of the target is determined by measuring the amplitude of the return signal. Velocity is measured by Doppler frequency measurements.

Three main types of RF pulses are used with pulse radar: pulsed CW, in which the frequency and phase of the signal remain constant during the short pulse period; pulsed FM, in which the frequency is swept in a linear fashion from f_1 to f_2 during the pulse period (chirp modulation); and pulsed phase code modulation (PCM), in which two-state coded phase modulation is used during the pulse period. The latter two methods of modulation are referred to as *pulse compression modulation* because the received pulses can be compressed into much shorter pulses than the transmitted pulses using appropriate circuits known as matched filters. That permits improved resolution and improved range accuracy compared to pulsed CW modulation. There are also other advantages including reduced vulnerability to electronic countermeasures.

With CW radar, range is measured by using either frequency or phase modulation of the CW signal and measuring the delay time between the transmitted waveform and the received, reflected waveform. Frequency-modulated CW is referred to as FMCW. Angles, radar cross-section, and velocity are measured in the same manner as in pulsed radar.

4.2 Radar Frequencies

Radars operate over a wide range of frequencies. Standard radar frequency bands and ITU allocations for radar are given in Table 4.1.

Each frequency region has its own set of characteristics that make it better than other frequencies for certain applications. The following discussion of characteristics is based in part on information presented in [2].

4.3 Types of Radars Based on Frequency

4.3.1 MF Radar (0.3–3.0 MHz)

With MF radar, the wavelength is in the range of 1000–100m. At these wavelengths a significant portion of the radiated energy can be propagated by ground wave and diffraction beyond the horizon. The lower the frequency, the lower the propagation loss. The main advantage for the MF radar is its over-the horizon capability. The disadvantages include requirement for very large antennas, large radar clutter levels,

Table 4.1 Standard Radar Frequency Bands

<i>Band Designation</i>	<i>Nominal Frequency Range</i>	<i>ITU Allocated Radar Frequency Bands for Region 2</i>
MF	0.3–3 MHz	1.85–2 MHz
HF	3–30 MHz	
VHF	30–300 MHz	138–144 MHz 216–225 MHz
UHF	300–1,000 MHz	420–450 MHz 890–942 MHz
L	1.0–2.0 GHz	1.215–1.4 GHz
S	2.0–4.0 GHz	2.3–2.5 GHz 2.7–3.7 GHz
C	4.0–8.0 GHz	5.25–5.925 GHz
X	8.0–12.0 GHz	8.5–10.68 GHz
K _u	12.0–18.0 GHz	13.4–14.0 GHz 15.7–17.7 GHz
K	18.0–26.5 GHz	24.05–24.25 GHz
W	26.5–40.0 GHz	33.4–36.0 GHz
V	40.0–75.0 GHz	59.0–64.0 GHz
W	75.0–110.0 GHz	76.0–81.0 GHz 92.0–100.0 GHz
Millimeter (mm) wave	110–300 GHz	126–142 GHz 144–149 GHz 231–235 GHz 238–248 GHz

Source: [3].

high ambient noise levels, and crowded electromagnetic spectrum. The result is that MF radar is not attractive for most radar applications.

4.3.2 HF Radar (3.0–30 MHz)

With HF radar, the wavelength is in the range of 100–10m. At this frequency range it is also possible to provide over-the-horizon radar with somewhat larger attenuation of the ground wave. It is also possible to provide very long range radar using ionospheric refraction. The military has used both bistatic and monostatic radars of this kind for detecting targets such as missiles in flight. Bistatic means that the transmitter and receiver antennas are not at the same location. In some bistatic systems, the antennas may be hundreds of miles apart. Monostatic means that the transmitter and receiver antennas are at the same location. In many cases, monostatic radar uses the same antenna for both transmit and receive functions.

The disadvantages of HF radar again include requirements for very large antennas, large radar clutter levels, high ambient noise levels, and crowded electromagnetic spectrum. The result is that HF radar, like MF radar is not attractive for most radar applications.

4.3.3 VHF Radar (30–300 MHz)

With VHF radar, the wavelength is in the range of 10–1m. VHF has been used for long range intercontinental ballistic missile (ICBM) early-warning radars and for satellite and aircraft surveillance radars. Such radars use very large array antennas and high transmitted power. External noise levels are much lower at VHF than at HF and lower frequencies. They are not as low as at higher frequencies such as L-, S-, and C-bands. VHF radars are not subject to unwanted weather echoes and atmospheric attenuation. Disadvantages of VHF radars include the need for large size antennas for good angle resolution.

4.3.4 UHF Radar (300–1,000 MHz)

With UHF radar, the wavelength is in the range of 1.0–0.3m. UHF is a good frequency band for reliable long range surveillance radar. It is free from weather effects and has good moving target indicator (MTI) capability. MTI capability is the ability to detect moving targets in a background of stationary radar clutter or ground return on the basis of Doppler frequency shift. MTI is discussed in more detail in Section 4.6. External noise levels for UHF radar are small and the required antenna size for good target angle resolution is not too great. Application of the UHF band for radar is limited by the wide frequency spectrum allotted to UHF television.

4.3.5 L-Band Radar (1.0–2.0 GHz)

With L-band radar, the wavelength is in the range of 30–15 cm. This frequency band is popular in the United States for aircraft surveillance radars. It has the advantages of improved angle resolution and low external noise. It is free from weather effects and atmospheric attenuation and has good MTI capability.

4.3.6 S-Band Radar (2.0–4.0 GHz)

With S-band radar the wavelength is in the range of 15–7.5 cm. S-band radars usually are not used for long-range surveillance but are used for precise target location and tracking. Good angular resolution can be achieved with reasonable size. External noise levels are low and the band is free from weather effects and atmospheric attenuation. S-band is a good compromise for medium range aircraft detection and tracking when a single radar must be used for both functions.

4.3.7 C-Band Radar (4.0–8.0 GHz)

With C-band radar the wavelength is in the range of 7.5–3.75 cm. C-band has been successfully used for moderate range surveillance applications in which precision information is necessary, as in the case of ship-navigation radar. It is also the frequency used for many precision long-range instrumentation radars as might be used for accurate tracking of missiles. Relatively long-range military weapon control radars also operate in this band.

4.3.8 X-Band Radar (8.0–12.0 GHz)

With X-band radar, the wavelength is in the range of 3.75–2.5 cm. This is a popular frequency band for military weapon control and for commercial applications. Civil marine radar, airborne weather-avoidance radar, and Doppler navigation radars are found at X-band. X-band radar antennas are fairly small and are favored where mobility and light weight are important. X-band permits large bandwidth capability and short pulses. Small-beam-angle antennas are also practical with small-size antennas.

At X-band there are significant weather effects and atmospheric attenuation. That makes it possible to use X-band for weather-detection radars. The two-way low-altitude attenuation at 10 GHz is about 0.024 dB/km.

4.3.9 K_U -Band Radar (12.0–18.0 GHz)

With K_U -band radar, the wavelength is in the range of 2.5–1.67 cm. K_U -band radars have the advantage of good resolution in both angle and range. High power is difficult to achieve, and the antennas are small. There is increased atmospheric attenuation and higher external noise. These factors result in relatively short range for K_U -band radar. Limitations due to rain clutter and attenuation are increasingly serious at the higher frequencies. The two-way low-altitude atmospheric attenuation is about 0.055 dB/km at 15.0 GHz.

4.3.10 K-Band Radar (18.0–27.0 GHz)

With K-band radar, the wavelength is in the range of 1.67 to 1.11 cm. The resonance frequency for water vapor is 22.2 GHz. At that frequency, the two-way low altitude atmospheric absorption is about 0.3 dB/km. High power is difficult to achieve, and antennas are small. These factors result in relatively short range for K-band radars. K-band radars have the advantage of good resolution in both range and angle.

4.3.11 K_a -Band Radar (26.5–40.0 GHz)

With K_a -band radar, the wavelength is in the range of 1.11 to 0.75 cm. The two-way atmospheric attenuation for K_a -band radar is about 0.14 dB/km at 35 GHz. High power is difficult to achieve, and antennas are small. These factors result in relatively short range for K_a -band radar. K_a -band radars have the advantage of good resolution in both range and angle.

4.3.12 V-Band Radar (40.0–75.0 GHz)

With V-band radar, the wavelength is in the range of 7.5 to 4.0 mm. V-band radar suffer the same limitations as the K-band radars, only more so. Oxygen molecules have resonance at 60 GHz, which is in the center of the V-band. At that frequency, the two-way atmospheric attenuation is about 35 dB/km. There are military radar applications where this very high attenuation is of value, for example, with preventing long-range detection of the V-band radars.

4.3.13 W-Band Radar (75–110 GHz)

With W-band radar, the wavelength is in the range of 4.0 to 2.73 mm. The two-way atmospheric attenuation at 95 GHz is about 0.8 dB/km. It suffers the same limitations as the other millimeter wave radars.

4.3.14 Millimeter-Wave Radar (110–300 GHz)

With millimeter-wave radar, the wavelength is in the range of 2.73–1.0 mm. The atmospheric attenuation at 140 GHz is about 1.0 dB/km. At 240 GHz, the atmospheric attenuation is about 15 dB/km.

4.4 Types of Radar

4.4.1 Surveillance Radars

Many ground-based surveillance radars are used throughout the world for the purpose of controlling air traffic flying to and from airports. That ability to detect and track is very important for the safety of the aircraft. A system of this type is illustrated in Figure 4.1.

Surveillance radars typically use L-band frequency with a rotating parabolic dish antenna that has a fan-shaped beam. The azimuth beam angle is small and the elevation beam angle is fairly large. It detects aircraft within line of sight, measuring their range, azimuth angle, and radar cross-section. The surveillance radar system used for aircraft surveillance may also include a secondary surveillance radar that works with an information friend or foe (IFF) beacon on the aircraft. The IFF unit is used to identify the aircraft and to receive status information from the aircraft. This includes such information as aircraft altitude, aircraft speed, direction of flight, and other information. The secondary surveillance radar operates at a frequency of 1.030 GHz for the ground-to-air link (interrogation) and 1.090 GHz for the air-to-ground link (reply). A fan-shaped beam is used here also.

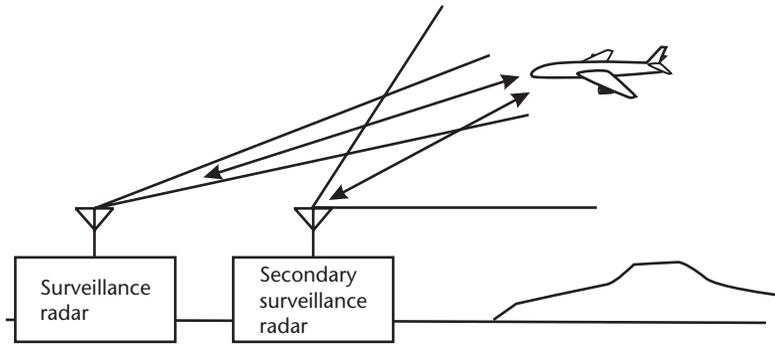


Figure 4.1 Ground-based surveillance radar.

A common type display used for surveillance radar is a cathode ray tube (CTR) plane position indicator (PPI), a map-like display. As the radar antenna scans in azimuth, the PPI display shows reflected signals received as bright spots on the circular screen. The degree of brightness is an indication of the radar cross-section of the target. The distance from the center of the display is an indication of the range to the target. The angle of the spot is an indication of the azimuth angle to the target.

4.4.2 CW Speed Measurement Radar

CW Doppler radar is widely used by the police and others to measure the speed of automobiles and other vehicles. These radars typically are small, solid-state microwave systems, which use Gunn diode oscillators as transmitters. Small antennas are used.

The two-way radar Doppler frequency is given by (4.1).

$$f_d = 2vf/c \quad (4.1)$$

where:

v = radial velocity of moving vehicle with respect to the radar

f = frequency of transmitted signal

c = speed of light (3×10^8 m/s)

If the target is moving toward the radar, the Doppler frequency is positive and the received signal frequency is higher than the transmit frequency. If the target is moving away from the radar, the Doppler frequency is negative, and the received signal frequency is lower than the transmit frequency.

4.4.3 Airborne Weather-Avoidance Radar

Weather-avoidance radar is used on commercial airlines and other civilian and military aircraft to identify regions of precipitation or clouds. Such information is important for the safety of the aircraft and the comfort of the passengers. A system of this type is shown in Figure 4.2.

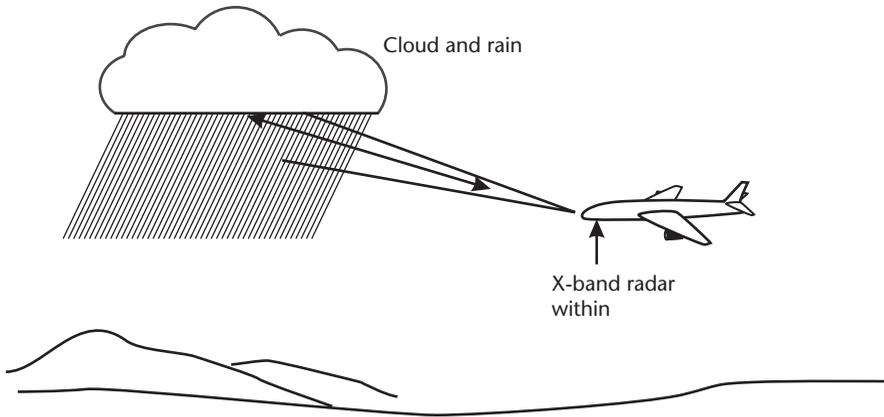


Figure 4.2 Airborne weather-avoidance radar.

Weather-avoidance radar can measure the range and the angle to clouds, rain, or snow. It also can indicate the rate of precipitation. X-band or higher frequency radars are needed to provide good reflection for clouds and precipitation.

4.4.4 Radar Altimeters

Radar altimeters, which are located on many aircraft, are used to accurately measure the altitude of aircraft. The radar type used typically is an FMCW radar and is simple in design and low in cost. The frequency band 4.2–4.4 GHz is reserved for radar altimeters. Figure 4.3 shows a radar system of this type.

4.4.5 Airborne Doppler Navigation Radar

Figure 4.4 shows an airborne Doppler navigation radar system. A system of this type may use three or four separate radar beams. In a four-beam system, two beams point down and forward, one on either side of the flight path, and two beams point rearward, one on either side of the flight path. The frequency of the return signals have a Doppler shift that is determined by the motion of the aircraft with respect to the ground. By comparing the Doppler frequency shift for each beam, it is possible to measure the aircraft's direction and rate of travel.

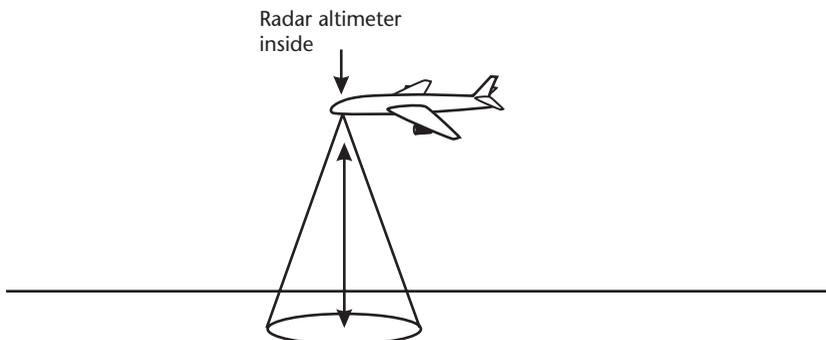


Figure 4.3 Airborne radar altimeter.

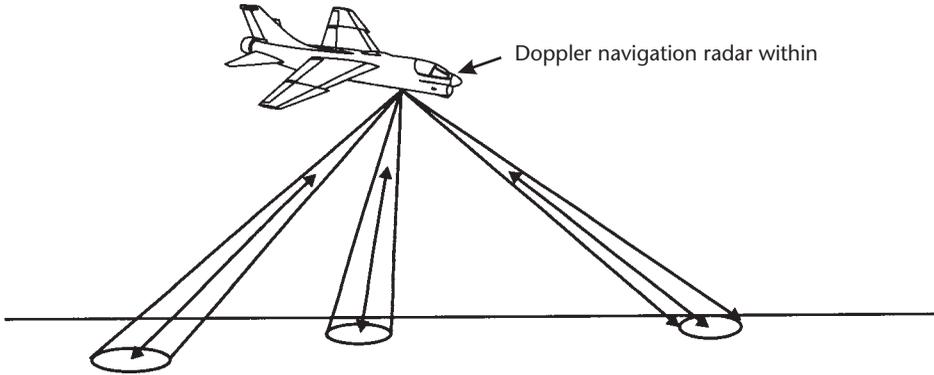


Figure 4.4 Doppler navigation radar.

4.4.6 Ship-Based Search and Surveillance Radar

Ship-based search and surveillance radars warn of potential collision with other ships or land objects. They also are used to detect navigation buoys. It must be able to do that in the presence of large radar clutter return from the water.

4.4.7 Shore-Based Search and Surveillance Radar

Shore-based radars of moderately high resolution are used for the surveillance of harbors as an aid to navigation. This radar is used to measure the range and angle to a ship or other target and the radar cross-section of the targets. It must be able to do that in the presence of large radar clutter return from the water.

4.4.8 Space Applications of Radar

Space vehicles use radar for rendezvous and docking and for tracking other man-made objects in space. Satellite-based radars also have been used for remote sensing of the Earth, planets, and other space objects. These radars measure the range to targets and the azimuth and elevation angles. They can also measure the radar cross-section of targets and the relative velocity of the targets with respect to the radar.

4.4.9 Ground-Based Instrumentation Radars for Locating and Tracking Missiles and Satellites

Some of the largest ground-based radars are used at test ranges for detection and tracking of missiles and satellites. These systems often use parabolic dish antennas with dish diameters of 60 to 90 feet. A radar of this type may operate at S-band or C-band with pulse powers of the order of 1.0 MW. Such systems sometimes use a pulse compression modulation known as chirp to improve the range resolution and tracking accuracy. As pointed out earlier, the term *chirp* refers to a type of modulation in which the frequency of the pulse is changed during the pulse period in a linear

sweep. A pulse compression circuit referred to as a *matched filter* is then used at the radar receiver to compress the return pulse.

4.4.10 Airborne Military Multiple Function Radars

Modern military aircraft use advanced multiple function radars. Functional capability includes weather detection, fire control, and terrain avoidance. Figure 4.5 illustrates a system of this type. Terrain avoidance is especially important for military aircraft, which often must fly at very high speeds very near to the ground.

4.4.11 Airborne Terrain-Following Radar

Cruise missiles and other military vehicles may use terrain-following radar to guide the missile to its target. Such a radar uses a down-looking radar to compare the terrain with a stored map in the vehicle's computer. The computer can then determine where the missile is at any given time. Errors in flight path thus can be detected and the necessary corrections made to allow the missile to fly to the desired target.

4.4.12 Airborne Side-Looking Radar

Figure 4.6 shows a front and a top view for an airborne side-looking radar. This type of radar may be used for high-resolution radar mapping of the ground. It is possible to take advantage of the forward motion of the aircraft to produce what is known as a synthetic aperture antenna of very large size. This allows the effective antenna azimuth beamwidth to be very small. The radar pulse width also is made very small by appropriate modulation to permit the range resolution to be small. This combination of small range resolution and small azimuth beamwidth permits the needed high resolution for a good quality radar map of the area to the side of the aircraft path. These radar mapping systems have the advantage over optical systems that they can “look” through clouds, rain, snow, fog, dust, and other cover. They have the disadvantage that their resolution is not as good as that of optical systems under clear conditions.

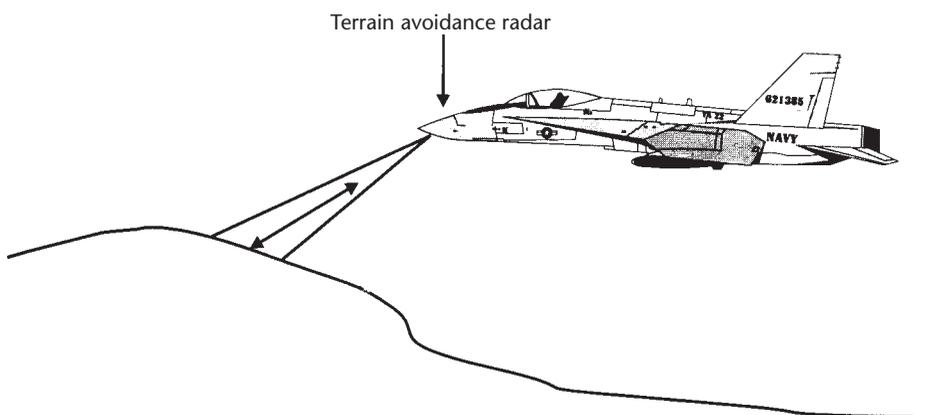


Figure 4.5 Airborne terrain avoidance radar.

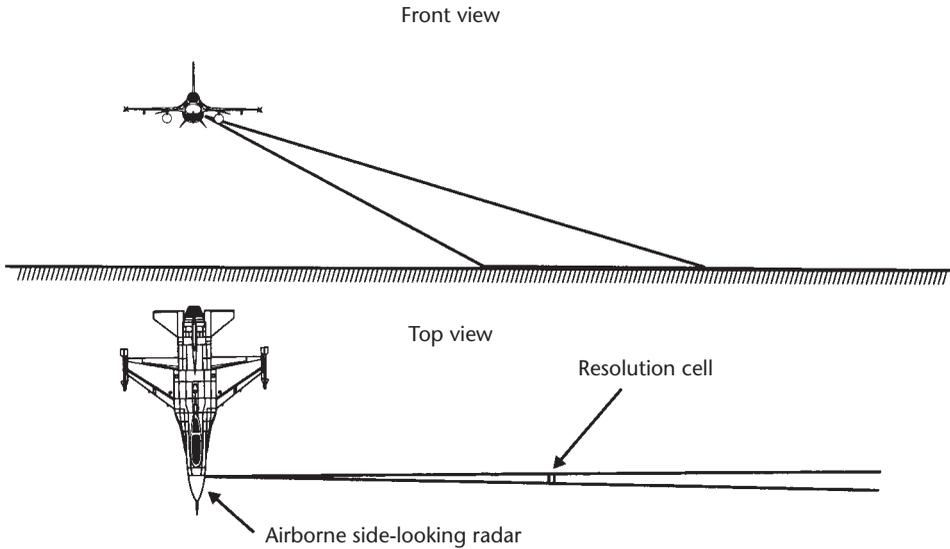


Figure 4.6 Airborne side-looking radar.

4.4.13 Ground-Based Military Radar for Locating and Tracking Aircraft and Missiles

A number of different types of military radars are used to detect and track aircraft and missiles. These are also used for fire control and pointing of weapons against such threats. These radars normally have pulse-Doppler or MTI capability.

4.4.14 Ground-Based Military Radars for Ground-Based Targets

Ground-based military radar is also used to locate and track moving targets on the ground such as tanks, trucks, and other vehicles. These radars must be able to see the targets in the presence of much larger return from the ground (referred to as ground clutter). Ground-based radars are also used to locate gun or mortar emplacements and for fire control and tracking.

4.4.15 Ship-Based Military Radar

Ship-based military radar is used to detect and track ships, aircraft, missiles, and other targets. It is also used for fire control and pointing of weapons against such threats. One of the U.S. Navy's most demanding tasks for radars is the detection and tracking of very low-flying missiles that travel only a few feet above the water. The radar must be able to see the missile in the presence of very much larger return from the water (referred to as sea clutter)

4.4.16 ICBM Defense Radars

Experimental ground-based radars have been developed and tested for ICBM defense systems. Types of radars for that function include long-range VHF early warning radars, UHF and L-band acquisition radars, and C-band target tracking

radars. The early warning radars are very large because of the large wavelength involved. Acquisition radar are also large but not so large as the VHF radars. They provide the needed information for the much smaller C-band tracking radars to lock on to the targets.

4.5 Radar Measurement

4.5.1 Radar Cross-Section of Targets

The RCS, σ , of a target is a measure of the effective cross-section area of the target if the assumption is made that the radar power to the target is scattered equally in all directions (through 4π steradians). The RCS depends not only on the size of a target but also on the shape of the target, the materials involved, the aspect angle at which the target is viewed, and the frequency of the radar signal. Some targets have very small radar cross-sections, even though they have large physical areas, because the signal is not scattered back toward the radar. On the other hand, some very small targets have very large RCSs because most of the energy that reaches the target is scattered back in the direction from which it came. A corner reflector is an example of that type of target.

The only target whose RCS is independent of viewing aspect angle is a conducting sphere. For that reason, metallic spheres often are used to calibrate radar systems. Figure 4.7 shows the ratio of the RCS of a sphere divided by the physical cross-section as a function of circumference and wavelength. As shown in the figure, there are three main regions of RCS. The first region is the Rayleigh region, where the RCS is much smaller than the physical cross-section of the sphere and falls off as the fourth power of the circumference-wavelength ratio. The second region is the resonance region, where the amplitude of the RCS oscillates up and down with increasing circumference over wavelength. The third region is the optical region, where the ratio of RCS to the cross-sectional area is nearly 1.

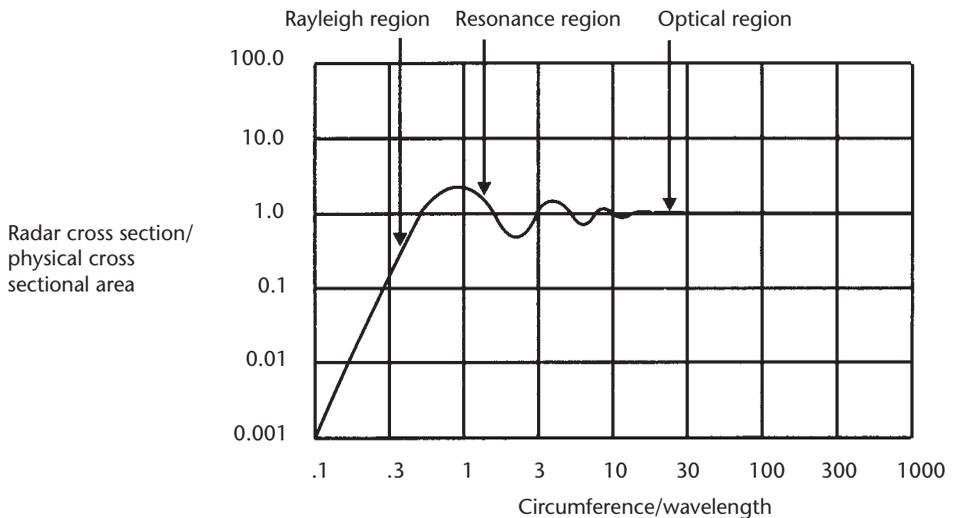


Figure 4.7 RCS for a sphere. (After: [1].)

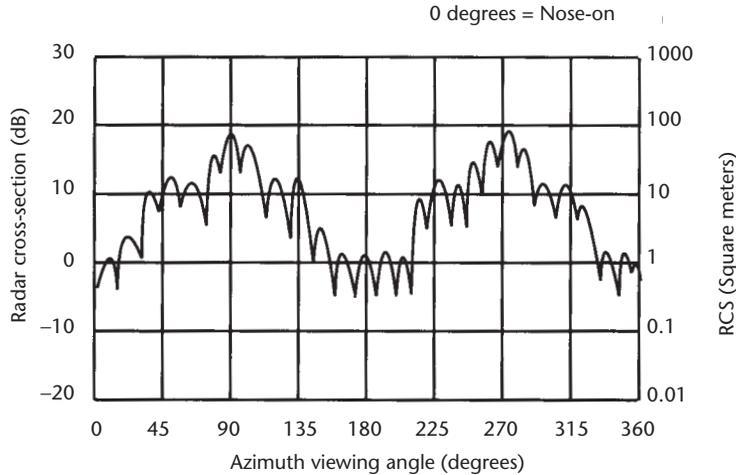


Figure 4.8 RCS for an aircraft as a function of azimuth viewing angle with 0-degree elevation angle.

For an example of the use of Figure 4.7, assume operation at a wavelength of 1.0m and a sphere diameter of 0.1m. The circumference is 0.314m, and the circumference-to-wavelength ratio is 0.314. Operation is thus in the Rayleigh region. From Figure 4.7, the ratio of the RCS to the physical cross-section area is about 0.1. The cross-section area is $7.85 \cdot 10^{-3} \text{ m}^2$; thus, the RCS is $7.85 \cdot 10^{-4} \text{ m}^2$. That is a very small RCS. On the other hand, if we assume operation at a wavelength of 0.03m and that same sphere, the circumference-to-wavelength ratio is 10.47. Operation is in the optical region, and the ratio of the RCS to the physical cross-section area is about 1.0. The RCS of the sphere thus is about $7.85 \cdot 10^{-3} \text{ m}^2$. We see a factor-of-10 difference in the RCS for the two frequencies. When a sphere is used to calibrate a radar, the size of the sphere usually is chosen so that operation will be in the optical region.

The general ideas regarding radar scattering in the three different regions apply to objects of other shapes as well as to spheres. Hence, all target RCSs are wavelength sensitive. All targets except spheres have RCS values that change with aspect angle. For example, an aircraft typically has a much larger RCS when viewed from a 90-degree aspect angle than when viewed nose on. Its RCS is characterized by many peaks and nulls as the azimuth aspect angle changes. A typical RCS pattern for an aircraft is shown in Figure 4.8.

The lobbing structure of the RCS signature or pattern changes greatly with frequency. At lower frequencies, there are fewer peaks and nulls than there are at the higher frequencies. The RCS can be viewed as the vector sum of many individual scatters located at different distances and therefore having different phase angles.

It is possible to reduce greatly the RCS of military aircraft by designing the aircraft so no surfaces provide specular reflections, such as is done with the Stealth bomber. It is also possible to reduce the RCS of targets by using special radar absorbing material (RAM) coatings. RAMs are frequency and thickness dependent. The thickness should be nearly one-quarter wavelength at the frequency of interest. Table 4.2 lists some RCSs for targets at microwave frequencies. The values given are average or typical values. Peak values could be 5–10 dB greater [1].

Table 4.2 RCSs of Some Targets at Microwave Frequencies (3–10 GHz)

<i>Type of Target</i>	<i>RCS (m^2)</i>
Conventional unmanned winged missile	0.5
Small single-engine aircraft	1.0
Small fighter aircraft	2.0
Four-passenger jet aircraft	2.0
Large fighter aircraft	6.0
Medium bomber or medium jet airliner	20.0
Larger bomber or large jet airliner	40.0
Jumbo jet airliner	100.0
Ships	
Small open boat	0.2
Small pleasure boat	2.0
Cabin cruiser	10.0
Large naval ship	10,000.0
Land Vehicles	
Bicycle	2.0
Automobile	100.0
Pickup truck	200.0
Nonmetallic Targets	
Insect	0.0001
Bird	0.01
Human	1.0

Source: [1].

The RCSs for most targets tend to increase with increasing frequency. They are greatest when large flat surfaces are viewed normal to the surfaces. Note that the RCS of a pickup truck is typically larger than that of even a large aircraft. That is because the pickup truck has flat surfaces that can be normal to the viewing angle, whereas the aircraft has curved or slanted surfaces with respect to the viewing angle. A large naval ship has a very large RCS for the same reason. A human has an RCS of about 1 m^2 , while a bird has an RCS of about 0.01 m^2 . A group of birds would have a peak RCS many times that of a single bird. In a similar way, a large swarm of insects would have a significant RCS, even though that of a single insect is very small.

It is typical to have large RCS fluctuations as a moving target is viewed over a period of time. That is due to the fact that at one period of time the target may present an RCS near a peak in the scatter pattern, while a short time later it may present a null in the scatter pattern.

4.5.2 Radar Clutter

Another important type of RCS is the RCS of the Earth and objects on the Earth as viewed by the radar. When that signal is not desired, it is referred to as clutter. Some of the information in this section is adapted from [1–3].

Figure 4.9 shows the geometry of main beam radar clutter. The clutter patch from which the RCS is determined is shown in the plan view. The size of the clutter patch depends on the antenna azimuth beam angle, the elevation or grazing angle, and the pulse width.

RCS for clutter is computed from an estimate of per-unit surface-area RCS times the projected antenna beam and range resolution footprint area. This area is given by (4.2).

$$A = 2R \tan(\theta_b/2) \cdot 150\text{m} \cdot \tau / \cos \theta_{el} \quad (4.2)$$

where:

R = range (meters)

θ_b = antenna azimuth beam angle (degrees)

τ = radar pulse length (microseconds)

θ_{el} = antenna beam elevation grazing angle (degrees)

For an example of the clutter RCS for land, assume the following:

Per-unit RCS = -30 dBsm = 0.001m^2

Range = 100 km

Antenna azimuth beam angle = 2 degrees

Radar pulse length = $1 \mu\text{s}$

Grazing angle = 10 degrees

Substituting those values into (4.2), the area of the clutter patch is

$$517,766\text{m}^2$$

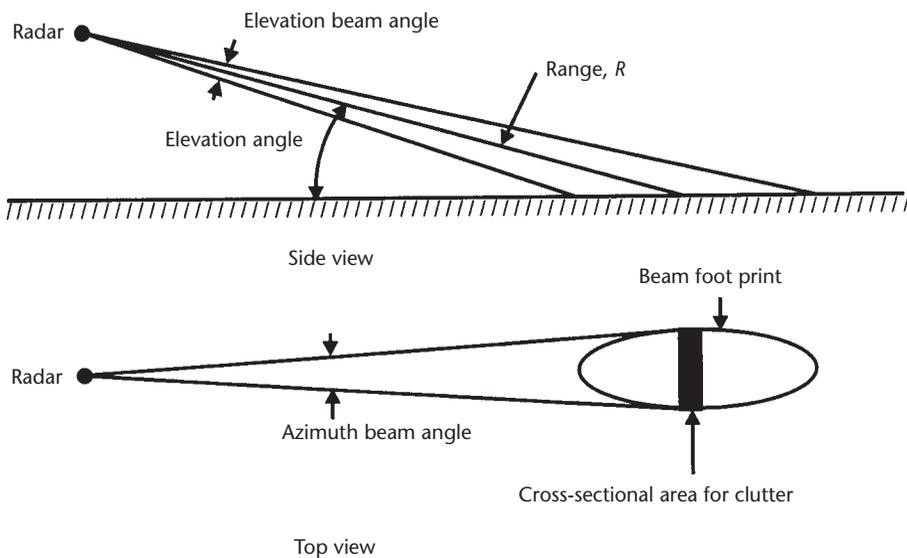


Figure 4.9 Geometry for radar clutter.

The RCS is

$$s = .001 \cdot 517,766 = 517.8\text{m}^2$$

There are a number of ways to help reduce the clutter in radar systems. One way is to use highly directional antennas with low sidelobes. Another way is to use short-duration pulses. Each of these methods helps reduce the area or the volume of the clutter and, therefore, the RCS of the clutter.

Figure 4.10 shows typical land clutter reflectivity or RCS per unit area expressed in decibels. The value depends on the grazing angle and the type of surface. For microwave frequencies, it tends to be independent of frequency. Types of surfaces shown are flatland, farmland, wooded hills, and mountains.

For an example of the use of Figure 4.10, assume a grazing angle of 3.0 degrees and wooded hills. For that condition, the indicated RCS per unit area is -20 dBsm. Thus, if the clutter patch is $100,000 \text{ m}^2$, the RCS is 0.01 times that area, or $1,000 \text{ m}^2$.

Figure 4.11 shows typical sea clutter reflectivity or RCS per unit area as a function of grazing angle and frequency. The values shown are for medium sea conditions. The values would increase as the wind speeds increase and the sea state becomes rougher. There is very little backscatter for a calm sea.

For an example of the use of Figure 4.11, again assume a grazing angle of 3.0 degrees and operation at X-band. We can see from the figure that the per-unit RCS is about -30 dBsm. If the clutter patch is $100,000 \text{ m}^2$, the RCS is 0.001 times that area, or 100 m^2 .

Another type of radar target is weather (clouds, fog, rain, or snow). Weather-detecting radars are used both on the ground and on aircraft to detect weather features. Rain or snow can be viewed as clutter when it interferes with the detection of other desired targets. Radar signals are backscattered from range

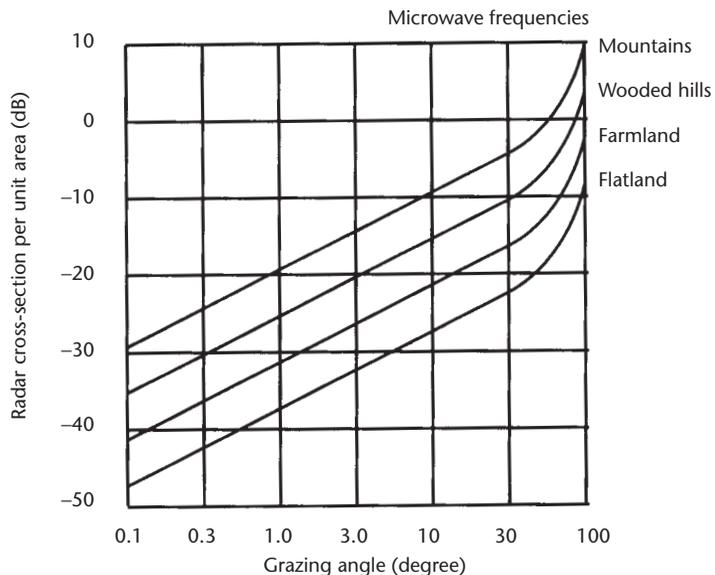


Figure 4.10 Land-clutter reflectivity versus grazing angle. (After: [2].)

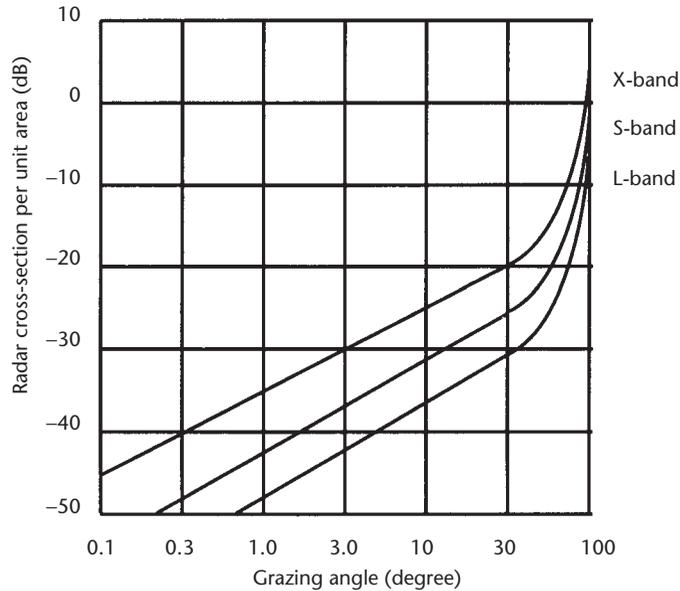


Figure 4.11 Sea-clutter reflectivity versus grazing angle. (After: [2].)

resolution cells or volumes having lengths defined by the pulse width ($150 \text{ m}/\mu\text{s}$), and cross-sectional areas are defined by the range and the azimuth and elevation beam angles. Thus, if for a range resolution cell at a range of 100 km , an elevation beam angle of 2 degrees, an azimuth beam angle of 4 degrees, and a pulse width of $2 \mu\text{s}$, the volume of the clutter cell would be $150 \text{ m}/\mu\text{s} \cdot 2 \mu\text{s} \cdot 2 \tan 1 \text{ degree} \cdot 2 \tan 2 \text{ degrees} = 300 \cdot 3,492 \cdot 6,993 = 7.3 \cdot 10^9 \text{ m}^3$. If the RCS per cubic meter is 10^{-8} m^2 per cubic meter of clutter cell, the clutter RCS would be 73 m^2 .

Figure 4.12 shows the volume reflectivity of rain as a function of precipitation rate and wavelength. We see that the RCS increases rapidly with increasing frequency. Thus, if we have a precipitation rate of 10 mm/hr , the per-unit RCS is about 10^{-12} m^2 per cubic meter with a wavelength of 100 cm (frequency = 300 MHz). At a wavelength of 2 cm (frequency = 15 GHz) and the same precipitation rate, the per-unit RCS is about 10^{-5} m^2 per cubic meter. That is a factor of 10^7 difference for a factor of 50 increase in frequency.

Snow has an RCS similar to that of rain for very low precipitation rates but can have as much as a factor of 10 or more increase in RCS for high precipitation rates ($>30 \text{ mm/hr}$).

4.5.3 Range Measurements

The main types of measurements made by radars include the range or distance to the target, the pointing angle to the target, the radial velocity of the target with respect to the radar, and the RCS of the target. Figure 4.13 shows concepts for range measurements using a pulse radar. (Block diagrams for radars of this kind are presented in Section 4.6.)

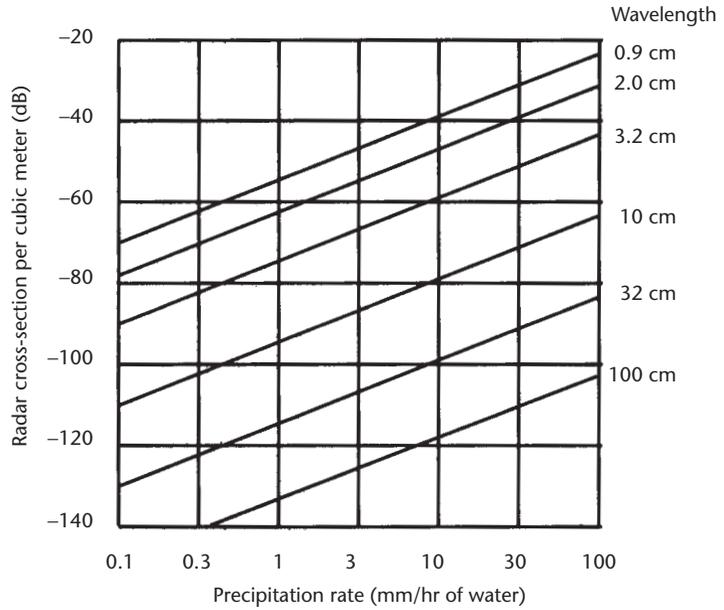


Figure 4.12 Volume reflectivity of rain. (After: [2].)

The radar transmits a train of short-duration RF pulses using a directional transmitting antenna. The pulse width must be less than the expected round-trip time delay.

The pulse width determines the range resolution of the radar. If small range resolution is needed, the pulse width must be small. The range resolution is the minimum distance between two targets that can be seen by the radar as separate targets.

A small part of the power transmitted is reflected back to the radar receiving antenna. The received signal is fed to the radar receiver, where it is amplified, down-converted in frequency, and detected. The two trains of pulses are shown in Figure 4.13. The radar measures the delay time between the two sets of pulses. The range can then be computed using (4.4).

$$R = TC/2 \tag{4.3}$$

where:

R = range (meters)

T = delay time (seconds)

c = speed of light = $3 \cdot 10^8$ m/s

From (4.3), we see that the radar range is equal to 150 m/ μ s of delay. For example, if the measured delay is 200 μ s, the radar range to target is $200 \cdot 150 = 30,000$ m = 30 km.

Figure 4.14 shows concepts for range measurements using FMCW-type radar. The radar transmits a CW signal with the frequency changing as a function of time. A triangular frequency pattern is shown in Figure 4.14. A directional transmitting antenna transmits the signal toward the target. A small part of the power

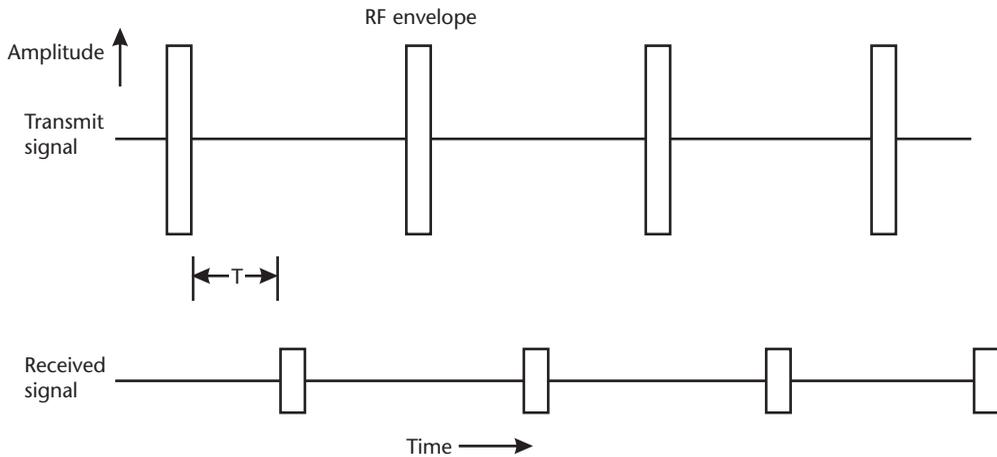


Figure 4.13 Concepts for range measurements using pulse-type radar.

transmitted is reflected back to the radar receiving antenna. The two frequency patterns are shown in Figure 4.14.

The received signal is fed to the radar receiver, where it is mixed with the transmitter signal, and the difference frequency is amplified and detected. The difference in frequency between the transmit signal and the receive signal is a measure of the time delay between the two signals. The range then can be computed using (4.4).

For example, assume that the transmitted signal changes linearly from 3.1 GHz to 3.2 GHz in a period of 0.001 second. The rate of change of frequency is thus $10^8/0.001 = 10^{11}$ Hz/s. If the measured frequency difference is 50 MHz, the corresponding delay is $5 \cdot 10^7/10^{11} = 5 \cdot 10^{-4}$ sec = 500 μ s. Substituting values into (4.3), the corresponding range is

$$R = TC/2$$

$$R = 500 \cdot 150 = 75 \text{ km}$$

4.5.4 Velocity Measurement Using CW Radar

Figure 4.15 shows the concepts for target radial velocity measurements using a CW radar. The radar transmits a CW signal with constant frequency and amplitude. A small part of the power transmitted is reflected back from the target to the radar receiving antenna. The signal has a frequency equal to the transmitted frequency plus or minus the Doppler frequency, depending on whether the target is moving toward the radar (plus) or away from the radar (minus).

The received signal is fed to the radar receiver mixer, where it is mixed with part of the transmitter signal that is reflected back from the antenna. A circulator or diplexer typically is used to direct the transmit signal to the antenna and the receive signal to the receiver mixer.

The difference frequency output from the mixer is the Doppler frequency. The signal is selected by a bandpass filter and then is amplified by a beat-frequency amplifier and fed to an indicator. The two-way radar Doppler frequency is given by

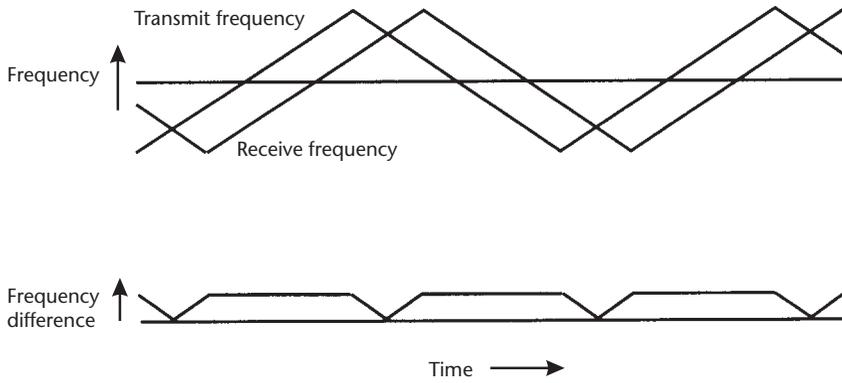


Figure 4.14 Concepts for range measurements using FMCW radar.

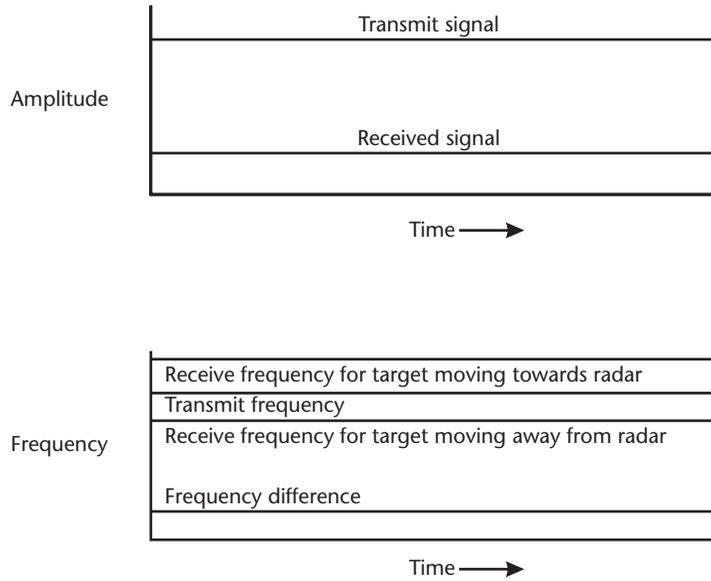


Figure 4.15 Concepts for target radial velocity measurements using a CW radar.

$$fd = 2V_r f_0 / c \tag{4.4}$$

where v_r = radial velocity with respect to the radar.

The radial velocity of the target with respect to the radar can be computed as follows:

$$v_r = f_d c / (2f_0) \tag{4.5}$$

where:

v_r = target radial velocity with respect to the radar

f_d = measured Doppler frequency

c = speed of light = $3 \cdot 10^8$ m/s

f_0 = transmitted frequency

For an example of the use of (4.5), assume that the transmit frequency is 9.2 GHz and the measured Doppler frequency is 1,000 Hz. Then, from (4.5)

$$v_r = 1,000 \cdot 3 \cdot 10^8 / (2 \cdot 9.2 \cdot 10^9)$$

$$v_r = 163 \text{ m/s} = 58.7 \text{ km/h} = 36.5 \text{ mi/h}$$

4.5.5 Velocity Measurements Using FMCW Radar

Figure 4.16 shows the concepts for target radial velocity measurements using an FMCW radar. The radar transmits a CW signal with triangular frequency modulation as a function of time, as shown in the figure. A directional transmitting antenna transmits the signal toward the target. A small part of the power transmitted is reflected back to the radar receiving antenna. The received signal is fed to the radar receiver, where it is mixed with the transmitter signal, and the difference frequency is amplified and detected. The transmitted signal frequency and the received signal frequency are shown for the condition where the received signal is shifted in frequency by the Doppler effect. The difference frequency for each half of the cycle is shown. The difference between the two difference frequencies is equal to twice the Doppler frequency. The measured Doppler frequency can be used to determine the radial velocity of the target with respect to the radar using (4.5).

4.5.6 Velocity Measurements Using a Pulse-Type Radar

Figure 4.17 shows the concepts for velocity measurements using a pulse-type radar. The radar transmits a train of short-duration RF pulses using a directional transmitting antenna. A small part of the power transmitted is reflected back to the radar receiving antenna. The received signal is fed to the radar receiver, where it is amplified, converted in frequency, and detected using a coherent phase-sensitive detector. The two trains of pulses are shown in Figure 4.17.

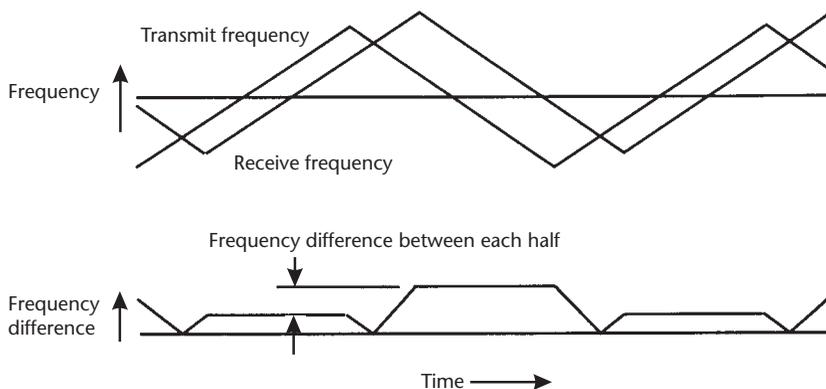


Figure 4.16 Concepts for velocity measurements using the FMCW-type radar.

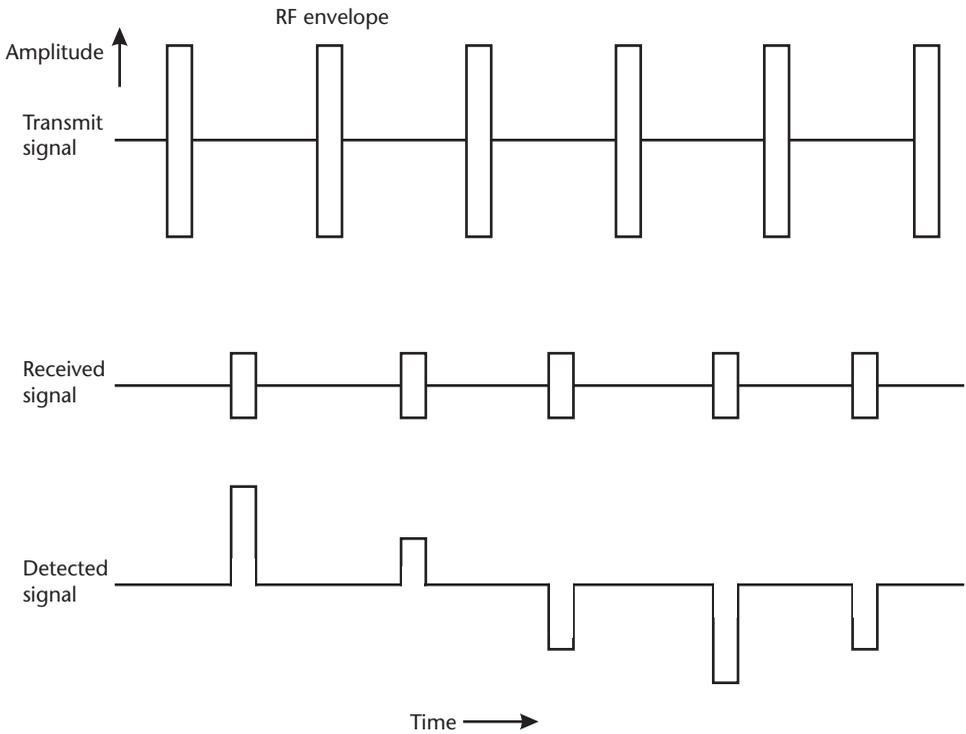


Figure 4.17 Concepts for velocity measurements using a pulse-type radar.

With a phase-sensitive detector, the amplitude of the detected pulses will be constant only if there is no radial velocity for the target. If there is a radial velocity, there will be a Doppler frequency shift, and the amplitude of the received pulse train will oscillate in amplitude at the Doppler frequency rate, as shown in, Figure 4.17. The rate of oscillation can be detected using appropriate signal processing circuits. The measured Doppler frequency then can be used with (4.5) to determine the target velocity.

4.5.7 Angle Measurements for Radars

Angle measurements for radars depend on having highly directional antennas and means for reading out the pointing angles for the antennas. Very high mechanical precision is possible for such radars as the old but reliable FPS-16 tracking radar as well as for many of the newer tracking radars.

Receiver antenna beams often are split, as shown in Figure 4.18. A process known as sequential lobbing, shown in Figure 4.18(a), can be used in tracking radars to permit pointing-error detection and pointing correction.

Another way to determine pointing errors is to use conical scan tracking shown in Figure 4.18(b). As the pencil beam is scanned in a circle with a squint angle, the return signals have amplitudes that change as a function of the scan angle. The information then can be used to correct the pointing for minimum error signal.

Monopulse tracking, shown in Figure 4.18(c), is frequently used for high-accuracy angle measurement and tracking. Monopulse radar uses split beams similar to

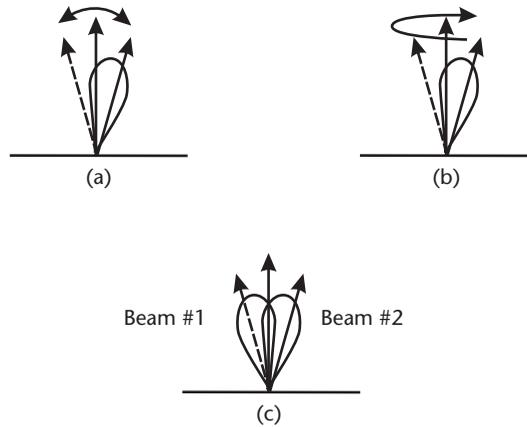


Figure 4.18 Concepts used for high-accuracy angle tracking and measurements: (a) single-beam sequential lobbing; (b) single-beam conical scan; and (c) two-beam monopulse tracking.

that used for sequential lobbing. The difference is that measurements can be made on a single pulse rather than on two or more sequential pulses. That leads to greater accuracy since errors due to fading are avoided.

Monopulse radars may use a single beam for transmit and two or more beams for receive. A four-beam system frequently is used for tracking in both azimuth and elevation. Again, the idea is to detect pointing errors by the difference in amplitude of the signals received in the several antennas. The detected errors then can be used to correct the pointing of the tracking radar antennas. Mechanical readouts of pointing angles are provided.

4.6 Moving-Target Indicator (MTI) and Pulse Doppler Radars

4.6.1 MTI Radar Types

Some of the information in this section is adapted from [4]. MTI radar takes advantage of the velocity-measuring capability of pulse radar to filter out unwanted radar signals that come from clutter. In an ideal MTI radar, only those signals having significant Doppler shift are received. In a practical MTI radar, there will be large clutter rejection, but filtering will not be perfect.

Pulse Doppler radars (PDRs) are similar in function to MTI radars. The main difference is in the PRF used by the two types of radars. MTI usually refers to a radar in which the PRF is chosen low enough to avoid multiple-time-around echoes. The problem here is that there can be ambiguous Doppler frequency measurements. PDR, on the other hand, has a high PRF, which may result in ambiguities in range. In other words, we do not know to which of several possible ranges the return signal corresponds. The main advantage of the PDR is that, because of the high PRF, there are no blind speeds or ambiguous velocity measurements.

Figure 4.19 is the block diagram for an MTI radar with a power amplifier transmitter. The power amplifier may be an amplifier chain with a modulator stage, several solid-state amplifiers in series, and a high-power output amplifier. Possible power amplifier types include CCTWT amplifiers, pulse klystron amplifiers, or twystron amplifiers.

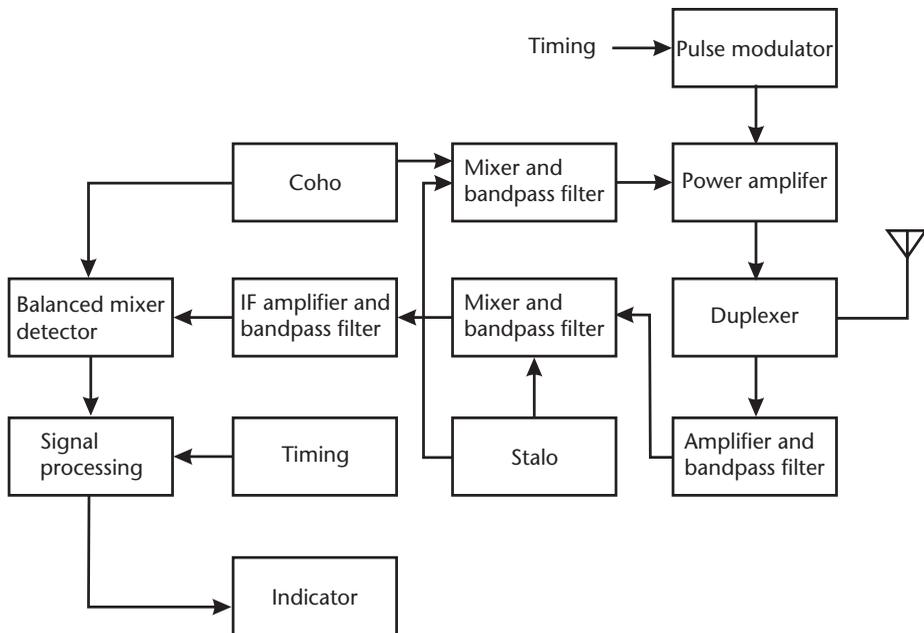


Figure 4.19 MTI radar with power amplifier transmitter. (After: [4].)

The power amplifier input CW signal is modulated by a pulse modulator, which receives inputs from the timing circuit of the radar. The CW RF input to the amplifier system is from an upconverter that consists of a mixer and a bandpass filter. One input to the mixer is a CW RF signal from a stable local oscillator (a “stalo”). The second input to the mixer is a CW IF signal from a coherent oscillator (a “coho”). The sum of the two frequencies is selected by the bandpass filter.

The pulse-modulated output signal from the power amplifier is fed through the duplexer to a high-gain directional antenna. The antenna then radiates the signal in the direction of the target.

The return signal from the target is received by the same antenna as used for transmit. It is routed by the duplexer to a downconversion mixer, which converts the received signal frequency to the difference between the stalo frequency and the received frequency. That difference is the coho frequency plus or minus the Doppler frequency, with the sign depending on the direction of travel of the target with respect to the radar. The signal is fed to the IF amplifier, where it is amplified to the desired level for detection and processing.

The output of the IF amplifier is fed to a phase detector, which has as its second input a CW coherent reference signal from the coho. The output of the phase detector is a train of bipolar video pulses, which are modulated in amplitude at the Doppler frequency.

The detected Doppler signal is then processed by a signal processing circuit that largely rejects clutter signals because they do not have the necessary Doppler frequency shift. Early radar systems used delay line cancellation circuits for that function. Digital processing and computer systems are used in many modern systems. The output of the signal processing system goes to the indicator systems, which may involve visual displays and data processing systems. Figure 4.20 is the block

4.6.2.1 Delay-Line Cancelers

The simple older-generation MTI delay-line canceler shown in Figure 4.21 is an example of a time-domain filter. The delay line has a delay equal to 1 divided by the pulse repetition frequency. For example, if the pulse repetition rate is 500 pulses/second, the delay must be equal to $2,000\ \mu\text{s}$ (2 ms). Delay times of this magnitude cannot be achieved with standard electromagnetic transmission lines, but they can be achieved with acoustic signals in fused-quartz delay lines. Delay lines of this type were used in the early days of radar. They have a number of disadvantages compared to more modern digital delay lines, including larger size and weight and high attenuation, which must be overcome by added gain. Loss in acoustic delay lines for radar can be 40–60 dB.

Analog acoustic delay lines were supplanted in the early 1970s by storage devices based on digital computer technology. The use of digital delay lines requires that the output of the MTI receiver phase detector be quantized into a sequence of digital words. The compactness and convenience of digital processing allow the implementation of more complex delay-line cancelers with filter characteristics not practical with analog methods. Quantization also provides greater accuracy and ability to process inferior signal-to-noise ratios.

Figure 4.22 shows the frequency response of a single and a double delay-line canceler. The delay-line canceler acts as a filter that rejects the dc component of clutter. Because of its periodic nature, the filter also rejects signals with frequencies

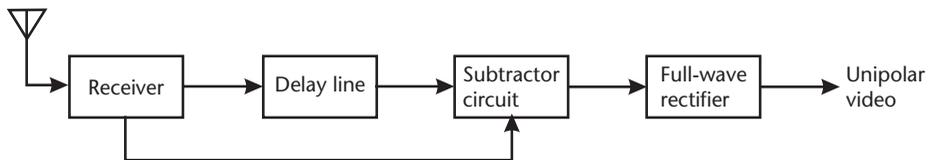


Figure 4.21 MTI receiver with delay-line canceler.

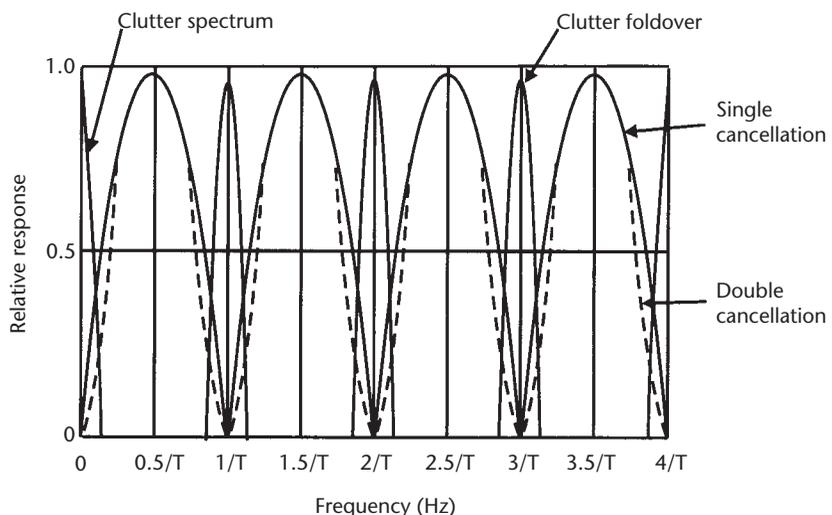


Figure 4.22 Frequency response of delay-line cancelers for single and double cancelers. (After: [6].)

in the vicinity of the pulse repetition frequency and its harmonics. The target velocities that result in zero MTI response are called blind speeds. In practice, long-range MTI radars that operate in the region of L- or S-band or higher and are primarily designed for the detection of aircraft must usually operate with ambiguous Doppler and blind speeds if they are to operate with unambiguous range. To have unambiguous range, we must have a period between pulses that is greater than the maximum delay expected for the targets of interest.

The use of a number of delay-line cancelers in series offers improvement in the amount of cancellation possible. Thus, it is common to use double or triple cancellation. The use of more than one PRF offers additional flexibility in the design of MTI Doppler filters. It not only reduces the effect of the blind speeds, it also allows a sharper low-frequency cutoff in the frequency response than might be obtained with a cascade of single-delay-line cancelers. This type of improvement is illustrated in Figure 4.23. Two delay-line cancelers are needed corresponding to the two PRFs, and they must be switched as the PRFs are switched.

4.6.2.2 Digital Signal Processing

A digital MTI processor is shown in Figure 4.24. Its operation is as follows. The output of the IF amplifier is fed to two phase detectors. The phase reference for the detectors is from a coho, with one of the reference signals shifted by 90 degrees from the other. One of the channels is the in-phase, or I , channel, while the other is the quadrature, or Q , channel. The outputs of the phase detectors are fed to sample and hold circuits followed by A/D converters. The outputs of the A/D converters then are fed to digital storage or digital delay line circuits as well as to subtraction circuits. The subtraction circuits compare the delayed signals with the undelayed signals in much the same way as is done with a delay-line canceler.

The output of each subtraction circuit is fed to a magnitude addition circuit. The method of combining signals is to add the absolute value of the I channel to the absolute value of the Q channel ($|I + Q|$). The addition of the Q channel removes the problem of reduced sensitivity due to blind phases.

The output of the magnitude addition circuit is fed to a D/A converter. The output of that circuit is fed to a data display unit.

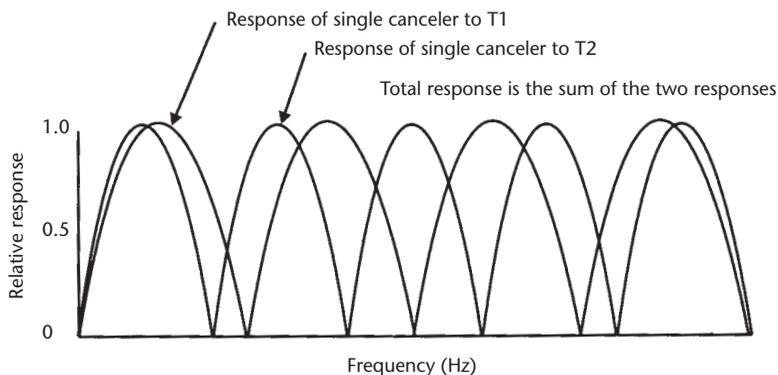


Figure 4.23 Frequency response of delay-line cancelers for $T_1/T_2 = 4/5$. (After: [6].)

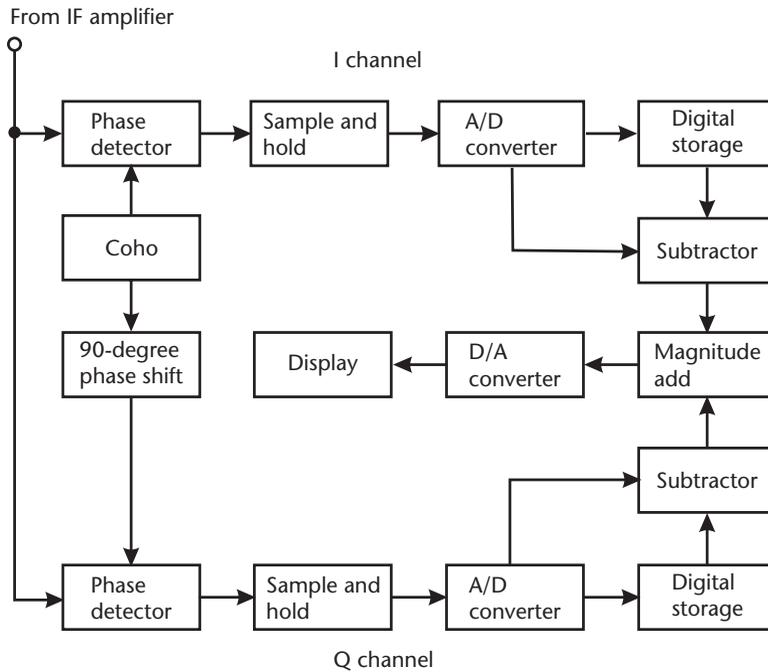


Figure 4.24 Simple digital MTI signal processor. (After: [6].)

Digital signal processing has some significant advantages over analog delay lines. As with most digital technology, it is possible to achieve greater stability, repeatability, and precision with digital processing than with analog delay-line cancelers. Thus, reliability is better. Other advantages include no special temperature control requirements, smaller size and weight, and larger dynamic range.

Digital filter banks are sometimes used with modern radars. These filters are based on the use of the fast Fourier transform (FFT). Since each filter occupies approximately $1/N$ of the bandwidth of a delay-line canceler, its signal-to-noise ratio will be greater than that of a delay-line canceler by $5 \log_{10} N$. The division of the frequency band into N independent parts by the N filters also allows a measure of the Doppler frequency to be made. If moving clutter appears at other than zero frequency, that may be rejected.

4.6.3 MTI from a Moving Platform

When the radar itself is in motion, as when mounted on a ship, an aircraft, or a police car, the detection of a moving target in the presence of clutter is more difficult than if the radar is stationary. The Doppler frequency of the clutter is no longer near zero hertz. It is determined by the speed of the radar platform. The frequency spectrum of the clutter is also widened by the platform motion.

An MTI radar on a moving platform is called AMTI (the A originally stood for “airborne”). There are two basic methods for providing the Doppler frequency compensation that is needed. In one method, the frequency of the cohο is changed to compensate for the shift in the clutter Doppler frequency. That can be accomplished by mixing the output of the cohο with a signal from a tunable oscillator having a

frequency approximately equal to the clutter Doppler frequency. The other method is to introduce a phase shift in one branch of the delay-line canceler.

4.7 Tracking Radars

Some of the information in this section is adapted from [7, 8].

4.7.1 Monopulse Tracking Radars

Figure 4.25 is a block diagram for a two-coordinate (azimuth and elevation) amplitude-comparison monopulse tracking radar (MTR). The radar uses a parabolic dish antenna with a cluster of four feed horns. The four horns provide four partially overlapping antenna beams. Four hybrid junctions are used with the four horns to provide sum and difference signals. Hybrid junction 1 provides the sum and difference signals for the left-side pair of horns, and hybrid junction 2 provides the sum and difference signals for the right-side pair of horns. Hybrid junction 3 has as its input the two sum outputs from hybrid junctions 1 and 2. Its output is the sum for all four horns. The difference output for hybrid junction 3 is the azimuth difference signal. Hybrid junction 4 has as its inputs the two difference outputs from hybrid junctions 1 and 2. The sum output from that hybrid junction is the elevation difference signal.

When the radar is transmitting, the transmitter signal is fed through the ATR unit and the TR unit, which direct the signal to the sum channel. Each antenna horn is fed with equal amplitude and equal phase. The resulting transmit antenna beam thus is centered on the axis of the antenna dish. When the radar is receiving, the transmitter is disconnected by the ATR and the TR units, and the sum channel signal

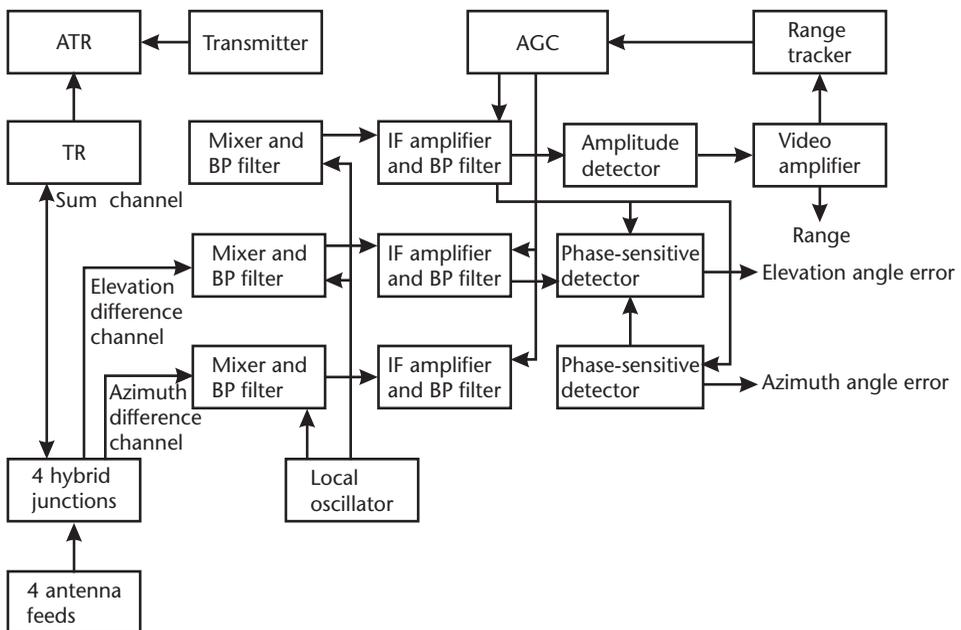


Figure 4.25 Block diagram of two-coordinate (azimuth and elevation) amplitude-comparison MTR. (After: [8].)

is fed to the radar receiver. In the simplified block diagram, Figure 4.28, the receiver consists of a mixer followed by an IF amplifier, a video amplifier, and a range tracker. In practice, the mixer would be preceded by a low-noise RF amplifier for improved sensitivity. The mixer has as its second input a CW signal from the LO. An AGC circuit is used to control the gain of the IF amplifiers.

The elevation difference channel is fed to a mixer, followed by an IF amplifier and a phase-sensitive detector. The phase-sensitive detector has as its reference input the output of the sum channel IF amplifier. The output of the detector is the elevation pointing angle error.

The azimuth difference channel is processed in a similar way as the elevation difference channel signal. It is fed to a mixer followed by an IF amplifier and a phase-sensitive detector. The output is the azimuth angle error.

The main advantage for this type of angle-tracking system is that an angle-pointing error is generated for each pulse. The system thus is much more accurate than a system that uses sequential lobbing or conical scan.

There are other monopulse tracking systems that use more than four antenna feed horns for improved accuracy. There are also phase-comparison monopulse systems that use two separate antennas. This type of system is sometimes called an interferometer radar. Angle error in this case is measured as the phase difference between the two received signals.

4.7.2 Tracking in Range Using Sequential Gating

One method for automatically tracking in range is based on the use of the split range gate. Figure 4.26(a) shows a detected echo pulse. Two range gates are used as shown in Figure 4.26(b). One is the early gate, the other the late gate. The gates are automatically positioned in time so the received-signal energy is equal in the two gates. If the gates are too late in time, the early gate will receive energy for a larger

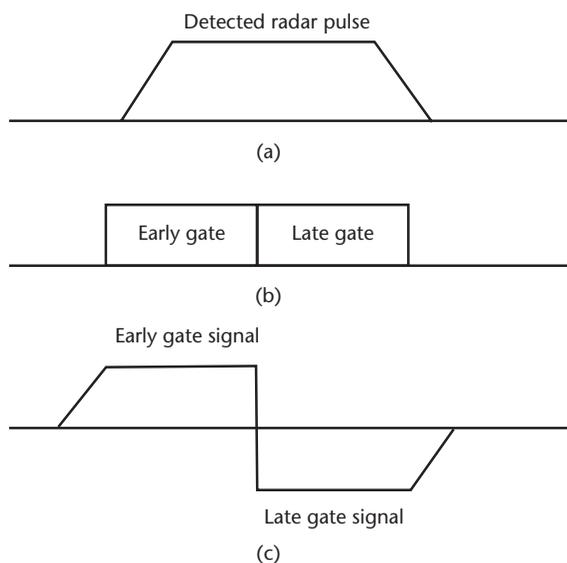


Figure 4.26 Split range gate tracking: (a) detected radar pulse; (b) early and late range gates; and (c) outputs from gates.

period of time than will the late gate. The integration of the two outputs results in a positive error signal, a condition illustrated in Figure 4.26(c). If the gates are too early in time, the early gate will see less energy than the late gate and a negative error signal will be produced.

The use of range gates for automatic tracking provides for high-accuracy range measurement. Reported accuracies are within 0.8 times the length of the target. Range gating also improves the signal-to-noise ratio since it eliminates the noise from other range intervals. The optimum gate width is equal to the pulse length. In ECM systems, range gate pull-off (RGPO) can be used to defeat these systems. RGPO jammers use TWT memory loops with delay lines to transmit pulses with increasing time delay.

References

- [1] Skolnik, M. I., *Introduction to Radar Systems*, 2nd ed., New York: McGraw-Hill, 1980, pp. 33–52.
- [2] Skolnik, M. I., *Introduction to Radar Systems*, 2nd ed., New York: McGraw-Hill, 1980, pp. 33–52.
- [3] Barton, D. K., *Modern Radar System Analysis*, Norwood, MA: Artech House, 1988, pp. 123–147.
- [4] Skolnik, M. I., *Introduction to Radar Systems*, 2nd ed., New York: McGraw-Hill, 1980, pp. 101–106.
- [5] Barton, D. K., *Modern Radar System Analysis*, Norwood, MA: Artech House, 1988, pp. 232–274.
- [6] Skolnik, M. I., *Introduction to Radar Systems*, 2nd ed., New York: McGraw-Hill, 1980, pp. 106–148.
- [7] Barton, D. K., *Modern Radar System Analysis*, Norwood, MA: Artech House, 1988, pp. 377–421.
- [8] Skolnik, M. I., *Introduction to Radar Systems*, 2nd ed., New York: McGraw-Hill, 1980, pp. 152–186.

Radio Frequency Propagation

This chapter provides information about RF propagation, including the mechanisms by which electrical signals are converted to electromagnetic waves at the transmitter, the characteristics of those generated electromagnetic waves, the way that the electromagnetic waves move from one point to another, the loss mechanisms involved in the movement, and the mechanisms by which electromagnetic waves are converted back to electrical signals at the receiver.

5.1 Antennas

5.1.1 Transmit Antennas

A transmitter antenna is a transducer that transforms electrical power delivered to the antenna into RF electromagnetic radiation. If the radiation from an antenna is uniform in all directions, it is called an isotropic antenna. Such an antenna is not possible in practice, but it is convenient to use as a reference.

The theoretical isotropic antenna is taken as the reference, and the gain of the antenna in a given direction is a measure of how the power level in that direction compares with that which would exist if the isotropic antenna had been present. That gain can be either less than or greater than 1. Expressed in decibels, it can be either positive or negative, since the logarithm of 1 is zero. For example, a typical tracking radar might use a parabolic dish antenna that produces a 1-degree azimuth by 1-degree elevation main beam. There are 41,300 square degrees in a spherical solid angle. The directional gain in that main beam then would be about 41,300, or 46 dBi, where dBi is decibels with respect to isotropic. At angles other than the main beam, there will be sidelobes. A typical power gain in the sidelobe directions might be 0 dBi or less.

The power gain of the antenna is less than the directional gain because of antenna losses and radiation through the sidelobes. In a typical case, the antenna with a 1-degree by 1-degree beam will have a power gain of about 27,000, or 44 dBi.

The typical relationship between antenna gain and beam angle is as follows:

$$G = 27,000 / (\theta_{az} \cdot \theta_{el}) \quad (5.1)$$

where:

G = numeric antenna gain

θ_{az} = the -3 -dB azimuth beam angle (degrees)

θ_{el} = the -3 -dB elevation beam angle (degrees)

Beam angles are measured at the half power (-3 dB) points on the beam. The antenna gain expressed in decibels with respect to isotropic (dBi) is $10 \log_{10} G$.

For an example of the use of (5.1), assume the case of a 2.0 -degree azimuth by 10 -degree elevation fan beam. The gain would be

$$G = 27,000 / (\theta_{az} \cdot \theta_{el})$$

$$G = 27,000 / (2 \cdot 10) = 1,350 = 31.3 \text{ dBi}$$

The effective radiated power (ERP) is defined as the product of the antenna gain and the radiated power. For example, if the numeric antenna gain is 10 and the radiated power is 200W ,

$$\text{ERP} = 200 \cdot 10 = 2,000\text{W}, \text{ or } 33 \text{ dBW}, \text{ or } 63 \text{ dBm}$$

As discussed in Chapter 2, dBW is decibels with respect to 1W , while dBm is decibels with respect to 1 mW .

5.1.2 Receive Antenna Gain and Capture Area

The antenna gain of an antenna when it is used as a receiver antenna is the same as the antenna gain when it is used as a transmitter antenna, that is, an antenna is bilateral. The effective capture area of the receiver antenna is the antenna gain times the area of an ideal isotropic antenna for the frequency of interest. The effective capture area of a receiver antenna is given by (5.2).

$$A_e = G_r \lambda^2 / 4\pi \quad (5.2)$$

where:

A_e = effective capture area of an antenna (square meters)

λ = wavelength (meters)

G_r = receive directional antenna gain

For an example of the use of (5.2), assume an antenna with a gain of 13 dBi operating at a wavelength of 2m . The effective capture area of the antenna would be

$$A_e = G_r \lambda^2 / 4\pi$$

$$A_e = 20 \times 4 / 4\pi = 6.37 \text{ m}^2$$

In the case of aperture-type antennas, such as horn antennas and parabolic dish reflector antennas, the effective capture area of the antenna is typically about half the actual aperture area. The reason is that the aperture is not illuminated uniformly by the feed system. For example, if we have a parabolic dish antenna with a diameter of 60 feet (18.3m), the aperture area is about 263 m^2 , and the effective capture area is about half that, or 132 m^2 .

5.2 Electromagnetic Waves, Fields, and Power Density

Figure 5.1 illustrates an electromagnetic wave traveling in free space. The wave is transverse electric and transverse magnetic, that is, a transverse electromagnetic (TEM) wave in which the E-field is at right angles to the magnetic field, and both are at right angles to the direction of propagation. Also, there is no electric or magnetic field in the direction of propagation, only transverse to the direction of propagation, hence the term transverse electromagnetic. Figure 5.1 shows a front view and a side view of the wave.

The essential properties of a radiated RF electromagnetic wave are wavelength, frequency, intensity, direction of travel, and direction of polarization. The wavelength is the distance the wave travels in one complete RF cycle period. The frequency is the number of cycles in a second. The intensity of the wave is the field strength given in volts per meter or the power density of the wave in watts per square meter. Performance calculations usually deal with the power density.

The direction of polarization is the direction of the free-space electrical (E) field. If the direction of polarization is horizontal with respect to the surface of the Earth, the wave is said to have horizontal polarization. If the free-space E-field is vertical with respect to the surface of the earth, the wave is said to have vertical polarization. These polarizations are considered linear.

It is also possible to have elliptical or circular polarization. Such waves have a rotating E-vector. They may be assumed to be composed of two linearly polarized waves that have orthogonal polarizations, that is, their E-field vectors differ by 90 degrees. In those cases, the polarization vector rotates in either the left or the right direction as the wave propagates, depending on whether the phase shift between the two linear polarization components is +90 degrees or -90 degrees.

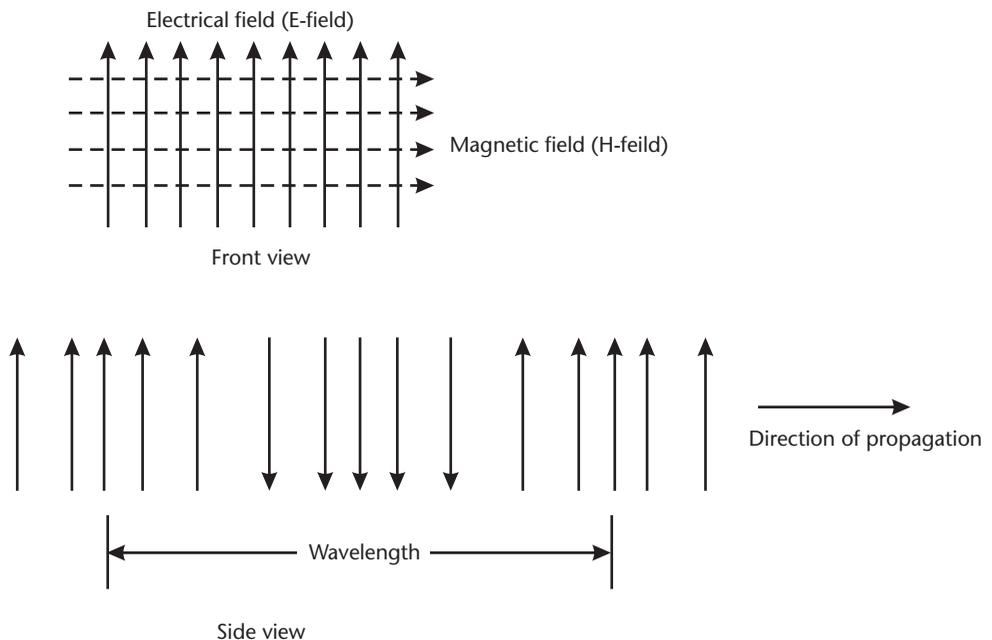


Figure 5.1 Illustration of a TEM wave.

The power density of a TEM wave at a point in space can be expressed in terms of its E-field and the characteristic impedance of free space as follows:

$$P_1 = E^2 / Z_0 \quad (5.3)$$

where:

P_1 = power density (watts per square meter)

E = E-field strength (volts per meter)

Z_0 = characteristic impedance of free space = 377Ω

The magnetic field strength, H , has units of amperes per meter. Alternatively, the power density in watts per square meter equals $377 H^2$.

For free-space propagation, the electric and magnetic field intensities decrease directly with range from the transmitter. Hence, the power density decreases as the square of the range. Thus, if point B is four times the distance from the transmitter as point A, the E-field intensity at point B will be one-fourth the E-field intensity at point A, and the power density at point B will be one-sixteenth the power density at point A.

The equation for power density in terms of range for free-space propagation is as follows:

$$P_1 = P_T G_T / 4\pi R^2 \quad (5.4)$$

where:

P_1 = power density at a distance R

P_T = transmitter power

G_T = transmitter antenna gain

R = range

As an example of E-field strength and power density for free-space propagation, assume a transmitter power output of 100W and an antenna gain of 10. The power density at a range of 100 km (10^5 m) thus would be

$$\begin{aligned} P_1 &= P_T G_T / 4\pi R^2 \\ P_1 &= 100 \times 10 / (4\pi \times 10^{10}) = 7.96 \times 10^{-9} \text{ W/m}^2 \end{aligned}$$

The RMS E-field strength would be

$$\begin{aligned} E &= (Z_0 P_1)^{0.5} \\ E &= (377 \times 8 \times 10^{-9})^{0.5} = 1.73 \times 10^{-3} \text{ V/m} \end{aligned} \quad (5.5)$$

5.3 RF Electrical Field Waveforms and Vector Addition

It is common to have two or more E-fields added. Examples are the addition of fields from the individual elements of a phased array antenna and the addition (or

subtraction) of the direct and reflected waves in the case of multipath propagation. Because E-fields have both magnitude and phase, they are vector quantities and must be added or subtracted by standard vector addition and subtraction techniques.

5.4 Free-Space Path Loss

The term free-space path loss refers to the spreading loss for a radiated signal between a transmitter antenna and a receiver antenna with gain equal to 1 (0 dBi) for both transmit and receiver antennas. It is given by

$$L_p = (4\pi)^2 R^2 / \lambda^2 \quad (5.6)$$

where:

L_p = Free-space path loss (numeric)

R = range (meters)

λ = wavelength (meters)

Free-space path loss increases as the frequency is increased because of the smaller capture area of the antenna. It also increases rapidly with increased range because of the range squared proportionality. For a fixed frequency, the one-way free-space path loss, as experienced by communication systems, increases 6 dB per octave (range doubling), or 20 dB per decade. The two-way free-space path loss, as experienced by radar systems, increases 12 dB per range doubling, or 40 dB per decade, since round-trip loss is proportional to R^4 . Twelve dB is calculated from $10 \log_{10}(2^4)$.

5.5 Excess Path Loss and Atmospheric Attenuation

The radiated signal sees a number of other losses in addition to free-space path loss. The sum of those losses is called the excess path loss. Examples of such losses are atmospheric attenuation, diffraction loss, multipath loss, ground-wave loss, ionosphere refraction loss, and scatter propagation loss. Excess path losses are discussed in the following paragraphs. Some of the information presented is adapted from [1].

5.5.1 Atmospheric Absorption

As electromagnetic waves pass through the atmosphere, they are attenuated by atmospheric absorption. At the lower frequencies, the additional attenuation is very small and propagation losses are nearly the same as for free-space propagation. At X-band and higher frequencies, atmospheric absorption loss becomes important and can become large. That is especially true at the resonance frequencies for water vapor and oxygen. The first resonance frequency for water vapor is about 22 GHz, while the first resonance frequency for oxygen is approximately 60 GHz. There are other resonance frequencies above 100 GHz for both water vapor and oxygen.

Figure 5.2 shows a plot of two-way (radar case) atmospheric attenuation coefficients versus frequency for clear atmospheric conditions, which includes oxygen plus water vapor, at sea level. Table 5.1 lists the attenuation coefficients in the so-called atmospheric windows and at the major peaks of Figure 5.2. It also lists attenuation coefficients at the center frequencies of the standard radar bands [1].

In the case of a communication system, a one-way path is involved. The atmospheric attenuation, therefore, is one-half the two-way value shown in the figure when expressed in decibels. For example, if the two-way atmospheric attenuation is computed as 26 dB for a given path, the one-way atmospheric attenuation would be 13 dB.

The attenuation coefficients given in Figure 5.2 are for sea-level atmospheric conditions. At higher elevations, the loss is less because of lower oxygen and water vapor densities. For example, at an altitude of 30,000 feet, the two-way atmospheric attenuation coefficient at 10 GHz is reduced to about 0.003 dB/km. That compares with 0.024 dB/km at sea level. A ground-based, X-band radar that is looking up at a 10-degree elevation angle and a range of 100 km will see an atmospheric loss of only about 0.9 dB, compared to the 2.4-dB loss at sea level for a low elevation angle.

5.5.2 Attenuation Produced by Rain, Snow, and Fog

Figure 5.3 shows a plot of attenuation coefficients for rain as a function of frequency [2]. The attenuation coefficients for one-way communication will be one-half the two-way value when expressed in decibels.

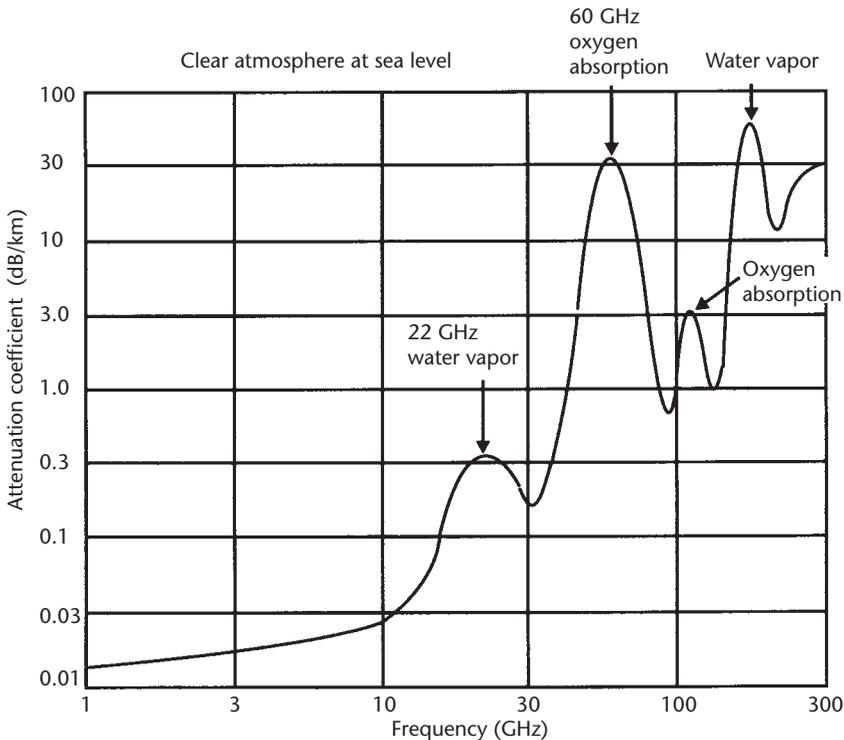
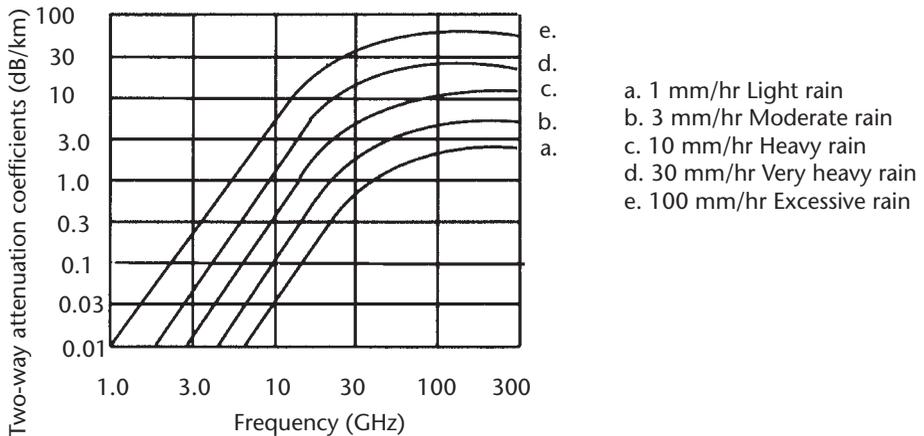


Figure 5.2 Two-way atmospheric attenuation coefficients versus frequency. (After: [2].)

Table 5.1 Attenuation Coefficients

Radar Band	Approximate Center Frequency or Major Peaks (GHz)	Attenuation Coefficient (dB/km)
L	1.3	0.012
S	3.0	0.015
C	5.5	0.017
X	10	0.024
K _u	15	0.055
K	22	0.3
K _a	35	0.14
V	60	35.0
W	95	0.8
mm	140	1.0
mm	240	15

**Figure 5.3** Two-way rain attenuation coefficients versus frequency and rate. (After: [2].)

At L-band (1.3 GHz typical), only the most intense rain can cause a doubling of the clear-atmosphere attenuation coefficient. Such rainfall cannot exist over any substantial path length. At S-band (3 GHz typical), the clear-atmosphere value is doubled by rain at 10 mm/hr. Higher frequencies are much more affected by rain.

The attenuation produced by ice particles in the atmosphere, whether occurring as hail, snow, sleet, or ice-crystal clouds, is much less than that caused by rain of an equivalent rate of precipitation. Fog also may produce attenuation at the higher frequencies. Attenuation from heavy fog is comparable to that of moderate rain.

5.6 Atmospheric Refraction

Another important effect of the atmosphere on electromagnetic wave propagation is refraction. As pointed out earlier, the velocity of the electromagnetic wave is inversely proportional to the square root of the dielectric constant of the medium.

Because the atmosphere is not of constant density as altitude varies and because the composition changes as a function of altitude, the propagation velocity also differs with altitude. A transmitted wave does not propagate in a straight line but follows a curved path. One consequence of that bending is that greater range is possible for communication and radar systems without the inherent horizon limitations.

If the bending of the ray path for the waves is just right, the bending just matches the curvature of the Earth, and there are no radio horizon limitations. Normally, however, the rate of bending is not that great. Even so, the horizon limit is extended significantly beyond that of straight-line propagation.

With the linear constant-gradient model and the so-called SBF method, the range capability is determined by plotting the ray path as a straight line and assuming that the radius of the Earth is increased. With the so-called standard atmosphere, the Earth's radius is assumed to be four-thirds times the true radius. It should be emphasized that this "four-thirds Earth" rule is only an approximation.

For an example of radio horizon extension using the four-thirds earth rule, assume an aircraft flying at 40,000 feet. With no refraction and 0-degree elevation angle at the Earth's surface, the radio horizon would be at a range of 245 miles. That is based on an Earth radius of 3,960 miles, or a circumference of 24,881 miles. With the "four-thirds Earth" model, the Earth "radius" is 5,283 miles. The radio horizon in that case would be 283 miles.

Under special cases, it is possible to produce ducting of the electromagnetic wave. At some elevated region above the Earth, the dielectric gradient becomes so large that the wave is refracted back to the Earth. There it is reflected, and the wave travels back to the strong refracting layer, where it is again refracted down. By that process, the wave may travel great distances without radio horizon limitations. Ducting is most common over water, but it may also appear over land areas.

5.7 Diffraction of Radio Waves

Figure 5.4 illustrates diffraction of radio waves. The term diffraction refers to the process by which an electromagnetic wave is bent in its path by a material edge. An example is a VHF or UHF communication system in which there are buildings that block the path of the electromagnetic wave. A fraction of the energy in the wave is bent around or over the buildings and may be received by a receiver that does not have optical line of sight back to the transmitter. There is loss in the diffraction process, but the signal may nevertheless be strong enough for good reception.

Figure 5.5 shows the approximate one-way attenuation for knife-edge diffraction [3]. The two-way attenuation, as would be experienced by radar, would be twice the one-way value expressed in decibels. The propagation factor or added attenuation for knife-edge diffraction depends on the vertical distance, h , between the top of the obstacle and the direct path between the radar and the target, and on the wavelength. In Figure 5.5, it is shown as a function of the diffraction parameter, v , which is calculated as follows:

$$v = \left[\frac{(2h\theta)}{\lambda} \right]^{0.5} \quad (5.7)$$

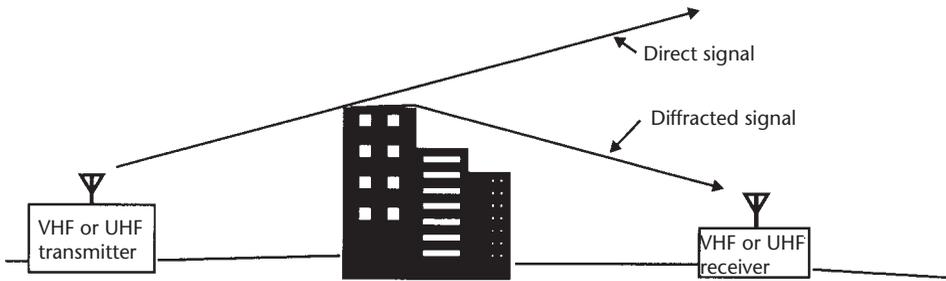


Figure 5.4 Example of diffraction with a communication system.

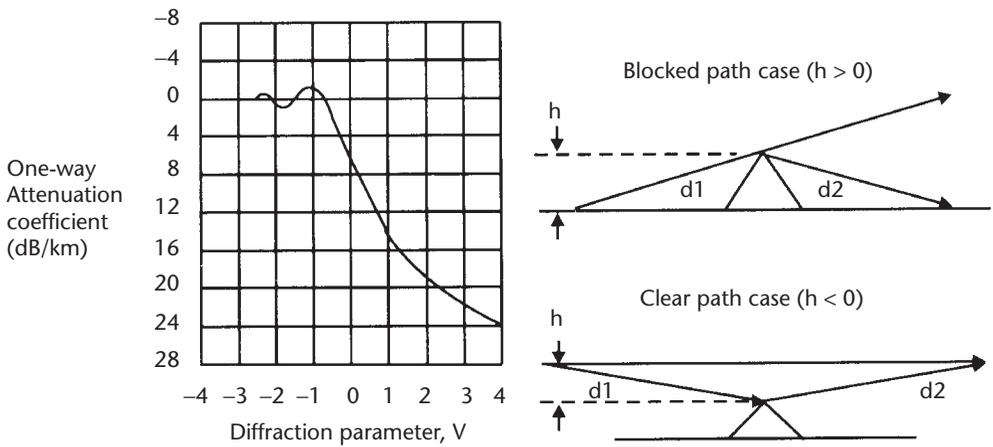


Figure 5.5 Attenuation for knife-edge diffraction. (After: [3].)

where θ = angle and λ = wavelength.

For h/d_1 and h/d_2 , each small the approximate value of v is given by (5.8).

$$v = h \left[(2/\lambda) (1/d_1 + 1/d_2) \right]^{0.5} \tag{5.8}$$

Positive v implies a blocked path, and negative v implies that the line-of-sight path clears the obstacle.

For an example of the use Figure 5.5 in predicting attenuation due to diffraction, assume the case of a communication system operating at a frequency of 1 GHz. Other assumptions are as follows:

$$h = 100\text{m}$$

$$d_1 = 20,000\text{m}$$

$$d_2 = 40,000\text{m}$$

$$\lambda = 0.3\text{m (1 GHz)}$$

Using (5.8),

$$v = h \left[(2/\lambda) (1/d_1 + 1/d_2) \right]^{0.5}$$

$$v = 100 \left[(2/0.3) (1/20,000 + 1/40,000) \right]^{0.5} = 2.23$$

From Figure 5.5 with $v = 2.23$, the one-way attenuation due to refraction is 20 dB. If the frequency had been 100 MHz instead of 1 GHz, the value of v would be 0.71. The attenuation in that case would be about 12 dB. Thus, we see that the diffraction loss is much lower at the lower frequencies than at the higher frequencies.

5.8 Multipath

Another important property of electromagnetic waves is that they reflect from the Earth's surface and other objects as well as travel directly to an antenna or target. The reflected signal adds to the direct signal, producing either a larger signal than the direct signal at the receiver antenna (constructive interference) or a smaller signal (destructive interference), depending on the relative phase angles of the two signals. That vector addition of a direct signal and a reflected signal is referred to as multipath. Multipath usually is considered undesirable because of the substantial reduction of signal in the nulls. In a few cases, it can be considered desirable because of the added signal in the peaks.

Figure 5.6 illustrates the concepts of multipath peaks extending beyond the free-space pattern and nulls extending below that pattern. In the limit where the direct and reflected signals are exactly the same amplitude, the peaks could extend 12 dB beyond the free-space pattern for radar and 6 dB for communication. The nulls in this case would extend to zero signal. In the typical case, the signals are not equal, peaks are not so great, and null is not so deep.

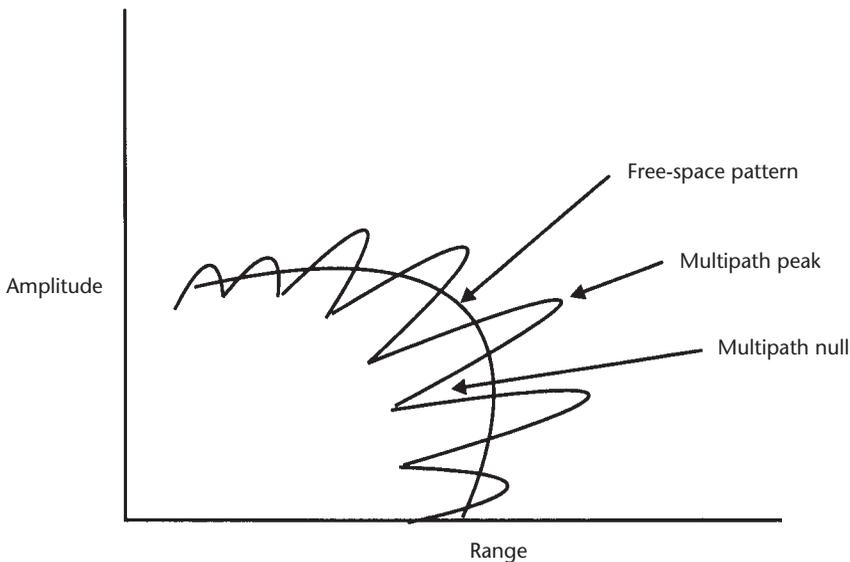


Figure 5.6 Coverage diagram with multipath lobbing showing constructive and destructive interference as a function of elevation angle. (After: [4].)

In the case of communication systems in which large elevation angles are used, there may be lobing at high elevation angles as well as at low angles. With radar and highly directional communication systems, multipath is important only at the lower elevation angles where the Earth is included in the main beam of the antenna. The surface reflection coefficient is the product of three factors: the Fresnel reflection coefficient, the specular scattering coefficient for a rough surface, and the coefficient for vegetative absorption. On reflection, the phase shift for horizontal polarization is 180 degrees, independent of grazing angles. For vertical polarization, the phase shift on reflection is 180 degrees for grazing angles less than the Brewster angle and near zero degrees for angles greater than the Brewster angle. Thus, for near-zero elevation angle and multipath, we should expect a null. How deep the null is depends on the magnitude of the reflection coefficient.

The Brewster angle is the angle that corresponds to the minimum reflection coefficient. That angle depends on the frequency and the type of reflecting surface. For example, for seawater the Brewster angle at 3 GHz is about 6 degrees. It is only about 1 degree for seawater and 100 MHz.

The scattering coefficient for a rough surface is much less than unity, with the exact value depending on the grazing angle and the root mean square (rms) deviation from a smooth surface. The same is true for a surface covered with vegetation. A layer of vegetation covering the surface absorbs much of the incident radiation and scatters the remainder in an irregular pattern [4].

Multipath can be very important in determining the propagation characteristics for both communication and radar systems. That is equally true over land and over water. It is most important when a surface is flat and free from vegetation.

5.9 Ionospheric Propagation

Some of the material presented in this and subsequent sections is quoted or adapted from [5, 6].

Important terms in talking about ionospheric propagation or HF communication are critical frequency, maximum usable frequency (MUF), and lowest usable frequency (LUF). The critical frequency is that frequency for a given layer that reflects a vertically incident wave. Frequencies higher than the critical frequency pass through the layer at vertical incidence. The MUF is the critical frequency times the secant of the angle of incidence at the reflecting layer times a correction factor for earth curvature. The MUF varies for each layer with local time of day, season, latitude, and throughout the 11-year sunspot cycle. In addition, ionization is subject to frequent abnormal variations, such as those caused by solar flares.

Ionospheric losses are a minimum near the MUF and increase rapidly for lower frequencies during daylight. An optimum working frequency is selected below the MUF to provide some margin for variations. The LUF is that frequency below which ionospheric absorption and radio noise levels make required radiated power impractical. At altitudes between about 50 km and 400 km, the atmosphere may be ionized by energy from the Sun. The so-called D-layer extends from altitudes of about 50 to 90 km and exists only during daylight hours. The D-layer reflects VLF and LF waves,

absorbs MF waves, and weakens HF waves through partial absorption. VHF and higher frequency waves pass through the D-layer with very little attenuation.

The E-layer exists at a height of about 110 km and may be present both day and night. It provides reflections for both MF and HF energy and small attenuation for higher frequencies.

The F_1 -layer exists at a height of about 175 to 250 km. It exists only during the daylight hours. It provides reflection of HF waves and low attenuation of higher frequency waves.

The F_2 -layer is at heights of about 250 to 400 km. This layer is the principal reflecting region for long-distance HF communication. It is present both day and night. Above about 100 MHz, the attenuation caused by the F_2 and the other ionized layers is less than about 1 dB and may be neglected.

It is possible to achieve very long-range communication using the ionosphere as a reflecting medium because the ionosphere is so high. The signal travels from the ground-based transmitter antenna to the ionospheric layer where the path is bent downward. Although it is really refraction rather than reflection that is involved, it is common to speak of the process as reflection. The downward directed wave then travels to the ground, where it is received. The process constitutes single-hop propagation. For two-hop propagation, the wave is reflected from the ground back to the ionosphere, where it is again refracted downward to the ground, where it is received. Three-hop (or more) propagation also is possible. By these means, it often is possible to communicate many thousands of miles.

Figure 5.7 illustrates HF ionospheric reflection propagation between Washington, D.C., and Chicago and between Washington, D.C., and San Francisco. In the case of communication between Washington and Chicago, single-hop paths are possible using either E-layer reflection or F_2 -layer reflection. In the case of communication between Washington and San Francisco, a two-hop path is possible using F_2 -layer reflection.

In evaluating the performance of HF communication systems, three main losses must be taken into account: the free-space path loss, the loss in the ionosphere, and

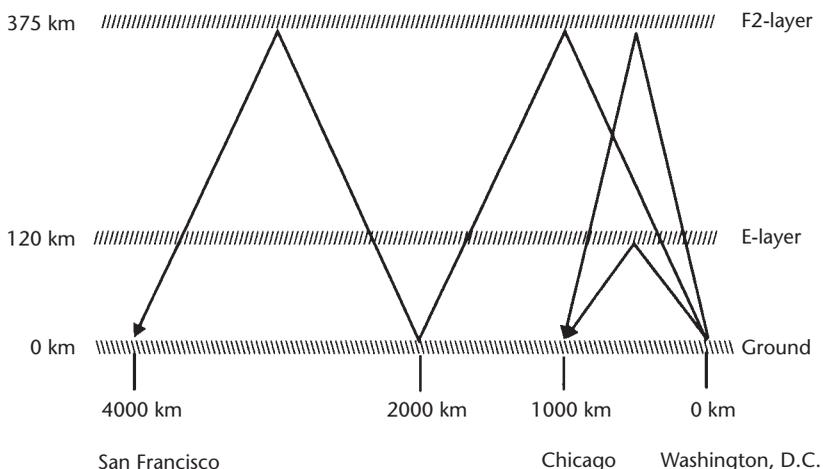


Figure 5.7 Example reflection paths for ionospheric propagation.

the loss through reflection on land or sea. With more than one hop, there will be losses each time the signal passes through the ionosphere and each time reflection takes place. Typical losses in the ionosphere range from as low as 2 dB per pass to as high as 30 dB, depending on the time of day and the frequency used. Losses due to reflection are typically less than 4 dB per reflection, depending on the frequency and the land type.

HF communication using the ionosphere as a means of reflection has only limited bandwidth capability. It has high noise levels and large losses. It has problems of multipath-induced fading unless operation is near the MUF. HF communication is not reliable over paths near the magnetic poles and has been largely replaced by satellite communication and other means for long-distance communication [5, 6].

5.10 Ground-Wave Propagation

Another type of propagation that is important at the lower frequencies is ground-wave propagation [6]. The ground wave glides over the surface of the Earth and is vertically polarized. Any horizontal component of the E-field in contact with the earth is short-circuited by the Earth and therefore is quickly attenuated. The ground wave induces charges in the Earth that travel with the wave and so constitute a current. As the ground wave passes over the surface of the Earth, it is weakened as a result of energy absorbed by the Earth. Losses are the lowest at the lower frequencies and increase with frequency. Ground-wave propagation is normally not used above HF frequencies because the attenuation is too great. Ground-wave losses are much lower over seawater than over land [6]. For example, with good ground, a frequency of 150 kHz, and a range of 1,000 miles, the ground-wave loss is about 30.5 dB greater than free-space loss. In the case of seawater and that same range and frequency, the ground-wave loss is about 22.5 dB greater than free-space loss. That is a difference of 8.0 dB. As another example, with good ground, a frequency of 500 kHz, and a range of 200 miles, the ground-wave loss is about 20 dB greater than free-space loss. In the case of seawater and the same range and frequency, the ground-wave loss is about 6 dB greater than free-space loss. That is a difference of 14 dB.

Ground-wave VLF and LF frequencies are used for high survivability, high value, military command links, and navigation systems. Such systems have range capabilities on the order of 1,000 to 2,500 miles. Ground waves also are used for standard AM broadcasting (535 to 1,605 kHz), where the range capability is approximately 60 to 100 miles. Ground-wave propagation also is used by the military at low HF frequencies where the range capability is 50 to 80 miles. Most ground-wave communication systems involve very large wavelengths. For example, at 30 kHz, the free-space wavelength is 10,000m. At 300 kHz, the free-space wavelength is 1,000m. At 3 MHz the free-space wavelength is 100m. At the higher frequencies, it is typical for a vertically polarized monopole antenna to be on the order of a quarter-wave high for good efficiency. At that height, the antenna would be resonant. That, however, would be impractical at 30 kHz, where the antenna height would be 2,500m. The same is true at 300 kHz, where the required

antenna height would be 250m. For those long wavelength systems, it is necessary to use electrically small antennas for transmitter antennas rather than quarter-wave monopoles. Such antennas require tuning and often use capacitive loading. With such tuning, the antennas can be made as small as 0.1 wavelength high or less. However, tuned antennas typically have low efficiency because of losses in the tuning systems.

In the case of receiver antennas for long wavelength signals, it is not necessary to have the best efficiency. That is due to the fact that the limiting noise for these systems is external to the receiver and is received by the antenna. Making a better receiver antenna increases the received signal, but it also increases the noise to the same degree. The signal-to-noise ratio thus is not improved.

Again, it must be emphasized that ground-wave propagation requires vertical polarization. Therefore, the antennas used must provide vertical polarization. Details about antennas for all bands and modes of propagation are discussed in Chapter 13.

5.11 Scatter Propagation

Three modes of scatter propagation have been used for communication: tropospheric scatter, ionospheric scatter, and meteor-burst scatter. Weak but reliable fields are propagated several hundred miles beyond the horizon in the VHF, UHF, and SHF bands by those scatter modes. High-gain directional antennas and high power are used for tropospheric scatter and ionospheric scatter because of the weakness of the signals.

In the case of tropospheric scatter, the signal is reflected by regions high above the Earth, where changes in the dielectric constant of the atmosphere take place. In a similar way, reflections can take place at those higher frequencies from the ionosphere where there are changes in the ionization density. Meteor burst scattering takes place when there is a meteor that enters the atmosphere, creating an ionized trail. Signals may be scattered from that trail as long as it exists. This type of system normally operates at VHF frequencies.

Figure 5.8 illustrates the concept of propagation via tropospheric scatter [5]. On the left of the diagram is a transmitter station with a high-power transmitter feeding a high-gain parabolic dish antenna directed at the sky. On the right is a receiver station with a high-gain parabolic dish receiver antenna. The two antenna beams are directed so there is a common volume high in the atmosphere for the two antenna main beams.

Although tropospheric scatter is subject to fading with little signal scattered forward, it nevertheless forms a reliable method of over-the-horizon communication. It is not affected by the abnormal phenomena that afflict HF ionospheric propagation. Path lengths typically are 300 to 500 km. Operating frequencies typically are centered around 900, 2,000, and 5,000 MHz. Losses include free-space path loss plus scatter loss in the troposphere. Scatter loss may be as high as 60 to 90 dB. High transmitting power, extremely good receiver sensitivity, and high antenna gain obviously are needed [5].

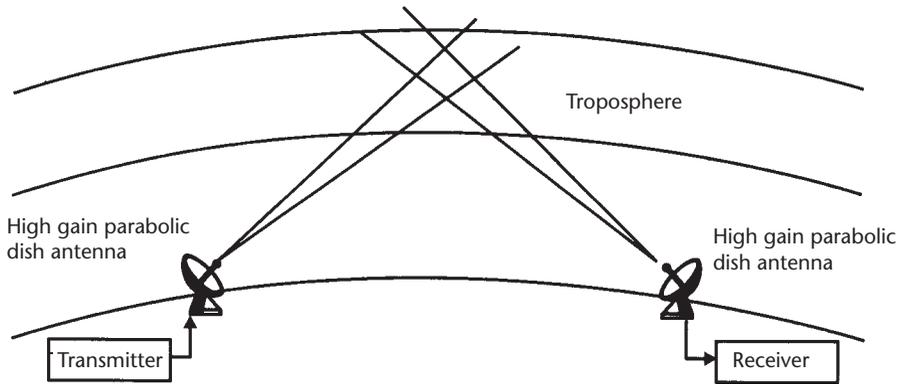


Figure 5.8 Tropospheric scatter propagation.

5.12 Fiber Optic Cable Propagation

In recent years, fiber optic communication has become an important mode of communication. Fiber optic cables have many advantages over other types of cable, including much larger bandwidth capability, smaller size and weight, reduced loss per unit length, and lower cost.

Many applications are now being found for fiber optic cables in addition to long-range communication links. Examples are fiber optic delay lines for radars and other RF systems, fiber optic control lines, and fiber optic lines used in the medical field for viewing inside the human body.

An optical fiber is a piece of highly pure glass of very small diameter. The fiber has an outside cladding of glass similar to the fiber itself but, because of a slightly different chemical composition, with a different refractive index. This type of optical fiber is known as a step-index fiber. Other types of optical fibers are possible, but they have higher attenuation than the step-index fiber.

Figure 5.9 shows attenuation in typical modern fiber optic cables as a function of wavelength. The majority of early commercial fiber optic links operated at a wavelength of about 0.85 μm . Attenuation at that wavelength is about 2.5 dB/km. Most of the fiber optic links now being installed operate in a second window, at 1.3- μm wavelength. Attenuation at that wavelength is about 0.4 dB/km. It is likely that future fiber optic systems will operate in the third window at a wavelength of about 1.6 μm to permit attenuations as low as 0.25 dB/km [7]. That lower attenuation permits reduced numbers of repeaters in long-distance links, higher reliability due to reduction in parts, and reduction in system cost.

5.13 Radar Cross-Section of Targets

The radar cross-section (RCS) of a real target is the cross-section that the target would have if the signal were scattered uniformly in all directions. A physically small target can have a very large RCS if the signal reaching the target is scattered back largely in the direction of the radar rather than uniformly in all directions. A

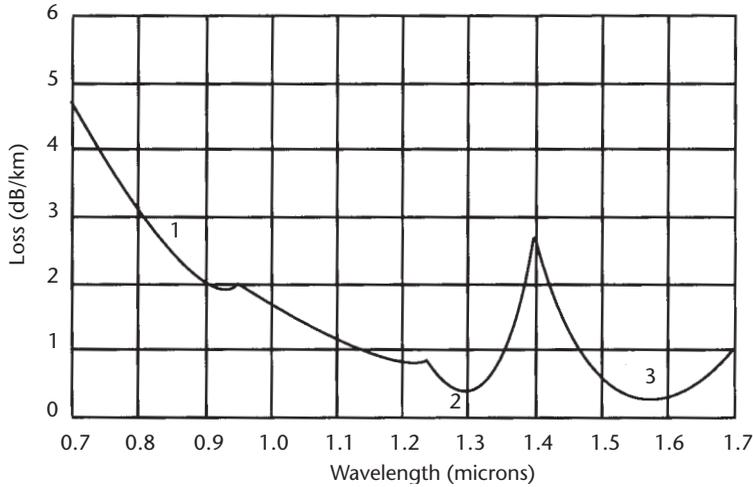


Figure 5.9 Attenuation in low-loss optic fibers. (Source: [7].)

corner reflector is an example of a target of this kind. On the other hand, a very large target can have a very small RCS if very little of the scattered signal is in the direction of the radar.

Figure 5.10 illustrates the concepts of propagation for radars using backscatter from targets. The figure shows that the signal experiences two-way spreading loss, two-way excess path loss, and reflection by an ideal spherical metallic target having the same back-scatter level as the real target but with uniform radiation in all directions. The radar is shown with separate transmitter and receiver antennas rather than a single antenna, as is normally used simply for ease in showing the loss functions. The radar transmitter has a transmitter antenna gain, G_T , which is normally quite high and which provides a fan or pencil beam, depending on the application. The receiver antenna has an effective capture area, A_e , that is equal to $G_R \lambda^2 / 4\pi$, where G_R is the receiver antenna gain and λ is the wavelength.

The radar may also see back-scatter signals from the ground or the sea and from weather such as rain or clouds. A more detailed discussion of RCS of targets and clutter for radar is presented in Chapter 4.

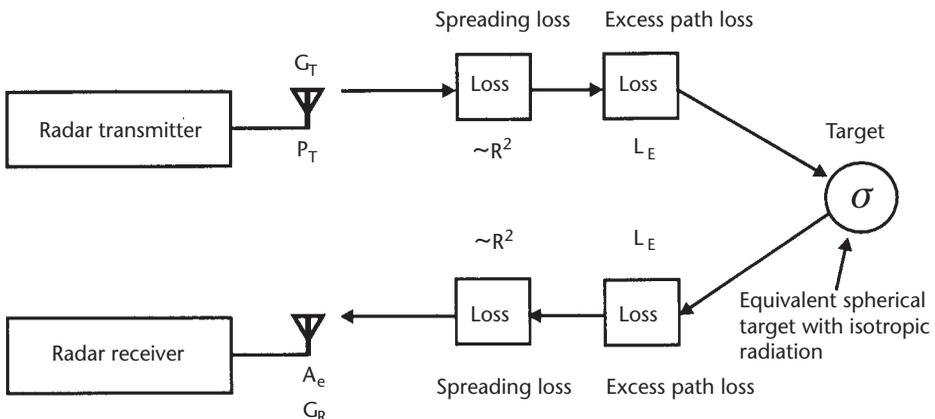


Figure 5.10 Concepts of propagation for radar with an equivalent ideal spherical target with RCS σ .

5.14 Equations for Calculating Propagation Performance for Communication Systems

The received power at the input to the receiver is given by (5.9).

$$P_r = P_T G_T G_R \lambda^2 / \left[(4\pi)^2 R^2 L_E L_{CT} L_{CR} \right] \quad (5.9)$$

where:

P_T = transmitter output power

G_T = transmitter antenna gain

G_R = receiver antenna gain

λ = wavelength

R = range (same units as wavelength)

L_E = excess path loss, numeric

L_{CT} = cable loss between transmitter and antenna

L_{CR} = cable loss between receiver and antenna

The following examples use (5.9) to determine the performance of communication systems.

5.14.1 Example 1: HF Ionospheric Reflection Communication System

Assumptions:

P_T = transmitter output power = 1,000W

G_T = transmitter antenna gain = 2 dBi

G_R = receiver antenna gain = 2 dBi

λ = wavelength = 30m

R = range = 2,000 km = 2×10^6 m

L_E = excess path loss = 15 dB

L_{CT} = cable loss between transmitter and antenna = 2 dB

L_{CR} = cable loss between receiver and antenna = 2 dB

Expressed in decibels, (5.9) is

$$P_r = P_T + G_T + G_R + 20 \log \lambda - 20 \log(4\pi) - 20 \log(R) - L_E - L_{CT} - L_{CR} \quad (5.10)$$

Substituting values:

$$P_r = 30 + 2 + 2 + 29.5 - 22 - 126 - 15 - 2 - 2$$

$$P_r = -103.5 \text{ dBW} = -73.5 \text{ dBm} = 4.47 \times 10^{-11} \text{ W}$$

5.14.2 Example 2: VHF Base Station to Mobile Unit Communication System

Assumptions:

$$P_T = \text{transmitter output power} = 100\text{W}$$

$$G_T = \text{transmitter antenna gain} = 3 \text{ dBi}$$

$$G_R = \text{receiver antenna gain} = 0 \text{ dBi}$$

$$\lambda = \text{wavelength} = 2\text{m}$$

$$R = \text{range} = 20 \text{ km}$$

$$L_E = \text{excess path loss} = 20 \text{ dB}$$

$$L_{CT} = \text{cable loss between transmitter and antenna} = 2 \text{ dB}$$

$$L_{CR} = \text{cable loss between receiver and antenna} = 1 \text{ dB}$$

Substituting values into (5.10):

$$P_r = 20 + 3 + 0 + 6 - 22 - 20 - 2 - 1$$

$$P_r = -102 \text{ dBW} = -72 \text{ dBm} = 6.3 \times 10^{-11} \text{ W}$$

5.14.3 Example 3: Microwave Uplink to Satellite Relay Located at Geostationary Orbit

Assumptions:

$$P_T = \text{transmitter output power} = 10,000\text{W}$$

$$G_T = \text{transmitter antenna gain} = 44 \text{ dBi}$$

$$G_R = \text{receiver antenna gain} = 24 \text{ dBi}$$

$$\lambda = \text{wavelength} = 0.05\text{m}$$

$$R = \text{range} = 40,000 \text{ km}$$

$$L_E = \text{excess path loss} = 1 \text{ dB}$$

$$L_{CT} = \text{cable loss between transmitter and antenna} = 2 \text{ dB}$$

$$L_{CR} = \text{cable loss between receiver and antenna} = 1 \text{ dB}$$

Substituting values into (5.10):

$$P_r = 40 + 44 + 24 - 26 - 22 - 152 - 1 - 2 - 1$$

$$P_r = -96 \text{ dBW} = -66 \text{ dBm} = 2.5 \times 10^{-10} \text{ W}$$

5.15 Equations for Calculating Propagation Performance for Radar Systems

The received power at the input to the radar receiver is given by (5.11).

$$P_r = P_T G_T G_R \lambda^2 \sigma / \left[(4\pi)^3 R^4 L_{E1} L_{E2} L_{GT} L_{GR} \right] \quad (5.11)$$

where:

P_T = transmitter output power

G_T = transmitter antenna gain

G_R = receiver antenna gain

λ = wavelength

σ = target RCS, same units as range and wavelength (i.e., meters squared)

R = range

L_{E1} = excess path loss for transmitted signal

L_{E2} = excess path loss for reflected signal

L_{GT} = waveguide or cable loss between transmitter and antenna

L_{GR} = waveguide or cable loss between receiver and antenna

The following examples use (5.11) to determine the performance of radar systems.

5.15.1 Example 4: L-Band Aircraft Surveillance Radar

Assumptions:

P_T = transmitter output power = 100 kW

G_T = transmitter antenna gain = 33 dBi

G_R = receiver antenna gain = 33 dBi

λ = wavelength = 0.23m

σ = target

RCS = 10 m²

R = range = 100 km

L_{E1} = excess path loss for transmitted signal = 2 dB

L_{E2} = excess path loss for back-scatter signal = 2 dB

L_{GT} = waveguide loss between transmitter and antenna = 2 dB

L_{GR} = waveguide loss between receiver and antenna = 2 dB

Expressed in decibels, (5.11) is

$$P_r = P_T + G_T + G_R + 20 \log \lambda + 10 \log \sigma - 30 \log(4\pi) - 40 \log(R) - L_{E1} - L_{E2} - L_{GT} - L_{GR} \quad (5.12)$$

Substituting values:

$$P_r = 50 + 33 + 33 - 12.8 + 10 - 33 - 200 - 2 - 2 - 2 - 2$$

$$P_r = -127.8 \text{ dBW} = -97.8 \text{ dBm} = 16 \times 10^{-13} \text{ W}$$

5.15.2 Example 5: X-Band Airborne Multiple-Function Radar

Assumptions:

$$P_T = \text{transmitter output power} = 100 \text{ kW}$$

$$G_T = \text{transmitter antenna gain} = 38 \text{ dBi}$$

$$G_R = \text{receiver antenna gain} = 38 \text{ dBi}$$

$$\lambda = \text{wavelength} = 0.033 \text{ m} \quad s = \text{target RCS} = 10 \text{ m}^2$$

$$R = \text{range} = 100 \text{ km}$$

$$\sigma = 10 \text{ m}^2$$

$$L_{E1} = \text{excess path loss for transmitted signal} = 2 \text{ dB}$$

$$L_{E2} = \text{excess path loss for back-scatter signal} = 2 \text{ dB}$$

$$L_{GT} = \text{waveguide loss between transmitter and antenna} = 2 \text{ dB}$$

$$L_{GR} = \text{waveguide loss between receiver and antenna} = 2 \text{ dB}$$

Substituting values into (5.12):

$$P_r = 50 + 38 + 38 - 29.6 + 10 - 33 - 200 - 2 - 2 - 2 - 2$$

$$P_r = -134.6 \text{ dBW} = -104.6 \text{ dBm} = 3.5 \times 10^{-14} \text{ W}$$

References

- [1] Barton, D. K., *Modern Radar System Analysis*, Norwood, MA: Artech House, 1988, Chap. 6.
- [2] Barton, D. K., *Modern Radar System Analysis*, Norwood, MA: Artech House, 1988, p. 283.
- [3] Barton, D. K., *Modern Radar System Analysis*, Norwood, MA: Artech House, 1988, p. 300.
- [4] Barton, D. K., *Modern Radar System Analysis*, Norwood, MA: Artech House, 1988, pp. 288–296.
- [5] Kennedy, G., *Electronic Communication Systems*, 3rd ed., New York: McGraw-Hill, 1985, pp. 221–231.
- [6] Crombic, D. D., “Electromagnetic-Wave Propagation,” in *Reference Data for Engineers*, 8th ed., M. E. Van Valkenburg, (ed.), SAMS, Carmel, IN: Prentice Hall Computer Publishing, 1993, pp. 33-3–33-12.
- [7] Kennedy, G., *Electronic Communication Systems*, 3rd ed., New York: McGraw-Hill, 1985, pp. 545–548.

RF Noise and Link Analysis

6.1 Concepts of RF Noise and Signal-to-Noise Ratio

Noise limits the performance of all receivers. To be effective, the received signal must be substantially greater than the RF noise. Figure 6.1 lists noise sources, including internal noise sources, where the noise is generated by the receiver itself, and external noise sources, in which the noise is received by the receiver antenna. External noise sources include both natural and man-made sources. Man-made noise sources include industrial noise, signals from radar or communication transmitters that provide interference, and jamming sources. Natural noise sources include atmospheric noise such as lightning storms, galactic noise, solar noise, atmospheric-loss noise, and hot-Earth noise. At the lower frequencies (VLF, LF, MF, and HF frequencies), external noise received by the receiver antenna from either a natural or a man-made source generally is much greater than internal noise and therefore is the limiting factor in system performance. At VHF and higher frequencies, internally generated noise generally is greater than external noise received by the antenna and therefore is the limiting factor in system performance.

The intended signal is transmitted by a signal transmitter and is received by the receiver antenna. At the receiver, the signal is added to the external noise, also received by the receiver antenna, and to the internal noise generated by the receiver. If the combined noise is much smaller than the signal, it will not significantly degrade the detected signal. However, if the noise is large compared to the desired signal, detection of the signal may not be possible.

Figure 6.2 illustrates detection of a signal in the presence of noise for a detected radar pulse. The assumption is made that both the signal and the noise have been filtered by a bandpass filter having a near-optimum bandwidth for detection of the radar signal. Thus, the frequency components for the noise are the same as those for the signal. Without filtering, the noise would have a wide range of frequency components with details, depending on the type of noise present.

Figure 6.2 is an amplitude-versus-time plot. Often a threshold is set as shown to permit a decision of whether a signal is present. The threshold level is set sufficiently above the noise level that the false alarm rate will be acceptable. Signals greater than that threshold level are classed as true signals. It cannot be set too high, however, or the probability of detection may be too low.

Power requirements are such to permit an acceptable probability of detection of the signal. A typical required minimum signal-to-noise ratio for radar after integration is about +15 dB. A typical required minimum signal-to-noise ratio for a cellular

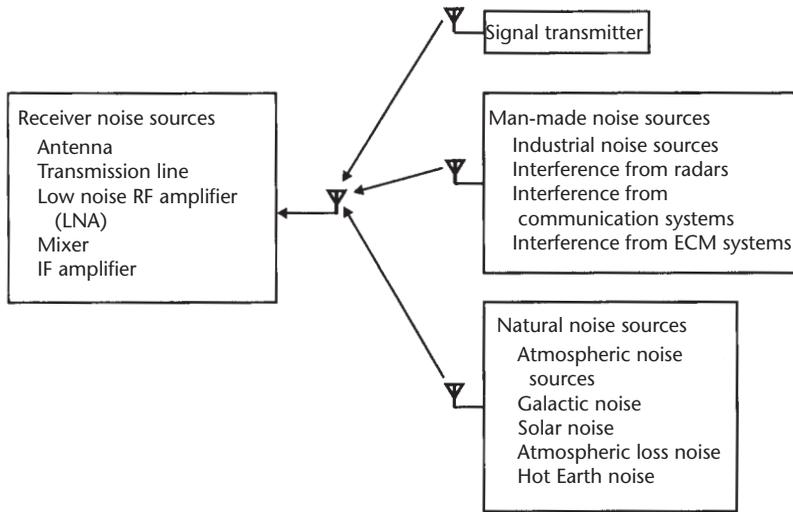


Figure 6.1 Noise sources.

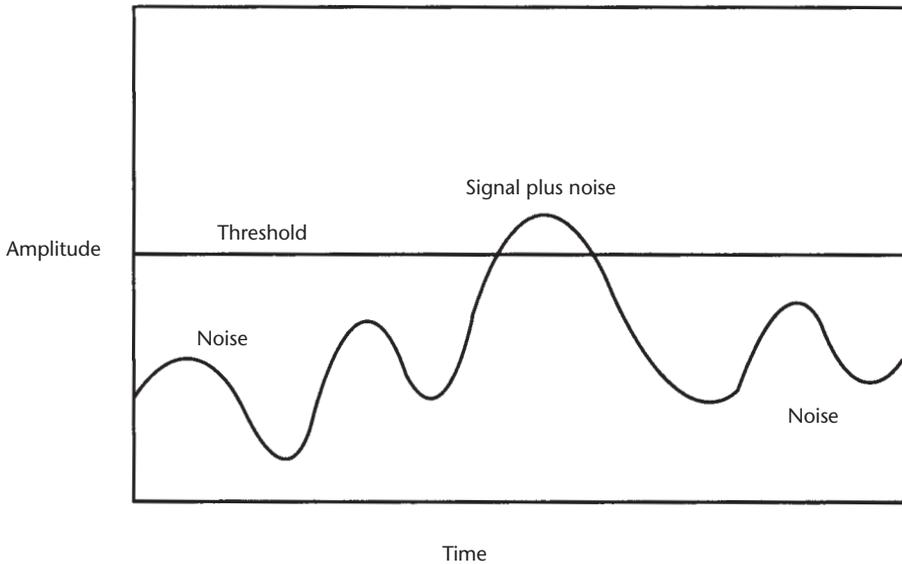


Figure 6.2 Detection of a radar signal in the presence of noise. (After: [3].)

telephone system might be +18 dB. Broadcast communication systems for entertainment would require +30 dB or higher signal-to-noise ratios.

6.2 Noise Power, Noise Temperature, and Noise Figure

The information in this chapter is based on [1–4].

Noise power is given by

$$P_n = kTB \quad (6.1)$$

where:

P_n = available noise power (watts)

k = Boltzmann's constant = $1.38 \cdot 10^{-23}$ J/K

T = noise temperature (kelvin)

B = effective receiver noise bandwidth (hertz)

For example, the noise power corresponding to the reference temperature $T_0 = 290\text{K}$ and a bandwidth of 1 Hz would be

$$P_n = 138 \cdot 10^{-23} \cdot 290 \cdot 1 = 4 \cdot 10^{-21} \text{ W}$$

Expressed in decibels, that would be

$$P_n = -204 \text{ dBW/Hz} = -174 \text{ dBm/Hz}$$

If a system had a noise temperature of 5,000K and an effective noise bandwidth of 1 MHz, the output noise power would be

$$P_n = 138 \cdot 10^{-23} \cdot 5,000 \cdot 10^6 = 69 \cdot 10^{-14} \text{ W}$$

Expressed in decibels, that would be -131.6 dBW or -101.6 dBm .

Another important expression is the noise power expressed in terms of the reference temperature, T_0 , and a noise figure F_n , where F_n is the ratio of the real noise temperature to the reference temperature. The expression is

$$P_n = kT_0BF_n \quad (6.2)$$

where:

P_n = available noise power (watts)

k = Boltzmann's constant = $1.38 \cdot 10^{-23}$ J/K

T_0 = reference noise temperature (290K)

B = effective receiver noise bandwidth (hertz)

F_n = noise figure

As an example of the use of (6.2), assume that a system has a noise figure of 7 dB or 5.1 numeric and an effective noise bandwidth of 25 kHz. Substituting values into (6.2), the noise power would be

$$P_n = 138 \cdot 10^{-23} \cdot 290 \cdot 25 \cdot 10^3 \cdot 5 = 5 \cdot 10^{-16} \text{ W}$$

Expressed in decibels, we have

$$P_n = 10 \log_{10}(kT_0) + 10 \log_{10}(B) + F_n$$

$$P_n = -204 + 44 + 7 = -153 \text{ dBW} = -123 \text{ dBm}$$

6.3 Multiple-Stage Systems with Noise

The information in this section is based on [2, 3]. It is common practice to use noise temperatures in performing noise calculations for complex systems involving noise stages and gain or loss stages in series. Figure 6.3 illustrates such a system. In the figure, T_1 , the antenna-noise temperature, represents the received noise from outside sources. T_2 is the thermal temperature of the transmission line system, usually assumed to be 290K. G_2 is the gain of the transmission line system. That gain is less than 1 and is actually a loss. T_3 and G_3 are the noise temperature and the gain of the low noise amplifier (LNA), a stage that may not be used in some systems. T_4 and G_4 are the noise temperature and the gain of the receiver.

It is common practice to provide the noise characteristics of LNAs and receivers in terms of noise figures, which are converted using (6.3).

$$T_e = (F_n - 1)T_0 \quad (6.3)$$

where:

T_e = noise temperature (kelvin)

F_n = noise figure (numeric)

T_0 = reference temperature (290K)

For example, if the noise figure of a receiver is given as 7 dB, we first would convert to a numeric equivalent, which for this case would be 5.0. Then we would calculate the noise temperature as follows:

$$T_e = (5 - 1) \cdot 290 = 1,160\text{K}$$

To convert a noise temperature to a noise figure, we use (6.4), a simple conversion from (6.3):

$$F_n = 1 + T_e / T_0 \quad (6.4)$$

For example, if the noise temperature of a system is 3,000K, the noise figure for the system would be

$$F_n = 1 + 3,000/290 = 11.3 = 10.5 \text{ dB}$$

The effective noise temperature of a receiver consisting of a number of networks in cascade is given by (6.5):

$$T_e = T_1 + T_2 / G_1 + T_3 / G_1 G_2 + \dots \quad (6.5)$$

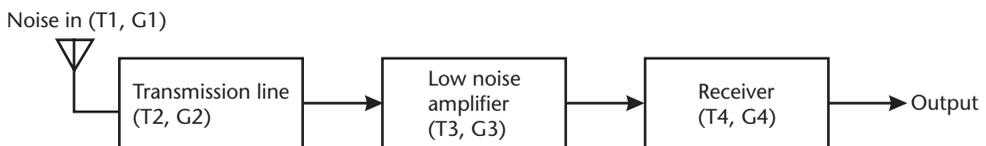


Figure 6.3 Multiple stage system with noise.

where T_i and G_i are the effective noise temperature and gain of the i th network. Both T_i and G_i are real numbers, not decibels.

An alternative equation for noise figure for a number of networks in cascade is given by (6.6):

$$F = F_1 + (F_2 - 1)/G_1 + (F_3 - 1)/G_1 G_2 + (F_4 - 1)/G_1 G_2 G_3 \quad (6.6)$$

where F_i is the noise figure, numeric.

For an example of the use of (6.5) for the system shown in Figure 6.3, assume the following:

$$\begin{aligned} T_1 &= 500\text{K} & G_1 &= 1 \\ T_2 &= 290\text{K} & G_2 &= 0.3 \\ T_3 &= 200\text{K} & G_3 &= 100 \\ T_4 &= 1,200\text{K} & G_4 &= 10^{10} \end{aligned}$$

Then,

$$\begin{aligned} T_e &= T_1 + T_2/G_1 + T_3/G_1 G_2 + \dots \\ T_e &= 500 + 290/1 + 200/(0.3 \cdot 1) + 1,200/(0.3 \cdot 1 \cdot 100) \\ T_e &= 500 + 290 + 667 + 40 = 1,497\text{K} \end{aligned}$$

The system noise figure would be found using (6.4), as follows:

$$\begin{aligned} F_n &= 1 + T_e/T_0 \\ F_n &= 1 + 1,497/290 = 62 = 79\text{dB} \end{aligned}$$

6.4 Types of Noise

The information in this section is based on [1, 4]. The types of external noise that are important to RF systems include the following:

- Atmospheric noise;
- Galactic noise;
- Solar noise;
- Ground noise;
- Man-made noise.

6.4.1 Atmospheric Noise

Figure 6.4 shows typical effective antenna noise figures, in decibels, above kT_0B for frequencies below 100 MHz and locations in the central region of the United States. Below 30 MHz, the strongest source of noise is atmospheric noise, generated mostly by lightning discharge in thunderstorms. The noise level depends on the frequency, the time of day, weather, the season of the year, and geographical location.

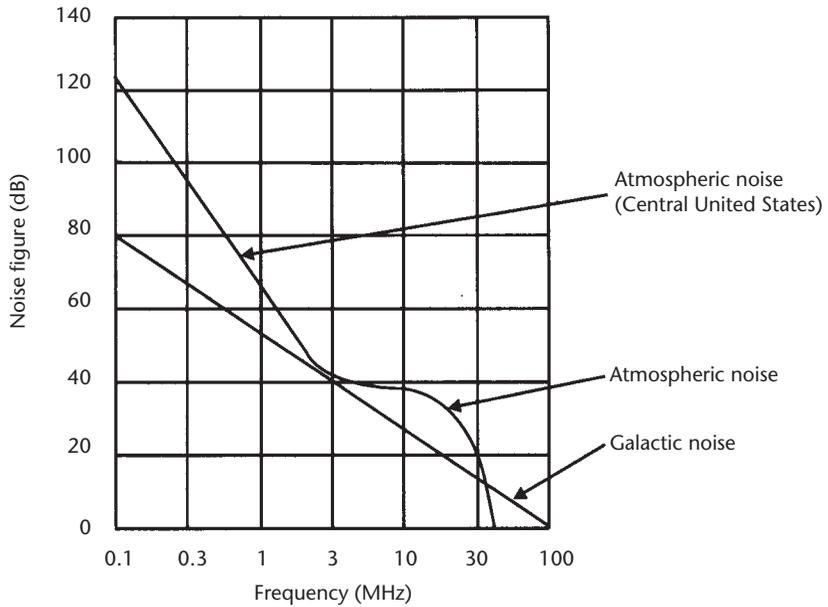


Figure 6.4 RF antenna noise figure below 100 MHz. (After: [1].)

Atmospheric noise is greatest at the lowest frequencies and decreases with increasing frequency. It propagates over the Earth in the same fashion as ordinary radio waves of the same frequencies. Thus, at HF and lower frequencies, it is not only the lightning storms in close proximity that cause the interference or noise, but storms any place in the world where propagation paths are possible.

Atmospheric noise becomes less severe at frequencies above about 30 MHz because of two factors. First, the higher frequencies are limited to line-of-sight propagation. Second, the mechanism generating the noise is such that very little of it is created in the VHF range and higher.

The region with the largest atmospheric lightning noise for the western hemisphere is the Caribbean near Panama. The median noise level for that area is about 85 dB above kT_0B at 1.0 MHz. The central and eastern regions of the United States have median values of noise power of about 70 dB above kT_0B at 1 MHz. The average atmospheric noise level for the western portion of the United States is substantially less, being about 60 dB in Utah and 50 dB in California at 1 MHz. The atmospheric noise levels decrease rapidly as we move northward toward the pole and as we move out over the Pacific and Atlantic oceans. The peak atmospheric noise can exceed the median values of average noise power by 10 dB or more. Atmospheric noise is lower than that in the summer by about 10 dB over much of the frequency range.

6.4.2 Galactic Noise

Figure 6.4 shows that above about 30 MHz and less than 100 MHz galactic noise is the largest type of natural noise. It is similar in magnitude to man-made noise in a quiet location.

Figure 6.5 shows typical antenna temperatures for frequencies above 100 MHz. Galactic noise is the largest natural noise component between 100 and 400 MHz. Above 400 MHz, other noise components dominate. Those other components are solar noise, atmospheric-loss noise, and hot-Earth noise. The plot in Figure 6.5 assumes that the Sun and the Earth are seen only through antenna sidelobes. If the earth is viewed with the main beam, the antenna noise temperature due to the hot Earth could be as high as 300K. The solar noise temperature could be much greater.

Galactic noise is most intense in the galactic plane and reaches a maximum in the direction of the galactic center. That center appears to be about 3 degrees in diameter. Antenna noise temperature resulting from galactic noise may be greater than 18,000K for a narrow-beam antenna pointed at the galactic center in the region of 100 MHz and less than 3K above about 1,000 MHz. Stronger than average noise radiation is received from a broader region within 30 degrees of the galactic plane. The minimum galactic noise antenna temperature that is measured when pointing at the galactic poles is about 500K at 100 MHz. The geometric mean of these maximum and minimum temperatures is 3,000K at 100 MHz.

For an example of galactic antenna noise temperature for directional antennas, assume the use of an antenna with 30-dBi antenna gain (beam angle 5 degrees diameter) operating at 100 MHz and pointing at the galactic center (3 degrees diameter). The fraction of the beam occupied by the source would be about 0.36. The noise temperature of the galactic center at 100 MHz is assumed at 18,000K, and the atmospheric loss is assumed to be 1 (0 dB). The antenna temperature due to galactic noise from only the galactic center may be found using (6.7).

$$T_a = a_c T_c / L_A \quad (6.7)$$

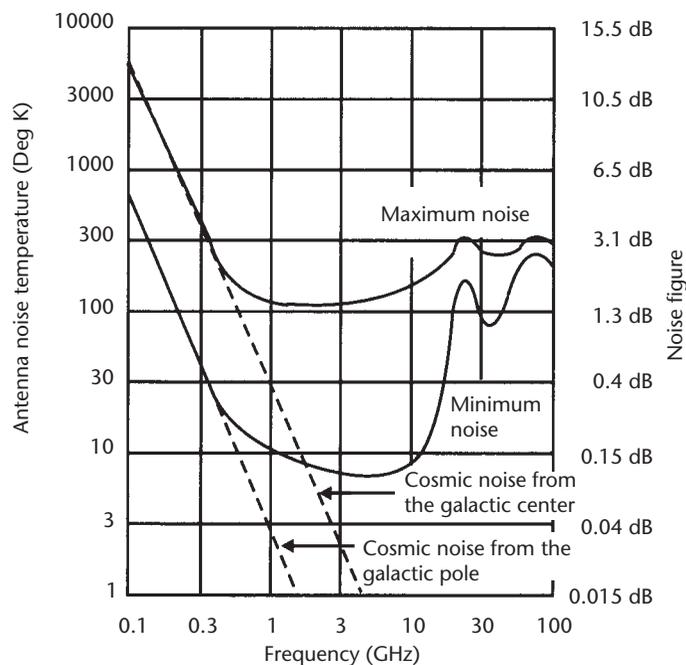


Figure 6.5 Radio frequency antenna noise temperature versus frequency above 100 MHz. (After: [4].)

where:

T_a = antenna noise temperature (kelvin)

a_c = fraction of beam filled by noise source

T_c = temperature at galactic center (kelvin)

L_A = Atmospheric loss (numeric)

Using those assumptions:

$$T_a = 0.36 \cdot 18,000/1 = 6,480\text{K}$$

If the antenna beam angle is less than 3 degrees (the approximate diameter of the galactic center) and the antenna points at the galactic center, the antenna temperature will be no greater than 18,000K. If the antenna beam angle is much larger than 3 degrees and points at the galactic center, it will see other parts of the galaxy in addition to the center.

If a high-gain (highly directional) antenna points at the galactic pole, it will see a noise temperature of about 500 degrees at 100 MHz. In no case will the antenna see less than the temperature of the galactic pole.

It is clear from Figure 6.5 that galactic noise usually is the strongest natural noise between 100 MHz and about 400 MHz. Other types of natural noise become important above 400 MHz and typically produce the maximum and minimum temperatures as shown.

6.4.3 Solar Noise

The Sun is a powerful noise source. If a directional antenna is pointed at the Sun, it will see a large antenna noise temperature. The Sun also can contribute appreciably to antenna noise through sidelobes of the antenna. During high levels of sunspot activity, noise temperatures from 100 to 10,000 times greater than those of the quiet sun may be observed for periods of seconds in what is called solar bursts, followed by levels about 10 times the quiet level lasting for several hours.

At microwave frequencies, the Sun's noise diameter is one-half degree. The Sun's effective noise temperature seen by an antenna of gain G_s is given by (6.8) and (6.9).

$$T_a = a_s T_s / L_A \quad (6.8)$$

$$T_a = 4.75 \cdot 10^{-6} G_s T_s / L_A \quad (6.9)$$

where:

T_a = antenna noise temperature

a_s = fraction of the beam filled by the Sun

T_s = noise temperature of the Sun

L_A = atmospheric loss (numeric)

G_s = antenna gain

Equation (6.9) can be derived from (6.8) as follows: The directional gain of the antenna is $G = 41,300/(\text{beam angle squared})$. The solid angle covered by the antenna is thus $41,300/G_s$.

The solid angle covered by the sun is 0.196 square degrees. Thus, the fraction of the beam filled by the Sun is

$$0.196/(41,300/G_s) = 4.75 \cdot 10^{-6} = a_s$$

The value for a_s is used in (6.8) to yield (6.9).

For example, if the frequency is 1,000 MHz and there are quiet sun conditions, the noise temperature of the Sun is about $(2 \cdot 10^5)\text{K}$. If we assume an antenna gain in the direction of the Sun of 31 dBi and an atmospheric loss of 1 dB, the antenna noise temperature will be

$$\begin{aligned} T_a &= 4.75 \cdot 10^{-6} G_s T_s / L_A \\ T_a &= 4.75 \cdot 10^{-6} \cdot 1,259 \cdot 2 \cdot 10^5 / 126 \\ T_a &= 949.3\text{K} \end{aligned}$$

6.4.4 Ground Noise

The Earth is a radiator of electromagnetic noise. The thermal temperature of the Earth is typically about 290K. In radar systems and directional communication systems, the Earth will be viewed mainly through the sidelobes of the antennas. The average sidelobe antenna gain typically is about -10 dBi. A rough estimate of antenna noise temperature in that case is 29K. The maximum possible ground antenna noise temperature is about 290K for directional antennas pointing at the ground.

6.4.5 Man-Made Noise and Interference

Man-made noise is due chiefly to electric motors, neon signs, power lines, and ignition systems located within a few hundred yards of the receiving antenna. In addition, there may be radiation from hundreds of communication and radar systems that may interfere with reception. The interference may include signals on assigned frequencies, harmonics of those signals, unwanted spurs (mixing products of signals generated within transmitters), subharmonics, and leakage from low-frequency oscillators used in multiplier chains.

Propagation of man-made noise and interference may be by means of power lines, by ground wave, and by any of the other propagation modes discussed in Chapter 5. The amplitude of man-made noise decreases with increasing frequency and varies considerably with location. Generally this type of noise is assumed to decrease with frequency as shown in the (6.10):

$$T_a = T_{100} (100/f_{\text{MHz}})^{2.5} \quad (6.10)$$

where T_{100} is the man-made noise temperature at 100 MHz.

For an example of the use of (6.10), assume an operating frequency of 400 MHz and a man-made noise temperature at 100 MHz of 300,000K ($F_a = 30.2$ dB above kT_0B). The calculated antenna noise temperature would be

$$T_a = T_{100} (100/f_{\text{MHz}})^{2.5}$$

$$T_a = 300,000 \cdot (100/400)^{2.5} = 9,375\text{K}$$

The temperature is about 15.2 dB above kT_0B .

To determine the man-made noise at a given site, it is necessary to make noise measurements, which will differ for time of day, season of the year, and direction. The man-made noise and interference signal typically exceed natural noise in the frequency range of 10 to 1,000 MHz.

6.5 Signal-to-Noise Improvement by Use of Integration

The information in this section is based on [3].

Figure 6.6 shows the signal-to-noise ratio improvement that is possible with the use of different signal processing techniques. Three cases are given in the figure. The greatest improvement results from coherent integration, where the term coherent refers to having phase reference and equal phase. That is referred to as a “perfect” integrator, because the processing improvement is equal to the number of pulses integrated. Coherent integration systems use synchronous detectors or phase detectors. The integration improvement factor (IIF), expressed in decibels, for this type of integration is

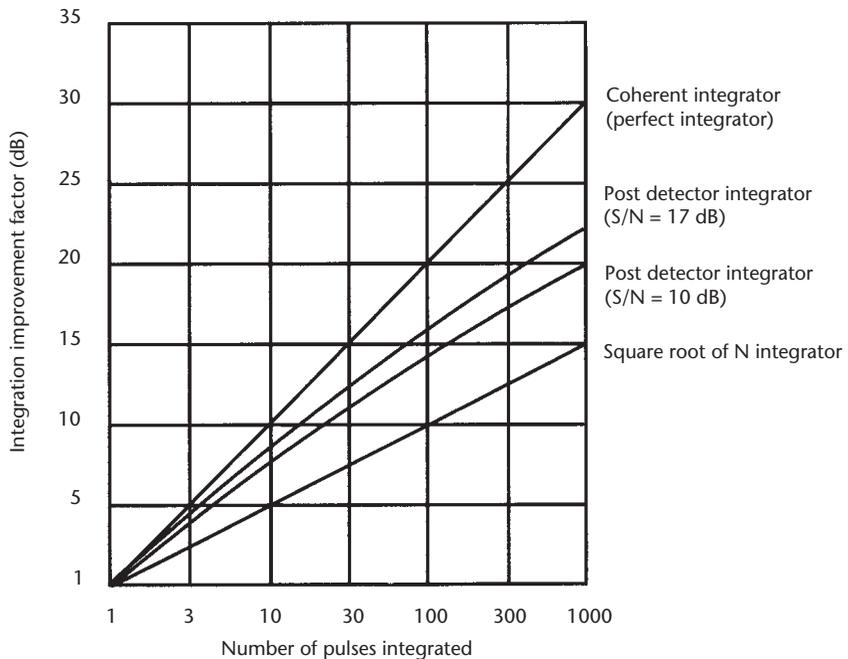


Figure 6.6 Integration improvement factor versus number of pulses integrated. (After: [3].)

$$\text{IIF} = 10 \log_{10} N$$

where N = number of pulses integrated.

The next level of improvement is for noncoherent integration. The type of detectors used in this case may be simple envelope detectors for AM. Here the improvement is less than perfect and is a function of the signal-to-noise ratio or the probability of detection.

The lowest level of improvement is the case where the integration improvement is proportional to the square root of the number of pulses integrated. An example of this type of integration is an operator viewing a CRT screen, where the IIF is

$$\text{IIF} = 10 \log_{10} (N^{0.5}) = 5 \log_{10} N$$

where N = number of pulses integrated.

Integration often is used with radar systems: many pulses are integrated to improve the probability of detection and to improve tracking accuracy. For example, assume that a radar integrates 100 pulses. If the radar uses coherent integration, the integration improvement factor, as given by Figure 6.6, is 20 dB. If the single-pulse signal-to-noise ratio is 17 dB and postdetector integration is used, the improvement factor as shown by Figure 6.6 is about 16 dB. If the single-pulse signal-to-noise ratio is 10 dB, the postdetector IIF is about 14 dB. If the integrator is a square root of the N integrator, the improvement factor, as shown by Figure 6.6, is only 10 dB.

6.6 Signal-to-Noise Ratio

The information in this section is based on [3]. The required signal-to-noise ratio for a radar system can be found in Figure 6.7. This figure shows the probability of detection, also called the probability of intercept (POI), for a sinusoidal signal in the presence of noise as a function of the signal-to-noise (power) ratio and the probability of false alarm. For an example of the use of Figure 6.7, assume that we want a probability of detection of 0.99 or greater and a probability of false alarm of no more than 10^{-7} . From Figure 6.7, we see that the required signal-to-noise ratio without integration, that is, postprocessing (single-pulse detection), is about 15 dB. For another example, assume a probability of false alarm of 10^{-6} and a probability of detection of 0.9. From Figure 6.7, the required signal-to-noise ratio is about 13.3 dB, again without any postprocessing.

The term false alarm means that the system indicates the presence of a signal when in reality it is merely noise. An interpretation of the false-alarm probability is as follows: On the average, there will be one false decision out of n_f possible decisions in the false-alarm time interval T_{fa} . The total number of decisions, n_f , in T_{fa} is equal to the number of range intervals per pulse period times the number of pulses per second (pulse repetition rate, or PRR) times the false-alarm time. Thus, the number of possible decisions is T_{fa}/τ , where τ is the pulse width.

A generally employed assumption is that the product of the receiver bandwidth and the pulse width ($B\tau$) is equal to 1. Errors in that approximation are not serious

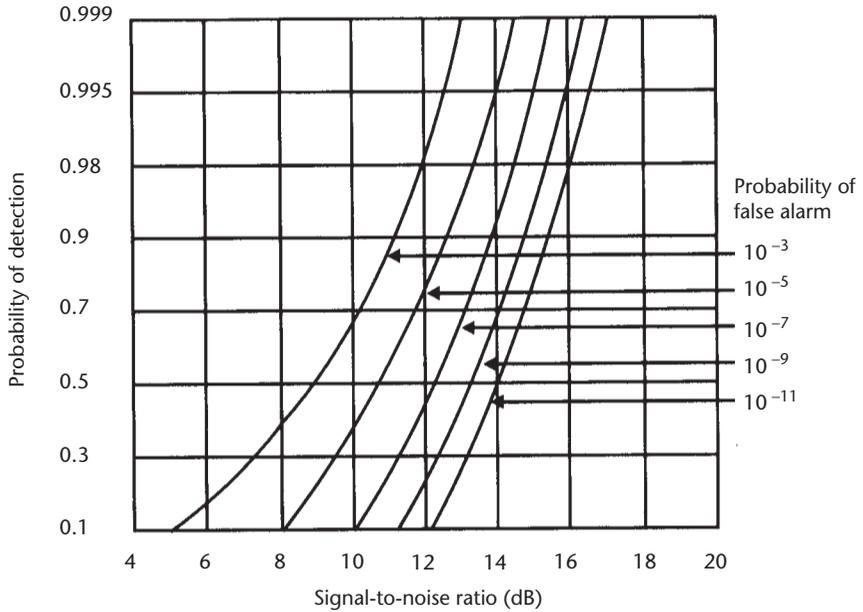


Figure 6.7 Required signal-to-noise ratio versus probability of detection and probability of false alarm. (After: [3].)

in practice because of the exponential relationship between the threshold and the probability of false alarm.

Using those assumptions, the probability of false alarm is

$$P_{fa} = 1/n_f = 1/T_{fa} B \tag{6.11}$$

where:

n_f = total number of decisions

T_{fa} = false-alarm time interval

B = receiver bandwidth

For an example of the use of (6.11), assume a radar with a pulse width of 1 μ s and a bandwidth of 1 MHz. If we want a false alarm of no more than once an hour (3,600 seconds), the false-alarm probability is calculated to be $1/(3,600 \cdot 10^6) = 2.8 \cdot 10^{-10}$.

If n pulses are processed so as to improve detection, the number of independent decisions in the time T_{fa} will be reduced by a factor of n . As a result, the equation for probability of false alarm becomes

$$P_{fa} = n/n_f = n/T_{fa} B \tag{6.12}$$

where the terms are the same as for (6.11) and n is the number of pulses.

For example, to compute the probability of false alarm with integration, assume the following:

$B = 0.5$ MHz

$T_{fa} = 1,000$ seconds

$$n = 10$$

$$P_{fa} = 10 / (1,000 \cdot 0.5 \cdot 10^6) = 2 \cdot 10^{-8}$$

The required signal-to-noise ratio for a communication system depends on the type of modulation used and the type of application. For AM and FM communication, it generally is sufficient to have a signal-to-noise ratio of 20 dB or greater. For broadcast of music and video, it usually is necessary to have a somewhat higher signal-to-noise ratio. A typical requirement might be 30 dB or more. These requirements are for analog signals. With digital signals, such as the newer DSS and DBS systems, the signal-to-noise requirement can be much less (perhaps as low as 15 dB).

6.7 Communication System Link Analysis

The information in this section is based on [2, 3].

A communication system link analysis is a set of calculations that shows the signal-to-noise ratio that is achieved at the receiver output based on a given set of assumptions for a communication system. This section shows the equations used in the communication system link analysis and an example link analysis for a typical communication system.

The equation for received power at the receiver antenna output is as follows:

$$P_R = P_T G_T G_R \lambda^2 / \left[(4\pi)^2 R^2 L_{E1} \right] \quad (6.13)$$

where:

P_R = receiver antenna power output

P_T = transmitter output power at the antenna input

G_T = transmitter antenna gain

G_R = receiver antenna gain

λ = wavelength (meters)

R = range (meters)

L_{E1} = one-way excess propagation loss (>1)

The equation for the receiver system noise power is as follows:

$$N = kT_0 B F_n \quad (6.14)$$

where:

$kT_0 = -174$ dBm/Hz = $4 \cdot 10^{-18}$ mW/Hz = $4 \cdot 10^{-21}$ W/Hz

B = receiver bandwidth (hertz)

F_n = receiver system noise figure (or noise factor) (numeric)

The equation for the signal-to-noise ratio is thus

$$S/N = P_T G_T G_R \lambda^2 / \left[(4\pi)^2 R^2 L_{E1} kT_0 B F_n \right] \quad (6.15)$$

where terms are as defined for (6.13) and (6.14).

For an example of the use of (6.15), assume the following:

P_T = transmitter output power = 100W = 50 dBm

G_T = transmit antenna gain = 3 dBi

G_R = receive antenna gain = 3 dBi

λ = wavelength = 0.1m = -10 dB

R = range = 100 km = 50 dB

LE_1 = receive excess propagation loss = 6 dB

kT_0 = reference noise level = -174 dBm

B = receiver bandwidth (20 kHz) = 43 dB

F_n = receiver system noise figure = 6 dB

Using those values and (6.15) expressed in decibels, we have

$$S/N = P_T + G_T + G_R + \lambda^2 - (4\pi)^2 - R^2 - L_{E1} - kT_0 - B - F_n$$

$$S/N = 50 + 3 + 3 - 20 - 22 - 100 - 6 + 174 - 43 - 6 = 33 \text{ dB}$$

6.8 Radar System Link Analysis

The information in this section is based on [3, 4].

A radar system link analysis is a set of calculations that shows the signal-to-noise ratio that is achieved at the receiver output based on a given set of assumptions for a radar system. This section shows equations that are used in the radar system link analysis and an example link analysis for a typical radar system.

The equation for received power at the receiver antenna output is as follows:

$$P_R = P_T G_T G_R \lambda^2 \sigma / \left[(4\pi)^3 R^4 L_{E1} L_{E2} \right] \quad (6.16)$$

where:

P_R = receiver antenna power output

P_T = transmitter output power

G_T = transmitter antenna gain

G_R = receiver antenna gain

λ = wavelength (meters)

σ = RCS (square meters)

R = range (meters)

LE_1 = transmit excess propagation loss

LE_2 = receive excess propagation loss

The equation for signal-to-noise ratio with integration gain, G_i , included is as follows:

$$S/N = P_T G_T G_R \lambda^2 \sigma G_i / \left[(4\pi)^3 R^4 L_{E1} L_{E2} kT_0 B F_n \right] \quad (6.17)$$

where terms are as previously defined.

Assumptions for the radar system are as follows:

P_T = transmitter output power (100 kW) 80 dBm

G_T = transmitter antenna gain 40 dBi

G_R = receiver antenna gain 40 dBi

λ = wavelength (0.063m) -12 dB

σ = RCS (1.6 m²) 2 dB

G_i = integration signal-to-noise gain 8 dB

R = range (100 km) 50 dB

LE_1 = transmit excess propagation loss 6 dB

LE_2 = receive excess propagation loss 6 dB

kT_0 = reference noise level -174 dBm

B = receiver bandwidth (100 kHz) 50 dB

F_n = receiver system noise figure 6 dB

Using those values and (6.16) expressed in decibels, we have

$$S/N = P_T + G_T + G_R + \lambda^2 + \sigma + G_i - (4\pi)^3 - R^4 - L_{E1} - L_{E2} - kT_0 - B - F_n$$

$$S/N = 80 + 40 + 40 - 24 + 3 + 8 - 33 - 200 - 6 - 6 + 174 - 50 - 6$$

$$S/N = 20 \text{ dB}$$

We see that the signal-to-noise ratio changes by 12 dB per octave or 40 dB per decade in range. It is directly proportional to the RCS of the target.

6.9 Performance Calculations for Radar Systems with Electronic Countermeasures

When ECM is present with a radar, the jamming signal may be much greater than the noise signal. For that case, a different equation must be used to calculate system performance. In the equation, the signal due to jamming is substituted for $kT_0 B F_n$. The jamming signal is as follows:

$$J = P_J G_J G_{RJ} \lambda^2 / \left[(4\pi)^2 (R_J) L_J B_J \right] \quad (6.18)$$

where:

P_J = jammer transmitter power

G_J = jammer transmitter antenna gain

G_{RJ} = receiver antenna gain for jammer signal

λ = wavelength

B_R = receiver bandwidth

R_J = jammer range to receiver

L_J = jammer loss in addition to free-space loss between the jammer transmitter and the receiver

B_J = jammer signal bandwidth

The equation for the signal-to-jamming ratio for radar with integration of a number of pulses is as follows:

$$S/J = P_T G_T G_R \sigma G_i (R_J)^2 L_J B_J / [(4\pi)R^4 L_{E1} L_{E2} P_J G_J G_{RJ} B_R] \quad (6.19)$$

where the terms are as defined previously.

For an example of the use of (6.18) for radar jamming, use the following assumptions.

Radar assumptions:

P_T = transmitter output power 80 dBm

G_T = transmitter antenna gain 40 dBi

G_R = receiver antenna gain 40 dBi

σ = RCS 2 dB

LE_1 = transmit excess propagation loss 6 dB

LE_2 = receive excess propagation loss 6 dB

B = receiver bandwidth 50 dB

G_i = integration gain of signal-to-noise ratio 8 dB

Jammer assumptions:

P_J = jammer output power 60 dBm

G_J = jammer antenna gain 10 dBi

G_{RJ} = receiver antenna gain seen by jammer -3 dBi

L_J = jammer excess propagation loss 6 dB

B_J = jammer noise bandwidth (200 MHz) 83 dB

R_J = jammer range (100 km) 50 dB

Expressing (6.18) in decibels and substituting values, we have

$$S/J = P_T + G_T + G_R + \sigma + G_i + (R_J)^2 + L_J + B_J - (4\pi) - R_4 - L_{E1} - L_{E2} - P_J - G_J - G_{RJ} - B_R$$

$$S/J = 80 + 40 + 3 + 8 + 100 + 6 + 83 - 33 - R^4 - 6 - 6 - 60 - 10 + 3 - 50$$

$$S/J = 198 - R^4$$

If the radar range is 100 km, the S/J ratio is $198 - 200 = -2$ dB. If it is 30 km, the signal-to-jamming ratio is $198 - 179 = 19$ dB.

References

- [1] Swenson, Jr., G. W., and A. R. Thompson, "Radio Noise and Interference," Chap. 34, *Reference Data for Engineers*, 8th ed., M. E. Van Valkenburg, (ed.), SAMS, Carmel, IN: Prentice Hall Computer Publishing, 1993.
- [2] Kennedy, G., *Electronic Communication Systems*, 3rd ed., New York: McGraw-Hill, 1985, Chap. 2.
- [3] Skolnik, M. I., *Introduction to Radar Systems*, 2nd ed., New York: McGraw-Hill, 1980, pp. 15–33.
- [4] Barton, D. K., *Modern Radar System Analysis*, Norwood, MA: Artech House, 1988, pp. 13–18.

Modulation Techniques

Modulation techniques can be either analog or digital. An analog modulation scheme is one in which the carrier signal is modulated or changed in amplitude, frequency, or phase in a way that is directly proportional to the amplitude of the modulating signal. This type of system is continuously variable. A digital modulation system is one in which the signal to be sent is first sampled on a periodic basis and converted to a digital signal using an analog-to-digital (A-to-D) converter. The digitized signal then is used to modulate the carrier signal in amplitude, frequency, phase, or a combination thereof. Both modulation techniques are used for communications and are discussed in this chapter, as are modulation types for radar.

7.1 Pulsed Continuous-Wave Signals

Figure 7.1 shows the waveform and frequency spectrum for a pulsed CW signal. This type of modulation is used for CW telegraphy, pulsed radar, and many other pulse-type systems. The frequency spectrum for rectangular-shaped RF pulses in a pulse train is a $(\sin x)/x$ -type voltage spectrum or a $[(\sin x)/x]^2$ power spectrum. The main spectral lobe (maximum amplitude lobe) is centered at the carrier frequency. It has a width at the nulls of two times the reciprocal of the minimum RF pulse width. For example, if the RF pulse width is $1 \mu\text{s}$, the width of the main spectrum lobe at the nulls is 2 MHz. Each of the sidelobes has half that width between nulls. The first sidelobe is reduced in amplitude by about 13.5 dB from the main spectral lobe. Higher-order sidelobes are progressively lower, falling off at a rate of about 20 dB per decade. The second and third sidelobes are down from the main lobe peak by about 17.9 dB and 20.8 dB, respectively.

Figure 7.1 shows only the envelope of the spectral lines, which are separated by the pulse repetition frequency. Thus, if the pulse repetition frequency were 1,000 Hz, there would be spectral lines every 1,000 Hz. The required -3 -dB bandwidth for a pulsed CW signal as used for radar is approximately the reciprocal of the pulse width. Thus, if the RF pulse width is $1 \mu\text{s}$, the required bandwidth is about 1 MHz.

Figure 7.1 also can be used to show the spectrum for pulsed phase modulation and phase code modulation. In phase code modulation, the pulse width is the shortest phase change interval.

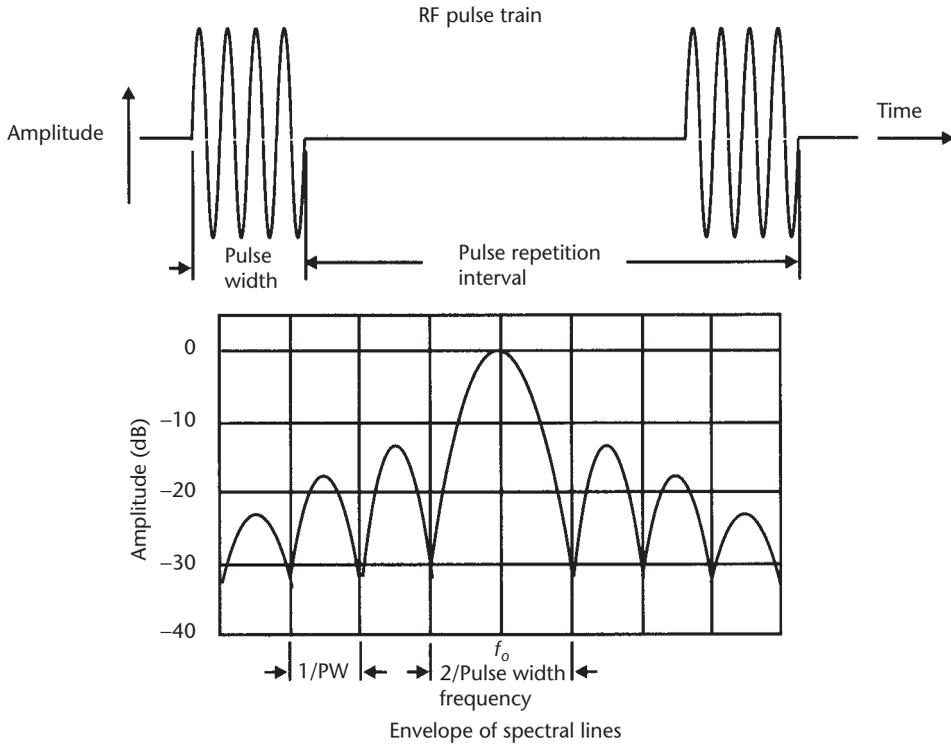


Figure 7.1 Waveform and frequency spectrum for a pulsed CW signal.

7.2 Conventional Amplitude Modulation

Some of the material presented in this and the following four sections is quoted or adapted from [1].

Figure 7.2 shows the waveform and the frequency spectrum for conventional AM as used for standard AM radio. The modulating signal in the figure is a single sine wave with 100% modulation. The RF frequency is constant, and the amplitude is made proportional to the amplitude of the information being sent.

The carrier amplitude may be indicated as V_c and the modulating signal amplitude as V_m . The ratio V_m/V_c is the modulation index (M) and is a number between 0 and 1. This ratio often is expressed as a percentage modulation varying between 0% and 100%. This is an example of analog communication in which signals are continuously variable rather than being digital or on-off in nature.

The relative power in the sidebands and the carrier depends on the percentage of modulation. If we had 100% modulation, the voltage of each sideband for single sine wave modulation would be one-half the voltage of the carrier. The power in each sideband therefore would be one-fourth the carrier power, and the total sideband power would be one-half the carrier power. The maximum power in the sidebands is thus only one-third the total power transmitted. Since it is only the sideband power that carries information, this is very poor efficiency compared to FM and some other modulation methods.

Figure 7.3 shows conventional amplitude modulation with 50% modulation ($M = 0.5$). Again, the modulating signal is a single sine wave. The sideband voltage

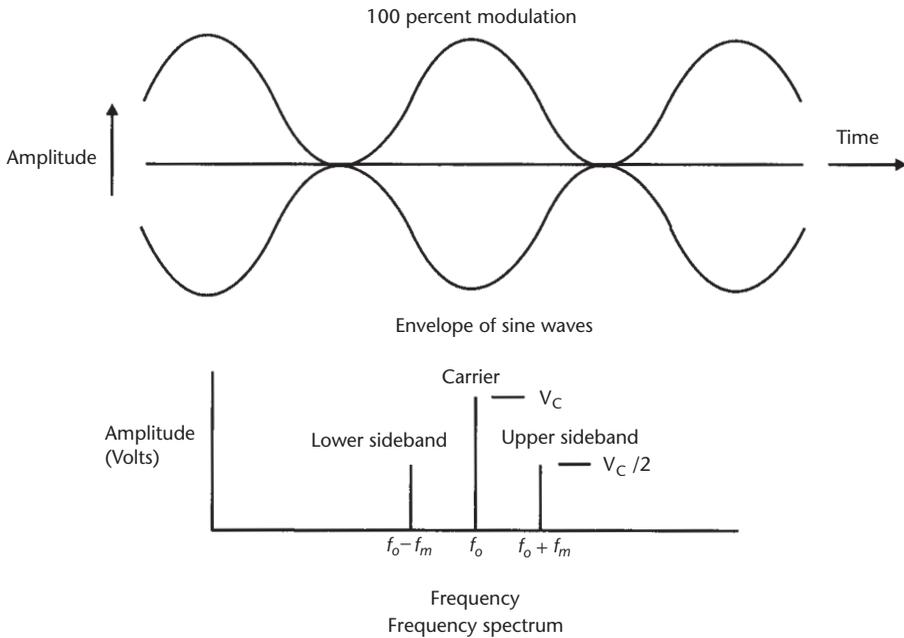


Figure 7.2 Waveform and frequency spectrum for a conventional AM signal with single sine wave modulation and 100% modulation.

amplitude for each sideband is one-fourth the amplitude of the carrier. The sideband power for each sideband is one-sixteenth the power of the carrier, and the total sideband power is one-eighth the carrier power.

If we had only 10% modulation, the voltage of each sideband for single sine wave modulation would be only one-twentieth the voltage of the carrier, and the power in each sideband would be only one-four hundredth the power in the carrier. The efficiency in this case would be very low.

The required bandwidth for conventional AM is typically twice the highest modulating frequency that we wish to pass. For example, with telephone and many voice communication radio systems, it usually is adequate to pass only a 3-kHz (300 to 3,300 Hz) bandwidth. The two sidebands on each side of the carrier thus would have a maximum width of 3 kHz and the total bandwidth would be 6.6 kHz.

7.3 Double Sideband Suppressed Carrier Modulation

Figure 7.4 shows the case of double-sideband (DSB) suppressed carrier modulation for a single sine wave modulation signal. The frequency spectrum thus shows only a lower sideband and an upper sideband but no carrier line.

The DSB signal usually is produced by a modulator known as a balanced modulator, which rejects the carrier frequency component. The operation of this circuit is as follows. Assume that the carrier input signal is sinusoidal and appears at the local oscillator (LO) input of the modulator as $\cos \omega_c t$ and that the baseband input signal to the modulator is $\cos \omega_m t$. The modulator performs a multiplication of the input signals, and the output is

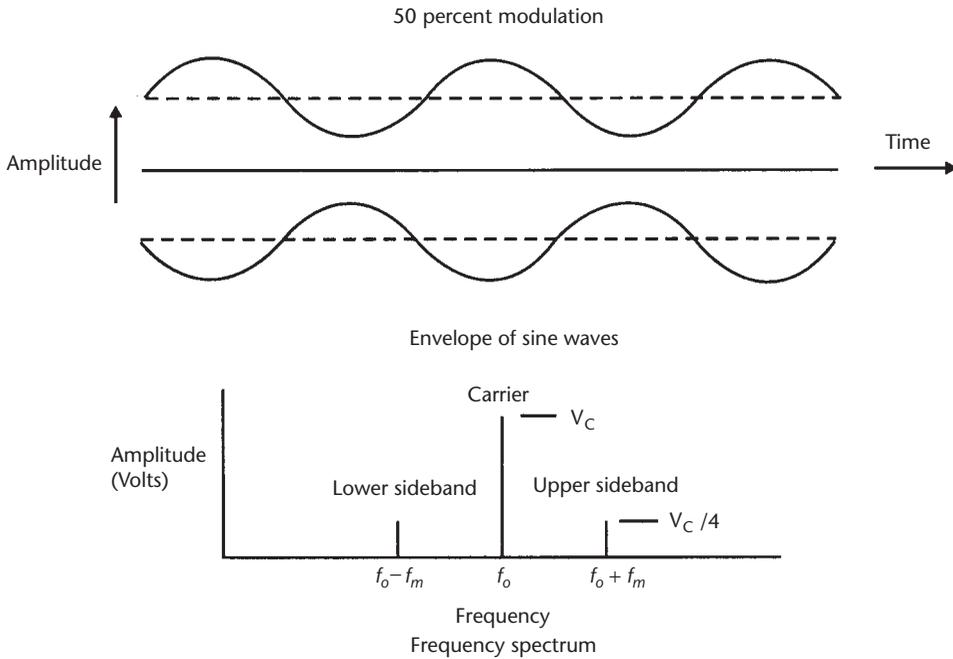


Figure 7.3 Waveform and frequency spectrum for a conventional AM signal with single sine wave modulation input and 50% modulation.

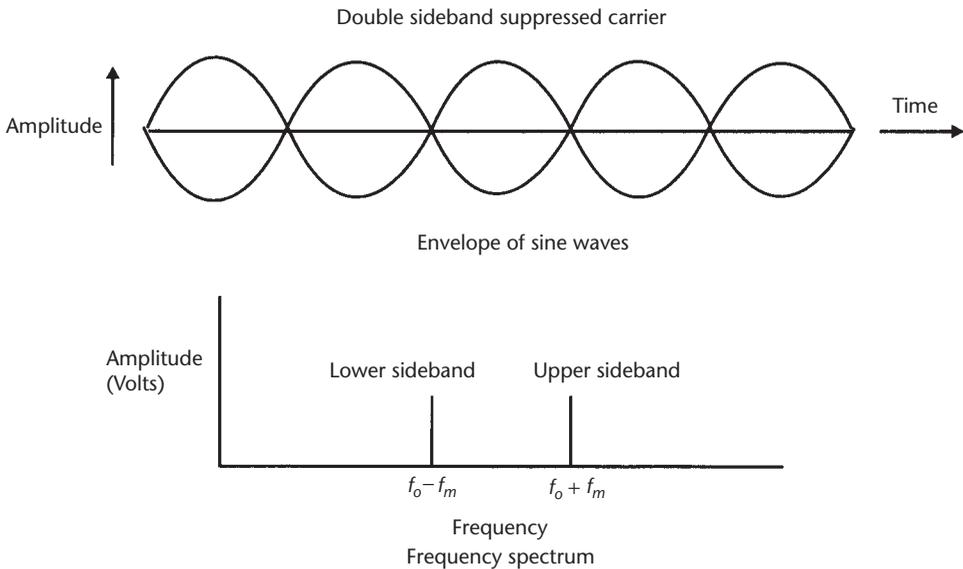


Figure 7.4 Waveform and frequency spectrum for a DSB suppressed carrier modulation (AM) signal with single sine wave modulation.

$$[\cos \omega_c t][\cos \omega_m t] = 1/2[\cos(\omega_c + \omega_m)t + \cos(\omega_c - \omega_m)t] \tag{7.1}$$

We see that the modulator has generated the sum and difference frequencies. It should be pointed out that that is the same circuit or device that could be used for a

mixer. The balanced mixer has as one of its inputs an LO signal and as the other signal a signal that we want to frequency convert. The mixer generates sum and difference frequencies as indicated by (7.1). A filter is used to select which of the two output frequencies we want to pass. If the output selected is the sum frequency, we say that the system is a frequency upconverter. If the output selected is the difference frequency, we say that the system is a frequency downconverter. (More details about mixers are presented in Chapter 8.)

7.4 Vestigial Sideband Modulation

Vestigial-sideband (VSB) modulation may be derived from a conventional AM signal by filtering out part of one of the two sidebands as well as part of the carrier.

Figure 7.5 shows the frequency spectrum for VSB as used for television video transmission. We see that 1.25 MHz of the lower sideband is transmitted with about 0.75 MHz of the lower sideband passed undiminished. This makes sure that the lowest frequencies in the wanted upper sideband are not distorted in phase by the VSB filter. By using this method, 3.0 MHz of the needed frequency spectrum is saved. Thus, instead of requiring 9-MHz bandwidth, as would be the case for AM, it requires only 6.0 MHz [1].

7.5 Single-Sideband Modulation

One of the more important types of analog modulation used for communications is single-sideband (SSB) modulation. This type of amplitude modulation is much more efficient than standard AM because it is not necessary to transmit the full carrier. It also uses only half the frequency bandwidth.

Figure 7.6 shows the spectra for two types of SSB modulation. For simplicity, a single sine wave modulating signal is assumed. Figure 7.6(a) shows an SSB reduced carrier. In this case, sufficient carrier is transmitted to permit the receiver to have a

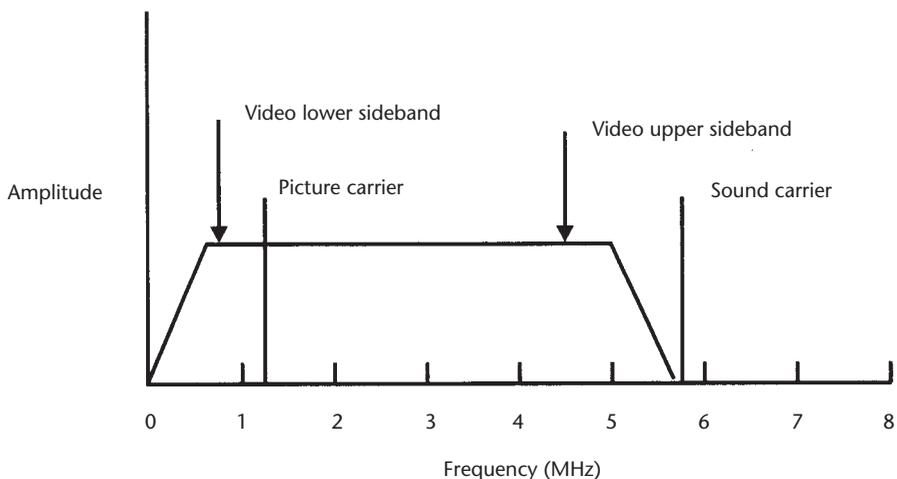


Figure 7.5 Spectrum for transmitted video signals for television. (After: [1].)

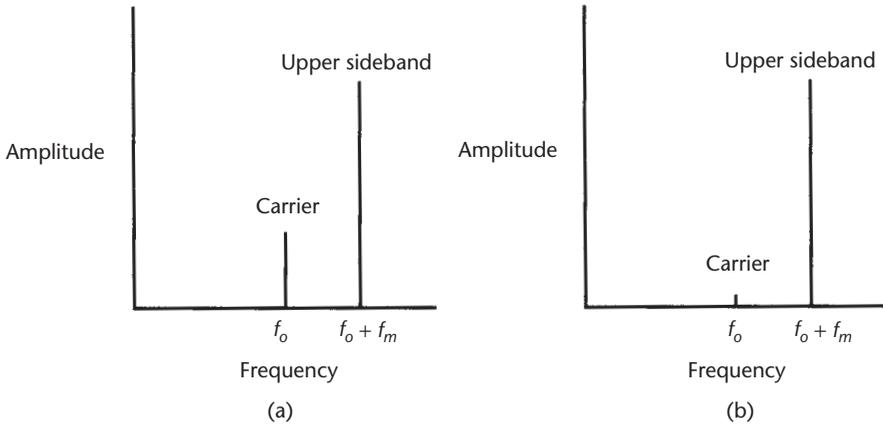


Figure 7.6 SSB modulation: (a) SSB reduced carrier frequency spectrum and (b) SSB suppressed carrier frequency spectrum.

carrier reference on which to lock. The receiver then generates the necessary carrier signal for demodulation of the SSB signals.

Figure 7.6(b) shows the second type of SSB modulation, SSB suppressed carrier. In this case, there is essentially no carrier signal for the receiver to lock on.

In some cases, SSB modulation can be produced with a balanced modulator and a bandpass filter. The balanced modulator eliminates the carrier; the filter must eliminate nearly all the signals on one side of the carrier. In some cases, that is done easily. In other cases, adequate filtering is difficult to realize. That is especially true for the case of low baseband frequencies that are very close to the carrier. An alternative approach for SSB modulation is use of a phase cancellation method. Both methods are discussed in detail in Chapter 9.

7.6 Standard Frequency Modulation

With standard FM, the waveform amplitude is constant, and the RF frequency is made proportional to the amplitude of the information being sent. This type of modulation is used for voice for television, broadcast FM radio, and many other types of communication systems. The frequency spectrum for FM is more complex than that for AM. In the case of a simple single sine wave modulation signal, the FM voltage is given by

$$v = A \sin(\omega_c t + m_f \sin \omega_m t) \quad (7.2)$$

where:

A = magnitude

$\omega_c = 2\pi \times$ carrier frequency

$\omega_m = 2\pi \times$ modulating frequency

t = time

m_f = modulation index for FM = frequency deviation/modulating frequency

Note that we have the case of a sine of a sine. This requires Bessel functions of the first kind for solution. An expansion of (7.2) using Bessel functions is shown as (7.3).

$$\begin{aligned}
 n = A \{ & J_0(m_f) \sin \omega_c t \\
 & + J_1(m_f) [\sin(\omega_c + \omega_m)t - \sin(\omega_c - \omega_m)t] \\
 & + J_2(m_f) [\sin(\omega_c + 2\omega_m)t - \sin(\omega_c - 2\omega_m)t] \\
 & + J_3(m_f) [\sin(\omega_c + 3\omega_m)t - \sin(\omega_c - 3\omega_m)t] \dots \}
 \end{aligned}
 \tag{7.3}$$

It can be seen that the output consists of a carrier and an apparently infinite number of pairs of sidelobes, each preceded by a J coefficient. These are Bessel functions of the first kind and of the order denoted by the subscript, with the argument m_f . The spacing between spectral lines is the modulation frequency. Thus, if the modulation frequency is 3 kHz, the spacing between spectrum lines is 3 kHz. Fortunately, only a limited number of spectral lines are large enough in amplitude to be important. It is important to note that the carrier term is not always the largest; in fact, it can be zero amplitude.

Table 7.1 shows Bessel functions of the first kind as a function of the modulation index, m_f . The number of lines on either side of the carrier that are significant is shown by this plot as well as the magnitude and sign (phase) of the components.

For an example of the use of Table 7.1, assume a narrowband FM system such as used for many types of voice communication. The maximum deviation is assumed to be 5 kHz, and the modulation frequency is assumed to be 1 kHz. The FM modulation index is the deviation divided by the modulation frequency, so the modulation index for this example is 5.0. From Table 7.1, we see that there are six spectrum lines with significant amplitudes on either side of the carrier. Because each

Table 7.1 Bessel Functions of the First Kind

x	Order (n)									
	J_0	J_1	J_2	J_3	J_4	J_5	J_6	J_7	J_8	J_9
0.00	1.00									
0.25	0.98	0.12								
0.5	0.94	0.24	0.03							
1.0	0.77	0.44	0.03							
1.5	0.51	0.56	0.23	0.01						
2.0	0.22	0.58	0.35	0.13	0.03					
2.5	-0.05	0.50	0.45	0.22	0.07	0.02				
3.0	-0.26	0.34	0.49	0.31	0.13	0.04	0.01			
4.0	-0.40	-0.07	0.36	0.43	0.28	0.13	0.05	0.02		
5.0	-0.18	-0.33	0.05	0.36	0.39	0.26	0.13	0.05	0.02	
6.0	0.15	-0.28	-0.24	0.11	0.36	0.36	0.25	0.13	0.06	0.02
7.0	0.30	0.00	-0.30	-0.17	0.16	0.35	0.34	0.23	0.13	0.06
8.0	0.17	0.23	-0.11	-0.29	-0.10	0.19	0.34	0.32	0.22	0.13
9.0	-0.09	0.24	0.14	-0.18	-0.27	-0.06	0.20	0.33	0.30	0.21

line is spaced by the modulation frequency of 1 kHz, the required bandwidth to pass all spectrum lines shown is 12 kHz.

Some of the terms in Table 7.1 are negative and some are positive. A negative sign indicates a change of phase by 180 degrees. That is the case for the carrier and the first sidebands for the example in the preceding paragraph. The rule of thumb known as *Carson's rule* states that the approximate bandwidth required to pass an FM signal is twice the sum of the deviation and the highest modulating frequency. Using the previous example and Carson's rule, the required bandwidth to pass this FM wave is $2 \times (5 + 1) = 12$ kHz, the same answer that we obtained using the Bessel plot.

Figure 7.7 shows plots of the FM frequency spectrum for modulation index values of 0.5, 2.5, and 5.0 with amplitudes determined by Table 7.1 [1]. The modulation frequency is assumed constant at 1 kHz for those plots.

Wideband frequency modulation is used for FM broadcast and television. Maximum permissible deviation is 75 kHz for FM broadcast. The permissible deviation for the sound accompanying television transmissions is 25 kHz in the United States and 50 kHz in Europe. The modulation frequency range for FM broadcast is 30 Hz to 15 kHz. Thus, the maximum modulation index ranges from 5 to 2,500 in entertainment broadcasting.

Narrowband FM with modulation bandwidths of 3 kHz or less is used by the so-called FM mobile communication services, which include police, ambulance, taxi cabs, fire departments, radio-controlled appliance repair services, ground-to-air communication services, and short-range ship-to-shore services. The higher audio frequencies (above 3 kHz) are attenuated as they are in most long-distance telephone systems, but the resulting speech quality is still perfectly adequate. Maximum deviations of 5 to 10 kHz are permitted, and channel spacing is not much greater than for AM broadcasting (of the order of 15 to 30 kHz) [1].

Preemphasis and deemphasis are used with all FM transmissions. Noise has a greater effect on the higher modulating frequencies than on the lower ones. If the

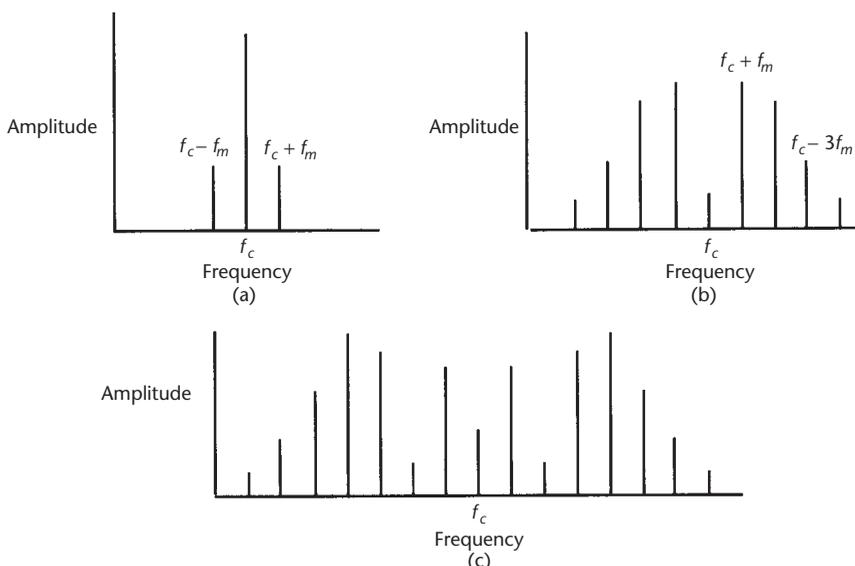


Figure 7.7 Example frequency spectra for FM: (a) spectrum for a modulation index of 0.5; (b) spectrum for a modulation index of 2.5; and (c) spectrum for a modulation index of 5.0.

higher audio frequency signals are artificially boosted in amplitude at the transmitter and correspondingly cut in amplitude at the receiver, an improvement in noise immunity can be expected [1]. This boosting in amplitude of the higher frequencies is called preemphasis. The corresponding reduction in amplitude of the higher frequencies at the receiver is called deemphasis.

Other FM spectra of interest are those for frequency sweep modulation (chirp modulation), as used for radar. They have complex multiple-line spectra with frequency components spaced by the pulse repetition frequency and a spectral width somewhat greater than two times the frequency deviation.

7.7 Modulation for Telemetry

Some of the following material is adapted from [2].

Telemetry, as the name suggests, consists of performing and reporting measurements on distant objects. With radio telemetry, there are always a number of channels of information that must be transmitted in parallel. Thus, either frequency division multiplexing (FDM) or time division multiplexing (TDM) is used. With FDM, a number of separate frequency channels are used in parallel to transmit the various channels of data. With TDM, the data alternates, using a single channel. First data channel 1 uses the single-frequency channel, then data channel 2 uses the frequency channel, then data channel 3 uses the frequency channel, and so on, until all data channels have been sampled. Then the process repeats.

In many telemetry systems, there is a need for both narrowband channels and relatively wideband channels. The mixture is achieved by a process known as subcommutating. Subcommutating consists of taking one of the wideband channels and subdividing it into several narrowband ones. For example, a system might use FDM in general and TDM for the subcommutated channels.

A large number of different types of telemetering systems have been used in the past. The FM/FM FDM system is used quite widely in the United States and is extended to pulse amplitude modulation (PAM)/FM/FM when subcommutating is required. It has the advantages of reliability and flexibility but requires greater bandwidth and carrier strength than the purely pulse systems. Systems such as pulse position modulation (PPM)/AM, pulse width modulation (PWM)/FM, PAM/FM, and pulse code modulation (PCM)/FM all have been used in certain applications, with varying advantages for each.

In PPM, the amplitude and the width of the pulse are kept constant, while the position of each pulse is varied with respect to a reference pulse. The position delay is proportional to the amplitude of the sampled signal. In PWM, the amplitude and the starting time for each pulse are fixed, but the width of each pulse is made proportional to the amplitude of the sampled signal. In PAM, the pulse width and the starting time are constant, and the amplitude of the pulse is made proportional to the amplitude of the sampled signal. In PCM, the total amplitude range that the signal may occupy is divided into a number of standard levels. The amplitude level of the sampled signal is transmitted in a binary code. For example, if 128 levels were used, it would take seven pulses of 1s or 0s to represent the amplitude of the sampled signal.

7.8 Combination Communication and Range-Measurement Systems

Phase-shift keying and high-bit-rate pseudorandom code modulation sometimes are used in systems in which both communication and range measurement are required. Examples are the Position Location and Reporting System (PLARS) and the Joint Tactical Information Distribution System (JTIDS) military communication systems. Very short duration, minimum phase-change intervals, and complex pseudorandom codes are used for modulation, producing a spread spectrum signal in which the energy is spread over very large bandwidths. The use of spread spectrum provides reduced vulnerability to jamming and interference by forcing a potential jammer to spread its energy over a large bandwidth. While the communication or radar system using the spread spectrum modulation also has its energy spread over a large bandwidth, it has the advantage that coherent integration is used to provide an effective narrow bandwidth.

Spread spectrum modulation using binary phase-shift keying (BPSK) or quaternary phase-shift keying (QPSK) with direct sequence pseudorandom code modulation also is used in some communication systems in which range measurement is not necessary. For those cases, a noise-like spectrum is produced that has antijam features and low observable signal capability.

As mentioned earlier in this chapter, simple phase-shift modulation that does not use wave shaping has a $[\sin x/x]^2$ spectrum with the first sidelobes down 13.5 dB and a rolloff rate of approximately 6 dB per octave (20 dB per decade) of frequency. The null-to-null main lobe frequency spectrum width is twice the clock frequency. Thus, with a code clock rate of 100 MHz, the null-to-null main lobe frequency spectrum width is 200 MHz. The -3 -dB spectral bandwidth is 0.88 times the code clock frequency. For a code clock rate of 100 MHz, the -3 -dB spectral bandwidth would be 88 MHz.

In the case of JTIDS, it was determined to be necessary to provide a rolloff rate faster than 6 dB per octave and lower sidelobes due to interference with other communication systems. The solution was to use a type of phase-shift keying called minimum shift keying (MSK). With this type of phase-shift keying, the phase is not changed abruptly at each phase change position in the code but is changed over a period of time. The result is a somewhat narrower main lobe (1.5 times code frequency for null-to-null main lobe width rather than 2.0 times code frequency) and a falloff rate of 12 dB per octave (40 dB per decade).

Spread spectrum phase-shift modulation may also find important applications for radar. It is one way to achieve high range resolution with long pulses, thereby providing large average power from a peak power-limited system.

7.9 Modulation for Radar

7.9.1 Pulsed CW Modulation

The most common waveform for radar is pulsed CW modulation. For example, a radar might use $1\text{-}\mu\text{s}$ duration pulses of RF at a pulse repetition frequency (PRF) of 1,000 pulses per second. The pulse repetition interval (PRI) would be 1,000 ms. The number of RF cycles in the $1\text{-}\mu\text{s}$ pulse would depend on the radar RF frequency. If,

for example, the radar operated at a frequency of 5 GHz, there would be 5,000 cycles of RF in the 1- μ s pulsed waveform. Typical radar pulse power output is in the range of 100 kW to 5 MW.

Figure 7.1 showed the waveform and frequency spectrum for a pulsed CW signal. The spectra are $(\sin x)/x$ -type voltage spectra or $[(\sin x)/x]^2$ power spectra, with the maximum main spectrum lobe width at the nulls equal to two times the reciprocal of the minimum RF pulse width. For example, if the RF pulse width is 1.0 μ s, the width of the main spectral lobe at the nulls would be 2.0 ms. The sidelobes each have half that width. The first sidelobe is down in amplitude by about 13.5 dB from the main spectrum lobe. Higher-order sidelobes are progressively lower, falling off at a rate of about 20 dB per decade. The second sidelobe is down by 17.9 dB, the third side lobe is down by 20.8 dB, and the fourth side lobe is down by 23.0 dB.

Spectral lines are separated by the PRF. Thus, if the PRR is 1,000 Hz, there would be spectral lines every 1,000 Hz. The required bandwidth for this type of modulation is approximately the reciprocal of the pulse width.

For most radar and communication system applications, it is not possible or desirable to use a rectangular pulse with zero rise time and zero fall time. The more common approach is to provide a nearly trapezoidal pulse. Such a pulse has the advantage of reducing interference to other systems by reducing the energy that is transmitted in far-out sidelobes.

Knowledge of the exact spectrum for the trapezoidal pulse usually is not required. What is important is to know the envelope of the spectrum so that the maximum level of sidelobe signals can be determined as a function of frequency. This envelope can be determined using (7.4) and (7.5) [3].

$$F_1 = 1/\pi[\tau + (D_1 + D_2)/2] \quad (7.4)$$

where:

F_1 = first frequency break point (megahertz)

D_1 = rise time of trapezoidal pulse (microseconds)

D_2 = fall time of trapezoidal pulse (microseconds)

τ = -3 dB pulse width (microseconds)

The equation for F_2 is as follows:

$$F_2 = (1/D_1 + 1/D_2)/2\pi \quad (7.5)$$

where:

F_2 = second frequency break point (megahertz)

D_1 = rise time of trapezoidal pulse (microseconds)

D_2 = fall time of trapezoidal pulse (microseconds)

The predicted envelope of the frequency spectral has a 0-dB per frequency decade slope from the spectrum center to F_1 . The envelope then falls off at a rate of 20 dB per decade (6 dB per octave) between F_1 and F_2 . At frequencies greater than F_2 , the envelope amplitude falls off at a rate of 40 dB per decade, or 12 dB per octave. The three examples in Figure 7.8 show the use of (7.4) and (7.5) in predicting the

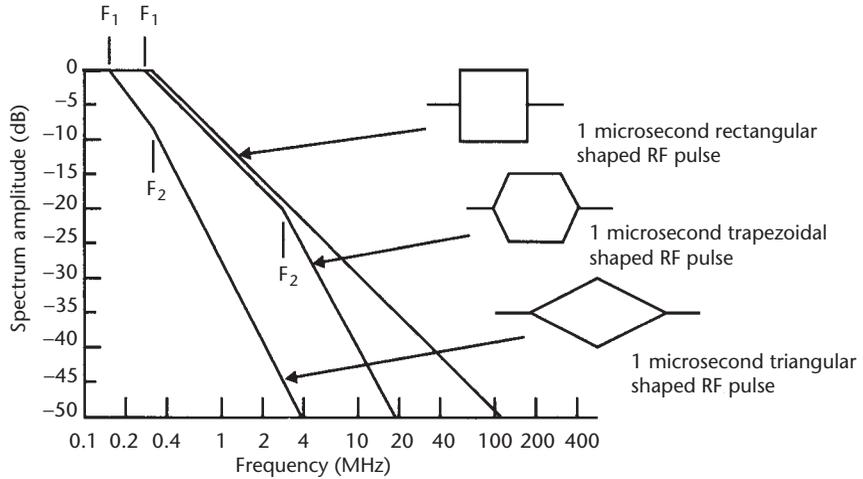


Figure 7.8 Examples of spectrum envelopes for RF pulses.

envelope of the frequency spectrum for pulse radar. The first example is the rectangular pulse with zero rise and fall times. For the case of a $1\text{-}\mu\text{s}$ pulse width, the value of F_1 is computed to be 0.318 MHz. The value of F_2 is infinity. Between those values, the slope of the envelope is 20 dB per decade. For this example, the spectrum is down only 30 dB at 10 MHz and 50 dB at 100 MHz on either side of the main spectral lobe.

In the second example in Figure 7.8, the same $1\text{-}\mu\text{s}$ half-power pulse width is assumed, but the rise and fall times are each $0.1\ \mu\text{s}$. For this case, F_1 is computed to be 0.289 MHz, and F_2 is 3.183 MHz. At frequencies greater than F_2 , the falloff rate for the spectrum envelope is 40 dB per decade. The spectral envelope has dropped to about -65 dB at 40 MHz offset from the carrier or spectral center. In the third example in Figure 7.8, the same $1\text{-}\mu\text{s}$ pulse is assumed, but this time the rise and fall times are each $1\ \mu\text{s}$. We thus have a triangular-shaped pulse. For this case, the value of F_1 is calculated to be 0.159 MHz, and the value of F_2 is 0.318 MHz. We see from the plot that the spectrum envelope has dropped to -50 dB at 4.0 MHz from the spectral center.

7.9.2 High-Power Impulse Generators and Ultra-Wideband, High-Power Microwave Generators

Recent developments in high-power impulse generators and ultra-wideband, high-power microwave (HPM) generators have made possible some new types of pulsed CW radars with very short duration pulses and wide frequency spectra. In the case of impulse type generators, it is possible to generate pulses with peak powers of more than 10^{10}W with pulse widths of less than 1 ns (10^{-9} seconds). The spectrum for the generated pulse that feeds the antenna has its peak at 0 frequency and frequency lines that extend to many gigahertz. If we assume a 0.5-ns pulse width, the first null in the spectrum is at 2 GHz, the second null is at 4 GHz, the third null is at 6 GHz, and so on.

Very wideband antennas are used with video pulse systems. A TEM horn antenna is one of the better antennas for such systems. A typical lower frequency for such a system is 50 MHz. The result is that the antenna converts the pulse to a single-cycle RF pulse with some ringing. The result is a modified, but very wide, frequency spectrum. The HPM systems also generate very high powers (10^9 W is typical). An RF pulse is generated at microwave frequencies, which may contain only a few RF cycles and may have pulse durations of only a few nanoseconds. Again very wide frequency spectra are generated.

It is expected that there may be many applications for very short duration, extremely high power pulse systems. A chief advantage for such systems, in the case of radar, is very high range resolution. These very high-power, short-pulse systems also may have important applications as impulse jammers. They also may be used for “zapping” electronic components via electromagnetic pulse (EMP).

7.9.3 Chirp Pulse Modulation

Another type of modulation often used when high range resolution is needed is known as chirp. The RF frequency is swept in a linear fashion during the time that the pulse is present.

As an example of a chirp pulse modulation, a 10- μ s pulsed RF signal might be used with the frequency changed from 5.0 GHz to 5.5 GHz during the 10- μ s period. That produces a wideband signal with a frequency spectral width in excess of 0.5 GHz. The received chirp signal may be processed by a pulse compression filter, yielding a very short duration output pulse. That type of pulse has the advantage of providing high range resolution with signal processing gain.

7.9.4 Phase Code Modulated Pulse Modulation

A third type of pulsed RF waveform that can be used for improved resolution is a phase code modulated pulse of RF. This type of waveform is shown in Figure 7.9. A biphasic modulated system would use two phase states, 0- and 180-degree phase shift. Phase change intervals could be as short as a few nanoseconds or could be much longer, depending on the code used and the duration of the pulsed RF signal. When the reflected signal is received, it can be processed by a phase detector and matched filter to produce a very short duration output pulse.

The type of code illustrated in Figure 7.9(a) is a seven-unit Barker code, only one of many possible codes that can be used. It has the advantage that the matched filter detected output will have all sidelobes at unity level in one direction, and the main-lobe seven units high in the opposite direction. The detected output is illustrated in Figure 7.9(b). When received and processed by the matched filter, the main output pulse will be one-seventh the width of the input pulse.

Other longer codes usually are used for larger pulse compression ratios and larger integration improvement. The longest Barker code is a 13-unit code. The peak-sidelobe ratio in that case is $-20 \log N = -22.3$ dB. Other types of codes are much longer than Barker codes and can be used if larger compression ratios are desired.

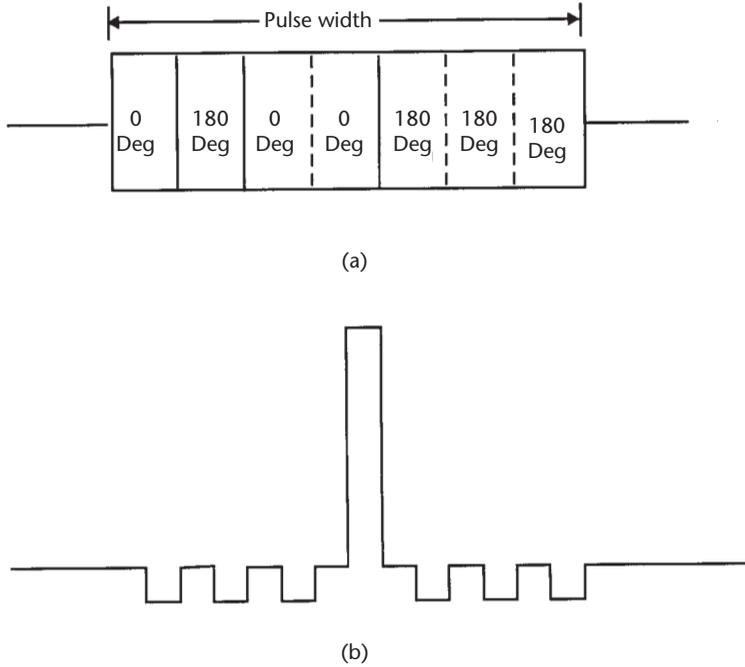


Figure 7.9 Example of phase code modulated pulse for radar: (a) details of modulated pulse and (b) matched filter response following detection.

7.9.5 Continuous-Wave Modulation

Radars also may use CW modulation, in which case the amplitude, frequency, and phase are held constant. This type of modulation does not permit range measurements, but it does permit angular measurements and measurements of radial velocity.

7.9.6 Frequency-Modulated CW Modulation

Another type of waveform used for radar is FMCW modulation. The amplitude of the wave is held constant and the frequency is swept from a low frequency to a high frequency within a band in a sawtooth fashion. FMCW modulation permits measurements of range as well as measurements of angle and radial velocity.

Figure 7.10 illustrates concepts for FMCW radar. The range measurement is proportional to the frequency difference between the signal being transmitted and the received signal reflected from the target.

7.10 Single-Channel Transmitter System

Figure 7.11 shows a simplified system block diagram for a single-channel transmitter system. The input system is assumed to be a microphone for voice. The output of the input device is at baseband frequencies and is assumed to be bandpass or lowpass filtered to a frequency band of 300 to 3,300 Hz. The frequency band is consistent with the bandwidth of a telephone channel. The baseband signal is fed as one

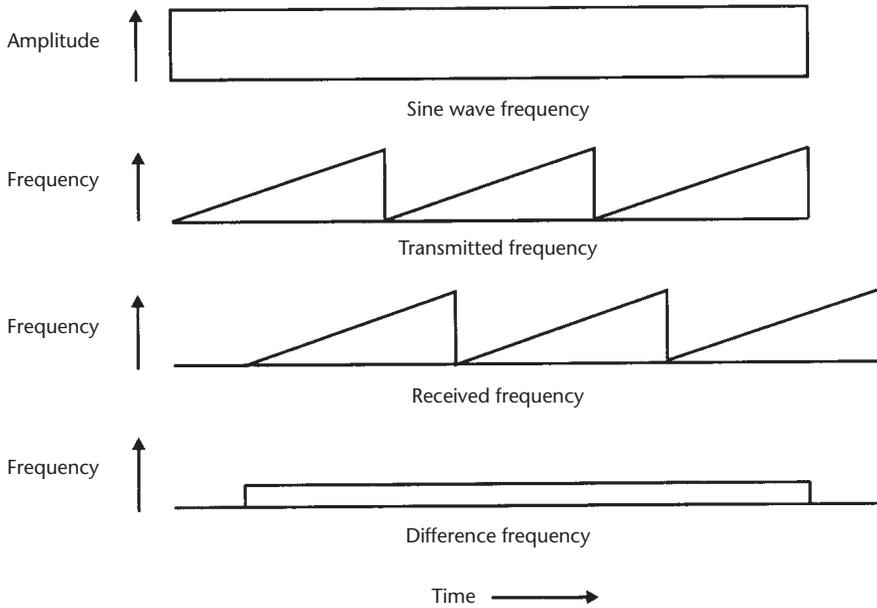


Figure 7.10 Example of FMCW modulation for radar.

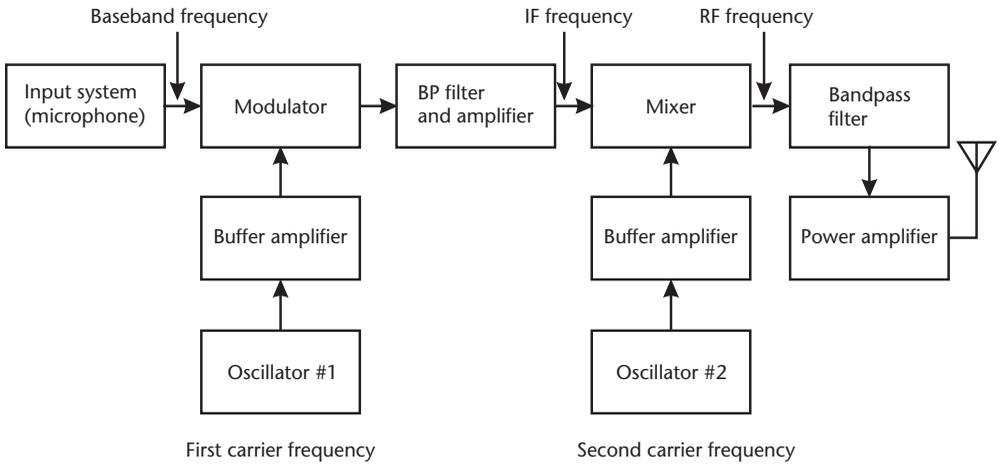


Figure 7.11 Single-channel transmitter system block diagram.

input to a modulator. The second input is a local oscillator operating at an IF frequency, for example, 30 MHz. The carrier is modulated by the baseband signal. The modulation type may be amplitude, frequency, or phase modulation.

The output of the modulator is first amplified and filtered, then fed as one input to a mixer circuit that acts as a frequency upconverter. The second input to the mixer is the output of an RF oscillator, commonly referred to as the LO. A buffer amplifier is used between the oscillator and the mixer for reverse isolation and frequency pulling considerations.

The output of the mixer is an RF signal with the desired carrier frequency being the sum of the two oscillator frequencies. For example, if the frequency of the oscillator that feeds the modulator is 30 MHz and the frequency of the second oscillator that feeds the mixer is 120 MHz, the bandpass-filtered mixer output will be 150 MHz.

There also will be undesired outputs from the mixer, including the difference frequency, which for this case would be 120 MHz minus 30 MHz or 90 MHz, the two oscillator frequencies, harmonics of the oscillator frequencies, and spurs or intermodulation products. The spurs or intermodulation products can be the result of mixing of harmonics with fundamental signals and can include both sum and difference frequencies. The mixer type is chosen to eliminate, to the extent possible, the two oscillator signals by using balanced mixers. The bandpass filter also is designed to suppress harmonics, fundamental frequencies, difference frequencies, and spurs.

The modulated RF signal is then fed to a power amplifier chain, where the power level is increased to the desired transmitter output power level. The output power may be 100W or more. The output of the power amplifier feeds the transmit antenna.

7.11 Frequency Division Multiplex Transmitter System

Some of the material presented in this section and in Sections 7.12 and 7.13 is quoted or adapted from [4].

In communication systems such as microwave relay systems, satellite relay systems, and undersea cable systems, the requirement is that many hundreds of telephone channels or other signals be transmitted on a single communication link. One way to accomplish that task is to use FDM. For such types of systems, many subchannels are used in parallel in the main allocated channel. In the past, analog modulation has been used for each of those subchannels. Modulation types are often SSB suppressed carrier AM or narrowband frequency modulation FM.

Figure 7.12 shows a simplified system block diagram for a FDM transmitter. The figure assumes that the inputs are telephone channels. Each channel is fed to a modulator, which has as its second input a CW signal from an LO. Each oscillator has a different frequency to provide the necessary separation of frequency channels. A typical group of modulators would have about 20 channels and 20 different crystal oscillators. Crystal oscillators are used rather than LC oscillators because of greatly improved frequency accuracy and stability. The oscillator frequencies typically may be separated by only about 10 kHz.

The outputs of each of the modulators in the group plus a pilot channel oscillator signal are added using an RF add circuit. The combined signal feeds a mixer circuit acting as a frequency upconverter.

The next step is to combine the outputs from the different groups of modulators. The signal from each group is different, since the mixers associated with each group use different oscillator frequencies. The combined signal is then frequency converted by another mixer to the final microwave frequency band used by the system. A power amplifier chain is used to increase the power level of the signal to the desired transmitter output level. The power amplifier connects to the transmit antenna.

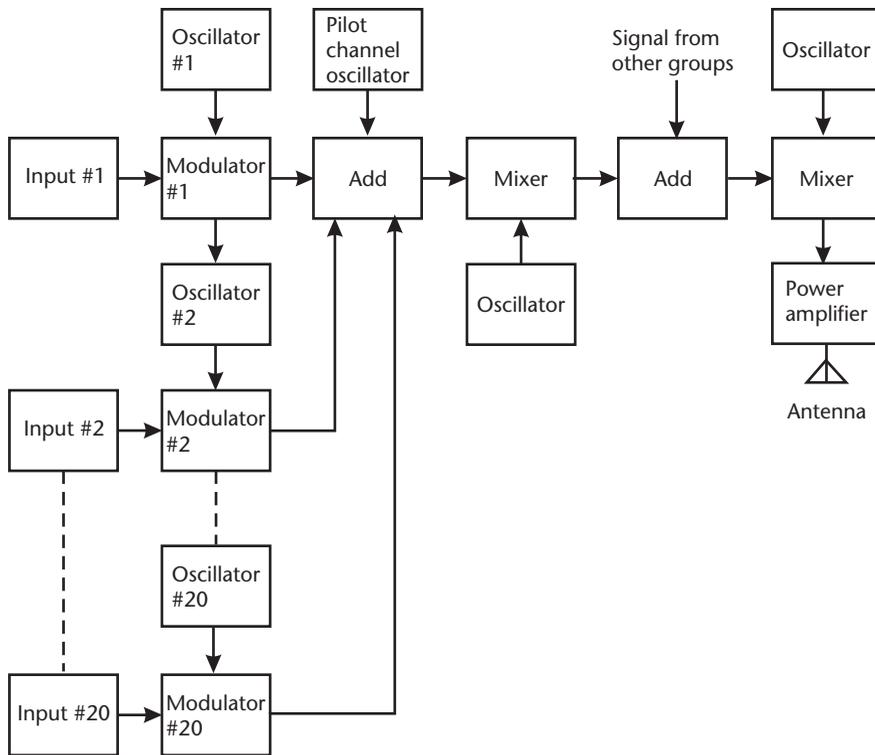


Figure 7.12 Simplified FDM system block diagram.

7.12 Sample Circuits and Analog-to-Digital Converter Concepts

Figure 7.13 shows an analog signal being sampled by a gate circuit. The output of the gate is connected to an A-to-D converter system. The sampling rate must be more than twice the highest frequency component in the analog signal and is termed the Nyquist sampling rate. Higher sampling rates are often used. The measured amplitude is expressed as a binary code with the number of bits required depending on the necessary resolution. For example, if 128 amplitude levels are used, the required number of bits per sample is 7, that is, $2^7 = 128$. The more bits used, the more accurate will be the representation of the analog signal. Using more bits, however, means more bandwidth is required and therefore lower signal-to-noise ratio.

The binary bits next are fed to a modulator. The modulator is frequently a phase modulator in which the carrier frequency is supplied by an IF signal oscillator, and the phase of the signal with respect to the oscillator signal is determined by the digital signal.

7.13 Time Division Multiplex Transmitter System with Pulse Code Modulation

Figure 7.14 shows key elements and concepts for a TDM transmitter system with PCM. In communication systems such as microwave relay systems, the requirement is that many hundreds of telephone channels or other signals be transmitted on a

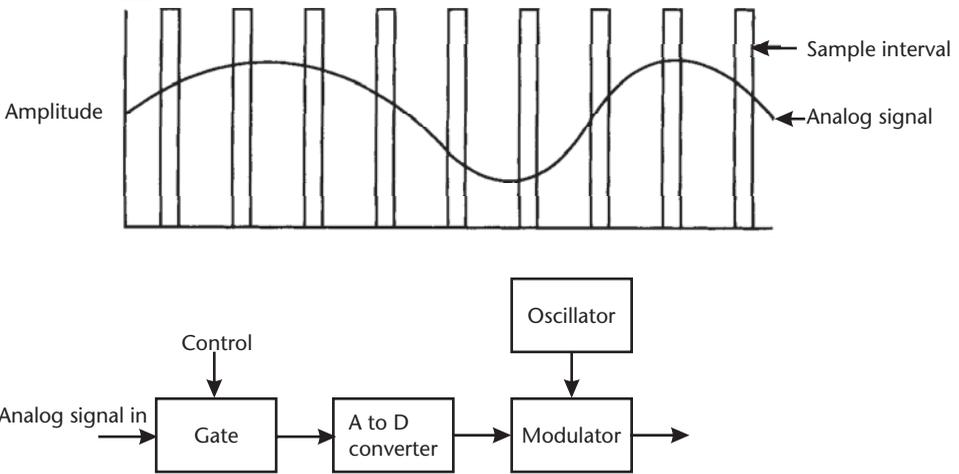


Figure 7.13 Sample and A-to-D converter concepts.

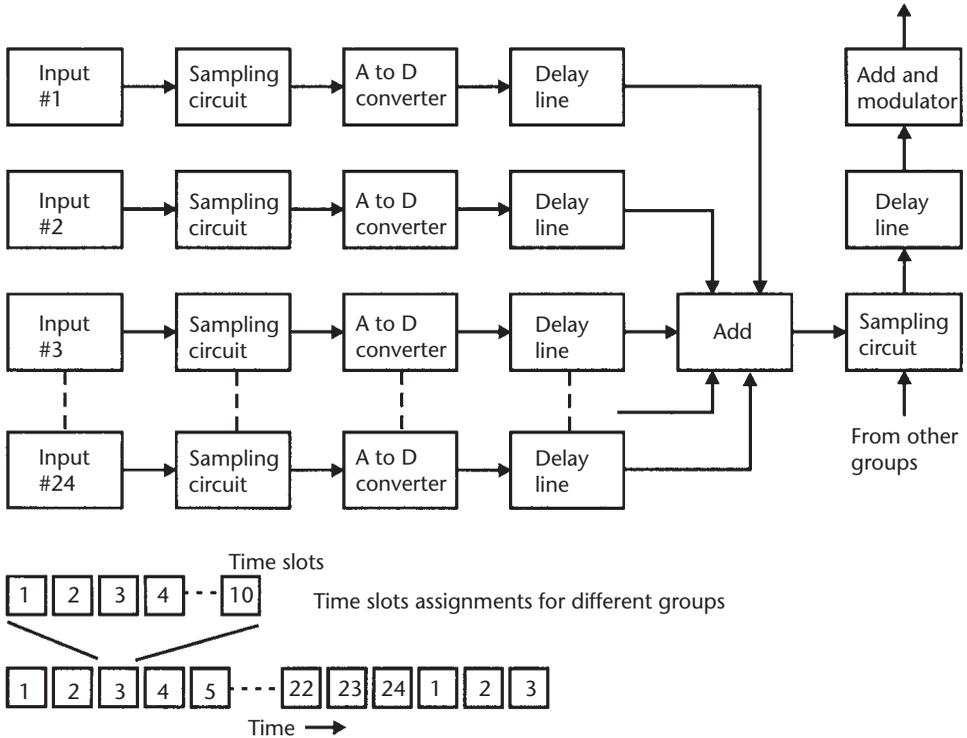


Figure 7.14 TDM system with pulse code modulation.

single communication link. A second way to do that, in addition to FDM is to use TDM. This approach has been found to be superior to FDM. Advantages include reduced cost, improved reliability, and improved signal-to-noise ratio. For that reason, many FDM systems used for long-distance telephone service recently have been replaced by TDM systems.

With TDM, a single wideband channel is used for the many hundreds of input channels by assigning time slots to each of the input channels of data. In essence, the system takes turns using the channel. This concept of time sharing the channel is illustrated at the bottom of Figure 7.14, which shows 24 major time slots and 10 time slots in each major time slot. Thus, there are a total of 240 time slots for digital data. The repetition rate for major time slots must be greater than twice the highest frequency component in the signals that are being sampled and A-to-D converted. Thus, if we are dealing with telephone signals with 3.3 kHz as the highest frequency component, the time-slot repetition rate must be greater than 6.6 kHz.

A simplified system block diagram for a TDM transmitter of this type is shown at the top of Figure 7.14. Inputs are fed to gate circuits to select the times for viewing the signals. The gates are followed by A-to-D converters and delay lines for spacing the digital data in time. The 24 delay line outputs are combined using an add circuit. A sampling circuit and set of delay lines follow that add circuit and are used to place digital data pulses from the different groups in their proper time slots in each major time slot. For simplicity, Figure 7.14 does not show the other elements of the system that follow the modulator. Those elements include a mixer and associated oscillator, a power amplifier chain, and the transmitter antenna.

7.14 Two-State Modulation Types for Binary Signals

A number of two-state modulation types can be used for binary signals. Three such systems are illustrated in Figure 7.15: on-off amplitude keying, frequency shift keying, and two-state phase-shift keying.

7.14.1 On-Off or Two-State Amplitude Keying

One possible two-state modulation type for binary signals is on-off or two-state amplitude keying, one of the earliest types of modulation used. It has been used in

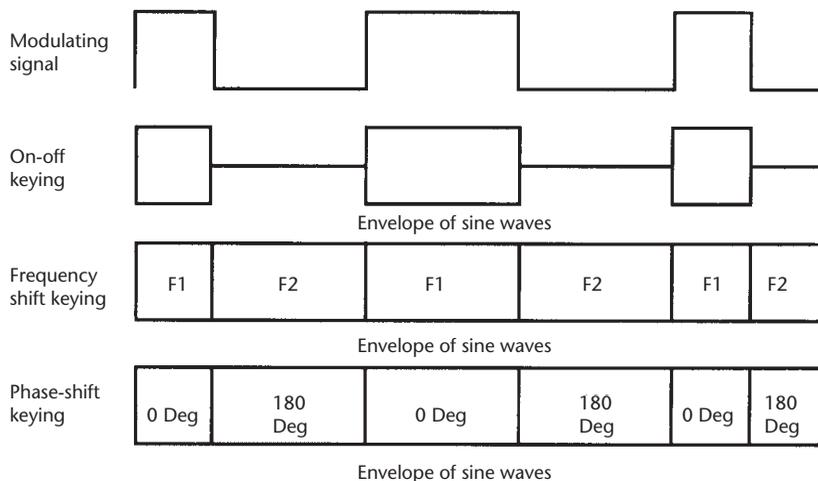


Figure 7.15 Two-state modulation types for binary signals.

the past for such communication modes as telegraphy and teletype or telex. In modern systems, two-state and four-state amplitude modulation can be used in combination with pulse phase modulation to increase the number of bits per baud or bits per pulse that are sent.

7.14.2 Frequency Shift Keying

Frequency shift keying (FSK) is a type of frequency modulation in which only two distinct frequency states are used. The amplitude of the RF waveform is constant. In practice, the frequency shift is only a few kilohertz [1]. FSK is used for transmission of many types of digital data. One important application is for telegraphy and telex, two forms of communication that employ typewriter-like machines operating at a maximum speed of about 60 words per minute to send written messages from one point to another [1].

7.14.3 Binary Phase-Shift Keying

BPSK is another form of two-state modulation. Both the frequency and the amplitude remain constant, and only the phase of the RF waveform is changed. The two states are 180 degrees apart. The reference normally is a signal derived from an oscillator and is sometimes referred to as coherent-phase signal. A similar reference is used in both the transmitter and the receiver.

It is also possible to send BPSK as a differential phase-shift signal. The information sent is determined by the phase change between successive pulses. If there is no change in phase, the signal may be called a *zero*. If there is a 180-degree change in phase, the signal may be called a *one*. With differential phase-shift modulation, the phase reference at the receiver is the signal itself delayed by one pulse period. This delay may be achieved by either analog delay lines or digital delay lines.

The coherent method of BPSK is known to be about 1 dB better in terms of signal-to-noise ratio than that of the differential method of BPSK. The advantage of using a coherent reference approaches 3 dB for multilevel modulation. Accordingly, coherent detection is preferred in those applications in which small losses in signal-to-noise ratio are significant, as in the case of the downlink for satellite communication [3].

In practice, we seldom have a true noise-free coherent reference. Usually, the phase reference uses a phase-locked loop (PLL) oscillator that has some jitter with respect to the transmitter reference. The result is that only partially coherent reception actually can be claimed [3].

In some cases, the differential-phase system has advantages over the coherent system. One case is when there are multipath induced phase changes during the signal pulse period. In that case, having a fixed phase-reference is not as good as having a reference that has been effected by the same multipath phase changes as the signal being detected.

7.15 Four-State and Eight-State Phase-Shift Keying

It is also possible to use a four-state modulation with the four states being 90 degrees apart. An example would be $+0$, $+90$, ± 180 , and -90 degrees. This type of modulation provides two bits of information for each pulse sent. The performance of this type of modulation is the same as that of BPSK for thermal noise, but it degrades more rapidly with CW interference and linear delay distortion [5]. Four-state phase modulation is illustrated in Figure 7.16.

Both coherent and differential QPSK can be produced with the 00 signal represented by a change of $+0$ degree from the phase reference, 01 can be represented by a change of $+90$ degrees, 10 can be represented by a change of -180 degrees, and 11 can be represented by a change of -90 degrees.

Another possible phase-shift keying system is the eight-state phase modulation or octonary phase-shift keying (OPSK). In OPSK, phase-shift steps are separated by 45 degrees. Example states are $+45$, $+90$, $+135$, $+157.5$, ± 180 , -135 , -90 , -45 , and 0 degrees. With OPSK, three bits of information are sent per pulse. This mode of modulation is used with a number of digital microwave radio systems operating in the 6- and 11-GHz bands.

7.16 Sixteen Phase-State Keying (16-PSK)

In 16 phase-state keying (16-PSK), 16 distinct phase states are used. Each phase step is separated by 22.5 degrees. With 16-PSK four bits of information are sent per pulse. No 16-PSK systems are in commercial use at this time because of the superior performance of the 16 amplitude-phase keying (16-APK) modulation method.

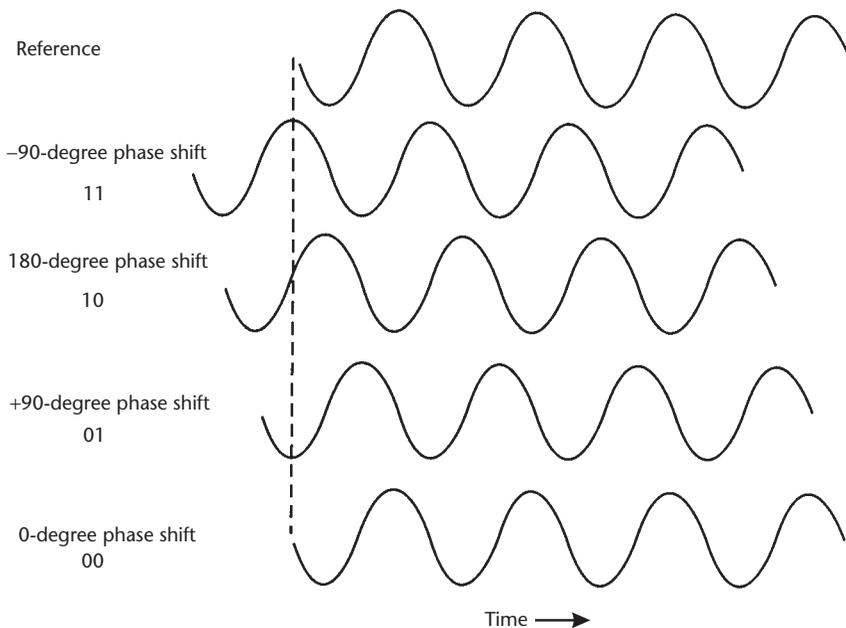


Figure 7.16 Four-state phase modulation for sending two bits per pulse.

7.17 Sixteen Amplitude-Phase Keying

Figure 7.17 shows the 16 possible state combinations for a modulation system that uses four possible phase states and four possible amplitude states per pulse. This type of modulation permits four bits per pulse to be sent. This type of modulation has been found to be a reliable method of communication for high-capacity TDM systems.

At the top of Figure 7.17, we show a reference and two of the 16 possible phase and amplitude states. These are -45 degrees phase shift with level 1 amplitude and -135 degrees phase shift with level 2 amplitude. Each of the 16 possible combinations of phase shift and amplitude is shown at the bottom of Figure 7.17 in a state table.

7.18 Direct Sequence Spread Spectrum Modulation

7.18.1 Pseudorandom Noise (PN) Generators

This section presents a brief discussion of pseudorandom noise generators. This type of noise generator is used in spread spectrum systems, discussed in detail in Section 7.18.2. Much of the information about pseudorandom noise generators presented in this section is quoted or adapted from [7].

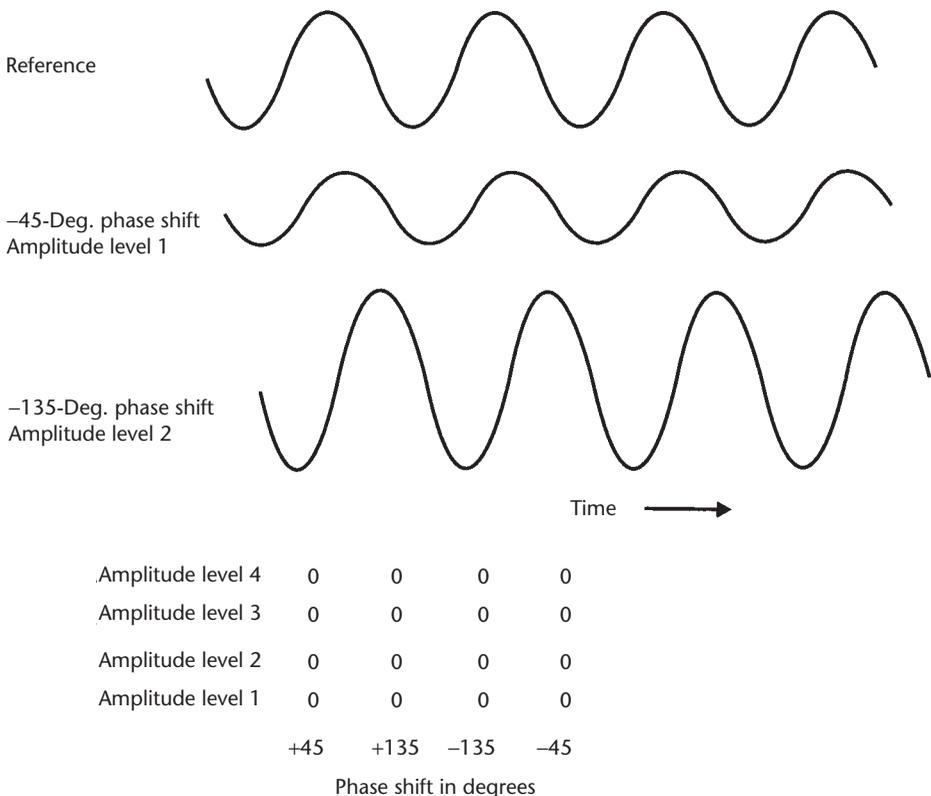


Figure 7.17 Phase angle and amplitude level combinations for 16-APK.

The type of generator discussed here is a digital code generator that produces outputs of ones and zeros in a noise-like fashion. The codes appear to be random in nature, but they are not. They have a controlled pattern and a finite repetition length, thus the name pseudorandom noise generators rather than random noise generators. Some pseudorandom codes are short with repetition times of only a small fraction of a second, while others have repetition times of many weeks, as in the case of one of the codes used in GPS.

The PN code generators are generally made up of shift registers with feedback circuits and modulo-2 adders. In maximal sequence binary shift registers, the maximum length sequence that can be generated is $2^n - 1$ chips, where n is the number of stages in the shift register. Feedback connections have been tabulated for maximal code generators from 3 to 100 stages, so that some sequences of any length from 7 through $2^{100} - 1$ chip are readily available.

Figure 7.18 illustrates the general form of a simple linear sequence generator. Outputs from the last delay stage and an intermediate stage are combined in a modulo-2 adder and fed back to the input of the first delay element. In this case the code sequence generated, 1110010, is cyclic (repetitive) with a total period seven times the period of a single delay element.

A generic equivalent and much preferable sequence generator configuration places the feedback adders between stages as in Figures 7.19 and 7.20.

Often a large number of code sequences are needed for applications such as code-division multiplexing. Under these conditions multiple feedback points are

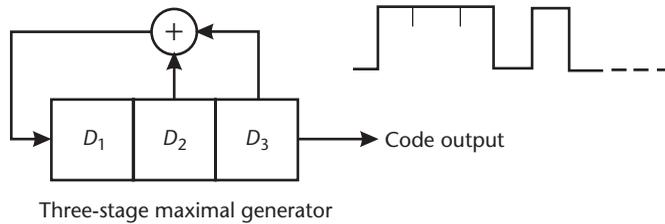


Figure 7.18 Typical simple code sequence generator.

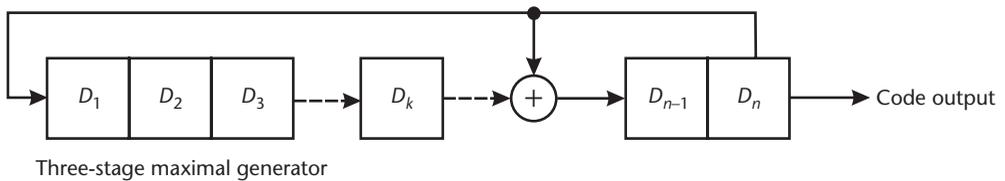


Figure 7.19 Alternate generic sequence generator configuration (modular type).

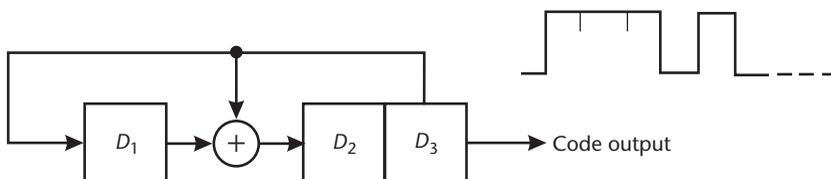


Figure 7.20 Code sequence generator equivalent to that of Figure 7.18.

necessary. A system of this type is shown in Figure 7.21. Shift register sequence generators made up in this way are called *modular shift register generators* (MSRGs).

One particularly applicable integrated circuit capable of acting as an MSRG is the MC8504 developed by Motorola. This IC includes four D flip-flops and four modulo-2 adders, plus a gate that can be used to sense the non-allowable all zeros condition. For longer generators, the output of one MC8504 can be cascaded as an input to the next, and a code generator that operates to 17 Mcps can be readily constructed. Figure 7.22 is a block diagram of the MC8504 [8].

There are many other types of code sequence generators but the foregoing examples should be sufficient to show the general concepts involved. For more information in this area, the author recommends that the reader consult [7].

7.18.2 Direct Sequence Spread Spectrum Systems

A spread spectrum modulation is one in which the transmitted signal is spread over a wide frequency band, much wider in fact than the minimum bandwidth required to transmit the information being sent. A direct sequence spread spectrum system is one in which a carrier is modulated by a digital code sequence whose bit rate is much higher than the information signal bandwidth. The information being sent may be added to the carrier signal but is usually imbedded in the spread spectrum signal by adding the information to the spectrum spreading code before its use for spreading modulation. The information being sent must be in digital form because addition to a code sequence involves modulo-2 addition to a binary code.

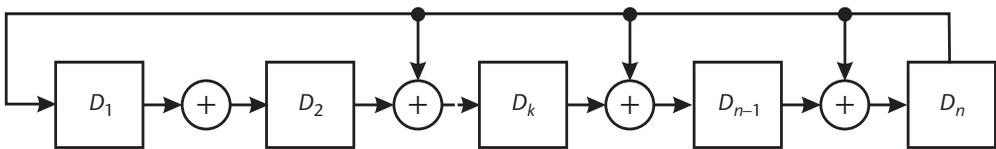


Figure 7.21 Modular multiple-tap sequence generator (MSRG).

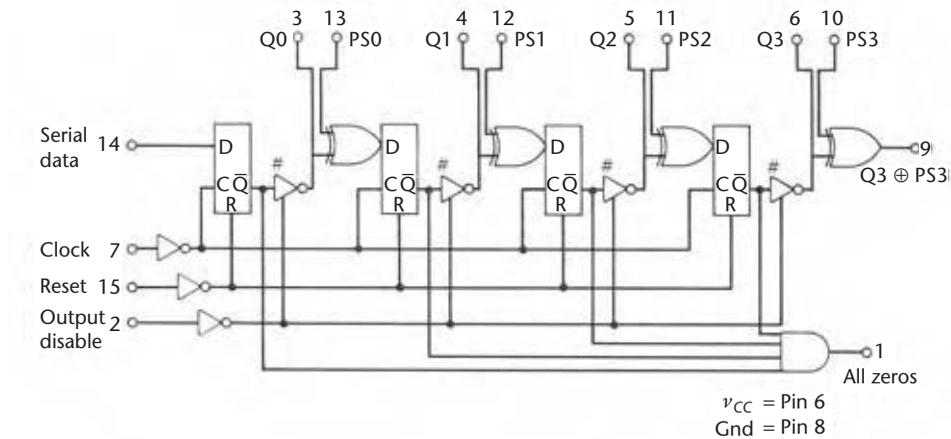


Figure 7.22 Block diagram of an MC8504 integrated circuit.

The most common way to modulate the carrier in the case of direct sequence spread spectrum modulation is to use two-state or biphasic phase shift keying. The resulting power spectrum is a $(\sin x/x)^2$ spectrum in which the main lobe null-to-null bandwidth is twice the clock rate of the code sequence used as a modulating signal. Each of the sidelobes has a null-to-null bandwidth equal to the clock rate. For example, if the modulating waveform has a 5-Mbps operating rate, the main lobe bandwidth (null-to-null) will be 10 MHz, and each sidelobe will have a null-to-null bandwidth of 5 MHz.

Typically, the direct sequence biphasic modulator has the form shown in Figure 7.23. A balanced mixer, discussed in Section 8.4, has the carrier signal as one input and the code sequence input as the other input. The output is the biphasic modulated output signal. Figure 7.23 shows the frequency spectrum for each of the inputs as well as the frequency spectrum for the output signal.

Figure 7.24 shows a simplified block diagram for a direct sequence communication link (transmitter and receiver) showing waveforms. The code sequence in each case is a pseudorandom noise generator. This type of code generator was discussed in Section 7.18.1.

In practice, the carrier is not usually modulated by baseband information. The baseband information is digitized and added to the code sequence. For this discussion, however, we will assume that the RF carrier has been modulated before code modulation because this somewhat simplifies discussion of the modulation-demodulation process.

After being amplified a received signal is multiplied by a reference with the same code, and assuming that the transmitter's code and the receiver's code are synchronized, the carrier inversions transmitted are removed and the original carrier restored. This narrowband restored carrier can then flow through a bandpass filter designed to pass only the baseband modulated carrier. Undesired signals are also treated in the same process of multiplication by the receiver's reference that maps the received direct sequence signal into the original carrier bandwidth. Any incoming signal not synchronous with the receiver's coded reference (a wideband signal) is

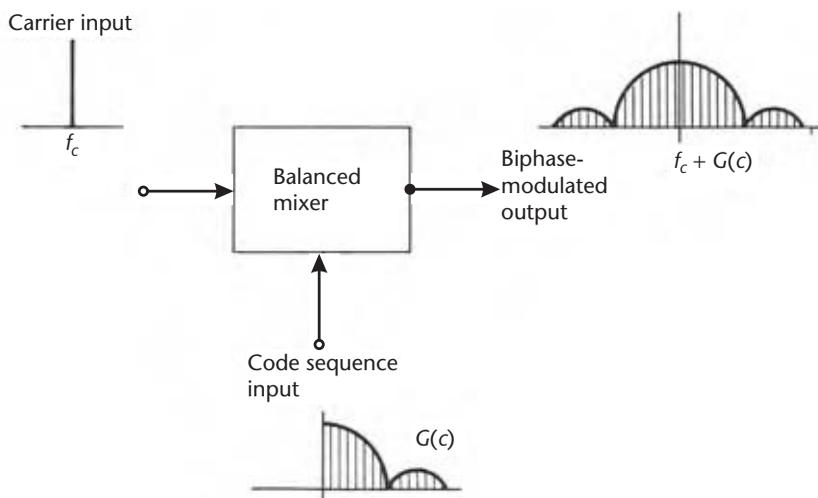


Figure 7.23 Direct sequence biphasic modulator.

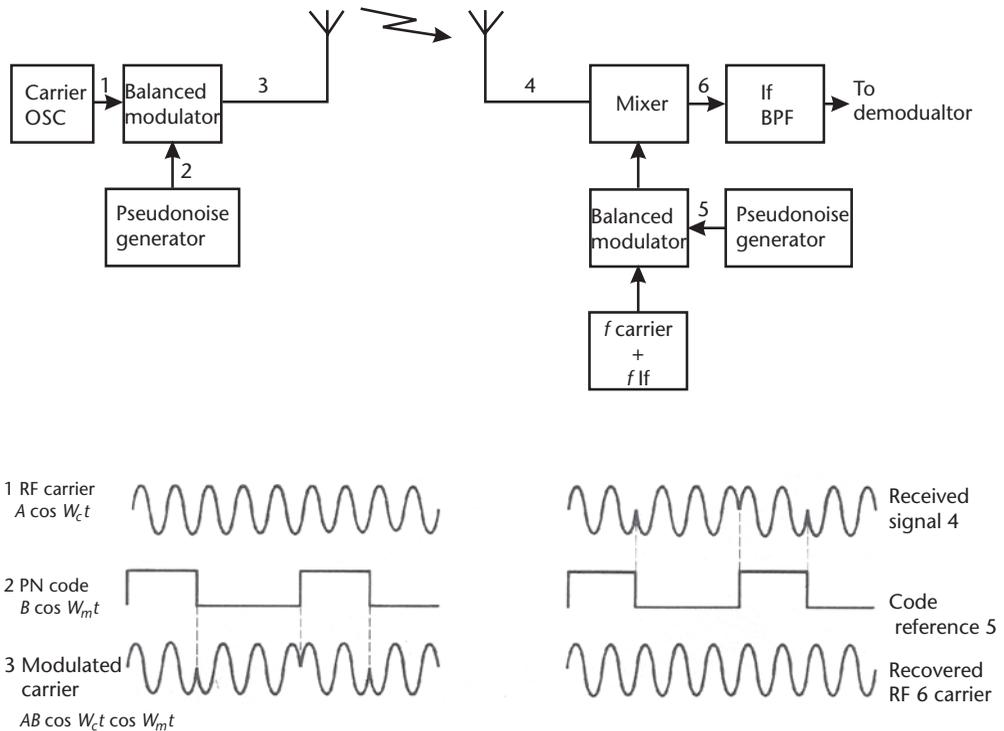


Figure 7.24 Overall direct sequence system showing waveforms.

spread to a bandwidth equal to its own bandwidth plus the bandwidth of the reference.

Because an unsynchronized input signal is mapped into a bandwidth at least as wide as the receiver’s reference, the bandpass filter can reject almost all the power of an undesired signal. This is the mechanism by which process gain is realized in a direct sequence system. The receiver transforms synchronous input signals from code-modulated bandwidth to the baseband-modulated bandwidth. At the same time nonsynchronous input signals are spread at least over the code-modulated bandwidth.

The spread spectrum processing gain (G_p) available may be estimated by the following equation:

$$G_p = BW_{RF} / R_{info} \tag{7.6}$$

where the RF bandwidth (BW_{RF}) is the bandwidth of the transmitted spread spectrum signal and the information rate (R_{info}) is the data rate in the information baseband channel. For an example of the use of (7.6), assume the RF bandwidth of the spread spectrum signal is 100 MHz and the data rate in the baseband channel is 10,000 bits per second; the processing gain is 10,000, or 40 dB.

The jamming margin in decibels is given by (7.7)

$$\text{Jamming margin} = G_p - L_{sys} - (S/N)_{out} \tag{7.7}$$

where:

G_p = processing gain in decibels

L_{sys} = system implementation losses in decibels

$(S/N)_{out}$ = signal-to-noise ratio at the information output in decibels

For example, a system with 30-dB processing gain, $(S/N)_{out}$ of 10 dB, and L_{sys} of 2 dB would have an 18-dB jamming margin.

The most difficult problem with spread spectrum systems is synchronization of the transmitter code generator and the receiver code generator. The simplest of all techniques for synchronization of these two code generators is to use a so-called *sliding correlator*. For this approach, the receiving system in searching for synchronization operates its code generator at a rate different from the transmitter's code generator. The effect is that the two code sequences slip in phase with respect to each other and slide past each other, stopping only when the point of coincidence is reached. The flow diagram of Figure 7.25 illustrates this process.

In practice, sliding correlation is almost always used, though the technique is often augmented by other methods for restricting the area of search.

One of the most effective techniques for making use of a sliding correlator employs special code sequences, short enough to allow a search through all possible code positions in some reasonable time. A well-chosen code sequence (called a *preamble* when used for synchronization) is a good solution to almost all synchronization problems. Typical synchronization preambles range in length from several hundred to several thousand chips depending on the specific system's requirements. Much more detail about spread spectrums and synchronization methods can be found in [7].

7.18.3 Applications for Direct Sequence Spread Spectrum

There are many applications for direct sequence spread spectrum. The antijam capability provided is very important for military communication systems. The reduced interference capability is used in the best of cellular telephone systems and the best of cordless telephone systems. The superior ranging capability provided by spread spectrum makes possible GPS navigation systems and other types of

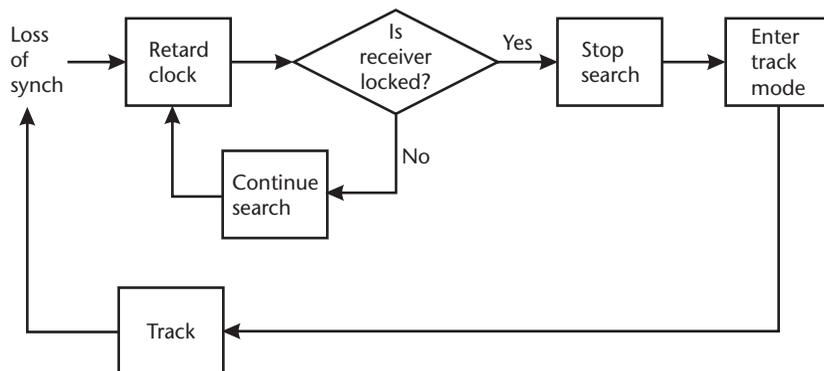


Figure 7.25 Flow diagram for sliding correlator synchronizer.

high-accuracy positioning systems. Spread spectrum is used for high-accuracy and high-resolution radar. The list of applications is infinite. There is every reason to believe that many more applications will be found for spread spectrum systems in the future.

References

- [1] Kennedy, G., *Electronic Communication Systems*, 3rd ed., New York: McGraw-Hill, 1985, Chaps. 3, 4, and 5.
- [2] Kennedy, G., *Electronic Communication Systems*, 3rd ed., New York: McGraw-Hill, 1985, pp. 484–486.
- [3] Murray, J. P., “Electromagnetic Compatibility,” in *Radar Handbook*, M. I. Skolnik, (ed.), New York: McGraw-Hill, 1980, pp. 29-18–29-22.
- [4] Kennedy, G., *Electronic Communication Systems*, 3rd ed., New York: McGraw-Hill, 1985, pp. 537–542.
- [5] Keiser, B. E., and E. Strange, *Digital Telephony and Network Integration*, 2nd ed., New York: Van Nostrand Reinhold, 1995, Chap. 6.
- [6] Dixon, R. C., *Spread Spectrum Systems*, New York: John Wiley & Sons, 1976.
- [7] Dixon, R. C., *Spread Spectrum Systems with Commercial Applications*, 3rd ed., New York: John Wiley & Sons, 1994.
- [8] Motorola. MC8504 Data Sheet, 1994.

Selected Bibliography

- Krauss, H. L., C. W. Bostian, and F. H. Raab, *Solid State Radio Engineering*, New York: John Wiley & Sons, 1980, Chap. 8.
- Matos, F., “Frequency Data,” Chap. 1 in *Reference Data for Engineers*, 8th ed., M. E. Van Valkenburg, (ed.), SAMS, Carmel, IN: Prentice Hall Computer Publishing, 1993.
- Noll, A. M., *Introduction to Telephones and Telephone Systems*, 2nd ed., Norwood, MA: Artech House, 1991, Chap. 3.
- Pursly, M. B., “Analog Communication,” Chap. 23 in *Reference Data for Engineers*, 8th ed., M. E. Van Valkenburg, (ed.), SAMS, Carmel, IN: Prentice Hall Computer Publishing, 1993.
- Pursly, M. B., “Digital Communication,” Chap. 24 in *Reference Data for Engineers*, 8th ed., M. E. Van Valkenburg, (ed.), SAMS, Carmel, IN: Prentice Hall Computer Publishing, 1993.

RF Amplifiers, Oscillators, Frequency Multipliers, and Mixers

This chapter briefly discusses RF amplifiers, oscillators, frequency multipliers, and mixers. These key components are found in communication, radar, and other important RF systems. The discussions here are at the systems level. More detailed discussions at the components and circuits level are provided in Part II.

8.1 Amplifiers

Amplifiers are important building blocks for RF systems, for example, in the following applications:

- Front-end, low-noise RF amplifiers for receivers;
- IF amplifiers for receivers;
- Audio, video, and other LF amplifiers;
- Buffer amplifiers for oscillators;
- Gain stages to augment gain lost due to passive components such as filters, mixers, and multipliers;
- Transmitter RF amplifier chains;
- Transmitter output power amplifiers.

Each of these applications is illustrated in Figure 8.1, which shows simplified system block diagrams for communication transmitter and receiver systems. Some of the information presented in this section is adapted from material presented in [1].

8.1.1 Front-End Low-Noise RF Amplifiers for Receivers

Front-end, low-noise RF amplifiers are an important part of many receiver systems. At MF and lower frequencies, there is no need for front-end, low-noise RF amplifiers for receivers. The external noise received by the antenna is so large that little is gained by having a low-noise amplifier precede the mixer.

Some, but not all, HF receivers also use low-noise, front-end amplifiers.

VHF and higher frequency receivers nearly always use low-noise, front-end amplifiers. The receiver noise from a mixer can be much greater than the noise that is received by the antenna. Having a low-noise amplifier that precedes the mixer thus can greatly improve the noise and sensitivity performance of the receiver

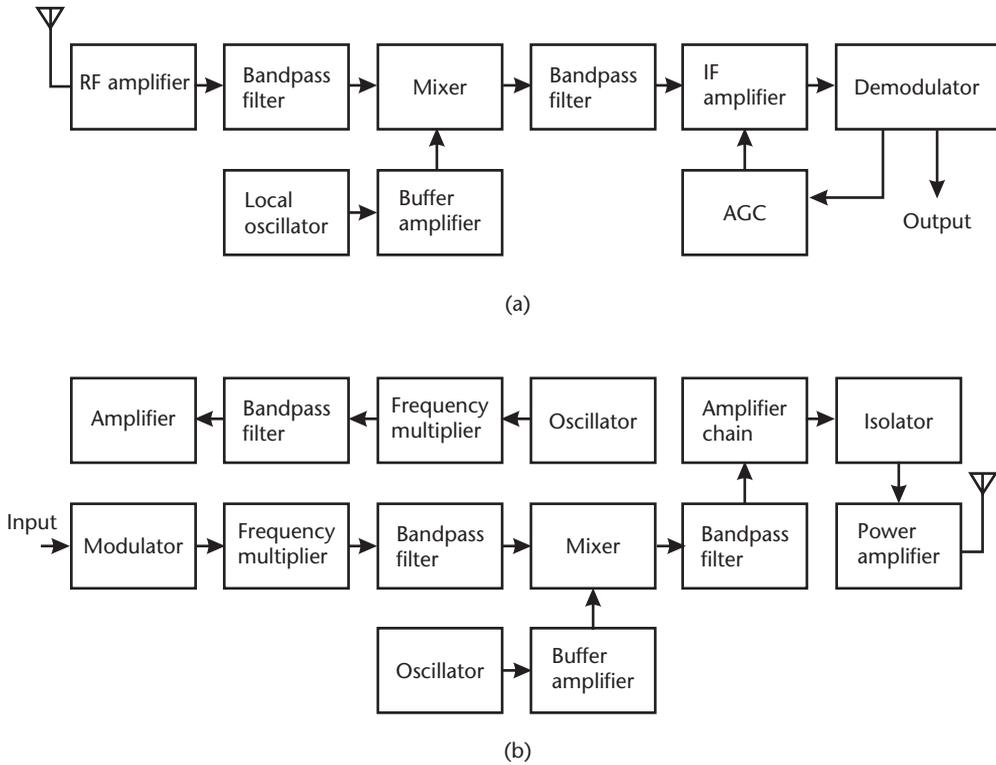


Figure 8.1 Simplified system block diagrams for receiver and transmitter systems: (a) receiver system and (b) transmitter system.

system. It was pointed out earlier that it is the first stage of a receiver that largely determines the system noise figure if that stage has reasonably high gain.

Low-noise RF amplifiers are constructed with either bipolar junction transistors (BJTs) or field-effect transistors (FETs). In each case, class A amplifiers are used. The term class A refers to the fact that the transistor is so biased that it is in conduction at all parts of the RF cycle. In contrast, a class B amplifier has a conduction angle of 180 degrees; a class AB amplifier has a conduction angle slightly greater than 180 degrees; and a class C amplifier has a conduction angle of much less than 180 degrees. The conduction angles for the four classes of amplifiers are illustrated in Figure 8.2.

Class A amplifiers usually consist of single-transistor gain stages. Class B and class AB amplifiers usually are two-transistor or two-tube “push-pull” type power amplifiers. Push-pull amplifiers are made up of two active devices, with one device operating for one-half of the RF cycle and the other device operating for the other half of the cycle. They frequently use transformers to combine the outputs. Class C amplifiers may be either single-transistor or single-tube circuits or push-pull type power amplifiers. In most cases, class C amplifiers use resonant circuits to convert the short-duration pulses to sine waves. Parallel devices are sometimes used in power amplifiers to increase the output power capability of the amplifiers.

Class A amplifiers have a maximum possible efficiency of 50%, whereas class B amplifiers have a maximum efficiency of 78%. Class AB amplifiers have maximum efficiency slightly greater than class B amplifiers. The efficiency of a class C

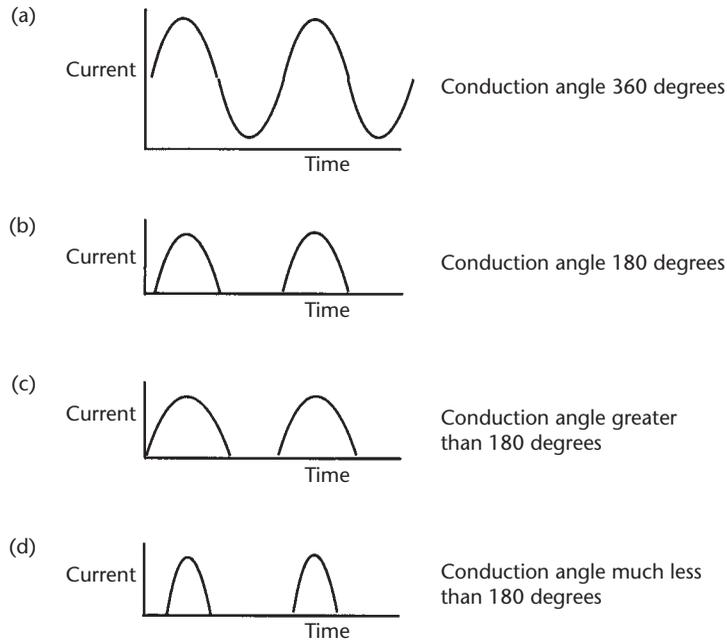


Figure 8.2 Conduction angles for different classes of amplifiers: (a) class A amplifier current; (b) class B amplifier current; (c) class AB amplifier current; and (d) class C amplifier current.

amplifier can be as high as 90% or more. Efficiency expressed as a percentage is 100 times the ratio of the RF output power to the dc input power.

Because class A amplifiers normally are used for small-signal amplification, efficiency is not very important. Class B and class AB amplifiers normally are used for large-signal amplification where efficiency is important. They normally are designed as linear amplifiers where the output is directly proportional to the input. They each may be wideband amplifiers. Class C amplifiers, on the other hand, are nonlinear amplifiers that may use tuned circuits on the output. Those tuned circuits are sometimes called *tank circuits*.

Silicon BJT RF amplifiers typically are used at the lower RF frequencies because of higher gain capability and more linear operation. The frequency is limited to 8–10 GHz, depending on the application. FET RF amplifiers made from gallium arsenide (GaAs) typically are used at microwave and higher frequencies because of their higher frequency capability and lower noise figure. The main reason for superior performance for these FETs at higher frequencies is the higher mobility of GaAs compared to silicon. FETs can use either silicon or GaAs in their construction, whereas BJTs use only silicon. BJTs and FETs are discussed in detail in Chapter 18.

The BJT RF amplifier is a current-controlled amplifier with the collector current determined by the base current. The FET amplifier, on the other hand, is a voltage-controlled amplifier with the drain current determined by the gate voltage. Each of these devices is diagrammed in Figure 8.3.

In each case, it is necessary to bias the transistor to the desired operating current or voltage using dc bias networks. It is also necessary to provide both input and output impedance matching circuits. Some matching circuits are designed to maximize

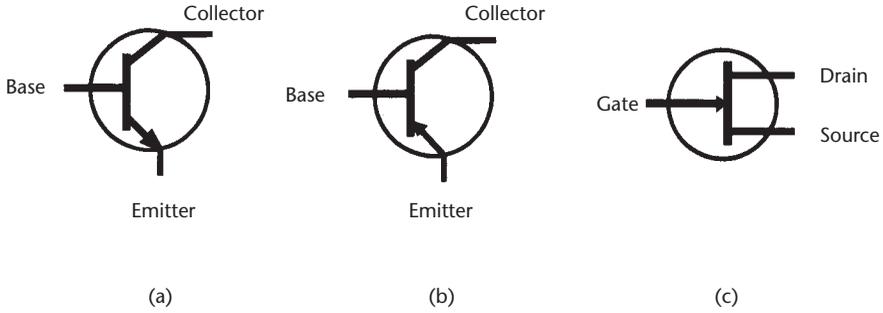


Figure 8.3 Diagrams for BJT and FET devices: (a) NPN BJT; (b) PNP BJT; and (c) FET.

power output, and some are designed for flat gain over a large bandwidth. In the case of the front-end low-noise amplifier, the matching may be for minimum noise figure. The design of impedance matching circuits for transistors using S-parameters is discussed in Chapter 18. The input impedance matching requirement for maximum gain is that the impedance of the transmission line that feeds the amplifier be transformed to the complex conjugate of the device input impedance. The output impedance matching requirement is that the impedance of the transmission line or load fed by the amplifier be transformed to the complex conjugate of the device output impedance. In each of these cases, the assumption is that S_{12} be essentially zero. For example, if the transistor input impedance at a given frequency is $25 + j 15\Omega$, the complex conjugate impedance would be $25 - j 15\Omega$. With a 50Ω characteristic impedance for the transmission line that feeds the amplifier, the required transformation would be from $50 + j 0\Omega$ to $25 - j 15\Omega$.

Impedance matching is done at VHF and lower frequencies using lumped constant LC circuits. At UHF and microwave frequencies, very small LC circuits sometimes are used in monolithic and MIC amplifier chips and ceramic packages. It is also possible to use microstrip and coax impedance matching circuits at microwave frequencies.

Figure 8.4(a) shows an example design for a low-noise transistor RF amplifier. It uses a separate bias source to set the bias current and operating point for the transistor. It also uses lumped element input and output impedance matching networks.

Some of the following material is quoted or adapted from [2].

In the ideal case, the signal input for a small-signal RF amplifier is simply increased in amplitude at the output without changing the frequency components of the signal. In practice, there will be some distortion and harmonic generation due to the nonlinear characteristics of the transistor. Bandpass filters are used to minimize or eliminate (suppress) signals outside the desired signal band.

Figure 8.4(b) shows the transfer function for a typical transistor. Because of its nonlinear nature, the transistor generates harmonics. It also can act as a mixer for multiple frequency signals, generating sum and difference frequencies. The mixing products form nonlinearities known as second- and third-order intermodulation distortion. Other forms of distortion generated by a transistor amplifier include cross-modulation distortion, and composite triple-beat distortion. As an example of the frequency components that are involved with a three-frequency signal at the input, the second-order distortion components include 3 dc components, 6 sum and

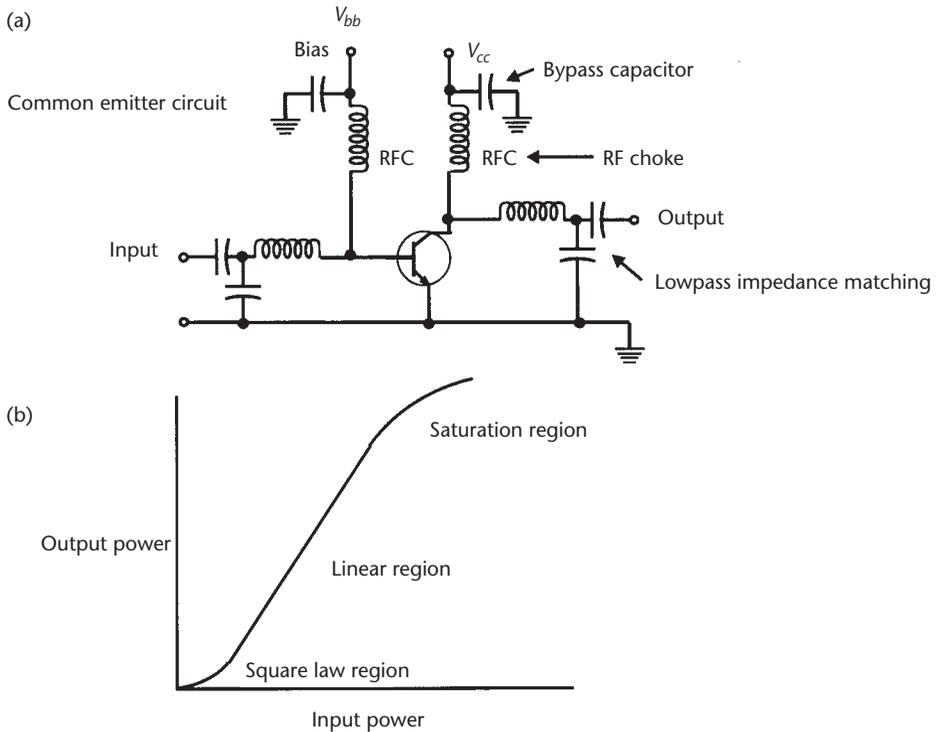


Figure 8.4 Example of a BJT RF amplifier: (a) RF transistor amplifier using impedance matching and (b) transfer function for a typical transistor.

difference components, and 3 second-harmonic components. The third-order distortion components include 3 third-harmonic components, 12 intermodulation components, 4 triple-beat components, 3 self-compression components, and 6 cross-compression or cross-expansion components.

In Figure 8.5, a typical transistor transfer curve shows the relationship between the fundamental, the second-order, and the third-order components of the signal. If the fundamental input power versus output power response of an amplifier is plotted on a log-log scale, it will have a 1-to-1 slope in the linear operating region. A plot of the second-order intermodulation products of the amplifier, plotted on the same scale, will have a slope of 2 to 1 (corresponding to a square relationship), and the third-order products will have a slope of 3 to 1 (corresponding to a cubic relationship).

The third-order spurious products are the most troublesome, because they may fall within the bandpass of even moderate bandwidth amplifiers.

The intercept point for the curves in Figure 8.5 generally is defined as the point where the extensions of the fundamental and the third-order responses intersect on the output power scale. The second-order response generally intersects at about the same point as well, unless the amplifier design suppresses even-order responses (e.g., push-pull stages).

The response curve shown in Figure 8.5 is for a device with a gain of 30 dB, a power output of +20 dBm at the 1-dB gain compression point, and a third-order intercept point of +30 dBm. As an example of the use of this plot, assume that the

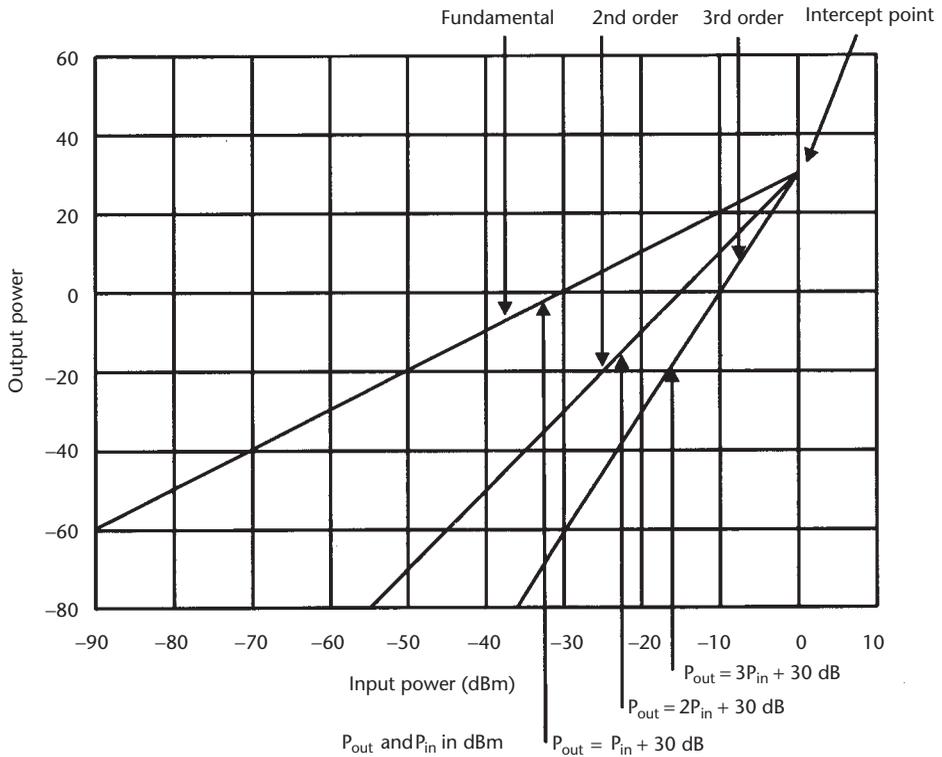


Figure 8.5 Fundamental, second-order, and third-order amplifier response curves for an RF amplifier. (After: [2].)

amplifier is driven to +15 dBm output. The second-order product will be suppressed 30 dB below the intercept point to 0 dBm. The third-order product will be suppressed 45 dB below the intercept point to -15 dBm.

For another example, assume that the amplifier is driven to +0 dBm output power. The second-order product will be suppressed 60 dB below the intercept point to -30 dBm. The third-order product will be suppressed 90 dB below the intercept point to -60 dBm.

One of the problems that we have with high-gain RF amplifiers is that they like to oscillate. Any student who has designed and built an RF amplifier knows about this problem.

A number of ways are used to prevent oscillations and provide the needed stability for RF amplifiers. Amplifiers are always designed so stages and overall systems have stability factors greater than 1.0, usually greater than 1.5 for unconditional stability. It is common practice to place amplifiers in metal containers or shields, so individual stages are shielded from each other. This type of shielding greatly reduces unwanted coupling between stages. Also, use of Eccosorb or other radar absorbing material (RAM) attached to covers (approximately 20–30 mils thick) usually reduces feedback and eliminates possible oscillations.

Another way to improve stability is to use neutralizing circuits. The goal is to provide a negative feedback signal that is equal in amplitude and opposite in phase to the positive feedback signal that is causing oscillations.

It is possible to buy packaged RF amplifiers from suppliers that specialize in amplifier design and production. Many such sources are available. These amplifiers may be in packages involving small 50Ω coaxial connectors, or they can be purchased as chips or flatpack devices with pins. Table 8.1 lists the characteristics for a low-noise GaAs MMIC amplifier manufactured by the ANZAC division of M/A-COM, which is owned by AMP.

8.1.2 IF Amplifiers

Most of the gain of a receiver is provided by the IF amplifiers. That is because it is easier to construct narrowband bandpass filter at IF frequencies than at RF frequencies, and amplifier stages have higher gain at the lower frequencies. These amplifiers follow the mixer and the LO, which converts the signal frequency from RF to IF.

At lower frequencies, it is possible to use single conversion and only one IF frequency. For example, at MF the AM receiver uses only one IF frequency at 455 kHz. At the higher frequencies, it is common to use two or more frequency conversions

Table 8.1 Characteristics for ANZAC Low-Noise GaAs MMIC Amplifier

<i>Characteristics</i>	<i>Values</i>			
Model Number	AM-280			
Guaranteed specifications (from -55°C to $+85^{\circ}\text{C}$)				
Frequency range	1.1–1.7 GHz			
Minimum gain (at 25°C)	1.1–1.3 GHz	20 dB		
	1.3–1.5 GHz	18 dB		
	1.5–1.7 GHz	16 dB		
Gain variation with temperature	± 3 dB max			
VSWR				
Input	1.1–1.5 GHz	2.0:1		
	1.5–1.7 GHz	2.3:1		
Output	1.1–1.7 GHz 2.0:1			
Maximum noise figure	At 25°C At 85°C At -154°C			
	1.1–1.5 GHz	1.7 dB	2.1 dB	1.5 dB
	1.5–1.7 GHz	2.0 dB	2.5 dB	1.8 dB
Output power for 1 dB compression	10 dBm min			
Operating characteristics				
Impedance	50Ω nominal			
Intermodulation intercept point (for two-tone output power up to 0 dBm)				
Second order	+35 dBm typ			
Third order	+22 dBm typ			
Bias power				
VD1 = 2.0–5.0 Vdc at ID1 = 20–50 mA				
VD2 = 2.0–5.0 Vdc at ID2 = 20–50 mA				
(Set VG1 and VG2 in the range of 0 Vdc to -3 Vdc to achieve desired ID1 and ID2 bias setting.)				
Die size	0.058 \times 0.048 \times 0.010 inch			
	(1.47 \times 1.22 \times 0.25 mm)			

and two or more IF amplifier frequencies. For example, a communication system might use a first IF at 30 MHz and a second IF at 455 kHz. The higher frequency for the first IF makes it easier to construct filters to suppress the image frequency, and the low frequency for the second IF makes it easier to make the required precision narrowband filters for providing optimum noise and signal bandwidth.

The image frequency is an undesired frequency that, when mixed with the LO signal, yields a signal at the IF frequency, just as the desired signal frequency does. Image frequency and its suppression are illustrated in Figure 8.6.

Figure 8.6(a) shows a simplified frequency spectrum for the output of the receiver antenna, along with a spectral line corresponding to the LO signal (f_{LO}). Two signals provide a frequency difference equal to the IF frequency: a signal frequency lower in frequency than the LO signal and an image frequency higher in frequency than the LO signal.

Figure 8.6(b) shows a simplified superheterodyne receiver that includes a bandpass filter for suppression of the image frequency. The filter has its bandpass centered at the desired signal and is sufficiently narrow that it presents large attenuation to the image frequency.

In some cases, the required IF bandwidth is very small. In other cases, such as for radar, this bandwidth must be quite large. In the latter case, it is common to use a second IF frequency such as 70 MHz and a first IF frequency of 300 MHz or more. In DBS receivers, the first IF frequency is 950–2,150 MHz, depending on the satellite system. The type of system discussed in the foregoing paragraphs that involves one or more mixers for downconversion and IF amplifiers is called a *superheterodyne* receiver. This is the most frequently used type of receiver. There are a few applications, however, where superheterodyne receivers are not used. The receiver used in those cases is called a *homodyne* receiver. In a homodyne receiver, all the gain is

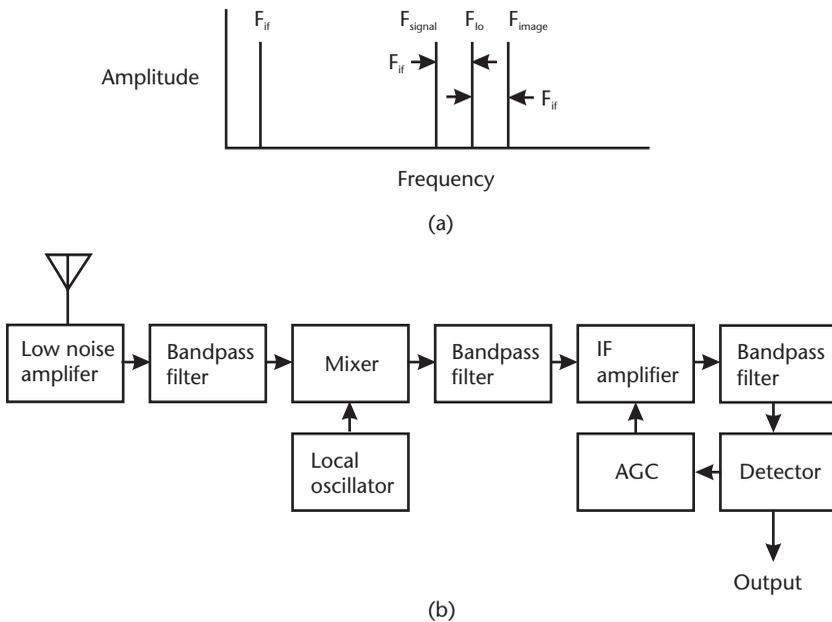


Figure 8.6 Superheterodyne receiver concepts: (a) frequency spectra showing desired signal frequency, LO frequency, and image frequency and (b) simplified superheterodyne receiver block diagram showing bandpass filter for suppression of image frequency.

provided at the RF frequency, and no IF amplifiers are used. An example of such a system is a very wide bandwidth receiver used in an ELINT system.

IF amplifiers usually include the capability for automatic gain control (AGC). A voltage is fed back from the demodulator following the last IF amplifier that is proportional to the amplitude of the detected signal. That voltage is applied to one or more amplifier stages to adjust the gain of those stages. An AGC loop is illustrated in the circuit in Figure 8.6(b).

The IF amplifiers used are sometimes BJT amplifiers involving a single transistor per stage, or they may be integrated circuits. In modern receivers, integrated circuit IF amplifiers are the most frequently used type.

Figure 8.7(a) shows a block diagram for a two-stage IF amplifier chain with the bandpass filtering and impedance matching provided by single-tuned magnetic transformers. Such a system frequently is used in simple AM receivers. Figure 8.5(b) shows a block diagram for an IF amplifier using discrete components and a three-electrode ceramic filter. This type of IF amplifier frequently is used in modern receiver systems rather than the older tuned transformer concept. Some televisions use SAW filters for better selectivity. Many other types of IF amplifier circuits could be shown, but these two are adequate to illustrate the main concepts of IF amplifiers.

8.1.3 Audio and Other LF Amplifiers

Many possible types of audio and other LF amplifiers are used in RF systems following demodulators or preceding modulators, including class A RC-coupled amplifiers, class A transformer-coupled amplifiers, emitter followers, and class AB push-pull amplifiers.

8.1.4 Transmitter RF Amplifier Chains

RF transmitters often use low-power modulators followed by amplifier chains that raise the power levels from at most a few milliwatts to the required drive level for

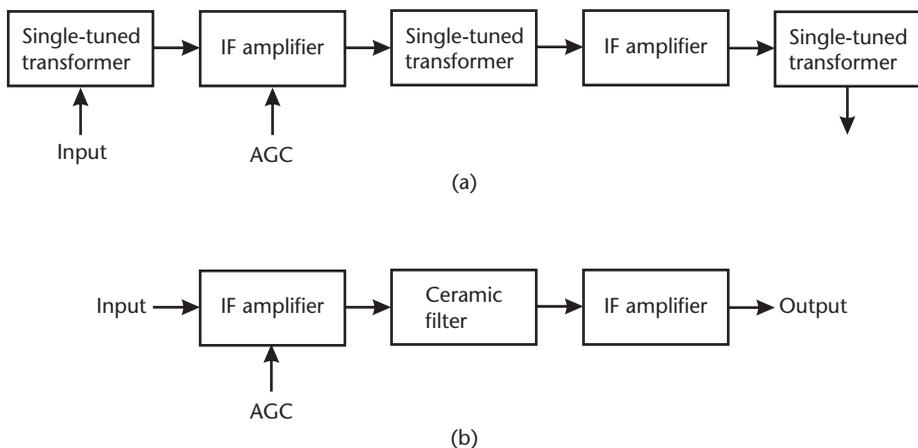


Figure 8.7 Example IF amplifiers: (a) two-stage IF amplifier with single-tuned magnetic transformers and (b) two-stage IF amplifier with a ceramic filter.

the final power amplifier. In many cases, this cascade of amplifier may include a frequency upconverter mixer stage for shifting the frequency from IF frequency to the desired RF one prior to transmission. In some cases, this amplifier chain may include a frequency multiplier stage. Block diagrams for these types of amplifier chains are shown in Figure 8.8.

The total gain for the CW amplifier chain is the sum of the gains for the individual amplifiers expressed in decibels. The output frequency is the same as the input frequency.

In a transmitter RF amplifier chain with a frequency converter, the output frequency is the sum of the input frequency and the LO frequency for upper sideband upconversion. It may also be the difference frequency, depending on LO frequency. For example, the input frequency may be 300 MHz and the LO frequency 1,000 MHz. The output frequency would be 1,300 MHz for the sum frequency or 700 MHz for the difference frequency. Again, the gain of the amplifier chain will be the sum of the gains or losses for each of the cascaded stages, expressed in decibels.

In a transmitter RF amplifier chain with a frequency multiplier, the output frequency is N times the input frequency, where N is the multiplication factor. For example, the chain may use a times-3 frequency multiplier. If the input frequency is 300 MHz, the output frequency will be 900 MHz.

8.1.5 Transmitter RF Power Amplifiers for Communication Systems

The final stage for the communication transmitter is the power amplifier. Transmitter RF power amplifiers for communication systems include both transistorized power amplifiers and tube-type power amplifiers. Transistor amplifiers include both

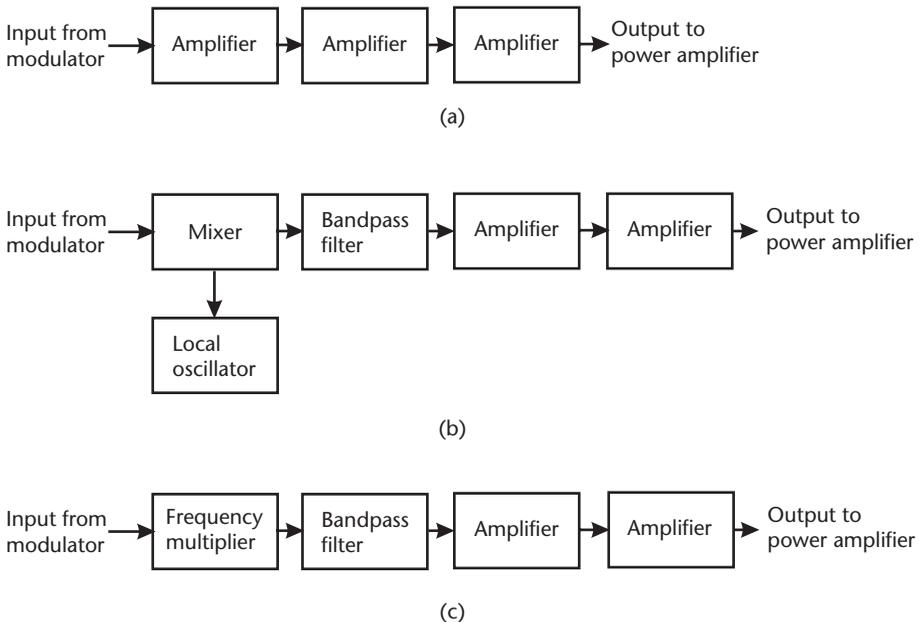


Figure 8.8 Example frequency amplifier chains for transmitters: (a) CW transmitter RF amplifier chain; (b) transmitter RF amplifier chain with frequency converter; and (c) transmitter RF amplifier chain with frequency multiplier.

single-ended RF amplifiers and push-pull amplifiers. Transistor types used in these amplifiers include both BJTs and FETs.

Figure 8.9 shows a schematic diagram for a class AB push-pull RF power amplifier. This amplifier uses two NPN BJTs. The bases for the transistors are connected to the secondary of an input transformer. This secondary winding is center tapped and is biased by means of a voltage supply, a series resistor, and a forward-conducting diode to the turn-on voltage for the transistor. That helps avoid distortion caused by the nonzero turn-on characteristic for BJTs. Having a conduction angle slightly greater than 180 degrees, as needed for class AB operation, is made possible by this type of bias.

The collectors for the two transistors are connected to the primary winding of the output transformer. This is a center-tapped winding, with the center tap connected to +V. With this type of amplifier, the two transistors take turns conducting for a half-cycle each. The result of the transformers is that a full-wave signal of low distortion is produced from the two half-wave signals.

Very high power transistor power amplifier systems may be made by combining the output power from many individual power amplifiers using RF power combiner circuits. With the use of a large number of transistors and combiners, it is now possible to provide output powers in the kilowatt range at frequencies up to about 1.0 GHz. Progressively smaller power levels can be provided at microwave frequencies.

Figure 8.10 shows a block diagram for a high-power transistor amplifier system consisting of 16 individual class AB power amplifiers. Groups of four amplifiers are combined using 4-to-1 RF power combiner circuits. The outputs of the four combiners are then combined using another four-way power combiner. The result is that the power output for the system is 16 times the power output of the single amplifiers less the losses in the power combiners.

Tube-type power amplifiers for communication transmitter systems are of two types. Grid-type vacuum tube amplifiers use triodes, tetrodes, or pentodes. At microwave frequencies, tubes used in communication transmitter systems include traveling wave tubes (TWTs) and klystrons. Vacuum-tube devices and circuits are discussed in Chapter 19. Some of the information here about tube-type power amplifiers is adapted from [3, 4].

Both communication-type, wideband TWT amplifiers and power solid-state amplifiers are used for point-to-point and mobile communication systems for surface installations. Transistor amplifier chains are used for the lower frequency bands, while TWTs are used for microwave frequencies. With TWTs in the 5–18-GHz range, the amplifier gains typically are about 50 dB. Output powers are

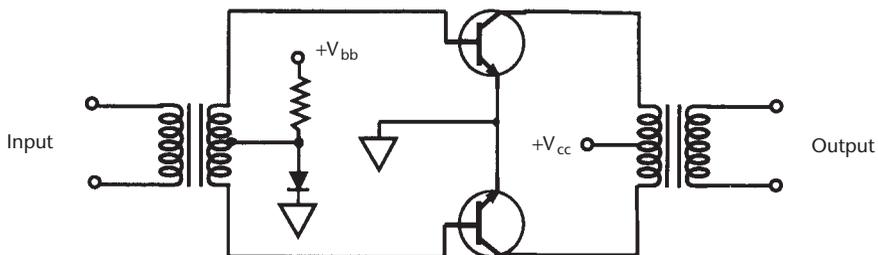


Figure 8.9 Class AB push-pull transistor amplifier circuit.

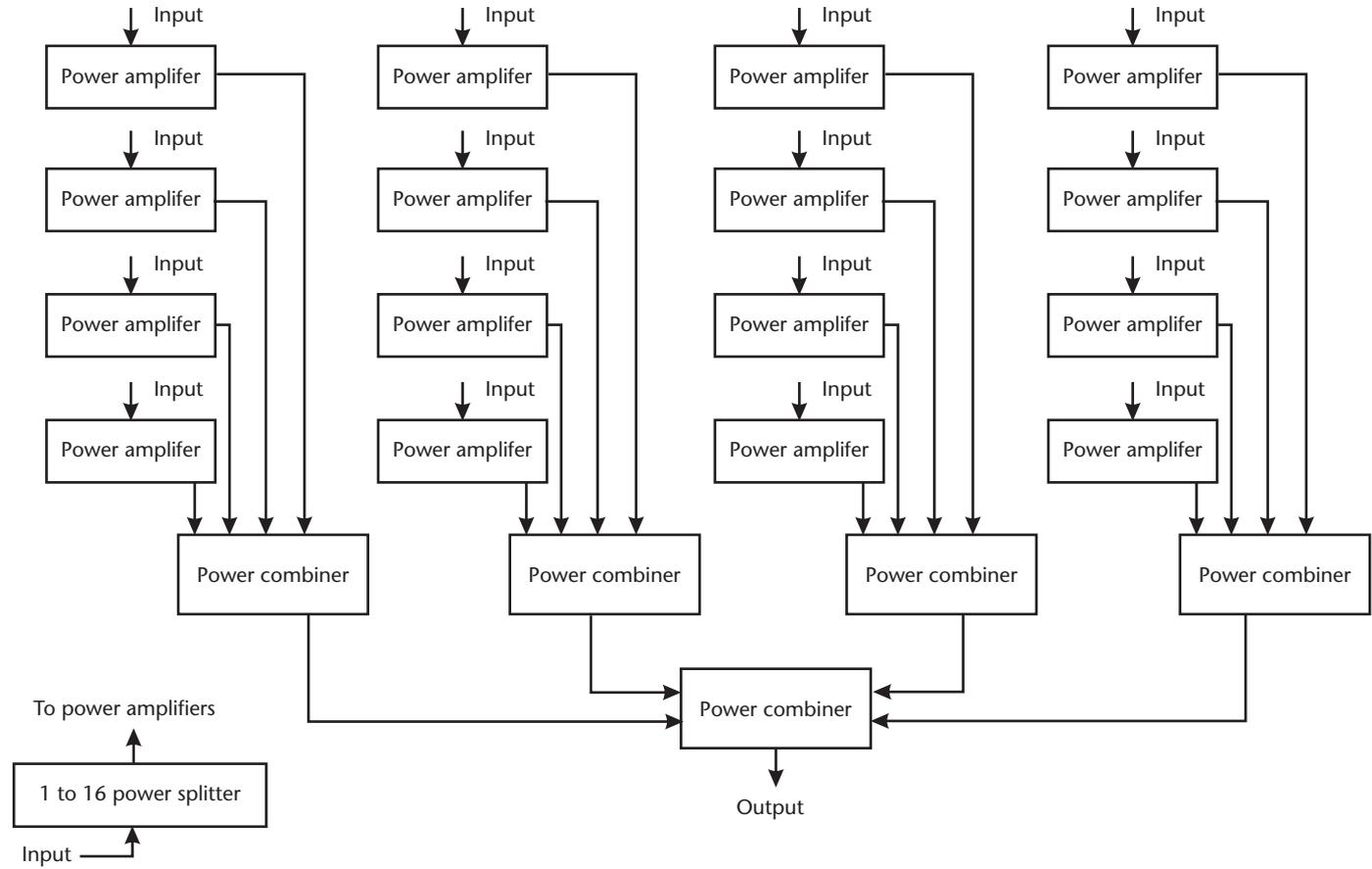


Figure 8.10 High-power transistor amplifier using 16 individual power amplifiers that are combined using RF power combiners.

available for as low as 25W to as high as 900W. These are driven by transistor amplifier chains.

Satellite uplinks, which require fairly high transmitter powers, usually use either TWT power amplifiers or klystron power amplifiers. Satellite downlinks currently use TWT amplifiers rather than solid-state amplifiers. The main reason for that is the higher efficiency for the TWT amplifiers. TWT amplifier efficiencies typically are as high as 60%, whereas solid-state amplifiers for the same frequencies typically are only about 30% to 40% efficient. The reason that efficiency is so important for satellite downlink transmitters is that the dc power for the amplifiers must be provided by solar cells. A saving in dc power means a big saving in cost of the satellite.

Klystrons are used for UHF television transmitters. Output powers typically are in the range of 30–60 kW. Typical gains are 30–50 dB. TWTs also can be used for broadcast, including both surface installations and spaceborne installations.

High-powered grid-type triodes and tetrodes are used as power amplifiers for broadcast at MF frequencies and for other low frequency applications. They produce output powers in the range of 10–500 kW.

8.1.6 RF Power Amplifiers and Oscillators for Radars, Navigation, and Electronic Countermeasure Applications

One type of power amplifier used for surface radar installations is a TWT/crossed-field amplifier (CFA) chain. A CFA is a very efficient, high-power amplifier. Because a typical gain for this type of device is between 10 and 20 dB, it is necessary to drive this tube with a fairly high-powered amplifier of high gain such as a TWT. Thus, the term TWT/CFA refers to a CFA driven by a TWT amplifier.

As an example of power levels and gains, a typical CFA has an output power of 1 MW (60 dBw) and a gain of 16 dB. The required input power from the TWT thus would be $60 - 16 = 44$ dBw, or 25 kW. A typical coupled-cavity TWT (CCTWT) could have an output power of 25 kW and a gain of 44 dB. The required input to the TWT then would be 1W, which can be supplied by a transistor amplifier chain. If an upconverter is the RF signal source that produces 1.0 dBm of output power, the required gain of the transistor amplifier chain would be 30 dB.

A second microwave tube used for surface installation search and surveillance radar is a high-power klystron. This type of high-powered microwave tube amplifier is available over the frequency range from UHF to K_a -band. Varian Associates makes narrowband, high-powered, pulsed klystrons with peak output powers in the range of 1–5.5 MW over the frequency range of 0.9–5.9 GHz. An example is an amplifier operating in the 2.7–3.0-GHz band with an output peak power of 1.5 MW (61.8 dBw), an average power output of 3 kW (34.7 dBw), and a typical gain of 50 dB. The required input or driving power from a transistor amplifier chain would be $61.8 - 50 = 11.8$ dBw, or 14.4W.

The third microwave tube possibility for the surface installation search and surveillance radar is the high-power coupled-cavity traveling wave tube (CCTWT). This type of high-power microwave-tube amplifier also is available over the frequency range from UHF to K_a -band. For example, a tube provides a peak output power of 240 kW (53.8 dBw) at UHF. Gains are in the range of 30 to 65 dB. For

example, assume a gain of 50 dB. The required drive power from a transistor amplifier chain is thus 3.8 dBw or 2.4W.

The fourth microwave-tube possibility for the surface installation search and surveillance radar is a high-power twystron. A twystron has an input section that is like that of a klystron and an output section like that of a TWT. It has the advantage of having a larger bandwidth capability than a klystron but with similar power and efficiency characteristics. Twystrons are available over the same frequency range as the klystron (UHF to K_a-band). A typical output power level is about 1 MW. A typical gain is about 50 dB. Again, the required driving power is only about 10W, which may be provided by a transistor amplifier chain.

A fifth way to provide the needed power for a search and surveillance radar operating at L- or S-band is to use many solid-state, pulse-power amplifiers that feed individual elements of a phased-array antenna. If, for example, 100 amplifiers are used, with each providing 100W output, the total power level would be 10 kW if properly phased.

Search and surveillance radars are also used with airborne and spaceborne installations. The types of microwave amplifiers used are TWTs and klystrons. The power levels used for these applications typically are lower than the examples shown for the surface installations. TWTs may be of the wideband type using helix slow-wave structures. Such tubes are available with pulse power outputs of 1 to 2 kW over the frequency range from 2 to 18 GHz. Bandwidths may be of the order of 2:1 to 3:1 (6–18 GHz), so complete radar and ECM bands are covered.

A wide range of powers is available for pulsed klystron amplifiers used for airborne and spaceborne radars. These may be as small as 1 kW or as large as 100 kW. Such pulsed amplifiers are available over the frequency range of 0.4–36 GHz. Gains are in the range of 40–60 dB. Again, the drivers for such amplifiers can be transistor amplifier chains.

Surface-based fire-control radars use TWT/CFA amplifier chains and TWT amplifiers. Fire control radars for airborne installations also use pulsed TWTs. They also may use pulsed magnetrons. Magnetrons are small crossed-field microwave oscillators with good power-to-weight ratios. The required inputs for these devices are high-voltage dc pulses. Magnetrons are very efficient and rugged. Typical output power levels are in the range of 10 kW to 1 MW.

Lower power magnetrons also are the tubes used to supply the RF energy in most, if not all, microwave ovens.

Weather radars are used in aircraft for bad-weather avoidance. The systems may use either TWT amplifiers or magnetron oscillators. TWT amplifiers are driven by transistor amplifier chains.

Missile homing as used by airborne systems usually is done using small magnetron oscillators. These are the simplest and lowest cost type radar systems.

Speed measurement by the police usually is done using small solid-state microwave Gunn diode oscillators. These very simple transmitters produce a few tens of watts.

In the past, navigation beacons used only magnetrons. These are now being replaced by solid-state amplifiers.

Both surface-based and airborne ECM systems use wideband helix-type TWTs. CW power levels typically are on the order of 200W or less. Dual-mode systems may

be used with higher pulse power capability provided in addition to CW operation. Again, these devices are driven by low-power solid-state amplifier chains.

8.2 Oscillators and Frequency Synthesizers

This section discusses oscillators and frequency synthesizers. Some of the information presented in this section is adapted from information presented in [5].

8.2.1 Transistor Feedback Oscillators

Communication transmitters and receivers typically use transistor feedback oscillators as frequency generators. Basically they are transistor amplifiers with positive feedback so that a part of the output signal is fed back to the input as a signal with the proper phase and amplitude to permit stable oscillation. The frequency of oscillation is determined mainly by the resonant circuit that is used in the feedback circuit.

Two main types of transistor oscillators are used in communication systems at lower frequencies: LC oscillators and crystal oscillators. There are also dielectric resonator oscillators (DROs) and those that use ceramic resonators. Others use transmission-line elements to provide the necessary resonant or tank circuits.

8.2.1.1 LC Oscillators

LC oscillators are frequently used in communication systems to generate signals. Types of LC oscillators frequently used include the Colpitts oscillator, the Clapp oscillator, the tuned- input tuned-output JFET oscillator, and the differential-pair oscillator. These oscillators may be fixed-frequency oscillators that are mechanically tuned using variable capacitors, or they may be voltage-controlled LC oscillators (VCOs) that use varactor diodes to provide variable capacitance, occasionally in parallel with fixed capacitors.

LC oscillators can be used as LOs for mixers and frequency converters, as carrier frequency inputs to modulators, and as frequency modulators. Their main weakness is that they do not provide very good frequency accuracy or stability or good phase noise performance, necessary for digital communication systems.

8.2.1.2 Quartz Crystal Oscillators

Quartz crystal oscillators are used extensively in communication and radar system transmitters. They can operate either in a series resonant mode or in a parallel resonant mode, depending on the design of the oscillator. Such crystals are very high-Q devices that can provide very high accuracy for frequency control. (The Q of a crystal is the center frequency divided by the 3-dB or half-power bandwidth.) Accuracies are in the range of one part in 10^5 to one part in 10^7 , in other words, 0.1 to 10.0 parts per million (ppm). The frequency range for these oscillators is typically 10^5 Hz (100 kHz) to 10^8 Hz (100 MHz).

Crystal oscillators can be tuned over a small frequency range using trimmer capacitors. These may be either mechanically tuned capacitors or voltage-tuned capacitors (varactors). Crystal oscillators are used for the same functions as LC oscillators.

Figure 8.11 shows a Colpitts oscillator with the crystal in series with the feedback path to the emitter. The crystal in this case is operating in the series resonant mode. Other frequently used crystal oscillators are the Miller oscillator and the Pierce oscillator (discussed in Chapter 16).

8.2.1.3 Yttrium Iron Garnet Resonator Oscillators

One type of UHF and microwave oscillator that is sometimes used is a yttrium iron garnet (YIG) resonator oscillator. YIG oscillators use highly polished YIG spheres as resonant devices with magnets to determine their resonant frequencies. Electromagnets can be used to tune the devices to the desired frequency of oscillation. The oscillators are used by transmitters as LOs for mixers or frequency converters. YIG devices and circuits are discussed in more detail in Chapter 16.

8.2.1.4 Dielectric Resonator Oscillator

DROs are microwave oscillators that use the unique properties of high-Q dielectric materials as a resonator to stabilize the frequency of free-running sources. Typical frequency stability is about 5 ppm per degree centigrade. Typical output power levels for oscillators of this kind are 0 to +20 dBm. DROs can operate from 1 to about 35 GHz. They have good phase noise properties.

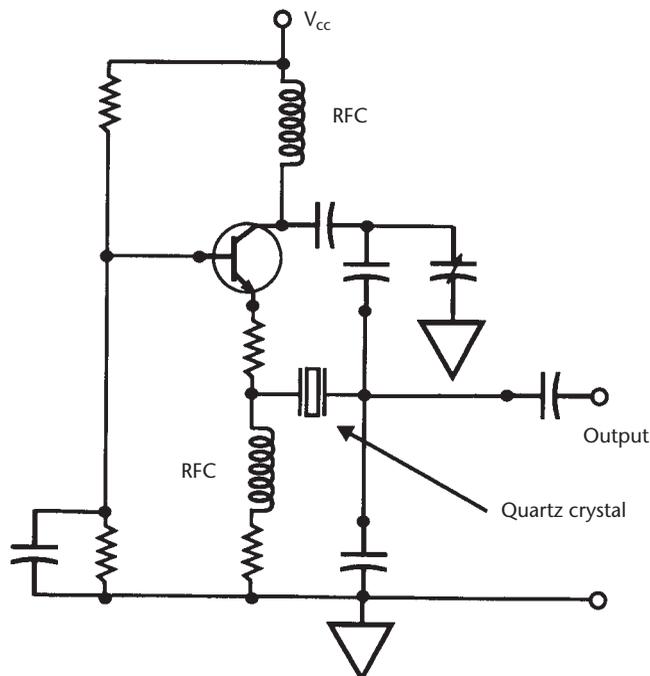


Figure 8.11 A Colpitts crystal oscillator.

8.2.2 Negative Resistance Two-Terminal Oscillators

A second class of oscillator sometimes used at UHF and microwave frequencies is a negative resistance oscillator. This class of oscillator includes tunnel diode oscillators, Gunn diode oscillators, IMPATT or avalanche diode oscillators, trapped plasma avalanche transit time (TRAPATT) diode oscillators, and limited space-charge accumulation (LSA) mode oscillators. The frequency of operation for these diode oscillators is determined by the use of cavity resonators. The Gunn diode and the IMPATT diode both are capable of a few watts output power. In general, the negative resistance-type oscillators have poor phase noise and poor frequency stability.

8.2.3 Frequency Synthesizers

Frequency synthesizers frequently are used with transmitters and receivers when there is a need for accurate multiple frequency selection. The two main types are direct frequency synthesizers and indirect frequency synthesizers. In direct frequency synthesizers, a number of crystal oscillators are used along with frequency multipliers, mixers, and switches. Frequency multipliers provide multiple frequencies from a single source that may be selected by an RF switch. Mixers provide sum and difference frequencies. By selection of the proper combination, it is possible to produce any one of hundreds of frequencies with the same accuracy as that of the crystal oscillators. Direct frequency synthesizers tend to be more complex, more costly, and have poorer noise characteristics than indirect frequency synthesizers.

Indirect frequency synthesizers frequently are used in all types of communication systems involving frequency selection. A system of this type is shown in Figure 8.12. The system in the figure uses a 100-kHz crystal oscillator as a reference frequency source for a PLL system. This oscillator is followed by a divide-by-4 frequency scaler, which provides a 25-kHz output signal. That 25-kHz signal is fed to a phase detector that compares its phase with that of the second input from a programmable frequency divider. The output of the phase detector is fed through a loop filter to a varactor-tuned LO. For the system in Figure 8.12, the oscillator frequency range is 98.8–118.6 MHz. The output of this oscillator initially is divided in frequency by 8 by a frequency prescaler. The output of the circuit is then fed to the programmable frequency divider, which has possible divider factors of 494 to 593. The output frequency is 25 kHz when the system is phase locked.

This circuit provides LO frequency output in the frequency range of 98.8–118.6 MHz tunable in steps of 25 kHz with an accuracy equal to that of the crystal oscillator. PLL circuits of this kind are available in integrated circuit form at low cost. They are, thus, finding application in many different types of low-cost receivers as well as in transmitters. They also are finding use in cellular devices.

PLLs and prescalers are available as integrated circuits from many companies. Prescalers are available that operate to 12 GHz (Hewlett-Packard Corporation).

Frequency synthesizers can be made for UHF and microwave frequencies as well as for VHF and lower frequencies. Some microwave systems use YIG-tuned oscillators. Most use varactor-tuned oscillators, since the tuning rate is faster and the tuning circuitry is easier to construct and simpler. Also, YIG-tuned oscillators can suffer from tuning hysteresis.

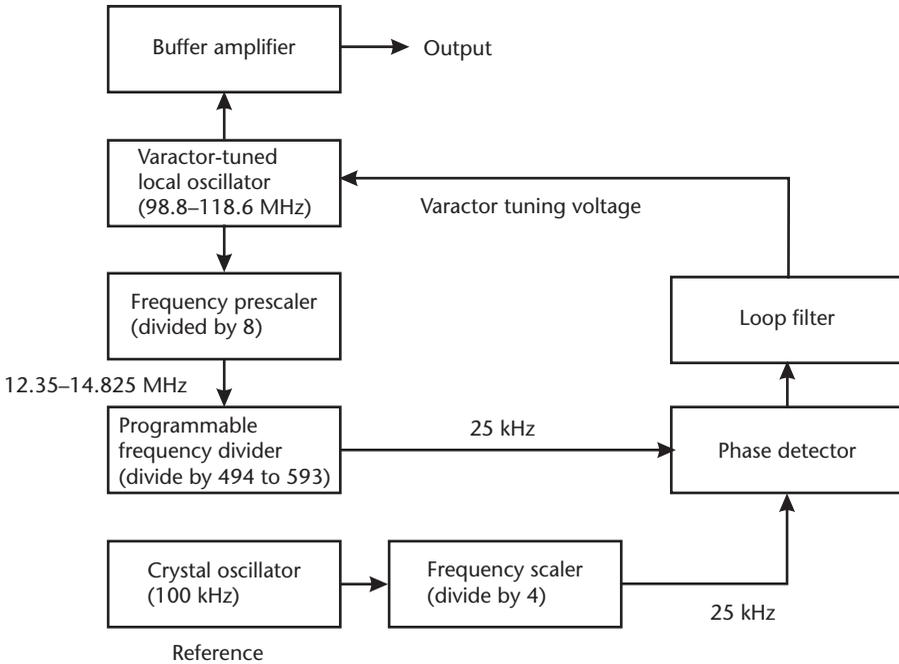


Figure 8.12 An indirect frequency synthesizer. (After: [6].)

It is common practice to use a buffer amplifier between the oscillator or frequency synthesizer that feeds the carrier frequency to the modulator circuit. Doing so provides isolation and a good impedance match between the VCO and the equipment to which it is connected. The use of a buffer amplifier also increases gain to provide sufficient power out for a mixer it may be driving. An alternative is to use an isolator (discussed in Chapter 11).

8.3 Frequency Multipliers

The nonlinearity inherent in any semiconductor diode or transistor can be used to multiply frequency. The most popular diode frequency multipliers use either varactor diodes or step-recovery diodes. Frequency multipliers of this type are discussed in this section. Some of the information presented here is adapted from [5, 7].

8.3.1 Varactor Diode Frequency Multipliers

Figure 8.13 shows a varactor-diode frequency tripler. Input and output impedance-matching circuits are used with this circuit. The input and output ports are coupled to the diode through series-tuned circuits, causing the input current and the output current and voltage to be essentially sinusoidal.

The varactor diode is a nonlinear voltage-variable capacitor. One or more so-called idler circuits are used with the circuit. These series-tuned circuits are placed in parallel with the diode and are resonant at harmonic frequencies other than the desired output frequency. Idler circuits are in practice empirically selected

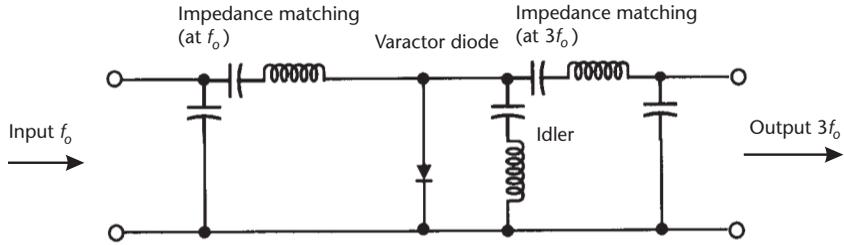


Figure 8.13 Varactor diode tripler circuit. (After: [5].)

and adjusted to improve efficiency. They are resonant at other harmonics, presenting an impedance to reflect unwanted harmonic energy back to the input.

Hyperabrupt snap-off varactor diodes multiply by high factors with better efficiency than ordinary varactor diodes, so they are used wherever possible. GaAs varactors often are used at the higher frequencies. A varactor multiplier of this type can have an efficiency for a 60-GHz doubler of greater than 50%.

The maximum output power for the varactor diode multipliers ranges from more than 10W at 2 GHz to about 25 mW at 100 GHz. Tripler efficiencies range from 70% at 2 GHz to about 40% at 36 GHz. One of the current applications for multiplier chains is to provide a low-power signal to phase-lock a Gunn or IMPATT oscillator.

8.3.2 Step-Recovery Diode Frequency Multipliers

A step-recovery diode is a silicon or GaAs p-n junction diode with construction similar to that of a varactor diode. It stores charge when conducting in the forward direction. When reverse bias is applied, the diode very briefly discharges the stored energy in the form of a sharp pulse that is rich in harmonics. The duration of the pulse typically is only 100 to 1,000 picoseconds (ps) ($1 \text{ ps} = 10^{-12} \text{ sec}$), depending on diode design.

Step-recovery diodes are frequently used in frequency multipliers. The circuit is essentially the same as that used for the varactor frequency multiplier, except that an inductor is used in series with the step-recovery diode. During the RF cycle, when the charge is completely drained from the diode, it switches from a conducting to a nonconducting state. This abrupt change in impedance then causes the inductor, which has stored energy, to generate a voltage impulse and hence a relatively flat frequency spectrum. Thus, the circuit is essentially an impulse generator. The desired harmonic frequency component is then extracted by the narrowband-tuned circuit and appropriate idler networks.

Figure 8.14 shows a typical multiplier chain using step-recovery diodes and varactors. The first stage is a transistor crystal oscillator followed by an amplifier with 35W output power at 160 MHz. That is followed by a step-recovery diode with a tuned circuit that is tuned to the tenth harmonic of 160 MHz, or 1.6 GHz. The output power of this stage is 3.5W, and the output frequency is 1.6 GHz. The next stage is a step-recovery diode with a tuned circuit that is tuned to the fifth harmonic of 1.6 GHz or 8.0 GHz. The output power of this stage is 0.7W, and the

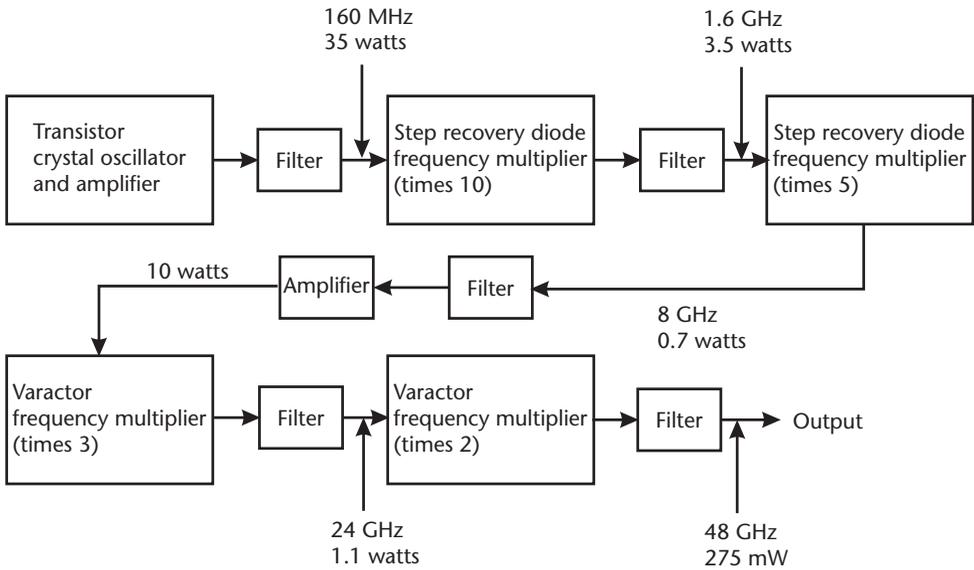


Figure 8.14 Frequency multiplier chain using step-recovery diodes and varactor diodes. (After: [7].)

output frequency is 8 GHz. The last two stages of the circuit use varactor diodes. The first of those stages is a tripler with an output frequency of 24 GHz and an output power of 1,100 mW. The last stage is a doubler with an output frequency of 48 GHz and an output power of 275 mW.

Frequency multipliers at microwave frequencies use microstrip, stripline, coaxial line, or waveguide. Step-recovery diodes are not available for frequencies above about 20 GHz, whereas varactors can be used well above 100 GHz.

Step-recovery diodes are available for powers in excess of 50W at 300 MHz, 10W at 2 GHz, and 1W at 10 GHz. Multiplication ratios up to 12 are commonly available. Efficiency can be in excess of 80% for triplers at frequencies up to 1 GHz. The efficiency drops to about 15% for a 5-times multiplier with an output frequency of 12 GHz.

8.3.3 Transistor Multipliers

Transistor multipliers also are frequently used in RF systems. These multipliers can exhibit less loss than varactors or SRDs. In a class A multiplier using a BJT, the circuit is similar to that of a small-signal amplifier with the output circuit tuned to the desired harmonic of the input signal.

A class A doubler can be made using an FET with a square-law transfer characteristic. Again, the circuit can be designed as a small signal amplifier with the output circuit tuned to the second harmonic of the input signal. A typical efficiency for this type of frequency multiplier is about 16%.

Class C frequency multipliers are possible using transistors. The design principles and procedures generally are the same as for class C power amplifiers. The output circuit is tuned to the desired harmonic of the input signal.

8.4 Mixers

Mixers are one of the most important of all RF components. Some of the information presented in this section is adapted from [6, 8, 9].

8.4.1 Diode Mixers

An important application for diodes is in mixer circuits. A number of different types of diode mixers are used in modern RF systems, including single-ended diode mixers, balanced diode mixers, double-balanced diode mixers, and triple-balanced diode mixers. Diode mixers can be used at both IF and microwave frequencies.

8.4.1.1 Single-Ended Diode Mixers

Figure 8.15(a) shows the circuit diagram for a single-ended diode mixer. A single diode is shown in series with the RF and LO inputs, a bias source, and a circuit tuned to the desired IF frequency. Such a circuit has a number of disadvantages compared with other types of mixers at the lower frequencies. It has a relatively high noise figure, a high conversion loss, high-order nonlinearities, no isolation between the LO and the RF inputs, and large output current at the LO frequency. Its main application is at microwave frequencies, where other types of mixers may not be practical.

At microwave frequencies, the single-ended diode mixer may use a Schottky-barrier type diode or, in some cases, a point-contact diode. The diode may be mounted in a waveguide so that it provides a complete dc path for rectification,

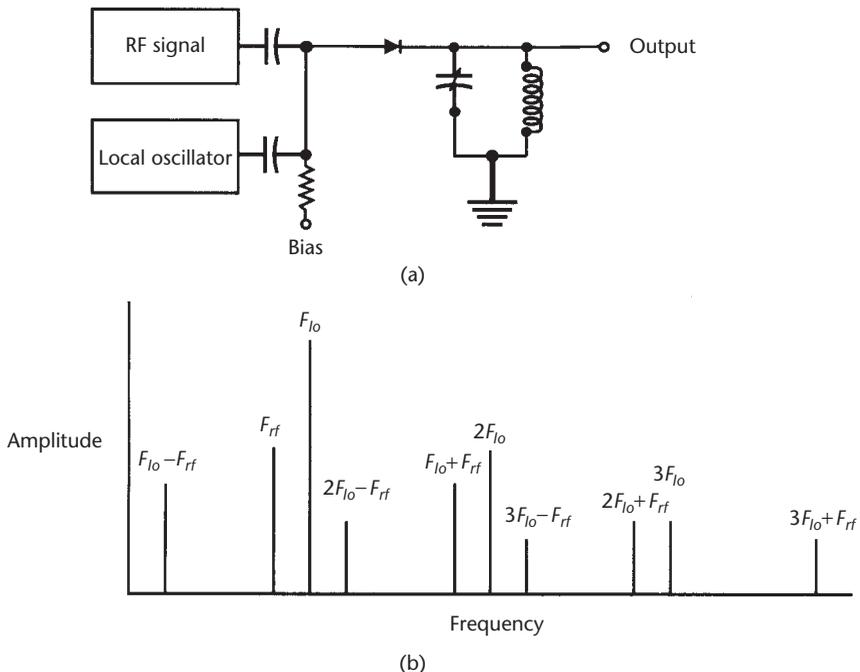


Figure 8.15 Single-diode mixer: (a) circuit diagram for a single-ended diode mixer and (b) partial frequency spectrum for a single-ended diode mixer.

without causing serious reflections in the waveguide. It is located a quarter-guide wavelength from a short where there is a region of high electric field strength. The diode output is passed through a dielectric RF bypass capacitor. The IF output to the IF amplifier is via a coax cable. The LO signal is introduced using a sidearm with a tuning screw to adjust tuning.

Waveguide is only one media that can be used for the single-ended diode mixer. Such mixers also are made in stripline, coax, and microstrip.

Figure 8.15(b) shows a partial frequency spectrum for a single-ended diode mixer. Note that the unfiltered spectrum includes the RF input frequency, the LO input signal, and all harmonics of the LO signal. It also includes the sum and difference frequencies on either side of the LO frequency and its harmonics. Other frequency components are not shown in the diagram. The amplitudes shown for the different frequency components are not to scale but are representative of the frequencies involved.

8.4.1.2 Single-Balanced Diode Mixers

Figure 8.16(a) shows a single-balanced diode mixer. Four diodes are used to provide a switching or clamping circuit. The RF signal can be fed to the mixer through a series resistor, which in turn is connected to both the diode switch and the output filter and load resistor. The control of the switch is by means of a transformer that is fed by the LO having an output voltage larger than the RF signal. During one-half of the cycle, the output is effectively zero. During the other half of the LO signal cycle, the diodes are back-biased and the switch is an effective open circuit. The output produced by

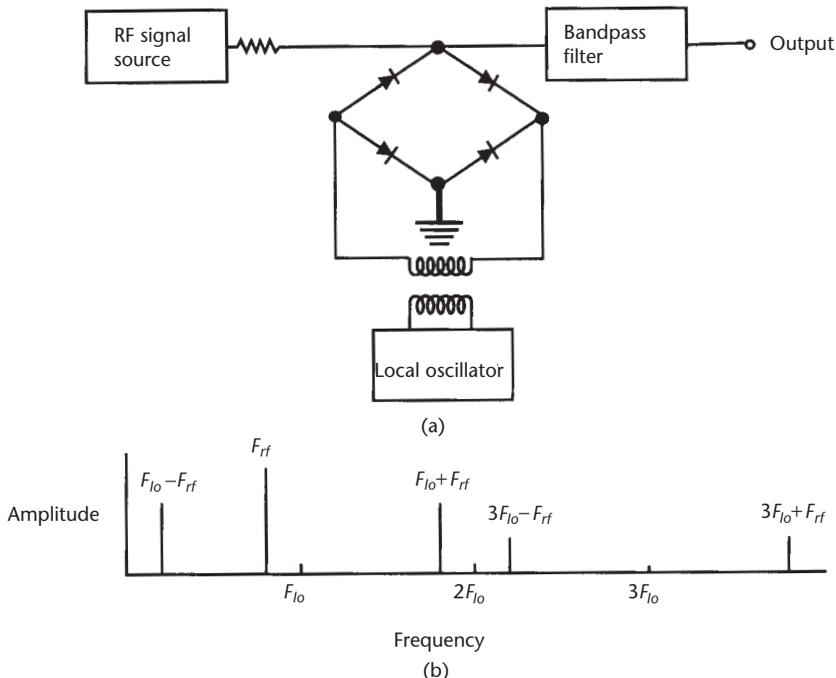


Figure 8.16 Single-balanced diode mixer: (a) single-balanced diode mixer circuit; (b) partial frequency spectrum for a single-balanced diode mixer. (After: [6].)

this mixer thus has a chopped waveform. If the RF frequency is higher than the LO frequency, the chopped signal output would be a series of RF pulses at the LO frequency repetition rate. Other versions of the single-balance mixer also are possible.

Figure 8.16(b) illustrates the frequency spectrum produced by the single-balanced diode mixer. The magnitude of the LO frequency and its harmonics are very small but are shown as frequency reference points. The components are largely eliminated by this type of mixer. The magnitude and the frequency of the RF signal and the sum and difference frequencies on either side of the LO frequency and odd harmonics of this frequency are shown. Other frequencies also are generated but are not shown. A bandpass filter is used to select the desired frequency components from the mixer and to eliminate any possible image response.

The single-balanced mixer has the disadvantage that a component of the RF frequency appears at the output. That is not the case for the double-balanced mixer.

8.4.1.3 Double-Balanced Diode Mixer

Figure 8.17(a) illustrates the double-balanced diode mixer, which uses two center-tapped transformers or power dividers along with four diodes. The RF signal is introduced using one transformer, and the LO signal is introduced using the other. As in the case of the single-balanced diode mixer, the LO signal is assumed to be larger in amplitude than the RF signal and controls the on-off cycle of the diodes. Only two of the four diodes conduct at a time for each half of the LO cycle. The result is a chopped waveform with no dc component.

Figure 8.17(b) illustrates the frequency spectrum produced by the double-balanced diode mixer. The output spectrum for the double-balanced diode mixer will

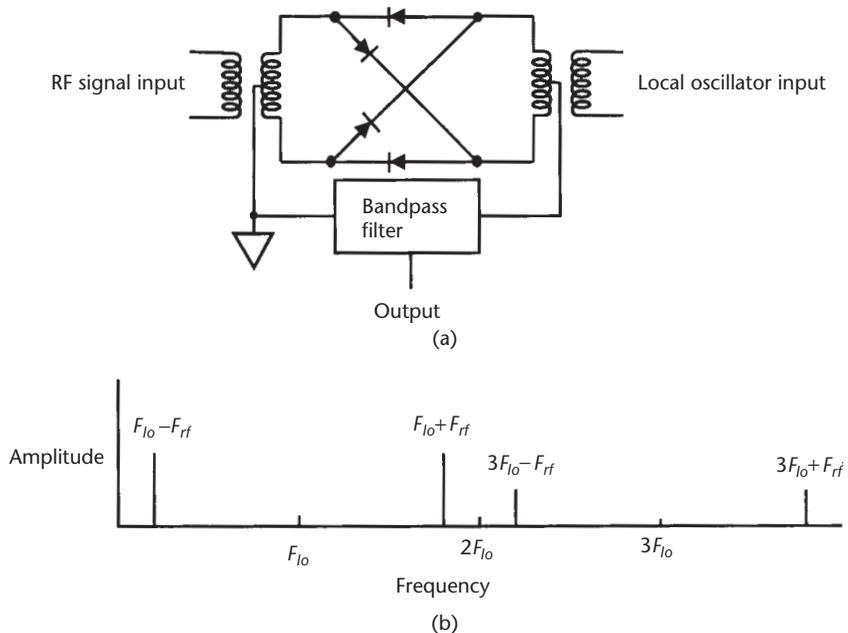


Figure 8.17 Double-balanced diode mixer: (a) double-balanced diode mixer circuit and (b) partial frequency spectrum for a double-balance diode mixer. (After: [6].)

contain only the frequencies $nF_{lo} \pm F_{rf}$, with n odd. Neither F_{lo} nor F_{rf} appears in the output (theoretically). That is an important advantage for this type of mixer. A bandpass filter is used to select the desired frequency components from the mixer. If ferrite toroidal-core transformers are used for lower frequency realizations, bandwidths of 400:1 can be achieved. Microwave mixers have smaller frequency ranges. For example, a low-frequency system might cover the 1–400-MHz range. For example, a microwave system might cover the 2–26-GHz range. These mixers typically have a conversion loss of about 7 dB and a SSB noise figure within 1 dB of the conversion loss. Frequency coverage is available from as low as 0.02 MHz to as high as 40 GHz. The isolation of the LO from the RF port is around 40 dB, decreasing at higher frequencies. The two-tone third-order intermodulation products typically are down 40 to 50 dB from the desired IF components.

The following information is adapted from a technical note in the *ANZAC RF and Microwave Signal Processing Components Catalog* [9].

Diode-type double-balanced mixers, as shown in Figure 8.17, belong to the general classification of resistive switching mixers, wherein an LO input signal is applied that is sufficiently large to cause strong conduction of the alternate diode pairs, thereby changing them from a low- to a high-resistance state during each half of the LO cycle. A virtual ground is, therefore, switched or commutated between the RF/IF transformer windings at a rate corresponding to the LO frequency. Since that switching causes a 180-degree phase reversal of the RF to IF port transmission during each half of the LO cycle, the mixing process is called biphas modulation.

For low-frequency operation, these devices typically use ferrite-core flux coupled transformers, which exhibit leakage inductance and stray capacitance. That limits upper frequency operation to approximately 4 GHz. For higher frequency operation, true transmission line realizations of the transformer functions will allow four-diode mixer operations to beyond 18 GHz. A system of this type is shown in Figure 8.18.

The low-frequency performance for the microwave mixer in Figure 8.18 is determined by the highpass nature of the RF and LO transmission line structure. A detailed discussion of transmission lines and terms is presented in Chapter 11.

Overlapping RF-IF or LO-IF frequency coverage is difficult to attain for the circuit in Figure 8.18 because the IF output encounters both the RF and LO structures

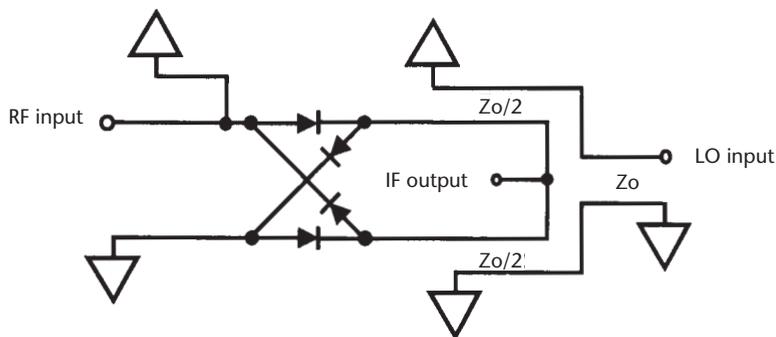


Figure 8.18 Typical four-diode microwave mixer schematic using transmission lines. (After: [9].)

in series for the IF signal path. To produce an overlapping IF range, a more complex eight-diode mixer was developed, as shown in Figure 8.19. Examination of this structure reveals that the LO is switching two diode pairs at a time, which are in series with the RF-IF signal path. By tracing out the RF-to-IF signal connections for each half of the LO input cycle, we see that biphase modulation is again being performed. The IF port can be seen to be an RF and LO null. The principle advantage of this design is its large RF-IF frequency range overlap. Its disadvantage is that it uses twice as many diodes and requires 3 dB more LO drive.

Figure 8.20 is a schematic diagram of a termination insensitive mixer (TIM), which consists of a transmission line hybrid network driving two sets of diodes. Isolation between each hybrid's opposite ports allows the LO to control independently the switching action of alternately conducting diode sets. The reverse bias applied to the off diodes is determined only by available LO input power and not by the diode's forward potential, as in the conventional ring-type mixers. An internal resistor absorbs mixer-generated, even-order LO frequency terms and improves LO VSWR by terminating the hybrid port opposite its LO input. This circuit feature improves performance by closely approximating a square-wave LO drive.

Other types of microwave mixers use hybrid junctions. A balanced mixer using a 3-dB quadrature (90-degree) coupler is shown in Figure 8.21. Other balanced mixers use the 180-degree "magic T." If all conditions were perfect, there would be little difference between the two types of hybrids. In practice, it is virtually impossible to match the mixer diodes perfectly at all frequencies and under all LO drive conditions. Therefore, some degradation in the hybrid performance can be expected as a result of reflections from the diode mounts. In the case of the 90-degree hybrid mixer, input VSWR at either port generally will be low since the reflections from the mixer diodes will be shunted out the opposite port. The isolation between the signal

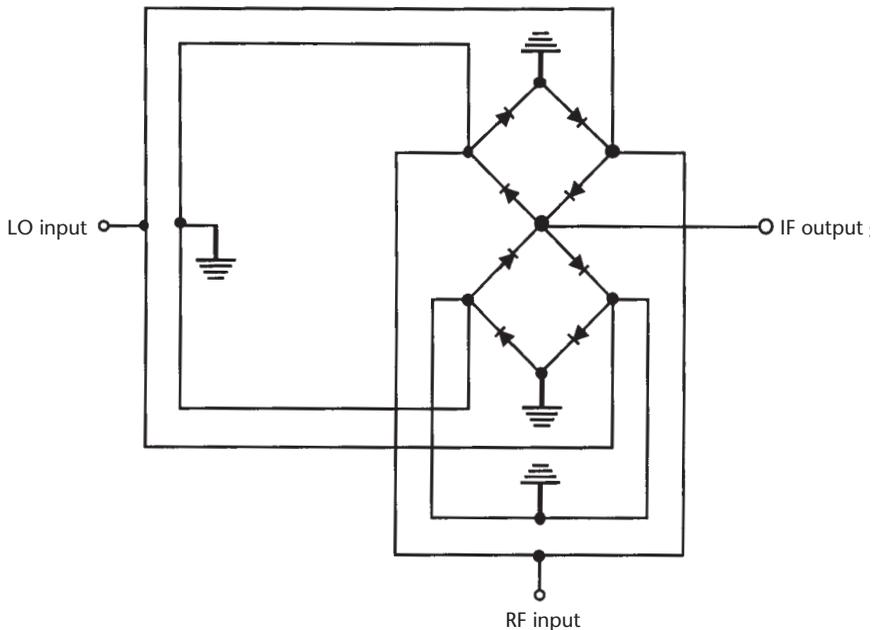


Figure 8.19 Typical eight-diode microwave mixer schematic. (After: [9].)

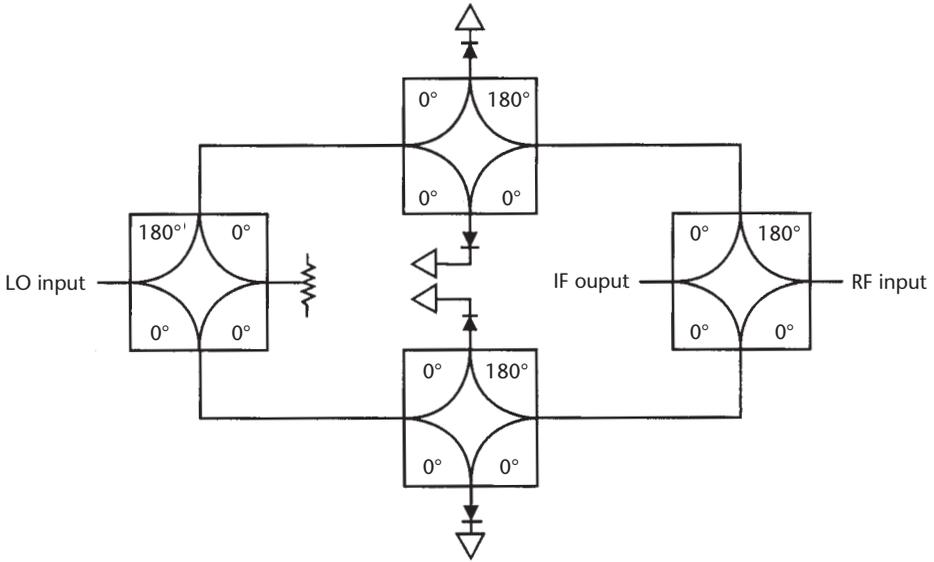


Figure 8.20 TIM schematic. (After: [9].)

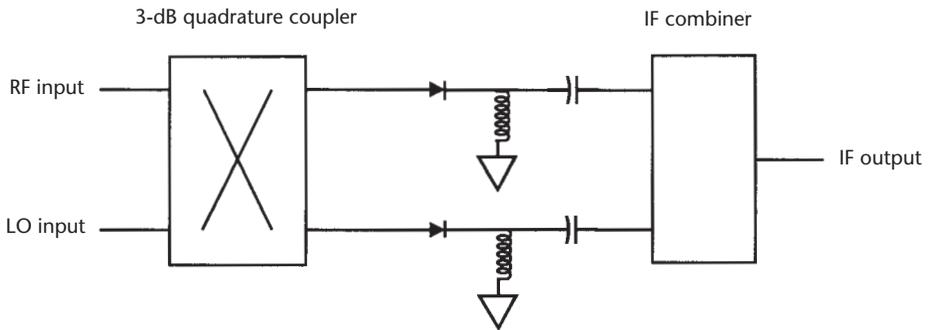


Figure 8.21 Balanced mixer using a 90-degree hybrid coupler and two diodes.

port and the LO port will be strictly a function of the return loss of the diode mount. Despite the fact that the isolation is low, the AM noise cancellation, which is a function of the amplitude and phase balance of the hybrid itself and not the return loss of the diodes, generally is as high with a 90-degree type hybrid as with the magic-T type hybrid. Those factors, combined with the ability to operate over octave and multioctave bandwidths with ease, make the 90-degree hybrid the most frequent choice for broadband microwave mixers [10].

The performance characteristics for a few diode mixers, as listed in the ANZAC catalog, are shown in Table 8.2 to illustrate mixer performance potential.

8.4.2 Transistor Mixers

It clearly is possible to make transistor mixers using either BJT or FET devices. These can be single-ended mixers, balance mixers, and double-balanced mixers.

Table 8.2 Some Performance Characteristics for Mixers

Characteristics	Standard RF Mixers			
	MD-108	MD-150	MD-162	MD-164
Model number	MD-108	MD-150	MD-162	MD-164
Frequency range				
RF/LO (MHz)	5–500	700–2,000	1,000–7,000	500–9,000
IF (MHz)	DC–500	DC–300	10–2,000	10–2,000
Conversion loss (dB typical)	5.6	6.2	6.0	6.5
Isolation				
LO-RF (dB typical)	45	35	25	22
LO-IF (dB typical)	40	20	20	27
RF-IF (dB typical)	25	24	22	25
LO drive (dBm)	+7	+7	+13	+13

Figure 8.22 shows two examples of single-ended FET mixers. The circuit in Figure 8.22(a) adds the LO signal and the RF signal at the gate input to the mixer. The circuit in Figure 8.22(b) has the LO applied to the source terminal of the FET and the RF signal applied to the gate terminal of the FET.

One advantage for transistor mixers is that they can provide conversion gain rather than conversion loss. BJTs can have conversion gains on the order of 20 dB, while FETs can have conversion gains on the order of 10 dB. FETs produce less intermodulation and cross-modulation distortion; for that reason, they are preferred over BJTs for high-frequency mixers. Both JFETs and MOSFETs are used, with MOSFETs generally exhibiting higher power gain.

MESFETs are used at microwave frequencies for transistor mixers.

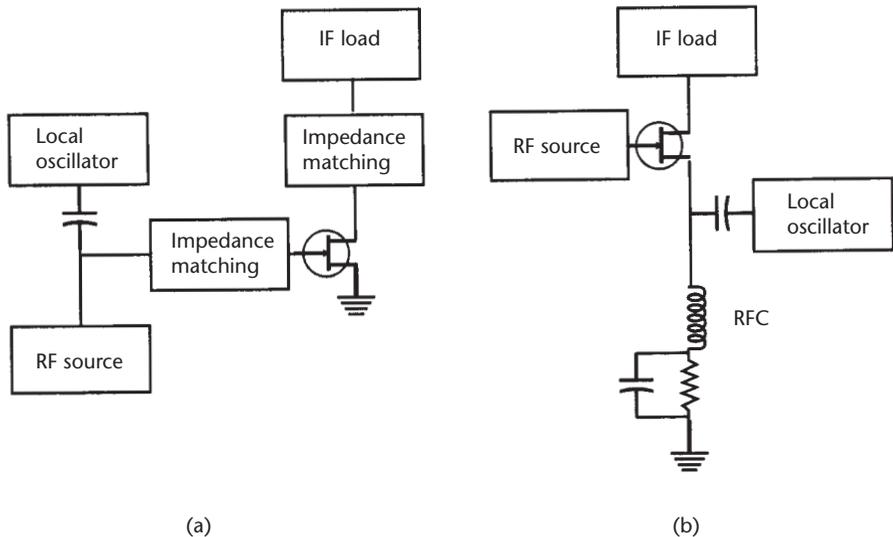


Figure 8.22 Example FET mixers: (a) single-gate-input-type mixer and (b) gate and source input type FET mixer.

References

- [1] Krauss, H. L., C. W. Bostian, and F. H. Raab, *Solid State Radio Engineering*, New York: John Wiley & Sons, 1980, Chaps. 4, 12, and 13.
- [2] Dye, N., and H. Granberg, *Radio Frequency Transistors: Principles and Practical Applications*, Boston, MA: Butterworth-Heinemann, 1993, Chap. 1.
- [3] Sivan, L., *Microwave Tube Transmitters*, London, England: Chapman and Hall, 1994.
- [4] Whitaker, J. C., *Power Vacuum Tubes Handbook*, New York: Van Nostrand Reinhold, 1994.
- [5] Krauss, H. L., C. W. Bostian, and F. H. Raab, *Solid State Radio Engineering*, New York: John Wiley & Sons, 1980, Chap. 5.
- [6] Krauss, H. L., C. W. Bostian, and F. H. Raab, *Solid State Radio Engineering*, New York: John Wiley & Sons, 1980, Chaps. 6 and 7.
- [7] Kennedy, G., *Electronic Communication Systems*, 3rd ed., New York: McGraw-Hill, 1985, Chap. 12.
- [8] Kennedy, G., *Electronic Communication Systems*, 3rd ed., New York: McGraw-Hill, 1985, pp. 341–343.
- [9] 1989 ANZAC RF and Microwave Signal Processing Components Catalog, Adams Russell Components Group, Burlington, MA, pp. 234–241.
- [10] Howe, H., Jr., *Stripline Circuit Design*, Dedham, MA: Artech House, 1974, pp. 268–269.

Selected Bibliography

- Maas, Stephen A., *Microwave Mixers*, Norwood, MA: Artech House, 1986.
- Medley, M. W., Jr., *Microwave and RF Circuits: Analysis, Synthesis, and Design*, Norwood, MA: Artech House, 1993.
- Vendelin, G. D., *Design of Amplifiers and Oscillators by the S-Parameter Method*, New York: John Wiley & Sons, 1982.
- Vendelin, G. D., A. M. Pavio, and U. L. Rohde, *Microwave Circuit Design Using Linear and Nonlinear Techniques*, New York: John Wiley & Sons, 1990.

Modulators and Demodulators

This chapter discusses two additional major devices used in communications, radar, and other RF systems: modulators and demodulators. Modulators are used in transmitter systems, demodulators in receivers. Demodulation devices are sometimes called detectors.

9.1 Modulators

This section discusses some of the more important types of modulator systems, including AM, FM, and phase modulators. Some of the information presented here is based on [1–3].

9.1.1 Modulators for Conventional Amplitude Modulation

9.1.1.1 Plate-Modulated Class C Amplifiers

Vacuum tube power amplifier stages are used predominantly for high-power AM modulators at VHF and lower frequencies. That is because of the need for transmitter power greater than that possible with transistor amplifiers. Tube types include triodes and tetrodes. A triode has a cathode for emitting electrons, a single control grid for controlling the flow of electrons, and a plate for collecting electrons. Thus, it has three electrodes. A tetrode has those same elements plus a second grid, called a screen grid, between the control grid and the plate. Tetrodes have the advantage that the screen grid greatly reduces coupling between the plate and the grid circuit, thereby reducing chances for oscillation and distortion and eliminating the need for a neutralizing feedback circuit. Power outputs for standard AM broadcast at MF typically are in the range of 10–100 kW. Thus, transmitter systems of this type require high-power audio frequency (AF) modulators, large dc power supplies, and large cooling systems. Because efficiency is important, class C operation normally is used for the output stage.

Figure 9.1 is a block diagram of a high-power AM transmitter system used for standard AM broadcast at MF. frequencies. The modulator is a plate-modulated class C RF power amplifier that uses a high-power triode or tetrode vacuum tube as the output stage. The output stage has a high-power audio transformer in series with the high-voltage supply. That permits the plate voltage to the amplifier to be modulated up and down in amplitude in accordance with the input signal.

The signal source for this transformer is an amplifier chain consisting of a first-stage AF processing and filtering circuit, an AF preamplifier, an AF class B

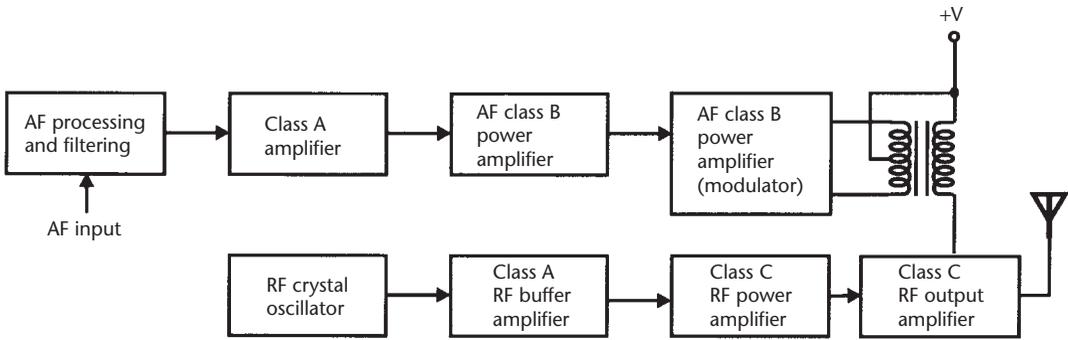


Figure 9.1 Block diagram of an AM transmitter.

power amplifier, and a high-power class B modulator amplifier. Up to one-third of the total output power for the system must be supplied by the high-power class B modulator amplifier.

In Figure 9.1, the carrier signal source starts with a crystal oscillator, which generates a CW signal at the desired RF frequency. That is followed by a class A RF buffer amplifier and a class C RF power amplifier. The RF carrier signal is fed to the control grid of the modulated class C power amplifier by means of a transformer with the appropriate signal level and bias so the amplifier operates in class C. A tuned output transformer is used to convert the pulse-type signal from the plate-modulated class C RF output amplifier to a modulated sine-wave output.

9.1.1.2 Grid-Modulated Class C Amplifiers

Another type of amplitude modulator sometimes used for high-power transmitters is a grid-modulated class C amplifier. In a grid-modulated triode class C amplifier, the RF input is provided by one transformer and the AF input by a second transformer connected in series with the first.

The grid of the triode is biased negatively, and a neutralizing feedback signal is provided to it from the output transformer. The output transformer is a single-tuned transformer with the primary tuned to the desired output frequency. The secondary typically is untuned.

The triode current is a series of current pulses corresponding in frequency to the RF input signal and in amplitude to the AF input. The tuned output transformer converts the signal to a sine-wave signal by means of the so-called fly-wheel action of the parallel resonant circuit.

9.1.1.3 Collector-Modulated Class C Transistor Amplifiers

Transistors also are used in AM class C power amplifiers. Figure 9.2 illustrates the case of a collector-modulated class C push-pull transistor amplifier. Such amplifiers can be used for AM for lower power applications. In the system shown in Figure 9.2, the modulating signal is fed to the primary winding of an untuned AF transformer. The secondary winding of the transformer is in series with the V_{cc} power supply. It is connected to the center tap of the primary winding of the output transformer with

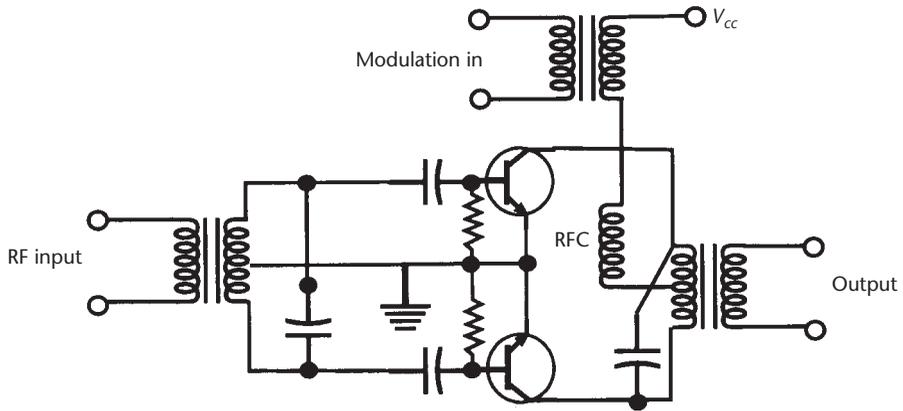


Figure 9.2 Collector-modulated class C transistor push-pull amplifier. (After: [2].)

the use of an RF choke. The voltage applied to the transistor collectors moves up and down in response to the modulation input. The RF signal is provided to the bases of the BJT devices with a tuned center-tapped RF transformer. The base of each transistor is biased to ground. With the bias set at 0V, the operation of the transistor amplifiers is large-angle class C.

9.1.1.4 Other Types of Conventional Amplitude Modulators

Other types of conventional amplitude modulators occasionally are used, including FET gate-biased AM systems in which the AF signal provides modulation of the gate bias for a class C FET power amplifier. This type of system typically uses an RF choke for dc power input and a capacitively coupled-tuned, parallel-resonant, output circuit. FETs also can be used in push-pull modulator configurations.

9.1.2 Modulators for Double-Sideband Modulation

Figure 9.3 is a simplified system block diagram for a DSB transmitter system. The first RF frequency is provided by a crystal oscillator. The low-power signal is fed to a balanced modulator through a buffer amplifier. A typical frequency for the signal might be in the range of 10–70 MHz.

The second signal to the balanced modulator is the input modulating signal, which is fed to the modulator through a buffer amplifier. The signal is assumed to be an audio, a video, or a digital signal. A balanced modulator is the same circuit as a balanced mixer, discussed in Chapter 8. The output of the balanced modulator is a set of sidebands on either side of the carrier, with the carrier having very low amplitude. The desired set of sidebands is selected by an output filter. An option might be to inject a small amount of carrier signal through an add circuit following the balanced modulator to make it easier for the receiver to lock on to the carrier frequency. A typical transmitter output frequency might be in the VHF, UHF, or microwave frequency range. The output of the modulator is then passed through an upconverter, which converts the signal to the desired output frequency.

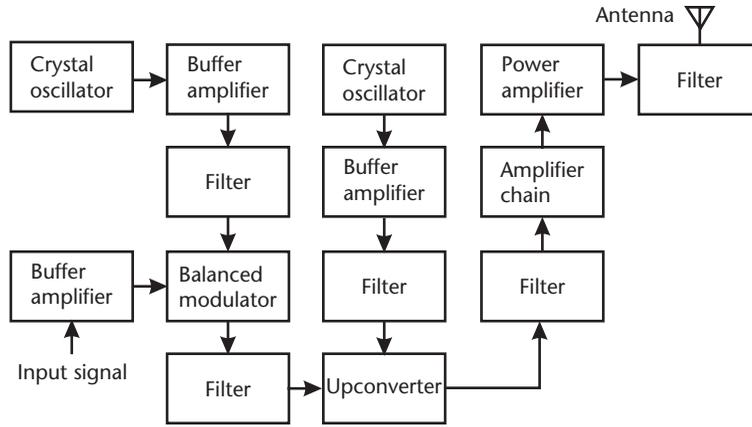


Figure 9.3 DSB transmitter system.

The next stage is an amplifier chain that increases the power level of the upconverted signal to the desired drive level for the power amplifier. Amplifiers for the chain are class A and class B amplifiers.

The power amplifier can be a solid-state amplifier, a triode or tetrode, or a microwave tube power amplifier such as a klystron or a traveling-wave tube. Output to the antenna is through a suitable lowpass or bandpass filter.

Figure 9.4 shows a balanced modulator and the associated frequency spectrum for a DSB system. For simplicity, a single-frequency sine wave is assumed for the input modulating signal. The passband of the modulator output filter is indicated on the spectrum diagram.

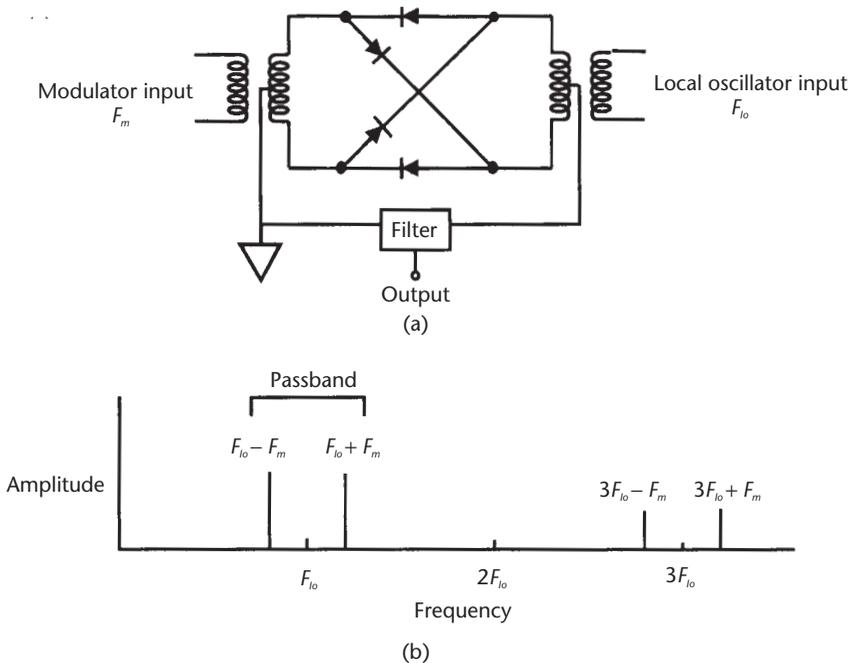


Figure 9.4 A balance modulator for a DSB system: (a) schematic diagram of balanced modulator and (b) frequency spectrum for DSB modulation.

9.1.3 Vestigial-Sideband Modulators

Figure 9.5 shows a VSB modulator. The block diagram looks like that of the DSB modulator. The difference is that the modulator includes a balanced modulator with sideband filter and reduced carrier injection. The sideband filter removes only part of one sideband. The remainder is sent along with the reduced carrier and the other full sideband. The upconverter, amplifiers, power amplifier, and output filter are the same as those described for DSB modulation.

9.1.4 Modulators for Single-Sideband Modulation

Figure 9.6 shows an SSB transmitter. The block diagram looks like that of the DSB transmitter. The difference is that the modulator includes a balanced modulator with either a complete sideband filter or a balanced modulator with SSB by a phase-shift method. The upconverter, amplifiers, power amplifier, and output filter are the same those described for DSB and VSB modulation.

Figure 9.7 shows a balanced modulator and associated frequency spectrum used for SSB modulation. The modulating signal is assumed to be a single-frequency sine wave. The passband for the output filter is indicated on the diagram for the upper sideband. The lower sideband also could be selected, if desired, rather than the upper sideband.

The sideband-suppression filter must have very sharp cutoff characteristics, and the IF must be quite low for most SSB applications. In a typical example, the filter’s response must change from near zero attenuation to near full (30 dB or more) attenuation over a range of only 600 Hz. To obtain a filter response curve with skirts as steep as those suggested, the Q of the filter (reactance/resistance) must be very high. Possible filter types include LC filters, crystal filters, ceramic filters, mechanical filters, and SAW filters. Because of Q limitations, LC filters cannot be used for IF values greater than about 100 kHz. Mechanical filters have been used at frequencies up to 500 kHz and crystal filters and ceramic filters up to about 30 MHz. SAW filters can be used up to 2 GHz.

The phase-shift method of producing SSB suppressed-carrier signal is shown in Figure 9.8. This method avoids filters and some of their attendant disadvantages.

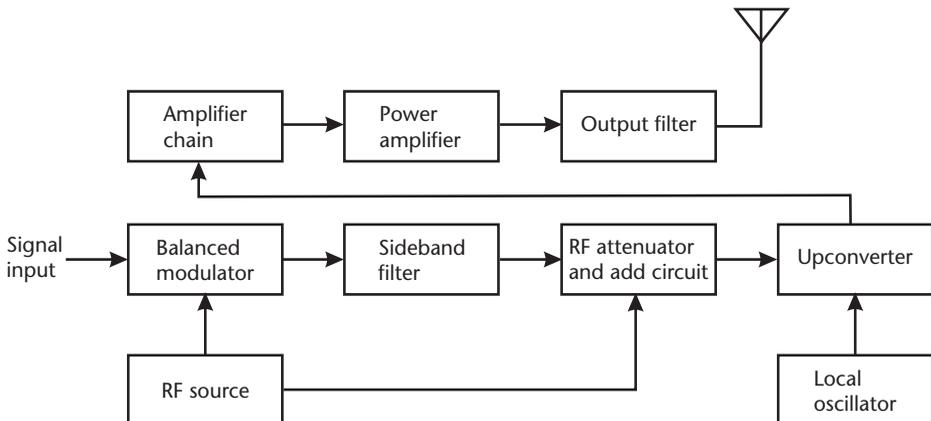


Figure 9.5 Block diagram of a VSB modulation system.

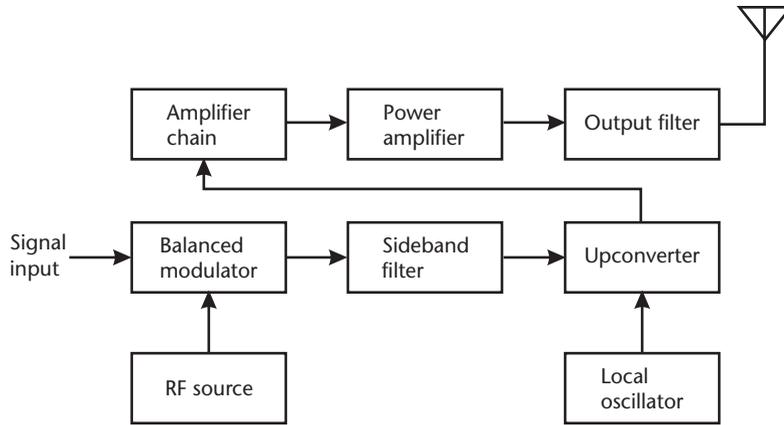


Figure 9.6 Block diagram of a transmitter using the filter method of SSB modulation.

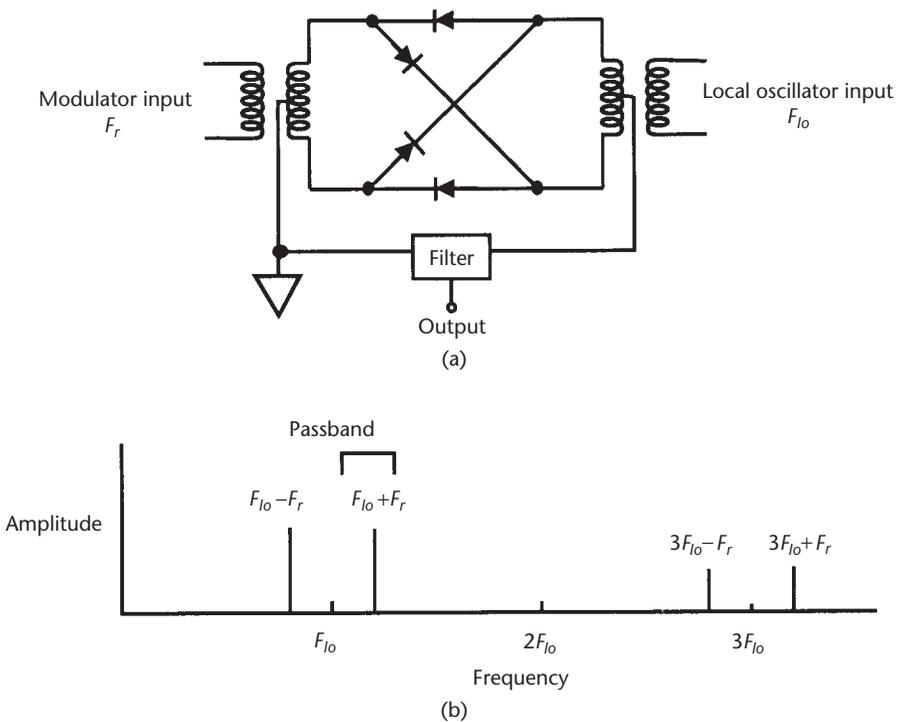


Figure 9.7 Example of a balance modulator: (a) balanced modulator circuit and (b) sample frequency spectrum for SSB suppressed carrier modulation.

The audio input signal is applied to two all-pass networks with phase shifts that differ by 90 degrees over the frequency range of interest. The signals are then applied to two balanced modulators along with in-phase and quadrature (90-degree out of phase) signals of the desired RF frequency. The in-phase and quadrature signals can be obtained by digital frequency division of the output of a variable-frequency oscillator operating at four times the output frequency. Not shown in Figure 9.8 is a gating circuit that sets the initial state of the flip-flops and thus determines which

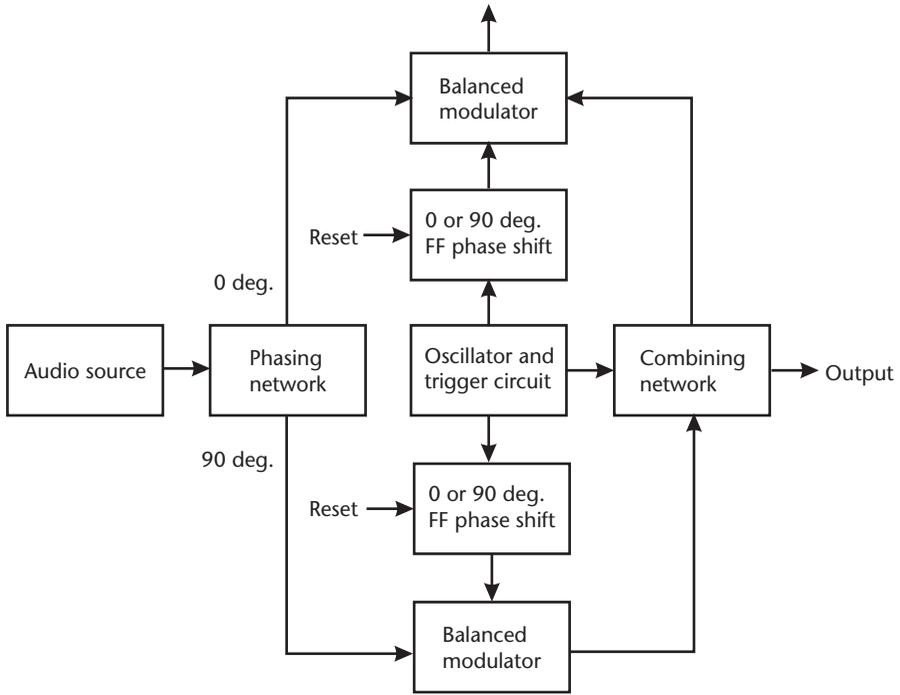


Figure 9.8 Phase-shift method of producing SSB suppressed-carrier modulation. (After: [1].)

sideband is produced. The outputs of the two balance modulators are summed and then amplified to the desired level.

The operation of an SSB modulator that uses the phase shift method is demonstrated as follows:

The equation of a wave with the carrier removed is

$$e_1 = mE \sin \omega_m t \sin \omega_c t \tag{9.1}$$

This is the case for the output of the first modulator. When both modulating and carrier frequencies are shifted 90 degrees, as in the case of the second balanced modulator, the equation for a wave with the carrier removed is

$$e_2 = mE \cos \omega_m t \cos \omega_c t \tag{9.2}$$

Adding (9.1) and (9.2) gives

$$\begin{aligned} e_1 + e_2 &= mE \sin \omega_m t \sin \omega_c t + mE \cos \omega_m t \cos \omega_c t \\ &= mE/2 [\cos(\omega_c - \omega_m)t - \cos(\omega_c + \omega_m)t] \\ &+ mE/2 [\cos(\omega_c - \omega_m)t + \cos(\omega_c + \omega_m)t] \\ &= mE [\cos(\omega_c - \omega_m)t] \end{aligned} \tag{9.3}$$

Equation (9.3) corresponds to the equation of the lower sideband. If the polarity of one of the modulating voltages or one of the RF voltages is reversed, the other sideband would appear at the output terminals.

Possible variations of SSB are SSB with full carrier and SSB with reduced carrier. The carrier can be added after generation of the SSB signal.

9.1.5 Modulators for Frequency-Division Multiplex

Figure 9.9 is a simplified block diagram of an FDM modulator system, used to send many signals in parallel by using different frequencies for each channel. In this system, the signals are sent using SSB suppressed-carrier modulation with the filter method for generation of SSB. Each channel has its own amplifier, oscillator, balanced modulator, and sideband filter. Each generated SSB signal is added along with a pilot signal and passed through a group bandpass filter. Table 9.1 lists oscillator frequencies and SSB modulator output frequencies for the 12 voice channels and pilot channel.

The next step up from a group is the basic supergroup, which consists of five groups. By proper choice of upconverter oscillator inputs, these groups occupy frequency channels as shown in Table 9.2.

These five groups are added along with a pilot channel at 547.92 kHz. Supergroups can be combined to form mastergroups, supermastergroups, and so on. The total set of sideband frequencies then can be shifted up in frequency to the desired transmit frequency.

9.1.6 Modulators for Standard Frequency Modulation

Frequency-modulated signals often are produced at low power levels and amplified by amplifier chains. The modulation can be accomplished either directly by variation of the frequency of an oscillator by the input signal or indirectly by phase

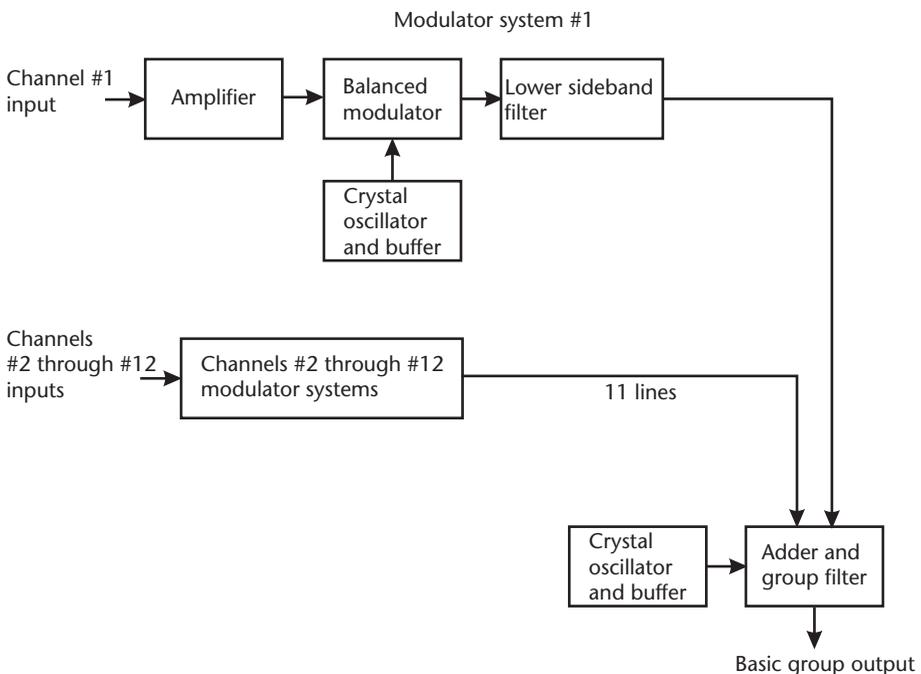


Figure 9.9 Basic 12-channel group translating equipment for FDM. (After: [4].)

Table 9.1 Frequencies for the 12 Voice Channels and Pilot Channel

<i>Channel Number</i>	<i>Crystal Oscillator Output Frequency (kHz)</i>	<i>SSB Modulator Output Frequency Band (kHz)</i>
1	108	104.6–107.7
2	104	100.6–103.7
3	100	96.6–99.7
4	96	92.6–95.7
5	92	88.6–91.7
6	88	84.6–87.7
7	84	80.6–83.7
8	80	76.6–79.7
9	76	72.6–75.7
10	72	68.6–71.7
11	68	64.6–67.7
12	64	60.6–63.7
Pilot channel	104.08	104.6–107.7

Table 9.2 Frequency Channels of Groups

<i>Group Number</i>	<i>Frequency Bands Used (kHz)</i>
1	312–360
2	360–408
3	408–456
4	456–504
5	504–552

modulation and other methods. To achieve good linearity, most frequency modulators produce a smaller modulation index or frequency deviation than is desired in the transmitter output. Frequency multipliers multiply frequency deviation and modulation index, as well as frequency. Thus, they often are used in the transmitter chain.

Figure 9.10 is an example of an FM modulator system. In this system, an AF source provides the modulating signal to an FM source consisting of a varactor-tuned crystal oscillator. The FM source has as its output a center frequency f with frequency deviation df . The signal is fed to the frequency multiplier having a multiplication factor of k . The output of this stage has a center frequency of kf and a deviation of $k df$. That signal is then upconverted to the desired transmit frequency and amplified with an amplifier chain and a power amplifier. The output then is filtered and fed to the antenna.

The varactor-tuned crystal oscillator in Figure 9.10 is a direct FM modulator, a system frequently used in portable and mobile transmitters. Frequency multiplication often is used with this system because of the small frequency deviation that is possible.

Another FM modulator is the phase-shift modulator. The phase-shift method of producing FM is shown in Figure 9.11.

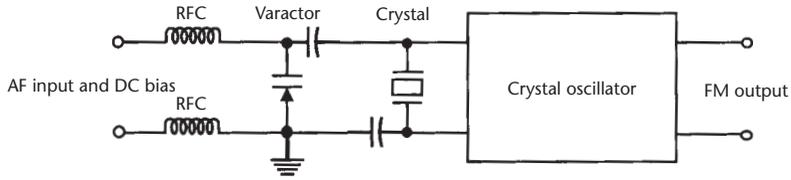


Figure 9.10 Varactor-tuned crystal oscillator FM modulator.

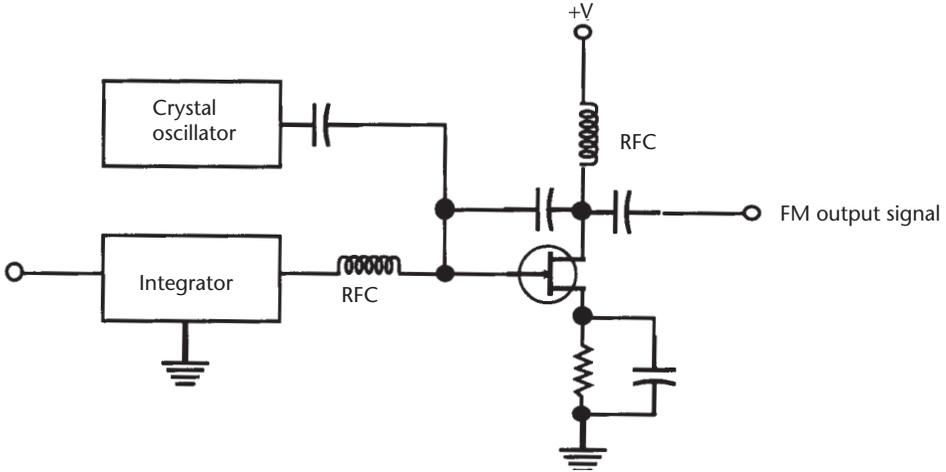


Figure 9.11 Phase-shift method of producing FM.

The phase-shift modulator in Figure 9.11 is an indirect FM modulator. This system uses a controllable conductance in combination with a fixed reactance to vary the phase delay of signals. Since this modulator directly varies the phase of the signal rather than the frequency, it is necessary to integrate the input AF signal before the modulator. The PLL method for producing FM is shown in Figure 9.12. This is also an indirect FM modulator. Again, the system produces phase modulation rather than FM, and it is necessary to integrate the AF signal before modulation. The signal is introduced into the system as an error voltage. The system responds to the injected error voltage by adjusting the VCO to produce the specified phase shift between its output and the reference signal.

It is common practice with FM modulator and receiver systems to use pre-emphasis and deemphasis as a means of improving the link signal-to-noise ratio. Noise has a greater effect on the higher modulating frequencies than on the lower ones. If the higher frequencies are artificially boosted in amplitude before modulation, noise will have less effect. This process is known as preemphasis. To recover the audio signal in its original form, we must reduce the amplitude of the higher frequencies at the receiver to the same degree they were boosted. That process is known as deemphasis. The net result of using this combination is that the signal-to-noise ratio is improved for the higher frequencies.

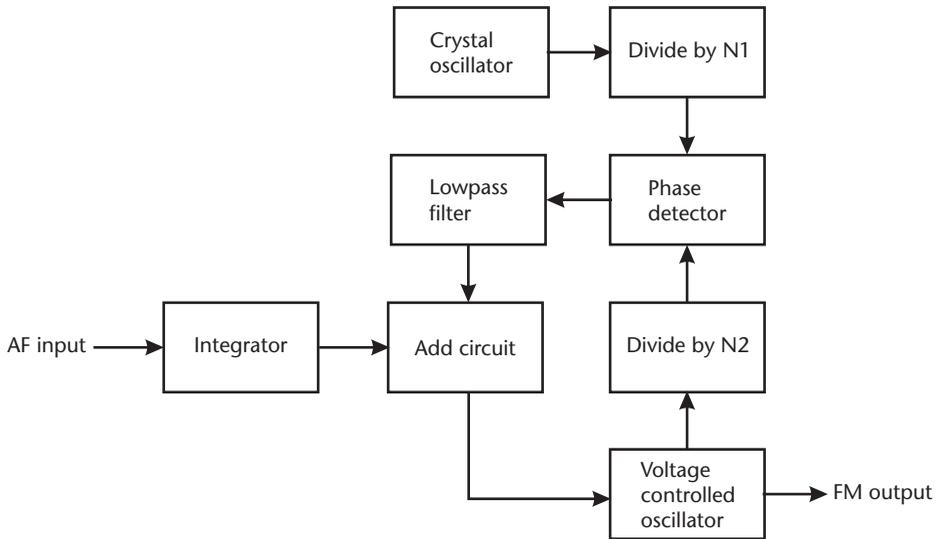


Figure 9.12 PLL method for producing standard FM.

9.1.7 Modulators for Frequency-Shift Keying

Frequency-shift keying (FSK) modulators work the same as standard frequency modulators except that only two frequency states are used. This type of modulation frequently is used for digital data transmission.

9.1.8 Modulators for Phase-Shift Keying

Modulators for phase-shift keying (PSK) were introduced in Chapter 7. Some of the information presented here is adapted from [3].

9.1.8.1 Binary Phase-Shift Keying

With binary phase-shift keying (BPSK), the signal phase is keyed between two phase states 180 degrees apart. Each pulse represents just one bit of information (1 or 0). An example of a modulator for such a system is shown in Figure 9.13.

In the system in Figure 9.13, a balanced modulator is used with one input from an LO at an assumed frequency of 70 MHz and the other input from a two-level amplitude modulator. The amplitude modulator is controlled by a logic circuit. The output of the balanced modulator normally is fed to an amplifier.

9.1.8.2 Quaternary Phase-Shift Keying

Four-phase states are used with quaternary phase-shift keying (QPSK) with phase states separated by 90 degrees. Each pulse represents two bits of information (00, 01, 10, or 11). An example of such a modulator is shown in Figure 9.14. In this system, an LO at an assumed frequency of 70 MHz feeds a 90-degree phase splitter. The outputs of the phase splitter, in-phase and quadrature signals, are fed to two balanced modulators, which are also fed by two two-level amplitude modulators.

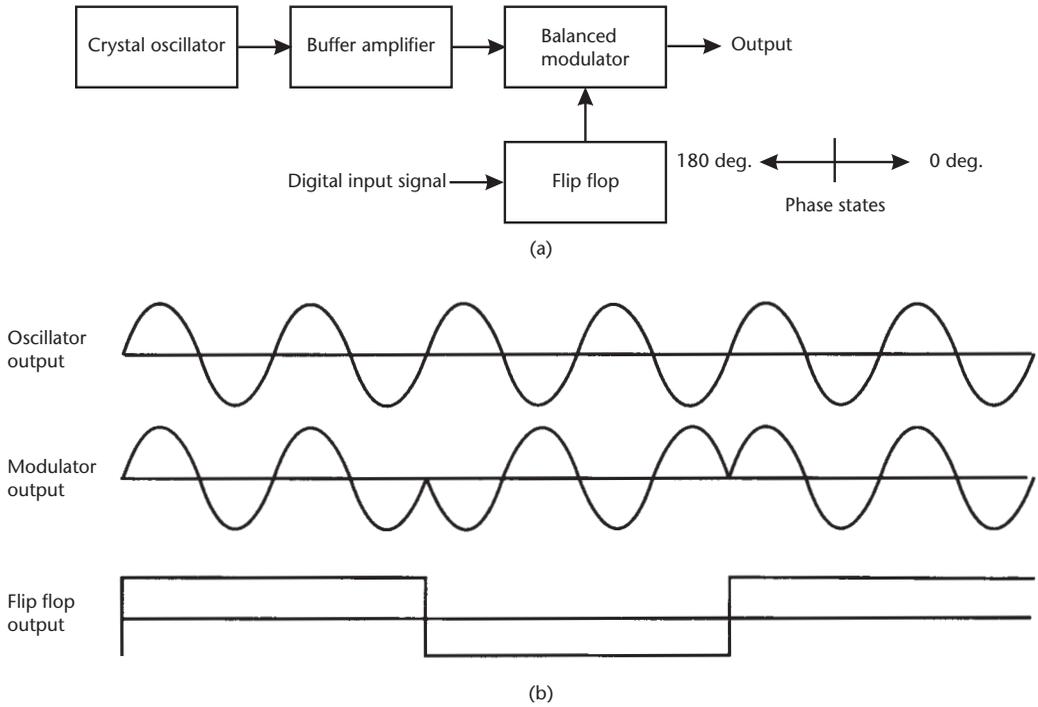


Figure 9.13 BPSK modulator: (a) circuit diagram and (b) sample waveforms.

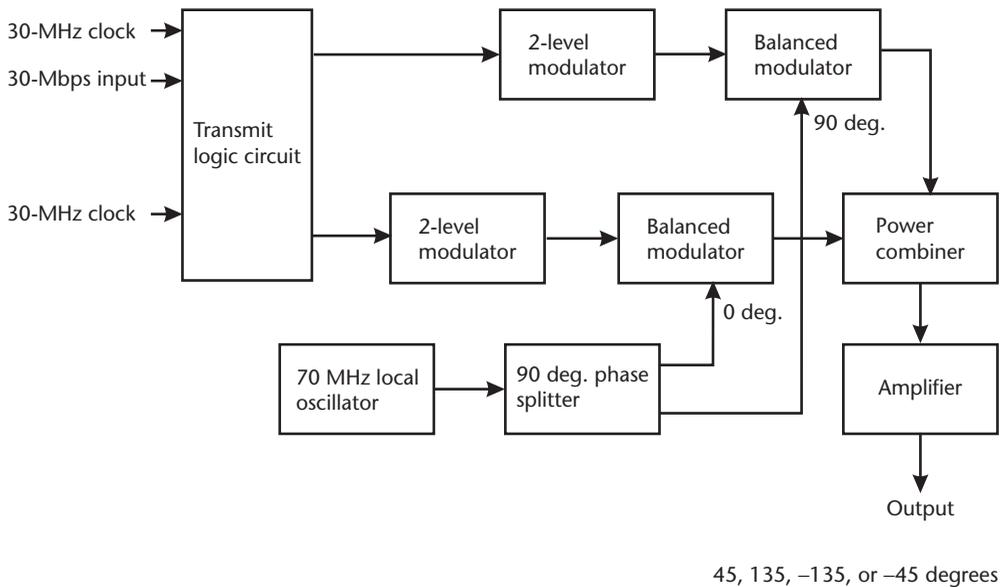


Figure 9.14 QPSK modulator. (After: [3].)

The amplitude modulators are controlled by a logic circuit. The outputs of the balanced modulators are fed to an in-phase power combiner circuit. The four possible vectors of +45, +135, -45, and -135 degrees are shown in Figure 9.14.

9.1.8.3 $\pi/4$ -DQPSK Modulation

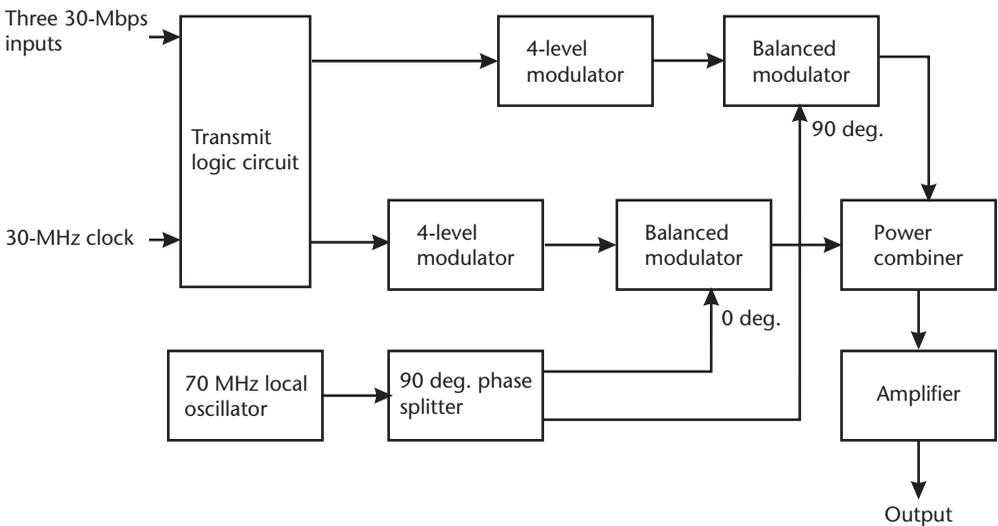
An important variation of QPSK modulation is $\pi/4$ -DQPSK modulation. This is the Telecommunication Industries Association (TIA) standard for digital cellular telephones. The D in DQPSK indicates that the four-state phase modulation is a differential phase modulation in which the phase reference is derived from the previous RF pulse. Information is sent by changes in phase from pulse to pulse rather than in absolute phase. The $\pi/4$ term means that the minimum phase-shift with respect to the reference is 45 degrees. Possible differential phase states thus are +45, +135, -135, and -45 degrees.

9.1.8.4 Eight-State Phase-Shift Keying

Eight phase states are used with eight-state phase-shift keying (8-PSK) with the phase states separated by 45 degrees. Each pulse represents three bits of information. An example of a modulator for such a system is shown in Figure 9.15. In this system, an LO at an assumed frequency of 70 MHz feeds a quadrature phase power divider. The outputs of the phase splitter are fed to two balanced modulators, which also are fed by two four-level amplitude modulators. The amplitude modulators are controlled by a logic circuit. The outputs of the balanced modulators are fed to a power combiner circuit. The output of that circuit is fed to an amplifier. The eight possible vector angles are +22.5, +67.5, +112.5, +157.5, -22.5, -67.5, -112.5, and -157.5 degrees.

9.1.8.5 16-Quadrature Amplitude Modulation

It is possible to use 16-PSK, providing four bits per pulse capability, but this type of system currently is not used because another 16-state keying system is better. The



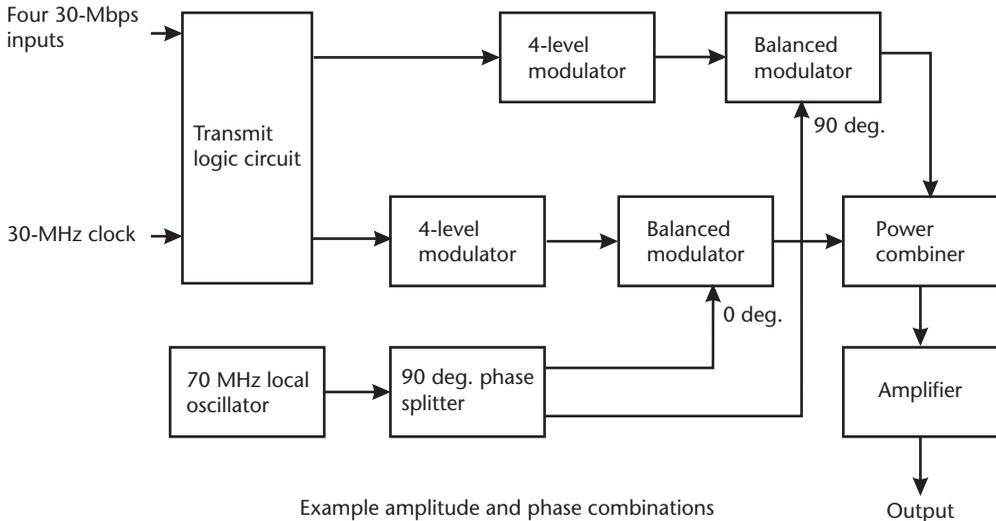
22.5, 67.5, 112.5, -22.5, -67.5, -112.5, or -157.5 degrees

Figure 9.15 8-PSK modulator. (After: [3].)

improved system is 16-QAM. This system uses four phase states and four possible amplitudes. This 4×4 combination results in 16 possible states. Figure 9.16 shows a block diagram for a 16-QAM system and lists 16 possible amplitude and phase combinations. Recently, 16-QAM modulation has been used extensively in microwave relay systems.

9.1.9 Modulators for Pulse Code Modulation Time-Division Multiplex Modulation

A 24-channel pulse code modulation time-division multiplex (PCM TDM) modulator system is shown in Figure 9.17. The 24-channel groups have a sampling rate of 8,000 samples per second, 8 bits (256 sampling levels) per sample, and a pulse width of about 0.625 ms. The sampling interval is 125 μ s, and the period required for each pulse group is 5 ms. Each 125-ms frame is used to provide 24 adjacent channel time slots, with the twenty-fifth slot used for synchronization. Sampling is done simultaneously on each channel. Sampling circuits are followed by delay lines, each with a different delay. Delays are 0, 5 μ s, 10 μ s, 15 μ s, and so on, to the twenty-fourth channel, which is delayed 115 μ s. The outputs of the delays are added to provide a single bit stream. The process is repeated 8,000 times per second. The bit stream is used to frequency- or phase-modulate the carrier of the transmitter.



Example amplitude and phase combinations

	In-phase signal amplitude	Quadrature phase signal amplitude	In-phase signal amplitude	Quadrature phase signal amplitude
1	1	1	5	3
2	1	-1	6	3
3	1	3	7	3
4	1	-3	8	3

A total of 16 combinations of amplitude and phase

Figure 9.16 Block diagram of a 16-QAM modulation system. (After: [3].)

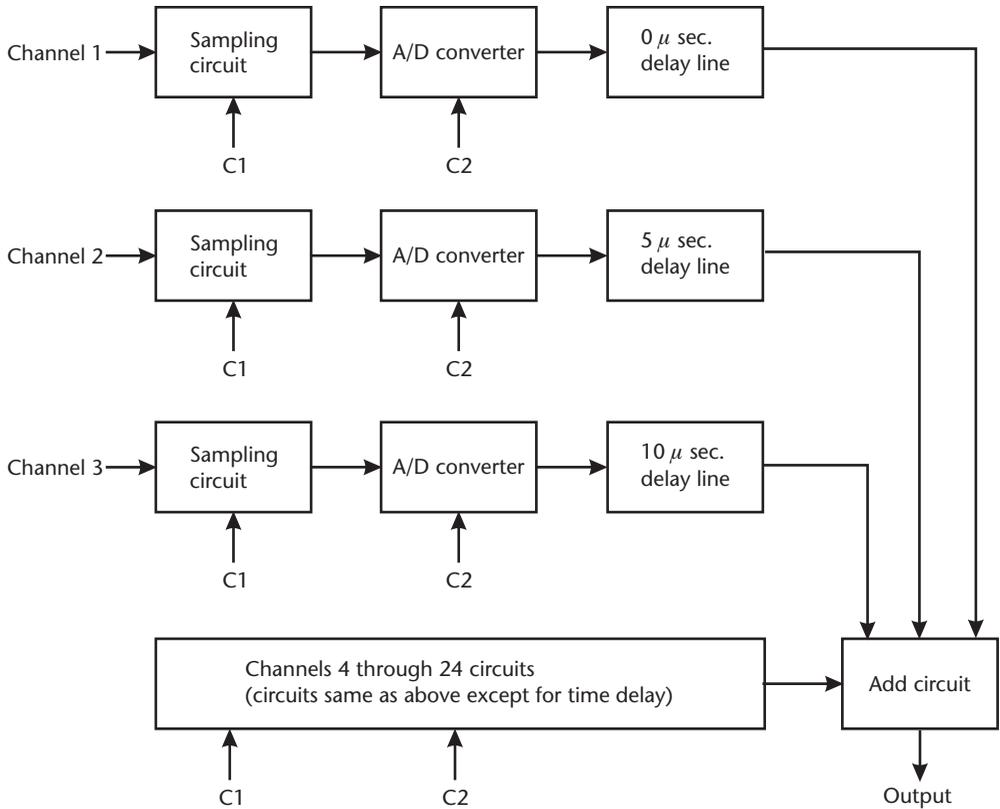


Figure 9.17 24-channel PCM TDM modulator system.

The second multiplex level, illustrated in Figure 9.18, provides 96 channels. The bit rate is 6.312 Mbps. Some of the bits are used for synchronization and others for housekeeping functions. The method of producing secondary multiplex levels consists essentially in dividing by 4 the pulse widths in the primary level signal and using the slots thus vacated to combine four primary streams, using delay lines and an adder to convert the four bit streams to one. The bit streams are then used to frequency- or phase-modulate the carrier of the transmitter.

9.1.10 Time-Division Multiple Access

Multiple access means that a multiplicity of participating signal sources can use the modulation system simultaneously by occupying time slots in a time frame. The TIA standard for North America digital cellular telephones uses TDMA. Details about this important system are presented in Chapter 1.

9.2 Demodulators or Detectors

This section discusses amplitude, frequency, and phase demodulators and detectors. These circuits follow the IF amplifiers and are used to recover the baseband signals. Some of the information presented in this section is adapted from [5, 6].

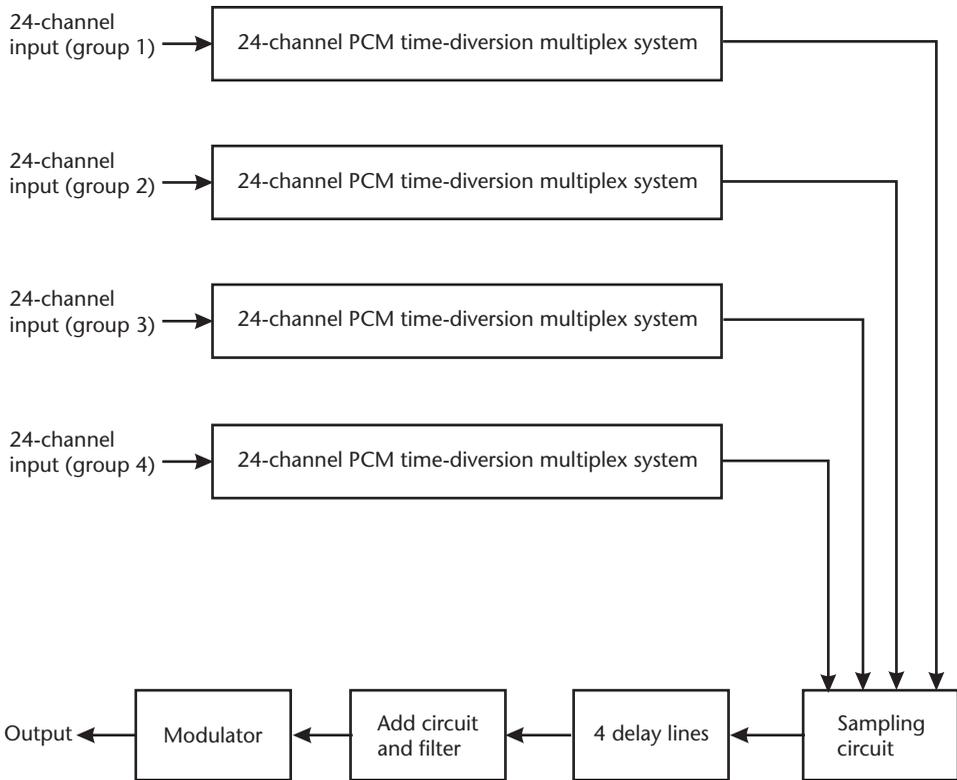


Figure 9.18 96-channel PCM TDM modulator system.

9.2.1 Amplitude Modulation Detectors

An AM detector is a circuit that converts an amplitude-modulated RF or IF signal to an audio, video, or pulse signal of the same form that was originally used to modulate the transmitted signal. One important type of AM detector is the envelope detector circuit. A simple form of that type of AM detector is shown in Figure 9.19(a). The circuit involves a diode rectifier followed by a lowpass filter. The rectifier converts the full-wave signal to a half-wave signal of either positive or negative polarity, depending on the direction of the diode. The lowpass filter is simply a resistor and capacitor in parallel following the rectifier. That allows for a fast rise time and a slow fall time, as desired. The combination makes the detector a peak-type detector, and the circuit output follows the envelope of the modulated signal. A form of distortion known as diagonal clipping results if the fall time is made too slow.

A slightly more complex envelope detector that is often used with AM receivers is shown in Figure 9.19(b). That circuit includes capability for providing volume control for the output signal and an AGC output to control the gain of the receiver.

9.2.2 Product Detectors

A product detector (Figure 9.20) is simply a balanced modulator or balanced mixer circuit, as discussed in Chapter 8. A product detector can be used to demodulate SSB signals, DSB signals, and standard AM signals. The mixer may be either a diode mixer or a transistor mixer. It has one input from the IF amplifier output and the

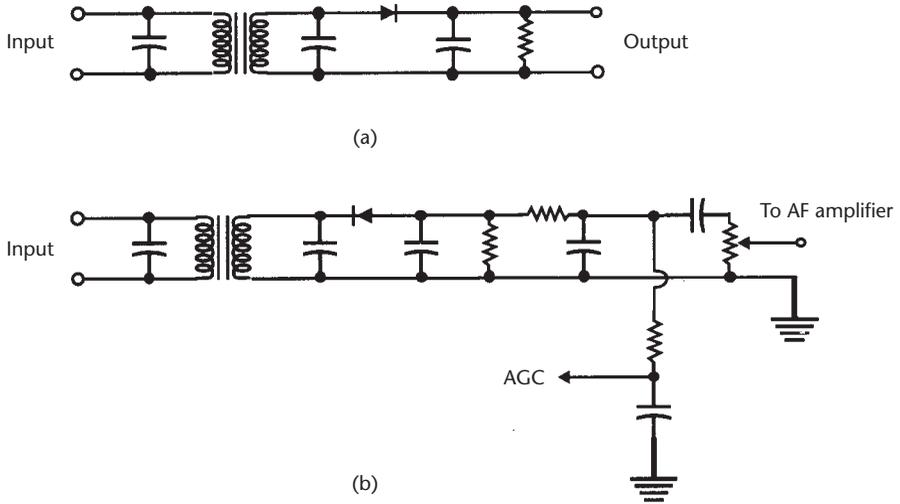


Figure 9.19 AM detectors or demodulators: (a) diode envelope detector and (b) practical AM receiver detector.

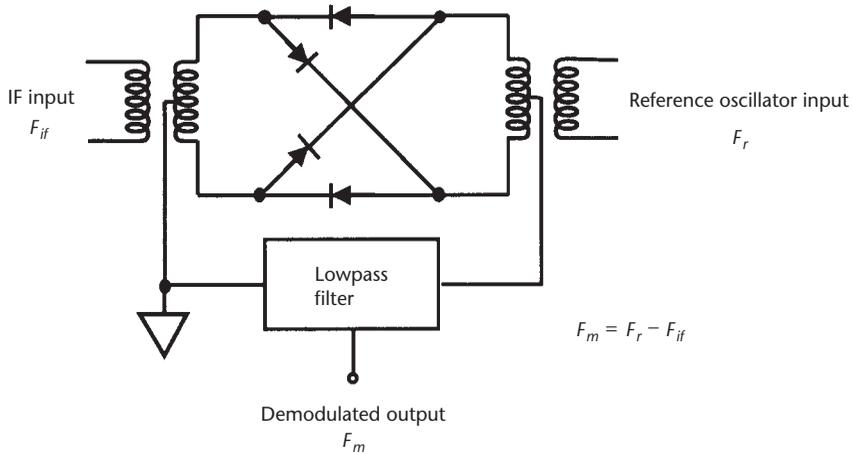


Figure 9.20 Product detector.

second input from a reference crystal or other oscillator. The frequency of the reference oscillator is the same as the carrier frequency. The difference frequency for the mixer is the desired modulation frequency, which can be obtained using a simple lowpass filter.

In the case of SSB detection, the requirements on frequency and phase of the reintroduced carrier at the phase detector are not severe, since only speech is transmitted. Small errors in frequency and phase of the recovered audio signal go unnoticed. The carrier oscillator frequency usually can be tuned manually for best reception.

Detection of DSB signals requires the reintroduced carrier to be in exact frequency and phase with the carrier that would have been there if the signal were

standard AM to avoid distortion. If the frequency is incorrect, the upper and lower sidebands will produce different beat frequencies with the carrier. If the phase of the carrier is not correct, the amplitude of the audio output is reduced. If a pilot reduced-amplitude carrier is transmitted with the DSB wave, it can be used to synchronize the VCO in a PLL.

Detection of AM waves with a product detector provides performance superior to that of the simple envelope detector because it is a coherent process. Although the carrier component is present in the AM wave, the product detector requires a separate carrier input of the same frequency and phase. That is obtained from the VCO output of a narrowband PLL that is locked to the carrier component of the AM wave.

A coherent I and Q detector, shown in Figure 9.21, is a circuit often used in radar receivers. The circuit does not use a PLL but rather a fixed frequency oscillator as the signal source for two balanced mixers. One of the two mixers uses a 90-degree phase-shift circuit so the two signals to the balanced mixers are in phase quadrature. The balanced mixers are followed by lowpass filters.

One reason for having I and Q outputs is so the phase angle of the signal can be detected as well as the amplitude. When there is a Doppler frequency shift, the amplitude of the detected signals will change at the Doppler frequency rate. By signal processing, the Doppler frequency can be measured using the two detected signals.

9.2.3 Frequency Modulation Detector Concepts

A number of FM demodulators or detectors use diodes and transformers. One important type is the Foster-Seeley discriminator, illustrated in Figure 9.22. The main shortcoming of the Foster-Seeley discriminator is that any AM on the incoming signal will be demodulated. A good limiter therefore must precede the discriminator for satisfactory operation. That requirement has eliminated the Foster-Seeley discriminator from almost all mass production entertainment receiver circuits in favor of the ratio detector.

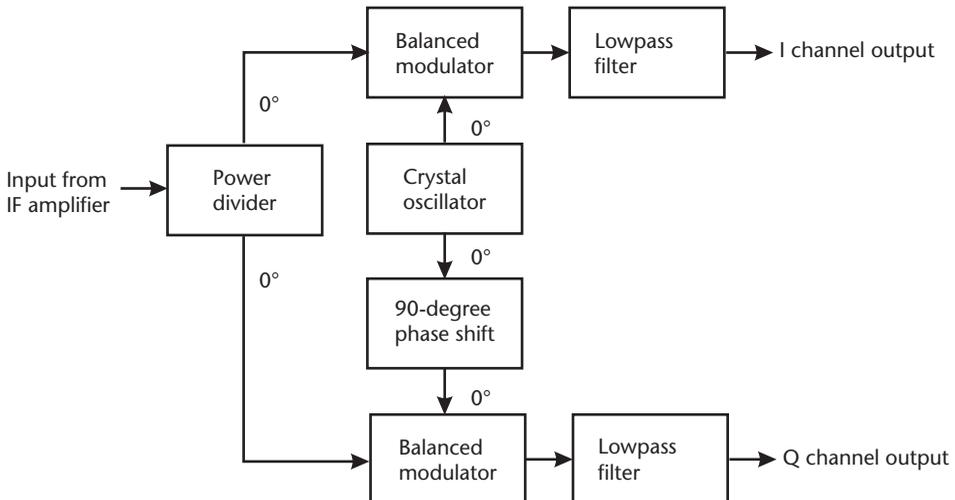


Figure 9.21 I and Q product detectors.

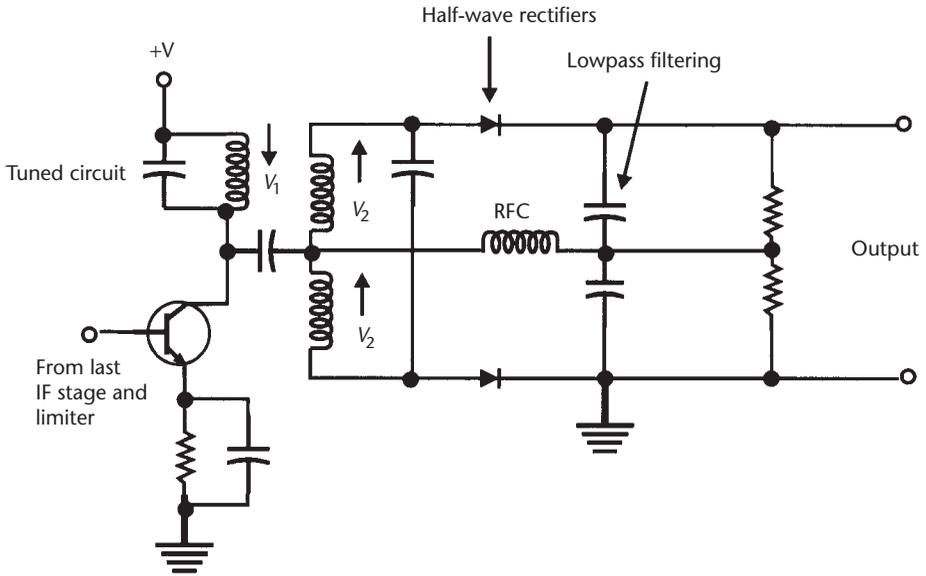


Figure 9.22 Foster-Seeley discriminator. (After: [5].)

The Foster-Seeley discriminator uses a transformer system that includes a tuned primary circuit, a tuned center-tapped secondary circuit, and a coupling capacitor between one side of the primary winding and the center tap of the secondary winding. The two ends of the secondary winding are connected to half-wave rectifiers and lowpass filters. The output is between the two filters.

The operation of the Foster-Seeley discriminator can be understood using the vector diagrams in Figure 9.23. In those diagrams, the two voltages V_2 and $-V_2$ are the voltages across the two parts of the secondary winding. The voltage V_1 is the voltage at the primary winding. The vector sum of V_1 and V_2 is V_a . Likewise, the vector sum of V_1 and $-V_2$ is V_b . At resonance, V_a and V_b are equal, as shown in Figure 9.23(a). Below resonance, the vectors are as shown in Figure 9.23(b). V_a in that case is less than V_b , as a result of the change in phase shift with respect to V_1 for V_2 and $-V_2$. Above resonance, the vectors are as shown in Figure 9.23(c). V_a in that case is greater than V_b . Again, that is a result of the change in phase shift with respect to V_1 for V_2 and $-V_2$.

Figure 9.24 shows the circuit diagram for one type of ratio detector. This circuit is not very sensitive to AM on the incoming signal and does not require a limiter.

The discriminator in Figure 9.24 uses the same phase-shift and vector addition concepts as the Foster-Seeley discriminator. The main differences are the shift in position for the ground, the addition of a resistor and capacitor network across the two rectifier outputs, and the location of the output terminals.

A number of other possible ratio detectors are also used as FM demodulator systems. They are similar to the system in Figure 9.24, with only small changes in implementation.

Another important type of FM detector is the PLL detector, illustrated in Figure 9.25. As the frequency changes, the error signal needed to track the frequency changes. Thus, that error signal is a measure of the signal frequency. It is used as the

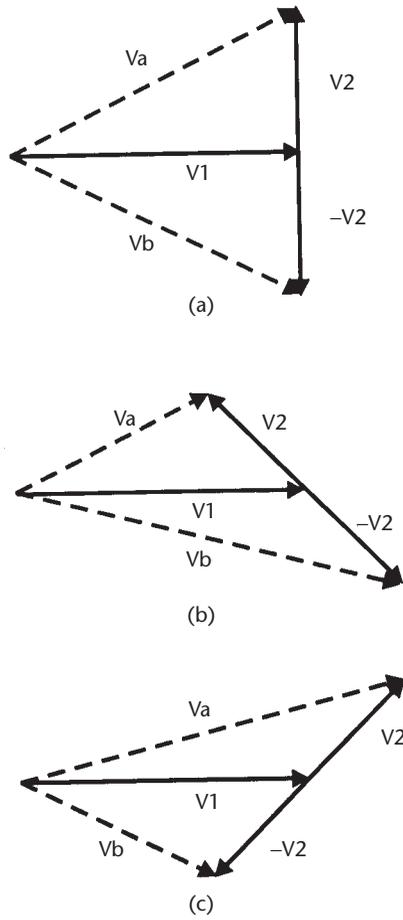


Figure 9.23 Vector diagrams for Foster-Seeley frequency discriminator: (a) at resonance frequency; (b) below resonance frequency; and (c) above resonance frequency. (See Figure 9.22 for V_1 and V_2 .)

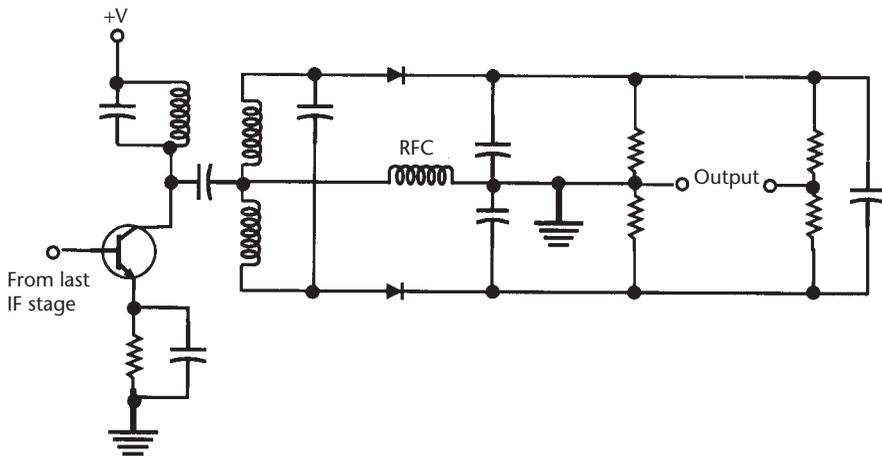


Figure 9.24 Basic ratio detector circuit. (After: [5].)

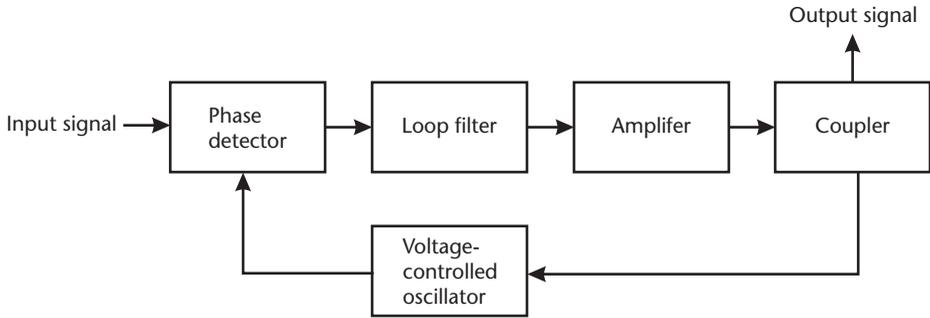


Figure 9.25 PLL FM detector.

output signal for the circuit. Performance is at least as good as that of the ratio detector.

Another FM detector sometimes used is a zero-crossing detector. In one version, the IF output signal is first hard-limited and converted into triggers at the zero crossings. The zero crossings are used to trigger a monostable multivibrator with a fixed pulse on time. The two outputs from the multivibrator are then fed to lowpass filters. The outputs of the filters are fed to a differential amplifier that indicates the difference between the two signals.

9.2.4 Phase Detectors

A double-balance diode mixer normally is used for the detection of BPSK. One input for the mixer is the phase-modulated carrier frequency from the IF amplifier. The other input is the reference carrier frequency from a crystal oscillator or other oscillator system. The output of the mixer is a detected signal. Synchronization systems can be used to obtain the correct frequency and phase angle for the phase reference. A system of this kind is shown in Figure 9.26.

Figure 9.27 shows the case of a coherent demodulator for QPSK. This demodulation provides two bits per baud (pulse).

In the circuit in Figure 9.27, the input from the IF amplifier is fed through a bandpass filter to a three-way power splitter. One output from the power splitter is

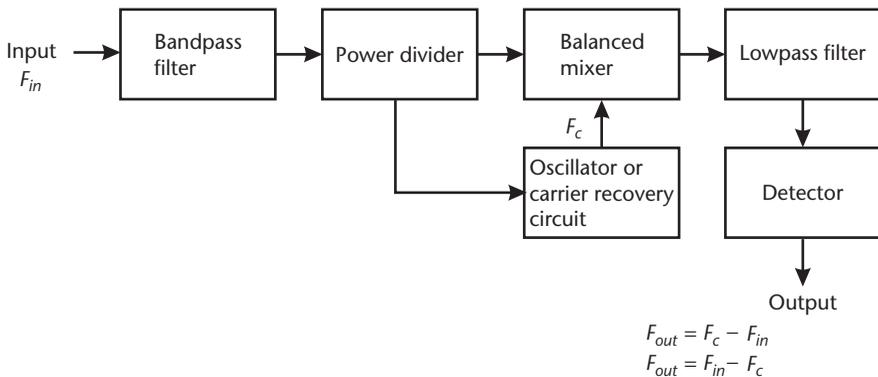


Figure 9.26 Phase detector for BPSK.

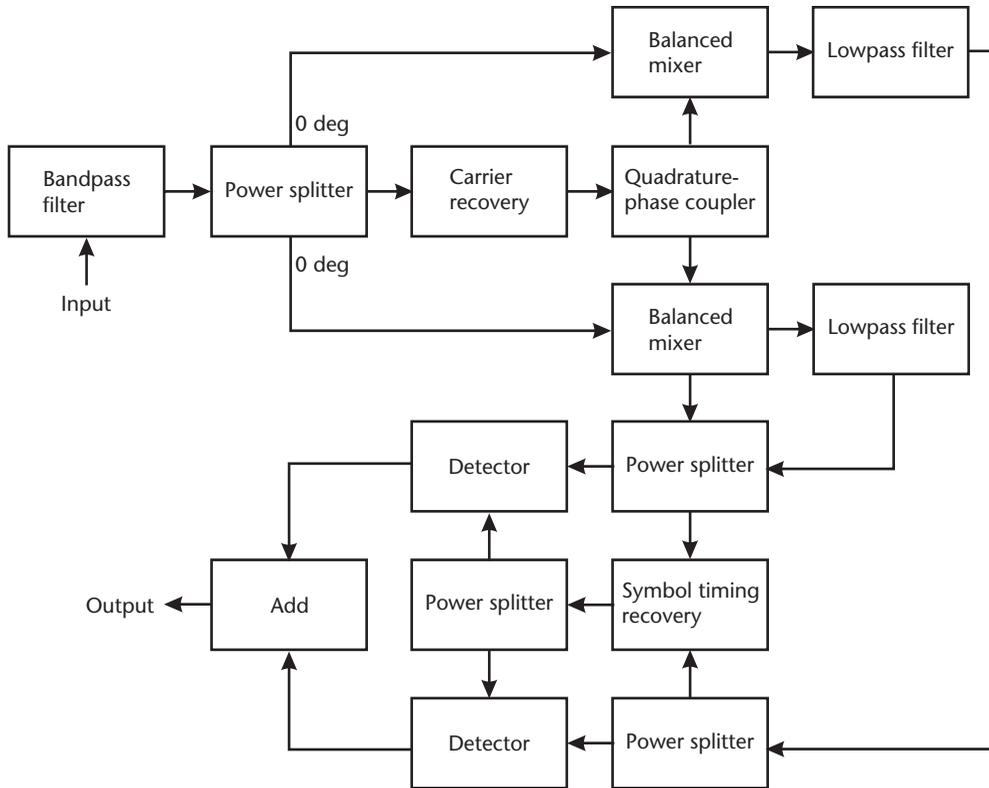


Figure 9.27 Block diagram of a QPSK demodulator.

used for carrier recovery. The carrier is then fed to a coupler whose outputs are in phase quadrature. The other two outputs from the input power splitter are fed to two balanced mixers. The two reference signals for those mixers are from the quadrature phase coupler.

The two mixers are followed by lowpass filters, which provide signals to two-way power dividers. One output from each divider is fed to a symbol-timing-recovery circuit. The other output is fed to detectors and timing circuits, which determine the one or zero states for the signals and provide the necessary time delay so that, when added, the parallel signal is converted to a serial digital signal.

References

- [1] Krauss, H. L., C. W. Bostian, and F. H. Rabb, *Solid State Radio Engineering*, New York: John Wiley & Sons, 1980, Chap. 8.
- [2] Kennedy, G., *Electronic Communication Systems*, 3rd ed., New York: McGraw-Hill, 1985, Chaps. 3–5.
- [3] Keiser, B. E., and E. Strange, *Digital Telephony and Network Integration*, 2nd ed., New York: Van Nostrand Reinhold, 1995, Chap. 6.
- [4] Kennedy, G., *Electronic Communication Systems*, 3rd ed., New York: McGraw-Hill, 1985, pp. 537–542.
- [5] Krauss, H. L., C. W. Bostian, and F. H. Rabb, *Solid State Radio Engineering*, New York: John Wiley & Sons, 1980, Chaps. 9 and 10.

- [6] Kennedy, G., *Electronic Communication Systems*, 3rd ed., New York: McGraw-Hill, 1985, Chaps. 3, 4, 6.
- [7] Krauss, H. L., C. W. Bostian, and F. H. Rabb, *Solid State Radio Engineering*, New York: John Wiley & Sons, 1980, Chap. 6.

Example Communication System Block Diagrams

This chapter presents several examples of some of the older types of communication systems, including HF ground-to-air communication systems, VHF and UHF ground-to-air communication systems, VHF and UHF mobile communication systems, FM broadcast systems, microwave relay systems, and satellite relay systems. Chapter 4 presents examples of radar systems. The components and subsystems used in each of these types of systems are mainly those discussed in Chapters 8 and 9. An important objective of this chapter is to show applications for the previously described components and subsystems.

10.1 HF Communication System Using Single-Sideband Modulation

Figure 10.1 illustrates an HF communication link between a ground station and an aircraft using ionospheric reflection propagation. This is an important means of communication for aircraft flying over an ocean. The types of communication systems involved are discussed in the following paragraphs.

Figure 10.2 is a block diagram of an HF communication system transmitter and receiver system operating in the 2–30-MHz frequency range. A combination transmitter and receiver system of this kind is called a transceiver. The system in Figure 10.2 is assumed to operate with SSB modulation.

The input to the transmitter is from an input system that includes a microphone and a buffer amplifier. The input system is followed by a SSB, suppressed carrier modulator. The second input to the modulator is a CW signal from a crystal oscillator operating at a frequency of 1.5 MHz. The SSB modulator is followed by a mixer upconverter. The LO for the mixer is a frequency synthesizer, which gives the transmitter the ability to operate on any one of many assigned frequencies with the precision of a crystal reference.

The mixer converts the modulated signal to the desired transmitting frequency. The mixer is followed by an amplifier chain consisting of a number of amplifiers in series, which bring the power level of the modulated signal to the level required to drive the output power amplifier. The power amplifier is assumed to be a push-pull class B power amplifier that uses either BJT or FET transistors. The output power to the antenna is assumed to be in the range of 100–1,000W.

Possible antenna systems for an HF communication system (discussed in Chapter 13) usually involve an impedance matching circuit and possibly a balanced to unbalanced (balun) transformer circuit. The same antennas can be used by both the

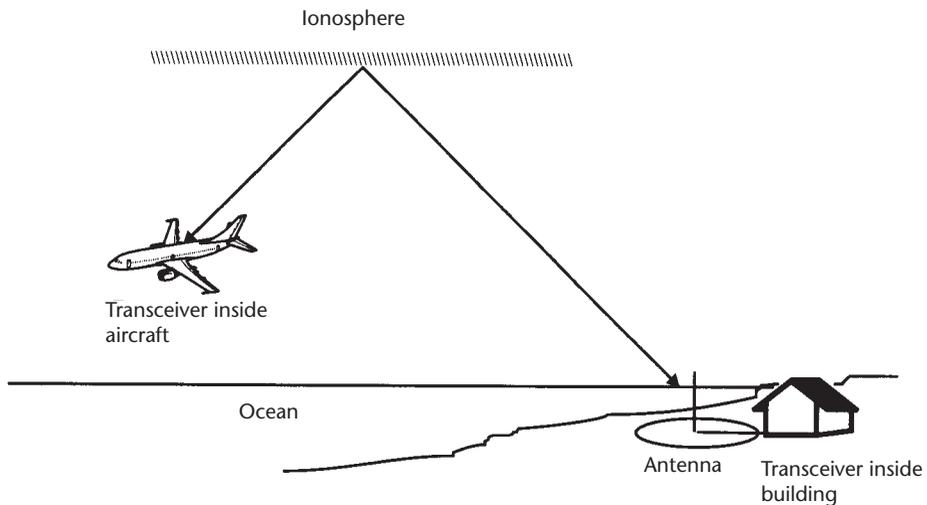


Figure 10.1 Illustration of an HF communication link between an aircraft and a ground station.

transmitter and the receiver system by using a duplexer to select either the transmit or receive mode. The transmit and receive antennas normally use a sparkgap-type lightning protection circuit to protect the transmitter and receiver circuits from lightning strikes.

The lower portion of Figure 10.2 is a block diagram of an HF receiver. The receiver is connected to the antenna system by way of the duplexer. The first stage of the receiver is a low-noise RF amplifier. The output of the RF amplifier is connected to a tunable bandpass filter and mixer circuit. The filter is used to limit the band of frequencies seen by the receiver to only frequencies near the desired signal. It is designed specifically to reject the image frequency, which, when mixed with the LO signal, will be converted to the IF amplifier frequency.

The system in Figure 10.2 uses dual conversion with two IF amplifier frequencies. Double conversion can provide better image rejection and better narrowband filtering capability than single conversion, at a cost of more complexity. The mixers are assumed to be diode-type double-balanced mixers. The first LO is a frequency synthesizer, which gives the system the ability to operate at any one of a number of assigned frequencies with the accuracy of a crystal oscillator.

A filter at the output of the mixer is used to select the lower sideband, which is at the IF amplifier frequency of 1.5 MHz. The IF amplifier is assumed to consist of two or more transistor class A amplifiers with bandpass filters. An alternative system uses integrated circuit IF amplifiers. Typical filters are single-tuned transformers, ceramic filters, SAW filters, or mechanical filters. The IF amplifier has AGC, with control signals provided by the output amplitude detector. The output of the first IF amplifier is fed to a second mixer and filter combination. The LO for this second mixer is a fixed-frequency crystal oscillator having a frequency of 1.4 MHz. The difference frequency output for the second mixer and filter combination thus is 100 kHz. That signal is fed to the second IF amplifier.

The output of the second IF amplifier is connected to the SSB demodulator. The second input to the demodulator is from a tunable 100-kHz oscillator (the same frequency as the IF amplifiers). That allows the oscillator to be tuned to essentially the

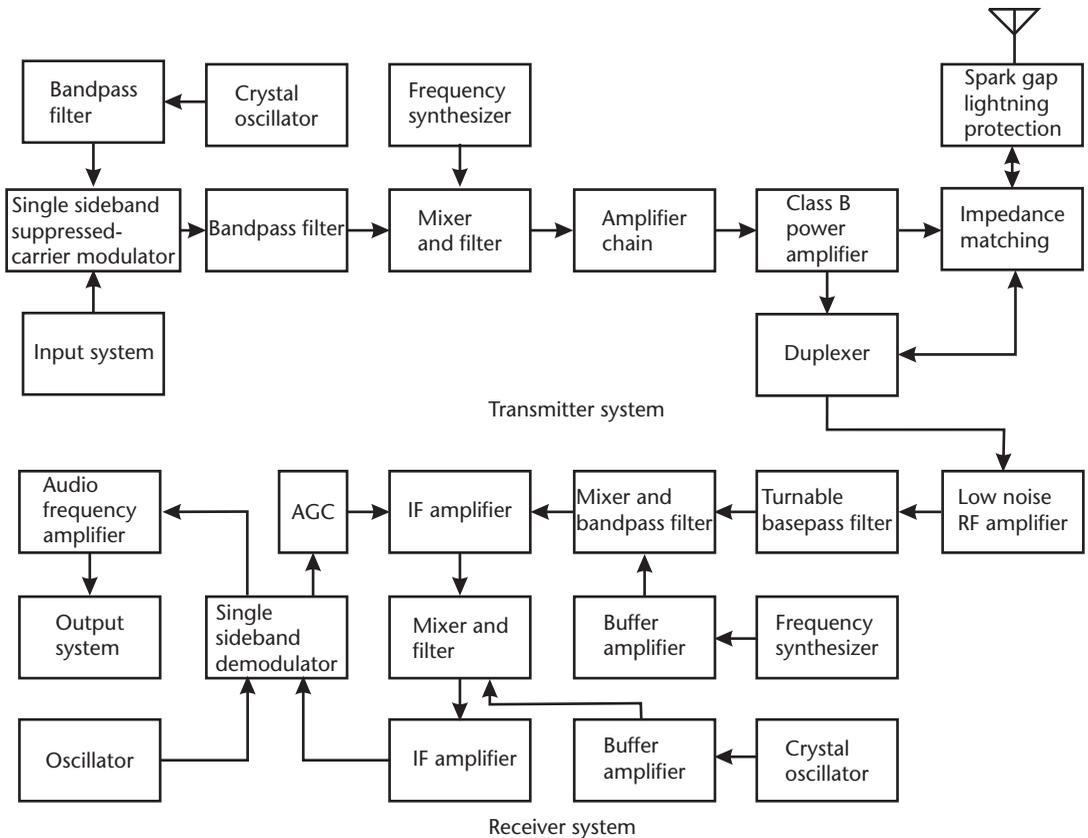


Figure 10.2 HF communication system using SSB modulation.

same frequency as the carrier frequency of the signal. The demodulator acts as a mixer downconverter to shift the signal down to baseband, and that is the recovered signal.

The output of the SSB demodulator is amplified by an audio frequency amplifier, the output of which is then fed to the output system. The output system may consist of a speaker and headphones.

An amplitude detector also is used at the output of the IF amplifier chain. The detector provides an AGC signal that is fed back to the IF amplifier to control the gain of the system.

10.2 VHF or UHF Ground-to-Air Communication System Using Either Amplitude Modulation or Narrowband FM

Figure 10.3 illustrates a VHF or UHF ground-to-air communication system. VHF systems normally use AM, while UHF systems normally use narrowband frequency modulation. Systems of this kind are used by all aircraft flying over land or near land. One or more transceivers are located in the ground stations and in the aircraft.

Figure 10.4 is an example of a VHF or UHF communication transceiver. The system shown uses narrowband FM modulation. A system of this type can be used

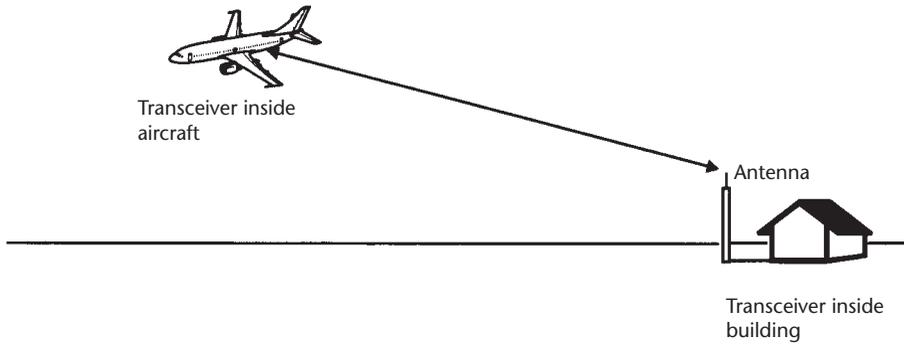


Figure 10.3 Ground-to-air communication system.

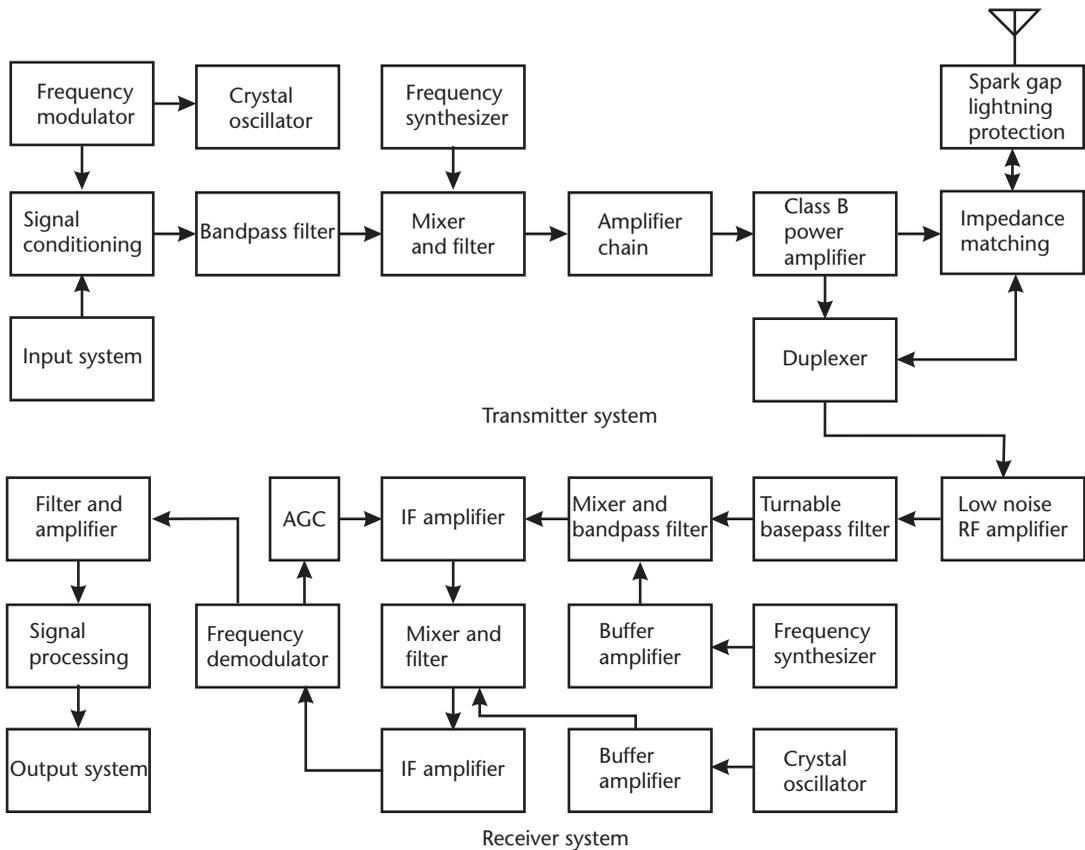


Figure 10.4 Example of a VHF or UHF FM mobile communication system.

for both ground-to-air communication and base station-to-ground mobile communications.

The frequency band used for ground mobile units is the 150–156.8-MHz range. The frequency band for FM ground-to-air communication is 225–400 MHz. The system depicted in Figure 10.4 is used for voice-only communication.

The input to the system normally is from a microphone. The main signal conditioning used is filtering and preemphasis. That is followed by a frequency modulator and a frequency multiplier. The output of the frequency multiplier is fed to a bandpass filter. The output of that filter is fed to a balanced mixer and filter. The second input to the mixer is from a frequency synthesizer. The balanced mixer feeds an amplifier chain, which in turn drives a class B transistor power amplifier. The output of the power amplifier is fed to a duplexer. In the transmit mode, the duplexer connects the power amplifier to the antenna system. This system includes impedance matching, lightning protection, and a monopole, dipole, or similar antenna.

In the receive mode, the antenna is connected by the duplexer to a low-noise RF amplifier. The output of this amplifier is fed to a tunable bandpass filter, which feeds a mixer and filter circuit. The second input to the mixer is from a frequency synthesizer and buffer amplifier. The output from the mixer is at the first IF frequency, assumed to be 10.7 MHz. The output is fed to an IF amplifier chain, which includes the necessary bandpass filtering. The amplifier chain has AGC, so the detected output will be essentially constant regardless of input signal strength.

The output of the first IF amplifier is fed to a second mixer and filter. The LO for that mixer is from a crystal oscillator and buffer amplifier. The output of the mixer is fed to the second IF amplifier, which is assumed to have a center frequency of 455 kHz.

The output of the second IF amplifier is fed to a frequency demodulator, which is followed by a filter and an AF amplifier. The output of the amplifier is fed to a signal processing circuit, which provides for deemphasis. The output of that circuit is fed to the output systems, including speakers and possibly headphones.

Many other two-way communication systems use narrowband FM, including ship-to-shore communication systems, ship-to-ship communication systems, and ship-to-air communication systems. The frequency of operation for these systems includes both VHF and UHF.

10.3 Frequency Modulation Broadcast Systems

Some of the information presented in this section is adapted from [1]. Figure 10.5 illustrates an FM broadcast system, which operates in the 88–108-MHz range. The transmitter generally uses an elevated site for the transmit antenna. That is done to extend the range of the system. The frequencies used are essentially line-of-sight frequencies. Even with an elevated site, signal blockage by buildings and trees is a possibility. In that case, the communication link will experience diffraction propagation loss as well as free-space propagation loss.

A block diagram of the transmitter system is shown in Figure 10.6(a). The system has the capability for stereo operation, so it is necessary to include a signal conditioning circuit with the transmitter for FM multiplex generation.

The input to the transmitter system is from the input system, which includes record players, tape players, compact disc players, switching circuits, controls, and microphones. Microphones for left and right positions with respect to the source are used. In addition, a subsidiary communication authorization (SCA) generator also may be used. The SCA system is used for transmitting background music for stores,

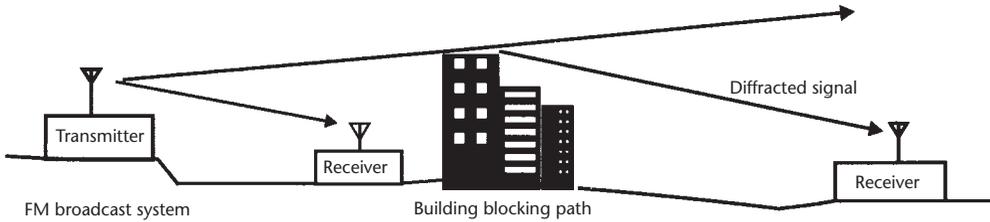


Figure 10.5 Illustration of an FM broadcasting system.

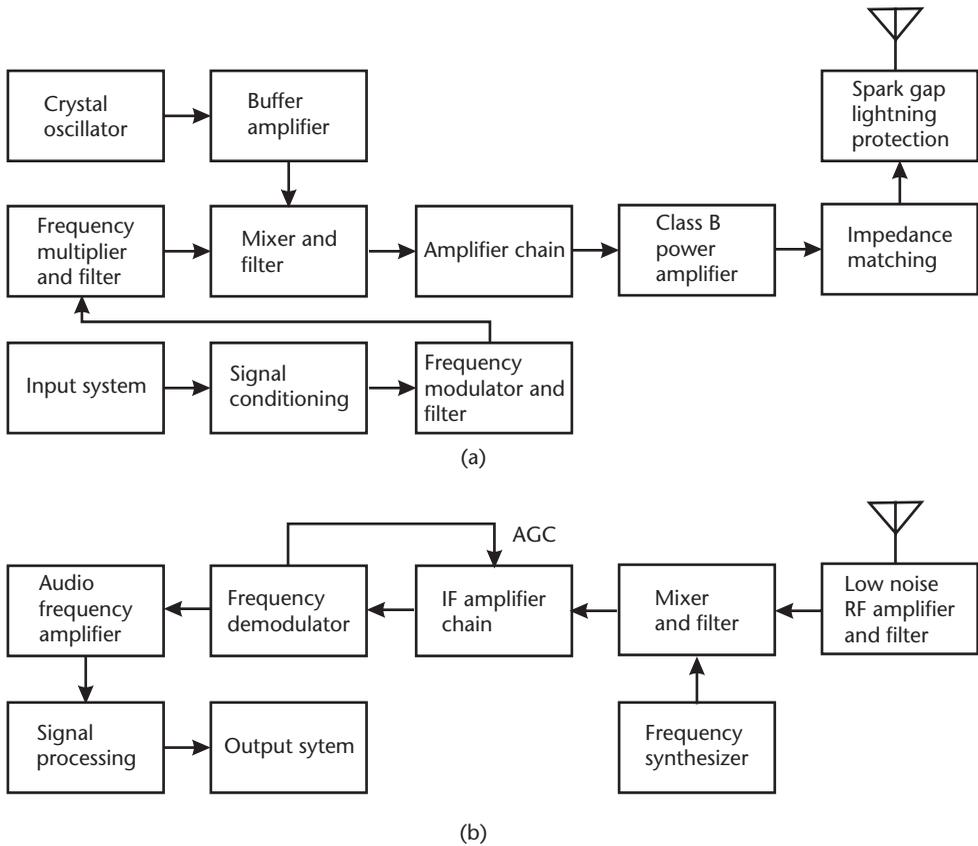


Figure 10.6 Example of an FM broadcast system: (a) transmitter system and (b) receiver system.

restaurants, and so on. The input system signal is fed to a signal conditioning circuit, and the output of this system is fed to a frequency modulator and filter.

The frequency modulator is assumed to be a direct frequency modulator using a varactor-modulated crystal oscillator. That is followed by a frequency multiplier and filter that feeds a balanced mixer. The frequency multiplier is used to increase the modulation index and the frequency of the signal.

The balanced mixer has as its second input a CW signal from a crystal oscillator and buffer amplifier. The frequency output of the balanced mixer is selected by a bandpass filter to be the desired sideband. The output signal is amplified by an

amplifier chain and then fed to a 10-kW class B tetrode-tube power amplifier. The output of that amplifier is fed through a coaxial cable to an impedance-matching circuit that includes the capability of feeding an array of six or more wideband turnstile antennas. The turnstile antenna consists of a pair of dipoles placed in the same plane at right angles to each other and fed with equal amplitudes and 90 degrees phase difference. This antenna array provides a pancake-type pattern that is omnidirectional in azimuth and provides a small elevation beam angle. The polarization is horizontal. The antenna covers the full 88–108-MHz frequency band for FM broadcasting.

An FM transmitter system uses frequency preemphasis. The higher frequencies are boosted at the transmitter and correspondingly reduced at the receiver. That is done because noise has a greater effect on the higher modulating frequencies than on the lower ones.

The FM receiver system shown in Figure 10.6(b) typically uses a dipole receiving antenna. Other possible types include Yagi antennas and whip antennas. The output of the antenna is connected to a low-noise RF amplifier, which is used to provide a low-noise figure for the receiver system. The output of the RF amplifier is fed to a tunable bandpass filter and a downconverter mixer. The mixer can be fed by a frequency synthesizer that acts as the LO for the mixer. The output of the mixer is connected to a bandpass filter. The output of that filter is fed to an IF amplifier chain. The IF amplifier chain includes bandpass filters to band-limit the received signal. That is followed by a frequency and amplitude demodulator system. The amplitude demodulator is used for generating the AGC signal. The frequency demodulator is used to demodulate the FM signal. Frequency deemphasis is used with this circuit.

The output of the FM demodulator is fed through an AF amplifier to a stereo FM multiplex demodulator. The AF amplifier amplifies not only audio frequencies but also frequencies up to about 75 kHz. The outputs from the signal processing circuits are fed to the output systems, such as speakers or headphones.

Signals from the input system are fed to the stereo FM multiplex generator shown in Figure 10.7. The left and right channel inputs are fed to a matrix circuit that provides sum and difference signals. The sum signals are in the frequency range of 50 Hz to 15 kHz. Those are fed to an adder circuit. The difference signals are fed to a balanced modulator that generates DSB modulation with no carrier. The second input to the balance modulator is a CW signal with a frequency of 38

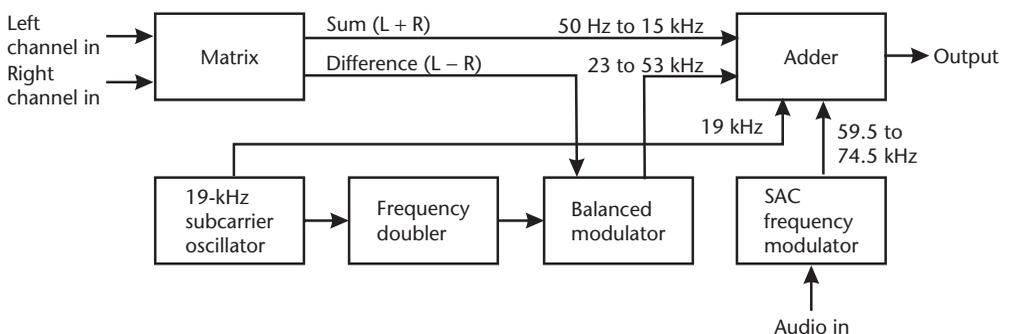


Figure 10.7 Stereo FM multiplex generator with SCA. (After: [1].)

kHz. The CW signal is generated by a 19-kHz oscillator followed by a doubler. The resulting DSB signal has a spectrum from $38 - 15$ kHz to $38 + 15$ kHz, or 23 to 53 kHz. This signal is also fed to the adder. The third signal to the adder is the output of the SCA frequency modulator, which has a subcarrier frequency of 67 kHz and a maximum deviation of ± 7.5 kHz. The resulting spectrum is in the range of 59.5–74.5 kHz. The fourth signal to the adder is the 19-kHz subcarrier oscillator signal that is used as a pilot signal. The sum of the four signals is then fed to the frequency modulator.

Figure 10.8 is a block diagram of a stereo FM multiplex demodulator circuit. The outputs of the stereo FM multiplex demodulator system are fed to the output systems, which are speaker systems, headphones, and others.

10.4 Microwave Relay Systems

An important type of microwave communication system that is used extensively is ground-to-ground microwave relay. Such systems are used to send telephone, radio, television, and other types of signals over long distances. Some of the information presented here regarding microwave relay systems was adapted from [2, 3].

A microwave relay system is illustrated in Figure 10.9.

Microwave relay systems currently operate in C-band and the K_u -band. The 4-GHz band uses frequencies from 3.7 to 4.2 GHz. The 6-GHz band uses frequencies from 5.925 to 6.425 GHz. The width of each of these bands is 500 MHz. Each band is subdivided into a number of channels, with channels in the 4-GHz band being 20 MHz wide, while channels in the 6-GHz band are 30 MHz wide.

The K_u -band systems use the 11-GHz band and the 18-GHz band. The 11-GHz band provides operation from 10.7 to 11.7 GHz. The 18-GHz band provides operation from 17.7 to 19.7 GHz.

The C-band systems have the advantage that they are not as affected by weather. The K_u -band systems, on the other hand, can experience significant attenuation due

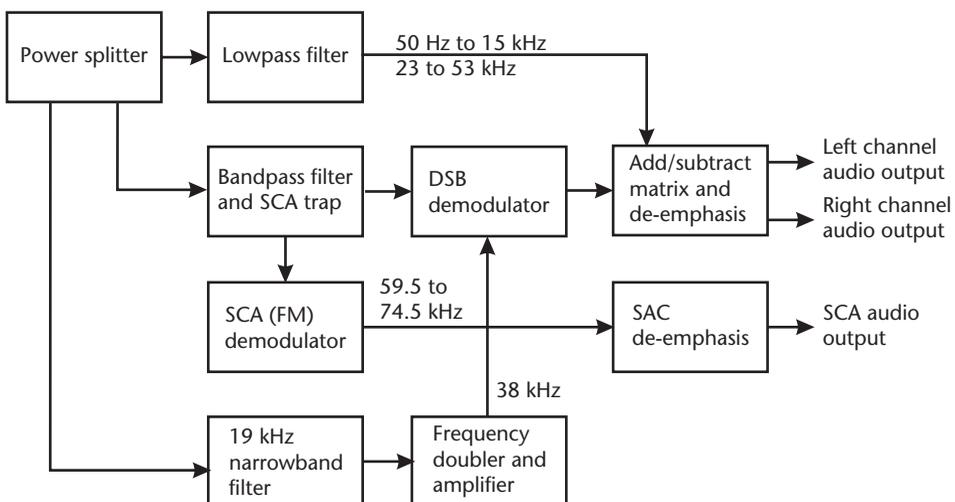


Figure 10.8 Stereo FM multiplex demodulator system. (After: [1].)

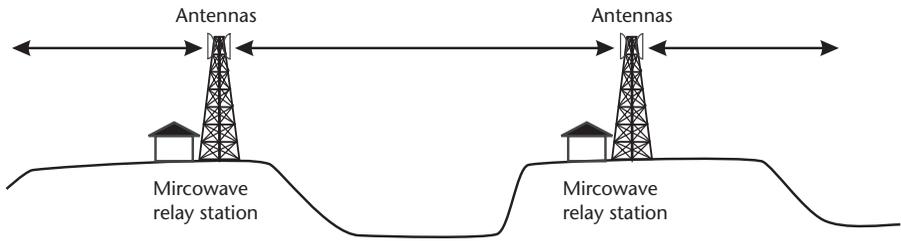


Figure 10.9 Illustration of a microwave relay system.

to rain or snow. Typical repeater spacings for C-band systems are in the range of 20 to 30 miles. K_u -band systems normally have spacings of about 15 miles.

Figure 10.10 is a signal-flow diagram for a 6-GHz microwave relay repeater system. The diagram shows three relay stations, each with two antennas. One antenna of the pair points in the A direction, while the other antenna points in the B direction. Each antenna has both transmission and reception capabilities. Two frequencies are assumed for each relay station: 6.0 GHz and 6.4 GHz.

Station 1 receives from the B direction at 6.4 GHz and transmits in the A direction at 6.0 GHz. It also receives from the A direction at 6.4 GHz and transmits in the B direction at 6.0 GHz.

Station 2 receives from the B direction at 6.0 GHz and transmits in the A direction at 6.4 GHz. It receives from the A direction at 6.0 GHz and transmits in the B direction at 6.4 GHz. The process of switching transmit and receive frequencies every station continues along the line of stations, as shown for relay station 3.

Figure 10.11 is a block diagram of a microwave relay repeater system. The microwave relay repeater receives a modulated microwave signal from one repeater and transmits the signal to the next repeater. Two-way operation is provided. The frequency difference for the two directions of communication is typically a few hundred megahertz at the 4- to 6-GHz frequency band. Two antennas are used, one facing in the A direction and the second facing in the B direction. The antenna types typically used are either parabolic dish antennas or hoghorn antennas. A hoghorn

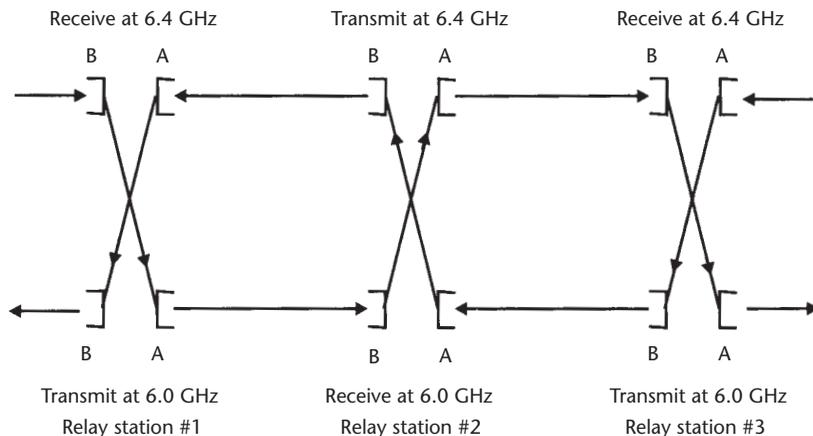


Figure 10.10 Signal-flow diagram of microwave relay systems.

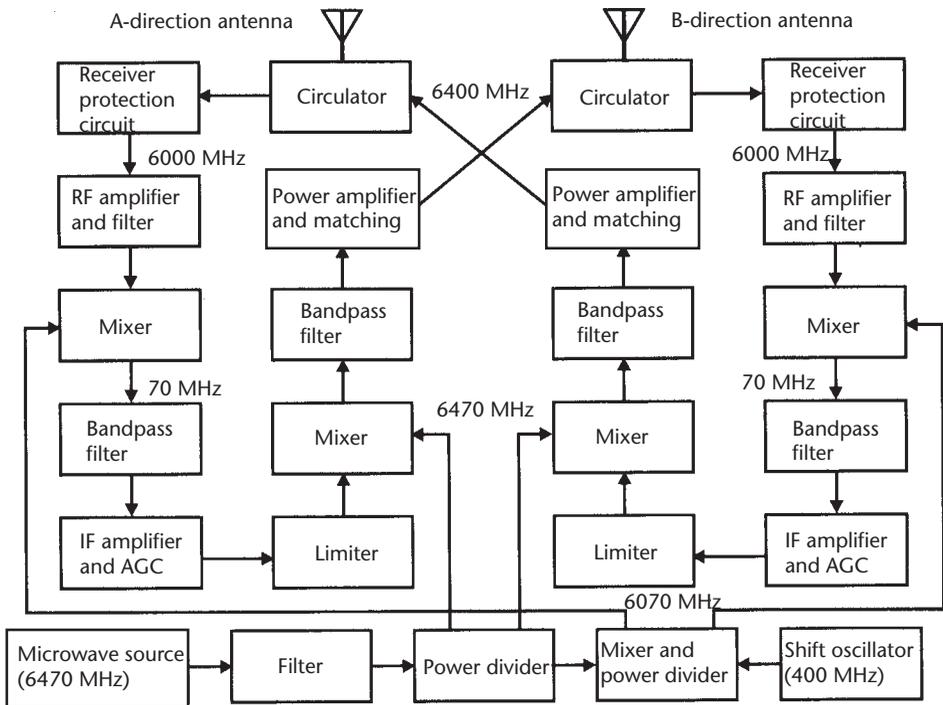


Figure 10.11 Block diagram of a microwave relay system. (After: [3].)

antenna is a combination of a parabolic reflector and a horn antenna with the reflector focus at the horn center directly below the reflector.

The antennas are connected by means of waveguide to circulators, which are used at the junction of receivers and transmitters. Thus, in the ideal case, the transmitter is connected to the antenna but not to the receiver, and the receiver is connected to the antenna but not to the transmitter. (Circulator systems are discussed in Chapters 11 and 12.) In the practical case, there is small coupling between the two systems but not enough to cause the receiver to be degraded in its operation by the transmit signal.

In the system shown in Figure 10.11, the transmit frequency is at 6,400 MHz and the receive frequency is at 6,000 MHz. The first block following the circulator is a receiver protection circuit. A protection circuit might include a sparkgap and zener diodes, which would conduct if the input voltage should exceed the breakdown or zener voltages.

The output of the protection circuit connects to a low-noise RF amplifier, which is followed by an image rejection bandpass filter. That is followed by the receiver mixer. For the example here, the LO frequency is 6,070 MHz and the input signal is 6,000 MHz. The difference frequency is 70 MHz.

The output of the receiver mixer is selected by a 70-MHz bandpass filter and fed to the 70-MHz IF amplifier chain. That is where most of the gain for the receiver takes place. An AGC circuit is used with the system to control the amplitude of the output signals. An amplitude limiter follows the IF amplifier to prevent spurious amplitude modulation.

The next stage following the limiter is the transmitter mixer. In the system shown, the LO frequency for this stage is 6,470 MHz. The difference frequency for the mixer is 6,400 MHz, which is selected by a bandpass filter that follows the mixer. The output of the filter is fed to the power amplifier. Examples of power amplifiers include FET power amplifiers and TWT amplifiers.

The output of the A channel power amplifier is fed to the circulator for the B-direction antenna. There, the signal is fed to the B-direction antenna. The operation of the two channels corresponding to the A and B directions is similar.

The source for LO signals is shown at the bottom of Figure 10.11. Two oscillators are used, one at the microwave frequency and one at the shift frequency. A typical modern system likely would use a VHF transistor crystal oscillator with a varactor frequency multiplier for the microwave oscillator. Multiplication factors are of the order of 20 to 40, and the power output is about 200 mW. The shift oscillator also may be of this type but with smaller multiplication factors required. In the system shown, the output of the microwave source is at 6,470 MHz. It is fed to a three-way power divider. Two of the outputs are fed to the transmitter mixers, and one is fed to a mixer that has as its second input the shift oscillator output. The shift oscillator frequency is 400 MHz, and the difference frequency in that case is 6,070 MHz. The difference frequency is selected by a bandpass filter and fed to a two-way power divider. The output of the divider is to the two receiver mixers.

Until recently, the type of modulation used by microwave relay system typically was analog FM, and the multiplex mode used was frequency multiplex. More recently, the telephone companies have switched over to time division multiplex systems, with phase modulation or QAM modulation used in the time slots. There clearly are reliability and performance advantages for switching to that type of modulation.

For an example of the type of performance achieved by modern microwave relay systems, consider the case of the AT&T DR6-30-135 system operating in the 6-GHz band and utilizing 64-QAM. This system was introduced in 1984, and 64-QAM modulation provides 4 bits per pulse. The total capacity for the system is 14,112 two-way digital voice circuits.

10.5 Satellite Relay Communication Systems

Satellite relay communication is one of our most important forms of communication. Its main use is for telephone and television relay. Some of the information presented here regarding satellite relay was adapted from [2, 4].

Figure 10.12 illustrates a satellite relay communication system. A single ground station transmits to a satellite relay, which receives the signal and then retransmits it to a number of ground stations. Each station uses high-gain parabolic antennas.

Figure 10.13 is a block diagram of one important type of early satellite relay system, the so-called transparent relay. In this system, the relay simply amplifies the signal and shifts the frequency.

In Figure 10.13, the signal from the ground station is received by the satellite using a high-gain directional antenna that points at the Earth. The signal received by the antenna is fed to a low-noise RF amplifier. The output of the amplifier is then

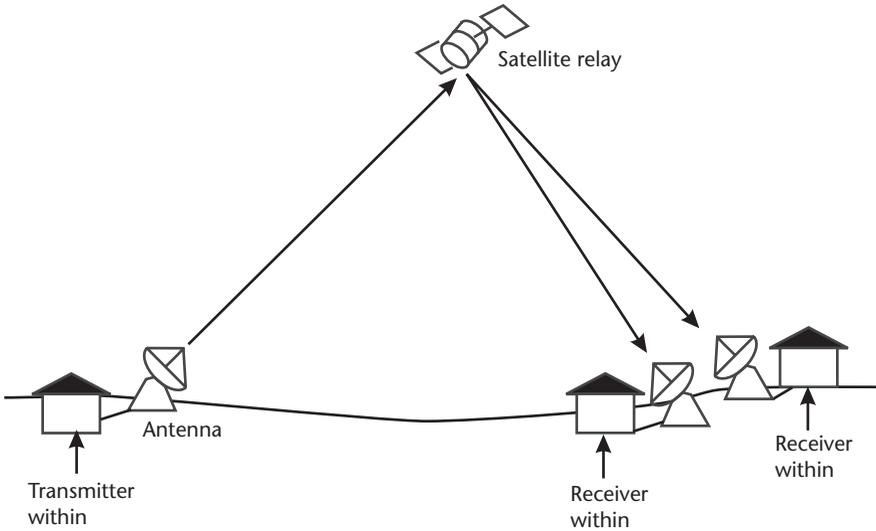


Figure 10.12 A satellite relay communication system.

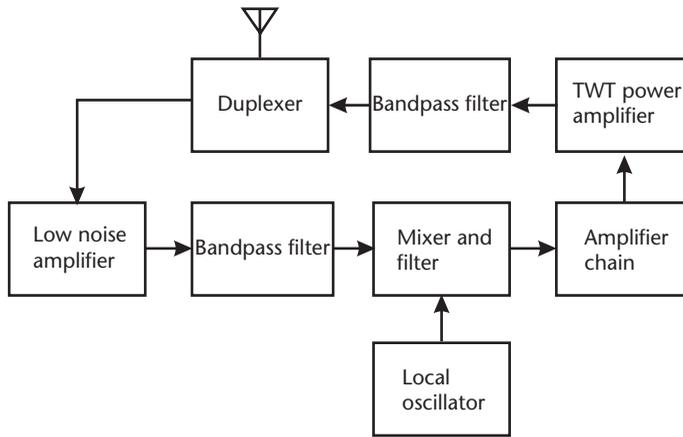


Figure 10.13 Single conversion type satellite relay (transparent relay).

fed to a balanced mixer and filter, where the signal frequency is changed to the transmitter frequency. A frequency shift of 1–2 GHz is typical. The signal is then amplified by an amplifier chain to the desired level for driving the TWT transmitter power amplifier. The output signal is bandpass filtered and then fed to the transmitting antenna. In some systems, the same antenna is used for both transmit and receive by use of a suitable duplexer or circulator.

It was recognized early that multiple-access ability was one of the most valuable features of satellite relay systems. Enhancement of that capability required the addition to the basic functions of the transparent relay of such functions as beam switching, transponder switching, and signal processing. Examples of signal processing additions included demodulation, remodulation, buffering, and storage.

The most important improvement in the evolution of satellite relay systems has been the availability of more weight and size for the satellites as a result of more

powerful launchers, resulting in more power and larger antennas. That, in turn, has resulted in a great reduction in the required size of ground-station antennas and a reduction in transmitting power. The introduction of very small antenna terminals (VSATs) allowed the development of numerous independent communications networks linking industries, banks, and other institutions. Currently most of the peripheral terminals operate with and through central stations, or hubs, on Earth. It is envisioned that the functions of the hub will be available onboard more sophisticated satellites.

The use of satellites has facilitated worldwide television distribution for special events in real time and, in most cases, with delays as required by the different time zones. In the United States, the nationwide distribution of television programs to cable network has become an industry of its own. For rural areas not reached by conventional television broadcast stations and not served by cable television systems, the higher levels of power provided by satellites have made possible direct reception of television from satellites by receive-only Earth terminals equipped with parabolic dish antennas with diameters of 2–5m.

The recent introduction of satellite relay digital television is an important step in reducing the required size of earth-station antennas for satellite television. That permits antennas with diameters of less than 1m.

Another development is direct television broadcast from satellites operating in the K_u -band (10.7–12.75 GHz) with transmitter power levels between 100W and 250W. These broadcasts can be received with antennas as small as 0.5m in diameter.

Satellite relay communication is very desirable for mobile stations. An early satellite relay mobile system was the MARISAT system, established in 1967, followed by the INMARSAT global system. Current systems are able to provide ship-to-shore and air-to-ground communication capability on a global scale. Satellite systems capable of serving Earth vehicles such as trucks, trains, and automobiles currently are being implemented. These systems will be able to provide not only communications but also such functions as determination of location, determination of status of critical shipments, and so on.

The majority of communications satellites are of the geostationary type. The orbital height is $35.863 \cdot 10^6$ m (22,278.5 miles). The orbital plane coincides with that of the Earth's equator. The transmission delay for a single hop ranges from 0.238 to 0.275 second. That has been found to be acceptable for telephone communications. Echo cancellers are used to keep echoes under control.

There are cases in which inclined orbits are used to provide high-latitude coverage. Highly elliptical orbits are typically used in such systems, with the apogee over the northern hemisphere. For that case, the satellite moves much slower near the apogee than it does near the perigee. The orbit is inclined 63.5 degrees to permit zero rotation of the line of apsides. Under those conditions, a satellite with a 12-hour period spends about 8 hours near geosynchronous altitude. Thus, three such satellites can provide continuous service at altitude. The former Union of Soviet Socialist Republics uses a system of this kind known as MOLNIYA.

Polar orbits are also used for noncommunications-type missions, such as Earth observation, weather, and surveillance.

The optimum frequency region for satellite communication systems can be shown to be between 0.8 and 8 GHz. Most of the early satellite relay systems have

operated in that region. The earliest frequency bands used for telephone relay were the same bands used for ground-based microwave relay. The 4-GHz band (3.7 to 4.2 GHz) is used for the downlink, and the 6-GHz band (5.925 to 6.425 GHz) is used for the uplink.

Expansion of satellite communications systems has resulted in the use of the 10–15-GHz band as well. Some of those systems operate at 10.95 to 11.2 GHz plus 11.45 to 11.7 GHz for the downlink and 14.0 to 14.5 GHz for the uplink.

In the near future, it is anticipated that there also will be much use of the 17- and 30-GHz bands for satellite relay.

Earlier types of transponders typically used bandwidths of 36 MHz. The number of transponders per satellite typically was limited to 12 for single-polarization satellites and 20 to 24 for dual-polarization satellites. Current systems use much larger numbers of transponders. For example, the INTELSAT VI uses 46 transponders. The introduction of the 11- and 14-GHz bands and high-speed digital transmission has brought about the use of wider bandwidth systems (80 to 200 MHz) in addition to the 36–40-MHz systems.

FM originally was used for radio signals transmitted to both Comstar I and Comstar IV. In 1982, the equipment at the Earth stations was changed to use SSB-AM of the uplink wave. SSB-AM is considerably more efficient in its use of spectrum than is FM. The number of voice channels handled by each transponder channel was increased to 7,800. With 12 transponders used, that increased the capacity of the satellites to 93,600 two-way voice circuits.

There is a trend now for converting satellite relay systems to TDM in much the same way that there was a switch over to this mode for ground-based microwave relays. This will not increase the number of voice circuits per satellite, but it will add to the reliability and the performance of the circuits.

The capacity of modern communication satellites is impressive. For example, Comstar IV has 24 transponder channels, each capable of carrying 670 Mbps. The total capacity of the satellite thus is 1.4 Gbps, which corresponds to a capability to transmit 220 million pages of text per hour.

Future satellites are expected to use onboard signal processing functions in addition to amplification. Digital transmission and digital techniques are expected to be used extensively. Digital transmission offers the following major advantages:

- Guaranteed error control;
- Reduced sensitivity to channel nonlinearity;
- Efficient tradeoff of power and bandwidth;
- Flexibility with regard to multiplexing diverse signals;
- Easy combination of the functions of transmission switching and routing;
- Signal regeneration capability;
- Implementation with rugged hardware.

10.6 Satellite Relay Earth Stations

There are three categories of Earth stations for satellite relay communications: stations that transmit and receive, stations that receive only, and stations that transmit only. The first type of station is used in two-way communication systems. The second type is used for receiving television broadcast from satellites. The third type is used in data collection systems. Satellite communication Earth-station transmit power ranges from a few watts to 10 kW or more. Solid-state systems are used at low power levels, and klystrons are typically used for high-power levels. Low-noise receivers are used in all cases.

Figure 10.14 is a simplified system block diagram for a two-way satellite communication system Earth station.

The antenna in a system like that depicted in Figure 10.14 typically is a parabolic dish antenna. The antenna size varies considerably. Some high-capacity systems use antenna diameters as large as 32m. Most antennas are somewhat smaller.

The antenna is followed by a duplexer. There are a number of possible types of duplexers, including circulators. The receive signal from the duplexer is first amplified by a low-noise amplifier, which is followed by downconverters and IF amplifiers. Dual conversion normally is used in the receiver system, with the first IF amplifier having a typical center frequency of about 300 MHz, and the second IF amplifier having a typical center frequency of about 70 MHz.

The IF amplifier is followed by a demodulator circuit. The type of demodulator depends on the type of modulation used. Examples are FDM/FM (single-carrier case), FDM/FM/frequency division multiple access (FDMA) (multiple-carrier case), and TDMA.

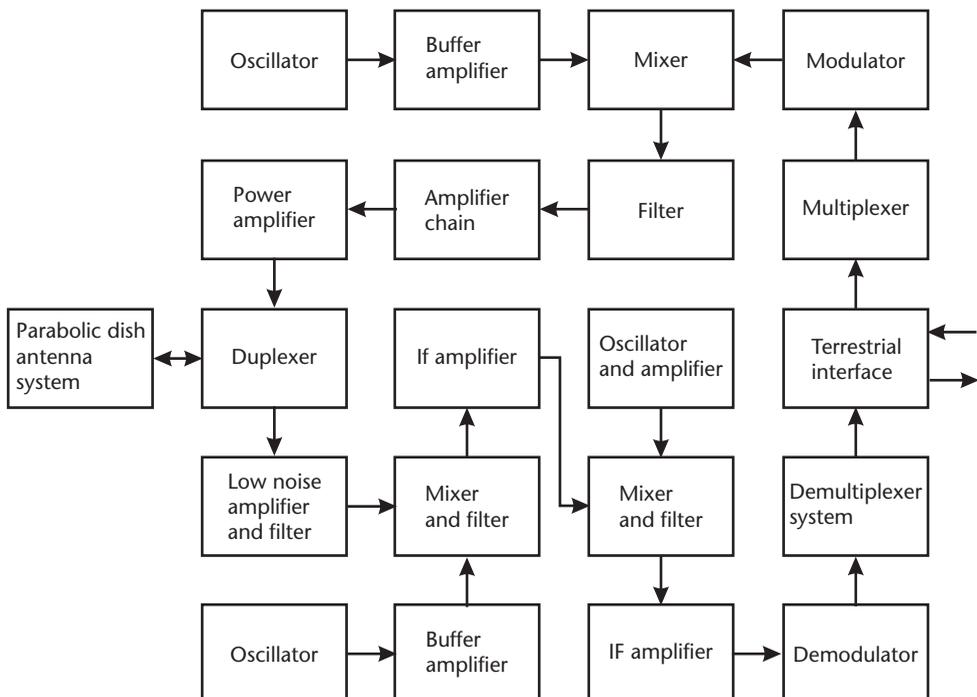


Figure 10.14 A block diagram of a two-way Earth station.

The demodulator is followed by a demultiplex system (deMUX), which is followed by a terrestrial interface unit with connections to and from the serving office.

In the transmit mode, signals are fed to the terrestrial interface unit from the serving office. The signals are routed to a multiplexer system. The output of this system is fed to a modulator. Again, the type of modulator used depends on the desired modulation. Again, example modulations are FDM/FM (single-carrier case), FDM/FM/FDMA (multiple-carrier case), and TDMA. The output of the demodulator is fed to the upconverter and amplifier system, which is followed by the power amplifier. A typical system uses a multiple-cavity klystron. The output of the power amplifier is fed to the duplexer, which routes the signal to the antenna.

For reasons of simplicity, a number of important system blocks are not shown in Figure 10.14, including pointing controls for the antenna, high voltage power supplies for the power amplifier, cooling systems, monitoring systems, control systems, among others.

References

- [1] Kennedy, G., *Electronic Communication Systems*, 3rd ed., New York: McGraw-Hill, 1985, Chaps. 5 and 6.
- [2] Noll, A. M., *Introduction to Telephones and Telephone Systems*, Norwood, MA: Artech House, 1991, Chap. 3.
- [3] Kennedy, G., *Electronic Communication Systems*, 3rd ed., New York: McGraw-Hill, 1985, pp. 548–552.
- [4] Bargellini, P., and G. Hyde, “Satellite and Space Communications,” Chap. 27 in *Reference Data for Engineers*, 8th ed., M. E. Van Valkenburg, (ed.), SAMS, Carmel, IN: Prentice Hall Computer Publishing, 1993.

PART II

RF Components and Circuits

Transmission Lines and Transmission Line Devices

11.1 Two-Wire Transmission Lines

Some of the information presented in this section is adapted from [1].

Many types of transmission lines can be used for RF applications. One type that is frequently used is the two-wire line. Figure 11.1 shows a two-wire transmission line.

Unshielded two-wire lines are often used when line pickup or radiation from the line is not too important. They have the capability for high-power handling and usually are constructed to provide either 300Ω or 600Ω characteristic impedance. The common two-wire television cable used for connecting a television antenna and the television receiver is an example of a low-power 300Ω two-wire line with a solid dielectric spacer between the lines. Higher power systems usually use an air dielectric with low-permittivity insulating spacers between the lines. The spacers are used to maintain the desired geometry of the line.

Two-wire lines are balanced lines. That means the fields around the conductors are symmetrical and there is no specific ground conductor. The coaxial line, on the other hand, is an unbalanced line.

The characteristic impedance for a transmission line is that impedance, which when used as the termination or load for the transmission line, produces no reflections but absorbs the total input power from the line. All other loads produce some reflection and standing waves on the line. The characteristic impedance for a two-wire transmission line is given approximately by (11.1). The equation applies where conductors are immersed in dielectric (the dielectric exists around the conductors and not just between them).

$$Z_0 = 276 \log_{10} (2D/d) / (\epsilon_r)^{1/2} \quad (11.1)$$

where:

ϵ_r = relative permittivity of dielectric medium

d = diameter of conductors

D = spacing between centers of conductors

For an example of the use of (11.1), assume the use of air as the dielectric spacer for the lines ($\epsilon_r = 1.0$), a conductor diameter, d , of 1/8 inch, and a spacing, D , of 1 inch. Substituting those values into (11.1), we have

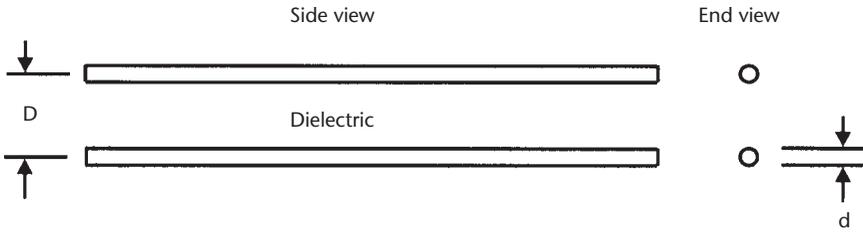


Figure 11.1 Two-wire transmission line.

$$Z_0 = 276 \log_{10}(2/0.125) = 332.3\Omega$$

11.2 Coaxial Transmission Lines

The coaxial line, or coax, is the most frequently used transmission line for conducting signals from point to point in RF systems. Figure 11.2 illustrates a coaxial transmission line.

The advantages of the coax transmission line include high power-handling capability and good shielding. Good shielding is important to prevent pickup of unwanted signals and to prevent radiation of signals that might interfere with other systems. Coaxial lines may be either flexible or rigid. The lines may use a solid dielectric, foam dielectric, or air dielectric, with spacers between inner and outer conductors. Most coax cables have either 50Ω or 75Ω characteristic impedances, although other values are available. The characteristic impedance for a coax line is given by (11.2).

$$Z_0 = \left[138 / (\epsilon_r)^{1/2} \right] \log_{10}(D/d) \tag{11.2}$$

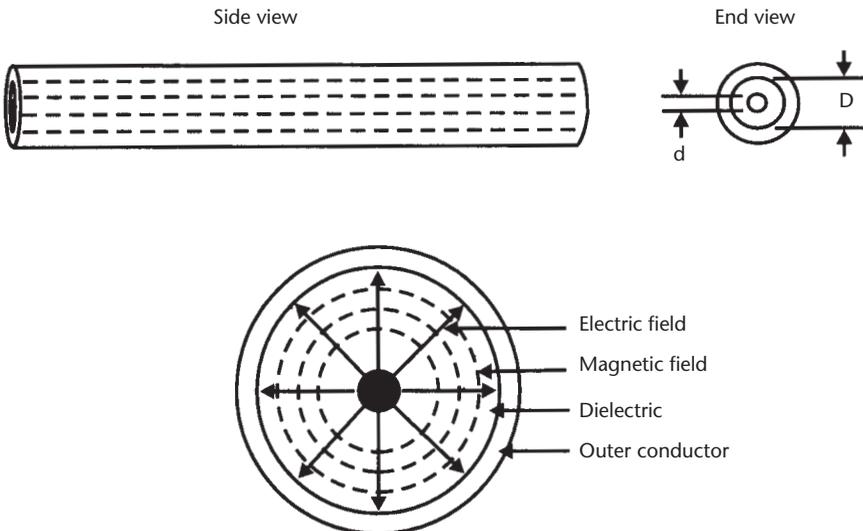


Figure 11.2 Coaxial cable transmission line.

where:

ϵ_r = relative dielectric constant of dielectric medium surrounding center conductor

d = diameter of inner conductor

D = diameter of dielectric or inner diameter of outer conductor

For an example of the use of (11.2), assume the use of polyethylene as the dielectric spacer ($\epsilon_r = 2.3$), an inner conductor diameter of 0.106 inch, and a dielectric outer diameter of 0.370 inch. (These values happen to be those for the type RG-217/U 50 Ω cable.) Substituting those values into (11.2), we have

$$Z_0 = (138/152) \log_{10}(0.370/0.106) = 49.3\Omega$$

Table 11.1 lists the approximate characteristics for several types of coax lines. The values in the table are based on information presented in [1].

Some of the coaxial lines listed in Table 11.1 use polyethylene as the dielectric material. The velocity factor for those cables is 0.66. By that is meant that the speed of the wave in the cable is 0.66 times the speed of light.

Large-diameter rigid coax cables are used for high-power applications where low attenuation and high power-handling capability are desired. Each of the four examples of rigid coaxial cables in Table 11.1 uses air as the dielectric material with teflon pins to support the center conductor. The velocity factor for those four cables is 0.81. Important large-size cables not shown in Table 11.1 are semirigid cables of the same diameter as the rigid cables listed. These cables have attenuations similar to those of the rigid cables at the lower frequencies but slightly higher attenuations at the higher frequencies. Examples of semirigid cables are Heliax cables, Styroflex cables, Spiroline cables, and Alumispline cables.

Very small diameter semirigid cables are used extensively in RF systems. They usually are made with solid copper outer and inner conductors and a teflon spacer ($\epsilon_r = 2.01$). These cables typically have characteristic impedance of either 50 Ω or 75 Ω . Small semirigid cables are used for the high microwave frequencies as well as for the lower frequencies.

Coax transmission lines have both an upper frequency limit and an upper power-handling limit, depending on the size of the coax. The larger the diameter of the coax cable, the lower the upper frequency limit but the higher the power limit. The upper frequency limit is the highest frequency that will support the TEM mode of operation, but it is just below the lowest frequency for which higher order modes (as supported in waveguide) are possible.

The maximum power limitation for cables depends on the size of the cables, the materials used, and the ambient temperature. The conduction losses of the cables result in heat, which must be dissipated. Operation of a polyethylene dielectric cable at a center conductor temperature in excess of 80°C is likely to cause permanent damage to the cable. A number of high-temperature cables, such as RG-211, RG-228, RG-225, and RG-227, can withstand inner conductor temperatures as high as 250°C.

The cables listed in Table 11.1 have upper frequency limits less than 6 GHz; the largest cable has a maximum, higher order mode-free frequency under 1 GHz.

Table 11.1 Approximate Characteristics of Coax Cables

<i>JAN or Other Types</i>	<i>Overall Diameter (inches)</i>	<i>Impedance (Ω)</i>	<i>Attenuation per 100 ft (dB)</i>	<i>Average Power Rating (kW)</i>
50 Ω single braid				
RG-58C/U	0.195	50	5.6 at 100 MHz 20 at 1,000 MHz	0.20 at 100 MHz 0.03 at 1,000 MHz
RG-213/U	0.405	50	2.3 at 100 MHz 8.5 at 1,000 MHz	0.7 at 100 MHz 0.2 at 1,000 MHz
RG-218/U	0.870	50	1.0 at 100 MHz 4.0 at 1,000 MHz	2.0 at 100 MHz 0.5 at 1,000 MHz
50 Ω double braid				
RG-55B/U	0.206	50	4.5 at 100 MHz 16 at 1,000 MHz	0.2 at 100 MHz 0.03 at 1,000 MHz
RG-214	0.425	50	2.0 at 100 MHz 8.0 at 1,000 MHz	0.7 at 100 MHz 0.2 at 1,000 MHz
50 Ω rigid				
Rigid	7/8	50	0.4 at 100 MHz 1.4 at 1,000 MHz	4.0 at 100 MHz 1.3 at 1,000 MHz
Rigid	1-5/8	50	0.2 at 100 MHz 0.7 at 1,000 MHz	15.0 at 100 MHz 3.0 at 1,000 MHz
Rigid	3-1/8	50	0.1 at 100 MHz 0.4 at 1,000 MHz	50.0 at 100 MHz 10.0 at 1,000 MHz
Rigid	6-1/8	50	0.05 at 100 MHz 0.15 at 1,000 MHz	140.0 at 100 MHz 30.0 at 1,000 MHz
75 Ω single braid				
11A/U	0.412	75	2.1 at 100 MHz 8.3 at 1,000 MHz	0.7 at 100 MHz 0.2 at 1,000 MHz
34B/U	0.630	75		1.3 at 100 MHz 0.36 at 1,000 MHz
59B/U	0.242	75		0.3 at 100 MHz 0.08 at 1,000 MHz
75 Ω double braid				
6A/U	0.332	75	3 at 100 MHz 12 at 1,000 MHz	0.4 at 100 MHz 0.12 at 1,000 MHz
216A/U	0.425	75	2.2 at 100 MHz 8.5 at 1,000 MHz	

Note: JAN = Joint Army-Navy.

Source: [1].

Microwave coax cables do exist for frequencies up to about 10 GHz. However, those cables have large per-unit length attenuations and low power ratings due to their smaller geometries. At microwave and higher frequencies, it is necessary to use waveguides for high-power and low-loss operation. Waveguides and waveguide-related circuits are discussed in Chapter 12.

11.3 Coaxial Cable Connectors

A number of different types of connectors are used with coax cables. The most commonly used connectors are BNC connectors, TNC connectors, type N connectors, UHF connectors, and subminiature A (SMA) series connectors.

BNC connectors have a bayonet lock-coupling mechanism to provide a fast connect/disconnect coaxial termination. The BNC series range covers cable entry (flexible), printed circuit board (PCB), bulkhead, panel, and adapter versions, and most are available in 50Ω and 75Ω characteristic impedances. The most popular cable connectors are those with a solder center contact and clamp outer contact or a crimp center contact and crimp outer contact. A typical straight cable plug has a diameter of 0.56 inch (14.3 mm) and a length of 1.08 inch (27.5 mm). The maximum working voltage dc at sea level for this connector is 500V.

TNC connectors are threaded-coupling versions of the BNC connectors. The increased rigidity of the threaded coupling gives the TNC connector a more consistent performance than the BNC connector under adverse operating conditions. The TNC series covers cable connectors, bulkhead and panel styles, adapters, and receptacles. Most are available in 50Ω and 75Ω characteristic impedances. A typical cable plug connector has a diameter of 0.57 inch and a length of 1.25 inches.

Type N connectors also have a threaded coupling mechanism to provide a rigid coaxial termination. The series range covers cable connectors (flexible and semirigid), bulkhead, panel, and adapter versions; most are available in 50Ω and 75Ω characteristic impedance. The most popular cable entry versions are those with a solder center contact and clamp outer contact or a crimp center contact and crimp outer contact. A typical cable plug has a diameter of 0.81 inch (20.6 mm) and a length of 1.44 inches (36.5 mm). The maximum working voltage dc at sea level is 1,000V. The upper frequency limit is in the range of 10–12 GHz.

UHF connectors have threaded mating coupling with interlocking serrations to prevent accidental uncoupling in even the harshest environments. The upper frequency limit for UHF connectors is 500 MHz. A typical cable plug has a diameter of 0.72 inch (18.2 mm) and a length of 1.56 inches (39.7 mm). The maximum working voltage is 500-V peak.

SMA connectors are semiprecision, high-frequency subminiature connectors. The inner diameter of the outer conductor is only 3 mm. They are characterized by a 1/4-36 mating thread and a butt-mating outer contact. They can accommodate semirigid and flexible cables with outer diameters from 0.047 to 0.250 inch, depending on the design. They provide repeatable electrical performance from 0 to 26 GHz and are widely used in microwave systems and subsystems where low attenuation and low voltage standing-wave ratio (VSWR) are required. This type of connector has a voltage rating of 355 to 500 Vrms, depending on the cable. A typical cable plug has an inside diameter of 0.14 inch (3.6 mm) and a length of 0.43 inch (10.9 mm).

Large-diameter rigid and semirigid cables use special connectors other than the types discussed here.

11.4 Microstrip Transmission Lines

Figure 11.3 is a drawing of a microstrip transmission line. Such transmission lines are frequently used in microwave circuits.

Microstrip consists of a thin, flat conductor above a ground plane with a solid dielectric spacer between the two lines. This type of line often is used in the construction of passive microwave circuits such as filters, couplers, power dividers, and power combiners. It also is used in the construction of active circuits such as amplifiers, mixers, switches, and phase-shifters. The characteristic impedance of these lines depends on the dimensions used. Often many different impedances are used on a single board (dielectric and ground plane). Application is limited to UHF, microwave, and higher frequencies where the wavelength is small.

Figure 11.4 shows the characteristic impedance of microstrip line as a function of the relative dielectric constant (ϵ_r) and the w-to-h ratio. The curves in the figure were calculated from quasi-TEM formulas presented in [1]. They assume that the thickness of the strip is negligible ($t = 0$).

For an example of the use of Figure 11.4, assume the use of glass-reinforced teflon (woven) as the dielectric spacer, sometimes called teflon fiberglass. It is one of the most commonly used materials for operation from UHF to K_u -band. It has a relative dielectric constant of about 2.55 and a dissipation factor or dielectric loss tangent of around 0.002. The thickness (h) for the dielectric material is assumed to be 0.10 inch, and the width for the top conductor (w) is assumed to be 0.20 inch. Thus, the w-to-h ratio is 2. The thickness of the top conductor is assumed to be 0. Using these values, we see from Figure 11.4 that the characteristic impedance for the line is about 60Ω .

Perhaps one of the best all-around dielectric materials is woven quartz reinforced teflon. It is similar to teflon fiberglass except that its electrical properties are greatly enhanced by the substitution of low-loss quartz for glass in the reinforcing cloth. Its dielectric constant is 2.47. Coupled with a loss tangent of 0.0006, that makes the material far superior electrically to other materials. It is, however, about six times more costly than teflon fiberglass.

With dielectrics, there is always a loss. The total current density is the sum of a conduction current and a displacement current density in time-phase quadrature. The dielectric-loss tangent is the ratio of the conduction current to the displacement current. For small loss tangents, the power factor is approximately equal to the loss tangent.

Line loss is the result of both dielectric losses and conductor losses. Dielectric losses increase proportionally to frequency, while conductor losses increase as the square root of frequency.

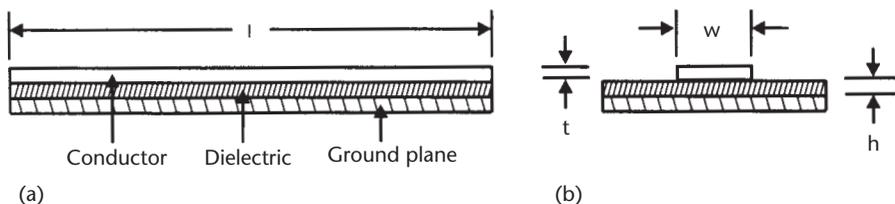


Figure 11.3 Microstrip transmission lines: (a) side view and (b) end view.

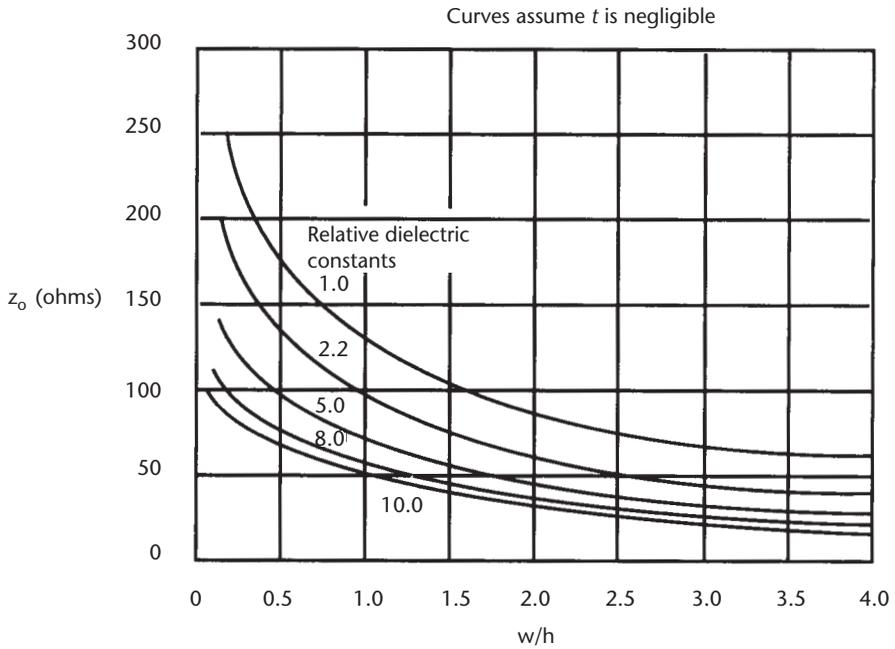


Figure 11.4 Characteristic impedance of microstrip line. (After: [1].)

Power-handling capability is small for microstrip due to the small physical size of the line. A typical 7/32-inch (0.22-inch) top line using a 1/16-inch (0.0625-inch) thick teflon-impregnated fiberglass dielectric has a 50°C rise in temperature above 20°C, ambient for 300W CW power at 3 GHz. With the same line and pulse conditions, corona effects appear at the edge of the strip conductor for pulse power of roughly 10 kW at 9 GHz.

11.5 Stripline Transmission Lines

Figure 11.5 is a drawing of a stripline transmission line.

Stripline transmission line, or stripline, differs from microstrip in that a second ground plane is placed above the conductor strip so that the center conductor is equally spaced from the pair of parallel ground planes. This type of line has a better shielding characteristic than microstrip. It can be used for many of the same types of applications as microstrip lines.

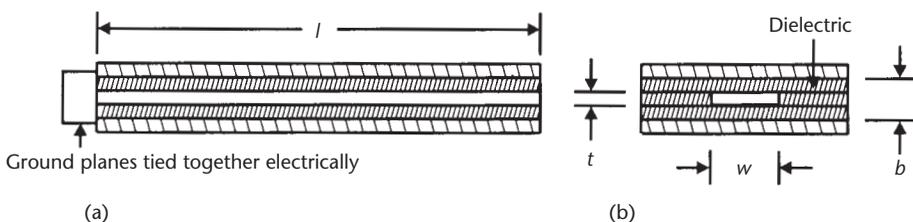


Figure 11.5 Stripline transmission line: (a) side view and (b) end view.

Figure 11.6 shows a plot of the square root of the relative dielectric constant of the dielectric times the characteristic impedance for stripline transmission lines as a function of w/b and t/b , where w is the width of the inner conductor, t is the thickness of this center conductor, and b is the spacing between outer conductors.

For an example of the use of Figure 11.6, assume a w -to- b ratio of 0.2 and a t -to- b ratio of 0.05. From the plot, the square root of the relative dielectric constant times Z_0 is seen to be 130. Assuming the use of a dielectric having a relative dielectric constant of 2.56, the characteristic impedance of the line is

$$Z_0 = 130 / (2.56)^{1/2} = 130 / 1.6 = 81.25 \Omega$$

The average power capability of stripline is primarily a function of the permissible temperature rise of the center conductor and surrounding laminate. It is therefore related to the dielectric used, its thermal conductivity, the electrical loss, the cross-section for the line, any case or supporting material, the maximum allowable temperature, and the ambient temperature. Reference [2] shows that for woven teflon fiberglass as the dielectric, a 0.125-inch ground plane spacing, a 50Ω line, an ambient temperature of 25°C, and an operating frequency of 2.5 GHz, the maximum average power that can be used is 560W. That is based on a calculated temperature rise per watt CW of 0.42°C and a maximum operating temperature of 265°C.

The only dielectric material that would provide higher power capability than woven teflon fiberglass for the same example is microfiber teflon fiberglass. For this case, the change of temperature per watt CW is 0.38°C and the maximum operating temperature is 260°C. Therefore, the maximum average power capability is 620W [2].

Figure 11.7 shows the attenuation per inch of line length versus frequency and ground-plane spacing for a 50Ω stripline transmission line made of woven teflon

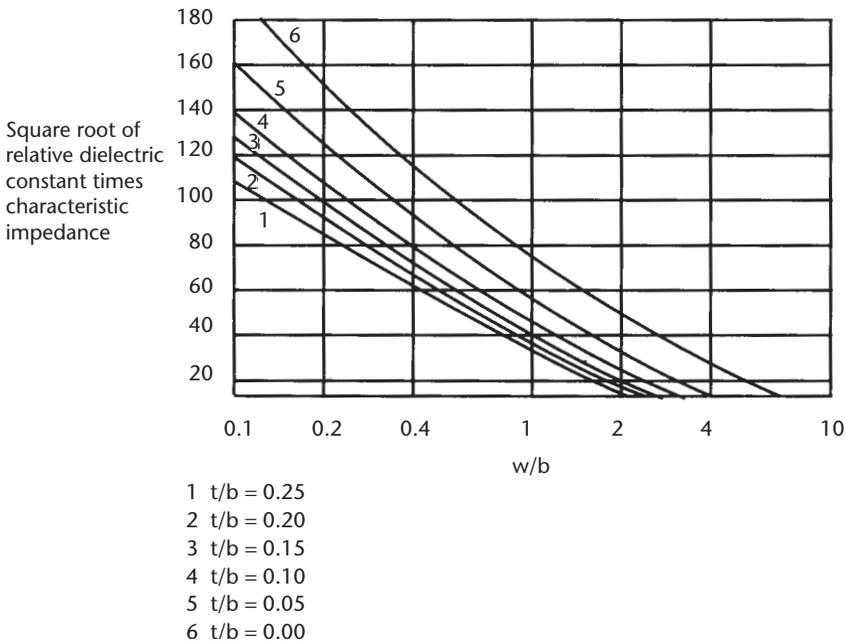


Figure 11.6 Characteristic impedance for stripline. (After: [1].)

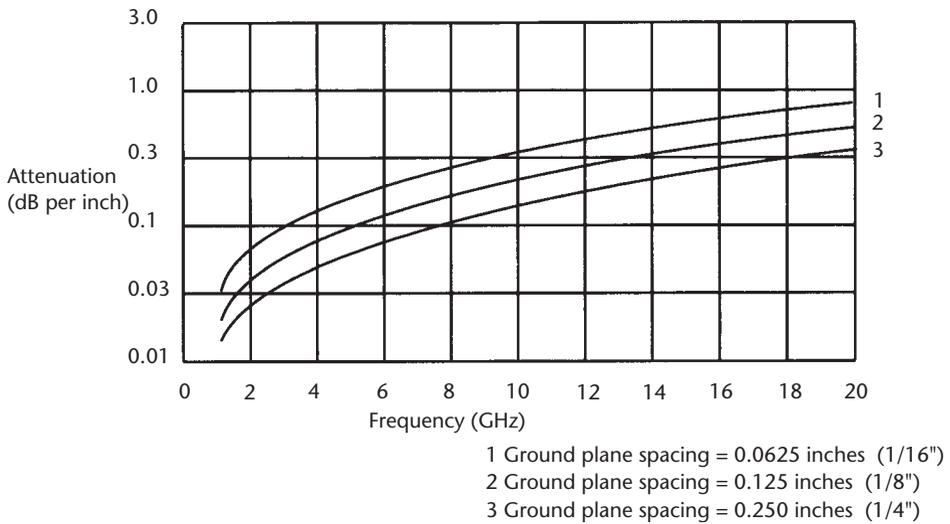


Figure 11.7 Loss versus frequency for a 50V stripline using woven teflon fiberglass as the dielectric material. (After: [2].)

fiberglass. For an example of the use of Figure 11.7, assume operation at a frequency of 10.5 GHz and a ground-plane spacing of 0.125 inch. From the plot, notice that the attenuation or insertion loss is about 0.2 dB/inch.

11.6 Characteristics of Transmission Lines

Some of the material in this section is quoted or adapted from [3].

11.6.1 Wave Velocity on Transmission Lines

The velocity of TEM waves on transmission lines, commonly called the phase velocity, is less than the free-space velocity of light. This velocity is given by (11.3).

$$v = c/(\epsilon_r)^{1/2} \quad (11.3)$$

where:

v = wave or phase velocity

$c = 3 \cdot 10^8$ m/s

ϵ_r = relative dielectric constant or relative permittivity

For example, if the dielectric for a coax cable is polyethylene, the relative dielectric constant is 2.3. The velocity of the wave on the line is 0.66 times the speed of light, or $1.98 \cdot 10^8$ meters per second. Thus, the velocity factor is 0.66.

11.6.2 Reflection Coefficients

The voltage reflection coefficient for a load or termination is given by (11.4).

$$\rho = E_2/E_1 = [(Z_L/Z_0) - 1]/[(Z_L/Z_0) + 1] \quad (11.4)$$

with

ρ = voltage reflection coefficient

E_2 = voltage of reflected wave

E_1 = voltage of incident wave

Z_L = load impedance

Z_0 = characteristic impedance of the system

The reflection coefficient has both magnitude and phase, because Z_L may be a combination of a resistance and a reactance, that is, $Z_L = R_L \pm jX_L$, and so is a vector quantity.

Another relationship of interest is

$$Z_L/Z_0 = (1 + \rho)/(1 - \rho) \quad (11.5)$$

where the terms are as defined for (11.4). This equation can be used to find Z_L if the voltage reflection coefficient is determined by measurement.

11.6.3 Standing-Wave Ratio

Reflections back down the line from an unmatched load give rise to reflection coefficients and standing waves on the line. At points on the line, the direct and reflected waves will be in phase. At those points, the two signals add, producing a maximum standing wave. At other points on the line, 90 degrees electrically away from the point of maximum standing wave, the direct and reflected waves will be 180 degrees out of phase. At those points, the two signals subtract, producing a minimum. Equations for the standing-wave ratio (SWR) are as follows:

$$S = |E_{\max}/E_{\min}| = |I_{\max}/I_{\min}| \quad (11.6)$$

$$S = (|E_1| + |E_2|)/(|E_1| - |E_2|) = (1 + |E_2|/|E_1|)/(1 - |E_2|/|E_1|) \quad (11.7)$$

$$S = (1 + |\rho|)/(1 - |\rho|) \quad (11.8)$$

where all terms are as previously defined.

The magnitude of the reflection coefficient can be found from the SWR, as follows:

$$|\rho| = (S - 1)/(S + 1) \quad (11.9)$$

Figures 11.8 and 11.9 illustrate voltage standing-wave pattern on a lossless line for different load conditions.

Figure 11.8(a) shows the case of an open-circuit load. The direct and reflected voltages are equal and in phase at the load. That results in a voltage maximum at the load equal to twice the direct wave voltage. The current is zero at the load. One-quarter of a wavelength toward the generator, the direct and reflected voltage waves are 180 degrees out of phase, so they add to zero for the case of a lossless line,

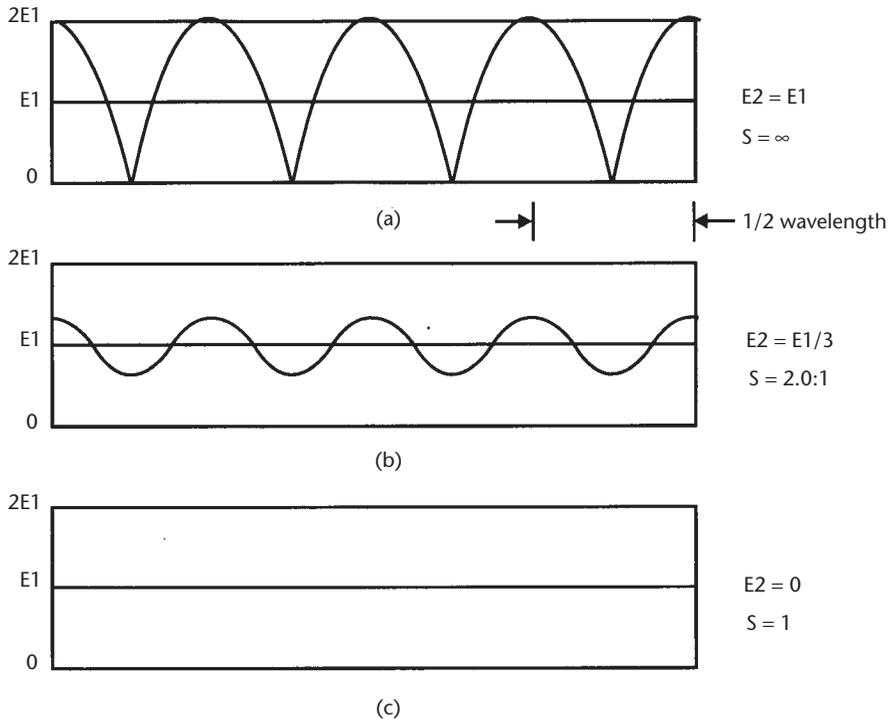


Figure 11.8 Standing waves for three resistive loads: (a) open-circuit load; (b) load is resistor with $R = 2Z_0$; and (c) load impedance = Z_0 .

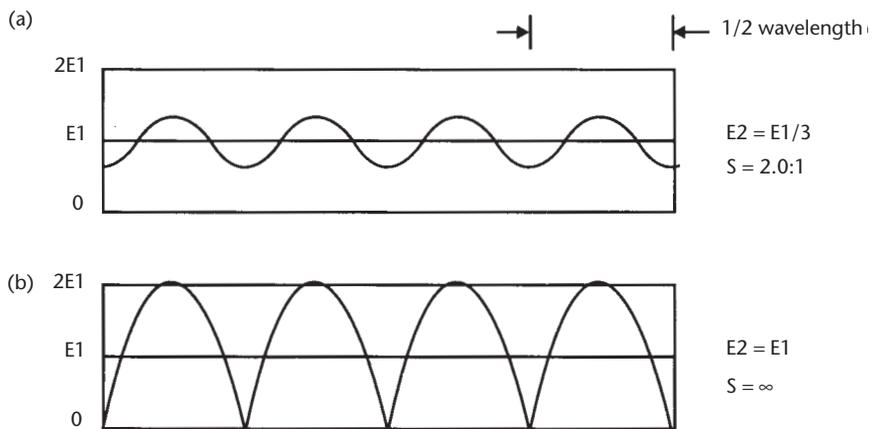


Figure 11.9 Standing waves for two additional resistive loads: (a) load is resistor with $R = Z_0/2$ and (b) short-circuit load.

thus producing a null. The currents are in phase at that point, so there is a current maximum at that point. One-half wavelength from the load, the voltages are in phase and the currents are out of phase, so there exists another voltage maximum and current minimum or null. The standing-wave pattern repeats every half-wavelength up the line. The voltage reflection coefficient case is 1.0, and the SWR is infinity.

If the line has loss, the magnitude of the SWR decreases as we move up the line away from the load. That is because the direct wave increases, while the reflected wave decreases in magnitude.

Figure 11.8(b) shows the standing-wave pattern when the load is resistive and two times the magnitude of the transmission line characteristic impedance. The voltage reflection coefficient in that case is $1/3$, and the SWR is 2. Again, voltage maximums occur at the load and at multiples of one-half wavelength up the line. Voltage nulls have one-half the voltage of the peaks. Current nulls are at the load and at multiples of one-half wavelength up the line.

Figure 11.8(c) shows that there are no voltage standing waves when the load is equal to the characteristic impedance of the line or the operating system. That is the case a designer wants to achieve through impedance matching. For this case, there is no reflected voltage and no reflected power. Thus, the VSWR is 1.0.

Figure 11.9(a) shows the case of a load that is one-half the characteristic impedance of the line. The voltage reflection coefficient in that case is again $1/3$, and the VSWR is again 2. The voltage minimum occurs at the load, and the first voltage maximum occurs one-quarter wavelength from the load. Voltage nulls have one-half the voltage of the peaks. Patterns are repeated every half wavelength up the line.

Figure 11.9(b) shows a load that is a short circuit or 0Ω impedance. For this case, the voltage is 0 at the load, and the current is a maximum. The voltage reflection coefficient is 1, and the SWR is infinity.

If the load is an inductive reactance, the voltage null is located at a distance from the load as though the inductive reactance were a short-circuited length of transmission line having a length less than a quarter-wavelength. If the load is a capacitive reactance, the voltage null is located a distance from the load as though the capacitive reactance were a length of line with an open-circuit for a load with a line length less than a quarter-wavelength.

11.7 The Smith Chart

11.7.1 Impedance and Admittance Coordinates

Figure 11.10 shows a simplified Smith chart [4]. The chart here differs in a number of ways from the usual Smith charts used in the solution of transmission line problems and in the design of impedance-matching circuits and other applications. It has the virtue that it is sufficiently small that it can be used in a handbook of this kind, and yet the numbers and letters are large enough to be easily read. Commercially available Smith charts are larger in size and have more lines and much better resolution and accuracy.

A Smith chart can be used as either an impedance chart or an admittance chart. In either case, the values are normalized to an impedance of $1 + j0\Omega$. At the center of the chart, the resistance or the conductance has a normalized value of 1.0. Toward the top of the chart, along the vertical center line, the resistance or conductance circles have normalized values less than 1. Conversely, toward the bottom of the chart, along the vertical center line, the circles have normalized values greater than 1. The vertical center line corresponds to normalized reactance or susceptance of 0. The

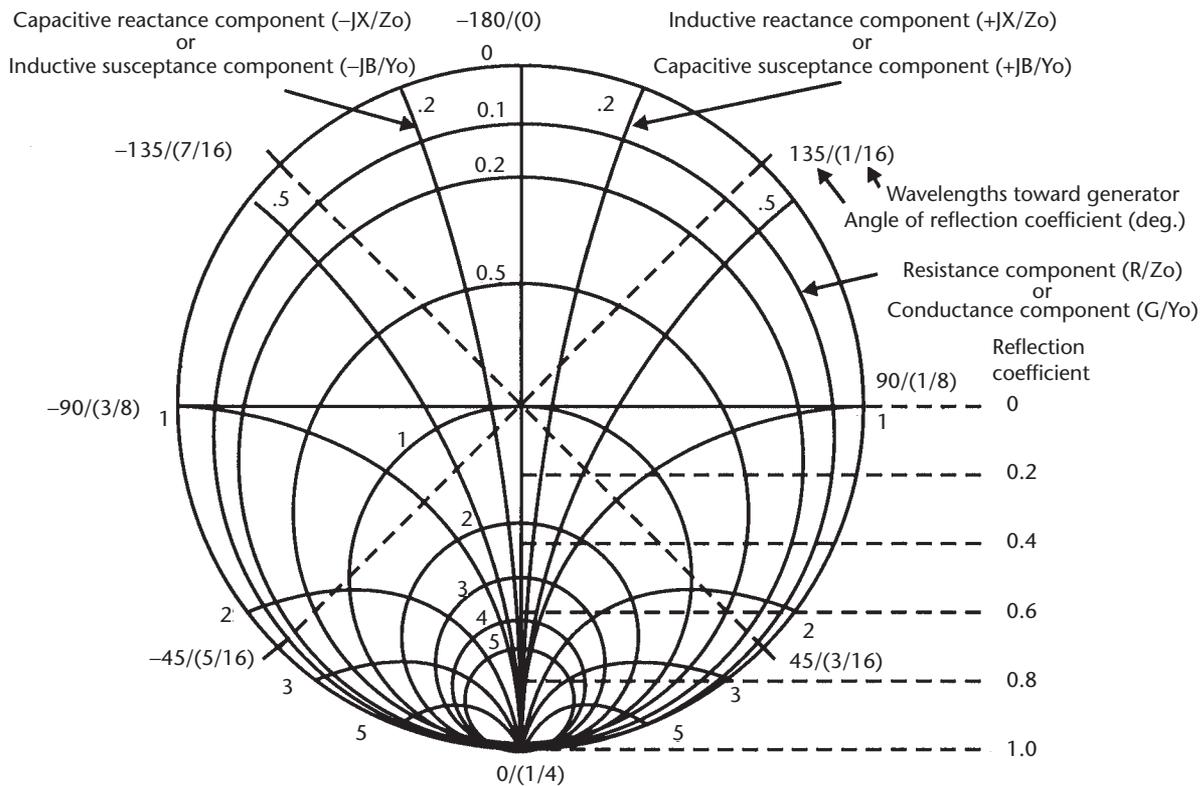


Figure 11.10 Simplified Smith chart.

curved lines to the right of the vertical center line correspond to normalized inductive reactance lines or capacitive susceptance lines. The curved lines to the left of the vertical center line correspond to normalized capacitive reactance lines or inductive susceptance lines. Thus, any point on the chart can be read as either a normalized impedance involving resistance and reactance or a normalized admittance involving conductance and susceptance.

11.7.2 Voltage Standing-Wave Ratio Circles

We can construct a voltage standing-wave circle on a Smith chart by drawing a circle with its center at the center of the chart and its radius such that the circle passes through the impedance or admittance of interest. The point on the center line where the circle crosses indicates the VSWR. If it crosses at 2.0, the VSWR is 2.0.

As a signal travels down a transmission line, its plot follows the standing-wave circle. The distance traveled can be determined with the scale identified as wavelength toward generator and located on the outer edge of the chart. Straight construction lines can be drawn from the center of the chart to the outside scales, with one passing through the load and the other passing through a point on the SWR circle of interest. The difference between wavelength readings indicates the wavelengths traveled.

11.7.3 Reflection Coefficients

We can determine the voltage reflection coefficient's angle for a given load impedance by first drawing a straight construction line from the center of the chart through the load to the scale on the outside of the chart. The angle is read from the angle of reflection coefficient scale. The magnitude of the coefficient is found by measuring the radial distance from the center of the chart to the impedance point of interest. A reflection coefficient scale is shown on the right of the chart.

The power reflection coefficient is the square of the voltage reflection coefficient. The return loss is a measure of how much of the transmitted power is returned to the transmitter due to reflection at the load. For example, if the voltage reflection coefficient is 0.316, the power reflection coefficient is 0.1, and the return loss is 10 dB.

11.7.4 Examples Using the Smith Chart

Figure 11.11 shows a normalized Smith chart with a 50Ω generator and a 50Ω coaxial cable connected to seven load impedances:

1. Open circuit;
2. $100 + j0\Omega$;
3. $50 + j0\Omega$;
4. $25 + j0\Omega$;
5. Short circuit;
6. $150 - j100\Omega$;
7. $20 + j30\Omega$.

The first five impedances are the same ones used for the plots in Figures 11.8 and 11.9.

Impedance 1 (open circuit) is located on the outer circle. The SWR is infinity and the voltage reflection coefficient is $1.0 \angle 0$ degree. The power reflection coefficient is 1.0, the return loss is 0 dB, the power transmission coefficient is zero, and the reflection loss is infinity.

Impedance 2 ($100 + j0\Omega$) is equal to $2.0 + j0.0$ when normalized to 50Ω and is located at $2.0 + j0.0$ on the Smith chart. The SWR is 2.0, and the voltage reflection coefficient is $0.333 \angle 0$ degree. The power reflection coefficient is 0.11, the return loss is 9.55 dB, the power transmission coefficient is 0.88, and the reflection loss is 0.52 dB. Thus, with a SWR of 2.0, only about 10% of the transmitted power is reflected back to the generator, and nearly 90% is transmitted. It is often a goal for impedance matching to have the resulting SWR 2.0 or less.

Impedance 3 ($50 + j0$) is equal to $1.0 + j0.0$ when normalized to 50Ω and is located at $1.0 + j0.0$, the center, on the Smith chart. The SWR is 1.0, and the voltage reflection coefficient is 0.0. The power reflection coefficient is 0.0, the return loss is infinity, the power transmission coefficient is 1.0, and the reflection loss is 0 dB. This is the ideal case for impedance matching where there is no power reflected to the generator and all the power is transmitted to the load.

Impedance 4 ($25 + j0$) is equal to $0.5 + j0.0$ when normalized to 50Ω and is located at $0.5 + j0.0$ on the Smith chart. The SWR is 2.0, and the voltage reflection coefficient is $0.333 \angle 180$ degrees. The power reflection coefficient is 0.11, the return loss is 9.55 dB, the power transmission coefficient is 0.88, and the reflection loss is 0.52 dB. We see that these values are the same as for impedance 2, except for the 180-degree phase shift for the reflection coefficient.

Impedance 5 (short circuit) also is located on the outer circle. The VSWR is infinity, and the voltage reflection coefficient is $1.0 \angle 180$ degrees. The power reflection coefficient is 1.0, the return loss is 0 dB, the power transmission coefficient is 0, and the reflection loss is infinity. These values are the same as for impedance 1, except for the 180-degree phase shift for the reflection coefficient.

Impedance 6 ($150 - j100$) is equal to $3.0 - j2.0$ when normalized to 50Ω and is located at $3.0 - j2$ on the Smith chart. The SWR is 4.4, and the voltage reflection coefficient is $0.625 \angle -18.5$ degrees. The power reflection coefficient is 0.4, the return loss is 4 dB, the power transmission coefficient is 0.6, and the reflection loss is 2.1 dB.

Impedance 7 ($20 + j30$) is equal to $0.4 + j0.6$ when normalized to 50Ω and is located at $0.4 + j0.6$ on the Smith chart. The VSWR is 3.4, and the voltage reflection coefficient is $0.55 \angle 112$ degrees. The power reflection coefficient is 0.3, the return loss is 5.2 dB, the power transmission coefficient is 0.7, and the reflection loss is 1.6 dB.

Figure 11.12 shows a 50Ω system with a 50Ω coax cable connected to four admittances. The normalized admittance for the chart is $1/50 = 0.02$ siemens (S).

The five admittances shown in Figure 11.12 are as follows:

1. Open circuit;
2. Short circuit;
3. $1/(50 + j0) \text{ S} = 0.02 + j0 \text{ S}$;

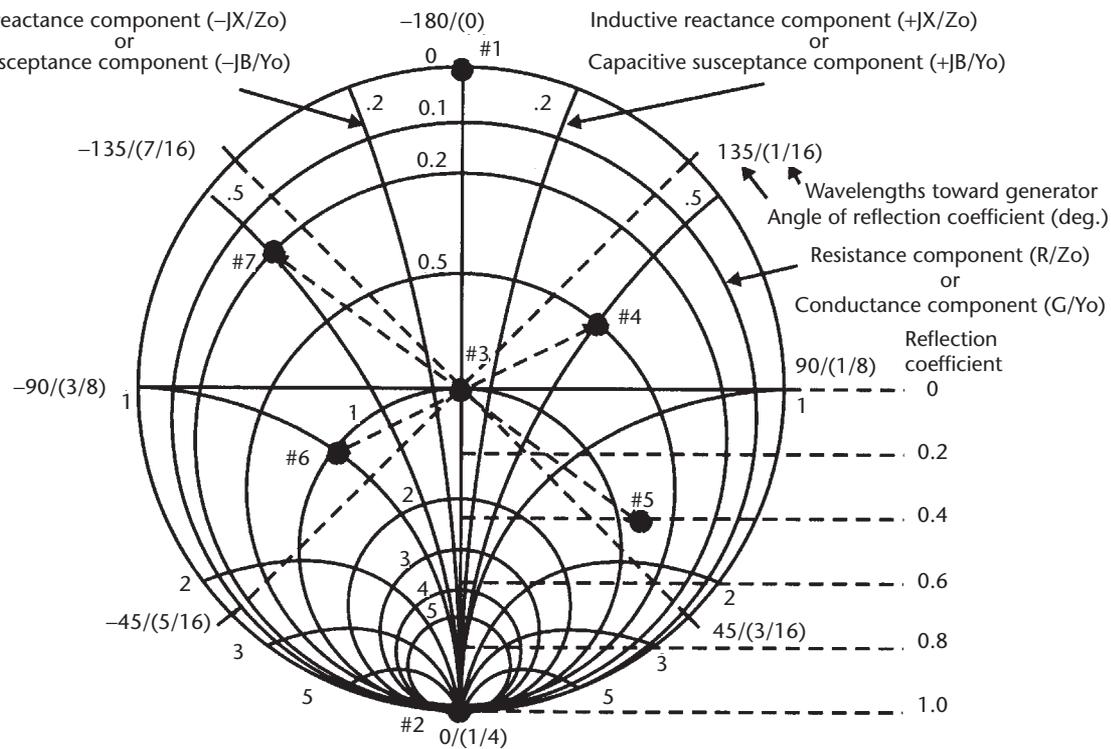


Figure 11.12 Example load admittances plotted on a Smith chart.

4. $1/(50 - j50) S = 0.01 + j0.01 S$;
5. $1/(10 - j25) S = 0.014 + j0.34 S$.

These are normalized to $0.02S$ as follows:

1. $0.0 + j0.0$;
2. Infinity $+ j0.0$;
3. $1.0 + j0.0$;
4. $0.5 + j0.5$;
5. $0.7 + j1.7$.

It is common practice when performing designs using the Smith chart to convert between an impedance chart and an admittance chart. That is done easily by constructing a straight line from the impedance point through the center of the chart to a point equal distance from the center of the chart as the impedance point. This is illustrated in Figure 11.12 for the case of admittances 4 and 5. First, impedance point 6 is located for $50 - j50\Omega$ normalized to $1 - j1$. That is converted to the admittance at point 4 by transferring through the center of the chart with a straight line to the corresponding point on the SWR circle. In a similar way, impedance point 7 is located for $10 - j25\Omega$ normalized to $0.2 - j0.5$. That is converted to the admittance at point 5 by transferring through the center of the chart with a straight line to the corresponding point on the SWR circle. The SWR values for the five admittances are as follows:

1. Infinity;
2. Infinity;
3. 1.0;
4. 2.62;
5. 4.2.

It is important that the impedance or the admittance at a point along a transmission line can be determined by locating that point on the SWR circle at the desired distance in wavelengths from the load, as illustrated in Figure 11.13.

In Figure 11.13, the load is assumed to have a normalized impedance of $Z_L = 1.0 - j1.9$ based on 50Ω , indicated by point 1 on the Smith chart. The SWR is about 5.0. If we move 0.0125 wavelength along the transmission line toward the generator, the normalized impedance is about $0.20 - j0.32\Omega$. This is found by moving on the 5.0 SWR circle from point 1 to point 2.

11.8 Impedance Matching Using the Smith Chart

11.8.1 Impedance Matching with a Quarter-Wave Transformer

The required characteristic impedance for a quarter-wave section of line to transform a load impedance to Z_0 of the main line is given by (11.10).

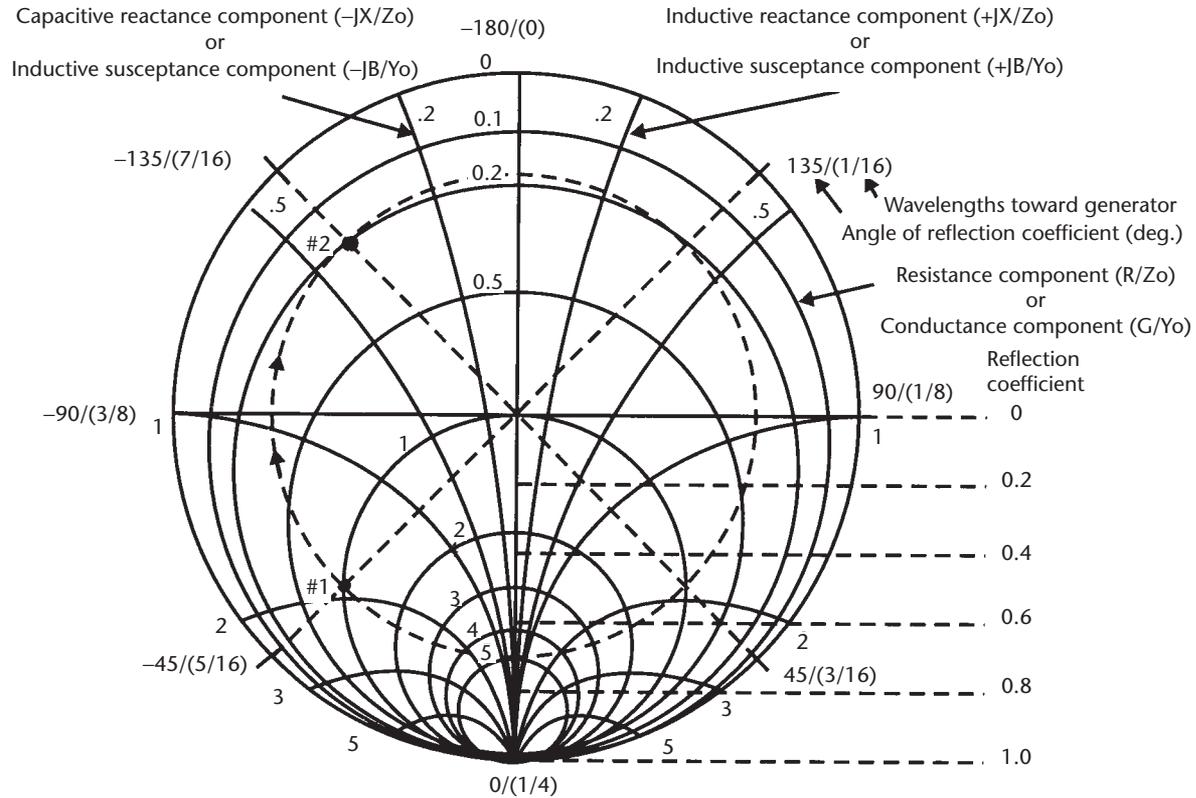


Figure 11.13 Impedance at a point along a transmission line, given the distance from a known impedance.

$$Z_T = (Z_0 Z_L)^{1/2} \quad (11.10)$$

where

Z_T = characteristic impedance of quarter-wave line

Z_0 = characteristic impedance of main transmission line

Z_L = impedance of the load

Equation (11.10) works only for purely resistive loads. For an example of the use of (11.10), assume that the load impedance is 39.8Ω and the main line characteristic impedance is 75Ω . Substituting values into (11.10), the characteristic impedance of the quarter-wave transformer line is

$$Z_T = (75 \cdot 39.8)^{1/2} = 54.6\Omega$$

Figure 11.14 shows an example of impedance matching with a quarter-wavelength transformer. In the figure, the input line has a characteristic impedance of 50Ω . The load has an impedance of $100 - j50\Omega$. That impedance normalized to 50Ω is $2.0 - j1.0$. By using a 50Ω line that is 0.213 wavelength long, the load impedance is transformed to a normalized impedance of $0.38 - j0.0$. The quarter-wave transformer can then be used to convert the normalized impedance to $1.0 + j0.0$. The characteristic impedance for the quarter-wave transformer is $(50 \cdot 19)^{1/2} = 30.8\Omega$.

Figure 11.15 illustrates the use of a Smith chart for determining the distance for the load that the quarter-wave transformer should be placed. The normalized impedance of the load is $2.0 - j1.0\Omega$. That impedance is entered on the chart as point A. An SWR circle is then drawn and point B marked on the chart where the impedance first becomes real as we travel from the load toward the generator. The normalized resistance at that point is 0.38 . The real resistance is $50 \cdot 0.38$, or 19.0Ω . The distance from the load traveled to reach that point is 0.213 wavelength.

Using (11.11), we compute the characteristic impedance for the quarter-wave transformer as follows:

$$Z'_0 = (Z_0 Z_L)^{1/2} \quad (11.11)$$

$$Z'_0 = (50 \cdot 19.0)^{1/2} = 30.8\Omega$$

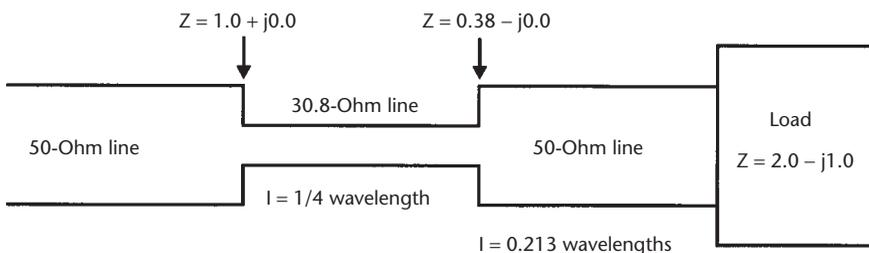


Figure 11.14 Example of impedance matching with a quarter-wavelength transformer.

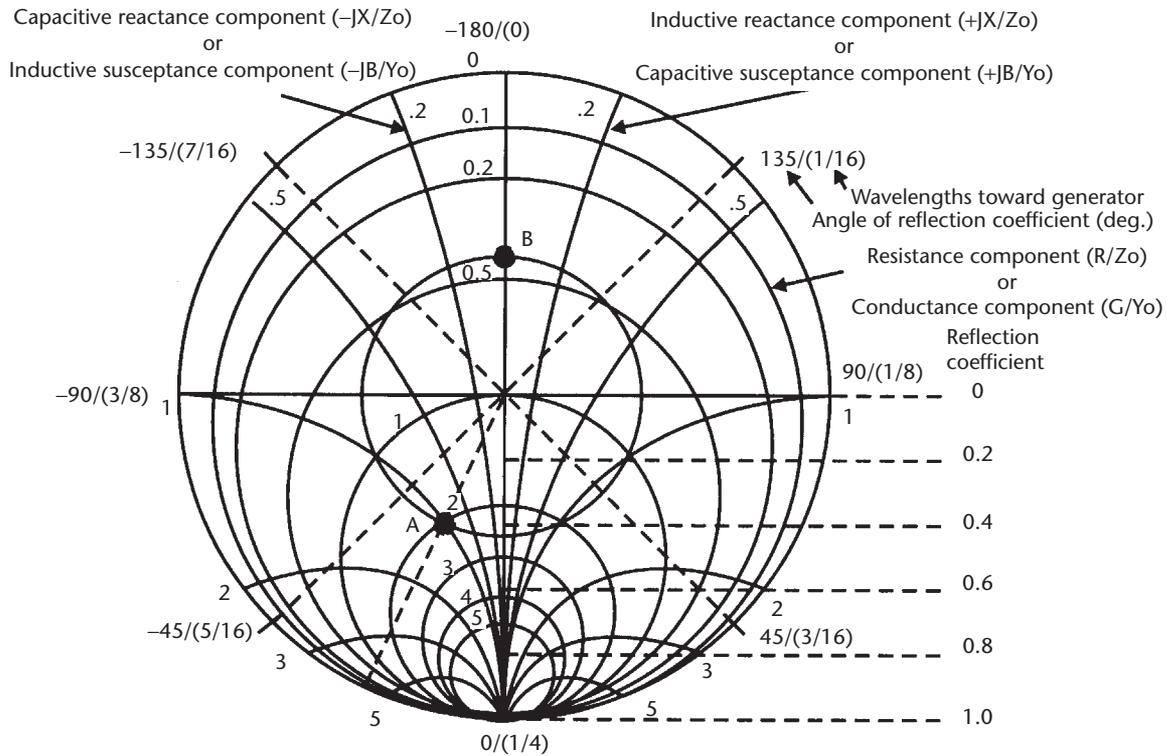


Figure 11.15 Example of the use of a Smith chart for impedance matching using a quarter-wavelength transformer.

The quarter-wave transformer provides a near-perfect match of impedance at only one frequency. It has a useful bandwidth that depends on the SWR that we can tolerate. An octave bandwidth is provided for the case of a VSWR of 1.5.

11.8.2 Impedance Matching with a Short-Circuited Stub

Figure 11.16 shows impedance matching using a single short-circuited stub. This method of impedance matching frequently is used with transmission lines at microwave frequencies.

The stub, which is a short section of line, is connected to the main line at a calculated distance from the load. It has the same characteristic impedance as the main line and a length that is calculated. The method of selecting the location and the length of the stub is illustrated in Figure 11.17.

For this example, assume a load of $75 - j100\Omega$ and a Z_0 of 50Ω . Thus, the normalized impedance is $1.5 - j2.0$. A VSWR circle drawn through that point shows that the VSWR without matching is about 4.7. The admittance for the load is found by transferring one-quarter wavelength around the VSWR circle to point 2. The normalized admittance is read as $0.2 + j0.3$. Next, move from that point on the VSWR circle to the point where the real part of the admittance is 1.0. The admittance at that point is $1.0 + j1.7$. The distance from point Q to point R is found from the chart to be about 0.13 wavelength. Therefore, the stub should be placed 0.13 wavelength from the load.

The susceptance of the stub should be $-j1.7$ at the frequency of interest to cancel out the $+j1.7$ of the admittance at point R. The length of the stub is found by moving from the position of the short when the chart is used as an admittance chart to the point on the outside of the chart where the susceptance is $-j1.7$. That distance is found to be about 0.08 wavelength. Adding $-j1.7$ to $1.0 + j1.7$ brings us along the 1.0 conductance circle to the $Y = 1.0 + j0.0$. That point is also the $Z = 1.0 + j0.0$ point.

Transmission line circuits normally are perfectly matched at only one frequency. At any other frequency, the match is not perfect. Usually, however, there is a substantial frequency range over which the match is good enough. Often, that frequency range is where the VSWR is 1.5 or less.

Figure 11.18(a) shows the case of a double-stub impedance-matching network. A double-stub matching system has the advantage that it can be placed almost any place on the main line that is convenient rather than at a critical point, as in the case of the single-stub system. The spacing of the stubs typically is one-eighth wavelength at the center frequency. The lengths of the shorted stubs A and B usually are selected

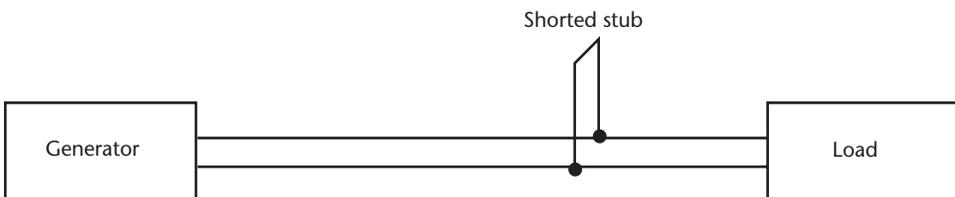


Figure 11.16 Short-circuited stub impedance matching.

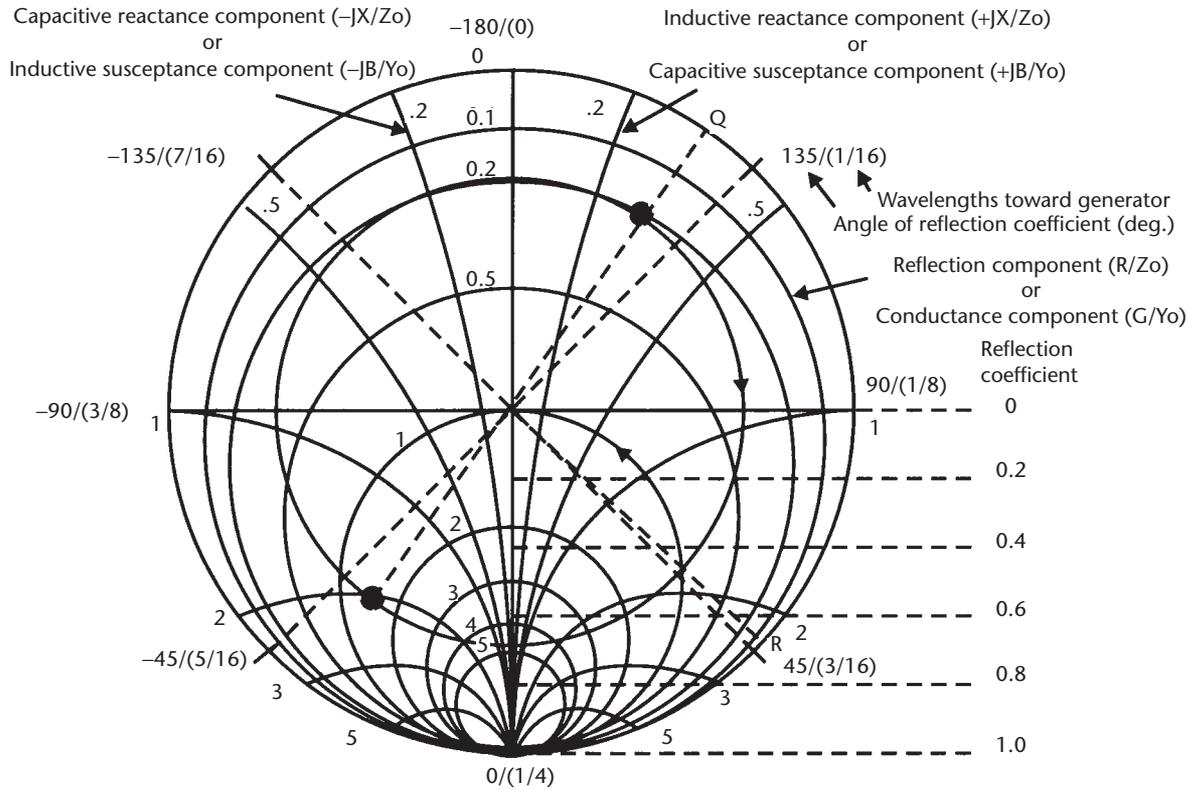


Figure 11.17 Example of the use of a Smith chart for impedance matching using a short-circuited stub.

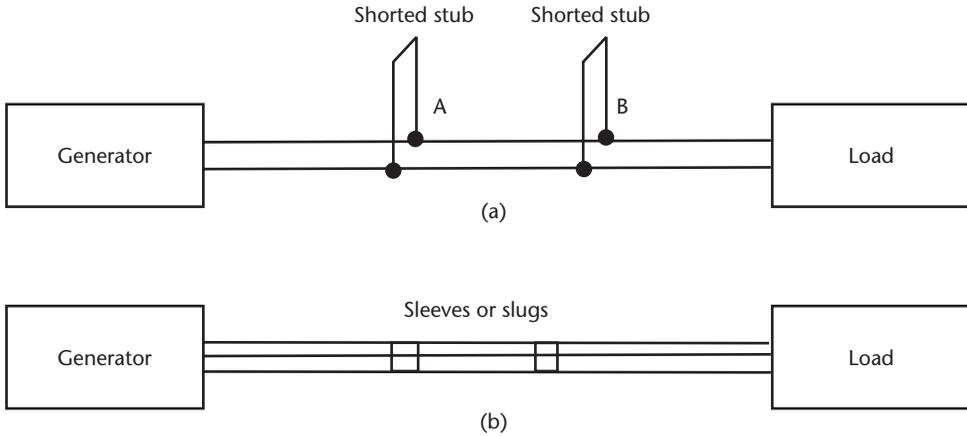


Figure 11.18 Double reflector matching: (a) double stub impedance matching and (b) double slug or double sleeve impedance matching.

by trial and error to reduce the VSWR to a low value. The type of stubs used typically involve adjustable shorting plungers. This type of system can achieve a perfect impedance match only if the conductive component of the impedance at the stub nearest the load and looking toward the load is less than $2/Z_0$. When this requirement is not satisfied, an impedance match still can be made by increasing the distance from the load by a quarter-wavelength. That is because of the impedance-transforming action of a quarter-wave line.

Figure 11.18(b) shows a double slug or sleeve tuner. The tuner uses either dielectric slugs or metallic sleeves, which can be moved along the line to adjust the phase of the reflections caused by the slugs or the sleeves. They can be adjusted in separation distance to adjust the magnitude of the reflections.

11.9 Coaxial Terminations

There are three main types of terminations for coax lines: short-circuit terminations, Z_0 terminations, and open-circuit terminations. Short-circuit terminations are used in impedance matching stubs and in some measurement systems for system calibration. Z_0 terminations also are used in measurement systems and in devices such as directional couplers and stripline isolators. High-power versions are used as transmitter dummy loads. Open-circuit line terminations are also used in impedance matching but not as often as shorted-stub terminations because of possible radiation problems. They are, however, used in coaxial measurement systems for calibration of vector impedance measurement systems, or network analyzers.

11.10 Coaxial Directional Couplers

Figure 11.19 shows two coaxial-type direction couplers. Figure 11.19(a) shows a loop-type directional coupler, which is a wideband coupler that can work over several octaves. The coupling from the main line to the output line is through a

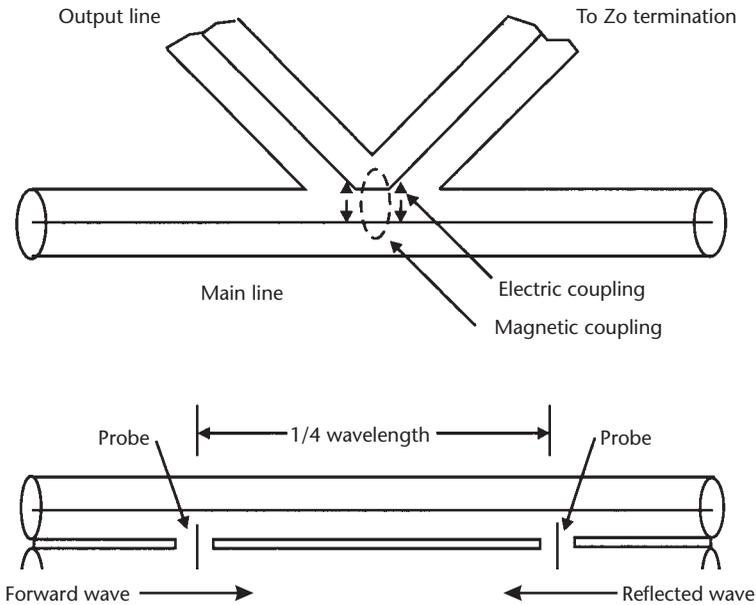


Figure 11.19 Coaxial-type directional couplers: (a) loop-type directional coupler and (b) two-hole probe-type directional coupler.

combination of capacitive coupling and magnetic or inductive coupling. By proper design, the two coupling mechanisms can be made equal such that for a forward wave on the main line the two equal amplitude and in-phase waves on the output line add, and the two equal amplitude and 180-degree out-of-phase signals on the terminated line add to zero. Typical coupling usually is in the range of 10–30 dB and is selected by loop design.

Figure 11.19(b) shows a two-hole probe directional coupler. In this type of coupler, the operating bandwidth is more limited. The probes are placed a quarter-wave apart at the center frequency. A forward wave on line A is measured at the right side of line B where the two equal amplitude signals coupled to line B by the probes add in phase. These signals are 180 degrees out of phase at the left side of B and add to zero. On the other hand, the reflected signal is measured at the left side of line B using the same addition and subtraction method. Coupling is usually in the range of 10–30 dB and is selected by probe design.

11.11 Baluns

Baluns are circuits or devices used to convert from a balance transmission line to an unbalanced transmission line or vice versa. A two-wire line is an example of a balanced line, while a coax is an example of an unbalanced line. Figure 11.20 shows two examples of baluns.

Figure 11.20(a) shows a very wideband balun using a modified length of coax. In this case, the transmission line on the left is a regular coaxial line, and the transmission line on the right is a two-wire line. The center portion of the line is the balun. It has a length of one wavelength or greater at the lowest frequency of

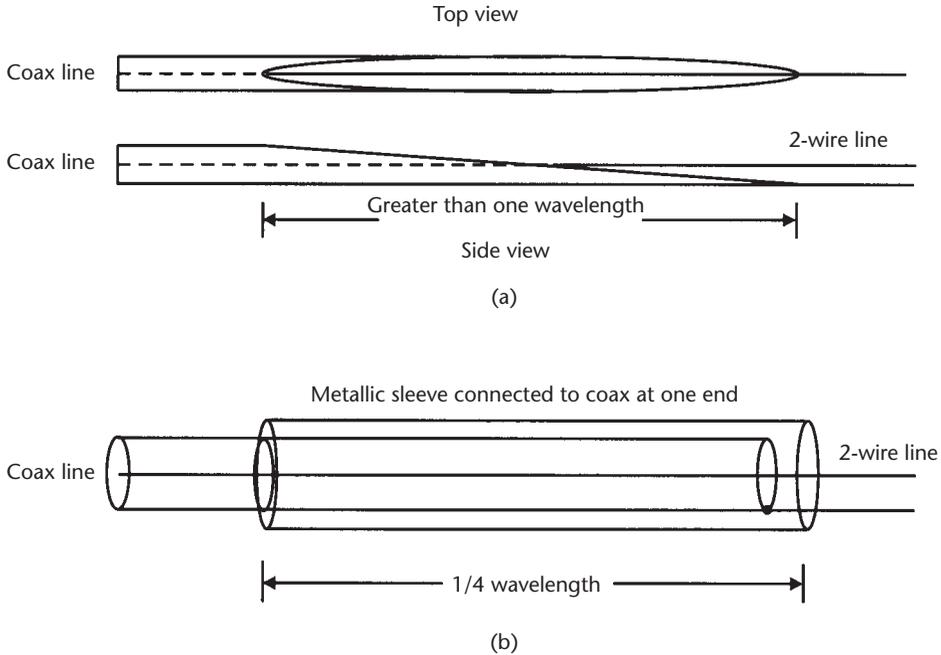


Figure 11.20 Two examples of baluns: (a) wideband balun and impedance transformer using a modified coax and (b) quarter-wave sleeve balun.

interest. The coax line is cut so the outer conductor gradually changes shape from a tube that surrounds the dielectric and the inner conductor to a half-sleeve and finally to a stripline to which a wire of the same diameter as the coax center conductor is connected. A balun of this type is sometimes used with log periodic or other wideband antennas having as much as a 10-to-1 frequency range.

Figure 11.20(b) shows a quarter-wave sleeve balun. In this case, a conducting sleeve is placed about the coaxial line spaced from the coax outer conductor over all but the left end where it is connected to the coax outer conductor. The short circuit of the sleeve and the outer conductor transforms to an open-circuit at the right end due to the quarter-wave transformer action. The inner conductor becomes one line of the two-wire line and the outer conductor is connected to a second line of the two-wire line. Again, the function of this circuit is to transform from an unbalanced line to a balanced line, where balanced and unbalanced refers to the electric field configurations.

11.12 Two-Wire Transmission Line Impedance Transformer

A wideband, two-wire line impedance transformer can be constructed by slowly changing the spacing between the two wires. This transformer has a length of at least one-half wavelength. It can be used over many octaves of frequency, such as a 10:1 frequency range. The length of the circuit is determined by the lowest frequency or the largest wavelength to be used. The characteristic impedance of the lines is determined by the size of the wires and the wire spacing. It also is possible to make a wideband

impedance transformer using microstrip or stripline using this same approach if the transition section is equal to or greater than a half-wavelength. It also is possible, but much more difficult, to make a coaxial wideband impedance transformer of this same type. Some of the information presented here is adapted from [3].

11.13 Stripline and Microstrip Circuits

Stripline and microstrip are used for many types of microwave circuits. This section discusses some of the more important of these circuits. Much of the information here regarding stripline circuits has been adapted from [2].

11.13.1 Shunt Stub DC Returns

Figure 11.21 shows a shunt stub as a dc return. The stub is connected to a ground at one end and to the main conductor at the other end. It has a length of one-quarter wavelength at the center frequency. With that length, the stub transforms from a short to an open-circuit at the main conductor attachment point for the microwave frequency, but it is a short circuit between the main line and the ground for dc. Shunt dc return stubs normally are formed using high-impedance lines.

11.13.2 Branch Line 90-Degree Hybrid Couplers

Figure 11.22 shows a two-arm branch line 90-degree hybrid coupler. The circuit consists of a main line, which is coupled to a secondary line by two quarter-wavelength lines spaced one-quarter wavelength apart, thus creating a square approximately one wavelength in circumference. The coupling factor is determined by the ratio of the impedance of the shunt and series arms, which also must be adjusted to maintain a proper impedance match over the band. Coupling to the secondary line typically is in the range of 3–26 dB. The phase difference between the two lines when used as a power divider is 90 degrees at the center frequency. The two-arm branch line 90-degree hybrid usually is used as a directional coupler, able to provide a measurement of both forward and reflected signals.

The circuit in Figure 11.22 shows a design in which $Z_0 = 50\Omega$ and the coupling factor is -3 dB. The direct output is also at -3 dB.

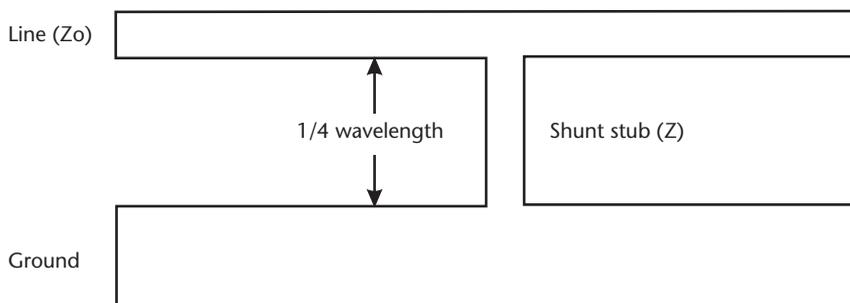


Figure 11.21 Shunt stub as a dc return. (After: [5].)

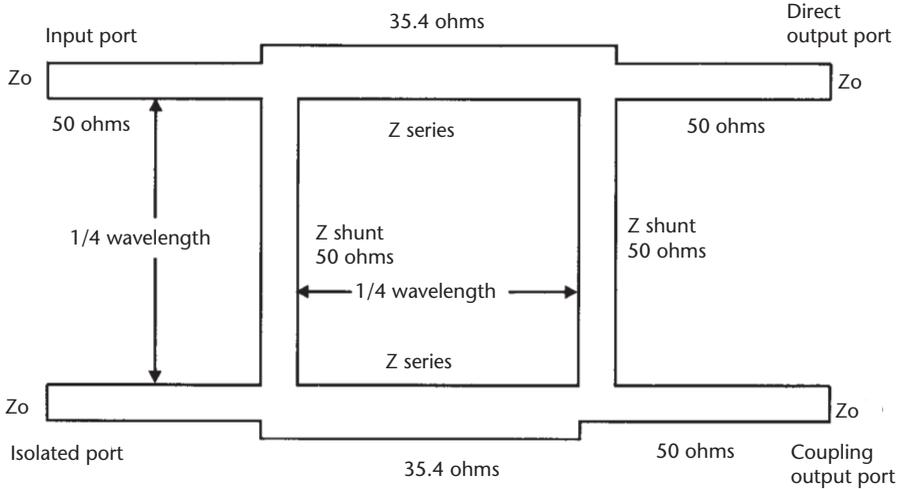


Figure 11.22 Branch line 90-degree hybrid coupler. (After: [6].)

11.13.3 Stripline or Microstrip Rat Race Hybrid Coupler

Figure 11.23 shows the case of a 1.5-wavelength “rat-race magic T.” It consists of a ring 1.5 wavelengths at the center frequency in circumference having four arms separated by 60 degrees of angular rotation. When used as a zero-degree phase-shift power divider, the input arm is arm 1. There are two output arms (arms 2 and 4) spaced one quarter-wavelength away, and a fourth terminated arm (arm 3) spaced a quarter-wavelength away from arm 4 and three-quarters of a wavelength away from arm 2. At center frequency, the output split from the common input arm 1 to the two

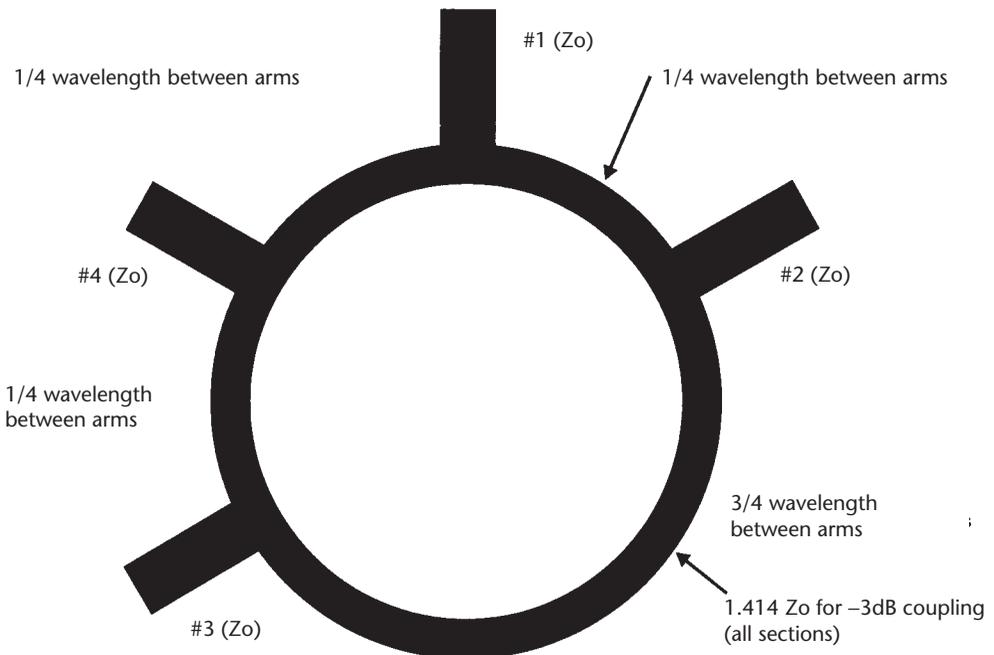


Figure 11.23 A 1.5-wavelength rat-race hybrid coupler. (After: [6].)

output arms 2 and 4 is equal, and the phase relationship between them is zero degrees.

When used as a 180-degree phase-shift power divider, the input arm is arm 2, and arm 4 is terminated. The power split to the two output arms 1 and 3 is also equal; however, the phase shift between them is 180 degrees. In general, the rat-race coupler has more bandwidth than the branch-line coupler, but the choice between couplers usually is made based on the phase difference needed for a particular application. The branch-line coupler provides a 90-degree phase shift between outputs, whereas the rat-race coupler provides either a zero-degree phase shift between outputs or, if desired, a 180-degree phase shift between outputs. For a 50Ω input line, the impedance of the one-quarter and the three-quarter wavelength sections is the square root of 2 times 50Ω for equal power division.

11.13.4 Split Inline Hybrid Dividers and Combiners

Figure 11.24 shows two types of split inline hybrid dividers or combiners. These circuits can be used either to split the input power between two loads or to combine two input signals into a single output.

Figure 11.24(a) is an uncompensated divider or combiner. The arm length is one-quarter wavelength at the center frequency. For a 50Ω input line, the impedance of the arms is 70.7Ω . A 100Ω resistor is placed between the output arms, as shown, for increased isolation between output ports.

Figure 11.24(b) shows a compensated inline divider or combiner. The difference between this combiner and the uncompensated system is the use of an

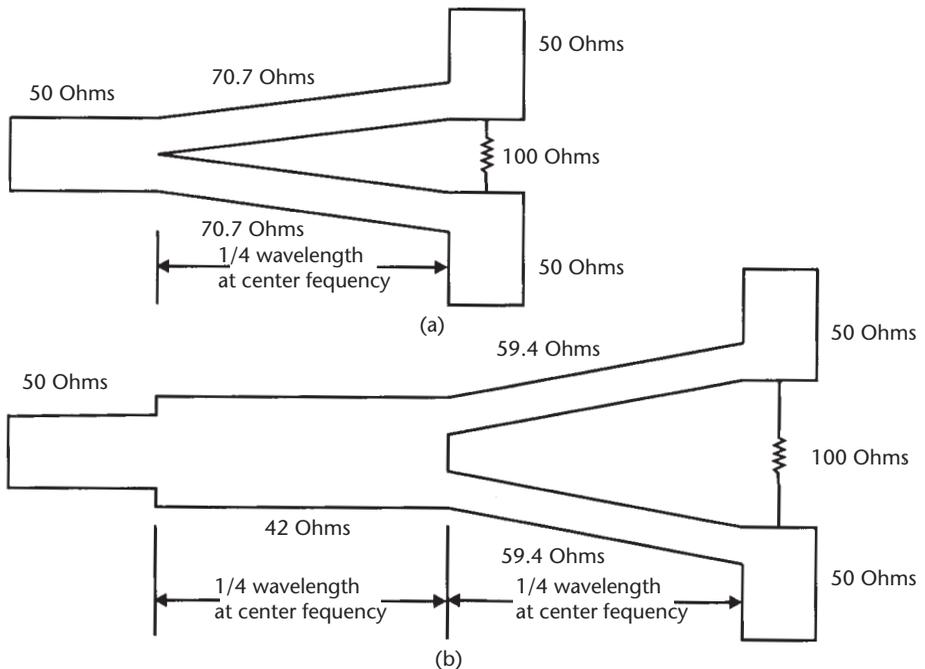


Figure 11.24 Two types of split inline hybrid dividers or combiners: (a) uncompensated inline hybrid divider and (b) compensated inline hybrid divider. (After: [6].)

additional quarter-wave section of line between the input line and the split. Different characteristic impedances are used for the arms, as shown. The VSWR characteristics for this divider or combiner are somewhat better than those of the uncompensated divider or combiner for the case of wideband operation. We see that a perfect VSWR is realized at only one frequency; however, a fairly wide bandwidth is available if we are willing to have a 1.7 or less VSWR. This is usually the case.

The systems shown are equal impedance systems for the two arms. It is also possible to make systems for coupling to lines of different impedance and to use more complex divider or combiner circuits that have multiple-octave coverage. Systems also are available that provide 1-to-4 and 1-to-8 division or combining ratios. These systems simply combine a number of the basic 1-to-2 ratio units.

11.13.5 Quarter-Wave Coupled-Line Directional Couplers

Figure 11.25 shows two stripline quarter-wave coupled-line directional couplers. The system shown in Figure 11.25(a) is an edge-coupled system; the system shown in Figure 11.25(b) has one line above the other separated by a dielectric spacer, normally considered broadside coupling in stripline. A fraction of the power on the main transmission line is coupled to the secondary line. This power varies as a function of the physical dimensions of the coupler and the direction of the propagation of the primary power. Port 1 is the input and port 4 is the output of the main line. Port 2 is the coupled output, while port 3 normally is terminated in the characteristic impedance and is isolated from the input port by 20 dB or so over an octave

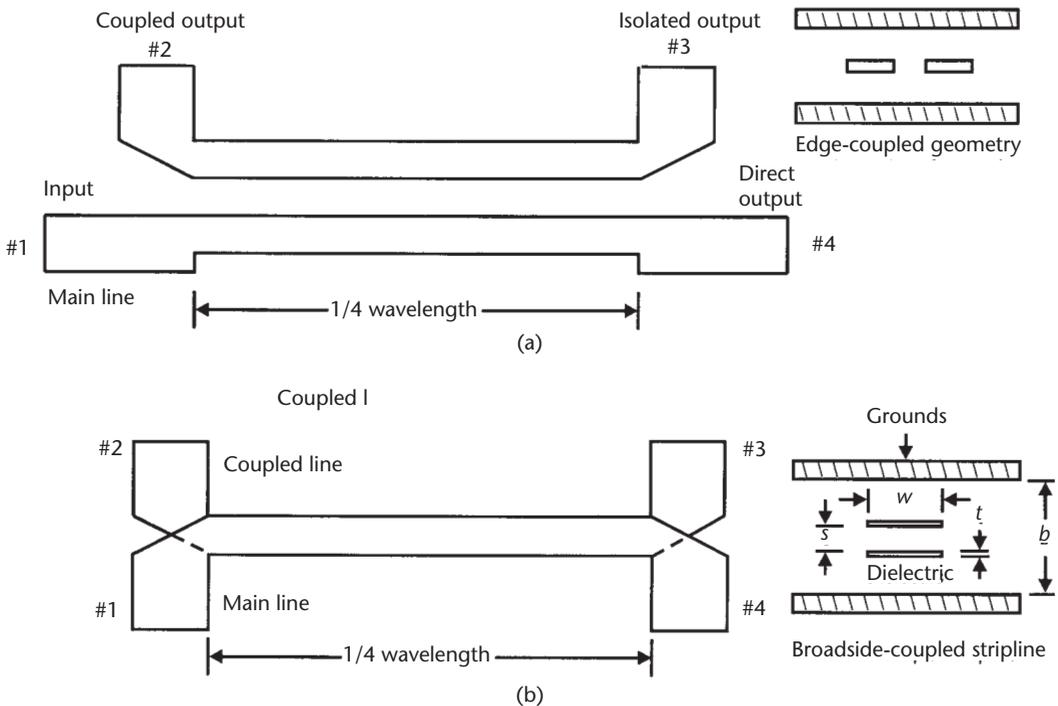


Figure 11.25 Stripline quarter-wave directional couplers: (a) edge-coupled directional coupler and (b) broadside-coupled directional coupler. (After: [7].)

bandwidth. The output power from port 2 is proportional to the power traveling from port 1 to port 4. The usable bandwidth of this coupler is about one octave for coupling, which varies by $\pm 10\%$ from its nominal value.

11.13.6 90-Degree Coupled-Line Hybrid Coupler

The broadside-coupled directional coupler in Figure 11.25(b) can be used as a 3-dB, 90-degree hybrid coupler. This 90-degree hybrid is perhaps the most useful of all hybrids and is used in many microwave circuits. The symbol for this hybrid is shown in Figure 11.26(a). A typical frequency response for this hybrid coupler is shown in Figure 11.26(b).

The single-section quarter-wave directional coupler has a 90-degree or quadrature phase relationship between the outputs independent of frequency for a perfect device. In reality, this phase relationship is 90 degrees, $\pm 2\text{--}3$ degrees, over an octave, which is useful for many applications. It generally is constructed as a completely overlapped coupler with a crossover. Input power splits equally between output 1 and output 2. The fourth arm is terminated in Z_0 . As in the case of the looser values of coupling, it frequently is desirable to couple at the center frequency in such a way as to provide a plus-or-minus tolerance around the nominal design frequency. Thus, an octave bandwidth -3-dB coupler normally is designed for -2.7 dB midband coupling. The output at the through port will be -3.3 dB with no other losses present. At extremes of 2:1 band, coupled output will be -3.3 dB and direct output will be -2.7 dB, providing 3 dB, ± 0.3 dB, over the octave bandwidth.

11.13.7 Stripline Lowpass Filters

Figure 11.27 shows a stripline lowpass filter. Also shown is the equivalent lumped element circuit. In the stripline circuit, inductors are narrow lines with small

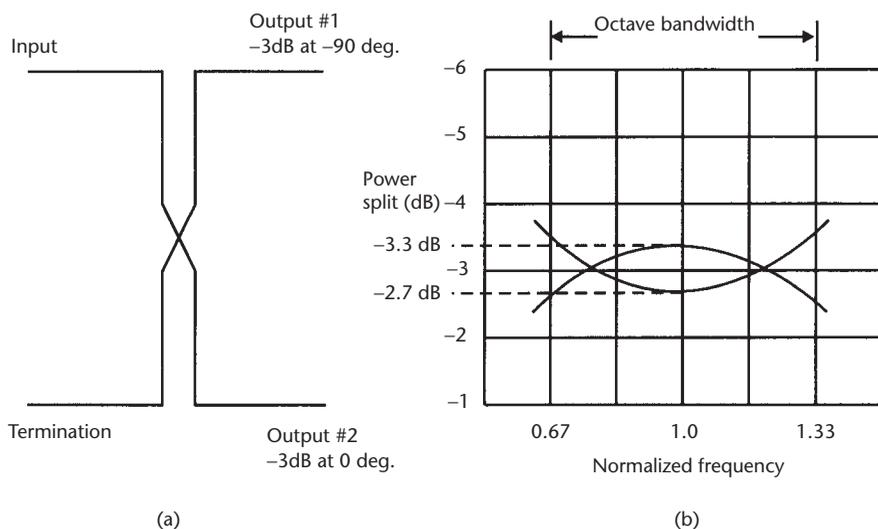


Figure 11.26 A 90-degree hybrid coupler: (a) symbol for 90-degree hybrid coupler and (b) typical frequency response for 90-degree hybrid coupler. (After: [7].)

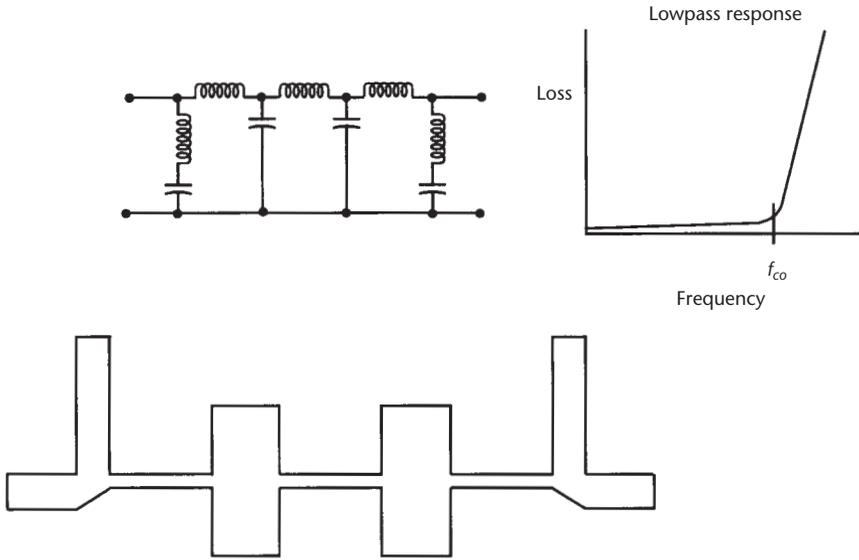


Figure 11.27 Example of a stripline lowpass filter. (After: [8].)

capacitance to ground and high characteristic impedances, while capacitors are fat lines with large capacitance to ground and low characteristic impedances. The two end lines are equivalent to LC series resonant circuits. All lengths are short compared to a wavelength. They usually are less than 0.1 wavelength, to resemble lumped elements.

A lowpass filter is designed to provide low loss to an RF signal for frequencies below a selected design frequency, f_{co} . Above that cutoff frequency, the loss or attenuation increases rapidly with an increase in frequency, reaching a high value at a selected second frequency. Above the second frequency, the attenuation remains high for increasing frequency. Designs can be Chebychev or Butterworth. The rate of attenuation depends on the number of sections in the design.

11.13.8 Stripline Highpass Filters

Figure 11.28 shows a stripline highpass filter. The shunt inductors are narrow lines with small capacitances to ground. The lengths are critical and are determined by analysis. Overlap capacitors are used on the main line between the shunt inductors. All elements are short compared to a wavelength and usually are less than 0.1 wavelength.

A highpass filter is designed to provide high loss or high attenuation to an RF signal for frequencies below a selected design frequency. Above that frequency, the loss or attenuation decreases rapidly with increase in frequency, reaching a low value at a selected second frequency. Above the second frequency, the attenuation remains low for increasing frequency. Again, attenuation versus frequency depends on the number of elements in the filter and whether a maximally flat (Butterworth) or equal-ripple filter prototype is used.

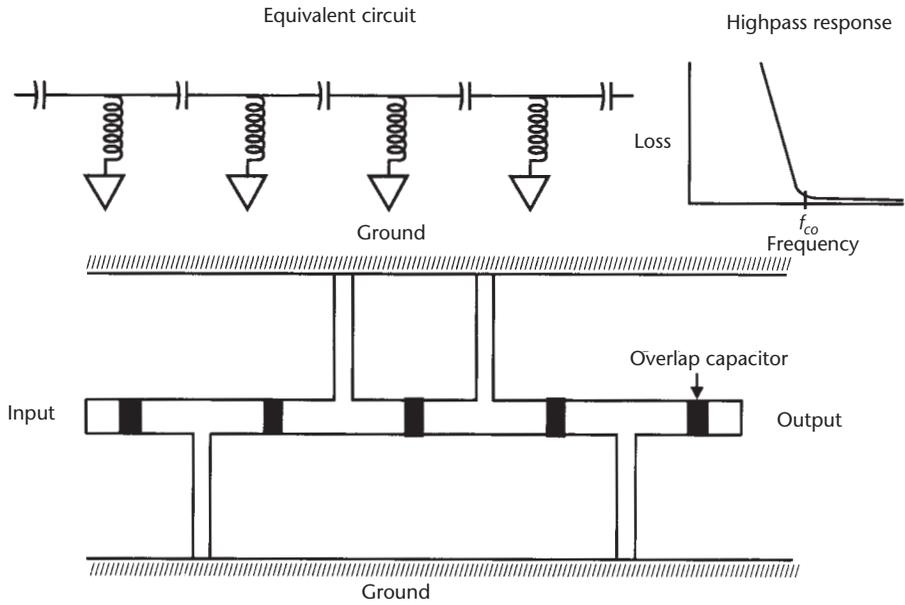


Figure 11.28 Example of a stripline high-pass filter. (After: [8].)

11.13.9 Stripline Bandpass Filters

Figure 11.29(a) shows a side-coupled half-wave resonator bandpass filter. This circuit can be readily printed and has the advantage of providing dc isolation. The type of system shown is a folded configuration. There are a number of other possibilities

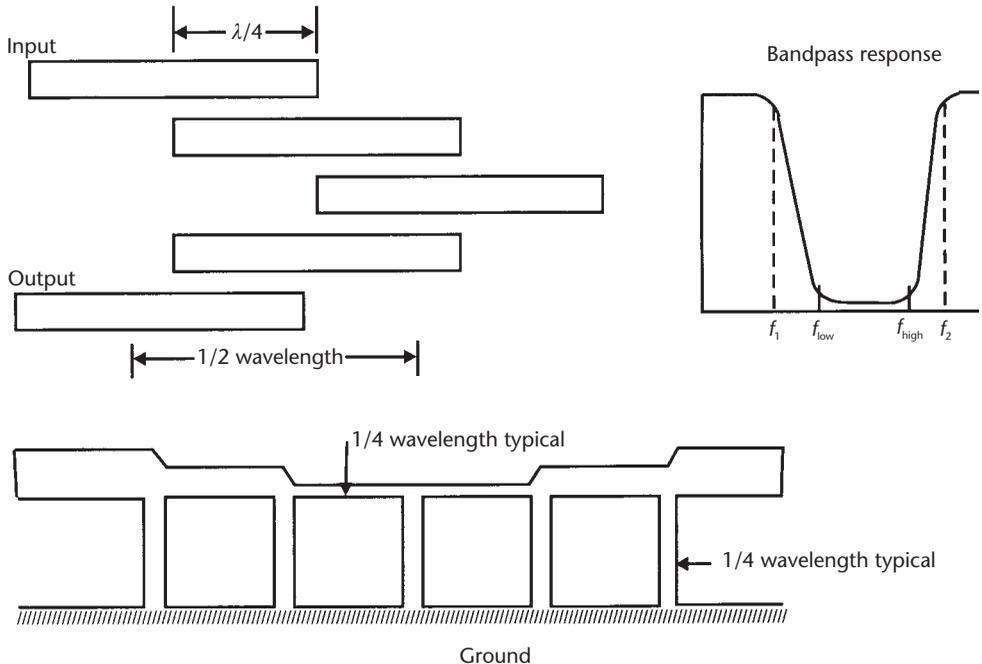


Figure 11.29 Stripline bandpass filters: (a) side-coupled half-wave resonator bandpass filter and (b) short-circuited quarter-wave stub bandpass filter. (After: [8].)

for implementing systems of this type involving multiple resonators. Bandwidths from 3% to approximately 20% or more are possible.

Figure 11.29(b) shows a short-circuited quarter-wave stub bandpass filter. This direct coupled filter consists of a series of short-circuited quarter-wavelength stubs separated by quarter-wavelength sections of line. This type of bandpass filter is used mainly for the wide passband filters (30% or greater), whereas the circuit of Figure 11.29(a) is used mainly for medium and narrow bandwidth filters.

A stripline bandpass filter is designed to provide high loss or high attenuation to an RF signal for frequencies below a selected design frequency, f_1 . Above that frequency, the loss or attenuation decreases rapidly with increase in frequency, reaching a low value at a selected second frequency, f_{low} . Above the second frequency, the attenuation remains low until a third frequency, f_{high} , is reached. Above the third frequency, the attenuation increases rapidly, reaching a high value of attenuation at a fourth frequency, f_2 . Above the fourth frequency, the attenuation remains high with increasing frequency. The bandwidth of stripline bandpass filter usually is defined for Butterworth or maximally flat filters as the width between half power points on the response curve. For Chebychev designs, it is the equal-ripple bandwidth, usually less than the 3-dB bandwidth.

11.14 Ferrite Circulators and Isolators

Figure 11.30 shows one version of a miniature Y-junction three-port ferrite circulator. This circuit consists of a stripline substrate, a circular ferrite disk, a single bias magnet in the case of microstrip or two magnets in the case of stripline, and a shielding and support can with top and bottom covers. With suitable magnetic field

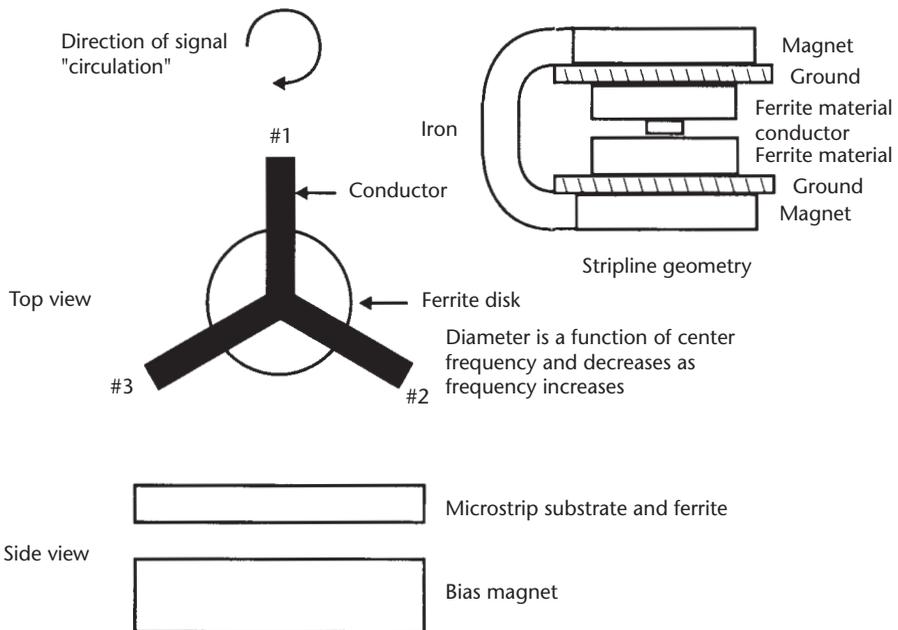


Figure 11.30 Diagram of a Y-type ferrite circulator.

strength, a phase shift is applied to any signal fed in to the circulators in Figure 11.30. If the three striplines are arranged 120 degrees apart, as shown, with proper dimension, the input signal splits into two counterrotating signals that, through the magnetic field application, add in phase at one port and out of phase at the third port. For input at port 1 with clockwise circulation, signals add at port 2, where all the signal emerges, and out of phase at port 3, which becomes isolated. Similarly, with port 2 as the input, all signal emerges from port 3. With port 3 as an input, all signal emerges from port 1. Thus, in a duplexer operation, the transmitter is connected to port 1, the antenna to port 2, and the receiver to port 3, as shown in Figure 11.31.

An isolator is a device that allows signals to pass in only one direction. Such an isolator can be obtained from the basic Y circulator by terminating one arm of the circulator in Z_0 . Improved isolation is possible by connecting two or more isolators in series.

A four-port circulator is obtained by joining two Y-junction circulators, as shown in Figure 11.32. This type of system can be used as a switch for transceiver systems.

Figure 11.32(a) is a schematic diagram for a four-port ferrite switch for the condition in which the transmitter is connected to the antenna for transmission. Any reflected power from the antenna is routed to a dummy load rather than to the receiver.

Figure 11.32(b) shows the condition in which the direction of the magnetic field on the right-hand circulator is switched, and the combination routes the received signal from the antenna to the receiver.

M/A-COM, Inc., Burlington, Massachusetts, makes a small ferrite circulator unit that is similar in function to the unit shown in Figure 11.30. The M/A-COM unit is constructed on a ferrite substrate known as a Ferrodisc. It is a true microstrip device using a single ferrite element with a circuit pattern on one face and the ground plane on the other. A two-element permanent magnet structure is included on the ferrite substrate. When one port is suitably terminated, typically in 50Ω , the device can be used as an isolator. Units are sold by M/A-COM covering the

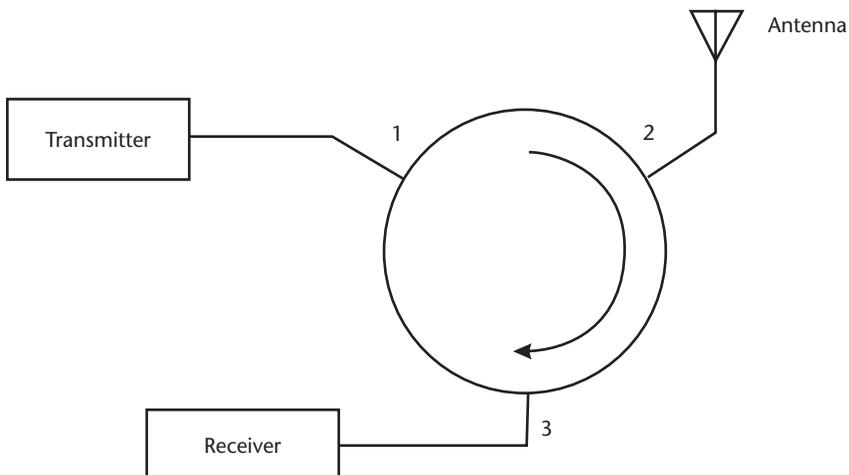


Figure 11.31 Duplexer arrangement using a circulator.

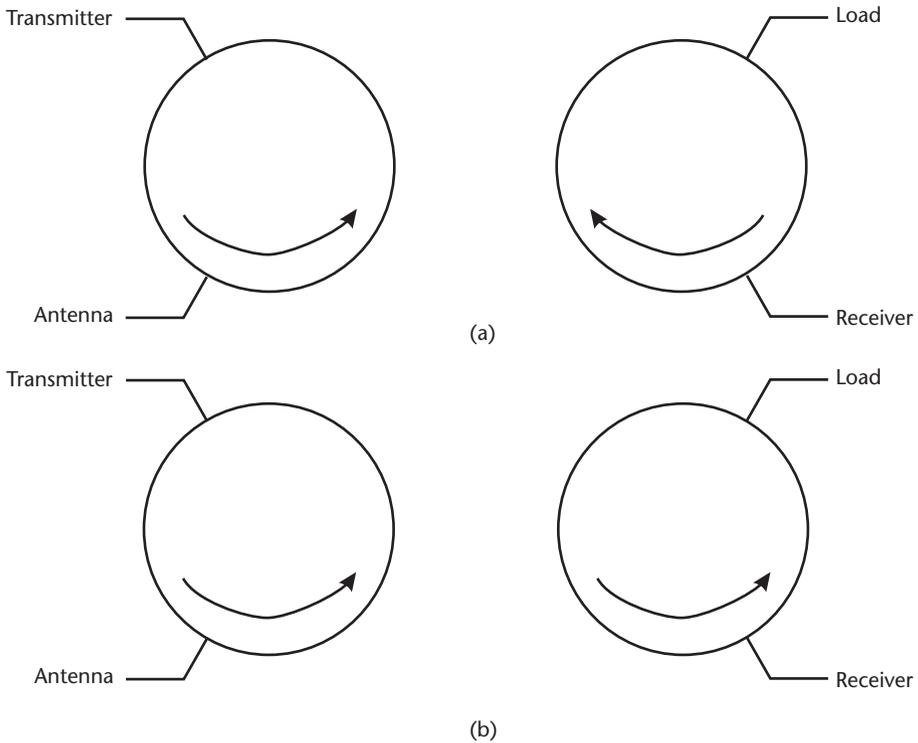


Figure 11.32 Schematic diagrams for ferrite switches: (a) transmit state and (b) receive state.

frequency range of 1.7–17.5 GHz. Typical insertion loss is 0.5 dB, and typical isolation is 18 dB. VSWRs of 1.3:1 are typical. Ferrite isolators are available from other sources that will work at frequencies as low as 100 MHz and above 17.5 GHz. Narrower bandwidth units typically exhibit lower insertion loss, higher isolation, and lower VSWR.

Some of the material in Section 11.14 includes information from [9].

11.15 Coaxial Electromechanical Switches

Coax electromechanical switches frequently are used in RF systems. The most common type is the single-pole, double-throw (SPDT) coaxial switch. An SPDT switch is used to switch a common coaxial input to either of two coaxial outputs or either of two coax inputs to a common output. An SPDT switch is shown in Figure 11.33.

A typical standard SPDT switch (M/A-COM part no. 7524-6132-00) is useful over the frequency range of dc to 12.4 GHz or beyond. It has a CW power rating of 100W below 4 GHz, 80W in the 4–8-GHz range, and 75W in the 8–12.4-GHz range. Typical insertion loss is 0.3 dB below 4 GHz, 0.4 dB in the 4–8 GHz range, and 0.5 dB in the 8–12.4-GHz range. Nominal isolation between ports is 70 dB below 4 GHz, 65 dB in the 4–8-GHz range, and 60 dB in the 8–12.4-GHz range. Switching time is less than 30 ms. Actuating voltage is 20–30V dc and actuating current is 311 mA at 28 volts dc. The life of this type of switch is greater than 1 million cycles.

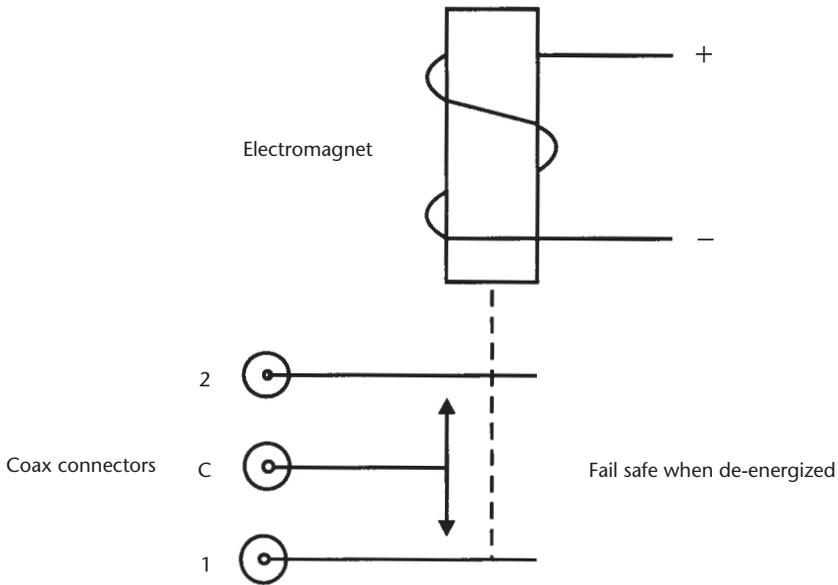


Figure 11.33 Coax electromechanical SPDT switch.

A typical miniature SPDT electromechanical switch using SMA connectors (M/A-COM part no. 7530-6412-00) is available for operation from dc to 26 GHz. The power ratings are in the range of 60–20W, depending on frequency; insertion loss is in the range of 0.15–0.6 dB, depending on frequency; and isolation between ports is in the range of 80–60 dB, depending on frequency.

Another type of electromechanical switch sometimes used in RF systems is a transfer switch. A switch of this kind is shown in Figure 11.34. It has four ports and provides two separate RF paths that switch simultaneously when actuated. The paths between terminals 1 and 2 and between terminals 3 and 4 normally are closed, while paths between terminals 1 and 4 and between terminals 2 and 3 normally are open. When actuated, the reverse condition exists.

A typical standard four-port transfer electromechanical switch (M/A-COM part no. 7525-6320-00) uses a type-N connector, and has a switching time of 50 ms max. Actuating voltage is 20–30-V dc, and actuating current is about 373 mA at 28 volts dc. Operating frequencies are in the range of dc to 12.4 GHz. The CW power rating is in the range of 100–75W, the insertion loss is in the range of 0.3–0.5 dB, and the isolation between ports is in the range of 70–60 dB, depending on frequency. Miniature four-port transfer switches using SMA connectors are available for operation from dc to 18 GHz at lower power ratings.

It is also possible to obtain miniature and standard-size single-throw multiposition coax switches. These switches have a common input with multiple outputs or many inputs with a common output. The switches are used for selecting or combining signals. A typical miniature SPMT electromechanical switch provides SP6T operation from dc to 26.5 GHz in a very small package (1.75 × 1.75 inches). The life of the switch is greater than one million cycles. Power ratings, insertion loss ratings, and isolation ratings are similar to the miniature SPDT switch.

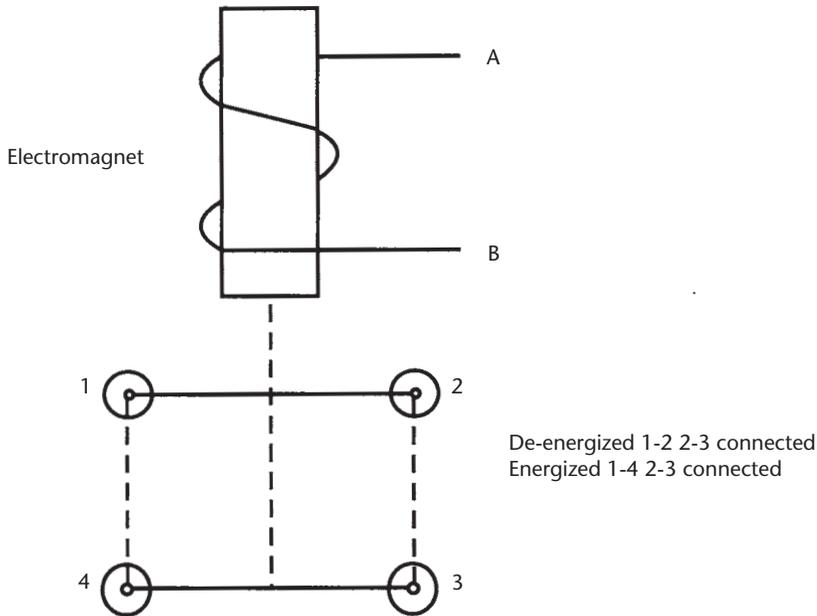


Figure 11.34 Electromechanical transfer switch.

11.16 PIN Diode Switches

A number of possible types of solid-state switches are used with transmission lines. Perhaps the most important of these are the PIN diode switches. Both series and shunt diode switches are used, as are series shunt combinations. When biased in the forward state, the equivalent circuit for the diode is a low-value resistor in series with package inductance. It acts like a short circuit to RF. When reverse biased, the equivalent circuit is a high-value resistor in series with the junction capacitance and that in parallel with the package capacitance. It acts almost as an open circuit to RF.

One important feature of PIN diode switches is that switching times are typically $1 \mu\text{s}$ or less. This short switching time is important for a number of radar applications.

An important application for PIN diode switches with radar is for phase-shifters used with phased array antennas. A system of this kind is shown in Figure 11.35.

An example phase-shifter system uses a 3-dB hybrid coupler plus two coax lines terminated with short circuits and two PIN diode switches. The 3-dB hybrid coupler has the property that a signal input at port 1 is divided equally in power between ports 2 and 3. No energy appears at port 4. The two PIN diode switches act to either pass or reflect the signals at ports 2 and 3. When the impedance of the diodes is such as to pass the signals, the signals are reflected by the short circuits located farther down the transmission lines. The signals at ports 2 and 3, after reflections from either the diode switches or the short circuits, combine at port 4. None of the reflected energy appears at port 1. The difference in path length with the diode switches open and closed is d . The two-way path, $2d$, is chosen to correspond to the desired increment of phase shift. An N -bit phase shifter can be obtained by cascading N such hybrids.

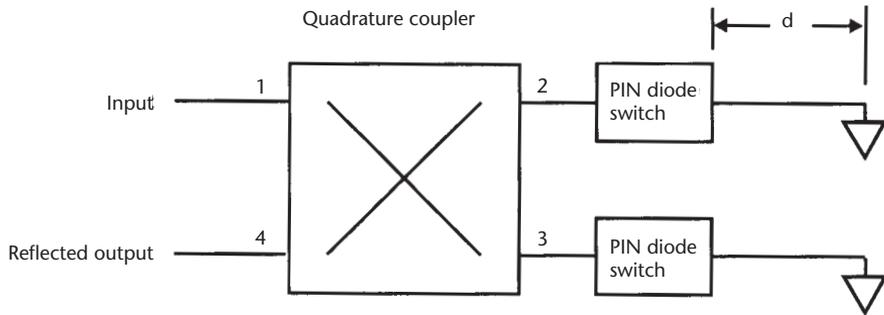


Figure 11.35 PIN diode phase-shifter.

11.17 Sparkgap Switches for Lightning Protection

Coaxial gas-filled tube, sparkgap switches often are used for lightning protection on coaxial lines connected between antennas and transmitters or receivers. Normally, they are open-circuits and pass the RF signals that enter the units. If lightning strikes the antenna and a high voltage is produced on the transmission line, the gas tubes fire and produce an effective short circuit across the line. The outside conductor of the switch is connected to the earth by a good grounding rod or other means. That permits the lightning current to pass to the earth. The sparkgap switches thus act as protective means to the transmitter and receiver systems connected to the transmission lines.

References

- [1] Itoh, T., "Transmission Lines," in *Reference Data for Engineers*, 8th ed., M. E. Van Valkenburg, (ed.), SAMS, Carmel, IN: Prentice Hall Computer Publishing, 1993, pp. 29-21-29-39.
- [2] Howe, H., Jr., *Stripline Circuit Design*, Dedham, MA: Artech House, 1974, pp. 14-31.
- [3] Kennedy, G., *Electronic Communication Systems*, 3rd ed., New York: McGraw-Hill, 1985, pp. 169-201.
- [4] Smith, P. H., "A Transmission Line Calculator," *Electronics*, Vol. 12, January 1939, p. 2931.
- [5] Howe, H., Jr., *Stripline Circuit Design*, Dedham, MA: Artech House, 1974, pp. 53-54.
- [6] Howe, H., Jr., *Stripline Circuit Design*, Dedham, MA: Artech House, 1974, Chap. 3.
- [7] Howe, H., Jr., *Stripline Circuit Design*, Dedham, MA: Artech House, 1974, Chap. 5.
- [8] Howe, H., Jr., *Stripline Circuit Design*, Dedham, MA: Artech House, 1974, Chap. 6.
- [9] Kennedy, G., *Electronic Communication Systems*, 3rd ed., New York: McGraw-Hill, 1985, pp. 334-340 and pp. 345-346.

Selected Bibliography

- Gupta, K. C., et al., *Microstrip Lines and Slotlines*, 2nd ed., Norwood, MA: Artech House, 1996.
 Kennedy, G., *Electronic Communication Systems*, 3rd ed., New York: McGraw-Hill, 1985, pp. 334-340.

Waveguides and Waveguide-Related Components

This chapter discusses waveguides and related components, including hybrid junctions, impedance-matching components, resistive loads and attenuators, directional couplers, ferrite isolators, ferrite circulators, ferrite switches, detectors, mixers, gas-tube switches, duplexers, and cavity resonators. Much of the information presented in this chapter is adapted from [1–4].

12.1 Introduction to Waveguides

A waveguide is a hollow conducting tube used to transmit electromagnetic waves. Waveguides are an alternative to transmission lines at UHF and higher frequencies, providing much lower losses and much higher power-handling capability than transmission lines. They are not used at lower frequencies because of the very large size that would be required.

Any configuration of electric and magnetic fields that exists inside a waveguide must be a solution of Maxwell's equations. In addition, those fields must satisfy the boundary conditions imposed by the walls of the guide. To the extent that the walls are perfect conductors, there can be no tangential component of electric field at the walls. Many different field configurations can be found that meet those requirements. Each such configuration is termed a mode.

A critical examination of the various possible field configurations or modes that can exist in a waveguide reveals that they all belong to one of two fundamental types. In one type, the electric field is everywhere transverse to the axis of the guide and has no component anywhere in the direction of the guide axis. The associated magnetic field does, however, have a component in the direction of the axis. Modes of this type are termed transverse electric (TE) modes. In the other type of distribution, the situation with respect to the fields is reversed. The magnetic field in that case is everywhere transverse to the guide axis, while at some places the electric field has components in the axial direction. Modes of this type are termed transverse magnetic (TM) modes. The different modes of each class usually are designated by double subscripts, such as TE_{10} .

The behavior of a waveguide is similar in many respects to the behavior of a transmission line. Thus, waves traveling along a waveguide have a phase velocity and are attenuated. When a wave reaches the end of a waveguide, it is reflected back unless the load impedance is the same as the characteristic impedance of the waveguide. An irregularity in a waveguide produces reflections, just an irregularity in

any other transmission line does. Again, reflected waves can be eliminated by the use of an impedance-matching network, exactly as with other forms of transmission lines. When both incident and reflected waves are present simultaneously in a waveguide, the result is a standing-wave pattern that can be characterized by an SWR.

In some respects, waveguides are different from other forms of transmission lines in their behavior. One difference is that a particular mode will propagate down a waveguide with low attenuation only if the wavelength of the wave is less than some critical value determined by the dimensions and the geometry of the guide. If the wavelength is greater than the critical cutoff value, the waves in the waveguide die out rapidly in amplitude even when the walls of the guide are of material that has high conductivity. Different modes have different values of cutoff wavelength. The particular mode for which the cutoff wavelength is greatest (lowest cutoff frequency) is termed the dominant mode.

12.2 Rectangular Waveguides

The most frequently used type of waveguide has a rectangular cross-section, as illustrated in Figure 12.1(a). The width of the guide (indicated by a) is usually twice the height of the guide (indicated by b). In such a guide, the preferred mode of operation

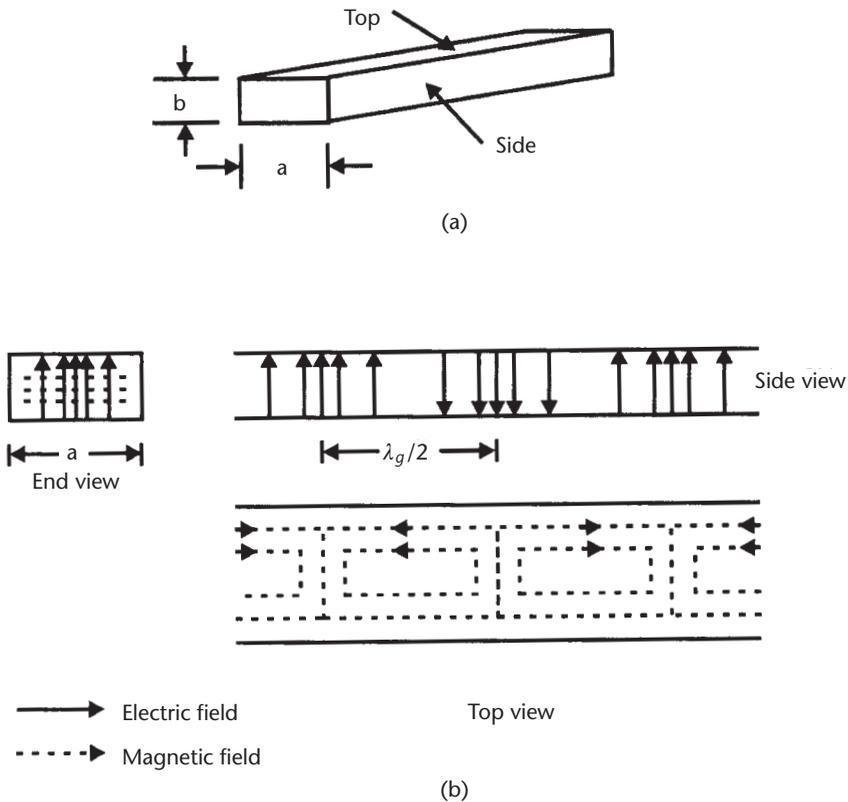


Figure 12.1 Rectangular waveguide and field configurations for dominant mode: (a) rectangular waveguide and (b) field configurations for the dominant or TE_{10} mode. (After: [4].)

is the dominant or lowest order mode. Field configurations for the dominant mode are illustrated in Figure 12.1(b). This mode is the TE_{10} mode. Here, the electric field is transverse to the guide axis and extends between the top and bottom walls. The intensity of the electric field is maximum at the center of the guide and drops off sinusoidally to zero intensity at the edges. The magnetic field is in the form of loops that lie in planes at right angles to the electric field, that is, planes parallel to the top and the bottom of the guide. The magnetic field is the same in all these planes, irrespective of the position of the plane along the vertical axis. This field configuration travels along the waveguide axis. As it travels down the guide, the amplitude is reduced as a result of energy losses in the walls of the guide. The wave drops back in phase with distance, just as it does in analogous transmission lines.

In the term TE_{10} , the subscript 1 means that the field distribution in the direction of the long side of the waveguide contains one-half cycle of variation. The subscript 0 indicates that there is no variation in either the electric or magnetic field strength in the direction of the short side of the guide. For the dominant mode in the rectangular waveguide, the cutoff wavelength, λ_c , is exactly twice the width, a , of the guide, that is, $\lambda_c = 2.0a$. For example, to transmit 300 MHz where the wavelength is 1m, the guide width must be greater than 50 cm. To transmit 3 GHz where the wavelength is 10 cm, the guide width must be greater than 5 cm. If the frequency is 100 MHz, the wavelength is 3m, and the width of the waveguide must be greater than 1.5m. Clearly, it is not practical to use a waveguide for VHF and lower frequencies, but it is practical to use a waveguide for UHF and higher frequencies.

As pointed out earlier, each mode that can exist in a waveguide has its own cutoff wavelength. The useful frequency range for a waveguide is somewhat less than the frequency range between the dominant mode cutoff frequency and the next higher mode cutoff frequency.

In Figure 12.1, the length, $\lambda_g/2$, indicates one-half guide wavelength. The axial length, λ_g , corresponds to one cycle of variation of the field configuration in the axial direction. It is related to the free-space wavelength, λ , and the cutoff wavelength, λ_c , according to the following equation:

$$\text{Guide wavelength} = \lambda_g = \lambda / \left[1 - (\lambda / \lambda_c)^2 \right]^{1/2} \quad (12.1)$$

where

$$\lambda = \text{wavelength} = c/f$$

$$c = \text{speed of light}$$

$$f = \text{frequency}$$

The phase velocity, v_p , is the distance the wave travels in 1 second. It is related to the velocity of light, c , by the following equation:

$$v_p / c = \lambda_g / \lambda \quad (12.2)$$

It is seen that the velocity of propagation always exceeds the velocity of light. As the frequency is lowered so that it approaches the cutoff value, the phase velocity increases and becomes infinite at cutoff.

The phase velocity v_p is an apparent velocity deduced from the rate of phase change with position along the axis. The actual velocity with which a pulse of energy travels is the group velocity, v_{gr} , and is related to v_p and c by the following equation:

$$v_p v_{gr} = c^2 \quad (12.3)$$

Thus, the group velocity is less than the velocity of light to the extent that the phase velocity is greater. Table 12.1 lists selected rectangular waveguides, their useful frequency range, their dimensions, and their designations. For an example of the use of Table 12.1, assume operation in X-band at a frequency of 10 GHz. The required waveguide would be number 6, which has a frequency range of 8.2–12.4 GHz. Outside dimensions for this waveguide are 25.4×12.7 mm, and the wall thickness is 1.3 mm. The inside dimensions can be determined by subtracting twice the wall thickness. The RETMA type number is WR90, and the JAN type number is RG-52/U.

Figure 12.2 is a physical picture of wave propagation in a rectangular waveguide. We can consider that the fields inside the waveguide are the result of a pair of electromagnetic waves that travel back and forth between the sides of the guide following a zigzag path, as illustrated in Figure 12.2(a). Each time such a wave strikes the conducting side wall, it is reflected with reversal of the electric field and with an angle of reflection equal to the angle of incidence.

The guide wavelength for the situation is the distance along the axis between the points in the guide where the positive crests coincide. In the top example in Figure 12.2(a), the free-space wavelength is much less than the cutoff wavelength, and the phase velocity in the waveguide is only slightly greater than the speed of light. The guide wavelength also is only slightly greater than the free-space wavelength. These

Table 12.1 Characteristics for Example Rectangular Waveguides

No.	Useful Frequency Range (GHz)	Outside Dimensions (mm)	Wall Thickness (mm)	RETMA Type No.	JAN Type No.
1	1.12–1.70	169 × 86.6	2.0	WR650	RG-69/U
2	1.70–2.60	113 × 58.7	2.0	WR430	RG-104/U
3	2.60–3.95	76.2 × 38.1	2.0	WR284	RG-48/U
4	3.95–5.85	50.8 × 25.4	1.6	WR187	RG-49/U
5	5.85–8.20	38.1 × 19.1	1.6	WR137	RG-50/U
6	8.20–12.40	25.4 × 12.7	1.3	WR90	RG-52/U
7	12.40–18.00	17.8 × 9.9	1.0	WR62	RG-91/U
8	18.00–26.50	12.7 × 6.4	1.0	WR42	RG-53/U
9	26.50–40.00	9.1 × 5.6	1.0	WR28	RG-96/U
10	40.00–60.00	6.8 × 4.4	1.0	WR19	—
11	60.00–90.00	5.1 × 3.6	1.0	WR12	RG-99/U
12	90.00–140.0	4.0 diam.	2.0 × 1.0	WR8	RG-138/U
13	140.0–220.0	4.0 diam.	1.3 × 0.64	WR5	RG-135/U
14	220.0–325.0	4.0 diam.	0.86 × 0.43	WR3	RG-139/U

Note: RETMA = Radio Electronic Television Manufacturers' Association; JAN = Joint Army-Navy.

Source: [2].

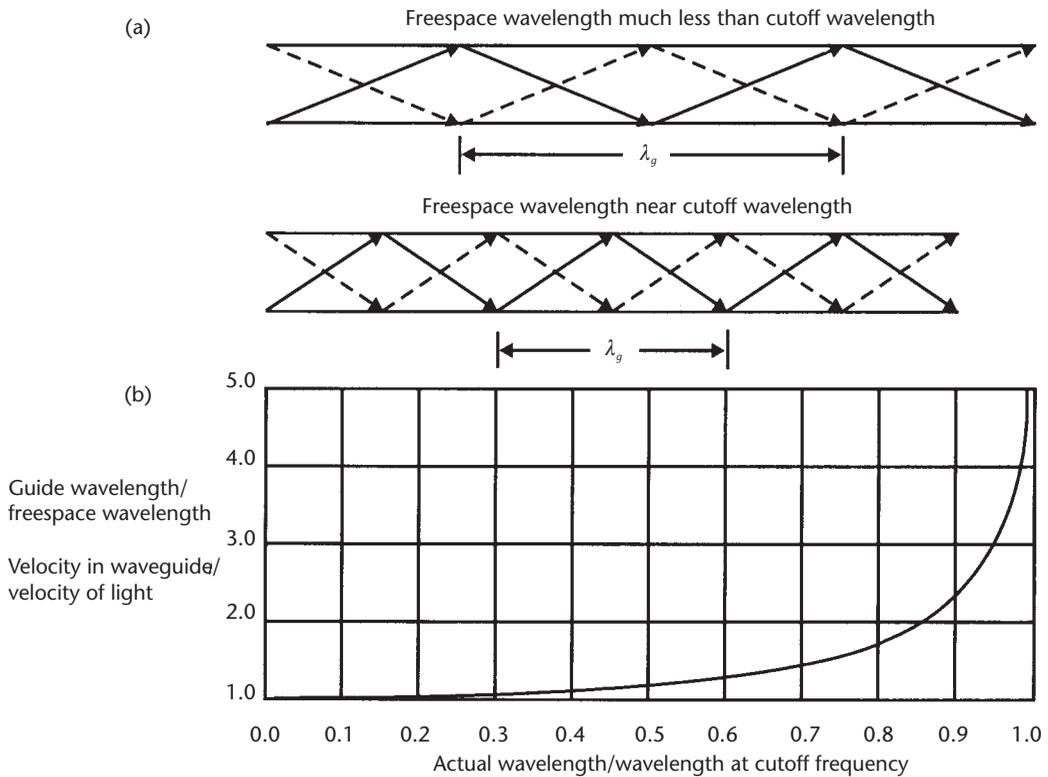


Figure 12.2 Guide wavelengths and velocities in waveguides: (a) paths followed by waves in waveguides and (b) plot of ratios of guide wavelength to free-space wavelength and velocity in wavelength to the speed of light. (After: [4].)

relationships are shown in Figure 12.2(b). In the bottom example in Figure 12.2(a), the free-space wavelength is near the cutoff wavelength, resulting in the guide wavelength and the phase velocity being much greater than the free-space wavelength and the speed of light. These relationships also are shown in Figure 12.2(b). When the free-space wavelength is equal to or greater than the cutoff wavelength, the wave does not propagate down the guide but simply moves back and forth between the sides of the guide.

The fields inside a waveguide induce currents that flow on the inner surfaces of the walls and that can be considered to be associated with the magnetic flux adjacent to the wall. The direction in which the current flows at any point in the wall is at right angles to the direction of the adjacent magnetic flux. In the sides of the guide, the current everywhere flows vertically for the case of the TE_{10} mode, since the magnetic flux in contact with the side walls lies in planes parallel to the top and bottom sides of the guide. In the top and bottom of the guide are a transverse component of current proportional to the axial component of the magnetic field and an axially flowing current component proportional at any point to the transverse magnetic field.

The current in the guide walls penetrates in accordance with the laws of skin effect (discussed in Chapter 14). Accordingly, the depth of penetration is inversely proportional to the square root of the frequency. At the very high frequencies at which waveguides are used, penetration is very small, and the walls provide practically perfect shielding.

A hole, joint, or slot in the waveguide wall introduces the possibility that energy will leak from the guide to the outside. When that happens, the fields inside the guide are affected, thereby introducing an irregularity with resulting reflection. The coupling thus introduced by a hole in the guide wall may be either to the electric or the magnetic field inside the guide. Electric coupling occurs when electrostatic flux lines that normally would terminate on the guide wall are able to pass through the hole to the outside or to another piece of the waveguide. Magnetic coupling results when the hole or the slot interferes with the current flowing in the guide wall. With either type of coupling, both electric and magnetic fields are present outside the main waveguide. Thus, electric flux leaking through the hole induces current on the outer surface of the guide that produces a magnetic field. Again, when magnetic flux leaks through a hole, the associated interference with the flow of current in the wall produces a voltage across the hole that gives rise to an electric field that extends beyond the guide primary.

The nature and the magnitude of the coupling in any particular case depend on the size, shape, and orientation of the coupling hole and on the thickness of the guide wall. The factors involved can be understood by considering the effects produced by long narrow slots oriented in various ways, as illustrated in Figure 12.3.

Slot 1, which is transverse to the magnetic field inside the guide, produces a minimum of interference with currents in the guide wall and introduces little or no magnetic coupling. It does, however, permit electric coupling if the slot width is great enough in proportion the wall thickness to permit a reasonable number of electric flux lines to pass through the slot. The electric coupling will be negligible if the slot is in the nature of a joint representing two surfaces fitted together or is very narrow. Similarly, the long, narrow slot 4 produces little magnetic coupling because it is transverse to the magnetic flux and therefore interferes only negligibly with the flow of current in the guide wall. It does not produce electric coupling because there is no electric field terminating on the side wall for the TE_{10} mode. Such a slot, therefore, has negligible effect, even if it is quite long.

In contrast, slot 5, while causing no electric coupling, introduces a substantial amount of magnetic coupling to outside space because its long dimension is parallel

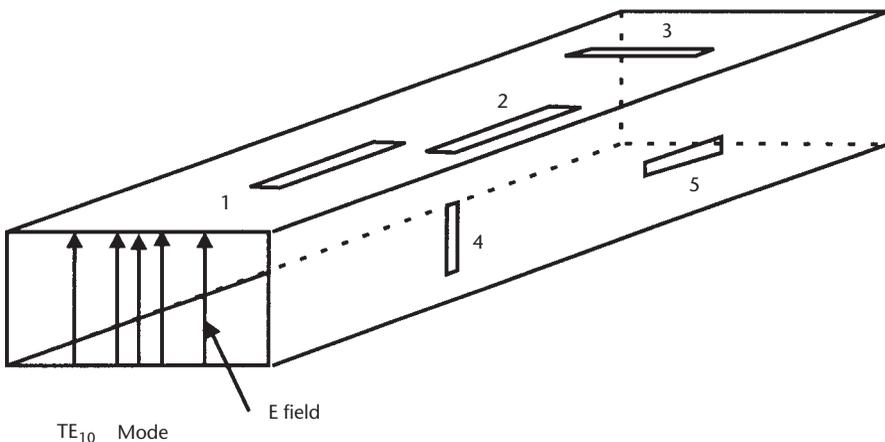


Figure 12.3 Waveguide with slots in walls and TE_{10} mode. (After: [4].)

to the magnetic field in the guide. Hence, this slot is oriented in a manner so as to permit easy escape of magnetic flux lines and to interfere to a maximum extent with the wall currents. This coupling is fully effective even if the slot is quite narrow; it is necessary only that the slot interrupt the flow of current in the wall.

Slots 2 and 3 in Figure 12.3 also give rise to magnetic coupling because they interfere with the flow of current in the guide wall. In the case of slot 2, the amount of coupling will be greater the farther the slot is to the side of the center line of the guide. Slots 2 and 3 will also simultaneously introduce electric coupling to the extent that the slot is wide enough in relation to the wall thickness to permit the passage of electric flux. In the case of slot 2, the electric coupling becomes less the farther the slot is from the center line because the intensity of the electric field terminating on the top and bottom sides of the guide becomes less as the side walls are approached.

Table 12.2 lists average attenuations for some rectangular waveguides operating in the dominant TE_{10} mode. The propagation of energy down a waveguide is accompanied by a certain amount of attenuation as a result of the energy dissipated by current induced in the walls of the guide. The magnitude of the current at any point is determined by the intensity of the magnetic field adjacent to the wall at that point. The resistivity that the induced currents encounter is determined by the skin depth of the wall and is therefore proportional to the square root of the frequency and to the square root of the resistivity of the material of which the wall is composed.

Skin depth and loss in conductors are discussed in Chapter 14.

Table 12.2 shows that in the frequency range of 1.12–1.70 GHz, the average attenuation is only 0.0052 dB/m. In contrast, in the 220.0–325.0 GHz frequency range, the average attenuation is 8.80 dB/m. Thus, we see the very strong effects

Table 12.2 Sample Average Attenuations and CW Power Ratings for Waveguides

<i>JAN Type</i>	<i>RETMA Type</i>	<i>Useful Frequency (GHz)</i>	<i>Average Attenuation (dB/m)</i>	<i>CW Power Rating (KW)</i>
RG-69/U	WR650	1.12–1.70	0.0052	14,600
RG-104/U	WR430	1.70–2.60	0.0097	6,400
RG-48/U	WR284	2.60–3.95	0.019	2,700
RG-49/U	WR187	3.95–5.85	0.036	1,700
aspnumRG-50/U	WR137	5.85–8.20	0.058	635
RG-52/U	WR90	8.20–12.40	0.110	245
RG-91/U	WR62	12.40–18.00	0.176	140
RG-53/U	WR42	18.00–26.50	0.37	51
RG-96/U	WR28	26.50–40.00	0.58	27
	WR19	40.00–60.00	0.95	13
RG-99/U	WR12	60.00–90.00	1.50	5.1
RG-138/U	WR8	90.00–140.00	2.60	2.2
RG-135/U	WR5	140.0–220.0	5.20	0.9
RG-139/U	WR3	220.0–325.0	8.80	0.4

Notes: RETMA = Radio Electronic Television Manufacturers' Association; JAN = Joint Army-Navy. The values shown are for copper waveguides except for the last five, which are for silver waveguides.

Source: [2].

that frequency and skin depth have on the attenuation in a waveguide. We also see the very large range of power ratings for waveguides as a function of frequency. An equation for maximum power in a waveguide is as follows:

$$P_{\max} \text{ for TE}_{10} = 3.6 \cdot a \cdot b \cdot \left(\lambda / \lambda_g \right) \text{ MW}$$

with a and b in inches. This holds for an air-filled waveguide at atmospheric pressure with the breakdown strength of air being 29 kV/cm.

It should be pointed out that the attenuation in the waveguide is not constant over the useful operating band. There is a particular frequency for which the attenuation is a minimum. On either side of that minimum, the attenuation increases, the result of two opposing tendencies. As the frequency is lowered, the skin depth becomes greater, causing the effective resistivity of the walls to decrease. At the same time, as the frequency approaches the cutoff value for the mode in question, the group velocity decreases. That causes the magnetic fields adjacent to the walls to become rapidly stronger for a given rate of energy flow down the guide. The operating range of the waveguide is thus chosen to be well above waveguide cutoff for the dominant or lowest order frequency mode and well below the next higher mode. Then either the average attenuation over this band is measured, or the theoretical average is predicted. From Table 12.2, we see that the useful bandwidth of a rectangular waveguide is much less than an octave. The ratios of the highest frequency to the lowest frequency in the band for the 14 waveguides shown in Table 12.2 are about 1.5.

A variation of the rectangular waveguide is a flexible waveguide. Flexible waveguides usually have about five times the attenuation of rigid waveguides due to the corrugations and so are used only for special cases where flexibility is needed.

12.3 Higher-Order Modes in Rectangular Waveguides

The dominant mode is only one of many field configurations that can exist in a waveguide. Configurations can be of either TE or TM types. Figure 12.4 shows field configurations in the transverse plane for the first four higher modes in a rectangular waveguide with sides $a = 2b$. Figure 12.4(a) shows the TE_{20} mode. The 2 in the subscript indicates that there are two half-cycles of the electric field configuration in the transverse plane in the direction of the long side of the rectangle. The 0 in the subscript indicates that there are zero half-cycle variations of the electric field in the direction of the short side of the rectangle. The cutoff wavelength for this mode is equal to a , where a is the dimension of the long side. Recall that for the dominant or TE_{10} mode the cutoff wavelength is $2a$. The TE_{11} mode is the first higher order mode that will be encountered.

Figure 12.4(b) shows the TE_{11} mode. The first 1 in the subscript indicates that there is only one half-cycle of the electric field configuration in the transverse plane in the direction of the long side of the rectangle. The second 1 in the subscript indicates that there is one half-cycle of the electric field in the direction of the short side of the rectangle. The cutoff wavelength for this mode is approximately equal to $0.89a$, where a is the long side dimension. Figure 12.4(c) shows the TM_{11} mode. Here

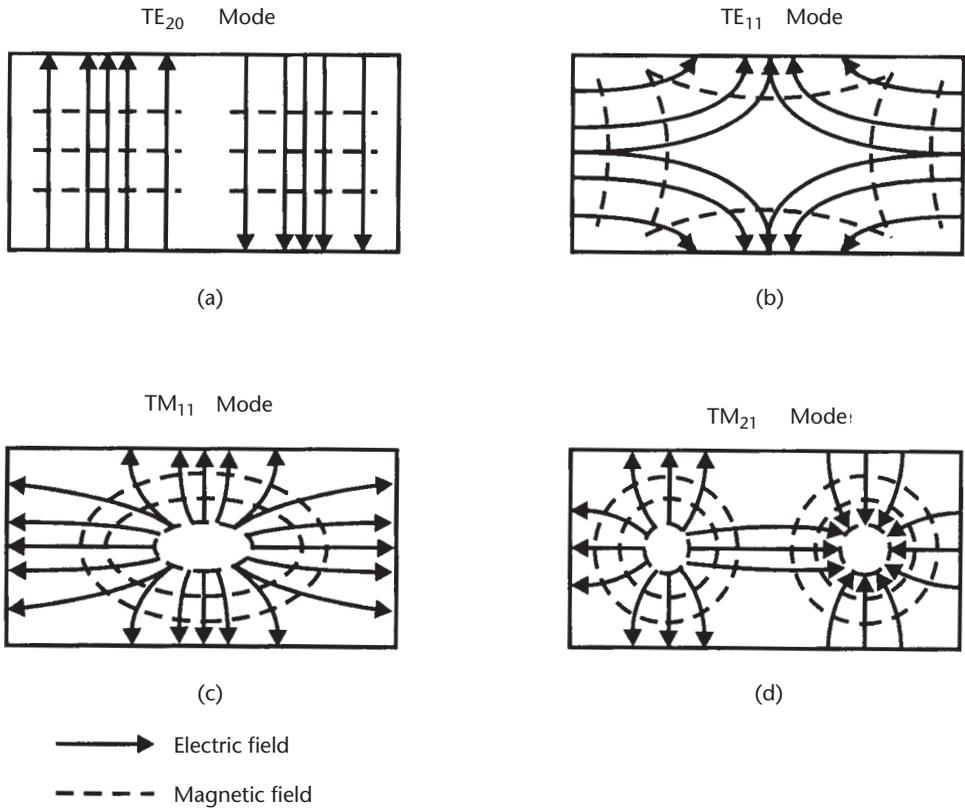


Figure 12.4 Field configurations in the transverse plane for the first four higher-order modes in a rectangular waveguide: (a) TE_{20} mode; (b) TE_{11} mode; (c) TM_{11} mode; and (d) TM_{21} mode. (After: [4].)

we see one half-period variation for the electric field in each of the directions. The cutoff wavelength for this case is the same as for the TE_{11} mode. Figure 12.4(d) shows the TM_{21} mode. Here we see two half-period variations for the electric field in the direction of the long side and one half-period variation for the electric field in the direction of the short side. The cutoff wavelength for this case is $0.71a$. The cutoff wavelength in the general case is given by the following relation:

$$\lambda_c = 2a / \left[(m^2) + (na/b)^2 \right]^{1/2} \tag{12.4}$$

Here a is the long dimension, b the short dimension for the rectangular guide, and m and n are the first and second subscripts describing the mode, respectively.

In the case of a square guide where $a = b$, the cutoff wavelengths are $2a$ for the TE_{10} and TE_{01} modes, $1.4a$ for the TE_{11} and TM_{11} modes, and a for the TE_{20} mode.

Any actual configuration of electric and magnetic fields existing in a waveguide can be regarded as the sum of a series of modes that are superimposed on one another. Modes in waveguides are thus analogous to the harmonics of a periodic wave, since a periodic wave of arbitrary shape always can be considered to be represented by the sum of a series of properly chosen harmonic components.

12.4 Launching the TE_{10} Mode Using a Coaxial Line Input

Figure 12.5 illustrates the launching of the TE_{10} mode using a coaxial line as the input with the center conductor extending into the waveguide. In the figure, the center conductor extends all the way to the top. A variation is to have the center line extension extend only part way into the inside of the guide. That type of arrangement is known as a voltage probe. A voltage probe acts as a small antenna to launch the wave.

Current in the center conductor generates a magnetic field in the guide that lies in a plane parallel to the top and bottom sides of the guide. At the same time, electric field lines are produced as shown. The TE_{10} mode is the largest single component in the field configuration. The difference between the field configuration of this mode and the actual field present is accounted for by the presence of higher order modes of smaller amplitude. Impedance matching is used to provide maximum power transfer from the coaxial line to the waveguide. (Characteristic impedance and impedance matching for a waveguide are discussed later.)

Figure 12.6 shows coupling from a coaxial line to a waveguide using a small loop. In the figure, the loop couples magnetic field into the waveguide.

One method of suppressing unwanted waveguide modes is to use metal vanes located on the short sides of the guide. Because of the field configuration, the vanes do not affect the TE_{10} mode, but they do interfere with both the electric and the magnetic fields of any TM or TE_{0n} modes that might be present. It is also important to choose waveguide sizes such that higher order modes see waveguide cutoff. Modes that are beyond cutoff and so cannot propagate are sometimes termed evanescent modes.

12.5 Characteristic Wave Impedance for Waveguides

In a TEM mode transmission line, one can define a characteristic impedance that is determined by the geometry of the line and that holds for all frequencies. In a similar manner, the waveguide has a characteristic impedance. However, that impedance is

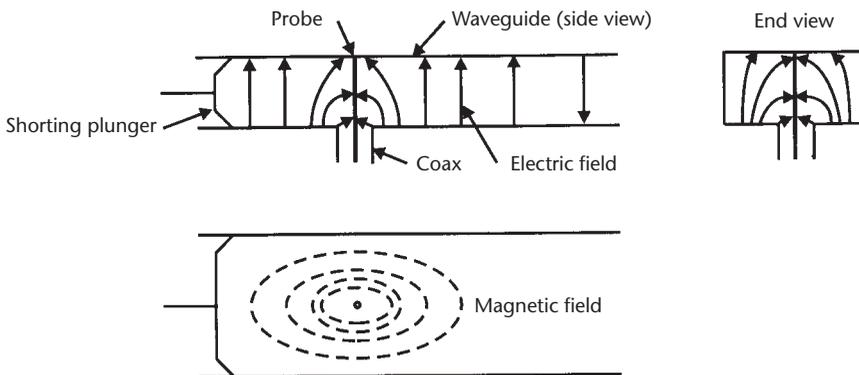


Figure 12.5 Launching of TE_{10} mode using a coaxial line and a voltage probe. (After: [4].)

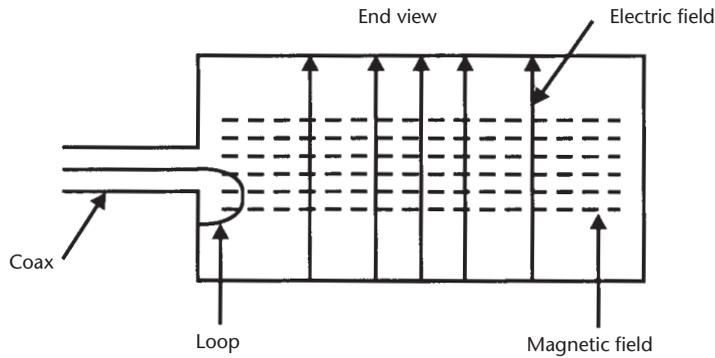


Figure 12.6 Coupling from a coaxial line to a waveguide using a small loop.

a function of frequency. For the TE_{m0} modes, the characteristic wave impedance for the rectangular waveguide is

$$Z_0 = 377 / \left[1 - (\lambda / \lambda_c)^2 \right]^{1/2} \quad (12.5)$$

where:

λ = free-space wavelength

λ_c = cutoff wavelength for the guide

For example, assume the cutoff wavelength for a waveguide is 10 cm and the free-space wavelength of the signal is 8 cm. Substituting values into (12.5), the characteristic wave impedance for the TE_{m0} modes would be

$$Z_0 = 377 / \left[1 - (8/10)^2 \right]^{1/2} = 6283\Omega$$

For the case of the TM_{mn} modes, the characteristic wave impedance for the rectangular waveguide is

$$Z_0 = 377 \left[1 - (\lambda / \lambda_c)^2 \right]^{1/2} \quad (12.6)$$

where the terms are the same as defined for (12.5). For example, assume the cutoff wavelength for a waveguide is 10 cm and the free-space wavelength of the signal is 8 cm. Substituting values into (12.6), the characteristic wave impedance for the TM_{mn} modes would be

$$Z_0 = 377 \left[1 - (8/10)^2 \right]^{1/2} = 2262\Omega$$

As in the case of TEM transmission lines, if the load impedance matches the characteristic impedance of the waveguide, there is no reflection at the load.

12.6 Other Types of Waveguides

12.6.1 Ridged Waveguides

Figure 12.7 shows the case of ridged waveguides. Rectangular waveguides are sometimes made with single ridges located inside the guide on either the top or the bottom walls, as shown in Figure 12.7(a), or they are made with double ridges located inside the guide on both the top and the bottom of the guide walls, as shown in Figure 12.7(b). The ridges are located midway between the two sides. The principal effects of such ridges are to lower the value of the cutoff frequency and to increase the useful bandwidth of the waveguide. By those means, it is possible to achieve nearly an octave bandwidth capability. It should be noted, however, that ridged waveguides generally have more attenuation per unit length than rectangular waveguides without ridges.

12.6.2 Circular Waveguides

Circular waveguides have the advantage that they can be used in application where rotation about the axis of the guide is required, as in the case of a rotating antenna with a waveguide feed. They have the disadvantage that there is only a very narrow range between the cutoff wavelength of the dominant mode and the cutoff wavelength of the next higher mode. Thus, the frequency range over which single-mode operation is assured is relatively limited. Also, because of the circular symmetry, the circular waveguide possesses no characteristic that positively prevents the plane of polarization of the wave from rotating about the guide axis as the wave travels. As a result, circular waveguides are used only where it is necessary to have a rotating joint in a waveguide system.

Field configurations for the two most important circular modes are illustrated in Figure 12.8: the TM_{01} mode in Figure 12.8(a) and the TE_{11} mode in Figure 12.8(b).

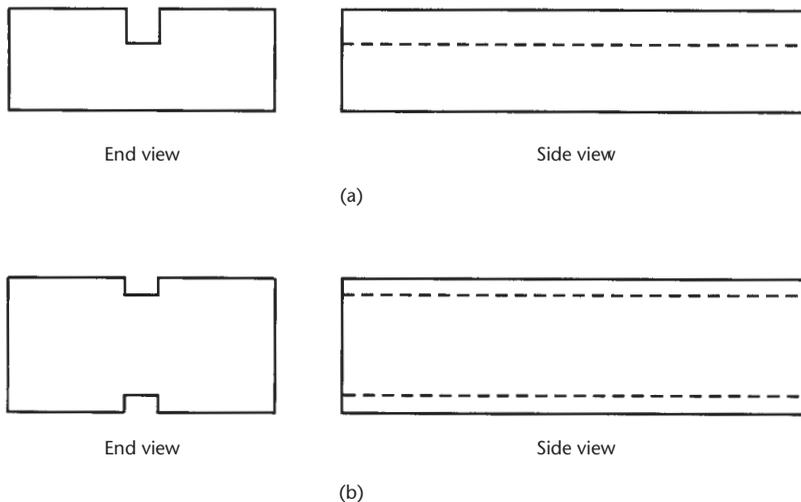


Figure 12.7 Ridged waveguide: (a) ridged waveguide with single ridge and (b) ridged waveguide with double ridge. (After: [2].)

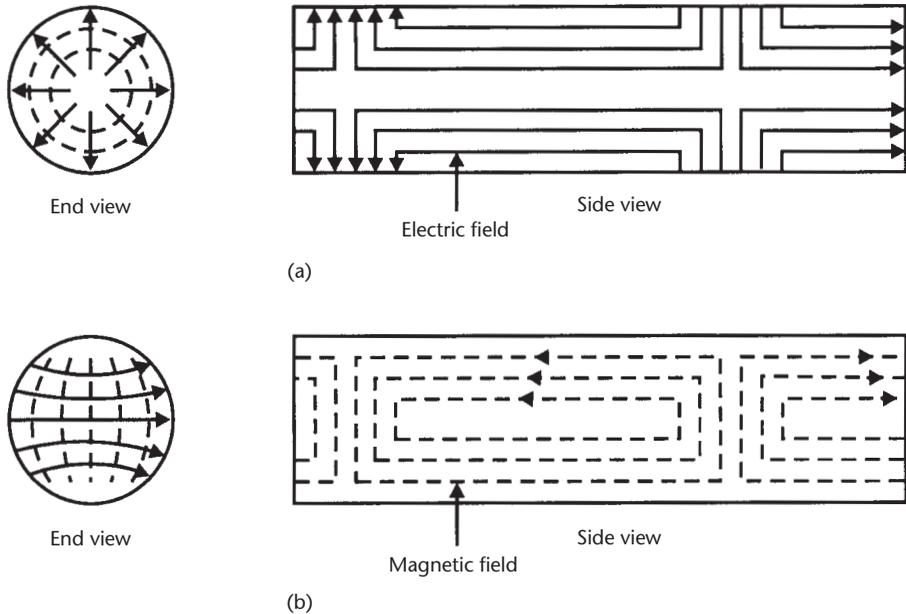


Figure 12.8 Circular waveguide: (a) circular waveguide with TM_{01} mode and (b) circular waveguide with TE_{11} mode. (After: [2].)

The TE_{11} mode is the dominant mode, and the TM_{01} mode is the first higher-order mode.

Table 12.3 lists the cutoff wavelengths in terms of the guide radius, r , for the first five circular waveguide modes.

The guide wavelength in a circular guide is greater than the wavelength in free space, just as in the rectangular guide. The velocity of phase propagation is λ_g/λ times the velocity of light in all cases. A wave traveling down a circular guide is attenuated as a result of power dissipated in the walls by the induced wall currents, exactly as in the case of a rectangular guide.

12.7 Waveguide Hardware

12.7.1 Waveguide Flanges

A typical waveguide has a flange at either or both ends. At the lower frequencies, the flanges are brazed or soldered onto the waveguide. At higher frequencies, a much

Table 12.3 Cutoff Wavelengths in Circular Waveguides

<i>Mode</i>	<i>Cutoff Wavelength</i>
TE_{11}	$3.42 r$
TM_{01}	$2.61 r$
TE_{21}	$2.06 r$
TE_{01}	$1.64 r$
TM_{11}	$1.64 r$

flatter, butted, plain flange is used. When two waveguides are connected, the two flanges are bolted together. Care must be taken to ensure that near-perfect mechanical alignment is achieved, so there are no undesirable reflections of the signal at those junctions. Waveguides with smaller dimensions sometimes are provided with threaded flanges, which are somewhat easier to align.

A second type of waveguide flange combination uses a plain flange on one side and a choke flange on the other end. With the choke flange, a short circuit is reflected to the junction of the waveguides using a half guide wavelength slot. Thus, an electrical short is placed at a surface where a mechanical short circuit would be difficult to achieve. Unlike the plain flange, the choke flange is frequency sensitive. A typical design for such a choke flange might provide a 10% bandwidth.

12.7.2 Rotary Joints

Rotating joints are often used in radar. An example would be where a radar is connected to a horn antenna feeding a parabolic reflector that must rotate for tracking. A rotating joint involving a circular waveguide is the most common type. The rotating part of the waveguide is circular and carries the TM_{01} mode, whereas the rectangular waveguide leading in and out of the joint carries the TE_{10} mode. The circular waveguide has a diameter that ensures that modes higher than the TM_{01} cannot propagate. The dominant TE_{11} in the circular guide is suppressed by a ring filter, which tends to short-circuit the electric field for that mode, while not affecting the electric field of the TM_{01} mode, which is everywhere perpendicular to the ring. A choke gap is left around the circular guide joint to reduce any mismatch that may occur. Impedance matching of some type is often provided at each circular-rectangular waveguide junction to compensate for reflections.

12.7.3 Tapered Transition Sections of Waveguides

A tapered section of waveguide is used when it is necessary to join waveguides having different dimensions or different cross-sectional shapes. Some reflections take place from a tapered section, but they can be reduced to low levels if the tapered section is made gradual, so the section has a length of two or more wavelengths at the lowest frequency of interest.

A tapered section of waveguide transforms the characteristic impedance of the waveguide. The impedance is directly proportional to the short dimension of the rectangular waveguide. The taper transition section thus can be used as an impedance-matching section for connecting to resistive loads.

12.7.4 Flexible Waveguides

Flexible waveguides are sometimes used when it is necessary to have a waveguide section capable of movement, which may include bending, twisting, stretching, and operation with vibration. The flexible waveguide must not cause undue attenuation or reflections and must be able to operate continuously for extended periods of time. Several different types of flexible waveguide are used, including copper or aluminum tubes having an elliptical cross-section, small transverse corrugations, and

transitions to rectangular waveguides at the two ends. These tubes transform the TE_{11} mode in the flexible waveguide into the TE_{10} mode at either end. Such a waveguide is of continuous construction; thus, joints and separate bends are not required. It may have a polyethylene or rubber outer cover for environmental protection; it also bends easily but cannot be twisted readily. Power-handling ability and SWR are fairly similar to those of rectangular waveguides of the same size, but attenuation in decibels per meter is about five times greater than in equivalent rectangular waveguides.

12.7.5 Waveguide Accessories

Manufacturers' catalogs show a large number of accessories that can be used in waveguide systems: 90-degree bends of several types, circular to rectangular waveguide transitions, 90-degree twist sections, H-plane T-junctions, and E-plane T-junctions. T-junctions (particularly the E-plane variety) often are used for impedance matching in a manner identical to the short-circuited TEM-transmission line stub. The vertical arm is provided with a sliding piston or plunger to produce a short circuit at any desired point. Other accessories include terminations, filters, couplers, power dividers, hangers, feedthroughs, pressure windows, gas inlets, straight sections, and flex-twist sections.

12.8 Waveguide Hybrid Junctions

Figure 12.9 shows a waveguide hybrid T-junction, also known as a magic T. A hybrid T-junction has some very interesting and highly useful properties. Its basic property is that arms 3 and 4 both are connected to arms 1 and 2 but are isolated from each other, provided that each arm is terminated in a correct impedance. If a signal is applied to arm 3, it divides equally between arms 1 and 2 but with opposite phase for the two arms. Ideally, none of the signal enters arm 4. An input signal at arm 4 likewise divides equally between arms 1 and 2, but with the same phase for

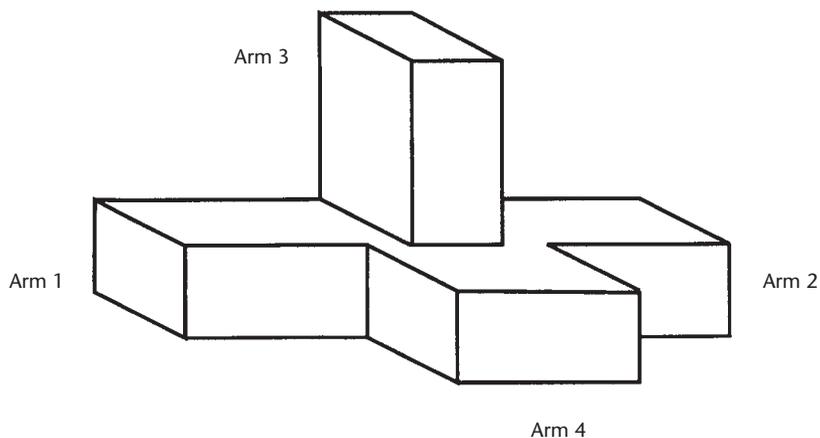


Figure 12.9 Waveguide hybrid T-junction (magic T). (After: [2].)

each arm. Again, ideally there is no signal coupled to arm 3. The operation for this coupler is the same as that of the stripline or microstrip rat-race coupler.

If two signals enter arms 1 and 2, they add in arm 4. On the other hand, the output in arm 3 for the two signals is the difference between the signals.

An example application of the magic T is as a front end for a microwave receiver. In that case, arm 3 is connected to the antenna, arm 4 is connected to the LO, arm 1 is terminated in a matched impedance, and arm 2 is connected to a mixer.

A second type of hybrid junction looks very different from the magic T and yet has very similar properties. This waveguide system, known as a hybrid ring or rat race and shown in Figure 12.10, consists of a rectangular waveguide bent in the E plane to form a complete loop whose median circumference is 1.5 guide wavelengths. It has four arms connected as shown with arms 1 and 4 separated by three-quarters guide wavelengths and the other arm spacings one-quarter guide wavelengths. If a signal is applied to arm 1, it divides evenly, with half traveling clockwise and the other half traveling counterclockwise. The two signals reaching arm 4 have traveled the same distance and add in phase. A part of the signal thus travels out through arm 4. A signal starting from arm 1 and reaching arm 2 has traveled a distance of $\lambda_g/4$ if traveling counterclockwise and $1.25 \lambda_g$ if traveling clockwise. Recall that phase repeats every wavelength. Thus, the two signals are in phase at arm 2 and add at that point. A part of the signal thus travels out arm 2. A signal starting at arm 1 and reaching arm 3 has traveled a half guide wavelength in the counterclockwise direction and a full guide wavelength in the clockwise direction. These two signals are 180 degrees out of phase and cancel. In a similar way, it can be shown that arm 3 is connected to arms 2 and 4 but not to arm 1.

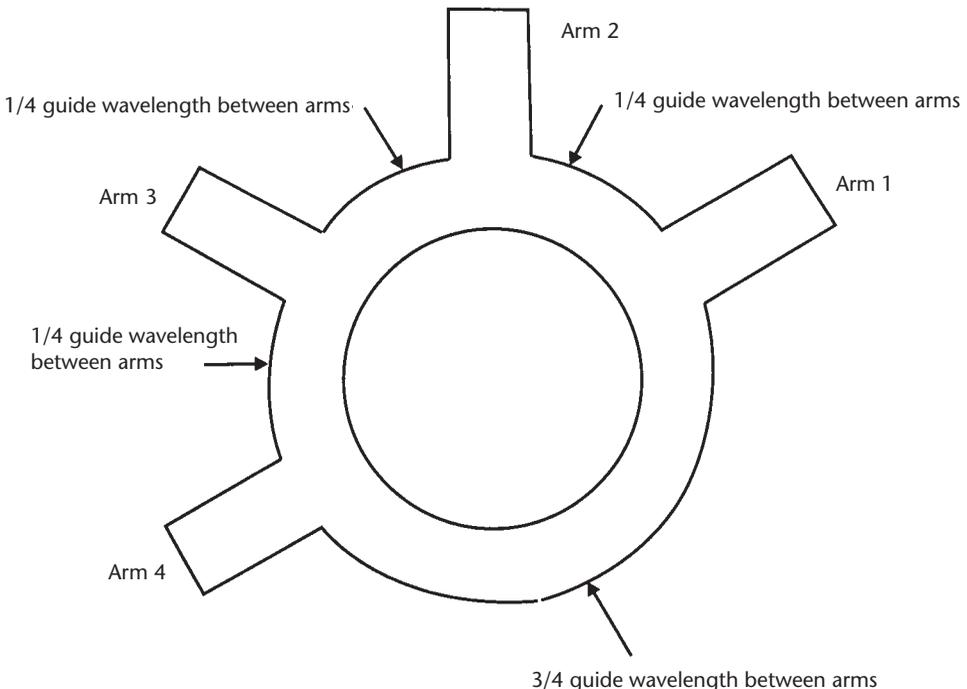


Figure 12.10 Waveguide hybrid ring or rat race. (After: [2].)

If one signal enters arm 2 and a second signal enters arm 4, they add at arm 3, since they each have traveled the same distance. On the other hand, on reaching arm 1, the signals subtract since they have traveled distances separated by one-half guide wavelength.

Thus, the behavior of the rat-race coupler is very similar to that of the magic T, although for different reasons. The two types of systems can be used interchangeably, with the magic T having the advantages of smaller bulk, ease of manufacture, and better isolation over waveguide band. The hybrid ring has dimensions that are frequency dependent, hence performance also is frequency dependent. That is not so in the case of the magic T.

12.9 Waveguide Impedance Matching

Impedance matching can be accomplished in waveguides by various types of obstacles placed in the waveguide. By those means, the equivalent of parallel capacitors, parallel inductors, or parallel resonant LC circuits can be added. They can be used in the same way that shorted or open-circuited stubs are used in TEM-mode transmission lines to provide impedance matching.

Three types of obstacles are illustrated in Figure 12.11. Figure 12.11(a) shows two types of waveguide irises or apertures with the openings parallel to the long side of the guide. This type of obstacle is used for the equivalent of parallel capacitance. Figure 12.11(b) shows two types of waveguide irises or apertures with the openings perpendicular to the long side of the guide. This type of obstacle is used for the equivalent of parallel inductance.

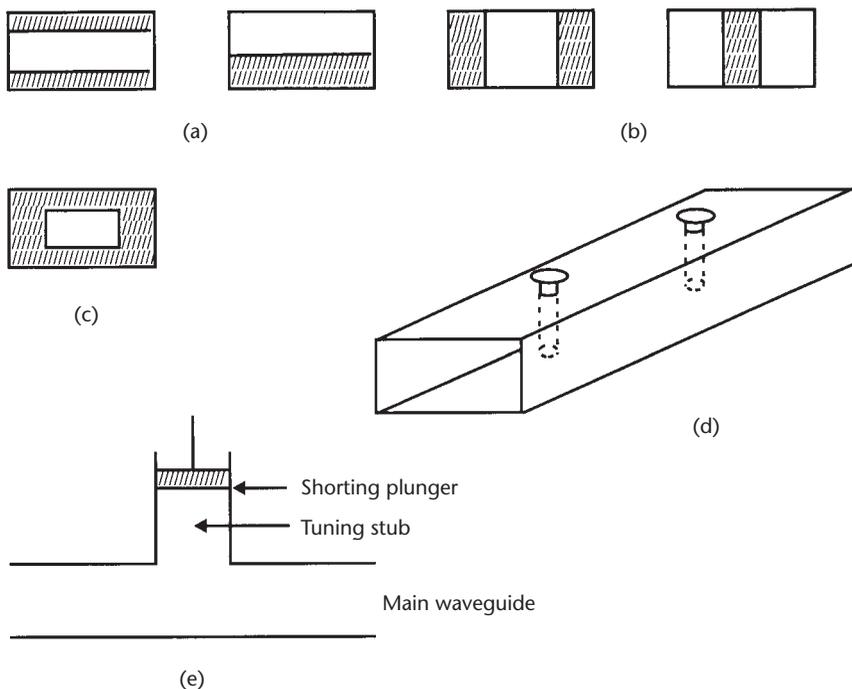


Figure 12.11 Waveguide impedance matching components: (a) capacitive waveguide irises; (b) inductive waveguide irises; (c) parallel resonant waveguide iris; (d) waveguide with two-screw tuner; and (e) stub tuner.

parallel to the short side of the guide. This type of obstacle is used for the equivalent of parallel inductance. Figure 12.11(c) shows a waveguide iris or aperture with a rectangular opening in the center of the guide with the long side of the opening parallel to the long side of the guide. This type of obstacle is used for the equivalent of a parallel resonant LC circuit. Irises are not easily adjustable and therefore normally are used to correct only permanent mismatches.

Another type of obstacle that can be used with waveguides for impedance matching is a cylindrical post extending into the waveguide from one of the broad sides. It can appear as either inductive or capacitive, depending on how far it extends into the waveguide. A short post appears capacitive, while a long post appears inductive.

Another type of obstacle that can be used when it is necessary to have adjustable matching elements is a screw extending into the waveguide from one of the broad sides of the guide. Again, the parallel impedance introduced by the screw depends on how far the screw extends into the guide.

It is common practice to use more than one screw for impedance-matching sections. These can be two-screw tuners or three-screw tuners. The spacing between screws typically is three-eighths guide wavelength at the center frequency in the waveguide band for the two-screw unit. A system of this type is shown in Figure 12.11(d).

The E-plane T also can be used in the same way as the adjustable TEM-mode transmission line shorted stub when it is provided with a sliding short-circuiting piston. A system of this type is shown in Figure 12.11(e).

12.10 Waveguide Resistive Loads and Attenuators

A common resistive termination for a waveguide is a length of epoxy-iron mixture fitted in at the end of the guide and tapered gradually with the point in the direction of the incoming wave. Such a termination absorbs the incoming waves and therefore does not cause reflections. Terminations of this type are not frequency sensitive if the length of the tapered section is greater than one-half the lowest frequency's guide wavelength.

A movable resistive vane can be used as a variable attenuator for a waveguide. In one type, the vane extends through a slot in the top of the waveguide. In another type, the vane is mounted on dielectric rods that extend through the sides of the guide.

12.11 Waveguide Directional Couplers

A waveguide directional coupler is illustrated in Figure 12.12. Waveguide directional couplers can be made by using a second waveguide parallel to the first with two holes providing coupling between them. The holes are spaced one-quarter guide wavelength apart. The operation and the use of waveguide directional couplers are similar to the operation and use of the TEM mode transmission line directional coupler. The couplers are used to measure simultaneously the forward and reflected

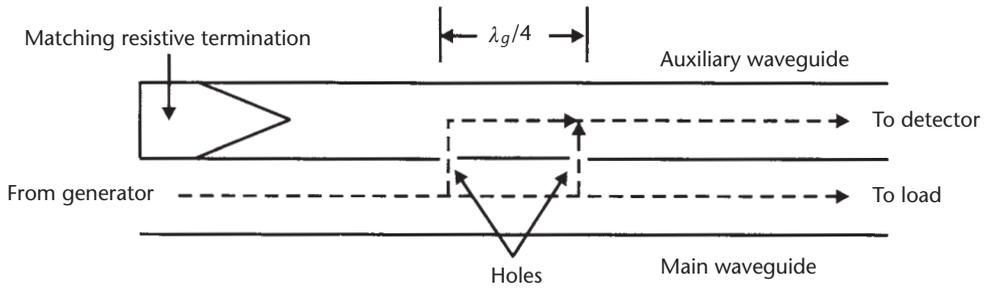


Figure 12.12 Two-hole waveguide directional coupler. (After: [2].)

power in the main waveguide. The amount of coupling depends on the size of the holes. A typical coupler has a coupling ratio of 20–40 dB and provides very low insertion loss.

12.12 Waveguide Ferrite Isolators, Circulators, and Switches

In many cases at microwave frequencies, it is desirable to have only one-way transmission of signals. An example is a microwave generator where the amplitude and frequency could be affected by changes in the load impedance, a phenomenon that is termed frequency pulling when applied to frequency variation. The solution to that problem is to use an isolator between the generator and the load. A second example is a semiconductor device used for microwave amplification that is a two-terminal device, in which the input and the output would interfere unless some means of isolation are used. In that case, a circulator usually is used. Ferrites often are used in such devices with magnetic fields involved.

Ferrite isolators may be based on Faraday rotation, in which the direction of polarization is angularly rotated by the ferrite device. A waveguide isolator of this type is illustrated in Figure 12.13.

The isolator in Figure 12.13 uses a center section of a circular waveguide operating in the TE_{11} mode, with transition sections connected to standard rectangular waveguides operating in the TE_{10} mode. One of the transition sections is rotated 45 degrees

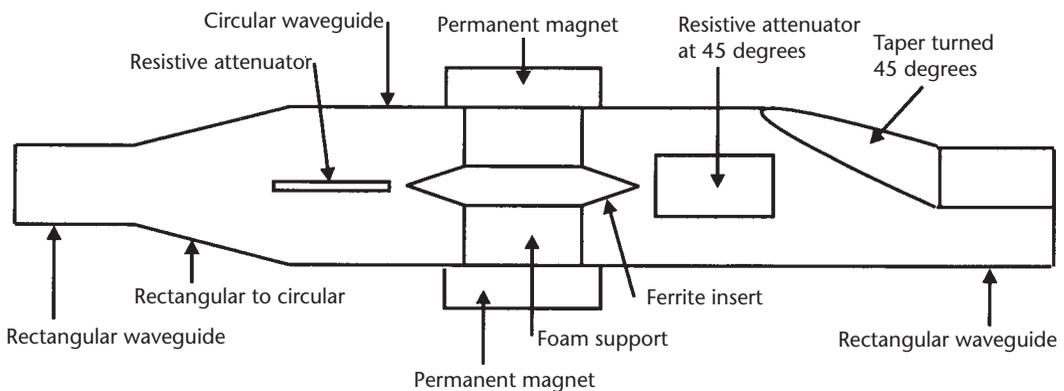


Figure 12.13 Faraday rotation isolator. (After: [2].)

degrees with respect to the other. A small-diameter ferrite rod is mounted by means of a foam support in the center of the circular waveguide. A permanent magnet is placed about the circular waveguide and the ferrite rod. Two other important components placed in the circular waveguide are flat resistive attenuators, with one rotated 45 degrees with respect to the other, as shown.

The operation of a Faraday rotation isolator is as follows. A wave passing through the ferrite in the forward direction (to the right) has its plane of polarization rotated clockwise 45 degrees. It passes the resistive attenuator with low loss and exits the transition section with the desired direction of E-field. A typical total insertion loss in the forward direction for an X-band isolator of this type is in the range of 0.5–1.0 dB. A wave passing through the isolator in the reverse direction also is rotated in polarization in the clockwise direction. As a result, the E-field direction is such that the wave is largely absorbed by the resistive vane. It also cannot propagate in the input rectangular waveguide because of its direction of E-field and the dimensions of the waveguide at right angles to the E-field (the wave is in waveguide cut-off). A typical loss for the reverse is 20–30 dB.

This type of isolator finds many applications for peak power levels less than about 2 kW. The power limitation is the result of nonlinearities in the ferrite, resulting in polarization shifts departing from the ideal 45 degrees.

A second popular type of high-power waveguide isolator is the resonant absorption isolator, illustrated in Figure 12.14. It uses a section of a rectangular waveguide operating in the TE_{10} mode. A piece of ferrite material is placed about a quarter of the way from one side of the waveguide and halfway between its ends. A permanent magnet is used with the ferrite, as shown, with a much stronger field than in the Faraday rotation isolator. At the location of the ferrite, the magnetic field of the TE_{10} wave is strong and circularly polarized. The polarization is clockwise in one direction of propagation and counterclockwise in the other. Therefore, there is unaffected propagation in one direction but resonance and large absorption in the opposite direction. Maximum power capability for this type of isolator is limited

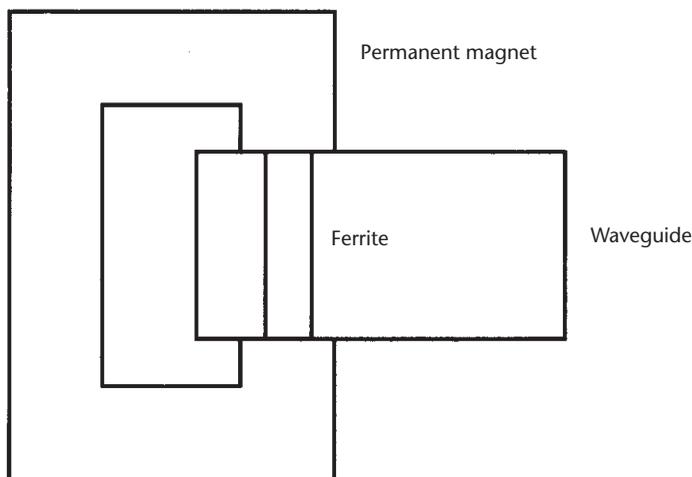


Figure 12.14 Resonance absorption isolator (end view). (After: [2].)

only by temperature rise that might bring the ferrite to its Curie point. Such systems can be built with peak power capabilities as high as 3,000 kW at S-band, 1,000 kW at C-band, and 300 kW at X-band.

There are waveguide, coaxial, and stripline versions of Y-junction ferrite circulators. Stripline versions were discussed in Chapter 11. Three-port, high-power, waveguide junction circulators and isolators are available in all standard waveguide sizes for WR28 through WR284. All circulators can be converted to isolators with the addition of a load. Peak power-handling capability depends on pressurization, duty cycle, pulse width, load mismatch, and altitude. Average power-handling capability can be increased by heat sinking, fins, or liquid cooling.

A waveguide can be placed at the junction of two or more intersecting waveguides to direct the flow of microwave energy. They can be used in any system where switchable isolation or duplexing is required. An SPDT switch is a junction circulator whose direction of circulation can be reversed on command by reversing the magnetic bias. That is achieved by replacing the permanent magnets found on a circulator with an electromagnetic source, such as an external electromagnet or an internal latch wire.

12.13 Waveguide Detectors and Mixers

A diode waveguide mount and detector are illustrated in Figure 12.15. Silicon-point contact diodes have been used for many years for microwave mixer and detector functions. More recently, Schottky-barrier diodes have been used for mixer

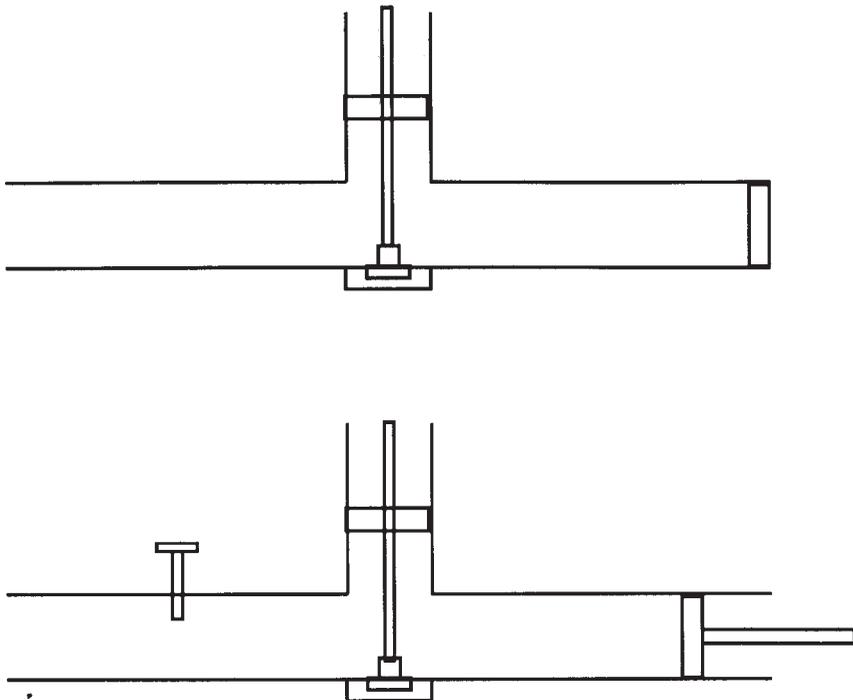


Figure 12.15 Diode waveguide mount and detector. (After: [2].)

functions because of their lower noise figure (below 6 dB at 10 GHz). In either case, it is necessary to use some form of diode mount to locate the diodes in the waveguide.

The detector in Figure 12.15 uses tuning for the waveguide, including a tuning plunger for providing the quarter guide wavelength location for the diode and a tuning screw. Figure 12.16 shows two types of waveguide mixers. If there is to be mixing in the waveguide, both the LO signal and the RF signal must be injected. One way to accomplish that is illustrated in Figure 12.16(a). The system includes a tuning screw in the LO line and a tuning plunger in the RF path. The diode is placed a quarter guide wavelength from the shorting plunger.

A balanced mixer configuration is illustrated in Figure 12.16(b). Here a magic T or hybrid T is used with two mixer diodes. This type of system is superior in performance to the single-ended mixer.

12.14 Gas-Tube Switches

Gas-tube switches are used in a number of different types of duplexers. These are sometimes called transmit/receive (TR) switches or antitransmit/receive (ATR) switches. The gas-tube switch is basically a piece of waveguide with glass windows on each end. The tubes are filled with a gas mixture, such as hydrogen, argon, water vapor, and ammonia, at low pressure. A pair of electrodes extending from the top and the bottom help to ionize the gas when high power is present.

At low power, the gas tube acts as an ordinary piece of waveguide, and the signal passes through with low insertion loss. When a high-power pulse arrives, the gas is ionized and becomes a poor conductor. The result is like placing a short circuit across the waveguide. Typical attenuations produced are greater than 60 dB. The switching action takes place rapidly (on the order of 10 ns). Quick deionization also is provided after the high-power pulse is ended.

12.15 Duplexers

A duplexer is a circuit that is designed to allow the use of a common antenna for both transmit and receive functions. Figure 12.17 shows the case of a branch-guide duplexer for radar using TR and ATR switches. For simplicity of illustration, the waveguides are shown as two-wire lines. When the radar transmitter pulse is present, it travels directly to the antenna. The TR and ATR switches also fire, and short circuits are produced. These are transformed by the quarter guide wavelength lines to open circuits at the main line. When no transmit pulse is present, the ATR is an open circuit that is transformed to a short circuit at the transmitter port. It is further transformed by the quarter guide wavelength main line to an open circuit at the antenna port. The TR switch acts as a normal waveguide, passing the receive signal from the antenna to the receiver.

The branch-line duplexer is a relatively narrowband device because it relies on the length of the waveguides that connect the switches to the main waveguide. This type of duplexer generally has been replaced by the balanced duplexer in modern radars.

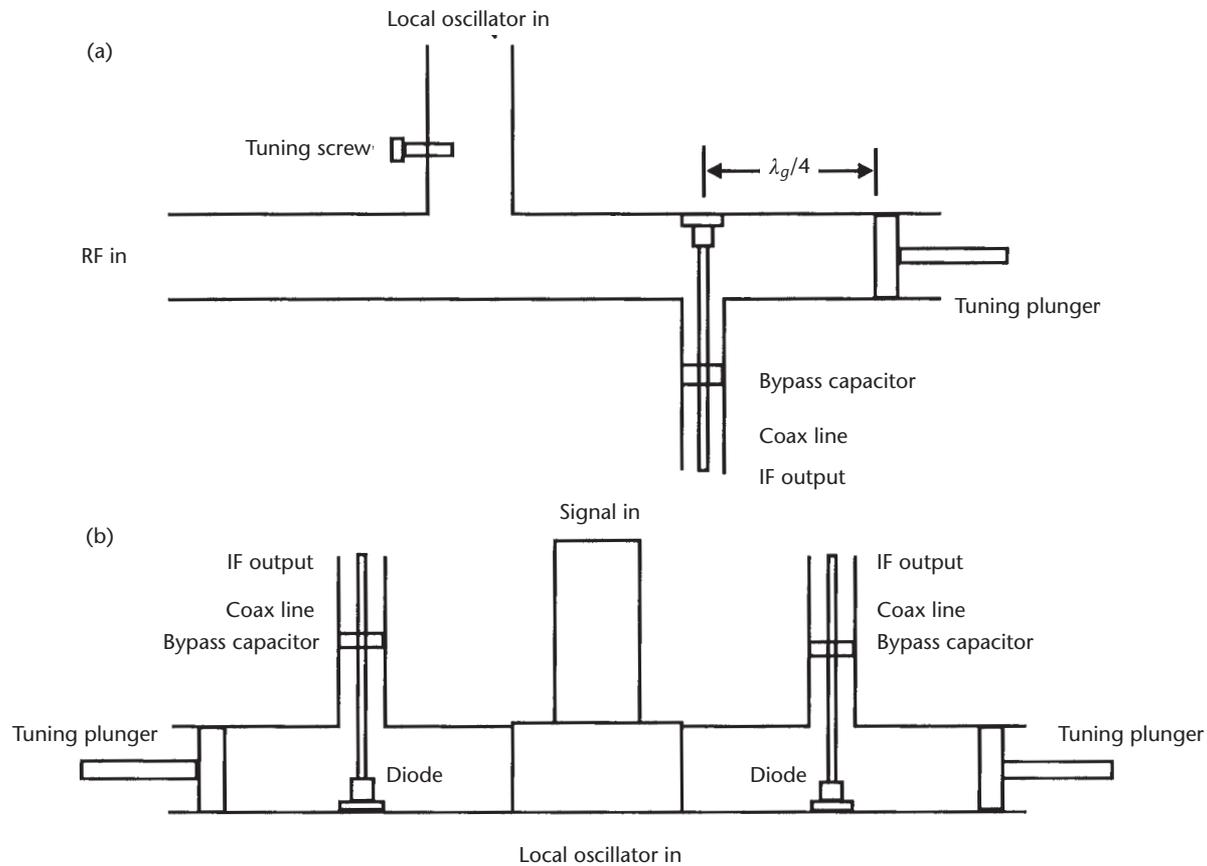


Figure 12.16 Two types of waveguide mixers: (a) single-ended waveguide diode mixer and (b) hybrid-T (magic-T) balanced waveguide mixer. (After: [2].)

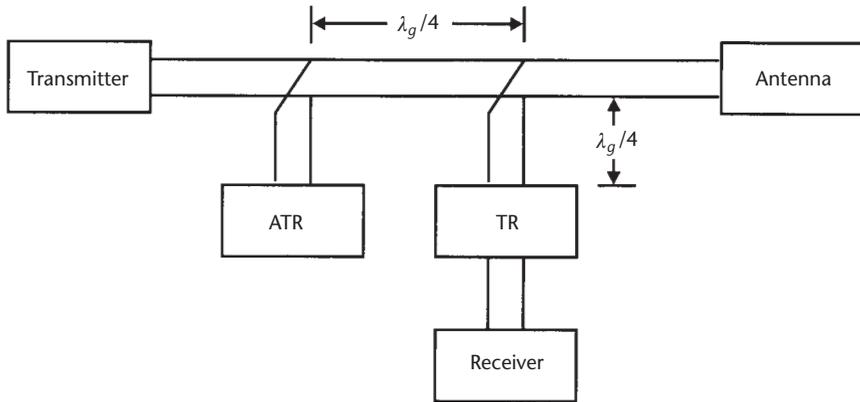
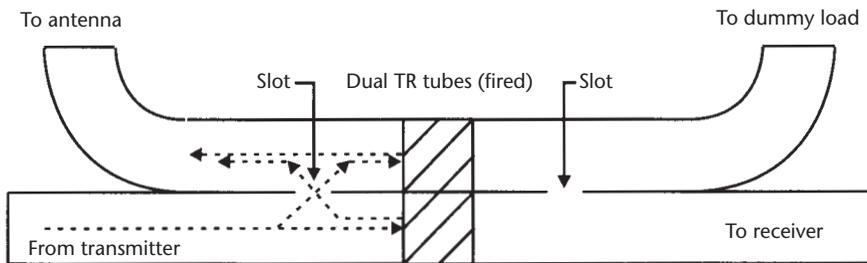


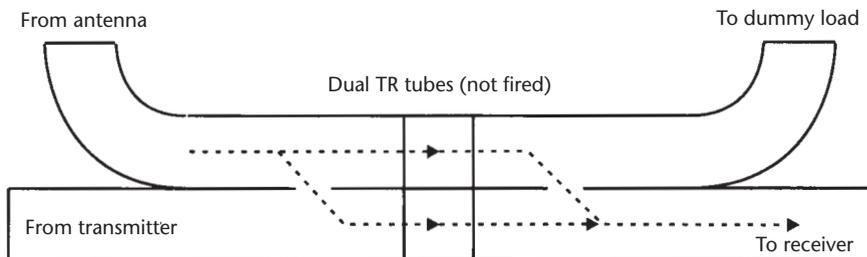
Figure 12.17 Branch-line type duplexer. (After: [3].)

Figure 12.18 shows a balanced duplexer using dual TR tubes and two short-slot hybrid junctions. The transmit condition of operation is shown in Figure 12.18(a). The TR tubes are fired, producing a near short circuit at their interface to the waveguides.

The short-slot hybrid junction consists of two sections of waveguides joined along one of their narrow walls with a specially designed pair of slots cut in the common narrow wall to provide coupling between the two guides. The short-slot hybrid can be considered a broadband directional coupler with a 3-dB coupling ratio.



(a)



(b)

Figure 12.18 Balanced duplexer using dual TR tubes. (After: [3].)

In the transmit condition, power is divided equally into each waveguide by the first short-slot hybrid junction. Both TR tubes break down and equally reflect the incident power out the antenna arm. The short-slot hybrid has the property that, each time the energy passes through the slot in either direction, its phase is advanced 90 degrees. Any energy that leaks through the TR tubes is directed to the arm with the matched dummy load and not to the receiver.

Figure 12.18(b) shows the operation of the balance duplexer for the receive condition. The TR tubes are unfired, and the return signals pass through the duplexer and into the receiver. The power splits equally at the first junction, and because of the 90-degree phase advance on passing through the slot, the energy recombines in the receiving arm and not in the dummy-load line.

Other types of balance duplexers are sometimes used. One of these is a four ATR system. During transmission, the ATR tubes located in a mount between the two short-slot hybrids ionize and allow high power to pass to the antenna. During reception, the ATR tubes present a high impedance, which results in the return signal power being reflected to the receiver. This type of system has higher power-handling capability than the system in Figure 12.18, but it has less bandwidth.

12.16 Cavity Resonators

Cavity resonators are used at microwave frequencies for much the same purposes as tuned LC circuits are used at the lower frequencies. Cavity resonators are characterized as having very high values of Q and therefore very narrow bandwidths.

Resonance occurs when the length of the resonator is some integer times the guide wavelength divided by 2. Figure 12.19 shows examples of cavity resonators. Figure 12.19(a) shows a halfwave waveguide cavity. The mode for that cavity is TE_{101} . The electric field is transverse to an axis in the length direction, as shown by the side view, and the variation of the electric field is one half-cycle, zero, and one half-cycle in the a , b , and l directions, respectively. The a dimension is the long dimension for the end view, the b dimension is the short dimension for the end view, and the l dimension is the long dimension for the side view.

Figure 12.19(b) shows a cylinder-type resonant cavity. The mode for this cavity is TM_{010} . Here, TM denotes that the magnetic field lies in planes transverse to the axis of the cylinder. The first and third subscripts denote that the variation of the magnetic field is zero with, respectively, radial direction and position along the axis. The second subscript indicates one half-cycle of variation in the field along a radial line passing from one edge of the cylinder to the other edge. The third number in the subscript indicates the number of half-cycles of electric field in the length dimension.

Figure 12.19(c) shows a reentrant-type resonant cavity. This type of cavity is frequently used in microwave tubes such as klystrons. Here the electric field is most intense in the gap or the region where the top and bottom sides are brought together to form a short space between walls. The magnetic field is most intense near the edges of that center section and falls off toward the outside edges of the cavity. When used with microwave tubes, an electron beam is sent through a hole in the center of the cavity where the top and bottom walls are close together. That couples

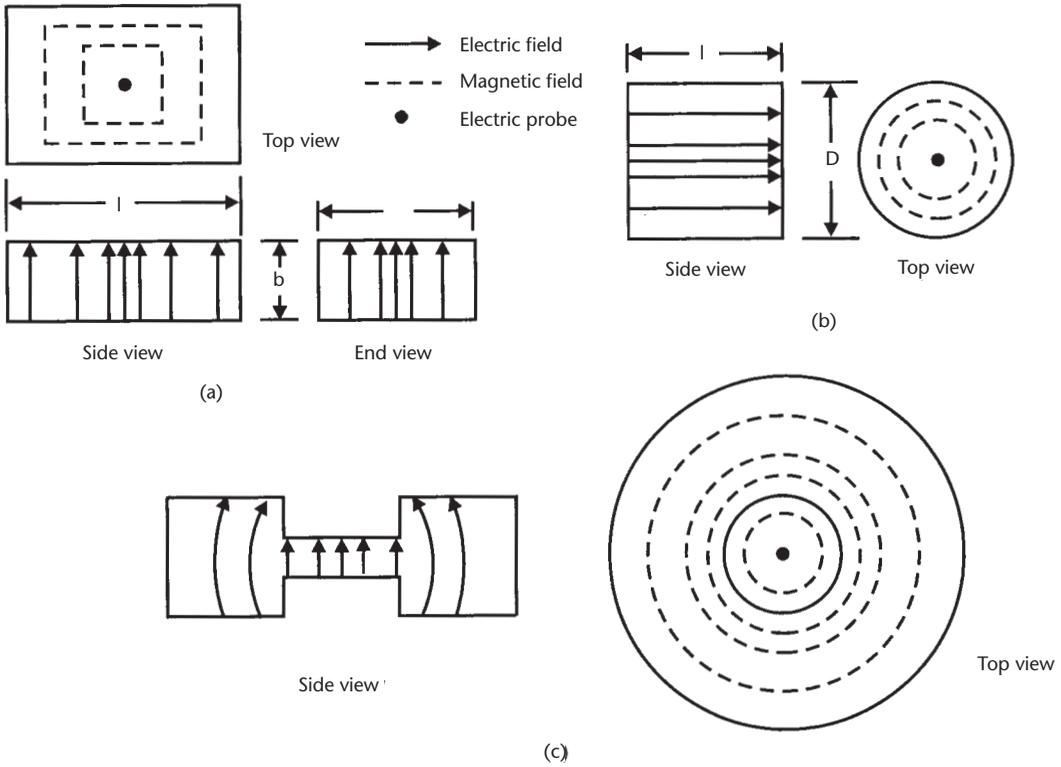


Figure 12.19 Examples of cavity resonators: (a) half-wave waveguide resonator; (b) cylindrical resonant cavity; and (c) resonant reentrant cavity. (After: [4].)

energy into the cavity, or the cavity transmits energy to the beam for velocity modulation.

Exactly the same methods can be used for coupling to cavity resonators as are used to couple to waveguides. Thus, slots, loops, and probes are used when coupling power into or out of a cavity. Electron-beam coupling also is common for energy coupling when the cavity is used in microwave tubes.

Tuning of cavities is done by the same methods as used for impedance matching in waveguides. Adjustable screws or posts are the most often used methods. Another method of tuning a cavity is to have walls that can be moved in or out by mechanical means.

The Q of a cavity resonator can be very high. The 3-dB bandwidth of the cavity is the center frequency divided by the Q . Cavities such as those shown in Figure 12.19(a) and (b) have typical values for Q of about 24,000 if silver-plated copper walls are used. The Q for a reentrant cavity such as that shown in Figure 12.19(c) is typically about 4,000.

References

- [1] Tatsuo, I., "Waveguides and Resonators," in *Reference Data for Engineers*, 8th ed., M. E. Van Valkenburg, (ed.), SAMS, Carmel, IN: Prentice Hall Computer Publishing, 1993, pp. 30-1–30-31.
- [2] Kennedy, G., *Electronic Communication Systems*, 3rd ed., New York: McGraw-Hill, 1985, pp. 288–347.
- [3] Skolnik, M. I., *Introduction to Radar Systems*, 2nd ed., New York: McGraw-Hill, 1980, pp. 359–366.
- [4] Terman, F. E., *Electronic and Radio Engineering*, 4th ed., New York: McGraw-Hill, 1955, pp. 127–165.

Selected Bibliography

Collin, R. E., *Field Theory of Guided Waves*, New York: McGraw-Hill, 1958.

Harvard University, Radio Research Laboratory, *Very High Frequency Techniques*, Vols. 1 and 2, New York: McGraw-Hill, 1947.

Marcuvitz, N., *Waveguide Handbook*, MIT Radiation Laboratory Series, Vol. 10, New York: McGraw-Hill, 1951.

Monreno, T., *Microwave Transmission Design Data*, New York: Dover Publications, 1958.

Southworth, G. C., *Principles and Applications of Waveguide Transmission*, New York: Van Nostrand, 1960.

Young, Leo, "Synchronous Branch Guide Directional Couplers for Low and High Power Applications," *IRE Trans.*, Vol. PGMTT-10, November 1962, pp. 459–475.

Antennas

Much of the material presented in this chapter is quoted or adapted from [1–11].

The same antenna can be used as a transmit and as a receive antenna. A transmit antenna is a transducer that transforms electrical power delivered to the antenna into RF radiation. If the radiation from an antenna is uniform in all directions, it is called an isotropic antenna. Such an antenna is not possible in practice; however, it serves as a useful reference.

The theoretical isotropic antenna is taken as the reference, and the gain of the antenna in a given direction is a measure of how the power level in that direction compares with the power level that would exist if the isotropic antenna had been present. The gain can be either less than 1 or greater than 1. Expressed in decibels relative to isotropic, it can be either positive or negative. For example, a typical tracking radar might use a parabolic dish antenna that produces a 1×1 degree main beam. There are 41,300 square degrees in a spherical solid angle. The directional gain in the main beam would then be about 41,300, or 46.2 dBi. At angles other than the main beam, there will be sidelobes. A typical power gain in the sidelobe directions might be -10 dBi or less, depending on the design of the antenna.

The power gain of an antenna is less than the directional gain because of losses and radiation through the sidelobes. In a typical case, an antenna with a 1×1 degree beam will have a power gain of about 27,000, or 44 dBi. That assumes an antenna efficiency of about 65%, or about 1.85 dB loss. The typical relationship between antenna gain and beam angle is as follows:

$$G = 27,000 / (\theta_{az} \cdot \theta_{el}) \quad (13.1)$$

where:

G = antenna gain as a number

θ_{az} = the azimuth beam angle in degrees

θ_{el} = the elevation beam angle in degrees

Beam angles ordinarily are measured and specified at the half power (-3 dB) points on the beam. The antenna gain expressed in dB with respect to isotropic is $10 \log_{10} G$.

For an example of the use of (13.1), assume the case of a 1.0×10 degree fan beam. Substituting values into (13.1), the gain would be

$$G = 27,000 / (1 \cdot 10) = 2,700 = 34.3 \text{ dBi}$$

The effective radiated power (ERP) is defined as the product of the antenna gain and the radiated power. For example, if the numeric antenna gain is 10 and the radiated power is 200W, the ERP is $200 \cdot 10 = 2,000\text{W}$, or 33 dBw (63 dBm).

A receive antenna is a transducer that captures electromagnetic energy radiated from a distant transmitter antenna and converts it into electrical energy. For most antennas, the gain when an antenna is used as a receive antenna is the same as its gain when used as a transmit antenna. The effective capture area of a receive antenna is its gain times the area of an ideal isotropic antenna for the frequency of interest. The expression for the effective capture area of a receive antenna is given by

$$A_e = G_r \lambda^2 / 4\pi \quad (13.2)$$

where

A_e = effective capture area of an antenna (square meters)

λ = wavelength (meters)

G_r = receive directional antenna gain (numeric)

For an example of the use of (13.2), assume an antenna with a gain of 13 dBi or 20 numeric operating at a wavelength of 2m. Substituting values into (13.2), the effective capture area of the antenna would be

$$A_e = 20 \cdot 4 / 4\pi = 6.37 \text{ m}^2$$

In the case of aperture-type antennas, such as horn antennas and parabolic dish reflector antennas, the effective capture area of the antenna typically is about half the actual aperture area. For example, a parabolic dish antenna with a diameter of 60 feet (18.3m) has an aperture area of about 263 m² and an “effective” capture area of only about 132 m².

The assumption has been made that the polarization of the receiver antenna is the same as the polarization of the signal being received. If that is not the case, there is a polarization loss. If the signal being received has a vertical polarization (direction of the E-field is vertical) and the receive antenna polarization is also vertical, there is no polarization loss. If, on the other hand, the polarization of the receive antenna is horizontal, the polarization loss could be 30 dB or more. For cases where the polarization angle difference is greater than 0 degrees but less than 90 degrees, the polarization loss will be between 0 and 30 dB, depending on the angle difference.

If the polarization of the signal being received is circular (E-field rotates) and the polarization of the receiver antenna is linear (horizontal or vertical), there is a 3-dB polarization loss. If the polarization of the signal being received is right-hand-circular and the polarization of the receiver antenna is left-hand-circular, the polarization loss could be 30 dB or more.

13.1 Monopole Antennas

13.1.1 Thin-Wire Monopole Antennas

One of the most commonly used antennas is the quarter-wave monopole antenna. This type of antenna can be thought of as being derived from the coaxial transmission line and is illustrated in Figure 13.1(a). Resonance occurs when the length of the monopole center element is about a quarter of a wavelength at the frequency of interest.

The radiation pattern for a quarter-wave monopole antenna is as shown in Figure 13.1(b). For a vertical monopole, the polarization is linear and vertical. The gain pattern is uniform in the azimuth plane. The elevation pattern has a peak a few degrees above zero elevation angle. It then slowly drops to a lower value at 90 degrees. The exact pattern depends on how large and how conductive the ground plane is: the larger the ground plane, the lower the elevation angle of the peak in the radiation pattern.

Other patterns are produced when the monopole antenna has a height greater than a quarter-wavelength. An example is the case of a $5/8\lambda$ -long monopole. The antenna pattern for this antenna is shown in Figure 13.1(c). The antenna gain in the horizontal direction is about 2 dB greater than that of the quarter-wave monopole. Such monopole antennas are frequently used for cellular telephones.

The peak antenna gain for a quarter-wave monopole is about 5 dBi. The exact antenna gain depends on elevation angle and losses in the antenna.

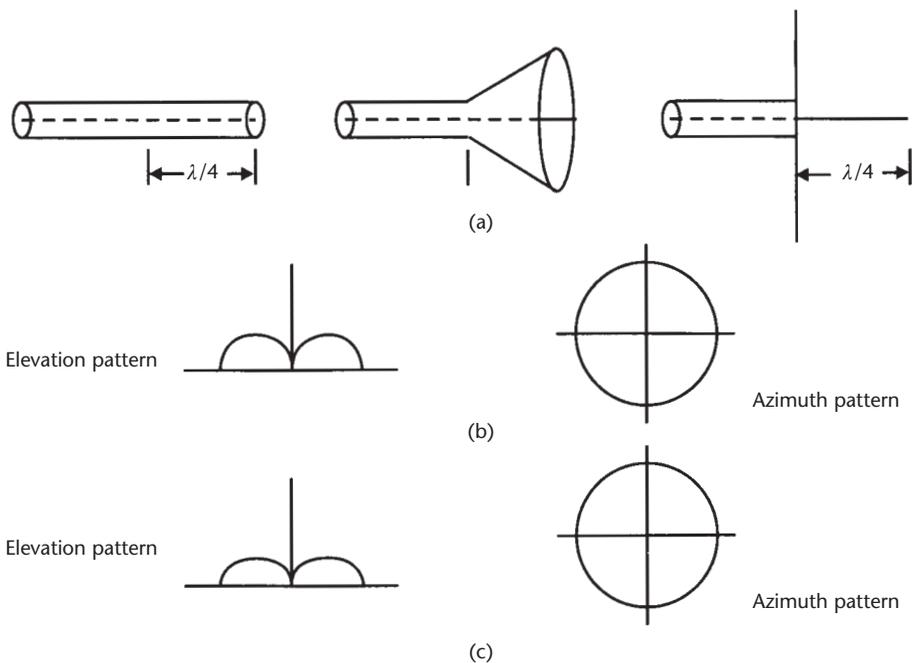


Figure 13.1 Monopole antennas: (a) monopole antenna being derived from the coax transmission line; (b) radiation and polarization for a quarter-wave monopole antenna; and (c) radiation pattern and polarization for a monopole antenna with $5/8\lambda$ height.

Figure 13.2(a) shows a quarter-wave monopole antenna with the ground plane at an angle of about 120 degrees from the center conductor. At that angle, the impedance at resonance has increased from about $37 + j0\Omega$ to about $50 + j0\Omega$.

The elevation pattern for this monopole antenna is shown in Figure 13.2(b). This type of monopole configuration is often used as a communication antenna at VHF and higher frequencies. It provides good near-horizon coverage, and the higher impedance facilitates proper and easy impedance matching between the antenna and a 50Ω coaxial cable.

13.1.2 Wideband Monopoles

Figures 13.2(c–e) show “fat” monopoles. Each of these antennas is a wideband antenna. The reason for wide bandwidth is that the impedance for these monopoles remains fairly constant over a large range of frequencies.

13.1.3 Impedance of Monopole Antennas

Figure 13.3 shows impedance trends for a conical monopole antenna. In those plots, the resistances and reactances are given as a function of antenna length in electrical degrees and cone angle. The length, L , in electrical degrees is $L = 360h/\lambda$, where h is the height of the monopole and λ is the wavelength. Notice that for cones with small cone angles there are vary large changes in both resistance and reactance as a function of antenna length. On the other hand, for cones with large cone angles, the variations are small, indicating a wider bandwidth capability.

Similar resistance and reactance plots can be provided for other shapes of monopoles, such as thin wires, fat cylinders, rectangular plates, and fan-shaped

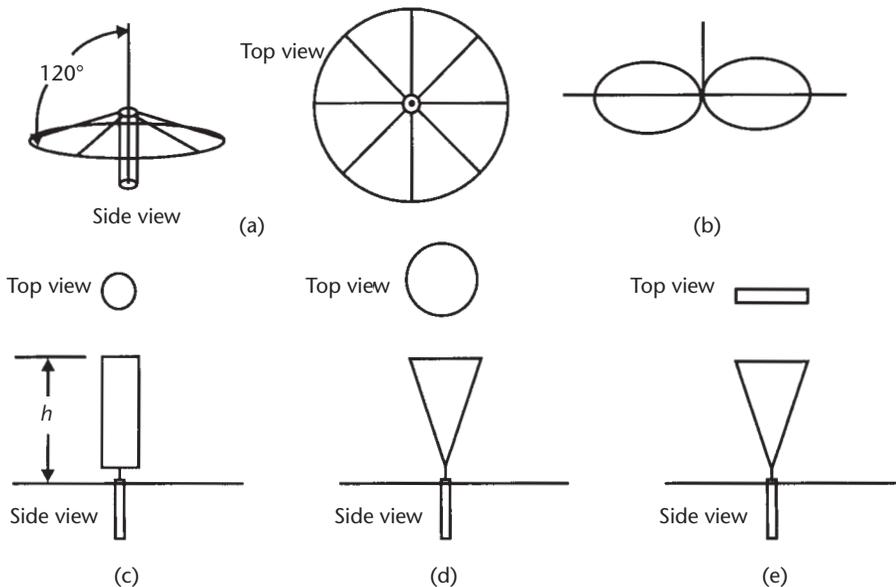


Figure 13.2 Example monopole antennas: (a) antenna with ground plane at 120 degrees from center conductor; (b) elevation radiation pattern for the antenna in (a); (c) fat cylinder monopole; (d) conical monopole; and (e) fan-shaped monopole.

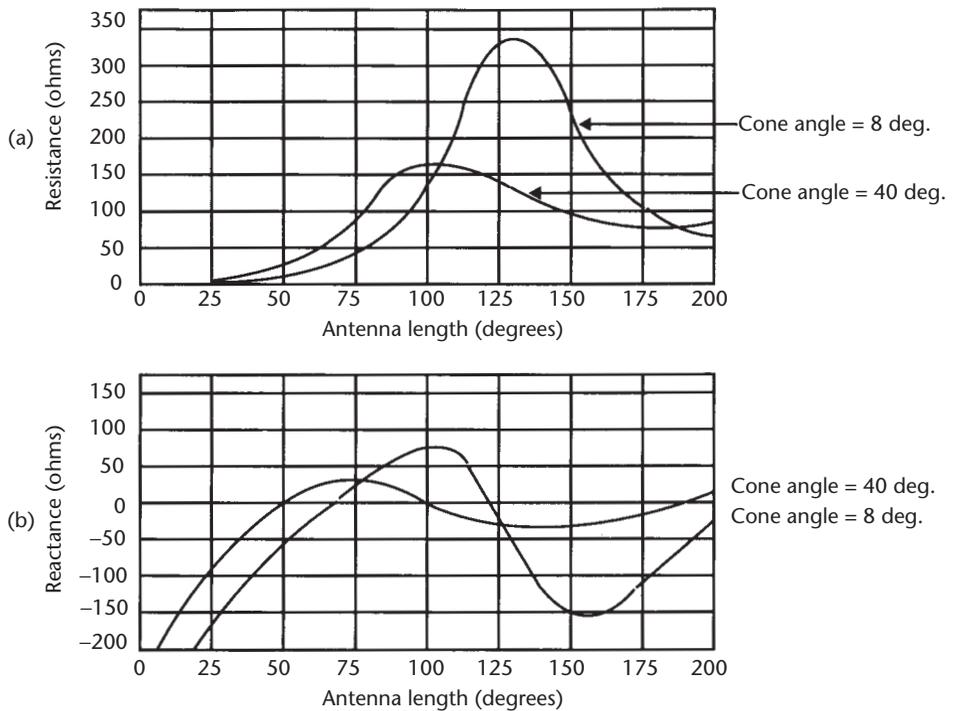


Figure 13.3 Resistance and reactance curves for conical monopole antennas: (a) resistance and (b) reactance. (After: [1].)

plates. The result would be the same general trend in impedance. Small-diameter monopoles may provide only 5% bandwidth, whereas large-diameter monopoles may provide 70–100% bandwidth.

13.1.4 Large-Size Monopole Antennas

At HF and MF frequencies, resonant monopole antennas are very large. For an example of the size of the quarter-wavelength monopole antenna, assume operation at 3 MHz. The wavelength at that frequency is 100m and a quarter-wavelength is 25m. Because of end effects, the required monopole length would be about 0.95 times that, or 23.75m. The ground plane should have a diameter of at least twice that, or 47.4m. An even larger ground diameter would be desirable for improving low-angle coverage.

Ground planes for MF and lower frequency monopole transmitter antennas frequently are constructed over highly conductive ground. In addition, radial copper wires usually are used in the ground planes to reduce ground-plane losses.

13.1.5 Electrically Small Monopole Antennas

It is not practical to construct full-size resonant monopole antennas at low MF, LF, and VLF because of the very large wavelength that is involved. For example, LF has a frequency range of 30–300 kHz and a wavelength of 10,000-1,000m. A quarter-wavelength monopole at LF would have a height of 2,500-250m.

Antennas that are much less than a quarter-wavelength high are referred to as electrically small antennas. An electrically small antenna has a large capacitive reactance and a very small radiation resistance. The approximate radiation resistance for an electrically small monopole antenna is given by

$$R_r = 800(h/\lambda)^2 \quad (13.3)$$

where

R_r = radiation resistance (ohms)

h = height of monopole (meters)

λ = wavelength (meters)

For an example of the use of (13.3), assume a monopole that is 50m high and operating at a wavelength of 1,000m. The antenna height is thus only 5% of a wavelength. Substituting values into (13.3), the approximate radiation resistance would be

$$R_r = 800(50/1,000)^2 = 2\Omega$$

One way to reduce the capacitive reactance of the electrically small antenna while at the same time improving the radiation resistance is to use capacitive or inductive loading. Examples of this are shown in Figure 13.4.

Figure 13.4(a) shows a typical LF antenna with flat-top loading consisting of a number of parallel wires connected between towers or masts, a vertical wire or set of wires that act as the monopole connected to the horizontal wires, and a series tuning inductance at the bottom.

Figure 13.4(b) shows a so-called umbrella-type loaded monopole in which conductors are extended at a large angle to the vertical using guy wires with insulators. The monopole antenna connects to the umbrella wires and has a tuning inductor at the bottom, that is, base loading. The same type of improvement is provided with this approach as that provided by the system in Figure 13.4(a) without the need for two masts or towers.

The tuning coils and transformers needed for impedance matching at LF typically are very large. The reason is the need for very low ohmic resistance for the coil and the need for very high Q . Since we are dealing with series resonance, very large currents are involved in a high-power system.

The monopole antenna normally is fed by a coaxial transmission line and does not require a balun. At the lower frequencies, impedance matching usually is accomplished using lumped constant inductor and capacitor circuits. Stub matching can be used at higher frequencies.

13.2 Dipole Antennas

13.2.1 Thin-Wire Dipole Antennas

Another commonly used antenna is the half-wave dipole antenna. The most common of this type is the thin-wire dipole. The thin-wire type of dipole antenna can be

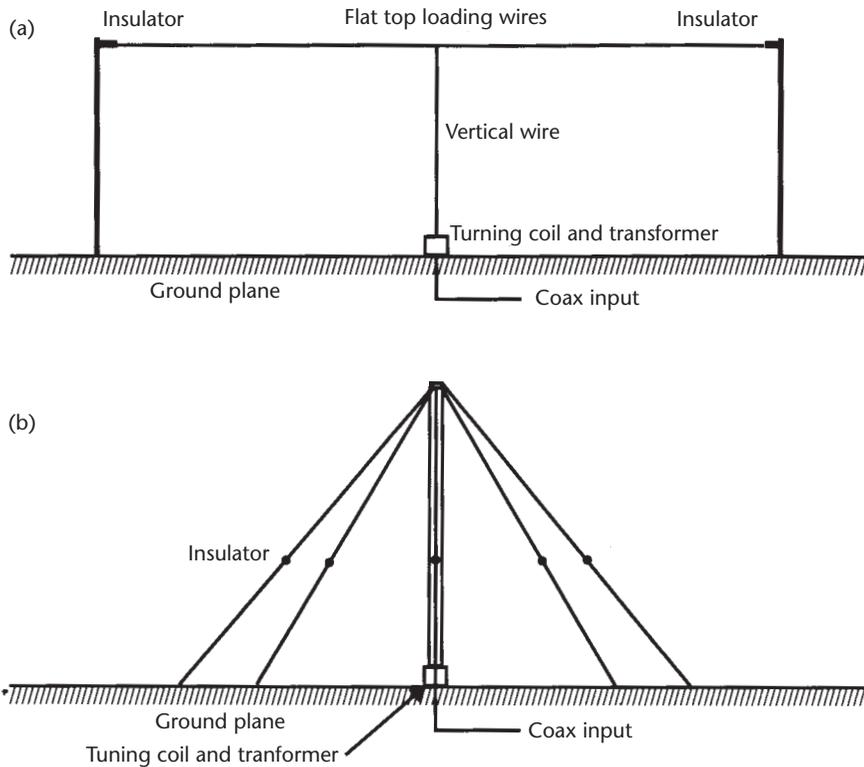


Figure 13.4 Examples of electrically small antennas for LF operation: (a) LF antenna with flat-top loading and inductive tuning and (b) umbrella-type loaded monopole.

thought of as being derived from the two-wire transmission line shown in Figure 13.5(a).

Figure 13.5(b) shows a quarter-wave section of a two-wire line spread into a V-shape, which becomes a V-type dipole. This form of dipole has fairly narrow beams in the direction of the center line of the V.

At Figure 13.5(c) is shown the quarter-wave section of two-wire line spread 90 degrees on either side of the center line for the transmission line. This is the normal dipole configuration.

The antenna pattern for a half-wave dipole antenna is shown to the right of the dipole in Figure 13.5(c). The azimuth radiation pattern is omnidirectional in the plane normal to the axis of the dipole. The elevation pattern is maximum normal to the dipole and varies as the cosine of the angle. The 3-dB elevation beam angle is about 78 degrees wide. The pattern has a deep null directly in line with the axis of the dipole. The maximum antenna gain for a half-wave dipole is about 2 dBi.

Other patterns result when the dipole antenna is other than a half-wave. In general, there are multiple lobes when the dipole is much larger than a half-wave. When shorter than a half-wave, the pattern is similar to that of the quarter-wave dipole.

Dipole antennas are fed by two-wire lines. When the transmission line to the antennas is coaxial, it is necessary to add a balun between the coax and the antenna because it is a balance (ungrounded) feed. A number of different types of baluns were discussed in Chapter 11.

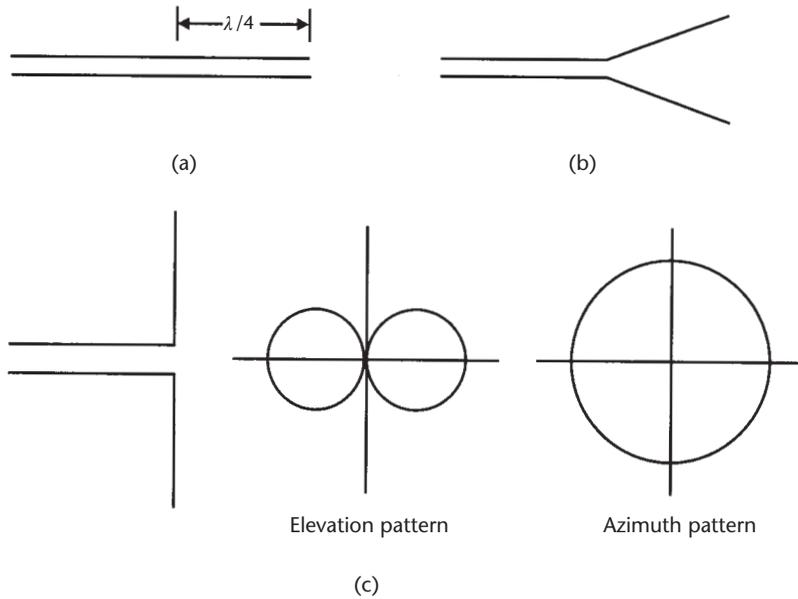


Figure 13.5 Dipole antennas derived from two-wire transmission line: (a) two-wire transmission line; (b) V-type antenna; and (c) dipole antenna.

13.2.2 Other Types of Dipole Antennas

Other types of dipole antennas are shown in Figure 13.6.

An electrically small dipole antenna, shown in Figure 13.6(a), has arms much less than a quarter-wavelength. It frequently is used at the lower frequencies, where it is not practical to use a full-size dipole. This type of antenna requires the use of impedance-matching circuits.

The antennas shown in Figures 13.6(b–e) are all so-called fat dipoles, used to provide broadband operation. They have essentially the same antenna patterns as thin-wire dipoles but different impedances.

Figure 13.6(f) shows the case of the folded dipole antenna. This antenna frequently is used as the feed element for television receiving antennas. A folded dipole antenna is made by connecting the ends of a regular antenna with a conductor that is spaced a short distance from the dipole. A transformer action is involved in the operation of the antenna. A typical system has the connecting conductor the same diameter as the dipole. When that is the case, the antenna has a radiation resistance about four times that of an ordinary dipole antenna. The resistance thus is about 292Ω at resonance, a decent value for matching to a 300Ω two-wire transmission line. Other radiation resistance values are possible with the folded dipole if different diameters are used for the two antenna elements. The folded dipole has a larger bandwidth than a simple thin-wire dipole, while having essentially the same antenna pattern. Polarization direction is again in line with the axis of the dipole.

13.2.3 Dipole Impedance

Figure 13.7 shows example resistive and reactive components of the antenna impedance as a function of the ratio of the dipole length to wavelength (L/λ) for a thin

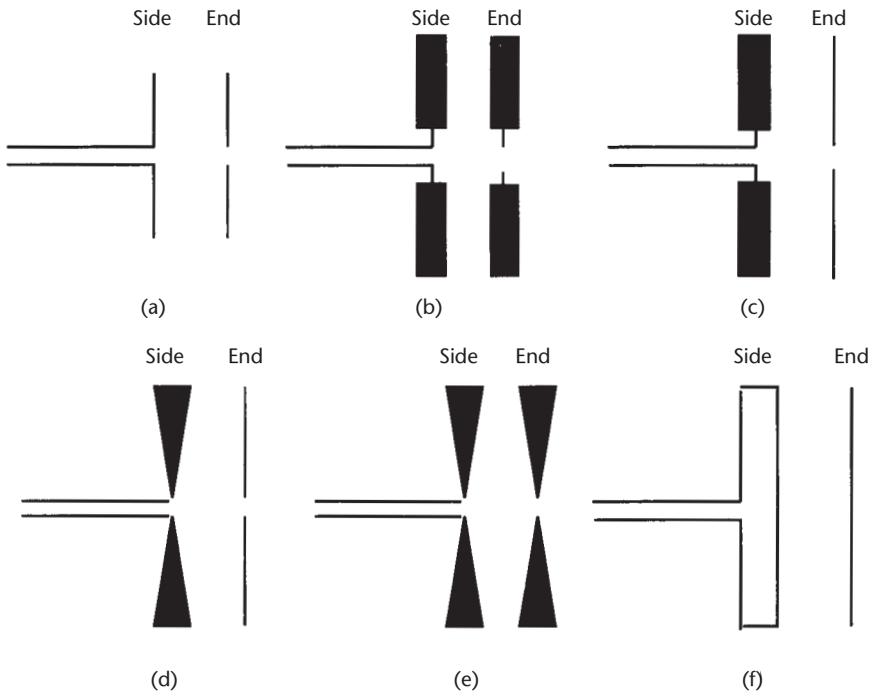


Figure 13.6 Other types of dipole antennas: (a) thin-wire electrically small dipole; (b) fat-cylinder dipole; (c) rectangular-plate dipole; (d) fan dipole; (e) conical dipole; and (f) folded dipole.

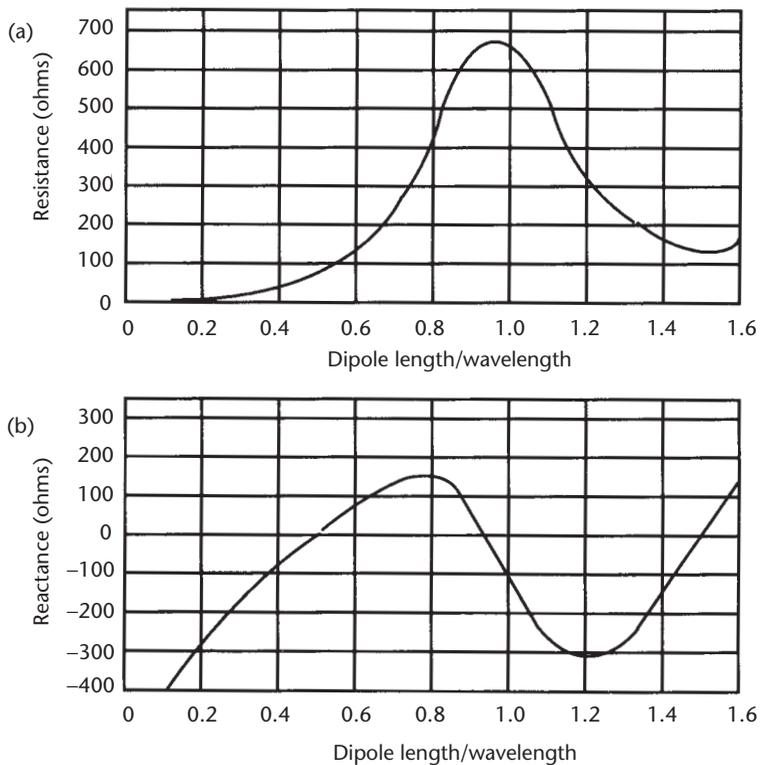


Figure 13.7 Resistance and reactance curves for a thin-wire dipole antenna: (a) resistance and (b) reactance. (After: [1].)

cylindrical dipole antenna. The impedance behaves very much like that of an open-circuit transmission line, except that radiation resistance is involved. Like the two-wire transmission line, the impedance of the dipole antenna is high and capacitive when the distance from the end of the dipole to the feed point is small compared to a quarter-wavelength. As the dipole is made larger, the capacitive reactance becomes less. When the arms of dipole are each a quarter-wavelength long, the reactance becomes zero and the antenna is at resonance. In the case of the thin-wire dipole, the resistive component is about 73Ω . The exact value depends on the length-to-diameter ratio of the dipole. The resistive component is made up largely of the radiation resistance of the antenna, which accounts for the power radiated.

As the arms of the dipole are increased beyond a quarter-wavelength but less than a half-wavelength, the impedance becomes inductive and larger. The resistive component continues to increase with increasing dipole length-to-wavelength ratio until it reaches a peak when the length of the dipole is near a wavelength (arms one-half wavelength). With increasing L/λ , the resistance drops and the reactance switches from inductive to capacitive. It then falls again, reaching 0 when L/λ is near 1.5. The resistance has dropped to a low value at that condition. We thus have another resonance. With larger L/λ the impedance switches to inductive, the resistance increases, and so on.

The thin-wire dipole has relatively narrow bandwidth, as small as 5%. The exact bandwidth depends on the length-to-diameter ratio: the smaller that ratio, the larger the bandwidth. On the other hand, as the dipole elements become large, the bandwidth can be quite large. Bandwidths near three-quarters of an octave are possible with wideband dipole antennas. The antenna gain, antenna patterns, and polarization characteristics are similar to those of the narrowband dipoles.

The peak antenna gain for a normal $\lambda/2$ dipole antenna is about 2 dB. Larger antenna gains are possible by using reflectors, directors, or both with the dipole. An example of a wideband UHF fan-type dipole with a corner reflector is shown in Figure 13.8. An example antenna of this type is useful over a frequency range of about 450–900 MHz. It has a gain in the range of about 8–12 dBi, depending on frequency.

The measured impedance characteristics for the antenna in Figure 13.8 are plotted on the Smith chart shown in Figure 13.9. For this case, the impedance of the Smith chart is normalized to 280Ω . It can be seen from this chart that the VSWR remains less than 2.5:1 over the full octave frequency range of 450–900 MHz. It is common practice to use a Smith chart in showing the impedance characteristic of antennas.

13.2.4 Dipole Current Distribution and Antenna Patterns for Different L/λ Ratios

Figure 13.10 illustrates current distributions on a thin-wire dipole for different L/λ ratios, where L is the full length of the dipole ($L = 2b$) and λ is the wavelength. This figure also shows corresponding antenna elevation patterns.

Figure 13.10(a) shows the case of $L < \lambda/2$. This is the electrically small antenna case. The currents at the ends of the dipole are zero. They increase toward the center but remain small compared to that of the half-wave resonant dipole. The elevation

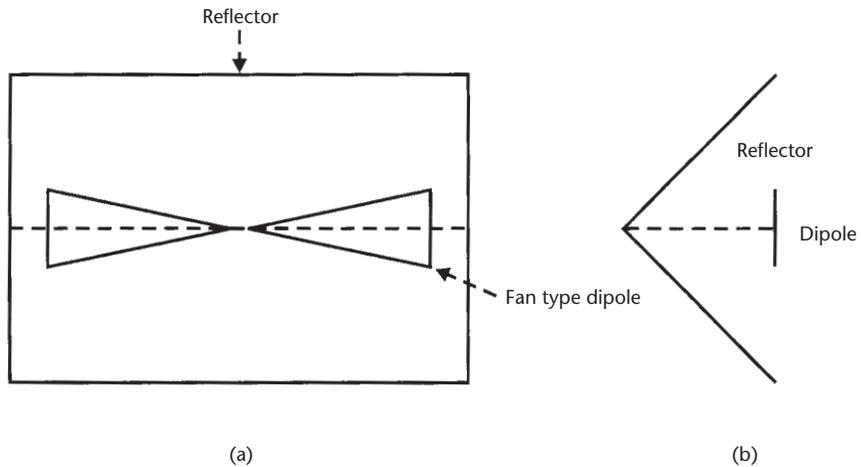


Figure 13.8 Fan dipole with corner reflector: (a) front view and (b) side view. (After: [3].)

pattern is shown as circles on either side of the dipole with the -3 dB beamwidth equal to about 90 degrees.

Figure 13.10(b) shows the case of $L = \lambda/2$, that is, the resonant antenna case. The currents at the ends of the dipole are zero. They increase toward the center in a sinusoidal fashion, reaching a maximum at the center of the antenna. The elevation pattern is shown as a fat ellipse on either side of the dipole, with the -3 -dB beamwidth equal to about 78 degrees.

Figure 13.10(c) shows the case of $L = 3\lambda/4$. The currents at the ends of the dipole are zero. They increase as we move toward the center in a sine wave fashion, reaching a maximum and then dropping off to about 70% of the peak at the center of the antenna. The elevation pattern is shown as an ellipse on either side of the dipole, with the -3 -dB beamwidth equal to about 60 degrees.

Figure 13.10(d) shows the case of $L = \lambda$. The currents at the ends of the dipole are zero. They also increase toward the center in a sine wave fashion, reaching a maximum and then dropping off to nearly zero at the center of the antenna. The elevation pattern is shown as an ellipse on either side of the dipole, with the -3 -dB beamwidth equal to about 47 degrees.

Figure 13.10(e) shows the case of $L = 3\lambda/2$. The currents at the ends of the dipole are zero. They increase toward the center in a sine wave fashion, reaching a maximum and then dropping off to zero at a point one-third the way from the center to the ends and then increasing in the opposite direction to a maximum at the center of the antenna. The elevation pattern is shown as a six-lobe pattern.

13.2.5 Turnstile Antenna

Figure 13.11(a) shows the turnstile antenna configuration. A turnstile antenna is made up of two dipole antennas placed 90 degrees from each other and fed 90 degrees out of phase. A quadrature coupler can be used to provide the 90-degree quadrature outputs. This antenna frequently is used to transmit or receive circular polarization. In a direction normal to the two dipoles, the polarization is circular. In

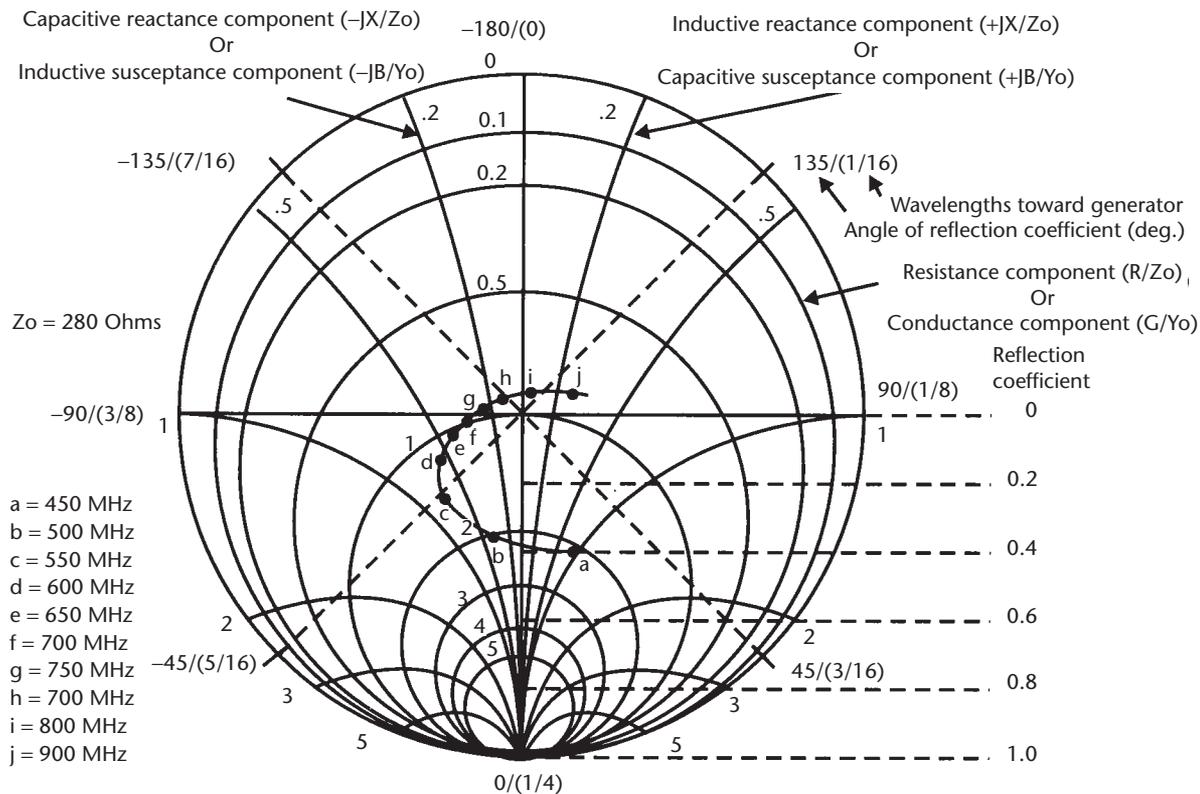


Figure 13.9 Impedance characteristics for a UHF corner reflector antenna. (After: [3].)

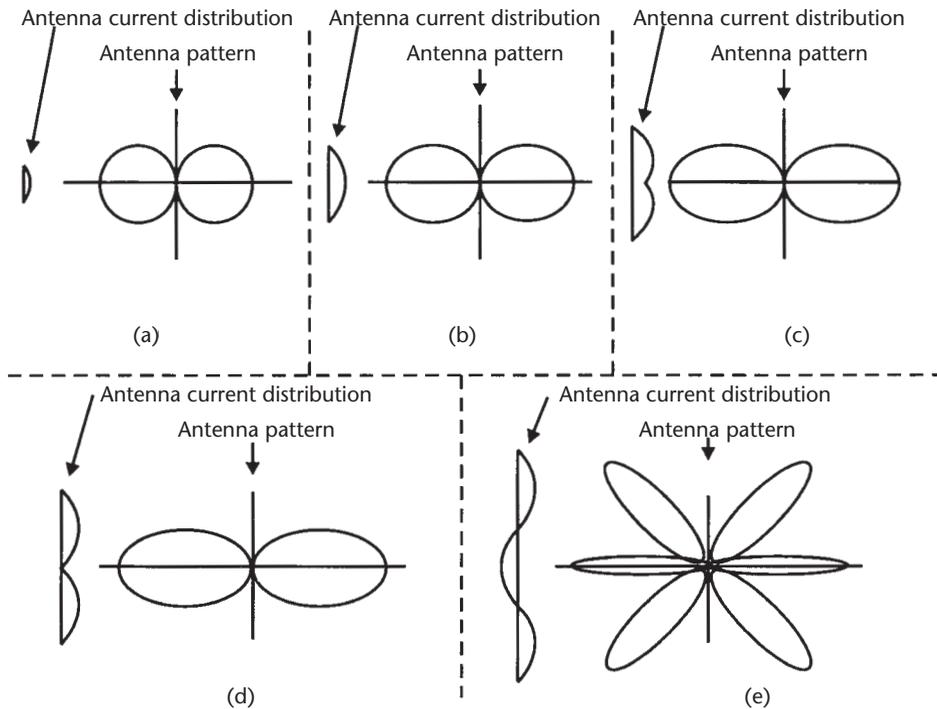


Figure 13.10 Current distribution and antenna patterns for different dipole sizes: (a) electrically small antenna ($L < 0.5$ wavelength); (b) resonant antenna ($L = 0.5$ wavelength); (c) $L = 0.75$ wavelength; (d) $L = 1$ wavelength; and (e) $L = 1.5$ wavelength.

the plane of the dipoles, the polarization is linear with the E-field in the plane of the dipoles.

Figure 13.11(b, c) shows the azimuth and elevation patterns for the turnstile antenna. The azimuth pattern shows the individual patterns for the two dipole antennas plus the vector sum or superposition of the two patterns, respectively. The elevation pattern shows that the coverage is nearly isotropic. In this pattern, the polarization is shown to be circular. Circular polarization means that the direction of the electric field vector rotates in either the right-hand or left-hand direction, depending on the direction of the phase shift. If the two antennas have different effective radiated powers, the polarization is elliptical.

13.3 Yagi-Uda Antennas

Figure 13.12 shows a Yagi-Uda antenna, or Yagi antenna as it is usually called. This antenna is an important example of a directional antenna that uses a dipole as its feed element. Often this dipole is a folded dipole. This type of antenna frequently is used as a receiving antenna for television. It also is used for many other applications where moderate directional gain is required.

On one side of the dipole are a number of director elements that are shorter than the feed dipole by 5–10%. The elements, which are spaced by 0.15–0.25 wavelength, are not dipoles but single rods that act as parasitic elements. The number of

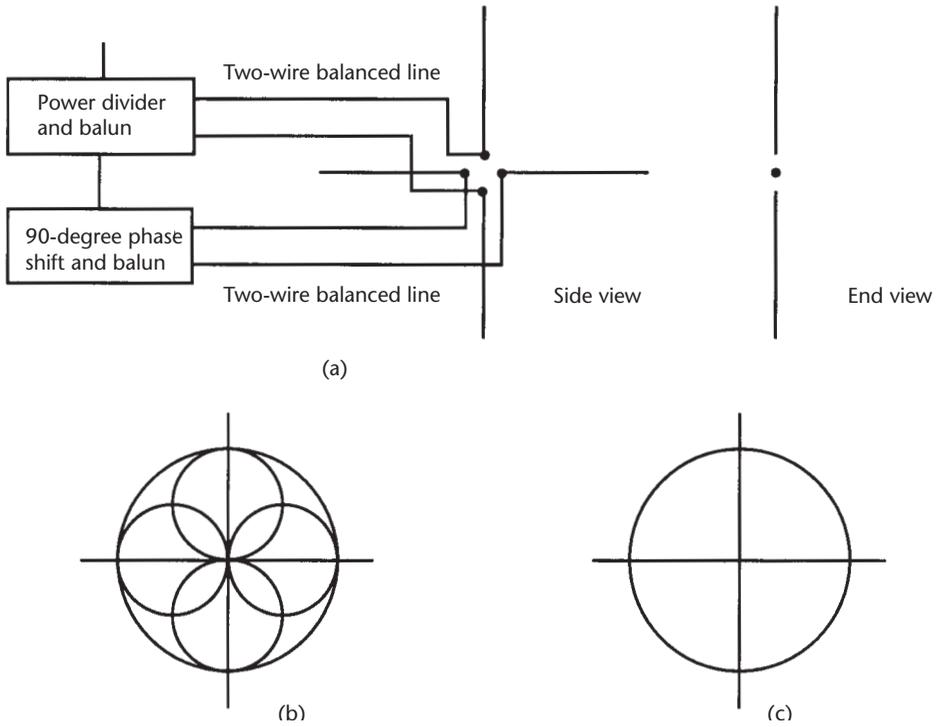


Figure 13.11 Turnstile antenna with antenna patterns: (a) turnstile antenna configuration; (b) azimuth pattern; and (c) elevation pattern.

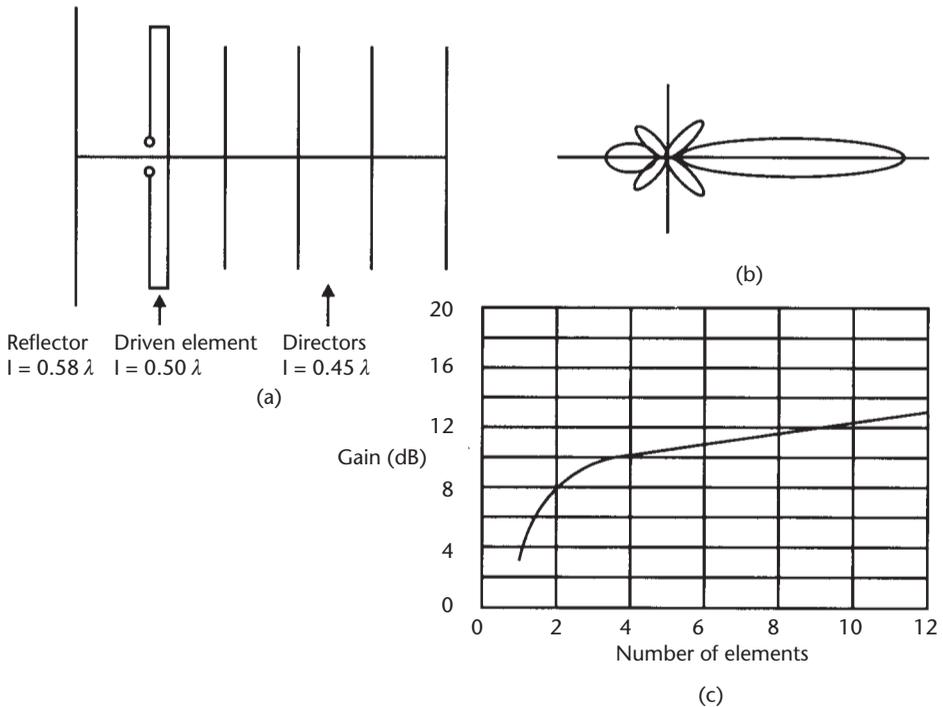


Figure 13.12 A Yagi-Uda, or Yagi, antenna: (a) example antenna configuration; (b) typical antenna pattern; and (c) gain characteristics for a Yagi antenna. (After: [4].)

elements depends on the desired gain for the antenna. On the other side of the dipole is a single reflector element. The length of that rod is 5–10% greater than that of the feed dipole. Again, the spacing from the dipole is in the range of 0.15–0.25 wavelength.

A typical antenna pattern for a Yagi antenna is shown in Figure 13.12(b). It has a main forward lobe with a typical peak antenna gain in the range of 10–12 dBi. It also has a back lobe with a peak gain of the order of 0–3 dBi. Thus, the typical front-to-back ratio is 9 to 10 dB.

A Yagi antenna is an end-fire traveling-wave antenna. Because the director array of elements is parasitic, the currents on the elements farther out from the driver have decreasing current amplitudes. Figure 13.12(c) shows how the gain of the antenna increases as the number of elements increases. We see that there is little value in increasing the number of director elements beyond about 10.

13.4 Sleeve Antennas

The addition of a sleeve to a monopole or dipole antenna can increase the bandwidth to more than an octave.

13.4.1 Sleeve Monopoles

Figure 13.13(a) illustrates a sleeve monopole configuration that has a 4:1 pattern bandwidth. The antenna is fed by a coaxial transmission line. The first sleeve monopole resonance occurs when $L + 1$ is approximately $\lambda/4$. The physical length is set by the low end of the frequency band to be covered. Thus, if we want to operate over a band of frequencies from 150 MHz to 600 MHz, the length $L + 1$ would have a value of about $2m/4 = 0.5m$.

The next requirement is to select the ratio $1/L$. It has been found experimentally that a value of $1/L$ of 2.25 yields optimum radiation patterns over a 4:1 band. The ratio D/d should be about 3.0. The VSWR is less than 8:1 for this 4:1 band. In most

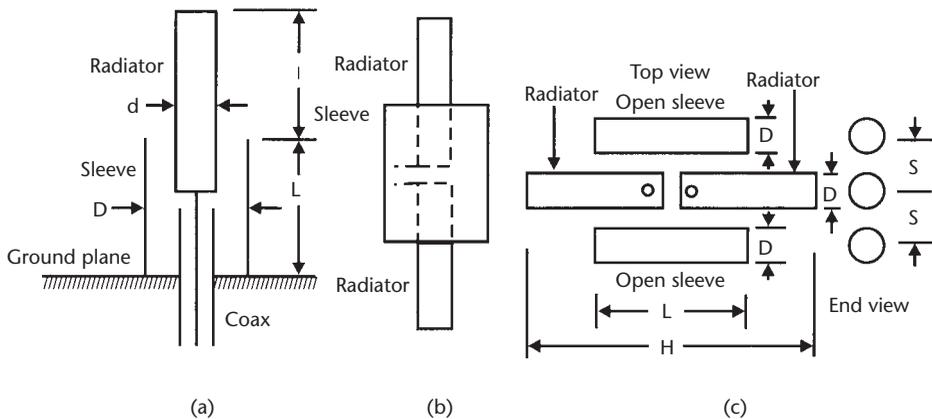


Figure 13.13 Sleeve antennas: (a) sleeve monopole; (b) sleeve dipole configuration; and (c) open-sleeve dipole antenna with reflector. (After: [5].)

applications, that is too high, requiring a matching network for proper performance and to minimize mismatch losses.

13.4.2 Sleeve Dipoles

Two forms of sleeve dipole antennas are shown in Figure 13.13. The operation of these antennas is similar to that of sleeve monopoles. Figure 13.13(b) shows the case of a sleeve dipole with the two-wire transmission line input brought into the sleeve through a hole in the sleeve. Figure 13.13(c) shows the case of an open sleeve dipole antenna mounted above a reflector surface. In that case, the tubular sleeve is replaced by two parasitic conductors that simulate the sleeve. The length of the conductors is approximately one-half that of center-fed dipoles. Details for the feed for this open-sleeve dipole are shown in the figure. The antenna is fed by a coaxial cable through one support arm for the dipole. The outer conductor for the coax is connected to the left dipole element. The center conductor is connected to the right dipole element.

For an example design, assume an antenna designed to operate over the 225–400-MHz band. The dipole-to-reflector spacing, S_d , is chosen to be 0.29λ at 400 MHz. All the dimensions required for the antenna expressed in wavelengths at the lowest frequency are as follows:

- $D = 0.026$;
- $H = 0.385$;
- $L = 0.216$;
- $S = 0.381$;
- $S_d = 0.163$.

These design values yield a VSWR that is less than 2.5 over the full operating range from 225 to 400 MHz.

13.5 Loop Antennas

Loop antennas seldom are used as transmit antennas. They are, however, used extensively as receive antennas at lower frequencies and sometimes in direction-finding systems. The two main types of loop antennas are air-core loop antennas and ferrite-core loop antennas.

13.5.1 Air-Core Loop Antennas

Figure 13.14(a) is a small air-core loop antenna with a few turns. This type of antenna sometimes is referred to as a magnetic dipole.

The antenna pattern for loop antennas is shown in Figure 13.14(c). The antenna pattern is similar to that of an electric dipole antenna, with the maximum gain in the plane of the loop and a deep null normal to the plane of the loop. The polarization is in the plane of the loop.

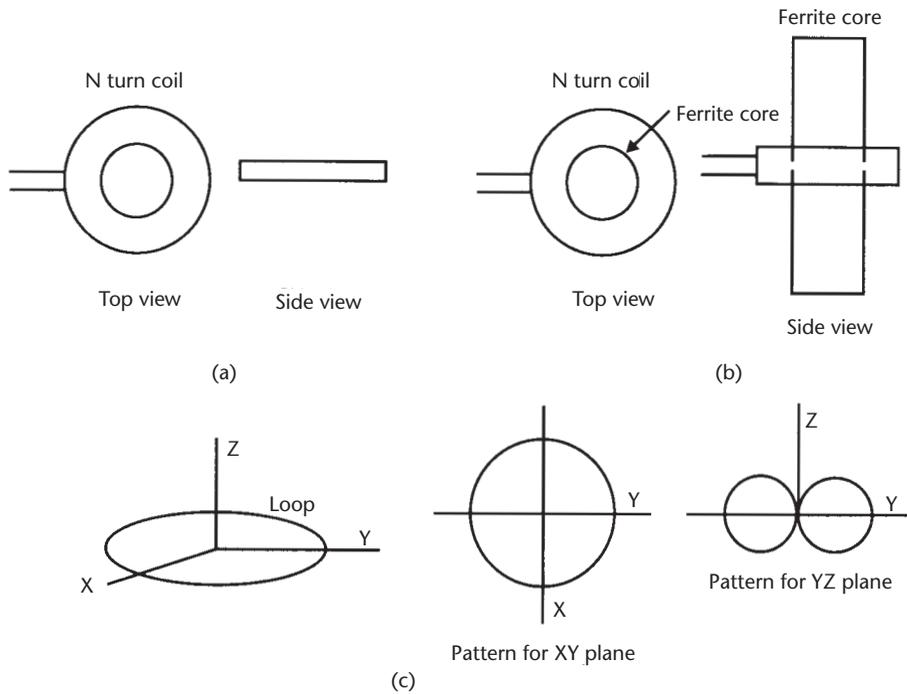


Figure 13.14 Loop antennas: (a) air-core loop antenna; (b) ferrite-core loop antenna; and (c) antenna pattern for loop antennas.

The electrically small loop antenna has a large inductive reactance and a small radiation resistance. It can be shown that when the perimeter of a circular loop antenna is less than about three-tenths of a wavelength, the approximate radiation resistance of this antenna is given by

$$R_r = 19,000 N^2 (D/\lambda)^4 \quad (13.4)$$

where:

D = loop diameter

λ = wavelength

N = number of loops or turns

For example, if $D = 0.1\lambda$ and $N = 4$, then

$$R_r = 19,000(16)(0.1)^4 = 30.8\Omega$$

Equation (13.4) shows that the radiation resistance of a loop antenna is increased by using a number of turns rather than just one. However, the total length of the wire involved must be less than about three-tenths of a wavelength.

The voltage induced in a small loop receiving antenna is given by

$$V = (2\pi/\lambda)nSE_\phi \sin \theta \quad (13.5)$$

where:

n = number of turns

S = surface area of loop

E_ϕ = component of electric field in the plane of the loop

θ = angle measured from the axis of the loop

Other terms are as defined for (13.4).

13.5.2 Ferrite-Core Loop Antennas

Another way to improve the radiation resistance of the loop antenna is to use a ferrite core for the loop. A ferrite-core multiturn loop is often called a loop-stick antenna. An antenna of this type is shown in Figure 13.14(b). Ferrite-core loop antennas are used for many AM broadcast receivers.

Antenna pattern and polarization for the ferrite-core loop antenna are the same as that of the air-core loop antenna. The radiation resistance of a coil of N turns wound on a ferrite core is given by

$$R_r = 19,000N^2 \mu_{\text{eff}} (D/\lambda)^4 \quad (13.6)$$

where μ_{eff} = effective permeability of the core material and other terms are as defined for (13.4). The inductive reactance term can be canceled using a series capacitor of equal reactance magnitude.

13.6 Helical Antennas

Figure 13.15 shows a helical antenna operating in the axial mode of radiation. In this mode, the helix radiates as an end-fire, traveling-wave antenna with a single maximum along the axis of the helix and a phase velocity along the helix axis less than the speed of light. The helix is thus a slow wave structure. The radiation has circular polarization. The antenna may have either left-hand or right-hand circular polarization, depending on the sense of the winding. If the turns go counterclockwise toward the end of the antenna (in the direction of radiation), polarization is left-hand circular polarization and vice versa.

A helical antenna has a coaxial input with the outer conductor connected to a ground plane and the inner conductor connected to the helix. The ground-plane size is not critical but should be made wider than a half-wavelength at the lowest frequency of interest. The circumference is about a wavelength at the center frequency, and the antenna is effective with the circumference in the range of three-quarters wavelength to four-thirds wavelength. The ratio of the lowest operating frequency to the highest operating frequency thus is about 1.8 to 1. The spacing between turns is about 0.21 wavelength, and the number of turns is approximately 12.

Equation (13.7) is an empirical formula for the half-power beamwidth for a helical antenna.

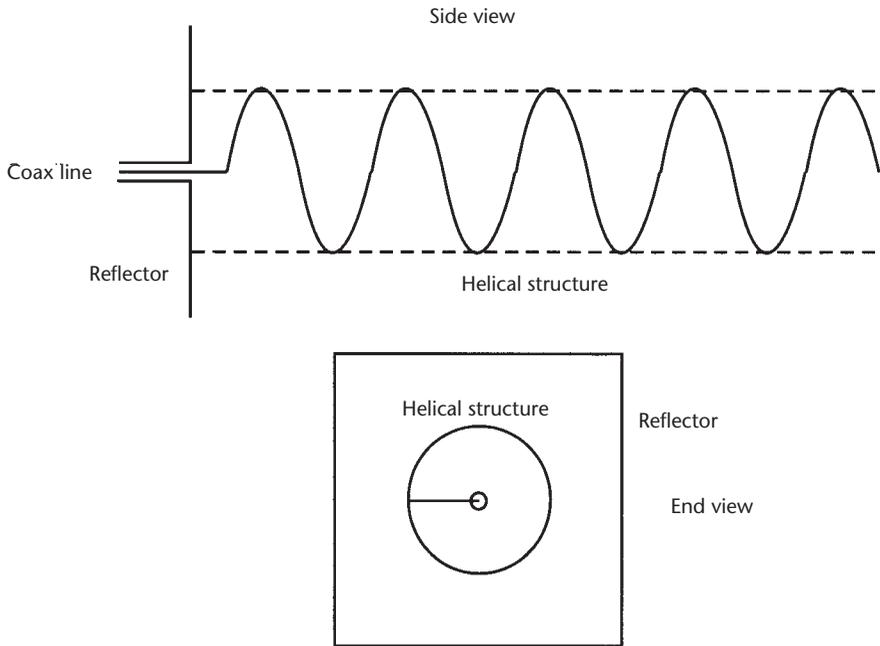


Figure 13.15 Axial-mode helical antenna.

$$\theta_{\text{hp}} = 52 / \left[(C/\lambda)(NS/\lambda)^{1/2} \right] \quad (13.7)$$

where

C = circumference of helix

λ = wavelength

N = number of turns for helix

S = spacing between turns

For an example of the use of (13.7), assume that $C/\lambda = 1$, $S/\lambda = 0.21$, and $N = 12$. Substituting those values into (13.7) gives

$$\theta_{\text{hp}} = 52 / \left[(1)(12 \cdot 0.21)^{1/2} \right] = 32.8 \text{ degrees}$$

The directional gain for this example antenna is

$$\begin{aligned} G &= 41,253 / \theta^2 \\ G &= 41,253 / (32.8 \cdot 32.8) = 383 = 15.8 \text{ dBi} \end{aligned} \quad (13.8)$$

The terminal impedance of a helical antenna in the axial mode is nearly purely resistive. An empirical formula for the input resistance for the helical antenna is given by

$$R_{\text{in}} = 140 C/\lambda \quad (13.9)$$

where the terms are as previously defined. A typical helical antenna thus would have a center frequency input resistance of about 140Ω .

Helical antennas often are used in arrays with four or more antennas mounted on the same ground plane and separated from each other by about a wavelength. These are all fed in phase so that the powers add.

13.7 Spiral Antennas

Frequently it is desirable to have the pattern and the impedance of an antenna remain constant over a wide range of frequencies, such as 10:1 or higher. An antenna of this type is often referred to as a frequency-independent antenna. Actually, there is no such thing as a frequency-independent antenna. They always have a lower frequency limit set by the largest dimensions that are possible or desirable and an upper frequency limit set by the smallest dimensions that are practical. Frequency-independent antennas generally are of two types: spiral antennas and log periodic antennas. This section describes three types of spiral antennas.

13.7.1 Equiangular Spiral Antennas

Figure 13.16(a) illustrates a planar equiangular spiral antenna. The figure shows metal sheets mounted on a dielectric support, which usually is done by printed-circuit techniques. Two arms start at the center and spiral outward. The antenna is fed

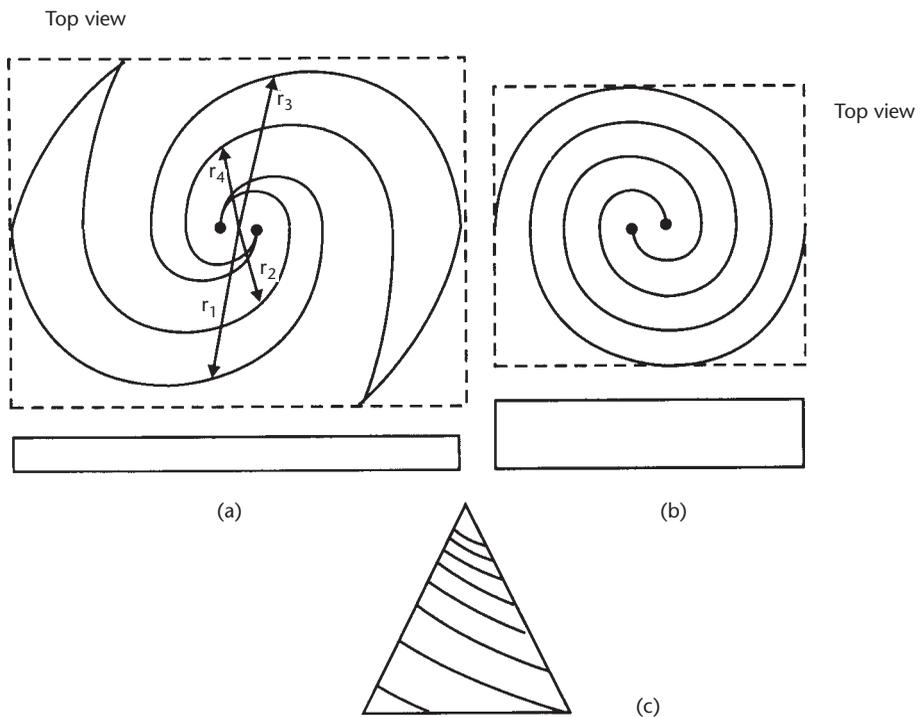


Figure 13.16 Spiral antennas: (a) planar equiangular spiral antenna; (b) Archimedean spiral antenna; and (c) conical equiangular spiral antenna. (After: [6].)

at the center by a coaxial feed line that is wound along one of the two antenna arms toward the feed points. The outer conductor of the coaxial feed line is connected at the feed location to one of the antenna arms, and the center conductor is connected to the other antenna arm.

The four edges of the metal antenna elements each have an equation for their curves. Edge 1 has its radius, r_1 , given by (13.10). Edge 2 has its radius, r_2 , given by (13.11). The other half of the antenna has edges that make the structure symmetric, that is, if one spiral arm were rotated one-half turn, it would coincide with the other arm. Edge 3 thus has its radius, r_3 , given by (13.12), and edge 4 has its radius, r_4 , given by (13.13).

$$r_1 = r_0 e^{a\phi} \quad (13.10)$$

$$r_2 = r_0 e^{a(\phi-\delta)} \quad (13.11)$$

$$r_3 = r_0 e^{a(\phi-\pi)} \quad (13.12)$$

$$r_4 = r_0 e^{a(\phi-\pi-\delta)} \quad (13.13)$$

A typical flare rate, a , is 0.22. The values of δ is $\pi/2$. The value of r_0 is the minimum radius of the antenna.

Spirals of one and one-half turns appear to be optimum. Bandwidths of 8:1 are typical; however, bandwidths as high as 20:1 can be obtained. A typical input impedance for the equiangular spiral antenna in Figure 13.16(a) is about 164Ω . The radiation pattern is bidirectional with two wide beams broadside to the plane of the antenna. The half-power beamwidth is approximately 90 degrees. The polarization of the radiation is close to circular over wide angles, out to as far as 70 degrees from broadside. The sense of the polarization is determined by the sense of the flare of the spiral. The spiral of Figure 13.16 radiates in the right-hand sense for directions out of the page and radiates in the left-hand sense for opposite propagation directions.

13.7.2 Archimedean Spiral Antennas

Figure 13.16(b) shows the Archimedean planar spiral antenna. This antenna is usually constructed using printed circuit techniques. The equations for the two spirals are given by

$$r = r_0 \phi \quad (13.14)$$

$$r = r_0(\phi - \pi) \quad (13.15)$$

The properties of the Archimedean spiral antenna are similar to those of the equiangular planar spiral antenna. It is a circularly polarized antenna. A single main beam can be obtained by placing a cylindrical cavity on one side of the spiral, thus forming a cavity-backed Archimedean spiral antenna. Commercially available antennas of this type have a nearly 90-degree half-power beamwidth, a 2:1 VSWR, a 1.1 axial ratio of polarization on boresight, over a 10:1 bandwidth. The circumference of such an antenna is roughly equal to a wavelength at the lowest frequency

of operation. In spiral antennas, most of the radiation comes from the region of the structure where the circumference is about one wavelength. Thus, as the frequency is changed, a different part of the spiral supports the majority of the current. This feature is responsible for the broadband performance.

13.7.3 Conical Spiral Antennas

Nonplanar types of spiral antennas are possible, for example, the conical equiangular spiral antenna in Figure 13.16(c). With that antenna, a single beam is produced off the tip of the cone. The typical front-to-back ratio of radiation is about 15 dB. A typical antenna has a half-power beam angle of about 80 degrees. The polarization is circular; however, the ellipticity does increase with off-axis angle. The antenna impedance is typically about 165Ω . The beamwidth is controlled by the cone angle and by the length of the cone.

13.8 Log-Periodic Antennas

Log-periodic (LP) antennas make up a second class of so-called frequency-independent antennas. These antennas are not really frequency independent, but they do have very large bandwidth capability. A 10-to-1 frequency range is typical for this class of antennas. This section discusses a number of different types of LP antennas.

13.8.1 Log-Periodic Dipole Array

The LP dipole array (LPDA) is a popular broadband antenna that is simple in construction, low cost, and lightweight. This antenna is shown in Figure 13.17.

The LPDA antenna is a series-fed array of parallel wire or thin rod dipoles of successively increasing lengths outward from the feed point at the apex. In the single plane version of this type of antenna, the interconnecting feed lines cross over between adjacent elements. The enclosed angle, α , bounds the dipole lengths.

One common way to construct an LPDA is to use a two-plane version of the antenna in which the two feed conductors are parallel and closely spaced, with monopole arms alternating in direction. A coaxial transmission line is run through the inside of one of the feed conductors. The outer conductor of the coax is attached to that conductor at the apex, and the inner conductor of the coax is connected to the other conductor of the LPDA transmission line at the apex. This type of construction is illustrated in Figure 13.17(b).

The scale factor, τ , for the LPDA is given by

$$\tau = R_{n+1}/R_n < 1 \quad (13.16)$$

$$\tau = L_{n+1}/L_n \quad (13.17)$$

$$\tau = d_{n+1}/d_n \quad (13.18)$$

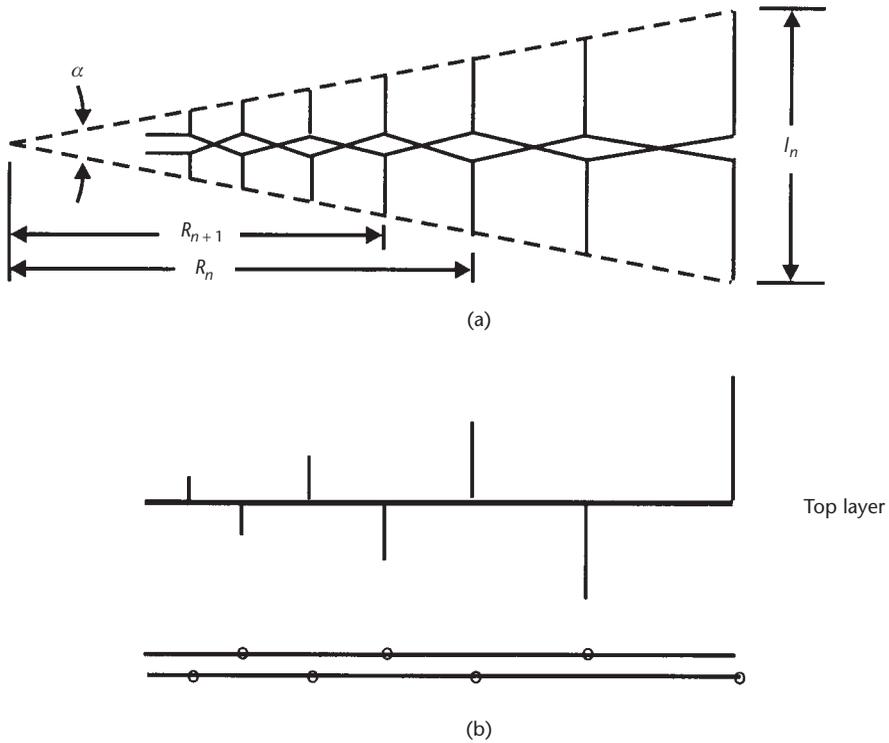


Figure 13.17 LPDAs: (a) LPDA configuration and (b) two-layer construction for LPDA. (After: [7].)

The ratio of successive element positions equals the ratio of successive dipole lengths and the ratio of successive dipole spacings. The spacing factor for the LPDA is defined as

$$\sigma = d_n / 2L_n \tag{13.19}$$

The apex angle, α , is given by

$$\alpha = 2 \tan^{-1} [(1 - \tau) / 4\sigma] \tag{13.20}$$

There is an active region for the LPDA where the few dipoles near the one that is a half-wavelength long support much more current than do the other radiating elements. The longer dipole behind the most active dipole behaves as a reflector, and the shorter dipole in front of the most active dipole acts as a director. The radiation is thus off the apex.

The pattern, gain, and impedance of an LPDA depend on the design parameters τ and σ . Figure 13.18 shows a plot of the antenna gain and optimum values of the scale factor, τ , and the spacing factor, σ . Note that there is an optimum combination for a given antenna gain, as indicated by the gain lines in the plot.

For an example of the use of Figure 13.18, assume the need for an antenna to cover the frequency range 54–216 MHz with a gain of 8.5 dB. From the figure, the

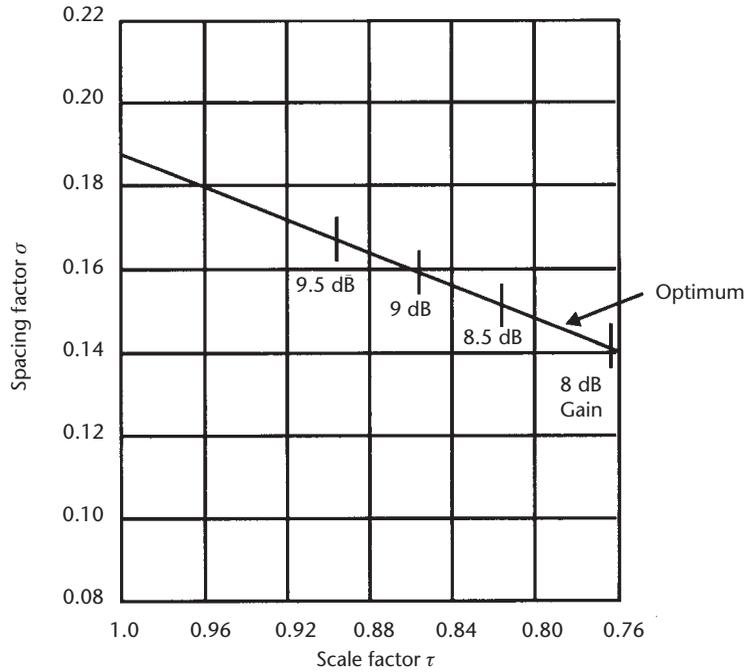


Figure 13.18 Antenna gain and optimum values of the scale factor and the spacing factor for an LPDA. (After: [7].)

optimum value of τ is about 0.822, and the optimum value of σ is about 0.149. Substituting values into (13.20) gives the apex angle, α :

$$\alpha = 2 \tan^{-1}[(1 - 0.822)/4 \cdot 0.149]$$

$$\alpha = 2 \tan^{-1}(0.30) = 33.25 \text{ degrees}$$

The length of the longest dipole is near a half-wavelength at 54 MHz. That would be 2.78m. The shortest radiating dipole should have a length of about a half wavelength at 216 MHz. That would be about 0.69m. The distance from the apex to the longest element is given by

$$R = 139/\tan 16.6 \text{ degrees} = 4.65\text{m}$$

The element lengths are found from the longest element by using (13.17), which is repeated here:

$$\tau = L_{n+1}/L_n$$

$$L_{n+1} = \tau L_n$$

For example,

$$L_2 = (0.822)(2.78) = 2.28\text{m}$$

and

$$L_3 = (0.822)(2.28) = 1.88\text{m}$$

Completing this process yields

$$L_4 = 1.54\text{m}$$

$$L_5 = 1.27\text{m}$$

$$L_6 = 1.04\text{m}$$

$$L_7 = 0.856\text{m}$$

$$L_8 = 0.704\text{m}$$

$$L_9 = 0.578\text{m}$$

$$L_{10} = 0.475\text{m}$$

The element spacing is given by $2\sigma L_n = 0.298 L_n$. Using the above element lengths, the spacings are

$$d_1 = 0.828\text{m}$$

$$d_2 = 0.679\text{m}$$

$$d_3 = 0.560\text{m}$$

$$d_4 = 0.459\text{m}$$

$$d_5 = 0.378\text{m}$$

$$d_6 = 0.310\text{m}$$

$$d_7 = 0.255\text{m}$$

$$d_8 = 0.210\text{m}$$

$$d_9 = 0.172\text{m}$$

13.8.2 Trapezoidal-Toothed Log-Periodic Antennas

Figure 13.19 shows construction details for two types of a trapezoidal-toothed LP antenna. Figure 13.19(a) shows the sheet metal-type antenna, while Figure 13.19(b) is shown the wire-type antenna.

Wedge angles (ψ) range from about 20 degrees to as much as 80 degrees. This sheet-metal version of the trapezoidal-toothed LP antenna often is used at microwave frequencies. This type of antenna is unidirectional, having a single main beam in the direction of the apex. A typical E-plane beam angle is about 65 degrees for apex angles (α) in the range of 20 to 60 degrees. The H-plane beam angle is determined largely by the wedge angle, ψ , and ranges from about 100 degrees, with $\psi = 20$ degrees, to 60 degrees, with $\psi = 90$ degrees. Over this wedge-angle range, the gain of the antenna typically is in the range of 8–10 dB with the smallest gain for the smallest wedge angle. The front-to-back ratio is greatest for small wedge angles. It is about 20 dB for a wedge angle of 30 degrees and about 10 dB for a wedge angle of 60 degrees. The polarization for this antenna is linear and is in the same direction as the teeth of the antenna.

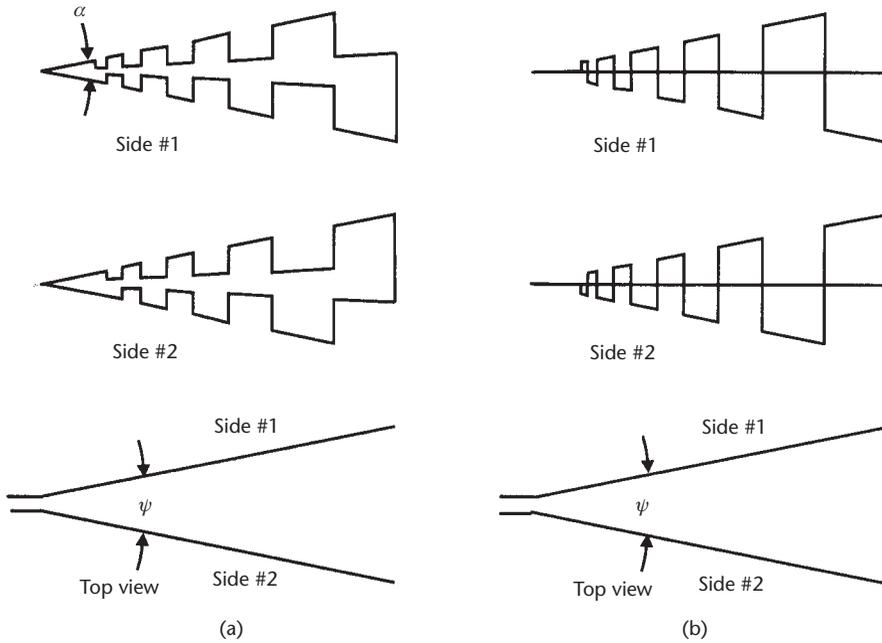


Figure 13.19 Trapezoidal toothed LP antennas: (a) sheet metal-type trapezoidal-toothed LP antenna and (b) wire-type trapezoidal-toothed LP antenna. (After: [7].)

The scale factor, τ , for the trapezoidal-toothed LP antenna is given by (13.21), (13.22), and (13.25), the same as (13.16), (13.17), and (13.18).

$$\tau = R_{n+1}/R_n < 1 \quad (13.21)$$

$$\tau = L_{n+1}/L_n \quad (13.22)$$

$$\tau = d_{n+1}/d_n \quad (13.23)$$

Thus, the ratio of successive element positions equals the ratio of successive element lengths and the ratio of successive element spacings. Because of the way trapezoidal-toothed LP antennas are made, a typical value for τ is much smaller than is the case for LPDA antennas. Typical values used are about 0.65.

There is an active region for the trapezoidal-toothed LP antenna where radiation takes place. That region is where elements are near a quarter-wavelength long. The longer element behind the most active element behaves as a reflector, and the shorter element in front of the most active element acts as a director. The radiation is thus off the apex.

Measurements for a trapezoidal-toothed LP antenna with a scale factor, τ , of 0.65, an apex angle, α , of 60 degrees, and a wedge angle, ψ , of 45 degrees have yielded E- and H-plane half-power beamwidths of 66 degrees, a gain of 9.2 dB, and a front-to-back ratio of 12.3 dB. The average input impedance has been measured as 110Ω with a VSWR of 1.45 over a 10:1 band. The radiation is linear polarized.

The usable bandwidth of the trapezoidal-toothed LP antenna is 10:1 or greater, which is one of the main advantages for this antenna.

One practical way to feed the trapezoidal-toothed LP antenna is to use a coaxial wideband balun that converts a 50Ω coaxial to a 110Ω balanced two-wire line over a distance of greater than a half-wavelength at the lowest frequency of operation. That usually is done with the balun located halfway between the two sides of the wedge and the two-wire line connected to the antenna at the apex.

Figure 13.19(b) shows construction details for a wire version of a trapezoidal-toothed LP antenna. This antenna is similar in many ways to the sheet metal-type trapezoidal-toothed LP antenna. It has the advantages of being lower in weight and with lower wind resistance than the sheet-metal version. This wire version of the trapezoidal-toothed LP antenna often is used at HF, VHF, and UHF frequencies.

Wedge angles (ψ) range from about 20 degrees to as much as 80 degrees. A typical E-plane beam angle is about 65 degrees for α values in the range of 20–60 degrees. The H-plane beam angle is determined largely by the wedge angle, ψ , and ranges from about 100 degrees, with $\psi = 20$ degrees, to 60 degrees, with $\psi = 90$ degrees. Over this wedge-angle range, the gain of the antenna typically is in the range of 8 to 10 dB, with the smallest gain for the smallest wedge angle. The front-to-back ratio is greatest for small wedge angles. It is about 20 dB for a wedge angle of 30 degrees and about 10 dB for a wedge angle of 60 degrees. The polarization for this antenna is linear and is in the same direction as the teeth of the antenna.

As in the case of the sheet-metal version, there is an active region for the wire trapezoidal-toothed LP antenna where radiation takes place. This region is where elements are near a quarter-wavelength long. The longer element behind the most active element behaves as a reflector, and the shorter element in front of the most active element acts as a director. The radiation is thus off the apex.

The performances for the two types of trapezoidal-toothed LP antennas are nearly the same.

13.8.3 Triangular-Toothed Log-Periodic Antennas

Figure 13.20 shows two types of triangular-toothed LP antennas. The performance of these types of antenna is similar to that of the trapezoidal-toothed LP antenna. Again, the typical value of the scale factor, τ , is about 0.65, the typical value of the apex angle, α , is about 60 degrees, and the typical value for the wedge angle, ψ , is about 45 degrees.

Figure 13.20(b) shows the wire version of the triangular-toothed LP antenna. The performance for this type of antenna is similar to that of its corresponding sheet-metal version, and design parameters are as given in the preceding paragraph.

13.9 Slot Antennas

13.9.1 Open-Slot Antennas

Figure 13.21(a) shows a slot antenna in a metallic plane. The slot has a length of about 0.5 wavelength and a width of about 0.1 wavelength. Many of the properties

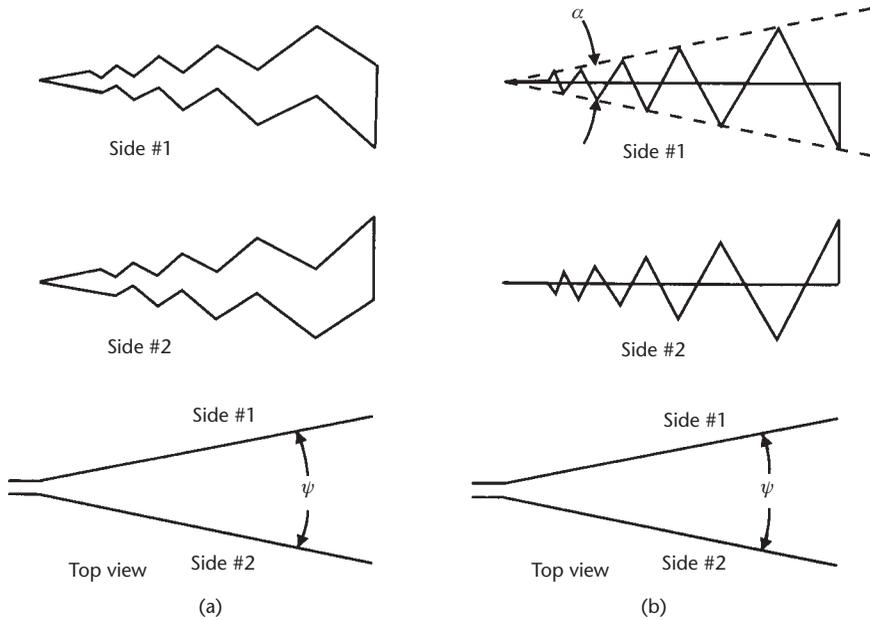


Figure 13.20 Triangular-toothed LP antennas: (a) sheet metal-type triangular-toothed LP antenna and (b) wire-type triangular-toothed LP antenna.

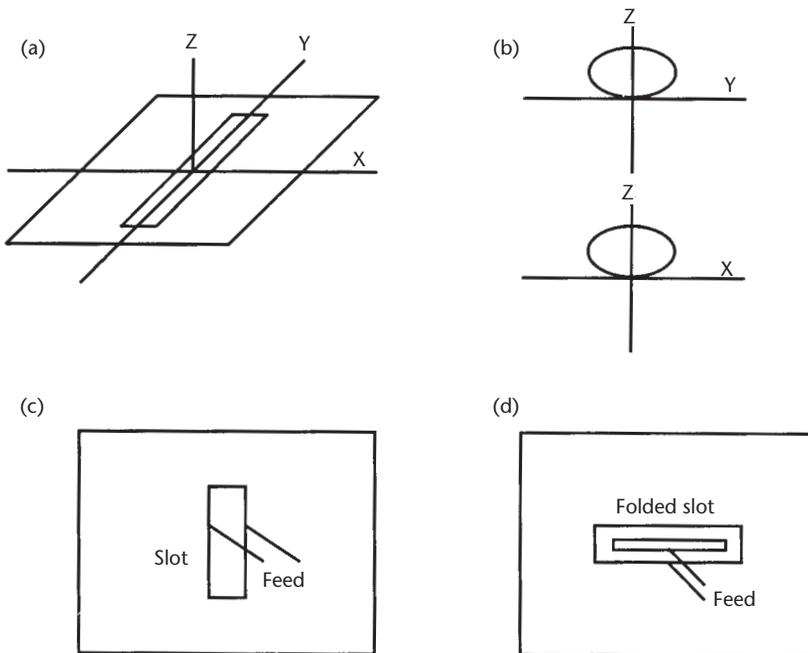


Figure 13.21 Slot antennas: (a) slot antenna in a metallic plane; (b) principal plane field diagrams for slot antenna in a metallic plane; (c) offset feed for lowering slot impedance; and (d) folded slot for lowering slot impedance. (After: [8].)

of the slot antenna can be deduced from the properties of the complementary metallic strip dipole antenna by the use of Babinet's equivalence principle. The E-plane pattern of the slot and the H-plane pattern of the dipole are omnidirectional, while the slot H-plane pattern is the same as the dipole E-plane pattern. The principal plane field diagram for the slot in Figure 13.21(a) is shown in Figure 13.21(b).

Slots can be excited by a transmission line connected across the slot, by an energized cavity placed behind it, or by a waveguide. In the case of feeding by a transmission line, it is usually necessary to provide some type of impedance matching. Resonance occurs at the frequency for which the slot length is near a half-wavelength. The resistance of the center-fed resonant slot antenna is about 363Ω . Two methods for lowering slot resistance are shown in Figure 13.21(c, d).

It is possible to transform the slot impedance by feeding the slot at some point off center, an approach shown in Figure 13.21(c). The closer the feed point is to the side of the slot, the lower the impedance will be. That is just opposite of what we would have with the metallic-strip dipole antenna. In the dipole case, the closer the feed is to the ends of the dipole, the higher the antenna impedance will be.

Figure 13.21(d) shows the case of a folded slot fed by a 75Ω coaxial line. By folding the slot, the impedance of the slot is reduced by a factor of 4, from about 363Ω to about 90Ω . Again, that is just opposite of the case of a folded dipole, where folding increases the impedance by a factor of 4.

13.9.2 Cavity-Backed Rectangular-Slot Antennas

Cavity-backed slot antennas often are used where surface-mounted antennas are needed with radiation in only one hemisphere. Such an antenna is shown in Figure 13.22(a).

One way to achieve a fairly wideband cavity-backed rectangular-slot antenna is to use a T-fed slot antenna, like the one shown in Figure 13.22. The input to the antenna is by means of a coaxial cable. The center conductor of this line is connected to the T-shaped element located in the bottom of the cavity. The length of the slot is about 0.62 wavelength, the width about 0.20 wavelength, and the depth of the slot about 0.20 wavelength. The diameter of the T-feed element, positioned in the center of the cavity, is about 0.073 wavelength. Measurements show that this type of antenna has a VSWR less than 2:1 over a frequency range of about an octave. Other types of wideband feed systems also are used with cavity-backed slot antennas.

Figure 13.22(b) shows a typical antenna pattern for a cavity-backed slot antenna on a ground plane. The larger the ground plane, the wider is the pattern in the xz -plane. The pattern is narrower in the yz -plane than in the xz -plane.

13.9.3 Waveguide-Fed Slot Antennas

Many radar and other microwave systems use waveguide-fed slot antennas. These systems often involve arrays of many slot antennas. Systems of this kind are shown in Figure 13.23.

Figure 13.23(a) shows a section of rectangular waveguide with longitudinal slots in the top wall of the waveguide. The slots are positioned at points one-half

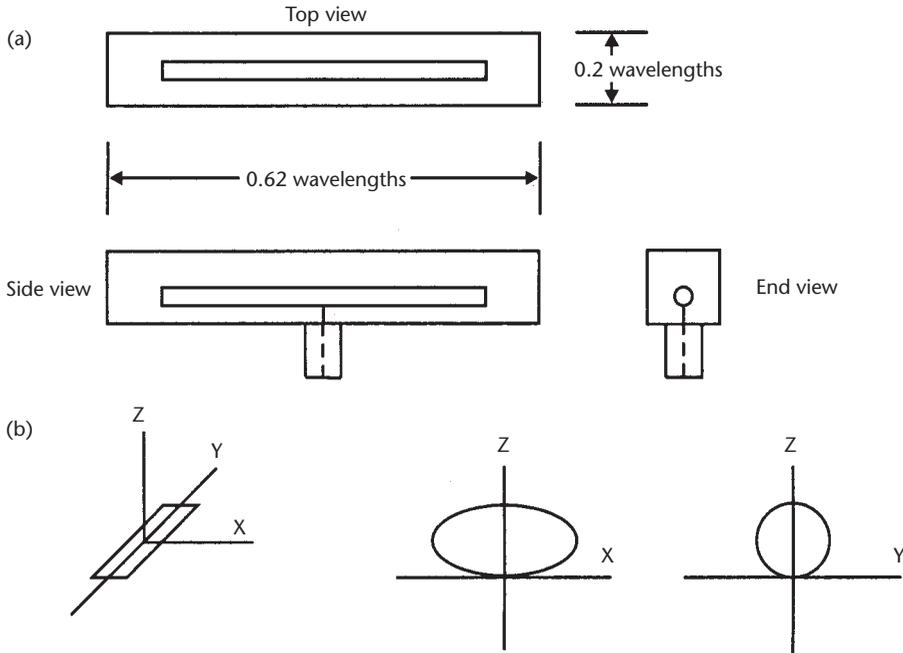


Figure 13.22 Cavity-backed rectangular slot antenna with T-feed: (a) antenna configuration and (b) antenna patterns. (After: [8].)

guide wavelength along the guide, with alternate slots on either side of the center line. With the slots positioned in this manner, the slots are in phase as needed for the array. The longitudinal slots are in position to interrupt current flow for the dominant TE_{10} mode and are, therefore, good radiators. The slots are each a half-wavelength long. The selection of the distance from the center line often is done experimentally so the desired radiation is provided by all slots.

Figure 13.23(b) shows a second type of waveguide-fed slot array antenna, one that has slots on the side wall of the waveguide. Each slot is slanted with respect to the vertical. Again, they are positioned so they are one-half guide wavelength apart. Every other slot is in a different direction, so the slots are in phase. Each of these slots is a good radiator because each blocks current flow in the side wall of the waveguide. The larger the angle between the slot and the vertical, the larger the radiation is. For small angles, it is necessary for the slot to cover not only the side wall but also part of the top and bottom walls if it is to be a half-wavelength slot. That is because the side wall of the guide normally is much less than a half-wavelength high.

The polarization for the slots is at right angles to the long axis of the slots. With small slant angles from vertical, as is normally the case, there is considerable mutual coupling between the slots, which must be taken into account. The slant angle that will be used for each slot in an array usually is selected experimentally so the desired radiation is provided by all slots.

The choice of whether to use a side-wall or a top-wall slot antenna is based mainly on the ease of fabrication for the type of system used.

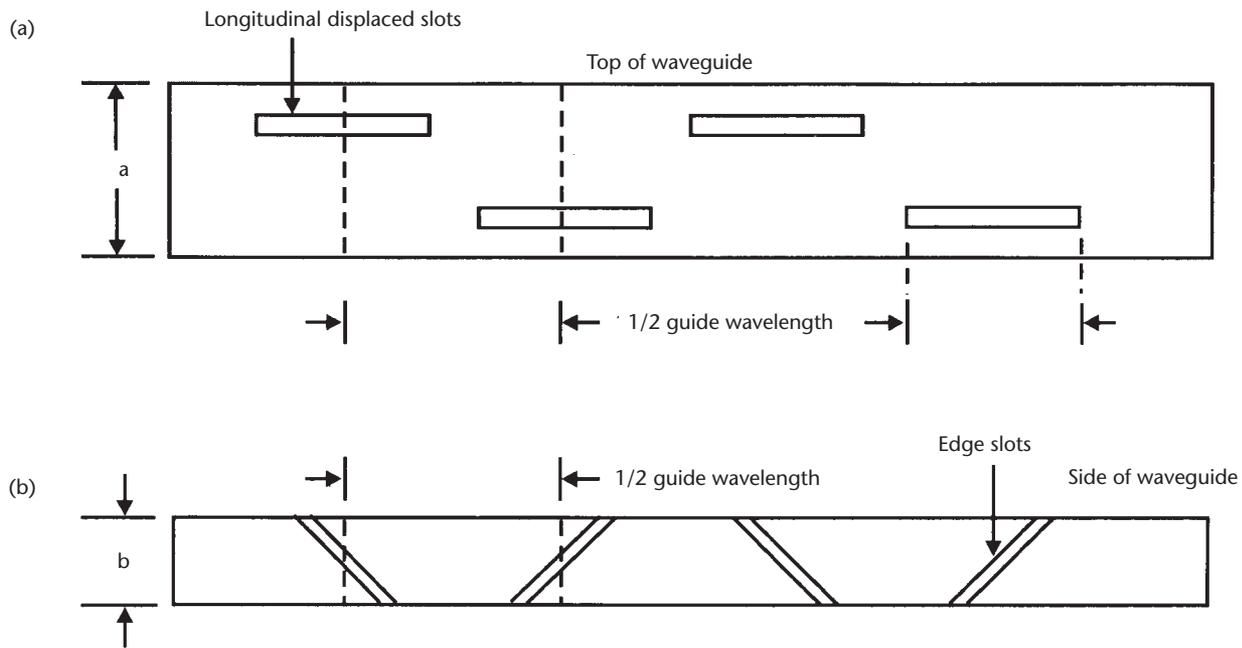


Figure 13.23 Waveguide-fed slot antenna array: (a) top-wall waveguide-fed slot antenna array and (b) side-wall waveguide-fed slot antenna array.

13.10 Notch Antennas

A notch antenna, shown in Figure 13.24, is a broadband radiator that is often used with aircraft. For example, the edge of a wing or a rudder can be used for a notch antenna.

Figure 13.24(a) shows the notch antenna configuration. The depth of the notch is a quarter-wavelength. The feed points are across the notch. The exact impedance depends on the distance that the feed is placed from the edge of the ground plane. It is greatest at the edge and decreases as we approach the end of the notch.

Figure 13.24(b) shows the radiation patterns for a notch antenna where the ground plane is large with respect to a wavelength.

13.11 Horn Antennas

Horn antennas are one of the more important types of antennas for use with waveguides. Horn antennas often are used as feeds for parabolic dish antennas. They are used separately where the gain requirements are not too high and also as standards in measuring performance of other antennas.

Figure 13.25 show three types of rectangular horn antennas.

The pyramidal horn and the conical horn emit pencil-like beams that have high directivity in both horizontal and vertical planes. Fan-shaped beams result when one

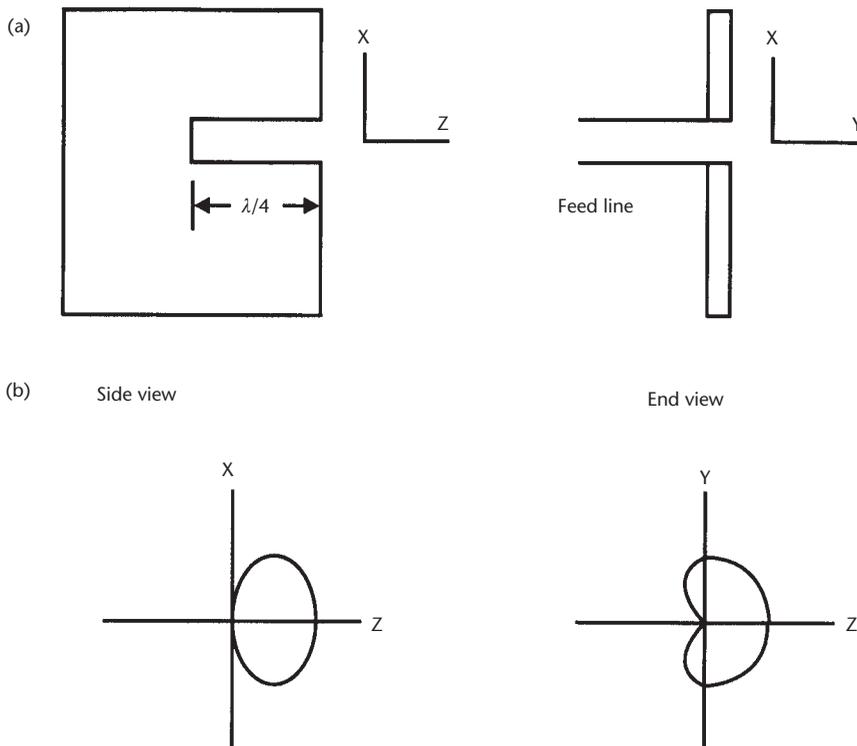


Figure 13.24 Notch antenna configuration and antenna patterns: (a) antenna configuration and (b) antenna radiation patterns. (After: [8].)

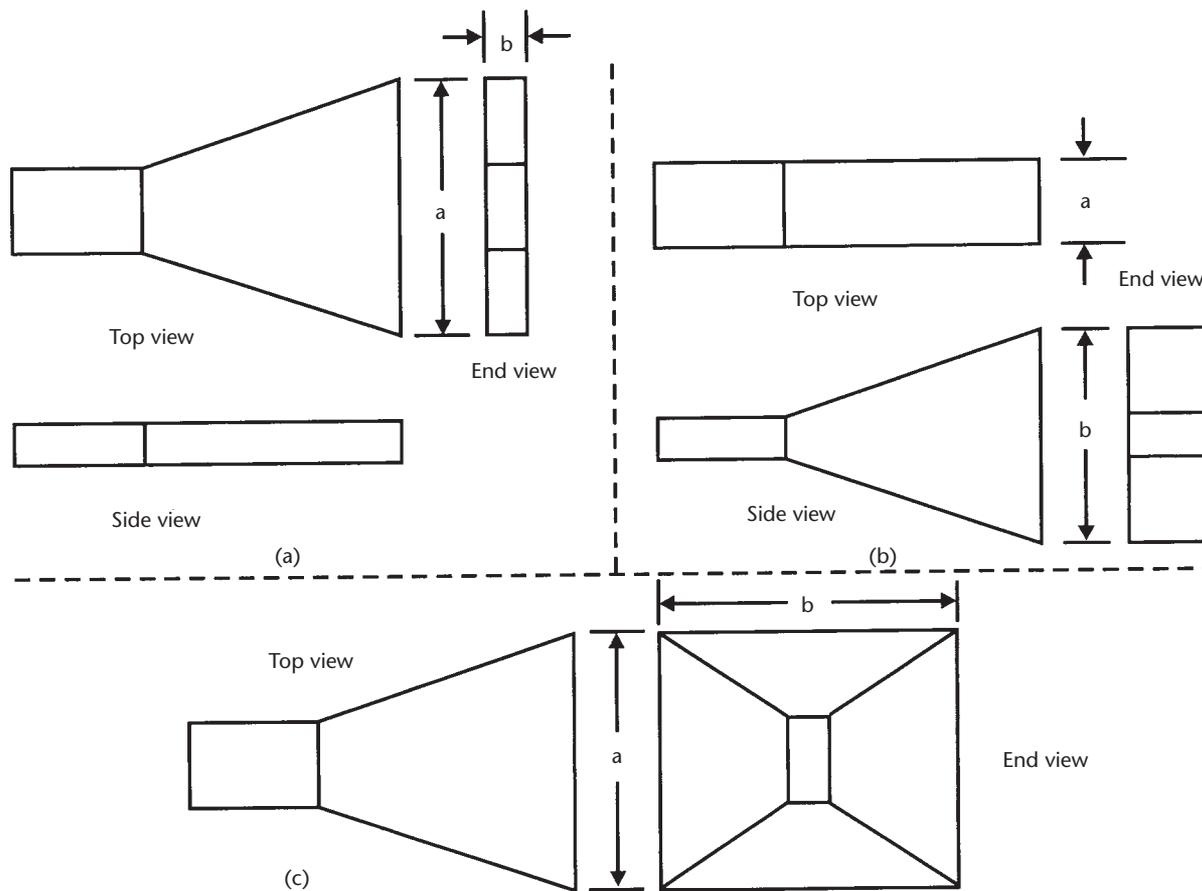


Figure 13.25 Three types of rectangular horn antennas: (a) H-plane sectoral horn; (b) E-plane sectoral horn; and (c) pyramidal horn.

dimension of the horn mouth is much smaller than the other. The sectoral horns formed by flaring in only one dimension exhibit this behavior.

In Figure 13.25, the dimension of the long width of the rectangular horn is a , and the dimension of the short width is b , just as in the case of the rectangular waveguide. Not labeled in the figure, the horn length from mouth to apex is L . The diameter of the mouth of the conical horn is d . Formulas for beam widths and power gains applicable to optimum horns of various types are given in Table 13.1. Directive gain is given by

$$\text{Directive gain} = 4\pi A_0 k / \lambda^2 \tag{13.24}$$

For an example of the use of Table 13.1, assume the design of a pyramidal horn in which the property that is optimized is the gain. Further assume a wavelength of 10 cm and a length, L , of 30 cm. The required values for dimensions a and b are

$$a = (3L\lambda)^{0.5} = (3 \cdot 30 \cdot 10)^{0.5} = 30 \text{ cm}$$

$$b = 0.81a = 0.81 \cdot 30 = 24.3 \text{ cm}$$

The value of A_0 is

$$a \cdot b = 30 \cdot 24.3 = 729 \text{ cm}^2$$

The half-power beam widths and the directional gain are:

$$\text{H-plane beam width} = 80\lambda/a = 80 \cdot 10/30 = 26.7 \text{ degrees}$$

$$\text{E-plane beam width} = 53\lambda/b = 53 \cdot 10/24.3 = 21.8 \text{ degrees}$$

$$\text{Directive gain} = 4\pi A_0 k / \lambda^2 = 4 \cdot \pi \cdot 729 \cdot 0.50 / 100 = 45.8 = 16.6 \text{ dB}$$

13.12 Lens Antennas

Figure 13.26 shows three types of lens antennas: dielectric lens antenna, Luneburg lens antenna, and a metal-plate lens antenna.

Table 13.1 Formulas for Optimum Horns

Horn Type	Optimized Property	Optimum Properties	HP Beam Widths (degrees)		Values of k for (13.24)
			H-Plane	E-Plane	
Pyramidal	Gain	$a = (3L\lambda)^{0.5}$ $b = 0.81a$ $G = 15.3 L/\lambda$	$80\lambda/a$	$53\lambda/b$	0.50
H-plane sectoral	Beam width in H-plane	$a = (3L\lambda)^{0.5}$	$80\lambda/a$	$51\lambda/b$	0.63
E-plane sectoral	Beam width in E-plane	$a = (2L\lambda)^{0.5}$	$68\lambda/a$	$53\lambda/b$	0.65
Conical	Gain	$a = (2.8L\lambda)^{0.5}$	$70\lambda/d$	$60\lambda/d$	0.52

Source: [9].

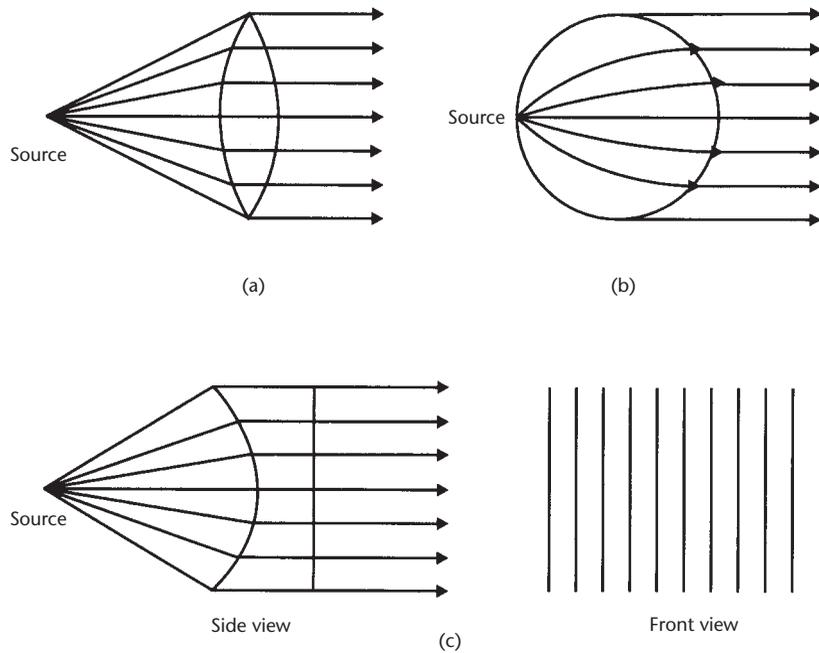


Figure 13.26 Lens antennas: (a) dielectric lens antenna; (b) Luneburg lens antenna; and (c) metal-plate lens antenna.

13.12.1 Dielectric Lens Antenna

Figure 13.26(a) shows the design and operation of a dielectric lens antenna. The objective of the dielectric lens is to convert a spherical wavefront from a point source such as a horn to a planar wavefront. To do that, the dielectric lens is made thick near the center and progressively thinner toward the edges of the lens. The speed of the RF wave is less in the dielectric than in air, being reduced by the square root of the relative permittivity of the dielectric. With the correct thickness, the total delay for the wave in traveling from the point source to the back of the lens is the same over all parts of the lens. The result is a planar wavefront.

The gain and the beam angle for a dielectric lens antenna depend on the area of the lens and the wavelength. The antenna gain is given by (13.24), with k equal to approximately 0.65, and A_0 is the area of the lens.

While the dielectric lens can be made with ordinary dielectric, it is possible to use a much lower weight artificial dielectric that consists of conducting small rods or spheres embedded in a low-weight dielectric such as styrofoam.

13.12.2 Luneburg Lens Antenna

A special case of a dielectric lens is a Luneburg lens, which is shown in Figure 13.26(b). A Luneburg lens is spherical in shape. The index of refraction (the square root of the dielectric constant) is a maximum at the center of the sphere, where it is equal to 1.414 and decreases to a value of 1.0 on the periphery. A practical Luneburg lens can be constructed from a large number of spherical shells, each of constant index of refraction. Discrete changes in index of refraction approximate a

continuous variation. In one example of a Luneburg lens, 10 concentric spherical shells are arranged one within the other. The dielectric constant of the individual shells varies from 1.1 to 2.0 in increments of 0.1. The Luneburg lens focuses on a point source on the periphery of the sphere. It is possible to use many such point sources for receiver antennas simultaneously with a single Luneburg lens.

13.12.3 Metallic-Plate Lens Antenna

Figure 13.26(c) shows the design and operation of a metal-plate lens. The objective of this lens is to convert a spherical wavefront from a point source such as a horn to a planar wavefront. To do that, the metal-plate lens is thin near the center and progressively thicker toward the edges of the lens. The phase velocity of the RF wave is greater in a metal-plate lens than in air. By selection of the correct thickness, the total delay for the wave in traveling from the point source to the back of the lens is the same over all parts of the lens. The result is a planar wavefront.

The metal-plate lens consists of conducting strips that are placed parallel to the electric field of the wave that approaches the lens and are spaced slightly in excess of a half wavelength. To the incident wave, such a structure behaves like a large number of waveguides in parallel.

One weakness of this type of lens antenna is that its focusing action is frequency sensitive. That is because the phase velocity in a waveguide depends on the frequency. The gain and the beam angle for a metal-plate antenna depend on the area of the lens and the wavelength.

13.13 Antenna Arrays

An antenna array is made up of a number of individual antennas, such as monopoles, dipoles, helical antennas, slots, and open-ended waveguide. These are positioned and phased in such a way as to provide a higher gain antenna of the desired beam shape and other desired characteristics.

13.13.1 End-Fire Line Antenna Arrays

Figure 13.27(a) illustrates a two-element end-fire line array. The array might be made up of two monopole antennas, each with omnidirectional azimuth patterns. These antenna elements are spaced one-quarter wavelength apart and are fed 90 degrees out of phase. If the phase of the right-hand element lags the phase of the left-hand element, the two radiated signals will be in phase, and their powers will add in the right-hand direction, as shown in the antenna pattern for the array. In the left-hand direction, the two radiated signals will be 180 degrees out of phase. If their amplitudes are equal, their fields will add to zero. At other azimuth angles, the resulting radiated power will be the vector addition of the two signals. The resulting pattern shown in Figure 13.27(b) is known as a cardioid pattern.

The typical maximum spacing of elements is $3\lambda/8$ for end-fire line arrays where many antenna elements are used. There is a progressive phase difference between

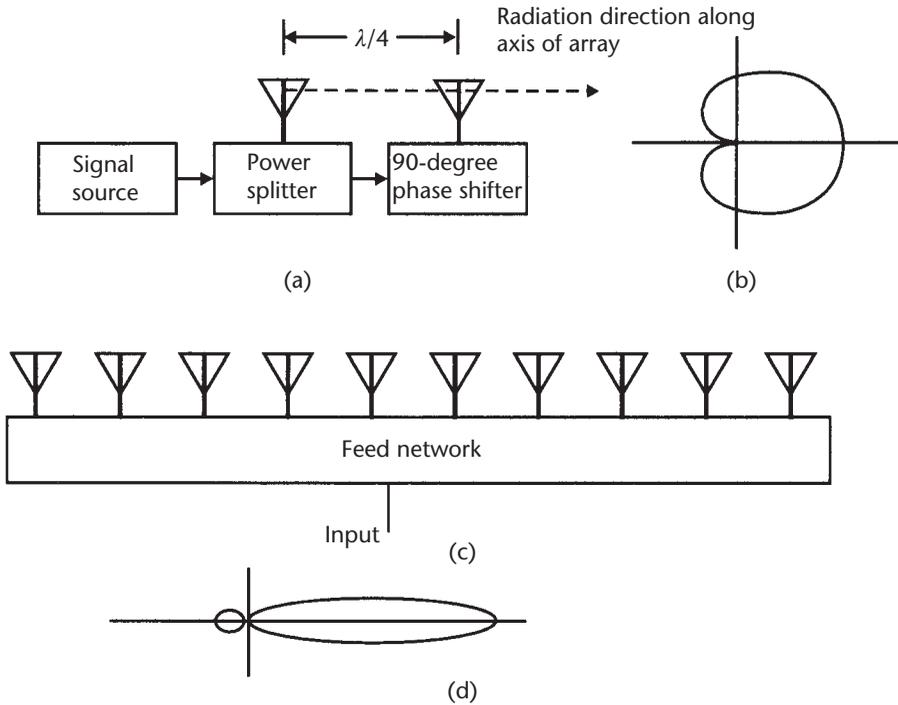


Figure 13.27 End-fire line array antennas: (a) simple two-element line array antenna; (b) azimuth antenna pattern for the two-element line array antenna; (c) example of 10-element end-fire line array; and (d) azimuth antenna pattern for 10-element end-fire line array.

adjacent antennas equal in cycles to the spacing between the antennas, expressed in wavelengths.

Directivity of the individual antennas of an end-fire array is taken into account by multiplying the array pattern obtained by postulating isotropic radiators by the actual directional characteristic of the antennas used. For example, assume that the array gain using isotropic radiators in the main beam is 10 and the element gain in that direction is 3. The total antenna gain in the main beam, then, is 30 (14.8 dB).

The following equations apply to the end-fire line array with isotropic elements, where L is the length of the array between the centers of the end antennas:

Width of the major lobe between nulls is

$$\text{Width} = 115 / (L/2\lambda)^{0.5} \text{ degrees} \tag{13.25}$$

$$\text{Directive gain (numeric)} = 4L/\lambda$$

For an example of the use of (13.25) and (13.26), assume the use of 20 isotropic antennas in a line spaced by $3\lambda/8$. The value of L is 7.125λ . Substituting values into (13.25) and (13.26) gives

$$\text{Width} = 115 / (7.13/2)^{0.5} = 60.9 \text{ degrees} \tag{13.26}$$

$$\text{Directive gain} = 4 \cdot 7.13 = 28.5 = 14.5 \text{ dB}$$

If each antenna element of the array had a gain in the desired direction of 3 dB, the total directive gain would be 17.5 dB. Figure 13.27(c) illustrates that array, while the resulting end-fire azimuth antenna pattern is shown in Figure 13.27(d).

13.13.2 Broadside Line Antenna Arrays

Figure 13.28(a) shows a broadside line antenna array. The typical antenna element spacing is $\lambda/2$ and all elements are fed in phase. If the elements are assumed to radiate uniformly in all directions, the azimuth antenna pattern might be as shown in Figure 13.28(b).

Equations (13.27), (13.28), and (13.29) apply to the broadside line arrays with isotropic elements, where L is the length of the array between the centers of the end antennas.

The width of the major lobe between nulls is given by

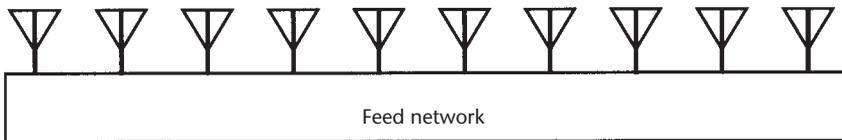
$$w_1 = 115/(L/\lambda) \text{ degrees} \quad (13.27)$$

The width of the major lobe between half-power points is given by

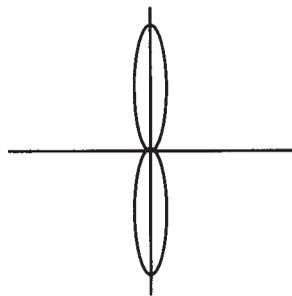
$$w_2 = 51/(L/\lambda) \text{ degrees} \quad (13.28)$$

The directive gain (numeric) is given by

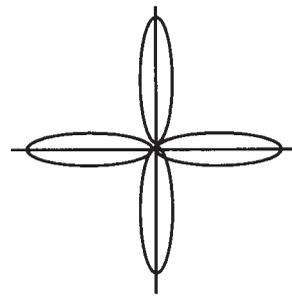
$$G_d = 2L/\lambda \quad (13.29)$$



(a)



(b)



(c)

Figure 13.28 Broadside line arrays: (a) antenna configuration; (b) azimuth antenna pattern for element spacing of one-half wavelengths and isotropic radiation by elements; and (c) azimuth antenna pattern for element spacing of one wavelength and isotropic radiation by elements.

For an example of the use of (13.27), (13.28), and (13.29), assume the use of 41 isotropic antennas in a line spaced by $\lambda/2$. The value of L is 20λ . Substituting values into (13.27) gives the width of the major lobe between nulls:

$$w_1 = 115/20 = 5.75 \text{ degrees}$$

Substituting values into (13.28) gives the width of the major lobe between half-power points:

$$w_2 = 51/20 = 2.55 \text{ degrees}$$

Substituting values into (13.29) gives the directive gain for isotropic radiators:

$$G_d = 2 \cdot 40 = 80 = 19 \text{ dB}$$

If the antenna elements are directional antennas with a gain of 3 dB in the desired broadside direction, the directive gain would total 22 dB.

Figure 13.28(c) shows the case of a line antenna with elements spaced by one wavelength. With that spacing, there will be major lobes in line with the antenna as well as broadside to the antenna. These lobes are called grating lobes. As the spacing is increased beyond one wavelength, the grating lobes occur at smaller angles from the main lobe. The angle for grating lobes is given by

$$\sin \theta = \pm n\lambda/d \quad (13.30)$$

where:

θ = the angle from broadside

n = the number of the grating lobe

λ = wavelength

d = element spacing

For an example of the use of (13.30), assume the following:

$n = 1$ (i.e., the first lobe)

$d = 2\lambda$

$\sin \theta = \pm n\lambda/d$

$\sin \theta = \pm 1 \cdot \lambda/2\lambda = \pm 0.5$

$\theta = \pm 30 \text{ degrees}$

when $d = \lambda$, $\sin \theta = \pm 1$, $\theta = \pm 90 \text{ degrees}$, which is the case shown in Figure 13.28(c).

Sometimes high-gain elements, such as helical antennas, are used in line arrays. In those cases, the spacing often is more than a wavelength and grating lobes will exist. They may not be of large amplitude, however, if the element pattern has low gain at those angles where there are grating lobes. Again, the antenna pattern will be the product of the array pattern and the element pattern.

13.14 Planar Arrays

Figure 13.29(a) shows an example of a planar antenna array. The antenna pattern for the array is shown in Figure 13.29(b). This type of array is a two-dimensional array with a “height” and a “width.” The example array is an 8×4 element array, which has a total of 32 elements. This might be viewed as four line arrays, with each line array having eight elements. Each element in the line array typically is spaced by $\lambda/2$, and each line array is spaced by $\lambda/2$.

Examples of antenna elements for the planar array are dipoles with reflector surfaces behind the dipoles, waveguides with slots, and open-ended waveguides. With dipoles having reflectors, the gain for each element may be about 4 dB. With 32 elements, the gain of the planar array would 19 dB.

A 100×100 element array using 4-dB gain elements would have a gain of about 44 dB and a beam angle of 1×1 degree. This antenna can be used as a reference in estimating the number of elements in a given direction needed for a given azimuth or elevation beam angle. This requirement is given by

$$n = 100/\theta \quad (13.31)$$

where θ is the desired azimuth or elevation -3 -dB beamwidth in degrees.

For example, a 2-degree azimuth \times 20-degree elevation fan-shaped beam would require $100/2 = 50$ antenna elements in the horizontal direction by $100/20 = 5$ elements in the vertical direction, or a total of $50 \cdot 5 = 250$ antenna elements.

13.15 Scanning Methods

13.15.1 Mechanically Scanned Arrays

There are a number of ways to point the high-gain small-width beams of array antennas. One way is to point the beam by mechanically positioning the array in the desired direction. In the case of a scanning system, the array can be rotated at the desired rate by mechanical means, provided that the desired rotational rate is not

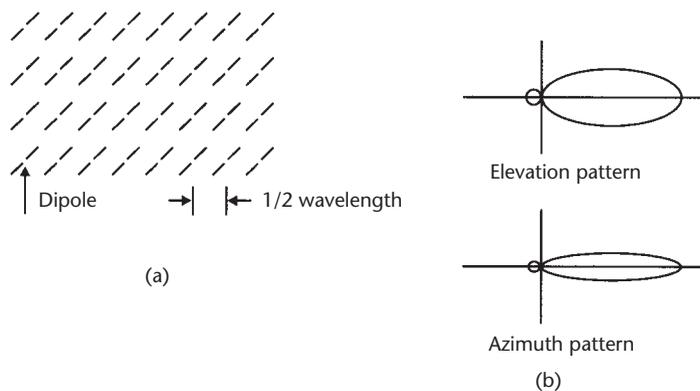


Figure 13.29 Planar array antenna: (a) antenna configuration and (b) antenna patterns.

too high. Maximum possible rotation rates depend on the size of the antenna. A typical maximum rate for a microwave antenna array might be on the order of one rotation per second.

Combinations often are used in which the array antenna is scanned or moved in one direction by mechanical means and in the other direction by either phase scan or frequency scan. These latter two scanning methods have the advantage of very rapid scan and positioning of the beams.

13.15.1.1 Phase-Scanned Arrays

Figure 13.30 shows an example of a phase-scanned line array. This system uses four antennas, with each antenna input having a variable phase shifter. The phase shifters are fed with equal amplitude signals of the same phase. This type of feed system is known as a corporate feed. Other possible feed systems include an end-fed series phase-shift arrangement and a center-fed series phase-shift arrangement.

Figure 13.30 also shows how the beam angle can be changed by a change in the phase angle of the individual phase shifters. When the phase shift is progressively larger from left to right, the beam is shifted to the left. When the phase shift is progressively larger from right to left, the beam is shifted to the right. The wave front shown in Figure 13.30 is the line where the signals from each antenna have the same phase. The beam angle is normal to that line.

A planar array with phase-shift volumetric scan in two angular coordinates can be used for some military applications. Such a system requires a phase-shifter for each antenna and is costly because of the large numbers of antennas and phase shifters involved.

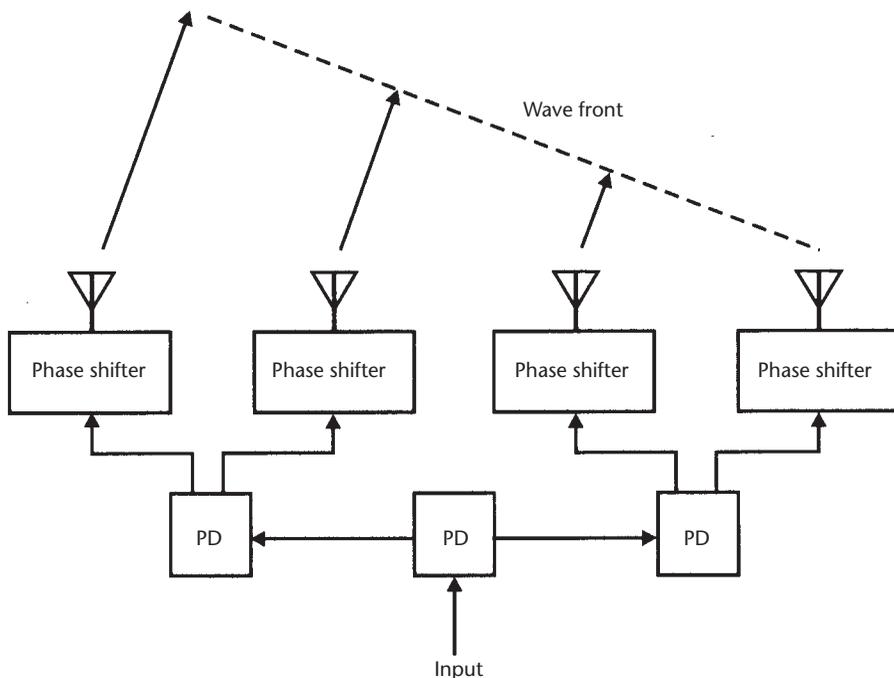


Figure 13.30 Phase-scanned line array antenna.

$$\sin \theta = (c/08c)(6.4 d/d)[1 - (f_0/1.1f_0)]$$

$$\sin \theta = 12.5 \cdot 6.4(1 - 0.91) = 0.72$$

$$\theta = 46.05 \text{ degrees}$$

It is possible with a frequency-scanned array to use a number of frequencies simultaneously to provide multiple beams. That can be done to permit a higher scan rate than would be possible with a single-beam antenna. The U.S. Navy AN/SPS-48 radar, which is used on many naval vessels, uses a mechanically and frequency-scanned antenna of this type.

Frequency scanning can also be used in a planar array. In one type of system, the horizontal line arrays are frequency-scanned arrays that use delay lines between elements. Many of these lines are stacked one above another. Each line has a phase shifter in the input to the line, and all lines are fed in parallel using suitable power dividers. Frequency change is used to steer the beam in the azimuth direction, and phase change is used to steer the beam in elevation. This type of antenna is sometimes called a phase-frequency array. Another type of system uses frequency scanning in one direction and mechanical scanning in the other direction.

13.15.2 Arrays with Space Feeds

The feed systems for antenna arrays can be complex and costly. They usually involve many sections of coaxial line or waveguide and many coax or waveguide power dividers and combiners. An alternative method is to use so-called space feeds.

Figure 13.32 shows two examples of arrays with space feeds. Figure 13.32(a) is a lens array. With this system, a primary feed antenna, such as a horn, is used to

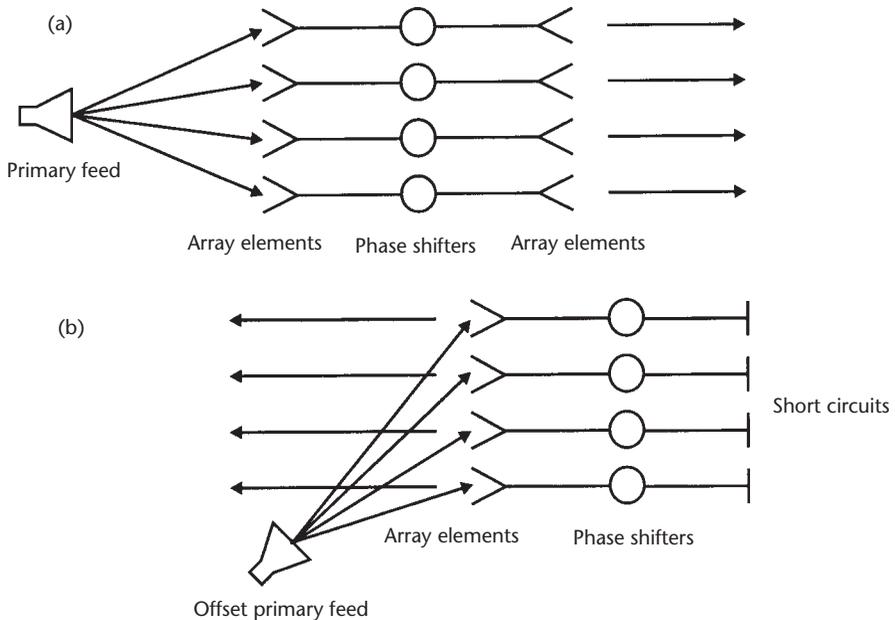


Figure 13.32 Arrays with space feeds: (a) lens array and (b) reflectarray.

transmit the signal to an array of small receiver horns, each connected to transmission lines containing phase shifters. The output of the phase shifters is sent to an array of transmitting horn antennas. That array can be either a line array or a two-dimensional planar array. The beam is pointed by adjusting the phase of all the phase shifters.

Figure 13.32(b) shows a so-called reflectarray antenna, in which an offset primary feed horn is used to direct energy to the array of receive horns. The output of the receive horns is sent to transmission lines that include phase shifters and short circuits for reflection. The reflected waves are sent back through the phase shifters and thence to the same array of horns that now act as transmit horns. The beam angle is thus pointed by means of the phase shifters.

13.16 Flat-Plate Reflector Type Antennas

Many types of reflector antennas are used with RF systems. Two of these are discussed here.

13.16.1 Half-Wave Dipole Antennas with Reflectors

When a reflector such as a copper sheet is placed close to a half-wave dipole antenna, a unidirectional radiation pattern is obtained with an antenna gain of about 4 dB. With a quarter-wave separation between the dipole and the reflector and with other dipoles spaced by about a half-wavelength, the impedance of the dipole is about 153Ω . That is quite different from the impedance of a single thin dipole in free space, which has an impedance at resonance of about 73Ω .

13.16.2 Corner Reflector Antennas

The directivity of an antenna-reflector combination can be increased by bending the reflector to form a corner. A directional gain of about 12 dB is obtained from a dipole antenna with a 90-degree corner and the feed spaced from the corner by 0.25λ . A 60-degree corner reflector can provide a directive gain of about 14.5 dB. The dimensions of the reflector should be about one wavelength by one wavelength.

13.17 Parabolic Reflector Antennas

Parabolic reflectors are the most commonly used high-gain antennas for microwave frequencies. Figure 13.33 shows a parabolic reflector antenna with a circular aperture. When used for transmitting, the signal is radiated from a horn or other feed located at the focus point, as shown. When used for receiving, the parabolic dish collimates a planar wavefront to the focal point, at which a horn antenna is located.

Spherical wavefronts are shown before reflection as the wave travels to the reflector surface. The surface is so designed that it converts the spherical wavefront signal to a planar wavefront signal. That requires that the distance from the focus to

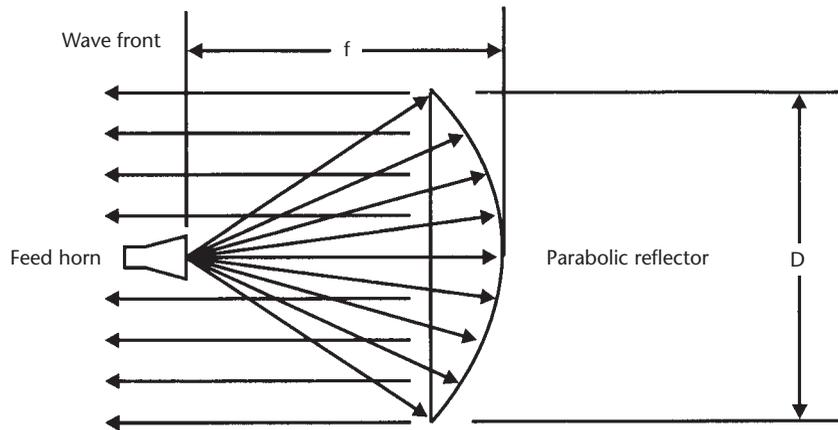


Figure 13.33 Parabolic reflector with a circular aperture and a front feed. (After: [10].)

the reflector and then to the planar wave front be the same regardless of position of reflection.

The width and the shape of the major lobe of the radiation pattern of a parabola depend on the size and shape of the mouth of the parabola and the variation of field intensity over the aperture defined by the mouth. If the shape of the mouth is circular and the field distribution across the mouth is uniform, the width of the main lobe between half-power points is $58 \text{ degrees}/(D/\lambda)$, where D is the diameter of the circular mouth. The width of the main lobe between nulls is $140 \text{ degrees}/(D/\lambda)$. If the shape of the mouth is rectangular and the field distribution is uniform, the width of the main lobe between half-power points is $51 \text{ degrees}/(D/\lambda)$, and the width between nulls is $115 \text{ degrees}/(D/\lambda)$. If the field distribution for the rectangular aperture is sinusoidal along D , the width of the main lobe between half-power points is $68/(D/\lambda)$, and the width between nulls is $182.5/(D/\lambda)$.

The field pattern of a parabolic antenna ordinarily possesses minor sidelobes, just like other types of antennas. These sidelobes can be minimized by tapering the field distribution across the aperture of the parabola so that the field intensity is maximum at the center and minimum at the edges.

The directive gain of the parabolic reflector antenna is given by (13.33), where A_0 is the actual area of the mouth of the antenna, and k is a correction factor that accounts for the nonuniform distribution of energy across the aperture due to tapering of the field. In a practical system, the value of k is on the order of 0.5 to 0.7.

$$\text{Directive gain} = 4\pi A_0 k / \lambda^2 \quad (13.33)$$

Feed systems for parabolic reflectors include horn antennas, dipole plus reflector antennas, Yagi antennas, and LP antennas. One important class of feeds is a monopulse feed system that consists of a group of horn antennas rather than a single horn. Such an arrangement allows a radar to measure more precisely the angular position of a target in azimuth and elevation on a single pulse. When transmitting, all four antennas radiate in a sum mode. When receiving, a sum channel is formed using all four antennas, an azimuth difference channel is formed, and an elevation difference channel is formed.

In many applications, only a section of a parabola is used, a situation sometimes referred to as a truncated paraboloid. When this shape is used, the resulting beam is a fan-shaped beam rather than a pencil beam as provided by a circular aperture.

Figure 13.34(a) shows a parabolic reflector antenna with an offset feed. This is a very important type of antenna system for communication systems as well as for radar. This type of antenna is normally used for a fan shaped beam.

Figure 13.34(b) shows a so-called *hohorn antenna*, which frequently is used in microwave relay systems. It consists of a parabolic reflector joined by a pyramidal horn.

Figure 13.35 shows a parabolic reflector with a Cassegrain feed. The primary feed is located behind the parabolic reflector. The signal from the feed is reflected by a hyperbolic subreflector; it then travels to the parabolic reflector, where it is reflected as a plane wave. This type of antenna has the advantage of permitting very short feed lines between the transmitter or receiver and the antenna. The advantage of short feed lines is reduced line loss, which improves the effective radiated power in a transmitter and reduces the system noise figure in a receiver. That is important for high-microwave and millimeter-wave systems, where the attenuation in waveguide can be large.

The use of Cassegrain feeds also permits a system with a much larger focal length than is possible with a front-feed system. Large focal length is desirable when dealing with short pulses. A disadvantage of this type of feed is that the subreflector causes aperture blocking, which reduces the efficiency of the antenna and increases the sidelobes.

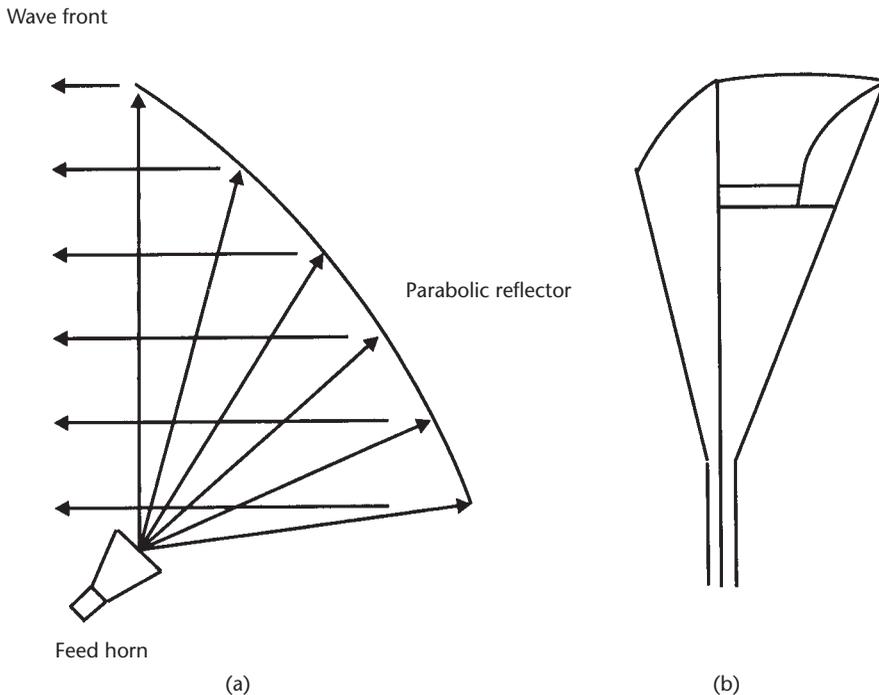


Figure 13.34 Parabolic reflector antenna with an offset feed: (a) system using a small horn feed and (b) hohorn antenna. (After: [2].)

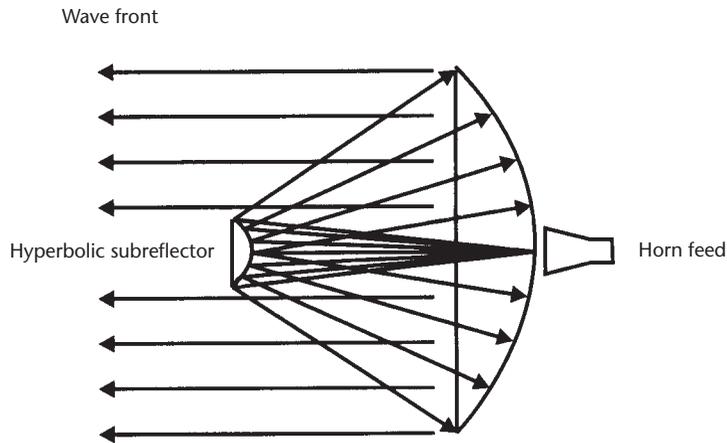


Figure 13.35 Parabolic reflector antenna with a Cassegrain feed. (After: [10].)

13.18 Patch Antennas

13.18.1 Introduction

Patch antennas are a relatively new type of microwave antenna that are finding use in GPS receivers and in microwave communication systems. These antennas are a type of microstrip antenna. Microstrip transmission lines are discussed in Section 11.4.

Microstrip antennas evolved more slowly than their circuit counterparts but have played an increasingly significant role in the antenna field since the mid-1970s. Microstrip antennas provide advantages of low profile, reduced weight, relatively low manufacturing cost, the potential to group many identical patches to build arrays and to integrate with circuit elements, and polarization diversity. Research in this field remains very active and is likely to continue for many years.

Figure 13.36 shows an illustration of microstrip resonators. These include a dipole and four examples of so-called patches. In each case an upper conductor is deposited on top of a dielectric substrate. Resonance occurs when the dipole or patch dimensions are of the order of a half guided wavelength. When the signal frequency is in the vicinity of a resonance, a microstrip resonator radiates a relatively broad beam, broadside to the plane of the substrate. Typically, a microstrip

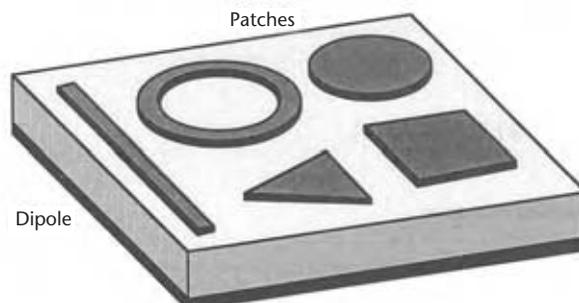


Figure 13.36 Microstrip resonators.

antenna has a gain of about 5 to 6 dB and has a beamwidth between 70 and 90 degrees. As an example, the E-plane and the H-plane radiation patterns shown in Figure 13.37 were measured for a 14×9.6 mm coaxially excited patch on a 0.8 mm thick epoxy substrate with a dielectric constant of 4.4 at its resonant frequency of 4.97 GHz.

The dielectric substrate and the ground plane are both square, with dielectric sides one wavelength wide. The antenna is fed by a coaxial feed located on the axis at 3.6 mm from the center of the patch.

13.18.2 Types of Patch Antennas

The patch shape most commonly used for patch antennas is the rectangular patch. The square patch is a special case of a rectangular patch. Figure 13.38 shows lines of surface current for three resonant modes on a square patch. These lines correspond, respectively, to the TM_{100} , TM_{010} , and TM_{110} modes of an equivalent square-shaped cavity having perfect magnetic conductor side walls.

The circular patch is another possible patch antenna. In this case the patch acts like a circular cavity. When the substrate thickness is much smaller than the radius of the patch, the first modes of resonance are the TM_{mn0} modes.

The microstrip ring resonator is a circular patch with a hole in the center. As a result, its mode patterns are similar to those of a circular patch as long as the center hole is small enough. The radiation patterns are also similar to those of the circular patch.

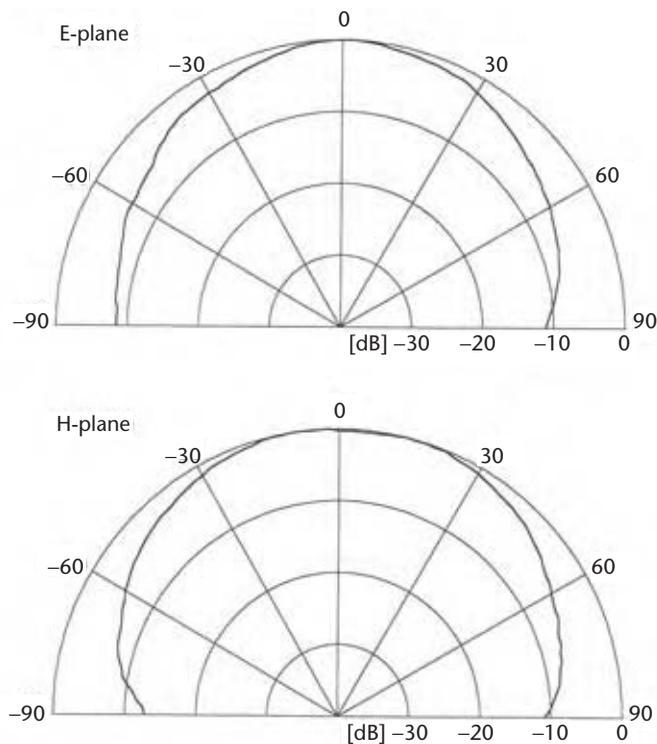


Figure 13.37 Measured radiation patterns of a rectangular microstrip patch antenna.

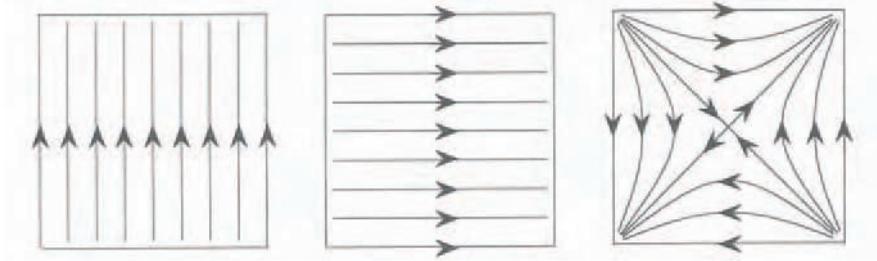


Figure 13.38 of surface current for three resonant modes on a square patch.

Other patch shapes are also possible. Triangular patches are found to provide radiation characteristics similar to those of rectangular patches but with smaller area. They are of particular interest for the design of periodic arrays because triangular radiating elements can be arranged in a manner that allows the designer to reduce significantly the coupling between adjacent elements of the array. This significantly simplifies array design.

Rectangular and circular patches are sometimes combined with short stubs, while slits and notches can be cut in the edges of the resonator and carefully located to reduce the antenna size, the cross-polarization, the sidelobe level, or the coupling between ports in the case of dual polarization antennas.

The resonant frequency of a microstrip resonator can be adjusted over a limited range by electronic means using a varactor diode in the resonant circuit. Such varactor diode circuits are discussed later in this book.

13.18.3 Antenna Feeds

There are a number of different types of antenna feeds that may be used with patch antennas. The simplest way to feed a microstrip patch is to connect a microstrip line directly to the edge of the patch with both elements located on the same substrate. Figure 13.39 shows a patch antenna with a microstrip line feed.

In this figure, the line is connected to the patch within an inset cut in the patch. It was found that cutting such an inset does not significantly affect the resonant frequency but that it modifies the input impedance. By properly selecting the depth of the inset, one can match the patch input impedance to the transmission line impedance without additional matching elements.

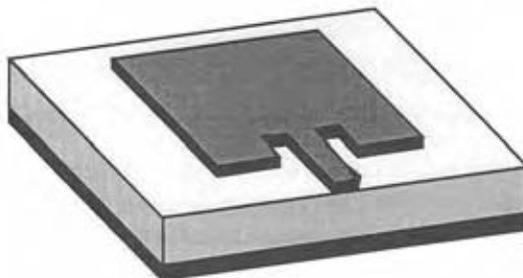


Figure 13.39 Patch antenna with microstrip line feed.

A quite different way to feed a patch is by means of a coaxial line that is set perpendicular to the ground plane. The center conductor of the coaxial line extends across the dielectric substrate and is connected to the patch. The input impedance of the patch depends on the position of the feed so that the patch can be matched to the line by properly positioning the feed.

It is not necessary for the feed line to touch the resonator, because the patch can be fed by proximity coupling. A feed line can be run along the edge of the patch. With this approach coupling takes place continuously along the edge of the patch rather than on a narrow portion of it. Several patches can be excited by the same line to realize linear arrays.

Proximity coupling of the patch to the feed line is also provided by placing the patch and feed line at different levels (the patch above the feed line rather than to the side). This offers some advantages by reducing radiation from the feed line.

Another feed approach is to place the ground plane between the patch and the feed line and using a slot in the ground plane to permit coupling.

13.18.4 Broadbanding

Bandwidth for an antenna may be defined as the percent of center frequency that a suitable impedance or voltage standing wave ratio can be realized. Without broadbanding, the bandwidth of a patch antenna is typically less than 5%. With broadbanding, it can be increased to well over 10%.

One method of increasing the bandwidth of patch antennas is to use parasitic elements on both sides of the central patch. Another approach is to use stacked patches (one above the other of different sizes). Another approach is to use a thick dielectric substrate of very low permittivity. This is done in the so-called *strip-slot-foam-invered patch* (SSFIP) antenna. Bandwidths well in excess of 10% can be achieved by these approaches.

13.18.5 SSFIP Antennas

A sandwich structure that combines the advantages of foam materials with nonresonant slot coupling and broadbanding techniques has been developed. It is called *strip-slot-foam-invered patch* (SSFIP). This structure is illustrated in Figure 13.40 and discussed next.

Ideally, the dielectric substrate of a microstrip patch antenna should be air, which has a relative permittivity of one. The next best material for the substrate is hard foam, some of which has a relative permittivity as low as 1.03. Some foam materials also exhibit good structural, thermal, and electrical properties and have fairly low cost. This material is therefore used for the substrate. The underside of a thin plastic layer is used to deposit the patch, hence the name inverted patch. The microstrip feed line and connecting circuitry on the underside of the ground plane are deposited on a regular microstrip substrate. Bandwidths of about 13% are reported for single patch SSFIP antennas of this type. Bandwidths as large as 33% are reported for SSFIP antennas with stacked patches. Other schemes can be used to increase the bandwidth of a SSFIP antenna besides the use of stacked patches. For instance, parasitic elements can be placed next to the main patch at the same level.

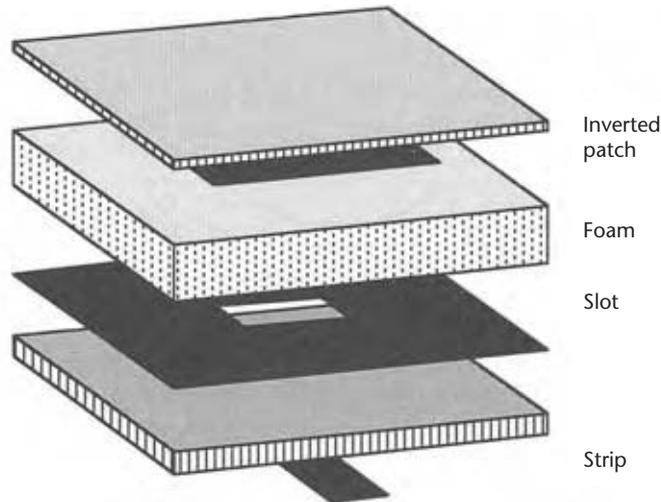


Figure 13.40 Strip-slot-foam-inverted patch antenna.

13.18.6 Dual Polarized and Circularly Polarized Patch Antennas

Figure 13.41 shows two approaches for excitation of dual linear polarization on square patch antennas. These use either pin feed points as would be provided with coax feed lines and edge feed lines. A circular polarized antenna is realized by feeding the patch with equal amplitudes on each line but having a phase difference of 90 degrees. A system of this type is shown in Figure 13.42 where the 90 degree phase difference is achieved at one frequency by increased line length. Other methods of producing the 90-degree phase shift are used where broad bandwidth is desired. An example would be a 90-degree hybrid coupler.

The circular polarization can also be obtained with a single input port that excites two orthogonal modes of a patch resonating at slightly different frequencies. Narrowband circularly polarized antennas of this type are shown in Figure 13.43.

Dual polarized wideband SSFIP antennas are also possible. These use dual slot excitation. Bandwidths of the order of 10.9% have been reported. A number of other dual polarized antennas are possible including antennas with multilayer feed structures. Such a system is shown in Figure 13.44.

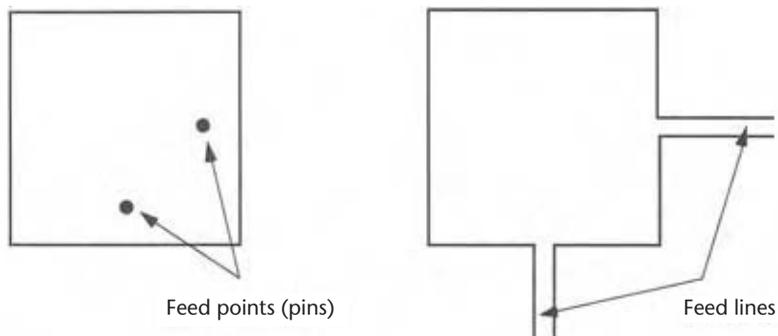


Figure 13.41 Excitation of dual linear polarization on patch antennas.

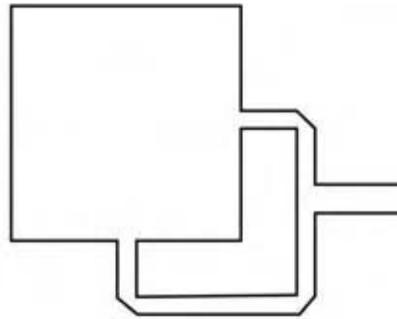


Figure 13.42 Circular polarization patch antenna with offset lines.

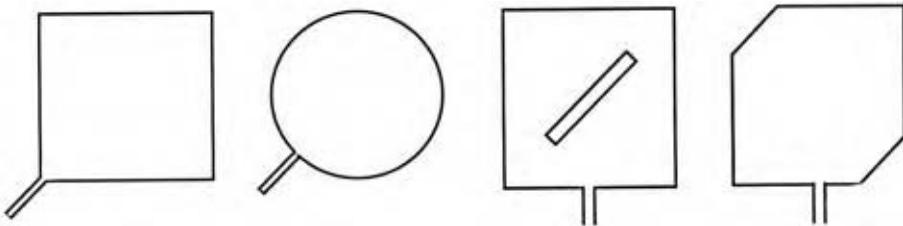


Figure 13.43 Narrowband circularly polarized patch antennas.

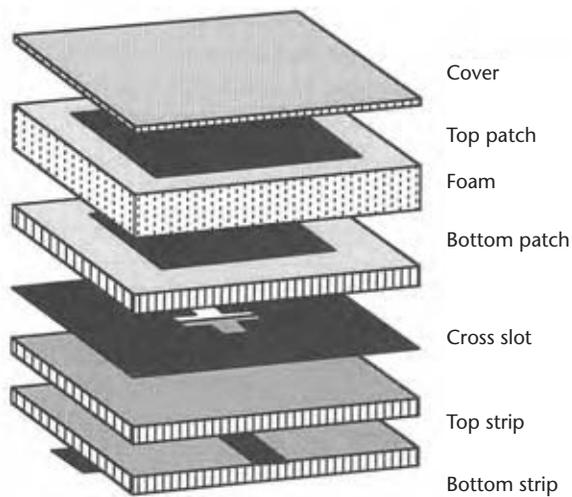


Figure 13.44 Patch antenna with multilayer feed structure.

13.18.7 Patch Antenna Arrays

Many examples of patch antenna arrays have been reported. These include both line arrays and planar arrays. Both wide band and narrow band patches have been used. An example X-band seven-element hexagonal array provided an 11% bandwidth. Sidelobe levels were below -16 dB.

Much of the material presented in this section is quoted or adapted from [12].

References

- [1] Jasik, H., (ed.), *Antenna Engineering Handbook*, New York: McGraw-Hill, 1961, Chap. 3.
- [2] Kennedy, G., *Electronic Communication Systems*, 3rd ed., New York: McGraw-Hill, 1985, Chap. 9.
- [3] Jasik, H., (ed.), *Antenna Engineering Handbook*, New York: McGraw-Hill, 1961, pp. 24-28–24-29.
- [4] Stutzman, W. L., and G. A. Thiele, *Antenna Theory and Design*, New York: John Wiley & Sons, 1981, pp. 220–226.
- [5] Stutzman, W. L., and G. A. Thiele, *Antenna Theory and Design*, New York: John Wiley & Sons, 1981, pp. 278–281.
- [6] Stutzman, W. L., and G. A. Thiele, *Antenna Theory and Design*, New York: John Wiley & Sons, 1981, pp. 281–287.
- [7] Stutzman, W. L., and G. A. Thiele, *Antenna Theory and Design*, New York: John Wiley & Sons, 1981, pp. 287–303.
- [8] Jasik, H., (ed.), *Antenna Engineering Handbook*, New York: McGraw-Hill, 1961, Chap. 8.
- [9] Terman, F. E., *Electronic and Radio Engineering*, 4th ed., New York: McGraw-Hill, 1955, p. 915.
- [10] Skolnik, M. I., *Introduction to Radar Systems*, 2nd ed., New York: McGraw-Hill, 1980, Chapter 7.
- [11] Hansen, R. C., “Antennas,” in *Reference Data for Engineers*, 8th ed., M. E. Van Valkenburg, (ed.), SAMS, Carmel, IN: Prentice Hall Computer Publishing, 1993, pp. 32-1–32-56.
- [12] Zurcher, J. -F., and F. E. Gardiol, *Broadband Patch Antennas*, Norwood, MA: Artech House, 1995.

Lumped Constant Components and Circuits

This chapter discusses resistors, inductors, and capacitors, so-called lumped constant components. The dimensions of these components generally are small compared to a wavelength. This chapter also discusses a number of circuits that use those components, including series resonant circuits, parallel resonant circuits, impedance-matching networks, and LC filters. Other lumped constant components and circuits are discussed in subsequent chapters.

14.1 Conductors and Skin Effect

The following is based on [1].

One of the important differences between LF systems and HF systems is that at high frequencies the current in a conductor is not uniform over the cross-section. Inductive reactance effects at high frequencies cause the current to flow only in the outer surface of the conductor. In computing resistance of a conductor at RF, it is common practice to first compute the skin depth. The skin depth is that distance below the surface of the conductor where the current density has diminished to $1/e$ its value at the surface, where e is the base of the natural logarithm (2.718). The thickness of the conductor is assumed to be several times the skin depth at the lowest frequency of interest to minimize ohmic or conductor losses. This value also applies to the thickness of microstrip or stripline conductors. The acceptable standard is ≥ 3 for δ low-loss microstrip or strip transmission line.

In computing resistance of a round conductor, the conductor is simulated by a cylindrical shell of the same surface shape but a thickness equal to the skin depth, with uniform current density equal to that which exists at the surface of the actual conductor.

The skin depth is given by

$$\delta = (\lambda / \pi \sigma \mu c)^{1/2} \quad (14.1)$$

where:

δ = skin depth (centimeters)

λ = free-space wavelength (centimeters)

σ = conductivity (siemens per centimeter) = $1/\rho$

ρ = resistivity (ohm-centimeters)

μ = permeability of conductor = $4\pi \times 10^{-9} \mu_r$ in henry/cm

μ_r = relative permeability of conductor material = 1.0

c = speed of light = 3×10^{10} cm per second

For numerical computations,

$$\delta = (6.61/f^{1/2})k_1 \text{ cm} \quad (14.2)$$

where

$$k_1 = [(1/\mu_r)\rho/\rho_c]^{1/2} \quad (14.3)$$

and

f = frequency (hertz)

ρ_c = resistivity of copper ($1.724 \times 10^{-6} \Omega\text{-cm}$)

At microwave and higher frequencies, the skin depth is very small and the resistance of even the best conductors may be high unless large dimensions are used.

14.2 RF Resistors

Resistors are used in many RF circuits. Applications include transistor biasing, loads for wideband amplifiers, gain control for receivers, and signal attenuators. Resistors used in RF circuits can have a number of shapes, with cylindrical being the most common at the lower frequencies. Resistors usually are made of carbon. At the higher frequencies, they may be in the form of carbon compound deposits on insulator surfaces or tantalum nitride (TaN) or nichrome (NiCr) for thin-film circuits. Platinum- or palladium-based compounds are used for thick-film resistors.

All resistors must have conducting leads connected to each end. At the higher frequencies, the length of the leads must be kept as short as possible. With chip resistors, mounting-pad size is determined by placement accuracy, part dimensional tolerance, and attachment method.

14.3 Inductors and Inductive Reactance

The approximate inductance for a straight short wire is given by

$$L = (\mu_0 l/2\pi)[\ln(4l/d)] \quad (14.4)$$

where:

L = inductance (henrys)

l = wire length (centimeters)

d = wire diameter (centimeters)

\ln = natural logarithm

For an example of the use of (14.4), assume a wire with a length of 3 cm and a diameter of 0.2 cm. Substituting values into (14.4), the inductance, L , would be

$$\begin{aligned} L &= 2 \times 10^{-9} [\ln(4l/d)] \\ L &= 2 \times 10^{-9} \times 3.0 [\ln(12/0.2)] \\ &= 24.5 \times 10^{-9} \text{H} \\ &= 24.5 \text{ nH} \end{aligned}$$

The inductive reactance is given by

$$X_L = \omega L = 2\pi fL \quad (14.5)$$

where:

f = frequency (hertz)

L = inductance (henrys)

At a frequency of 200 MHz, a wire with a diameter of 0.2 cm and a length of 3 cm would have a reactance of 30.8Ω . We thus see the reason that resistor or capacitor leads must be kept as short as possible at RF frequencies to avoid having an undesirably large inductance in series with the resistance or capacitance.

An air-core solenoid made of thin wire has a much greater inductance for the same length than the straight thin wire. At RF, such solenoids often have only one turn or, at most, a few turns.

The approximate inductance for an air-core solenoid of a number of turns is given by

$$L = n^2 \left[r^2 / (9r + 10l) \right] \quad (14.6)$$

where

L = inductance (microhenrys)

n = number of turns

r = solenoid radius (inches)

l = solenoid length (inches)

For an example of the use of (14.6), assume a solenoid with a radius of 0.25 inch, a length of 0.5 inch, and 5 turns. Substituting those values into (14.6), the inductance would be

$$L = 25 [0.0625 / (2.25 + 5)] = 0.216 \mu\text{H}$$

The inductive reactance for this solenoid at a frequency of 200 MHz would be 270.8Ω .

Figure 14.1 shows the flat square spiral inductor. This device frequently is used with printed transmission line circuits, that is, microstrip.

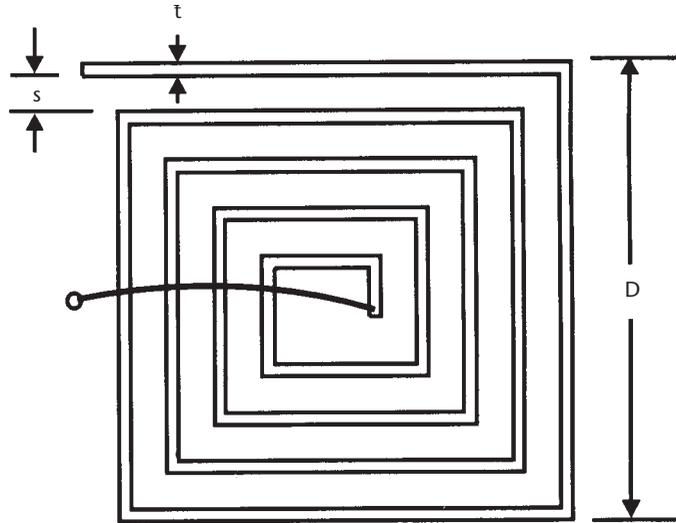


Figure 14.1 Flat square spiral inductor.

The approximate inductance for the flat square spiral inductance is given by

$$L = 85 \times 10^{-3} \times D \times N^{1.7} \quad (14.7)$$

where:

L = inductance (microhenrys)

D = maximum outside dimension in centimeters

$N = (D/2)/(w + s)$

s = space between lines (centimeters)

w = width of lines (centimeters)

Another important type of inductor used with RF circuits is a toroidal inductor, which is shown in Figure 14.2. It consists of a few turns of wire with a powdered iron or ferrite core. The core adds greatly to the inductance of the coil at VHF and lower frequencies.

The toroidal inductor with a ferrite or powdered iron core and a square cross-section has an inductance that is given by

$$L = (N^2 \mu_r \mu_0 / 2\pi) \ln(d_2 / d_1) \quad (14.8)$$

where:

L = inductance (henrys)

N = number of turns

μ_r = relative permeability of core

μ_0 = permeability of free space = $4\pi \times 10^{-7}$ H/m

t = toroid thickness (meters)

d_2 = toroid outer diameter (meters)

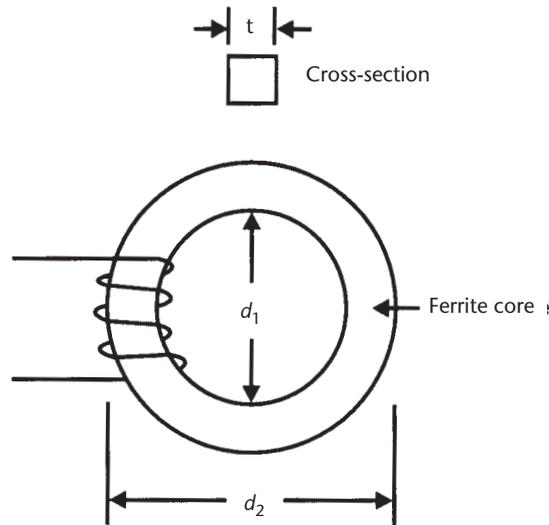


Figure 14.2 Toroidal inductor.

d_1 = toroid inner diameter (meters)

For an example of the use of (14.8), assume the following:

$$N = 4$$

$$\mu_r = 1,900$$

$$\mu_0 = 4\pi \times 10^{-7} \text{ H/m}$$

$$t = 0.0005\text{m}$$

$$d_2 = 0.02\text{m}$$

$$d_1 = 0.01\text{m}$$

$$L = (N^2 \mu_r \mu_0 t / 2\pi) \ln(d_2/d_1)$$

$$L = (16 \times 1,900 \times 4\pi \times 10^{-7} \times 0.0005 / 2\pi) \ln(0.02/0.01)$$

$$L = 2.10 \times 10^{-5} \text{ H} = 21.0 \mu\text{H}$$

The current in an inductor lags the applied voltage by 90 degrees. The inductive reactance, which is the voltage divided by the current, thus has a phase angle of +90 degrees.

The total inductance for two or more inductors in series is the sum of the individual inductances. The total inductance for two or more inductors in parallel is the reciprocal of the sum of the inductive susceptances, where the inductive susceptance is the reciprocal of the inductive reactance ($-jB_L = 1/+jX_L$). Unlike resistors, no power is dissipated in pure inductors. Energy is stored in the magnetic field and is transferred back to the conductor when the field strength is returned to zero.

The impedance of an RL circuit containing both resistance and inductance in series is given by

$$Z = R + jX_L \quad (14.9)$$

where:

j = operator indicating a phase angle of 90 degrees

Z = impedance

R = resistance

X_L = inductive reactance = $2\pi fL = \omega L$

f = frequency

ν = radian frequency

The operator j is also equal to the square root of -1 .

14.4 RF Chokes

An RF choke is another name for an inductor that is used to pass dc current but that blocks ac or RF current. Lumped-element chokes have capacitance between windings and are parallel resonant circuits at some frequency. This resonant frequency normally is chosen to be above the highest frequency with which we want to operate. This component is used in many RF circuits. In microstrip or stripline, chokes normally are printed sections of high-impedance transmission line a quarter-wavelength long at the center frequency of interest.

14.5 Capacitors and Capacitive Reactance

A capacitor typically consists of two conducting surfaces or plates separated by a dielectric insulation material that permits the storage of energy in the electric field between the plates. This component is illustrated in Figure 14.3. The dielectric prevents current flow when the applied voltage is constant, but a time-varying voltage produces a current proportional to the rate of voltage change.

The current in a capacitor is given by

$$I = C dv/dt \quad (14.10)$$

where C is the capacitance measured in farads.

The capacitance, C , of the parallel plate structure is given by

$$C = \epsilon A/d = \epsilon_0 \epsilon_r A/d \quad (14.11)$$

where:

ϵ = absolute permittivity of the dielectric = $\epsilon_0 \epsilon_r$

A = area of parallel plates or conductors

d = spacing of plates or conductors

ϵ_0 = permittivity of free space = 8.854×10^{-12} F/m

ϵ_r = relative permittivity or dielectric constant of dielectric medium

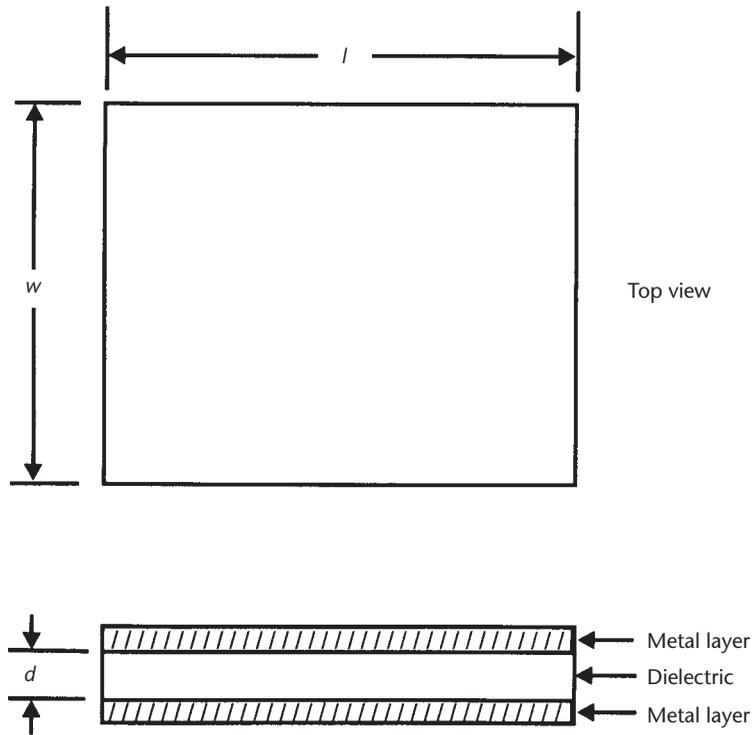


Figure 14.3 Illustration of a two-plate capacitor.

A typical dielectric such as mica has a permittivity of 5×10^{-11} F/m. Thus, the relative permittivity or dielectric constant of mica is 5.65. Materials used for printed circuits, microstrip, and stripline circuits have dielectric constants in the range of 2.3 to 2.6. FR-4 is about $\epsilon_r = 4.5$. Microstrip is also fabricated on alumina ($\epsilon_r = 9.6 - 10.0$) and beryllia ($\epsilon_r = 6.0$). The relative permittivity for fused silica is about $\epsilon_r = 2.1$.

For an example of the use of (14.11), assume the following:

Dielectric constant = 2.5

Area = 1 cm^2

Spacing = 0.1 cm

Substituting those values into (14.11) gives

$$C = 2.5 \times 8.854 \times 10^{-12} \times 1/0.1 = 2.21 \times 10^{-12} \text{ F}$$

$$C = 2.21 \text{ pF}$$

In many cases, parallel plates are used to provide larger capacitance. In this case, the total capacitance is given by

$$C = (N - 1)\epsilon_0 \epsilon_r A/d \quad (14.12)$$

where N = number of plates, and the other terms are as defined for (14.11).

There are many other possible shapes and types for capacitors, including mechanically adjustable and voltage-variable capacitors. One form of solid-state

diode has a capacitance that can be changed by changing the dc voltage across the diode. Such a voltage-variable capacitance diode is called a varactor.

Table 14.1 lists some of the more important types of available fixed-value capacitors with capacitance range, maximum voltage, and other characteristics. Comments about each type of capacitor follow.

Mica dielectric capacitors have excellent RF characteristics. Tubular, ceramic dielectric capacitors are available in very low values. Disk ceramic dielectric capacitors are small, inexpensive, and popular. Chip ceramic dielectric capacitors are used frequently in HF circuits. Mylar dielectric capacitors also are inexpensive and popular. Polystyrene dielectric capacitors are high quality but large. Polycarbonate dielectric capacitors are high quality. Glass and porcelain dielectric capacitors are good quality and inexpensive and have good long-term stability. Tantalum dielectric capacitors have high capacitance with acceptable leakage. They are polarized, meaning that there is a plus and a minus voltage assignment for terminals. They are small, have low inductance, and are popular. Electrolytic capacitors are used mainly for power supply filters. They are polarized. Oil-filled capacitors are used for high-voltage filters. They are large, have long life, and are not polarized.

The capacitive reactance is given by

$$X_c = 1/\omega C \quad (14.13)$$

where:

X_c = capacitive reactance (ohms)

$\omega = 2\pi \times f$

f = frequency (hertz)

C = capacitance (farads)

The total capacitance for two or more capacitors connected in series is the reciprocal of the sum of the reciprocals of the individual capacitors. The total capacitance for two or more capacitors connected in parallel is the sum of capacitance for the

Table 14.1 Characteristics for Fixed-Value Capacitors

<i>Type</i>	<i>Capacitance Range</i>	<i>Maximum Voltage</i>	<i>Accuracy</i>	<i>Temperature Stability</i>
Mica	1 pf–0.01 μf	100–600	Good	Good
Tubular	0.5–100 pf	100–600	Selectable	Poor
Ceramic disk and chip	10 pf–1 μf	50–1,000		
Ceramic mylar	0.001–10 μf	50–600	Good	Poor
Polystyrene	10 pf–0.01 μf	100–600	Good	
Polycarbonate	100 pf–10 μf	50–400	Good	Good
Glass	10–1,000 pf	100–600	Good	
Porcelain	100 pf–0.1 μf	50–400	Good	Good
Tantalum	0.1–500 μf	6–100	Poor	Poor
Electrolytic	0.1 μf –0.2 f	3–600	Poor	Poor
Oil	0.1–20 μf	200–10k		

individual capacitors. Capacitive susceptance is the reciprocal of the capacitive reactance. In other words, capacitances in parallel are like resistors or inductors in series and capacitances in series are like resistors or inductors in parallel.

Unlike resistors, no power is dissipated in pure capacitors. In a real capacitor, however, there is some loss or power dissipation due to a small conductance between the plates or the sides of the capacitor or the finite resistivity of the plate material itself.

The impedance of a series RC circuit containing both resistance and capacitance is given by

$$Z = R - jX_C \quad (14.14)$$

where:

Z = impedance (voltage/current)

R = resistance

jX_C = capacitive reactance

The j in the capacitive reactance expression is the operator that indicates a phase angle of 90 degrees. The $-j$ indicates a phase angle of -90 degrees.

If a dc voltage is applied across a circuit consisting of a resistor and a capacitor connected in series, the voltage across the capacitor changes slowly with a rise time of RC . It changes by 63% in one RC time. The time it takes to go from 10% to 90% of its final value is 2.2 RC . That usually is referred to as the rise time of a pulse in an electronic circuit.

Some of the preceding material is based on [2].

14.6 Series Resonant RLC Circuits

Some of the material in the following three sections is quoted or adapted from [3].

A series resonant RLC circuit is made up of one or more resistors, one or more capacitors, and one or more inductors, all connected in series. Figure 14.4 shows impedance characteristics for a series resonant RLC circuit. The Q in this case is 10, where $Q = X/R$ at resonance. At resonance, the impedance is a minimum, and the impedance angle is zero. Below resonance, the circuit becomes capacitive and the impedance angle is negative. Above resonance, the impedance becomes inductive, and the impedance angle is positive.

The input impedance of the series resonant circuit is given by

$$Z = V/I = R + j(\omega L - 1/\omega C) \quad (14.15)$$

where terms are as previously defined.

A large Q means that the resonant peak for Z will be narrow, while a small Q will produce a wide resonant peak. It can be shown that the -3 -dB bandwidth, BW_{3dB} , is equal to the center or resonant frequency, f_0 , divided by the Q , as shown in (14.16):

$$BW_{3dB} = f_0 / Q \quad (14.16)$$

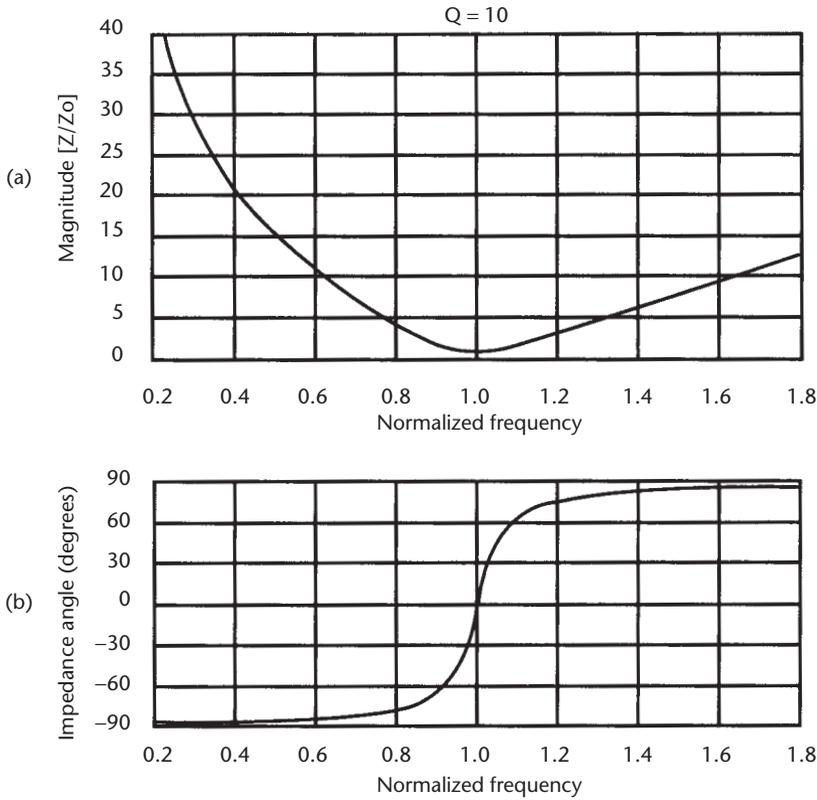


Figure 14.4 Series resonant RLC circuit impedance characteristics: (a) magnitude of normalized impedance as a function of normalized frequency and (b) phase angle of normalized impedance as a function of normalized frequency. (After: [3].)

At resonance, the voltage across the capacitor is Q times the applied voltage. That can be important from the point of view of possible voltage breakdown for the capacitor.

14.7 Parallel Resonant RLC Circuits

A parallel resonant RLC circuit consists of a resistor, an inductor, and a capacitor, all in parallel. Examples of impedance characteristics for such a circuit are shown in Figure 14.5. Notice that the impedance is maximum at resonance and falls off on either side of resonance. The impedance angle is positive below resonance, zero at resonance, and negative above resonance.

Parallel resonance occurs when the input voltage and current are in phase and $B_L = B_C$ or $1/\omega L = \omega C$.

The Q of the parallel circuit is given by

$$Q = \omega_0 C/G = \omega_0 CR = R/\omega_0 L \tag{14.17}$$

$$G = \text{conductance} = 1/R \tag{14.18}$$

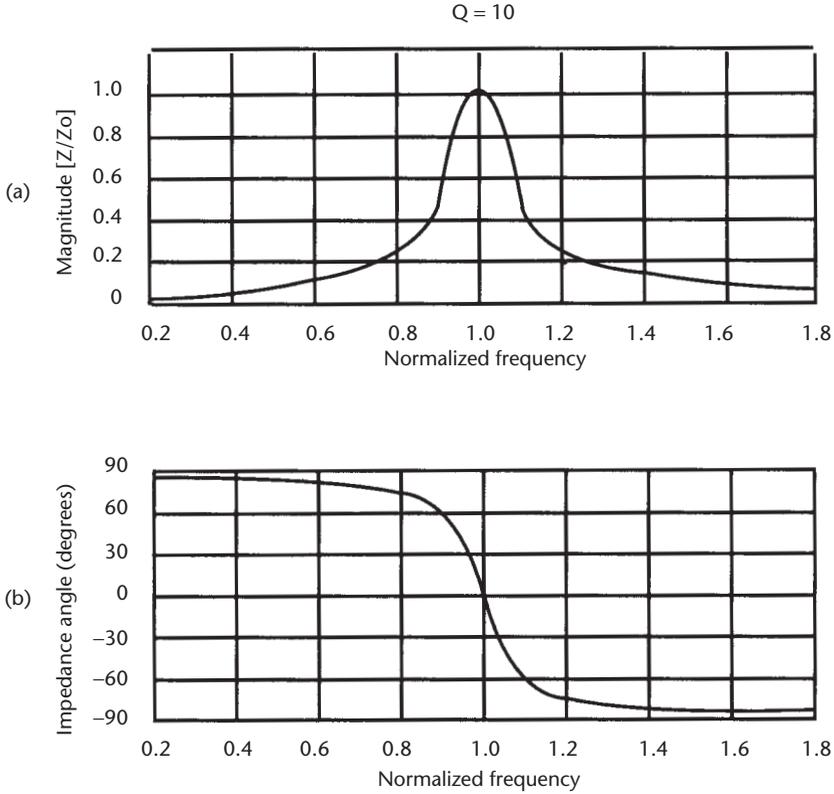


Figure 14.5 Impedance characteristics of parallel resonant RLC circuits: (a) normalized impedance as a function of normalized impedance and (b) phase angle of normalized impedance as a function of normalized frequency. (After: [3].)

where ω_0 = resonant radian frequency.

The admittance for the circuit is

$$Y = G + j(\omega C - 1/\omega L) \tag{14.19}$$

where $Y = 1/Z$. The resonance peak is narrow for high values of Q and wide for small values of Q . Again, the 3-dB bandwidth is equal to the resonant center frequency divided by Q . It can be shown that, at resonance, the current in the capacitor is Q times the input current to the resonant circuit.

The following are examples of calculations for the impedance of a parallel resonant circuit:

At $f/f_0 = 0.5$ and $Q = 10$

Normalized $G = 1$

Normalized $B_c = 0.5Q = 5$

Normalized $B_L = Q/0.5 = 20$

$$Y/Y_0 = 1 + j5 - j20 = 1 - j15 = 15.03 \angle -86.19 \text{ degrees}$$

$$Z/Z_0 = 1/(Y/Y_0) = 0.0665 \angle 86.19 \text{ degrees}$$

For an example of component calculations, assume a Q of 10, a frequency of 100 MHz, and a resistance of $1,000\Omega$.

$$Q = \omega CR = R/\omega L \quad (14.20)$$

$$C = 10 / (628 \times 10^8 \times 1,000) = 15.9$$

$$L = 1,000 / (10 \times 628 \times 10^8) = 0.159 \mu\text{H}$$

Parallel resonant circuits are used for many applications in RF circuits. One important application is for selecting the desired frequency spectrum for amplifiers, oscillators, filters, and other RF components.

14.8 Complex Resonant Circuits

In many practical cases, resonant circuits are more complex than simple series resonant or parallel resonant circuits. Four examples of such complex circuits are shown in Figure 14.6: (a) a circuit with a resistor in series with the inductor; (b) a circuit with a resistor in series with the capacitor; (c) a transformer-like circuit with two capacitors and the load resistor connected in parallel with one of them; and (d) a second transformer-like circuit but with two inductors and the load resistor connected in parallel with one of them. Figure 14.6 and the analyses that follow show how these circuits are converted to simple parallel resonant RLC circuits.

Circuit analysis can be used to calculate the equivalent circuits shown in Figure 14.6. Alternatively, it is possible to use a Smith chart for circuit analysis of complex circuits. Following is a brief discussion of a method of analyzing complex resonant circuits using Q values and formulas for converting from parallel to series equivalent circuits and from series to parallel equivalent circuits. The circuit in Figure

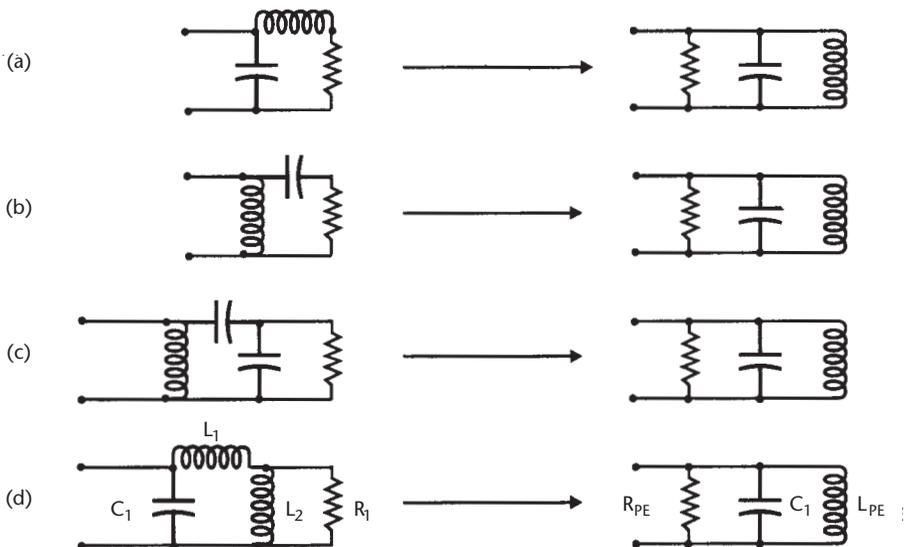


Figure 14.6 (a-d) Resonant circuits and circuit transformation. (After: [3].)

14.6(d) is used as the circuit that is converted to a simple parallel resonant RLC circuit. The assumed impedance values for the circuit are as follows:

$$R_1 = 50\Omega$$

$$X_{L1} = 183\Omega$$

$$X_{L2} = 125\Omega$$

$$X_C = 209\Omega$$

The normalized values of impedance based on 50Ω are as follows:

$$R_1 = 1$$

$$X_{L1} = 3.66$$

$$X_{L2} = 2.5$$

$$X_C = 4.19$$

The Q of the parallel and the series circuits are as follows:

$$\text{Parallel } Q = Q_P = R_P / X_P \quad (14.21)$$

$$\text{Series } Q = Q_S = X_S / R_S \quad (14.22)$$

The subscript P stands for parallel, and the subscript S for series. The exact formulas for conversion to equivalent circuit values are as follows:

$$R_{PE} = R_S (1 + Q_S^2) \quad (14.23a)$$

$$X_{PE} = X_S (Q_S^2 + 1) / Q_S^2 \quad (14.23b)$$

and

$$R_{SE} = R_P / (1 + Q_P^2) \quad (14.24a)$$

$$X_{SE} = X_P [Q_P^2 / (Q_P^2 + 1)] \quad (14.24b)$$

The approximate formulas for conversion to equivalent circuits are as follows. These formulas assume $Q \gg 1$.

$$R_{PE} = R_S Q_S^2 \quad (14.25a)$$

$$X_{PE} = X_S \quad (14.25b)$$

and

$$R_{SE} = R_P / Q_P^2 \quad (14.26a)$$

$$X_{SE} = X_P \quad (14.26b)$$

The exact circuits are next applied to the circuit in Figure 14.6(d). First, we convert the parallel circuits L_2 and R_1 to a series circuit. The value of Q_p is found as follows:

$$\text{Parallel } Q = Q_p = R_p / X_p = 1/2.5 = 0.4$$

The series equivalent circuit values are found as follows:

$$R_{SE} = R_p / (1 + Q_p^2) = 1 / (1 + 0.16) = 0.862$$

$$X_{SE} = X_p [Q_p^2 / (Q_p^2 + 1)] = 2.5 [0.16 / (0.16 + 1)] = 0.345$$

Adding X_{L1} to those values, we have

$$Z = 0.862 + j3.66 + j0.345 = 0.862 + j4.0$$

We now convert this series RL circuit to a parallel RL circuit. The value of Q_s is found as follows:

$$\text{Series } Q = Q_s = X_s / R_s = 4.0 / 0.862 = 4.64$$

The parallel equivalent circuit values are found as follows:

$$R_{PE} = R_s (1 + Q_s^2) = 0.862 (1 + 21.5) = 19.4$$

$$X_{PE} = X_s (Q_s^2 + 1) / Q_s = 4.0 [(21.5 + 1) / 21.5] = 4.19$$

The normalized value of X_c was given as 4.19, so reactance values for the equivalent parallel inductance and capacitance are the same. The total circuit Q is $19.4/4.19 = 4.6$.

Now we convert the normalized impedance to component values, assuming a radian frequency ω of 10^8 rps (frequency = 15.92 MHz), $Z_0 = 50\Omega$, and $Y_0 = 0.02S$.

$$R_1 = 50\Omega$$

$$L_1 = 183.0 \times 10^{-8} \text{ H} = 1.83 \mu\text{H}$$

$$L_2 = 125 \times 10^{-8} \text{ H} = 1.25 \mu\text{H}$$

$$C_1 = 4.8 \times 10^{-11} \text{ F} = 84 \text{ pF}$$

$$R_{PE} = 19.4 \times 50 = 970\Omega$$

$$X_{PE} = 4.19 \times 50 = 208.3\Omega$$

$$L_{PE} = 208.3 \times 10^{-8} \text{ H} = 2.08 \mu\text{H}$$

$$Q = R_p / X_{C1} = 970 / 208.3 = 4.66$$

$$B = 15.92 / 4.66 = 3.42 \text{ MHz}$$

14.9 The Use of the Smith Chart for Circuit Analysis

In Chapter 11, the Smith chart was used in transmission line analysis and design. Here, it is used in lumped constant circuit analysis and design.

Figure 14.7 shows an example of the use of the Smith chart for circuit analysis. This example also performs the analysis for the circuit in Figure 14.6(d).

The starting point will be R1. We will assume that this resistor has a normalized impedance R/Z_0 based on 50Ω of 1. It is thus located on the Smith chart at the center of the chart at point 1. The normalized conductance is also 1 and is located at the same point since the conductance is simply 1 over the resistance. To this conductance, we add the inductive admittance $-jBL2$. We will assume that this normalized admittance is $-j0.4$, which is equal to $1/jX_{L2} = -j/2.5 = -j0.4$. Next, we travel along the 1.0 conductance circle to the $-j0.4$ arc. The normalized admittance at that point is $1 - j0.4$, located at point 2.

Next, we convert to a series impedance. That is done so that this impedance can be added to the impedance of L_1 . The conversion is done by traveling along the straight line through the center of the chart to point 3, which is located the same distance from the center as point 2. Recall that to change from an admittance chart to an impedance chart, we simply move a quarter-wavelength on an SWR circle. That is equivalent to transferring through the center of the chart on a straight line to a point that is an equal distance from the center of the chart. The normalized impedance at this point is about $0.85 + j0.35$.

$L1$ has a normalized impedance of $+j3.66$. To add that value to $+j0.34$, travel along the $R = 0.85$ circle to point 4. The normalized impedance at that point is $0.85 + j4.0$.

This impedance is now converted to an admittance so it may be added to the admittance of C. This is done by moving along a straight line through the center of the chart to point 5. This point is located the same distance from the center of the chart as point 4. The normalized admittance, $G_3 - jB_{L3}$, at this point is $0.051 - j0.24$.

The next move is along the $G = 0.051$ circle to point 6 by adding a capacitive susceptance B_{C1} of $+j0.24$. The admittance at point 6 is $0.051 + j0$.

Next, again convert this admittance to an impedance. Again, move on a straight line through the center of the chart to point 7, which also is located the same distance from the center of the chart as point 6. The normalized impedance at that point is $19.4 + j0$. The real impedance based on a normalized resistance of 50Ω is $R_p = 50 \times 19.4 + j0 = 970\Omega$. This completes the graphical analysis and design.

Now, convert the normalized impedance to component values, assuming a radian frequency, ω , of 10^8 rps (frequency = 15.92 MHz), $Z_0 = 50\Omega$, and $Y_0 = 0.02S$.

$$R_1 = 50\Omega$$

$$L_1 = 183.0 \times 10^{-8} \text{ H} = 1.83 \mu\text{H}$$

$$L_2 = 125 \times 10^{-8} \text{ H} = 1.25 \mu\text{H}$$

$$C_1 = 4.8 \times 10^{-11} \text{ F} = 84 \text{ pF}$$

$$R_{PE} = 19.4 \times 50 = 970\Omega$$

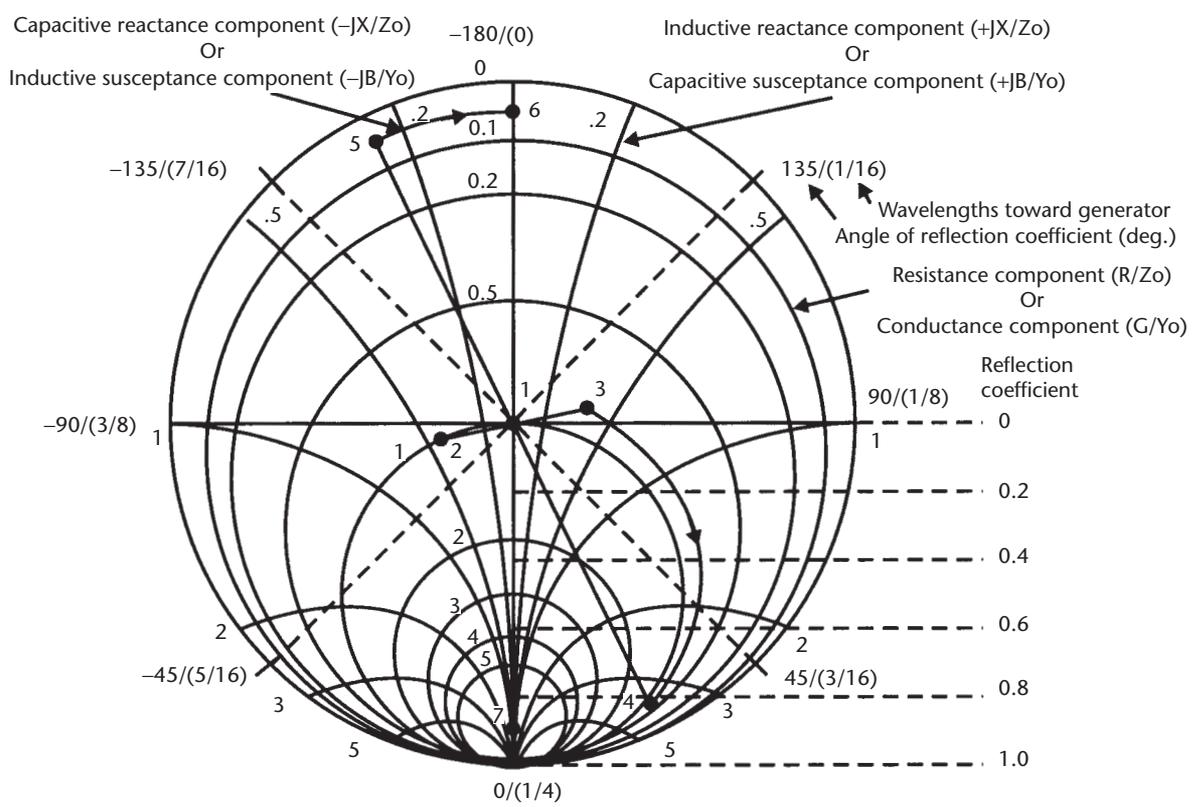


Figure 14.7 Example of the use of the Smith chart for circuit analysis.

$$X_{pE} = 4.19 \times 50 = 208.3\Omega$$

$$L_{pE} = 208.3 \times 10^{-8} \text{ H} = 2.08 \mu\text{H}$$

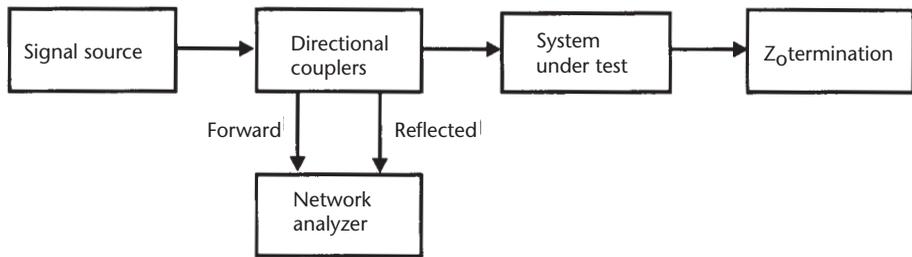
$$Q = R_p / X_{C1} = 970 / 208.3 = 4.66$$

$$B = 15.92 / 4.66 = 3.42 \text{ MHz}$$

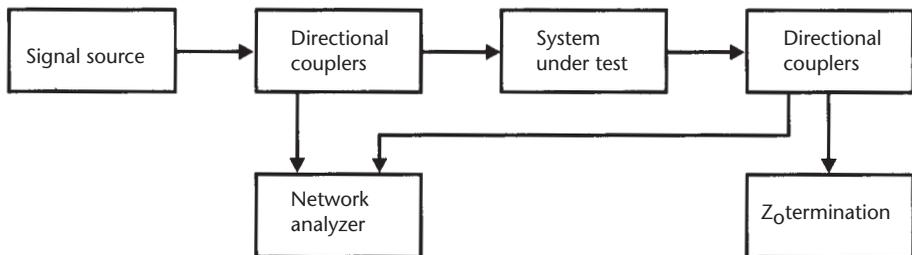
14.10 S-Parameters

Another way to represent a complex circuit is by means of scattering parameters, or S-parameters. With this approach, parameters are measured by placing the system under test in an S-parameter test set with an associated network analyzer and signal source. The S-parameter test set is a mounting and switching arrangement with directional couplers and line terminations that allow for measurement of voltage reflection coefficients and network power-gain measurements. The signal source usually has a 50Ω output impedance, and transmission lines and terminations also are 50Ω . This test system is illustrated in Figure 14.8 for two types of test conditions.

In the circuit in Figure 14.8(a), directional couplers are used to measure the forward and reflected signal voltages, thus providing input and output reflection coefficients. The system under test uses switches, so that for the measurement of the input reflection coefficient (reflected signal voltage/forward signal voltage), the



(a)



(b)

Figure 14.8 Illustrations of setups for measuring S-parameters: (a) setup for measuring S_{11} and S_{22} and (b) setup for measuring S_{21} and S_{12} .

output of the two-way directional coupler is connected to the input of the system under test; and for measurement of output reflection coefficients, the output of the two-way directional coupler is connected to the output of the system under test. In each case, the system under test is terminated in Z_0 . The input voltage reflection coefficient is called S_{11} , and the output voltage reflection coefficient is called S_{22} . These parameters have both magnitude and phase angle. They can be plotted on a Smith chart to obtain the input and output circuit impedances.

Figure 14.8(b) shows the setup for measuring S_{21} and S_{12} . S_{21} is the output signal power over the input signal power, as measured with the two directional couplers when the signal is fed to the input of the system under test. S_{12} is the output signal power over the input signal power measured the same way when the signal is fed to the output of the system under test. In each case, the system under test is terminated in Z_0 . These parameters have both magnitude and phase angle. Figure 14.9 illustrates the use of the Smith chart in converting S-parameters to input or output impedances. In this case, S_{11} is plotted. It is assumed to have a value of 0.5 at an angle of -140 degrees. It is plotted by constructing a radial line from the center of the chart to -140 degrees, as shown on the angle of the reflection coefficient circle. A reflection coefficient circle is also constructed using the radial-scaled parameters for a voltage reflection coefficient of 0.5. The intersection of the radial angle line and the reflection coefficient circle gives us the S_{11} location. The corresponding input impedance is read from the Smith chart as $(0.37 - j 0.32)Z_0$. If Z_0 is 50Ω , the input impedance would be $18.5 - j 16\Omega$. In many cases, the network analyzer used to measure the S-parameters includes a Smith chart CRT display. The impedance values thus can be read directly from this chart. The magnitude and the angle also can be given by meter displays. The information can then be plotted on paper versions of the Smith chart.

14.11 Impedance Matching Using LC Circuits

Impedance matching is important in RF systems. For example, the maximum power gain of an RF amplifier can be realized only if the input of the amplifier is impedance matched to the source impedance and the output of the amplifier is impedance matched to the load impedance. The same is true when matching the impedance of an antenna to the transmission line connected to it, so there will be no reflections.

With impedance-matching circuits, the objective is to move from the source impedance to the complex conjugate of the load impedance. The complex conjugate impedance has the same magnitudes for resistance and reactance components as the load impedance but the opposite sign for the reactance component. Thus, if the load impedance is $4.0 + j 1.0$, the complex conjugate impedance is $4.0 - j 1.0$.

Figure 14.10 shows four examples of discrete impedance-matching networks.

The circuit in Figure 14.10(a) shows a single L-section for output matching. It shows a transistor as the input or source impedance with dc current fed through an RF choke (RFC). This choke blocks RF signals and, if it is designed correctly, does not usually enter into the analysis. A low-impedance coupling capacitor, C_1 , couples RF signals from the transistor to the L-section network made up of L and C_2 . Again, this coupling capacitor does not enter into the analysis of the impedance-matching

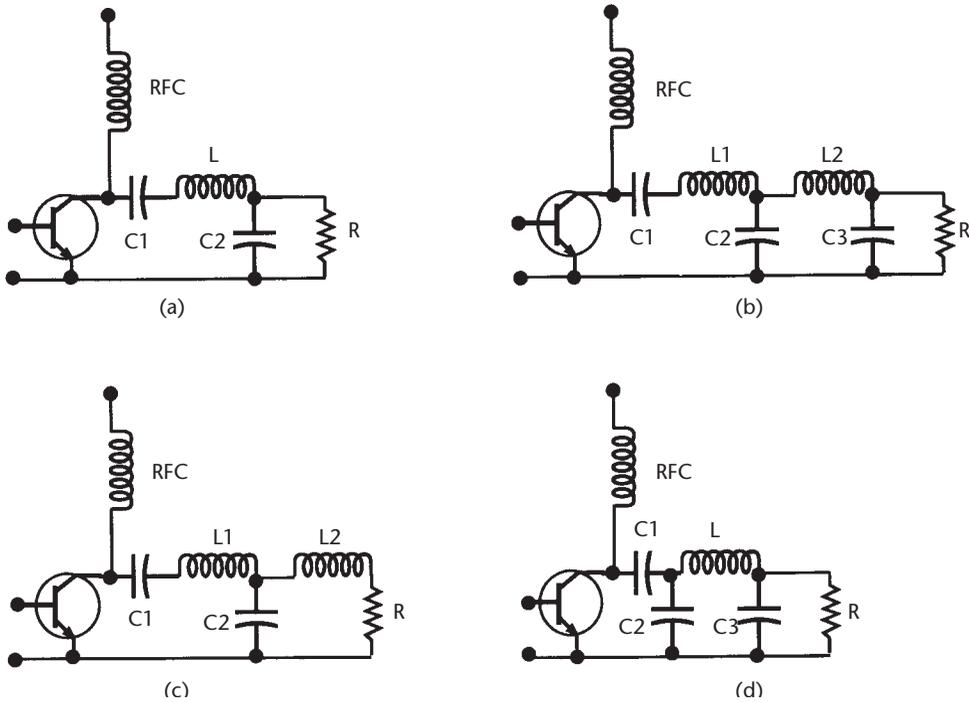


Figure 14.10 Examples of discrete impedance-matching networks: (a) single L-section; (b) double L-sections; (c) T-section; and (d) pi-section.

circuit. The load resistor, R_L , is connected across C_2 . This type of circuit is a wideband matching circuit and is not intended to be a resonant circuit. It is used in wideband amplifiers and other wideband systems. The single L-section has the advantage of simplicity. Its disadvantage is that it is not practical for all types of impedance combinations because of limited degrees of freedom in component choice.

The circuit in Figure 14.10(b) shows double L-sections for output impedance matching. Again, the source impedance is the output impedance of a transistor. The two L-sections are L_1 and C_2 and L_2 and C_3 . The load impedance is R_L . This type of circuit frequently is used with power amplifiers.

Another matching circuit sometimes used for output matching is the T-section shown in Figure 14.10(c). This network is used when a higher Q is required.

The circuit in Figure 14.10(d) is a pi network. This circuit frequently is used with power amplifiers and for impedance-matching of antennas.

Each of the circuits shown in Figure 14.10 also can be used as input matching circuits for transistors and other devices. Both the input and the output of transistor and tube amplifiers need to have impedance-matching networks if maximum gain and maximum power output are to be realized.

14.12 Impedance-Matching Design Using the Smith Chart

The design of impedance matching circuits frequently is done using S-parameters and the Smith chart. Example designs for L, T, and pi networks are shown in Figures

14.11, 14.12, and 14.13. Each of these designs is discussed briefly in the following paragraphs.

Figure 14.11 shows an example where the goal is to match the output of a transistor amplifier to a 50Ω transmission line using an L-section LC network. The normalized transistor output impedance is assumed to be $0.4 - j 0.4$ based on 50Ω , which is shown as point 1.

The selected starting point is point 1, located at $0.4 - j 0.4$ on the Smith chart. First, move to the right on the 0.4 resistance circle to point 2. In doing so, series inductance is being added. This point is chosen so an equal distance on a straight line through the center will transform to the 1.0 resistance circle. The inductive impedance required to move from point 1 to point 2 is $j 0.9$. The impedance at point 2 is $0.4 + j 0.5$.

Next, convert to an admittance chart by transferring an equal radial distance to point 3. This point has an admittance of $1 - j 1.2$. To provide an impedance match to the 50Ω load, a capacitive susceptance of $j 1.2$ must be added. That moves the point clockwise on the 1.0 conductance circle to point 4, which has an admittance of $1 + j 0$, or an impedance of $1 + j 0$, thus obtaining a perfect impedance match at that frequency.

Component values then can be determined by using the values of X and B that were selected. For example, if the operating radian frequency, ω , is 188 rps,

$$L_1 = Z_0 \times X_{L1} / \omega = 50 \times 0.9 / 10^8 = 0.45 \mu\text{H}$$

$$C_2 = 1/Z_0 \times B_{C2} / \omega = 0.02 \times 1.2 / 10^8 = 240 \text{ pF}$$

Figure 14.12 shows an example where the output of a transistor amplifier is to be matched to a 50Ω transmission line using a T-section LC network. The normalized transistor output impedance is assumed to be $0.15 - j 2.0$ based on 50Ω and is shown as point 1.

The starting point is point 1, located at $0.15 - j 2.0$ on the Smith chart. Next, move clockwise on the 0.15 resistance circle to point 2. This point is chosen so we will be in good position to reach the $R = 1$ circle after two transfers. To reach point 2, 0.78 normalized inductive reactance must be added.

Now, convert to an admittance chart and transfer through the center an equal radial distance to point 3 on the 0.1 admittance circle. The admittance at this point is $0.1 + j 0.8$. Now, move counterclockwise to point 4. This requires a B_c value of $-j 0.5$.

Next, convert to an impedance chart and transfer through the center to point 5. This point is on the $R = 1$ circle. Finally, add a series inductive impedance of $+j 3.0$. This represents a move clockwise to point 6, which is at an impedance of $1 + j 0$, creating an impedance match.

Component values then can be determined by using the values of X and B that were selected. For example, if the operating radian frequency, ω , is 18^8 rps,

$$L_1 = Z_0 \times X_{L1} / \omega = 50 \times 0.78 / 10^8 = 0.39 \mu\text{H}$$

$$C_2 = 1/Z_0 \times B_{C2} / \omega = 0.02 \times 0.5 / 10^8 = 100 \text{ pF}$$

$$L_2 = Z_0 \times X_{L2} / \omega = 50 \times 3.0 / 10^8 = 15 \mu\text{H}$$

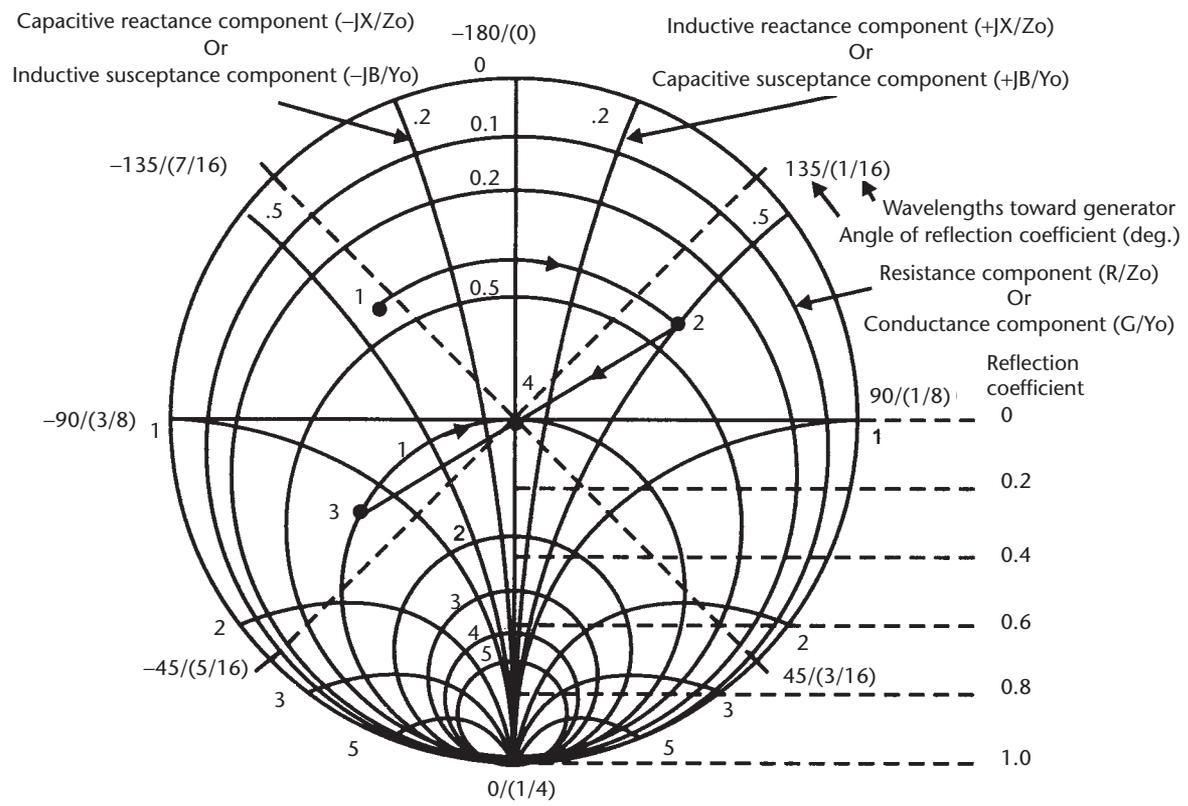


Figure 14.11 Example L-section impedance matching design using the Smith chart.

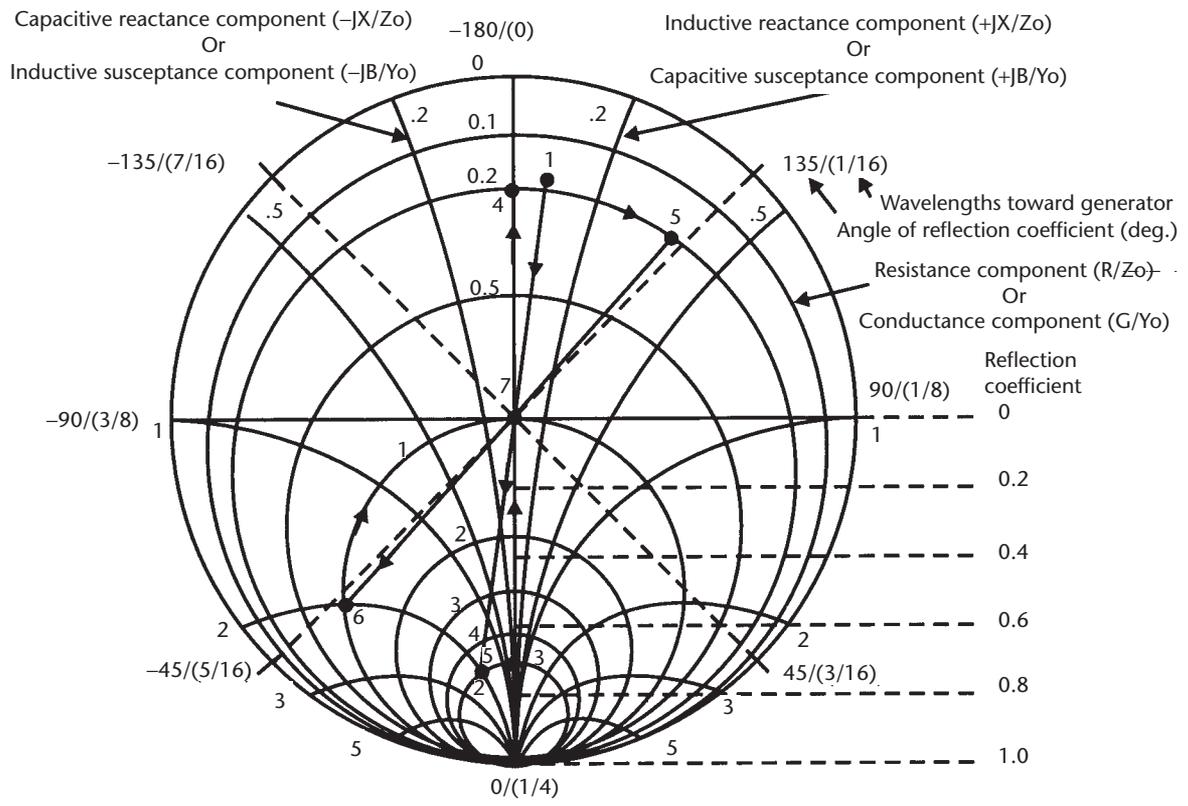


Figure 14.13 Example pi-section impedance-matching design using the Smith chart.

Figure 14.13 shows an example of matching the output of a transistor amplifier to a 50Ω transmission line using a pi-section LC network. The normalized transistor output impedance is assumed to be $0.17 + j 0.07$ in a 50Ω system, which is shown as point 1.

The selected starting point is point 1, located at $0.17 + j 0.07$ on the Smith chart. The first step is to convert to an admittance chart and transfer through the center to point 2. The admittance at this point is $5.0 - j 2.0$. Next, travel clockwise to point 3, which is reached by adding $-j 2.0$ capacitive susceptance.

Now, convert to an impedance chart by transferring through the center to point 4. The impedance at this point is $0.2 + j 0.0$. Then, add a series inductive reactance of $j 0.4$. to move clockwise to point 5. The impedance at this point is $0.2 + j 0.4$.

Next, convert to an admittance chart by transferring through the center to point 6. The admittance at this point is $1.0 - j 2.0$. Finally, move clockwise on the unity conductance circle to point 7, where the admittance is $1 + j 0$ and the impedance is also $1 + j 0$, forming a perfect impedance match.

Component values then can be determined by using the values of X and B that were selected. For example, if the operating radian frequency, ω , is again 18^8 rps,

$$C_2 = 1/Z_0 \times B_{C2} / \omega = 0.02 \times 2.0 / 10^8 = 400 \text{ pF}$$

$$L_1 = Z_0 \times X_{L1} / \omega = 50 \times 0.4 / 10^8 = 0.2 \text{ } \mu\text{H}$$

$$C_3 = 1/Z_0 \times B_{C3} / \omega = 0.02 \times 2.0 / 10^8 = 400 \text{ pF}$$

14.13 LC Filters

Some of the following material is derived from [4].

LC filters are used extensively in communication and radar systems. These filters sometimes are referred to as lumped constant or lumped-element filters. There are four main types of LC filters: bandpass filters, band-reject filters, lowpass filters, and highpass filters. Typical attenuation versus frequency plots for these filter types are shown in Figure 14.14.

In the case of the bandpass filter response shown in Figure 14.14(a), the attenuation or loss is a minimum at the center of design frequency and increases on either side. The difference between the top of the plot and the first dashed line is the insertion loss over the passband. A typical loss might be 0.5–1.0 at VHF. The next dashed line indicates the -3 -dB points on the attenuation curve. The difference between the two vertical lines is the -3 -dB bandwidth. This may be as small as 1% of the center frequency or as wide as an octave, depending on the design of the filter. The plot is in decibels and shows a rapid increase in attenuation for the filter on either side of the -3 -dB points. The slope of the curves depends on the number of sections or elements used in the filter.

The typical frequency response of the lowpass filter is shown in Figure 14.14(b). Attenuation is a minimum from dc to near some desired critical frequency. As the frequency approaches that frequency, the attenuation begins to increase. At the cut-off frequency, the added loss over insertion loss is 3 dB. Above that frequency, the

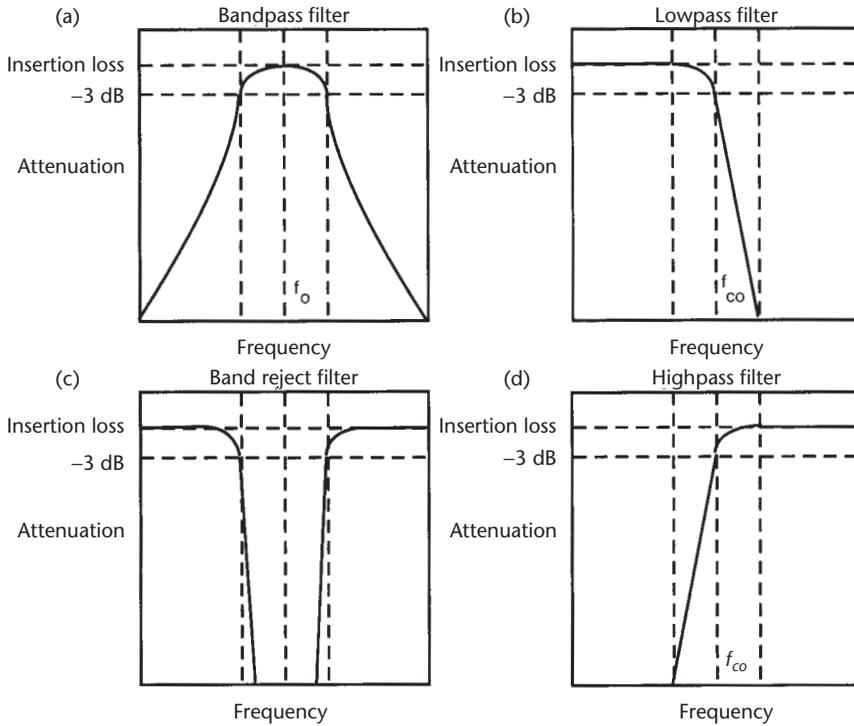


Figure 14.14 Attenuation versus frequency for LC filters: (a) bandpass filter; (b) lowpass filter; (c) band-reject filter; and (d) highpass filter. (After: [4].)

attenuation increases rapidly with the rate of increase, depending on the number of filter stages used and the ripple level for Chebyshev designs.

A Chebyshev filter is a filter that has an intentional small ripple in the passband. That allows use of designs with rapid increase in attenuation with change in frequency on either side of the -3 -dB passband. The other class of filter, known as a Butterworth filter, has a maximally flat response within the passband.

The typical response for the band-reject filter is shown in Figure 14.14(c). The response in this case is the opposite of that of the bandpass filter. The attenuation in this case is low at all frequencies except near a desired center frequency. As the frequency approaches the desired frequency, the attenuation becomes large and is maximum at the center frequency. Again, the second horizontal dashed line indicates the points on the curve where the attenuation in addition to the insertion loss has increased to 3 dB. The bandwidth of the band-reject filter is indicated by the vertical dashed lines. The bandwidth may be small or large, depending on the frequency to reject, the slope of attenuation versus frequency response, acceptable in-band loss, the ripple level, and so on.

The typical frequency response of the highpass filter is shown in Figure 14.14(d). In this case, the attenuation for the filter is very large from dc to a frequency near a selected cutoff frequency. Above that frequency, the attenuation decreases and becomes low for higher frequencies. Again, the slope of the response curve and the cutoff frequency depend on the design of the filter and the number of filter stages used.

The detailed design for multiple-stage LC filters is complex and beyond the scope of this book.

Figure 14.15 shows four examples of bandpass filter circuits made by K&L Microwave Inc., a Dover Technologies Company, 408 Cole Circle Salisbury, Maryland 21801. K&L uses these four circuits plus at least four other circuits for bandpass filters to meet design requirements. Obviously, the circuits shown are simplified and in general show only a fraction of the number of stages used (typically in the range of four to nine).

Figure 14.15(a) shows a two-stage resonant ladder bandpass filter. Each stage includes a series resonant LC circuit followed by a parallel resonant LC circuit. The output is from a tap between the two resonant circuits. A six-stage filter would contain six of these resonant circuit pairs. This circuit is used in HF wideband applications. Figure 14.15(b) shows a three-stage capacitively coupled “tank” circuits bandpass filter. The term tank circuit refers to a parallel resonant LC circuit. The output tap is between the coupling capacitor and the tank circuit. This circuit is an excellent structure for narrowband-use applications.

The circuit in Figure 14.15(c) is a cascade of highpass and lowpass filters. The lowpass filter is made up of an inductor followed by a capacitor with the output tap between the two components. The highpass filter is made up of a capacitor followed

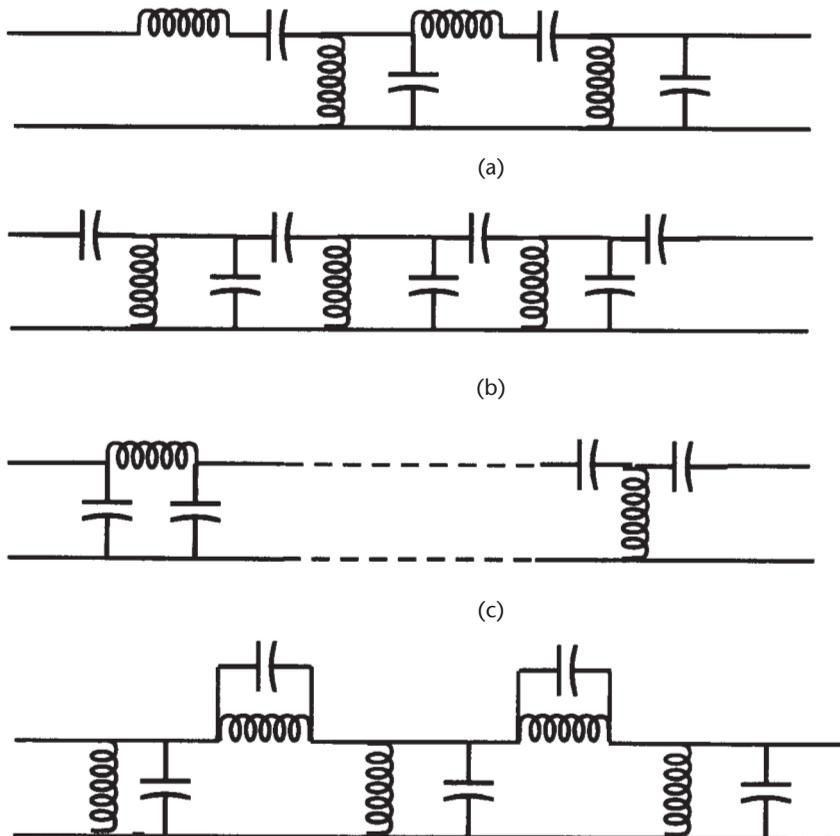


Figure 14.15 Examples of LC bandpass filters: (a) resonant ladder used in wideband applications; (b) capacitively coupled tank circuits; (c) cascade of highpass and lowpass filters; and (d) narrowband symmetrical Chebyshev structure. (After: [4].)

by an inductor with the output tap between the two components. A typical bandpass filter circuit might include four to six lowpass stages and four to six highpass stages. This circuit is well suited for bandwidths approaching an octave bandwidth or wider.

The circuit in Figure 14.15(d) shows a two-stage narrowband symmetrical Chebychev bandpass filter structure. Each stage consists of a parallel resonant tank circuit followed by a second parallel resonant tank circuit with the output tap located between the two resonant circuits. A typical bandpass filter may have four to eight such stages.

References

- [1] Van Valkenburg, M. E., (ed.), *Reference Data for Engineers*, 8th ed., SAMS, Carmel, IN: Prentice Hall Computer Publishing, 1993, Chap. 6.
- [2] Van Valkenburg, M. E., (ed.), *Reference Data for Engineers*, 8th ed., SAMS, Carmel, IN: Prentice Hall Computer Publishing, 1993, Chap. 5.
- [3] Krauss, H. L., C. W. Bostian, and F. H. Raab, *Solid State Radio Engineering*, New York: John Wiley & Sons, 1980, Chapter 3 and Appendix 3.1.
- [4] *K&L Microwave and RF Components Catalog*, pp. 4–12.

RF Transformer Devices and Circuits

This chapter discusses RF transformer devices and circuits. Topics include conventional transformers, core material for RF transformers, IF amplifier transformers, HF wideband conventional transformers, transmission-line transformers, and power combiners and splitters. Some of the information presented is based on [1–3].

15.1 Conventional Transformers

Figure 15.1 shows a conventional two-winding transformer with windings on a ferrite core. In an ideal transformer, the following relationships apply:

$$\text{Turns ratio} = a = N_1/N_2 = E_1/E_2 = I_2/I_1 \quad (15.1)$$

$$R_1 = a^2 R_2 \quad (15.2)$$

In practice in a real transformer, winding loss, leakage inductance, and magnetizing inductance must be taken into account. That may be done using the equivalent circuits in Figure 15.2.

Nomenclature for these circuits is as follows:

C_p = primary equivalent shunt capacitance

C_s = secondary equivalent shunt capacitance

E_g = rms generator voltage

E_{out} = rms output voltage

L_m = magnetizing inductance

L_p = primary leakage inductance

L_s = secondary leakage inductance

R_c = core-loss equivalent shunt resistance

R_g = generator resistance

R_L = load resistance

R_p = primary winding resistance

R_s = secondary winding resistance

The LF response of a conventional transformer is degraded by a shunt susceptance that appears across the primary winding. The magnetizing inductance, L_m , of the transformer is given by

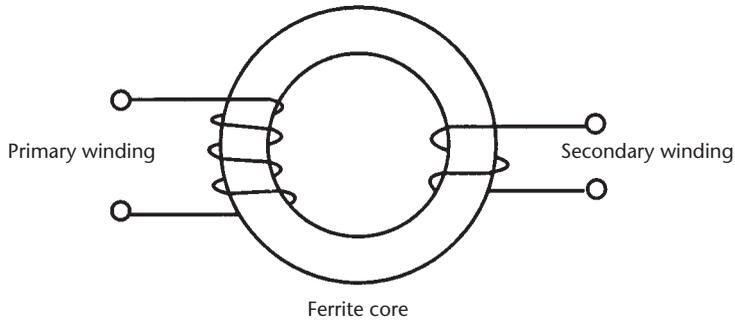


Figure 15.1 Conventional transformer with two windings.

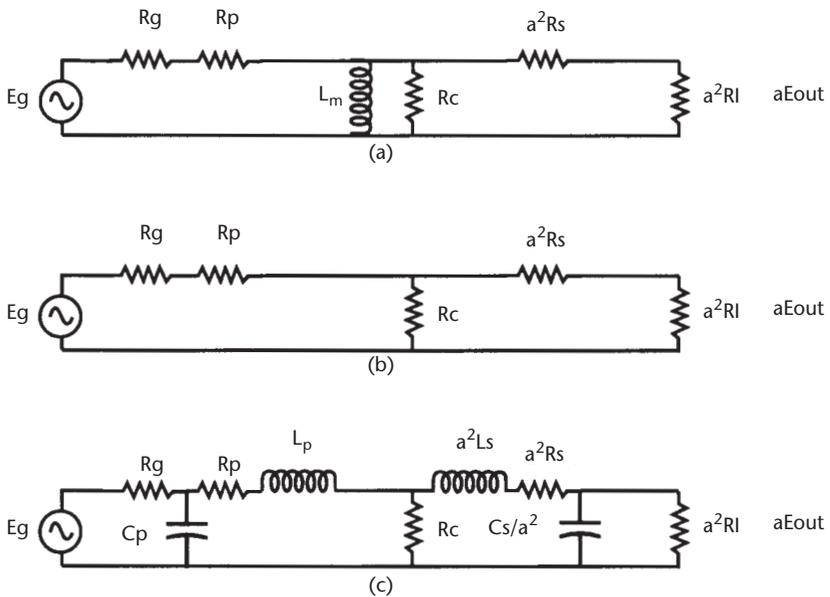


Figure 15.2 Equivalent circuits for conventional two-winding transformer: (a) low-frequencies equivalent circuit; (b) intermediate-frequencies equivalent circuit; and (c) high-frequencies equivalent circuit.

$$L_m = 4\pi n^2 \mu_r \mu_0 A_e / L_e \tag{15.3}$$

where:

n = number of turns in the primary winding

μ_r = relative permeability of the core

μ_0 = permeability of vacuum = $4\pi \times 10^{-7}$ H/m

A_e = effective area of the core (square meters)

L_e = average magnetic path length (meters)

For an example of the use of (15.3), assume the following:

$n = 4$

$$\mu_r = 100$$

$$\mu_0 = 4\pi \times 10^{-7} \text{ H/m}$$

$$A_e = 2 \text{ cm}^2 = 2 \times 10^{-4} \text{ m}^2$$

$$L_e = 6 \text{ cm} = 0.06 \text{ m}$$

Substituting those values into (15.3), we have

$$L_m = 4\pi \times 16 \times 100 \times 4\pi \times 10^{-7} \times 2 \times 10^{-4} / 6 \times 0.06$$

$$L_m = 8.4 \times 10^{-5} \text{ H} = 84 \mu\text{H}$$

The inductive shunt susceptance depends on the frequency. For example, at 200 kHz, the inductive reactance for L_m would be 105.5Ω and the susceptance would be 0.0095S . The LF cutoff would be at 200 kHz if the parallel resistance is also 105.5Ω .

The equivalent circuit shown in Figure 15.2(b) is for the case where the magnetizing inductive reactance is high, and the primary and secondary inductive reactances are low. This is the normal operating frequency band for the transformer.

The equivalent circuit shown in Figure 15.2(c) is for the case where the magnetizing inductive reactance is high, and the primary and secondary inductive reactances are also high. This is the upper cutoff band for the transformer.

Figure 15.3 shows the case of a conventional autotransformer. In this case, there is only one winding on the core instead of two. The input to the transformer is between one side of the coil and a tap point located between the ends of the coil. In the Figure 15.3, the tap is located at the center, which makes the primary winding have one-half the turns of the secondary winding. In the ideal case, the output has twice the voltage of the input and half the current. The output impedance is four times the input impedance. As we will see later, many of the RF transformers used are of this type. Basically, RF autotransformers can be compared to their LF counterpart, except that, with increasing frequency, leakage reactance becomes an important factor.

15.2 Magnetic Core Material for RF Transformers

Some type of magnetic core is required for the lower frequencies and for wideband RF transformers to extend coverage at the low end of the frequency band. Powdered

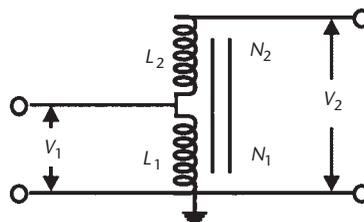


Figure 15.3 A conventional autotransformer.

iron often is used at the lower frequencies, while ferrites are the most common magnetic materials used for the higher frequency RF transformers.

Two main types of ferrites are used: nickel-manganese compositions and nickel-zinc compositions. Nickel-manganese compositions have higher permeabilities than nickel-zinc ferrites but larger losses at the higher frequencies. For that reason, nickel-zinc ferrites usually are used for the higher frequencies. One disadvantage of the nickel-zinc ferrites is that their Curie points can be as low as 130°C. The Curie point is the temperature at which damage may be done to the magnetic material by heat. Nickel-zinc ferrites can be manufactured only with relative permeability (μ_r) of less than about 1,000.

Because high-permeability ferrites saturate easier than low-permeability ones, it is good design practice to limit their maximum flux densities as shown in Table 15.1.

With RF transformers, either the primary or the secondary can be used for B_{\max} calculations, but the 50Ω side (if applicable) commonly is used for convenience and standardization. A general formula for calculating flux density for a ferrite core is given by

$$B = V_{rms} \times 10^8 / 4.44 fnA \quad (15.4)$$

where:

B = flux density (gauss)

V_{rms} = rms voltage on the winding

f = frequency (Hz)

A = core cross-sectional area (cm²)

n = number of turns

For an example of the use of (15.4), assume the following:

$V_{rms} = 50\Omega$

$f = 2.0$ MHz

$A = 1.0$ cm²

$n = 4$

Substituting those values into (15.4) gives

$$B = 50 \times 10^8 / (4.44 \times 2 \times 10^6 \times 4 \times 1)$$

$$B = 140.7\text{G}$$

Table 15.1 Maximum Flux Densities for High-Permeability Ferrites

μ_r	B_{\max} (G/cm ²)
400-800	40-60
100-400	60-90
<100	90-120

Equation (15.4) can be modified to show the required core area for a given flux density, as shown next:

$$A = V_{rms} \times 10^8 / 4.44 f n B \quad (15.5)$$

where the terms are as defined for (15.4).

15.3 Tuned Transformers

Tuned transformers, as used in IF amplifiers, are often constructed with adjustable rod-shaped cores, which can be moved in and out of the windings to change the value of the inductance, thereby changing the resonant frequency of the tuned circuit.

Figure 15.4 shows a transformer circuit with one side tuned. The circuit provides an alternative way to attain impedance matching. It can provide isolation between input and output circuits and introduce a phase reversal, if desired. The left-side coil, L_1 , is the primary winding of the transformer, and the right-side coil, L_2 , is the secondary winding. The coupling coefficient, k , is equal to the mutual inductance, M , divided by the square root of the product of L_1 and L_2 , as shown in (15.6).

$$k = M / (L_1 L_2)^{1/2} \quad (15.6)$$

Figure 15.5 shows the case of a double-tuned transformer. These circuits have been used extensively in receiver IF amplifier stages as the primary means of filtering signals. They permit flexibility in adjusting the shape of the selectivity curve. Although they now are being supplanted by ceramic, crystal, and surface acoustic wave filters, they still are used where widely different impedance levels must be matched and in some FM discriminators. Both single-tuned and the double-tuned transformer circuits are packaged inside small metal cans or shields. In some cases, the transformers are designed for operation at 455 kHz, as used in standard AM IF systems. In other cases, they are designed for operation at 10.7 MHz, as used in FM IF systems.

Figure 15.5(a) shows the double-tuned transformer circuit. Figure 15.5(b) shows the frequency response of the double-tuned transformer for three values of the transformer coupling coefficient, k . Those values are the critical coupling, the below-critical coupling, and the above-critical coupling. For k equal to or less than k_c , the response curve has a single peak. For k greater than k_c , the resonant curve has two peaks, as shown in Figure 15.5(b). This condition is called the overcoupled

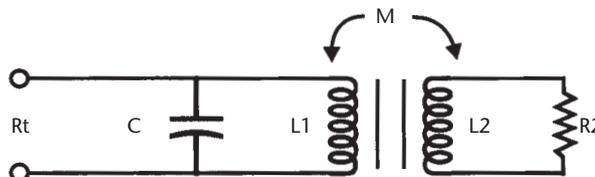


Figure 15.4 Single-tuned transformer circuit.

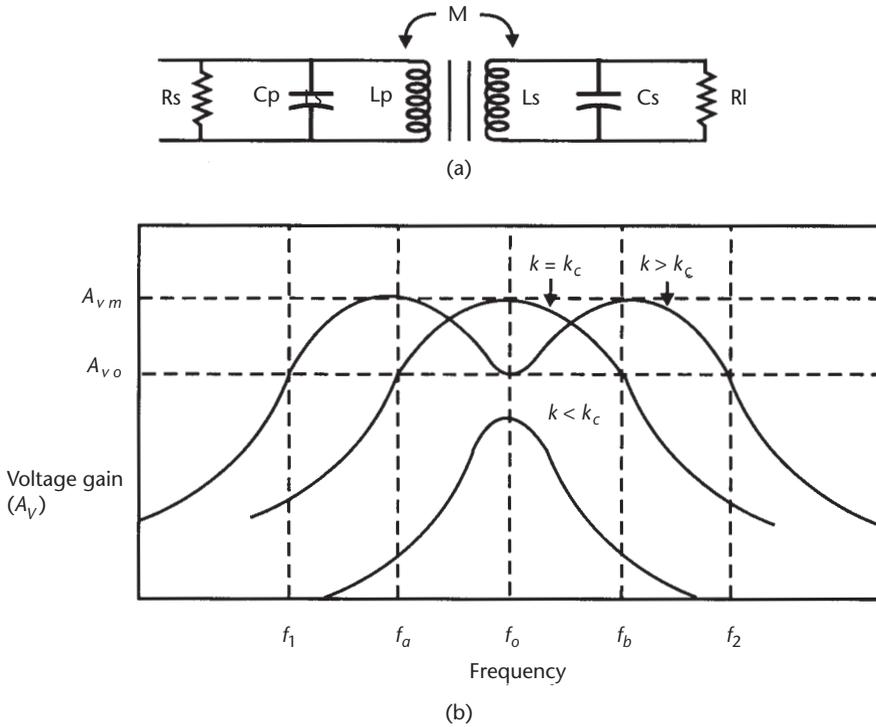


Figure 15.5 Double-tuned transformer: (a) transformer equivalent circuit and (b) frequency response of double-tuned transformer. (After: [1].)

case. The ratio of the peak gain to the gain at the valley at f_0 is given approximately by $0.5(kQ + 1/kQ)$ and is controlled by the choice of k and Q . The bandwidth of the circuit often is defined as $f_2 - f_1$, as shown in the figure.

For maximum power transfer at resonance, $R = (\omega_0 M)^2 / R_l$, or $\omega_0 M = R$. The circuit Q at resonance is defined by $Q = \omega_0 L / R$. It can be shown that the coefficient of coupling for the maximum power transfer condition (called critical coupling) is $k_c = 1/Q$.

It should be noted that if unequal impedances are to be matched on the two sides of the circuit, the primary or the secondary coil can be tapped.

Much of the foregoing discussion of tuned transformers is based on [1].

15.4 High-Frequency Wideband Conventional RF Transformers

High-frequency (HF), wideband RF transformers are used for the following functions:

- Impedance transformation;
- Baluns;
- Phase inversion;
- Power combining for push-pull amplifiers;
- Hybrid combiners for multiple power sources.

There are two main types of HF, wideband RF transformers: conventional transformers and transmission-line transformers. The conventional transformer is usually inferior in performance to a transmission-line transformer. The difference is mainly in the power-handling capability, loss factor, and bandwidth. The conventional transformer, however, can be constructed for a wider range of impedance ratios than the transmission-line transformer.

Figure 15.6 shows a popular, conventional RF transformer that is often used for HF and VHF applications. This transformer uses only one turn for the primary winding and two or more turns for the secondary. Its construction is two metal tubes with a connection between the tubes on one end, making a U-shaped single turn. The second set of wires are threaded through the two tubes to form a continuous multiturn winding. That results in a tight coupling between the two windings with relatively low mutual winding capacitance, thereby allowing its use at very low impedance levels. This transformer has the disadvantage that only integer-squared impedance ratios, such as 4:1, 9:1, and 16:1, are possible.

The bandwidth for the one-turn transformer is largely determined by the impedance ratio used. For example, a 9:1 impedance ratio transformer can have a bandwidth of up to about 60 MHz, a 25:1 transformer can have a bandwidth of up to about 30 MHz, and a 36:1 unit has a usable bandwidth of only about 15–20 MHz.

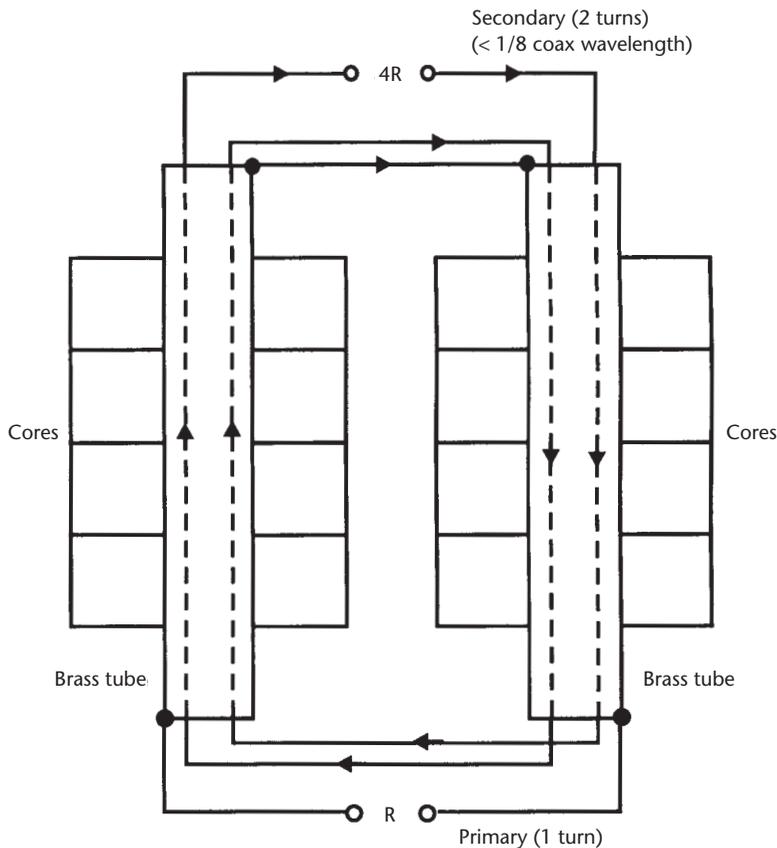


Figure 15.6 Conventional RF transformer using brass tubes. (After: [1].)

Figure 15.7 shows another variation of the conventional HF transformer. With this transformer, a U-shaped section of semirigid coax cable has two straight sections of semirigid coaxial cable soldered to it. The center wires of the coax cables are connected so they make two loops, as shown. This arrangement is mounted in a ferrite core consisting of an E and an I core joined. The transformer so produced has a 1:4 impedance ratio. The outer conductors at their open ends provide terminals for the single-turn winding, and the two coaxial center conductor ends provide terminals for the two-turn winding.

A 1:9 impedance ratio transformer of this type is shown in Figure 15.8. It is made with two U-shaped sections of semirigid coax cable plus two straight sections of semirigid coax cable soldered together to form a single U-shaped structure. The center wires of the coax cables are connected in series to form three loops. This concept can be extended to form four loop systems for a 1:16 impedance ratio transformer.

The types of transformers shown in Figures 15.6, 15.7, and 15.8 may or may not use ferrite cores to extend the LF coverage. At the higher frequencies, the core is of no value and would not be used. Transformers of the type shown in Figures 15.7 and 15.8 are usable up to about 300 MHz. The HF is limited by the fact that the total length of the high-impedance winding must be kept below about one-eighth of a wavelength at the highest frequency of operation to avoid major resonances. For operation at 200 MHz, the physical length of a U-shaped 4:1 unit is limited to about 3.5 cm. The length of a U-shaped 9:1 unit is limited to about 2.5 cm.

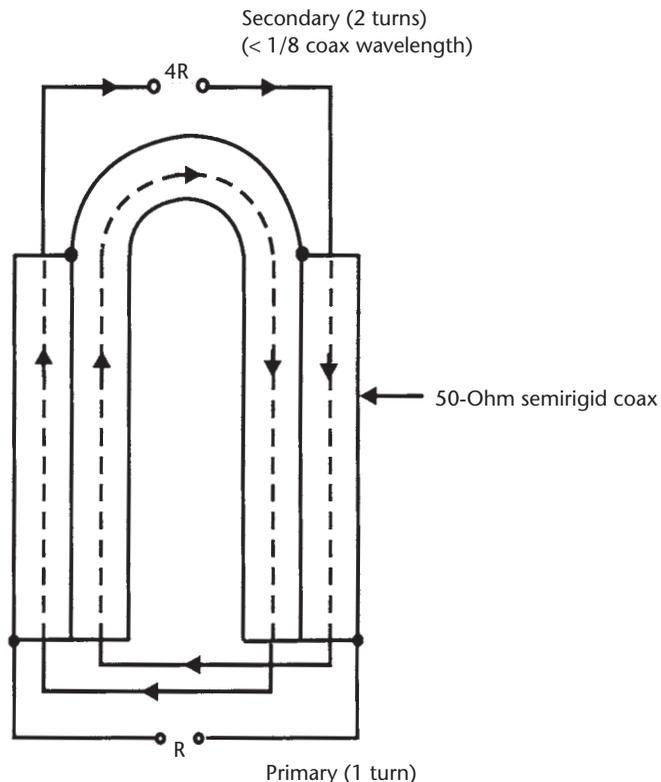


Figure 15.7 HF 4:1 transformer using coax cables. (After: [2].)

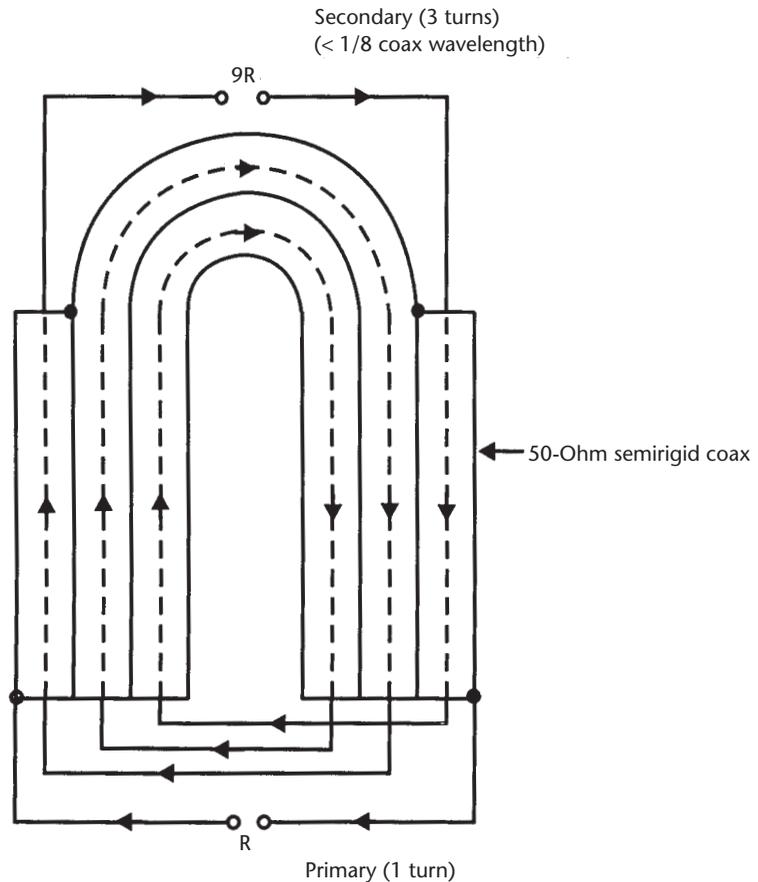


Figure 15.8 HF 9:1 transformer using coax cables. (After: [2].)

This type of transformer, having a typical length of about 3 cm, can be operated at frequencies as low as 3–10 MHz when a ferrite core is used. Without that core, the LF limit is about 100 MHz. The transformer has a typical power rating of 200–300W.

Another type of VHF transformer using semirigid coax is shown in Figure 15.9(a). This type of transformer often is used at frequencies as high as 1,000 MHz. No core is used at the higher frequencies. The operation of this transformer is as follows. A single-loop path is formed starting at terminal A and following lines 1, 2, 3, 4, and 5 to terminal B. A two-loop path is formed starting at terminal C and following lines 6, 7, 8, 5, 4, 3, 2, 1, 9, and 10 to terminal D. We thus have a 2:1 voltage transformation or a 4:1 impedance transformation transformer.

A 1:1 balun transformer of this type is shown in Figure 15.9(b). This balun transformer is used to convert from a coax unbalanced line to a balanced two-wire line. There are no loops in this case. The voltage out is equal to the voltage in.

The length of this type of transformer (balun) must be one-quarter guide wavelength in the coax. The coaxial cable impedance is $Z_0\Omega$ in and out. Because of the one-quarter wavelength property, this type of transformer is inherently a narrowband device, providing no isolation between the balanced and the unbalanced ends.

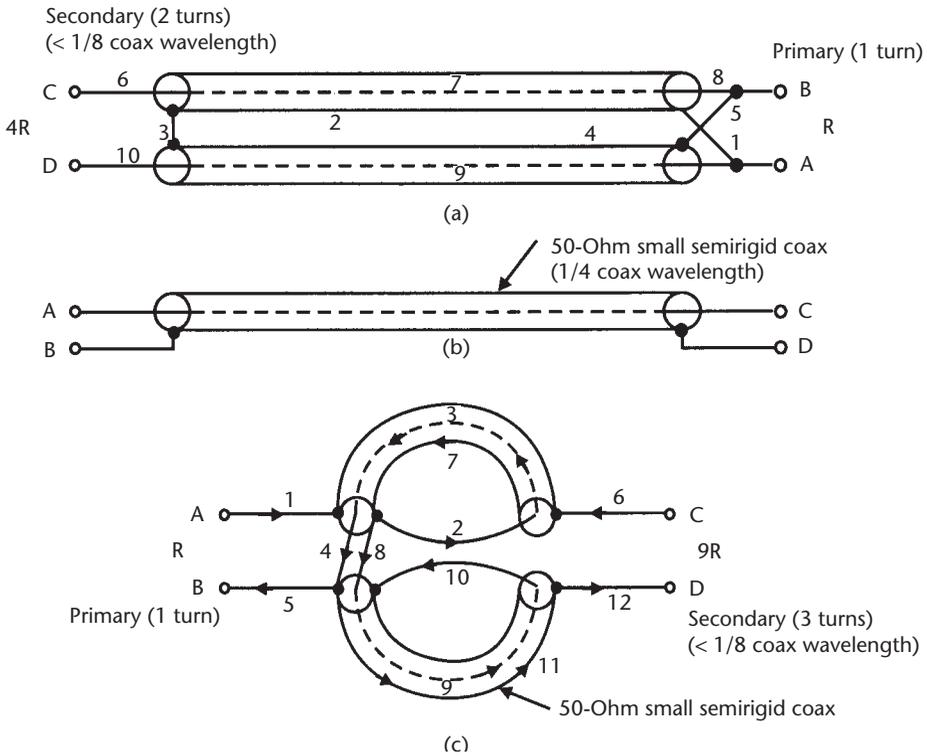


Figure 15.9 Coaxial transformers: (a) 4:1 coaxial transformer; (b) 1:1 coaxial balun transformer; and (c) 9:1 coaxial transformer.

Figure 15.9(c) shows the case of a 3:1 voltage transformer or a 9:1 impedance transformer of this type. A single-loop path is formed starting at A and following lines 1, 2, 3, 4, and 5 to B. A three-loop path is formed starting at C and following lines 6, 7, 8, 9, 10, 11, and 12 to D.

15.5 Transmission-Line Transformers

A number of RF applications require wideband transformers covering an octave or more, for example, hybrid combiners for wideband power amplifiers, hybrid power splitters, baluns, and impedance transformers. The type of transformers used are known as transmission-line transformers because they use transmission lines of either twisted-pair type or coaxial type with the characteristic impedance of the lines being an important part of the design. These transformers are not to be confused with the coaxial transformers discussed in Figure 15.4. Concepts for transmission-line transformers are shown in Figure 15.10.

To understand the bandwidth limitations inherent in conventional transformers, consider the conventional autotransformer shown in Figure 15.10(a). This unit is center tapped on the input to provide a two-to-one step-up in voltage and a four-to-one step-up in impedance. LF performance is degraded by the shunt susceptance of the windings, while HF performance is degraded by the series reactance of the windings.

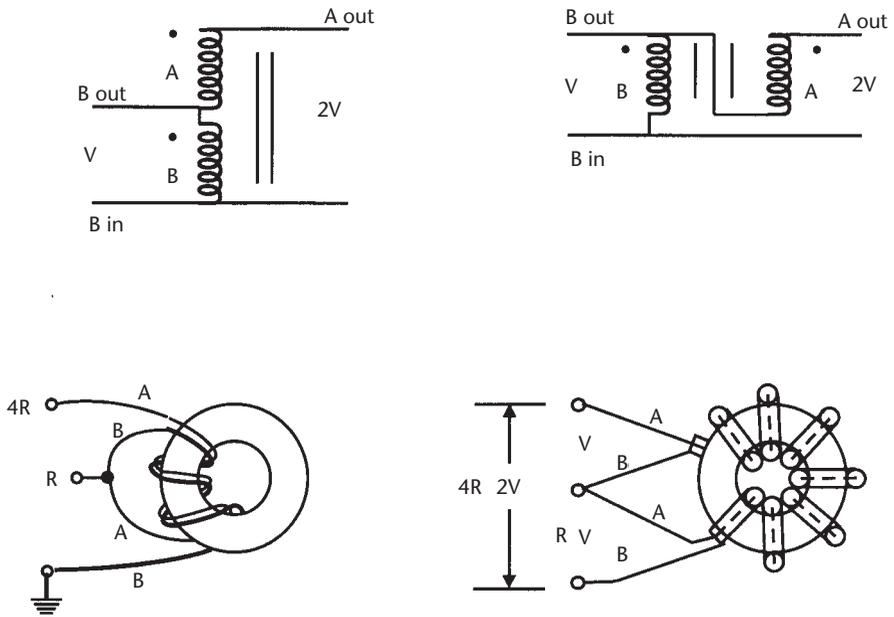


Figure 15.10 Concepts for transmission-line transformers: (a) autotransformer; (b) transmission-line transformer; (c) twisted-pair transmission-line transformer; and (d) coax transmission-line transformer. (After: [1].)

Figure 15.10(b) shows the autotransformer rearranged into a primary winding and a secondary winding with connection between windings. The impedance and voltage transformations are the same as in Figure 15.10(a). The two separate windings can be replaced by the two conductors in a single transmission-line, as shown in Figure 15.10(c). There, the transformer consists of a magnetic core with a two-wire, twisted-pair transmission line wound around the core. At low frequencies, the shunt susceptance degrades performance, as in the conventional autotransformer. At high frequencies, the circuit behaves as a transmission line, thereby greatly extending the HF limit for the transformer.

Notice the way in which the transmission lines are connected to provide the 4:1 impedance transformation. In the case of the twisted-pair transmission line in Figure 15.10(c), the two wires are labeled A and B. After looping around the core a number of times, the end of wire B is connected to the beginning of wire A. The low-impedance winding is the B-in to B-out connection. The high-impedance winding is the B-in to A-out connection. This winding has twice the path length and twice the number of turns as the low-impedance winding. Thus, the circuit is a 1:4 impedance ratio transformer.

Figure 15.10(d) illustrates a magnetic core with a coaxial transmission line wound around the core. The outer conductor is labeled B, and the inner conductor is labeled A. After looping around the core a number of times, the end of the outer conductor B is connected to the beginning of inner conductor A. The low-impedance winding is the B-in to B-out connection. The high-impedance winding is the B-in to A-out connection. This winding has twice the path length and twice the number of turns as the low-impedance winding. The circuit also behaves as a 1:4 impedance ratio transformer.

At low frequencies, the shunt susceptance degrades performance, as in the conventional autotransformer. At high frequencies, the circuit behaves as a transmission line, thereby greatly extending the HF limit for the transformer.

A number of possible transmission-line transformer configurations provide different impedance transformations. Four such circuits are shown in Figure 15.11: a balun with a 1:1 transformation ratio, a transformer with a 4:1 impedance ratio, a two-transformer combination with a 9:1 impedance ratio, and a three-transformer combination with a 16:1 impedance ratio. A second approach to providing a 16:1 impedance ratio is to use two 4:1 impedance transformers in series.

As in the case of the conventional transformer, the LF performance of the transmission-line transformer is determined by the ferrite core. With transmission-line transformers, the characteristic impedance of the transmission lines must be correct to take advantage of the optimum performance. At high frequencies, the series reactance combines with the interwinding capacitance, and the circuit behaves as a transmission line, greatly extending the HF response. The power transferred from the input to the output is coupled, not through the magnetic core except at very low frequencies, but rather through the dielectric medium separating the line conductors. It follows that a relatively small cross-section core can operate unsaturated at very high power levels.

The characteristic impedance of the line should be the geometric mean of the input and output impedances. For example, if the input impedance for the transmitter is 12.5Ω and the output impedance is 50Ω , the line characteristic impedance should be

$$Z_0 = (R_{in} \times R_{out})^{1/2} \tag{15.7}$$

$$Z_0 = (12.5 \times 50)^{1/2} = 25\Omega$$

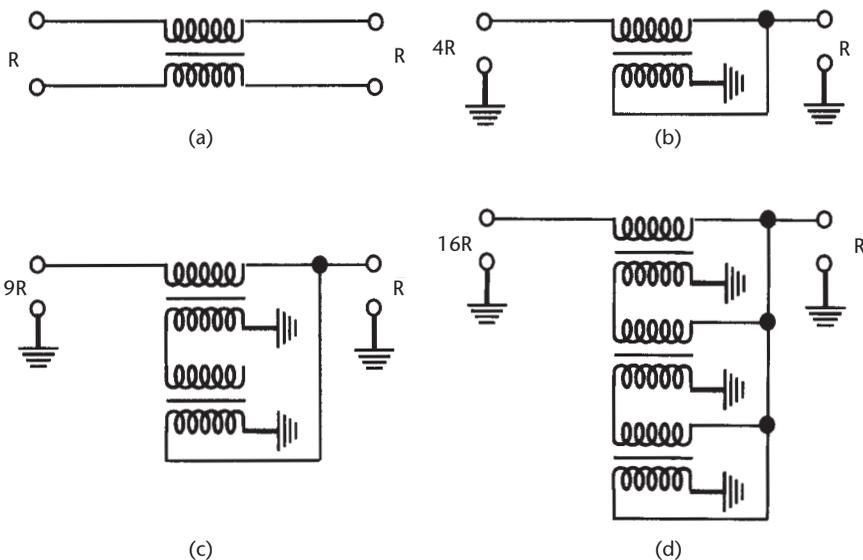


Figure 15.11 Transmission-line transformer configurations: (a) balun (1:1); (b) 4:1 transformer; (c) 9:1 transformer; and (d) 16:1 transformer. (After: [1].)

15.6 Power Combiners and Splitters

Hybrid coupler circuits can be used to combine two or more power amplifiers when a larger output is needed than can be obtained from a single amplifier. Direct parallel operation of the devices is usually unsatisfactory because the currents do not always divide equally among the devices.

Figure 15.12 shows hybrid couplers using both conventional transformers and transmission-line transformers. The basic hybrid combiner using a conventional center-tapped transformer, shown in Figure 15.12(a), is used to combine the outputs of two power amplifiers to achieve higher power.

The two amplifiers in Figure 15.12(a) are driven 180 degrees out of phase (push-pull) with respect to each other. The transformed sum of the two output currents is provided to the load R_o . The difference in the two output currents flows through the resistor R_h , provided for isolation purposes.

Figure 15.12(b) shows a hybrid splitter that uses a conventional center-tapped transformer. This type of circuit is used to drive two loads from a single source.

The circuit in Figure 15.12(c) is a wideband hybrid combiner that uses transmission-line transformers. By using many such combiners, it is possible to provide RF solid-state power amplifiers with power levels of more than a thousand watts at VHF and UHF frequencies.

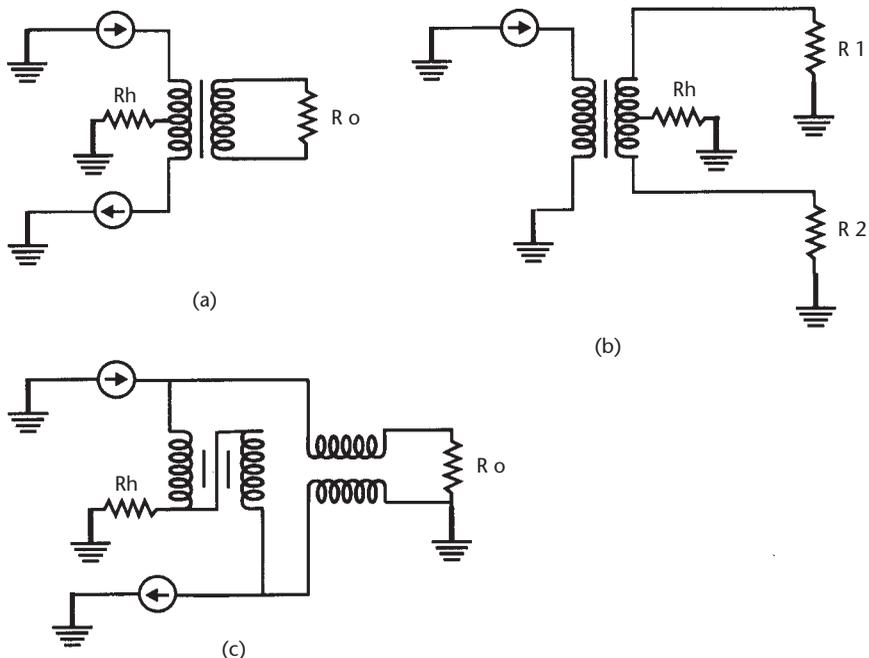


Figure 15.12 Hybrid couplers using transformers: (a) basic hybrid combiner; (b) hybrid splitter; and (c) broadband hybrid combiner. (After: [1].)

References

- [1] Krauss, H. L., C. W. Bostian, and F. H. Raab, *Solid State Radio Engineering*, New York: John Wiley & Sons, 1980, pp. 371–382.
- [2] Dye, N., and H. Granberg, *Radio Frequency Transistors: Principles and Practical Applications*, Boston, MA: Butterworth-Heinemann, 1993, Chap. 10.
- [3] Gottlieb, I. M., *Solid-State High-Frequency Power*, Reston, VA: Reston Publishing, 1982.
- [4] Van Valkenburg, M. E., (ed.), *Reference Data for Engineers*, 8th ed., SAMS, Carmel, IN: Prentice Hall Computer Publishing, 1993, Chap. 13.

Piezoelectric, Ferrimagnetic, and Acoustic Devices and Circuits

This chapter discusses piezoelectric, ferrimagnetic, and acoustic devices and circuits. Topics include quartz crystal oscillators, quartz crystal filters, ceramic filters, dielectric resonator oscillators, yttrium iron garnet (YIG) filters, YIG oscillators, surface acoustic wave (SAW) delay lines, and bulk acoustic wave (BAW) delay lines.

16.1 Quartz Crystal Resonators and Oscillators

Some of the material presented in this section is quoted or adapted from [1–3].

Quartz crystals are piezoelectric devices. The term piezoelectric refers to the property of a material in which mechanical deformation along one crystal axis results in an electrical potential or electric field along another axis. Conversely, an applied voltage deforms and produces mechanical stress in the crystal.

Quartz crystals are used extensively as the frequency reference or frequency controlling element for oscillators used in RF transmitters and receivers.

Figure 16.1(a) shows the symbol for the quartz crystal. Figure 16.1(b) shows the equivalent circuit for the crystal. We see that the crystal acts like a complex RLC resonant circuit.

Figure 16.2 shows an example of a terminal impedance characteristic for a quartz crystal. Notice that the crystal has two resonance frequencies. The first or lower frequency resonance is a series resonance in which the impedance is resistive and low. The second or higher frequency resonance is a parallel resonance in which the impedance is also resistive but very high. Between the two resonance frequencies, the impedance is inductive and has a positive phase angle near 90 degrees. Above the parallel resonance frequency, the impedance is capacitive and has a negative phase angle near -90 degrees.

Capacitors can be used for frequency correction or adjustment in conjunction with quartz crystals. The capacitors can be connected in either parallel or series and as either mechanically adjustable capacitors or electronically variable varactor diodes. Varactors also can be used for voltage-controlled frequency modulation of a crystal oscillator.

A quartz crystal can vibrate in a number of mechanical modes. The mode with the lowest resonance frequency is called the fundamental mode. Higher-order modes are called overtones or harmonics. Crystals intended for oscillator frequencies up to about 15 MHz normally operate in the fundamental mode and can be used in either series or parallel resonance. Above 15 MHz, an overtone normally is

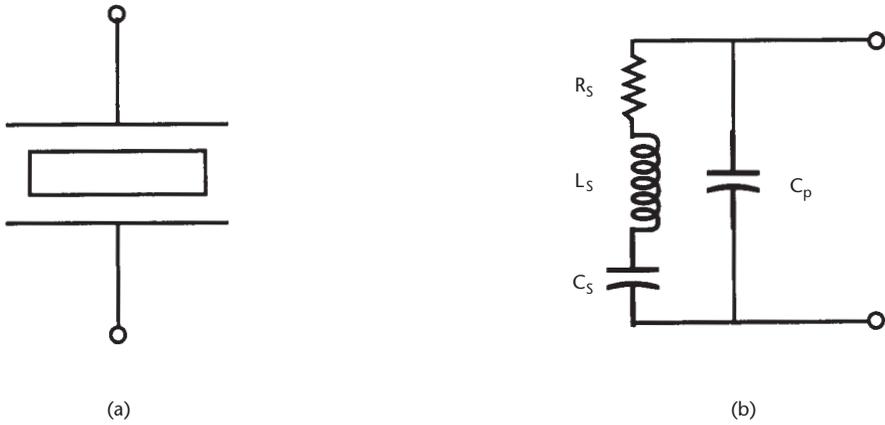


Figure 16.1 Quartz crystal: (a) crystal symbol and (b) equivalent circuit for a quartz crystal. (After: [2].)

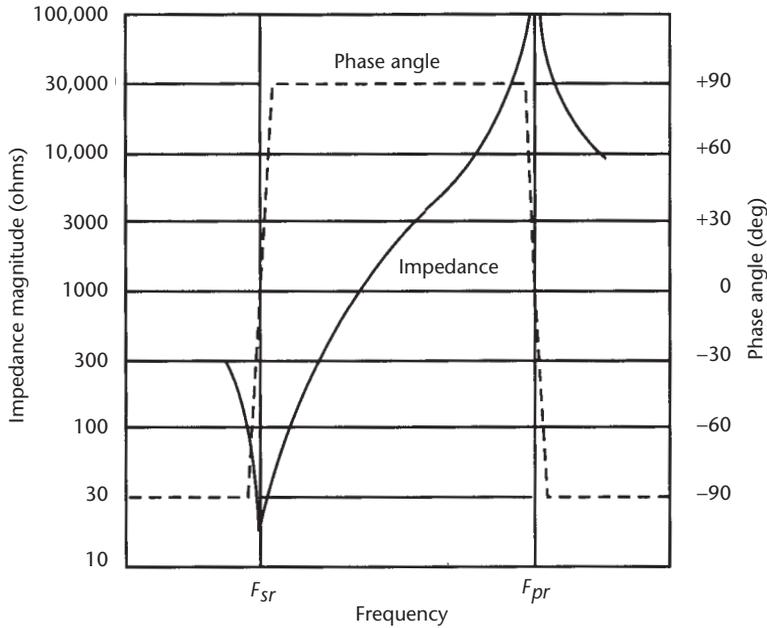


Figure 16.2 Example of a terminal impedance for a crystal (fundamental mode). (After: [2].)

utilized. For that condition, operation is always with in-series resonant mode. The highest frequency that can be obtained with an overtone crystal is about 180 MHz.

Quartz crystals have a number of different sizes and shapes and normally are encapsulated in metal holders. Many crystals for surface-mount construction are plastic encapsulated. Their size depends on operating frequency. A typical holder for a 5-MHz crystal has the dimensions $0.75 \times 0.75 \times 0.35$ inch. A typical size holder for a 15-MHz crystal has the dimensions $0.36 \times 0.43 \times 0.18$ inch, due to a smaller-size crystal being employed. For some crystals, pins are used as terminals. In other cases,

Table 16.1 TO-5 (HC-35) Packaged Crystals Made by CINOX

Crystal Part Number	Frequency Range (MHz)	Overtone	R_s (max) (Ω)	Operating Temperature ($^{\circ}\text{C}$)	Frequency Tolerance (ppm)
PC3125-0	30.0–70.0	3	40	25	± 5
PC5125-0	60.0–130.0	5	60	25	± 5
PC7125-0	130.0–170.0	7	150	25	± 5

flexible wire leads are provided. CINOX offers a complete line of TO-5 packaged crystals from 10 MHz to 180 MHz. Table 16.1 lists several characteristics of these crystals.

A temperature-compensated crystal oscillator is one to which a network has been added to vary the crystal's load impedance over temperature. The network typically consists of a thermistor-resistor network driving a varactor diode in series with the crystal. If the elements of the thermistor-resistor network are chosen properly, the output shift caused by the varactor diode will cancel almost exactly the crystal's temperature characteristic. The net result is a very small variation of crystal oscillator frequency with changes in temperature. For example, the frequency variation over a frequency range from -55°C to $+95^{\circ}\text{C}$ is less than 0.5 ppm.

An oven-controlled crystal oscillator is one that is operated inside a temperature-controlled oven. With such a system, it is possible to achieve extremely small frequency errors. Different oven and oscillator systems have different accuracies. The best unit reported by CINOX for operation in the 0 – 50°C range has a $\Delta f/f$ of 2×10^{-10} . Other units reported have $\Delta f/f$ values of 5×10^{-9} , 2×10^{-8} , and 1×10^{-7} . As always, better performance costs more.

Figures 16.3 and 16.4 show a number of examples of transistor oscillators using quartz crystals to determine the operating frequency. The circuit at Figure 16.3(a) is a Pierce oscillator. This is basically a common source Colpitts circuit with the crystal acting as an inductor and forming a resonant circuit with C_g , C_d , and the internal capacitances of the FET. Notice that in this case the crystal operates in the frequency range between the series resonance and the parallel resonance. The approach to designing a FET Pierce crystal oscillator often is a cut-and-try approach. Trim capacitors often are used for frequency adjustments.

Figure 16.3(b) shows a Miller oscillator. A Miller oscillator is similar to the tuned-input, tuned-output oscillator. Both the crystal and the output tank circuit (parallel RLC circuit) appear as inductive reactances at the oscillation frequency.

Figure 16.4(a) shows the Colpitts crystal oscillator with the crystal operating in the series resonant mode. At resonance, the crystal behaves as a small resistance. That serves as the necessary feedback from the collector of the transistor to the emitter at only the desired frequency. This circuit is particularly useful at higher frequencies, where series resonant, overtone crystals are normally used.

Figure 16.4(b) shows a second type of Colpitts crystal oscillator. The crystal is connected between the base and the ground. It also operates in a series resonant mode and grounds the transistor base at the operating frequency. In both of these Colpitts circuits, the transistor operates as a common-base amplifier, which requires that the base be ac grounded. Again, this type of circuit is useful at higher frequencies, where series resonant overtone crystals usually are utilized.

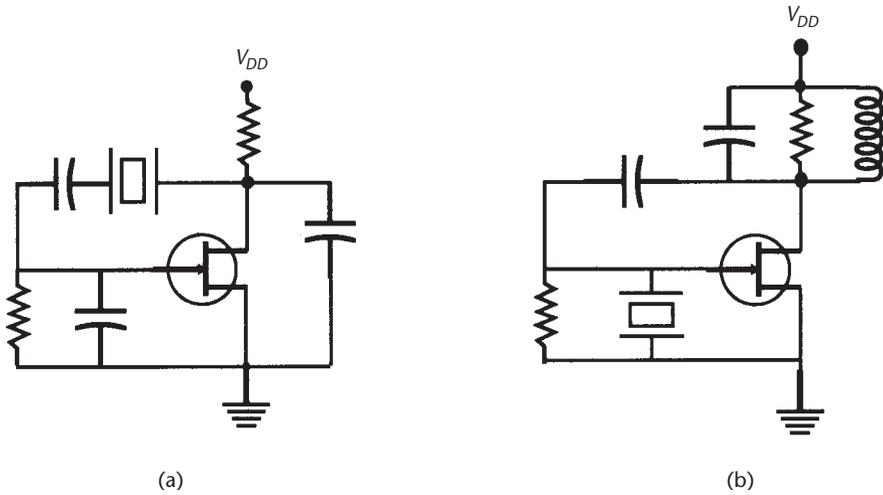


Figure 16.3 Examples of crystal oscillators: (a) Pierce crystal oscillator circuit and (b) Miller crystal oscillator circuit. (After: [2].)

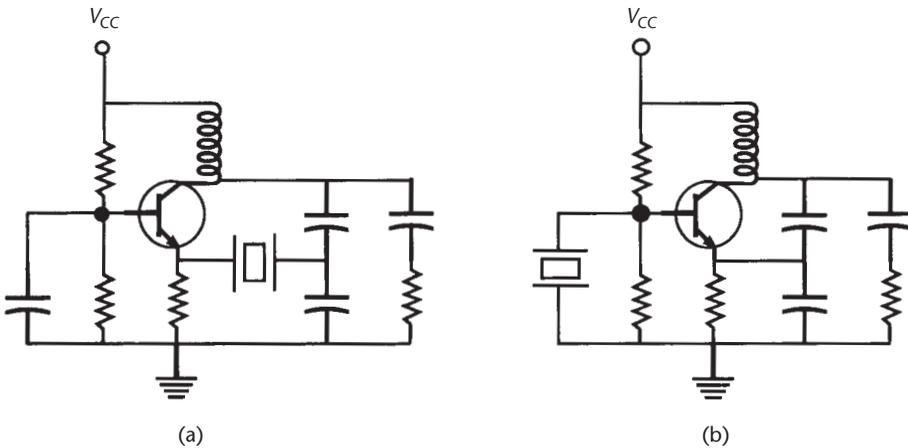


Figure 16.4 Two types of Colpitts crystal oscillators: (a) crystal in series resonant mode to provide feedback to emitter and (b) crystal in a series resonant mode to ground the base of the transistor. (After: [2].)

16.2 Monolithic Crystal Filters

Figure 16.5(a) illustrates a piezoelectric quartz crystal with three electrodes. The equivalent circuit for this filter is shown in Figure 16.5(b). Such a monolithic crystal can be used to fabricate a very high Q bandpass filter. The bandwidth of these filters is limited to a maximum of a few tenths of 1%. Their useful frequency range is about 5–350 MHz.

A typical frequency response of a four-pole monolithic crystal filter with center frequency at 75 MHz is shown in Figure 16.5(c). Quartz crystal filters have much higher Q values than ceramic filters and higher operating frequencies.

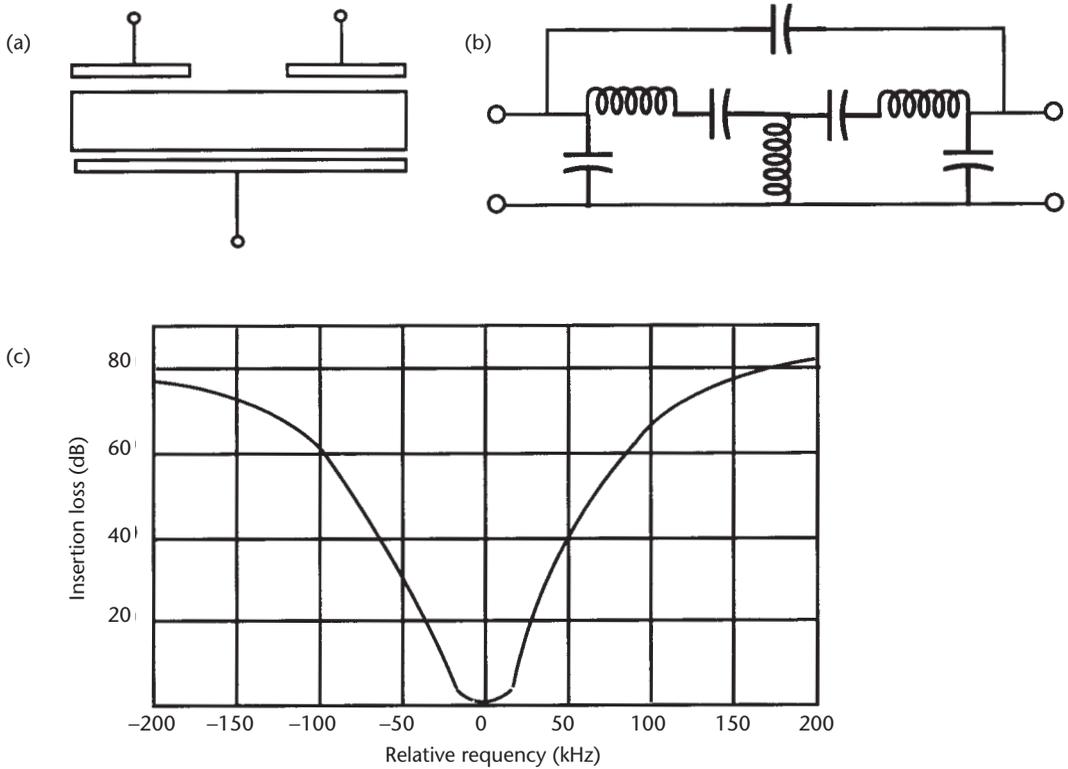


Figure 16.5 Characteristics for a monolithic crystal filter: (a) three-electrode monolithic crystal filter configuration; (b) approximate equivalent circuit; and (c) example frequency response for filter. (After: [5].)

16.3 Ceramic Filters

Some of the material presented in this section and Section 16.4 is quoted or adapted from [4–6].

Ceramic filters are made from piezoelectric ceramics. They are available with center frequencies ranging from a few kilohertz to more than 1.0 GHz. They have bandwidths ranging from 0.05–20%.

Figure 16.6 illustrates a ceramic disc resonator with two electrodes.

Figure 16.6(a) shows the ceramic disc resonator configuration. Figure 16.6(b) shows the equivalent circuit for this single resonator. Figure 16.6(c) shows the impedance magnitude as a function of frequency. Ceramic resonators have a series resonant frequency in which the impedance is low and a parallel resonant frequency in which the impedance is very high.

A three-electrode piezoelectric ceramic resonator is shown in Figure 16.7(a). Figure 16.7(b) shows the approximate equivalent circuit for this three-electrode ceramic resonator. Figure 16.7(c) shows its approximate frequency response. This type of device can be used as a bandpass filter for IF amplifiers operating at 455 kHz. They also can be used for this purpose at 10.7 MHz.

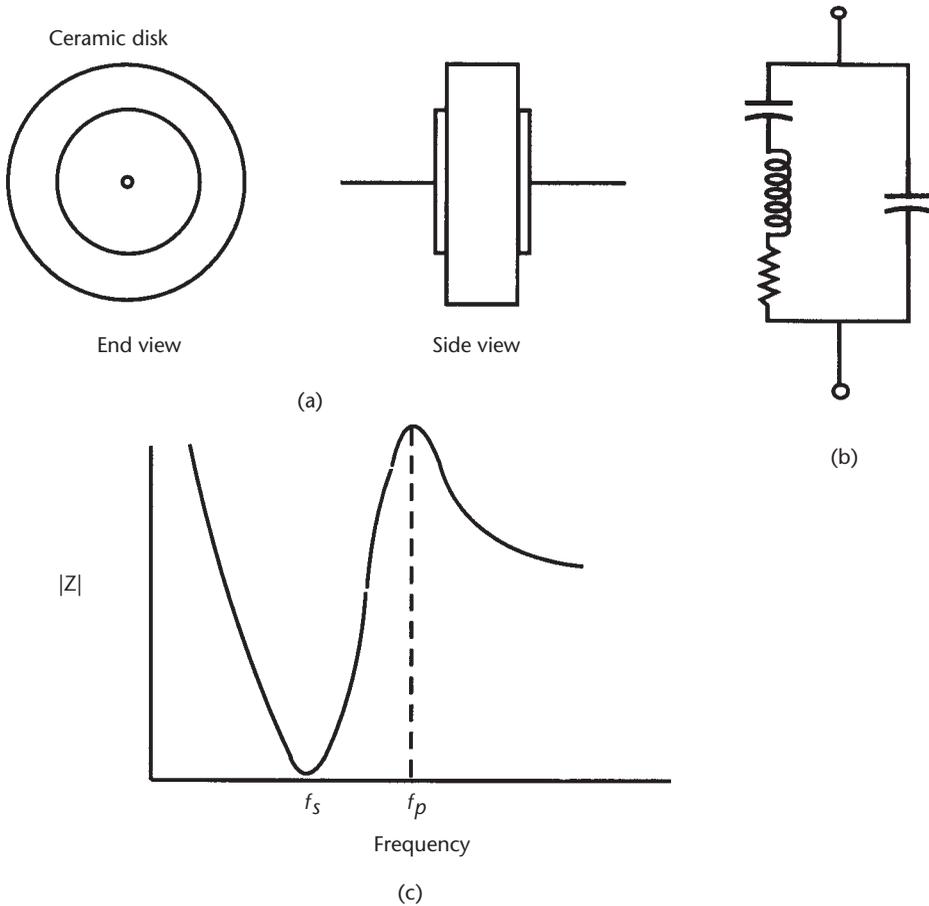


Figure 16.6 Ceramic disc resonator characteristics: (a) ceramic disc resonator configuration; (b) equivalent circuit; and (c) impedance magnitude versus frequency. (After: [5].)

16.4 Dielectric Resonant Oscillators

16.4.1 Dielectric Resonator Description and Parameters

A dielectric resonant oscillator is a free-running oscillator stabilized by the insertion of a high- Q dielectric resonator into the circuit. This dielectric resonator acts in many ways like an air-filled metallic cavity resonator with the advantage that it is much smaller in size. That is due to the higher dielectric constant of the titanate material used, compared to that of air. The wave velocity in the cavity is reduced by approximately the square root of the dielectric constant. The type of resonators used has dielectric constants of about 38, so dimensions for the cavity are reduced by a factor of about 6.2. The material it is made of is usually a barium titanate-based material. The unloaded Q of such a resonator is about 7,000 at 4 GHz.

Figure 16.8 shows one configuration used for the dielectric resonator. Example dimensions are given for the resonator as a function of frequency.

The height of the resonator, H , is one-half wavelength at the resonant frequency. For example, at 10 GHz, the wavelength in free space is 3 cm. The wavelength in the dielectric is that value divided by the square root of the dielectric

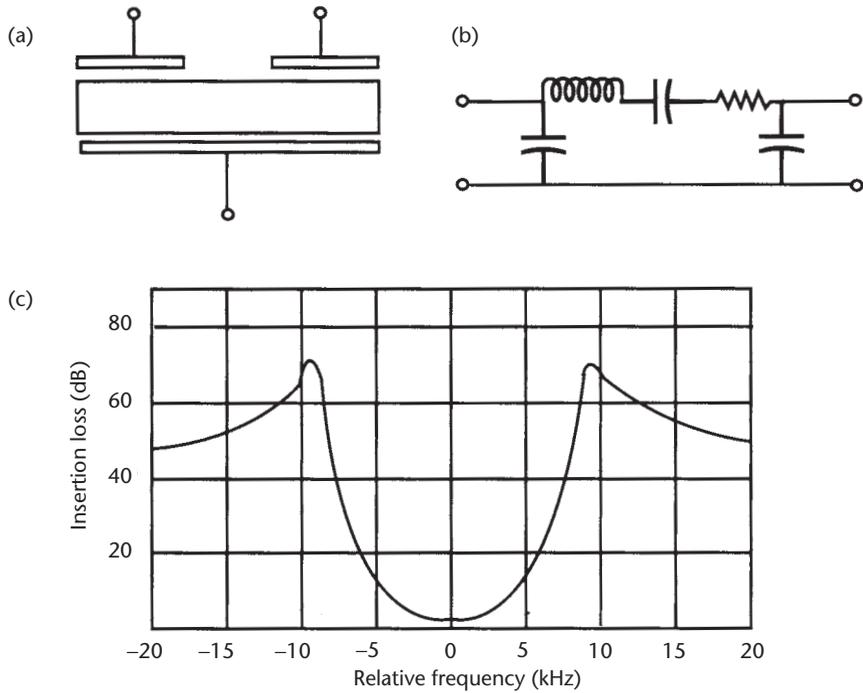


Figure 16.7 Characteristics for a ceramic IF amplifier filter: (a) three-electrode ceramic filter configuration; (b) approximate equivalent circuit; and (c) frequency response. (After: [5].)

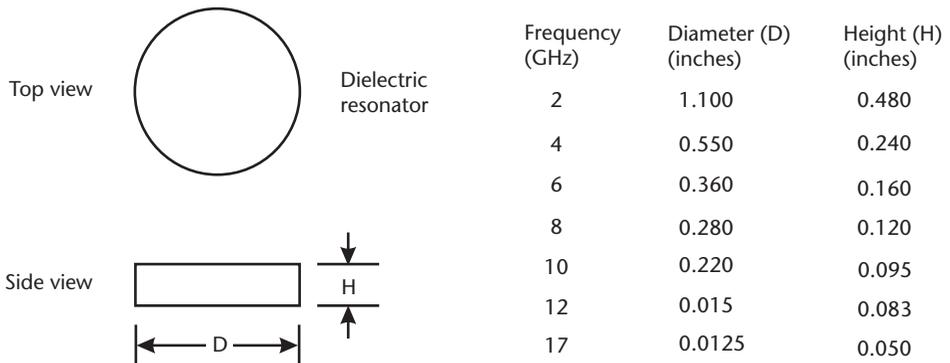


Figure 16.8 Dielectric resonator dimensions. (After: [6].)

constant. If we assume that the dielectric constant is 38, the wavelength in the dielectric at 10 GHz is 0.49 cm. One-half wavelength would be 0.24 cm, or 0.095 inch, as given in the table.

The diameter of the resonator is greater than one-half the cutoff wavelength. The cutoff wavelength in free space is given by

$$\lambda_0 = 2\pi r / (kr) \tag{16.1}$$

where

r = radius of resonator

(kr) = solution of a Bessel function equation

For the TE_{010} mode, $(kr) = 3.83$. Thus, the cutoff wavelength for this mode is

$$\lambda_{c0} = 2\pi r / (kr) = 1.6r = 0.8D$$

For example, at 10 GHz and a dielectric constant of 38, the cutoff wavelength is 0.49 cm = 0.19 inch and $D = 0.15$ inch. The required diameter to be well above cut-off is about 0.22 inch.

16.4.2 Coupling Between a Dielectric Resonator and a Microstrip Line

With a metallic cavity, coupling to the cavity is by means of a probe or a loop inserted into the inside of the cavity. With a dielectric resonator with no metallic walls, coupling usually is done using a microstrip line near the edge of the resonator. This configuration is shown in Figure 16.9(a).

In the Figure 16.9, only one edge coupling is used. In other cases, two or more microstrip lines are used for multiple-edge coupling.

The dielectric resonator oscillator (DRO) configuration shown in Figure 16.9(b) uses a metallic enclosure and a pedestal or spacer for the resonator. The spacer is added to improve the loaded Q by optimizing the coupling. The lateral distance between the resonator and the microstrip conductor primarily determines the amount of coupling between the resonator and the microstrip line. The metallic shielding is required to minimize radiation losses and reduce unwanted stray coupling to adjacent circuitry.

The dielectric resonator placed adjacent to the microstrip line operates like a reaction cavity that reflects the RF energy at the resonant frequency. It is similar to an open circuit with a voltage maximum at the reference plane at the resonant frequency.

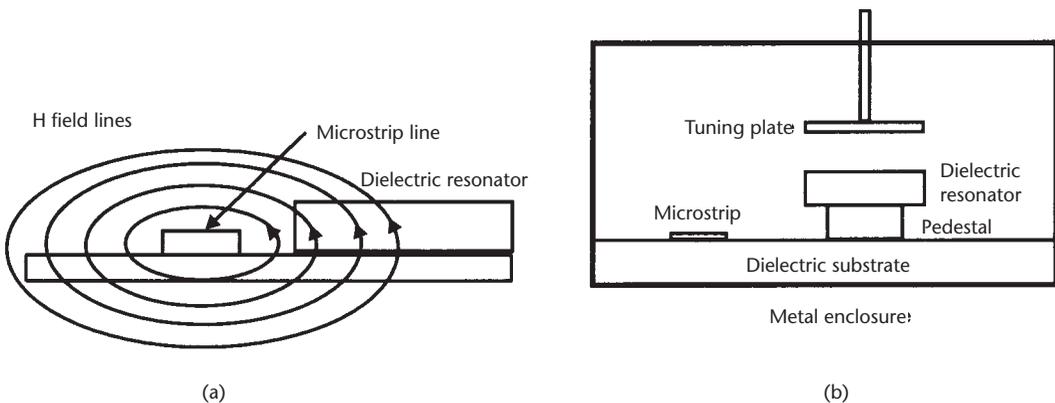


Figure 16.9 Dielectric resonator oscillator coupling: (a) edge coupling of a TE_{01} mode between the dielectric resonator and a microstrip line and (b) example of dielectric resonator oscillator configuration using a metallic enclosure and a pedestal or spacer for the resonator. (After: [6].)

16.4.3 Mechanical and Electrical Tuning of Dielectric Resonators

It is possible to tune the dielectric resonator by providing mechanical tuning, electrical tuning, or a combination of both. Mechanical tuning involves use of a capacitance plate that can be moved with respect to the resonator. Electrical tuning involves the use of a voltage variable capacitor (varactor diode) coupled to the resonator with a microstrip line. Mechanical tuning is illustrated in Figure 16.9(b).

16.4.4 Examples of Dielectrically Stabilized Oscillators

Examples of dielectrically stabilized oscillators are produced by Anzac, a division of Adams-Russell Corporation, Inc., in Burlington, Massachusetts. This company offers dielectric resonant oscillators from 2.65 GHz (dimensions $3.5 \times 1.92 \times 1.66$ inches) to 12 GHz (dimensions $1.88 \times 1.15 \times 0.77$ inch). Output powers are 10 dBm minimum. Mechanical tuning ranges from ± 5 MHz at 2.65 GHz to ± 15 MHz at 12 GHz. Harmonics typically are -25 dBc, and spurious signals typically are -90 dBc. Phase noise typically is -95 dBc/Hz at $f_0 \pm 10$ kHz. These units use SMA connectors for RF and solder feedthrough for dc.

16.5 YIG Resonators and Filters

Some of the material presented in this section is quoted or adapted from [7].

16.5.1 Ferrimagnetic Resonance in Yttrium Iron Garnet Crystals

One of the most interesting types of crystals is the YIG crystal. This crystal is a magnetic insulator that resonates at a microwave frequency when magnetized by a suitable dc magnetic field. A unique feature of a YIG crystal is that for a spherical configuration the resonant frequency is related to only the direct magnetic field, not to its dimensions. YIG resonators are thus constructed as small, highly polished spheres with diameters between 1.0 mm and 2.0 mm. YIG crystals of this type are used in magnetic-field-controlled bandpass filters, bandstop filters, oscillators, limiters, discriminators, and numerous microwave systems. The useful frequency range for such systems is about 500 MHz to about 40 GHz.

The basic ferrimagnetic resonance phenomenon in a YIG crystal can be explained in terms of spinning electrons, which create a net magnetic moment in each molecule of the crystal, illustrated in Figure 16.10(a). Application of a biasing magnetic field causes the magnetic dipoles to align themselves in the direction of the magnetic field, thus producing a strong net magnetization. Any microwave magnetic field at right angles to the dc magnetic field results in precession of the magnetic dipoles around the biasing field. If the frequency of the microwave field coincides with the natural precessional frequency of the YIG crystal, strong interaction results.

For spherical resonators, the resonant frequency, f_0 , is given by

$$f_0 = \gamma(H_0 + H_a) \quad (16.2)$$

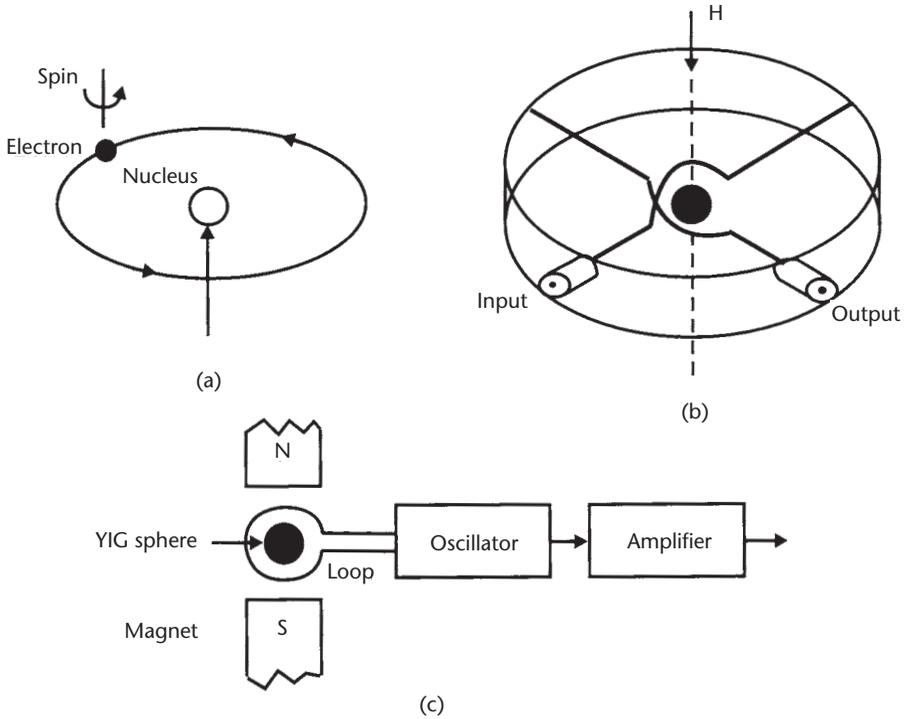


Figure 16.10 YIG crystal concepts and applications: (a) spin motion in YIG resonators; (b) YIG semiloop coupling structure; and (c) schematic diagram of YIG-tuned oscillator. (After: [7].)

where:

$$\begin{aligned} \gamma &= \text{charge-mass ratio of an electron or the gyromagnetic ratio} \\ &= 2.21 \times 10^5 \text{ (rad/s)/(A/m)} \end{aligned}$$

$$H_0 = \text{applied direct field (A/m)}$$

$$H_a = \text{internal crystal anisotropy field}$$

H_a usually is very small compared to H_0 and so is neglected in the following example.

For an example of the use of (16.2), assume $H_0 = 85.3 \text{ kA/m}$. Thus, the resonant frequency is

$$\begin{aligned} f_0 &= \gamma(H_0 + H_a) \\ f_0 &= 2.21 \times 10^5 \times 85.3 \times 10^3 = 19 \times 10^{10} \text{ rad/s} = 3.0 \text{ GHz} \end{aligned}$$

Thus, as shown in (16.2), the resonant frequency is directly proportional to the applied direct magnetic field. The upper frequency limit is determined by the difficulty of establishing the required direct magnetic field with reasonable power supply specifications.

The unloaded Q for the YIG sphere is given by

$$Q_U = H_0 / \Delta H \tag{16.3}$$

where:

H_0 = applied direct field (A/m)

ΔH = uniform mode line width = 40 A/m

For this example, the calculated unloaded Q is 4,600, which is as good as can be achieved with a conventional metallic cavity resonator at the same frequency. The advantage for the YIG resonator is that it can be tuned quickly by electrical means and over a large frequency range. Tuning time is not as fast as that of a varactor. Usually YIG systems require a heater to be built into the structure to maintain the YIG sphere at a constant temperature.

16.5.2 YIG Bandpass Filters

Figure 16.10(b) shows one version of a YIG bandpass filter using semiloop wires. The YIG sphere usually is mounted on a ceramic rod. The two semiloops are at right angles to each other and to the H_0 field. This field is supplied by a magnet, which is not shown in Figure 16.10.

In the absence of the direct magnetic field, the two circuits are decoupled and no transmission takes place between the input and the output terminals. When the magnetic field is applied, the microwave magnetic field at right angles to the direct field results in precession of the magnetic dipoles around the biasing field. The microwave magnetic field is applied by the input semiloop. Because of the precession of the magnetic dipoles around the biasing field, there is strong coupling to the output semiloop. That produces a bandpass filter structure with strong coupling in only one narrow frequency band. The center frequency can be shifted by changing the direct magnetic field by changing the input current.

Many other coupling structures can be used with YIG bandpass filters, including full-loop structures, stripline structures, coax line structures, and waveguide structures. There are single-sphere, two-sphere, and four-sphere filters.

16.5.3 YIG-Tuned Oscillators

Many possible oscillator configurations can use YIG resonators to control the frequency of oscillation. Figure 16.10(c) shows a diagram of a YIG-tuned oscillator. The magnet that supplies the direct field is shown. The loop about the YIG sphere is connected to an active device such as a BJT or FET oscillator or a Gunn-diode oscillator. (These types of oscillators are discussed in Chapter 17.)

16.6 Surface Acoustic Wave Delay Lines

Some of the material presented in this section is quoted or adapted from [8, 9].

16.6.1 Nondispersive Delay Lines

SAW delay lines are made using single large crystals of quartz, lithium niobate, lithium tantalate, and other piezoelectric materials. Such devices provide accurate

wideband delays in a very small volume. That is because the velocity of the acoustic wave generated with electrical transducers is approximately 100,000 times slower than that of an electromagnetic wave. The useful frequency range for SAW delay lines is about 10 MHz to 1,600 MHz at this time. The upper frequency limitation and achievable bandwidths are due to fabrication limitations for the transducers used to launch the acoustic waves. Table 16.2 lists some SAW delay line material parameters.

The simplest SAW delay line configuration uses an input interdigital transducer (IDT) and an output IDT that are separated on the crystal surface by some distance that determines the amount of the time delay. This type of configuration is shown in Figure 16.11.

The SAW is launched by application of an RF signal to the interdigital electrode transducer on the left. In a nondispersive delay line, the fingers of the transducer on a given side are a full wavelength apart at the acoustic wave velocity. Adjacent fingers are a half-wavelength apart. The presence of the electric field between the fingers causes mechanical stress in the piezoelectric crystal surface, and an acoustic wave is launched in both directions. The wave moving to the left is absorbed by the left absorber element, thus preventing reflections. The wave moving to the right travels across the highly polished piezoelectric surface to the receiving transducer. The acoustic wave is accompanied by an electric field. As the wave passes under the

Table 16.2 SAW Delay Line Material Parameters

<i>Material</i>	<i>Surface Wave Velocity</i>	<i>Temperature Dependency</i>	<i>Optimum Fractional Bandwidth</i>
ST quartz	0.124 inch/ μ s		0.1–0.5%
Lithium tantalite	0.129 inch/ μ s	–23 ppm/ $^{\circ}$ C	4–9%
YZ lithium niobate	0.134 inch/ μ s	–94 ppm/ $^{\circ}$ C	7–30%
Lithium niobate	0.153 inch/ μ s	–72 ppm/ $^{\circ}$ C	15–67%

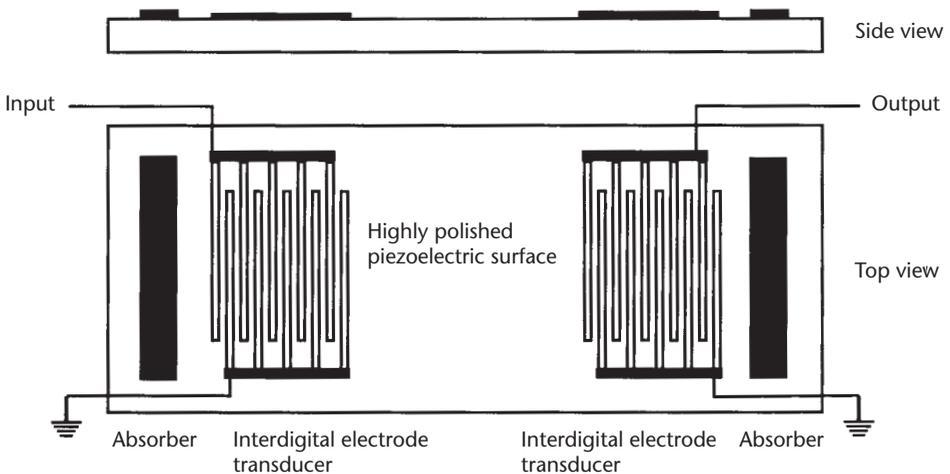


Figure 16.11 SAW in-line transversal filter or delay line structure. (After: [9].)

receiving transducer fingers, the electric field induces a delayed signal voltage, which is then fed out of the system. The acoustic wave that reaches the absorber element at the right is absorbed, thus preventing reflections.

The term nondispersive delay line means that the delay is constant and independent of frequency.

A dispersive delay line, on the other hand, means that the delay is a function of the frequency. Nondispersive delay lines are used in oscillators, frequency discriminator circuits, filters, and other signal processing applications. In filter configurations, the finger length usually is not uniform. The spacing determines the wavelength of the acoustic wave that is preferentially excited, the finger overlap determines the strength of each source of the waves and determines the shape of the filter response, and the number of fingers determines the bandwidth. An approximate equation for bandwidth for the SAW filter is

$$BW = f/N$$

where:

f = center frequency

N = number of finger-pairs

For example, if 20 finger pairs are used, the bandwidth of the filter is about 5%. Five finger pairs are shown in each of the transducers shown in Figure 16.11. The launching transducer usually has only a few fingers compared to the receiving transducer, which may contain many fingers.

The insertion loss of a SAW bandpass filter is in the range of 10–30 dB, 6 dB of which is due to the bidirectional transducers. That is somewhat larger than the loss of ceramic or crystal filters. However, the wider bandwidth capability and the ability to shape the transfer function of the filter make the SAW filter attractive for many receiver applications. The filters are now used in televisions, cellular telephones, radars, and so on, for highly selective filtering.

Nondispersive IDTs are used when relatively narrow bandwidths (<30%) are required, while dispersive IDTs allow the implementation of very large (<67%) fractional bandwidths with relatively low loss.

Although most nondispersive delay lines are less than 20 μ s in length, it is possible to produce delays of up to 150 μ s in an area smaller than 2 \times 2 inches. A minimum delay of 250 ns is recommended, even for very short delay lines.

Table 16.3 shows performance parameters for nondispersive delay lines made by Sawtek, Inc., in Orlando, Florida.

16.6.2 Tapped Delay Lines

The design of the basic delay line can be extended to include many output transducers at different locations along the piezoelectric substrate, each having a different delay to realize a tapped delay line designed for a particular PSK code. It is possible to implement correlators for biphase, quadriphase, and MSK.

A PSK device can serve as an expander that elongates a short impulse into a coded waveform with uniform amplitude over its time duration and as a

Table 16.3 Performance Parameters for Nondispersive Delay Lines

<i>Parameters</i>	<i>Values</i>
Center frequency	10 to 1,600 MHz
Fractional bandwidth	2 to 40%
Delay	0.25 to 150 μ s
Insertion loss	10 to 35 dB
Amplitude ripple	0.1 to 1.0 dB
Phase ripple	1 to 10 degrees
Group delay ripple	10 to 250 ns
umTriple-transit suppression	30 to 60 dB
Spurious suppression	40 to 70 dB

compressor that shortens the coded signal from an expander in a time-reversed code sequence to a short impulse with low sidelobes. The PSK device is most often used as a compressor because the phase-encoded signal can be generated easily by digital means. Table 16.4 shows performance parameters for Sawtek Inc. PSK delay lines.

SAW PSK correlators are widely used in spread spectrum communication systems, phase-coded radars, radio data links, communication modems, navigation and identification systems, and range differencing surveillance systems. These devices offer the advantages of compact size, real-time processing, and asynchronous operation.

16.6.3 Dispersive Delay Lines

SAW dispersive delay lines currently are used most extensively in pulse compression radars. They provide a high degree of flexibility in the implementation of different types of waveforms, which makes them suitable for the optimization of particular radar applications.

The basic concepts of a pulse compression system are shown in Figure 16.12. A short impulse is applied to the SAW expander to produce an elongated frequency-modulated signal at the output of the device. The coding sense can be chosen

Table 16.4 Performance Parameters for PSK Delay Lines

<i>Parameter</i>	<i>Values</i>
Center frequency	10 to 800 MHz
Fractional bandwidth	2 to 40%
Insertion loss	20 to 50 dB
Maximum length	<40 μ s
Chip rate	<200 MHz
Triple-transit suppression	30 to 60 dB
Spurious suppression	40 to 70 dB
Sidelobe level degradation from theoretical	<1 to 3 dB
Processing gain degradation from theoretical	<1 to 3 dB

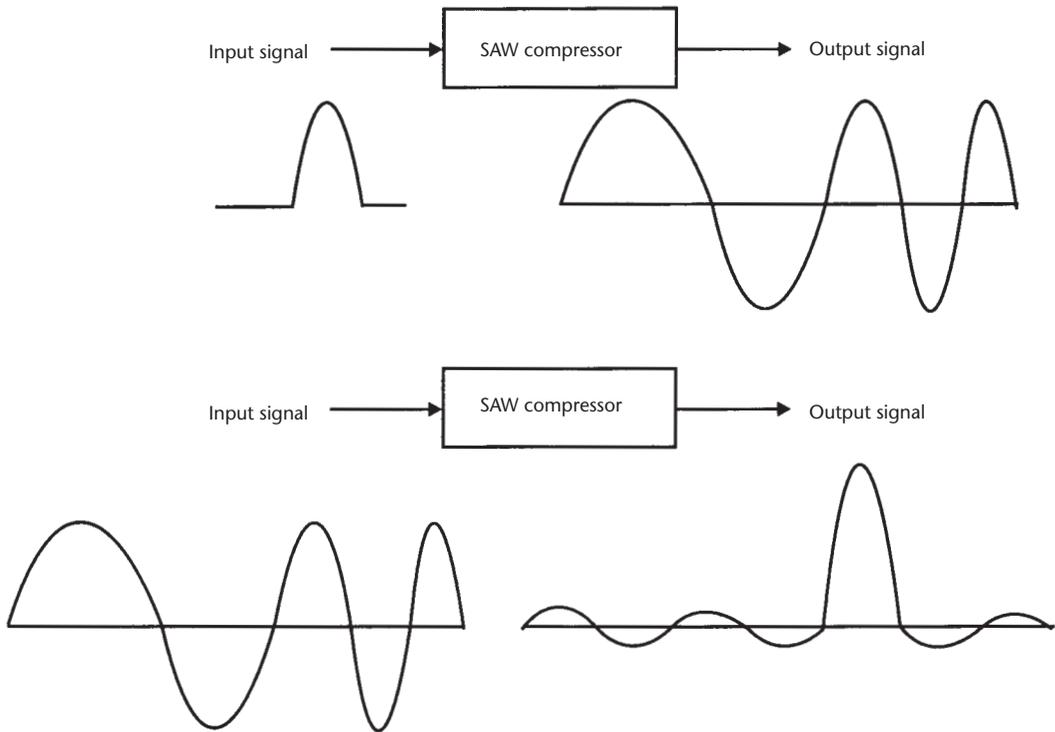


Figure 16.12 Basic operation of a pulse compression system. (After: [9].)

to be “up,” with high frequencies having longer delays than low frequencies, or “down,” with high frequencies having shorter delays than low frequencies. After the amplified expanded signal reflects from the target, it is received and fed into a SAW compressor having the complex conjugate frequency characteristic, or simply the reverse impulse time response, of the expander. The delays encountered by the different frequencies are opposite to the delays in the expander network, resulting in all frequencies being compressed in time.

There are a number of ways to implement SAW dispersive delay lines. Two such systems are shown in Figure 16.13: the in-line device with a broadband transducer and the in-line device with two dispersive transducers.

Figure 16.14 shows the reflective array compressor. This compressor is used when long delays (up to $120\ \mu\text{s}$) or large-bandwidth time delay products ($>1,000$) are needed. Shallow grooves etched in the delay path result in SAW reflections to form a delay that depends on the frequency. The reflective array compressor usually is more complex and costly than its IDT counterparts. It is also more susceptible to temperature effects and suffers from higher insertion loss than conventional IDT designs. On the other hand, it is more tolerant to manufacturing defects, allows higher BT products to be implemented, and offers a significant reduction in size.

Table 16.5 shows performance parameters for Sawtek dispersive delay lines. Parameters are presented for both expanders and compressors.

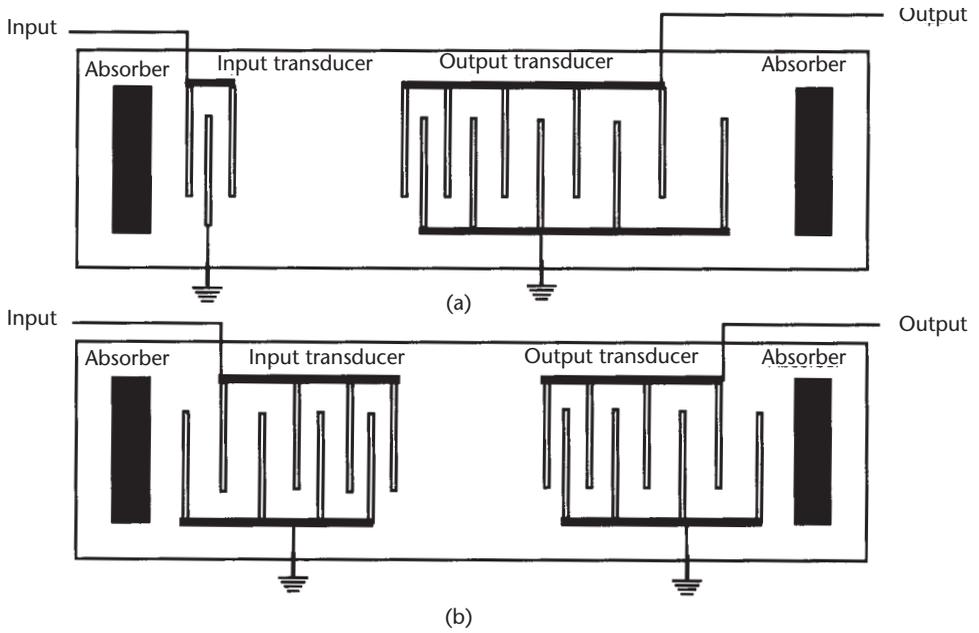


Figure 16.13 Dispersive delay lines: (a) in-line device with one dispersive transducers and (b) in-line device with two dispersive transducers. (After: [9].)

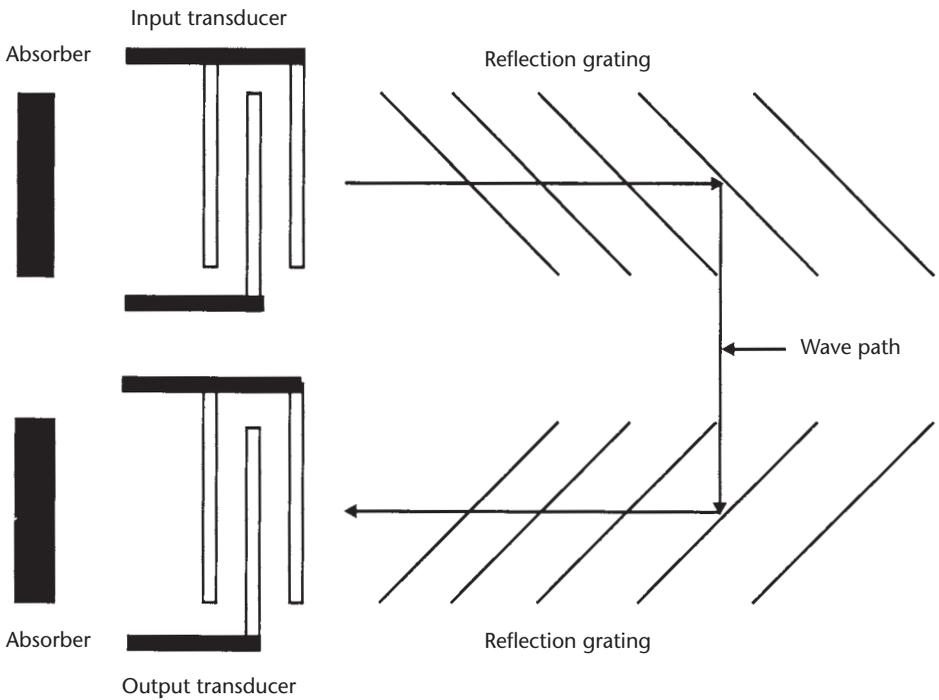


Figure 16.14 Reflective array compressor. (After: [9].)

Table 16.5 Performance Parameters for Dispersive Delay Lines

<i>Parameters</i>	<i>Values</i>
Expander	
Center frequency	20 to 1,000 MHz
Fractional bandwidth	2 to 67%
Pulse length	0.25 to 120 μ s
tblCoding type	LFM, NLFM, MNLFM
Amplitude ripple of expanded pulse	± 0.25 dB to ± 0.5 dB
I/O impedance	50V
VSWR	1.5:1
Compressor	
Compressed pulse width (τ)	3 to 1,000 ns
Close in time sidelobes ($t < 6\tau$)	<30 dB to <40 dB
Far out time sidelobes ($t > 6\tau$)	<40 dB to <45 dB
Mismatch loss	0.1 to 2 dB
Signal-to-noise improvement	10 dB to 35 dB
I/O impedance	50 Ω
VSWR	1.5:1

16.7 Surface Acoustic Wave Delay Line Oscillators

SAW oscillators are available from Andersen Laboratories in Bloomfield, Connecticut, at operating frequencies from 100 MHz to 2.6 GHz. These oscillators have low phase noise, compact size, good frequency stability, and rugged construction. They can be either fixed-frequency oscillators or voltage-controlled oscillators. Typical output power is about 10 mW. Figure 16.15 is a block diagram of a SAW delay line VCO. This oscillator uses a SAW device as the frequency-controlling element in the feedback loop of an amplifier. In the circuit shown, the output of a transistor amplifier is fed to a power divider or coupler. Part of the output power is fed to a buffer amplifier, and part is fed to a SAW delay line. The buffer amplifier isolates the oscillator from the load and provides the needed power output.

The delay line is specifically designed to set the fundamental operating frequency of the loop and provide the necessary phase noise characteristics for VCO requirements. The overall gain of the loop is ≥ 1 . The phase shift around the loop is an integral number of 2π radians. The oscillator supports a comb of output frequencies spaced by $1/\tau$, where τ is the delay time of the SAW delay line. The delay line has a line frequency passband that selects the desired frequency line.

By introducing a predictable voltage-controlled phase shift into the feedback loop of the amplifier, the frequency of oscillation can be varied (pulled) from the center frequency over some specified operating range.

Table 16.6 shows performance parameters for standard product oscillators made by Andersen Laboratories. Oscillator specifications depend on each individual application and vary widely. Table 16.6 should be used only as a guide.

The frequency stability for the SAW oscillator is approximately 12.5 ppm for the temperature range of 0–50°C. That can be improved by limiting the temperature range or by incorporating a heater. The packaged size for a standard Andersen

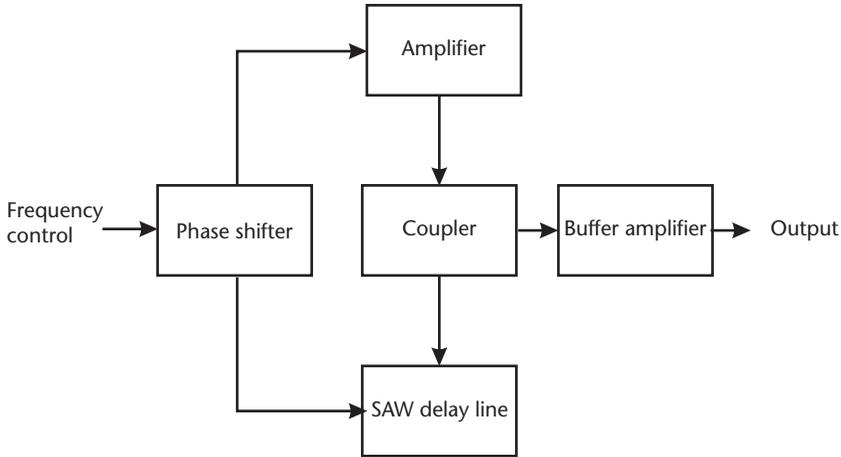


Figure 16.15 Block diagram of SAW VCO.

Table 16.6 Performance Parameters for SAW Oscillators

Parameters	Values $f < 1,300$ MHz	Values $f > 1,300$ MHz
Operating frequency	100 to 1,300 MHz	1,300 to 2,300 MHz
Tuning range	Up to 1,000 kHz	Up to 1,500 kHz
Modulation bandwidth	Up to 500 kHz	Up to 500 kHz
Tuning voltage	0 to 12V	0 to 12V
Output power	+10 dBm nominal	+10 dBm nominal
Power variation over temperature	± 1.5 dB maximum	± 2.0 dB maximum
Spurious outputs		
Harmonic	-30 dBc maximum	-30 dBc maximum
Nonharmonic	-60 dBc	-60 dBc maximum
Frequency accuracy at 25°C	± 20 ppm	± 20 ppm

Laboratories SAW oscillator is only about $1 \times 1.5 \times 0.25$ inch in the dual in-line package (DIP) configuration and slightly larger for other configurations.

16.8 Bulk Acoustic Wave Delay Lines

BAW delay lines are another important class of crystal delay line. A basic BAW delay line comprises two piezoelectric transducers bonded to a low-velocity medium such as quartz. The time delay is determined by the acoustic velocity and the length of the path.

Figure 16.16 shows some BAW delay line configurations. The simplest geometry, illustrated in Figure 16.16(a), is that of a rectangular bar with piezoelectric transducers bonded to each end. An electric signal is applied to the input transducer and subsequently converted to an acoustic signal. The acoustic signal is then transmitted through the crystal to the output transducer, where it is detected and converted back to an electrical signal.

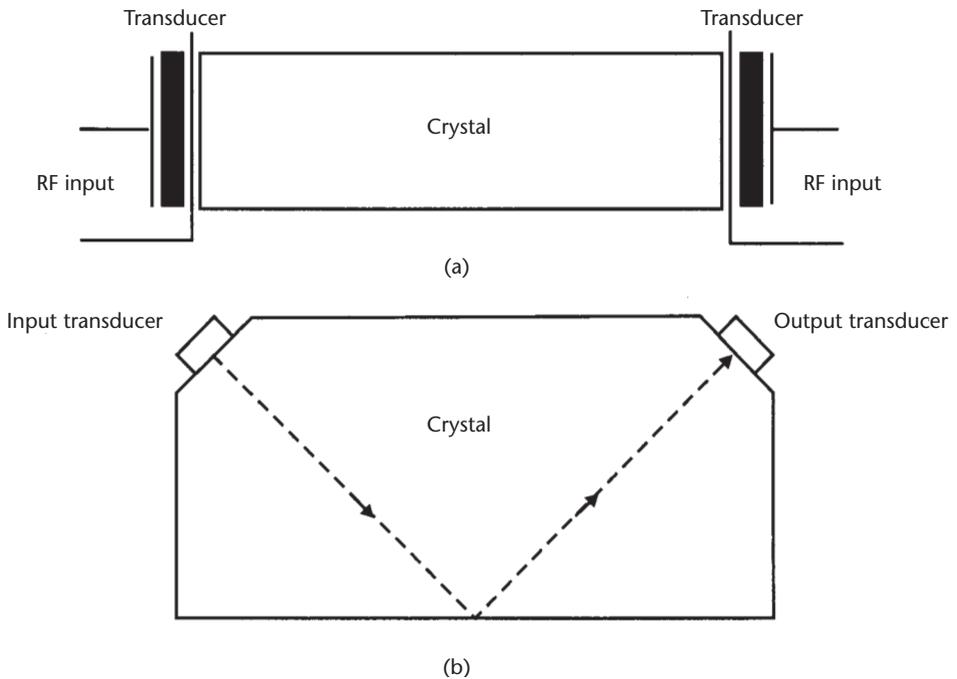


Figure 16.16 BAW delay line configurations: (a) schematic of rectangular bar BAW delay line and (b) reflection type BAW delay line.

It is possible to provide long delays by using multiple reflection type crystals. Figure 16.16(b) shows the case of a single reflection crystal. There are other configurations that have large numbers of bounces.

The delay media used by Andersen Laboratories for BAW delay lines is fused quartz, crystalline quartz, or glass. Fused quartz is the delay medium most frequently used for standard BAW delay lines. The acoustic loss constant for fused quartz is low, making possible large delay lines with reasonable insertion loss. By using multiple internal reflections, delays up to $5,000 \mu\text{s}$ can be achieved with a single device. The wave velocity in fused quartz is about $4.3 \mu\text{s}/\text{inch}$ in the compression mode, which is normally used for BAW delay lines.

Table 16.7 shows typical performance parameters for standard BAW delay lines made by Andersen Laboratories.

Teledyne Microwave in Mountain View, California, produces BAW delay devices that operate at microwave frequencies. The RF input is through an impedance-matching network. The output of this network is a thin-film acoustic transducer. The delay crystal material is usually single-crystal sapphire or quartz. At the ends of the rod are thin-film acoustic transducers for converting from acoustic energy to electrical energy and vice versa. An impedance-matching network is used at the output.

This transducer is made up of three layers: a metalized counterelectrode, a piezoelectric layer, and a metalized top electrode. The counterelectrode is composed of Cr-Au composite metalization, which acts as the ground plane of the transducer. The piezoelectric film is a sputtered ZnO thin-film that converts electromagnetic energy to acoustic energy. The top electrode is also a Cr-Au composite metalization.

Table 16.7 Performance Parameters for LF BAW Delay Lines

<i>Parameters</i>	<i>Values</i>
Delay	0.25 to 5,000 μ s
Center frequency	5 to 150 MHz
3-dB bandwidth	60%
Insertion loss	5 dB to 60 dB
Feedthrough	Suppressed to >50 dB
Triple transit	Suppressed to >40 dB
Random signal	Suppressed to >50 dB
Cross-talk	Suppressed to >60 dB

Delay devices using bulk acoustic waves can be built for operation in the 300-MHz to 18-GHz range. Bandwidths in excess of one octave have been achieved, but the narrower the bandwidth, the lower the insertion loss.

Time delay between 100 ns and 30 μ s are achievable with a single crystal acting as a two-port device. The longer time delays can be realized only at lower frequencies since acoustic propagation losses increase with frequency and insertion losses become prohibitive at higher frequencies for long delays. For example, a delay device with 0.5 μ s delay with a center frequency at 16 GHz and a bandwidth of 1 GHz will exhibit an insertion loss of about 60 dB. That loss can be overcome by amplification.

References

- [1] Kinsman, R. G., *Crystal Filters: Design, Manufacture, and Application*, New York: John Wiley & Sons, 1987.
- [2] Krauss, H. L., C. W. Bostian, and F. H. Raab, *Solid State Radio Engineering*, New York: John Wiley & Sons, 1980, pp. 151–161.
- [3] Matthys, R. J., *Crystal Oscillator Circuits*, New York: Wiley Interscience, 1983.
- [4] Kajfez, D., and P. Guillon, *Dielectric Resonators*, Norwood, MA: Artech House, 1986.
- [5] Krauss, H. L., C. W. Bostian, and F. H. Raab, *Solid State Radio Engineering*, New York: John Wiley & Sons, 1980, pp. 273–279.
- [6] *ANZAC RF and Microwave Signal Processing Components Catalog*, Burlington, MA: Adams-Russell Co., pp. 455–459.
- [7] Helszajn, J., *YIG Resonators and Filters*, New York: John Wiley & Sons, 1985.
- [8] Kennedy, G., *Electronic Communication Systems*, 3rd ed., New York: McGraw-Hill, 1985, pp. 391–393.
- [9] Skolnik, M. I., *Introduction to Radar Systems*, 2nd ed., New York: McGraw-Hill, 1980, pp. 420–434.

Semiconductor Diodes and Their Circuits

This chapter discusses semiconductor diodes and circuits. Topics include semiconductor materials, “ordinary” junction diodes, Zener diodes, Schottky-barrier diodes, PIN diodes, varactor diodes, step-recovery diodes, tunnel diodes, Gunn-effect diodes, IMPATT diodes, LEDs, IR laser diodes, and IR photodiodes.

Some of the information presented here is adapted from [1, 2].

17.1 Semiconductor Materials

Some of the following material is quoted or adapted from [1].

Semiconductors are materials that have electrical conductivities intermediate between those of metals and insulators, that is, they are “semi” conductors. These materials are found in column IV and neighboring columns of the periodic table. The column IV materials of interest are silicon (Si) and germanium (Ge), with Si being the most important, and are referred to as elemental semiconductors. The elements in column III of interest are boron (B), aluminum (Al), gallium (Ga), and indium (In). These elements are used as doping materials for Si or other semiconductor materials for producing p-type material. The elements in column V of interest are phosphorus (P), arsenic (As), and antimony (Sb). These elements are used as doping materials for Si or other semiconductor materials for producing n-type materials.

Compounds of materials from columns III and V make up part of the intermetallic or compound semiconductors. These III-V semiconductors include AlP, AlAs, AlSb, GaP, GaAs, GaSb, InP, InAs, InSb. Of these, gallium arsenide (GaAs) is the most important and has been used for microwave and higher frequency FET devices and Gunn diodes. GaAs has also been used for Schottky-barrier diodes, tunnel diodes, varactors, and step-recovery diodes. The elements in column II that are of interest are zinc (Zn) and cadmium (Cd). The elements in column VI that are of interest are sulfur (S), selenium (Se), and tellurium (Te). Semiconductor compounds that use elements from columns II and VI are ZnS, SnSe, ZnTe, CdS, CdSe, and CdTe. Some of the compound semiconductors are used in LEDs.

17.2 “Ordinary” Junction Diodes

Figure 17.1 shows an ordinary junction diode. This diode consists of n- and p-doped Si with two terminals. The current-voltage (I-V) characteristics for the diode are shown in the figure. When the p-side of the diode is positive with respect to the

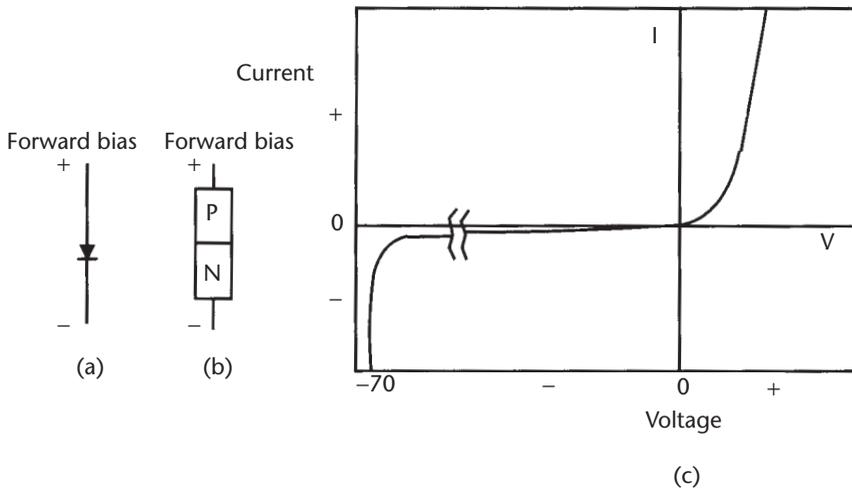


Figure 17.1 Ordinary junction diode: (a) symbol; (b) diode configuration; and (c) current-voltage characteristics. (After: [1].)

n-side, the resistance of the diode is low. In the case of a Si diode, it takes about 0.7V to turn on the diode in the forward direction.

Two diodes in series would require about 1.4V to turn on the diodes.

When the p-side of the diode is negative with respect to the n-side, the resistance of the diode is at first very high and little current flows. At some high negative voltage, reverse breakdown takes place, and the reverse diode resistance becomes small. The mechanism for such high voltage breakdown is known as *avalanche breakdown*.

17.3 Zener Diodes

A Zener diode consists of n- and p-doped Si with two terminals. The current-voltage characteristics for this diode look very much like the I-V characteristics for the ordinary diode except that the reverse bias breakdown occurs at much lower voltages. Typical breakdown voltages for Zener diodes are more like 5–20V rather than 70V or more for an ordinary diode. Zener breakdown occurs in heavily doped junctions in which the transition from p to n material is abrupt. The mechanism involved is called tunneling.

The Zener diode finds application as a voltage reference. A resistor usually is placed in series with the diode to limit the current. The reverse voltage across the diode is fixed by the design of the diode.

It should be pointed out ordinary diodes occasionally are used as voltage references in bias circuits for transistor amplifiers. In that case, the diode is forward biased, and the voltage drop across each diode is again about 0.7V. Two or more diodes sometimes are connected in series to increase the voltage of the reference.

17.4 Schottky-Barrier Diodes

Much of the remainder of this chapter is quoted or adapted from [2].

The Schottky-barrier diode is an extension of the oldest semiconductor device of them all, the point-contact diode. With the Schottky-barrier diode, the metal-semiconductor interface is a surface rather than a point contact. Like the point-contact diode, the Schottky-barrier diode has no minority carriers in the reverse-bias condition. Thus, the delay present in junction diodes, due to hole-electron recombination time, is absent. Because of the larger contact area between the metal and the semiconductor, the Schottky-barrier diode has much lower resistance and lower noise than the point-contact diode. That makes this diode desirable for microwave and higher frequency applications, where ordinary junction diodes are not effective.

The most commonly used semiconductor materials for Schottky-barrier diodes are N-type Si and N-type GaAs. GaAs has the lower noise and the higher operating frequency limits. On the other hand, Si is easier to fabricate and is consequently used at X-band and lower frequencies in preference to GaAs. The metal at the interface with the semiconductor is often a thin layer of titanium surrounded by gold for protection and low ohmic resistance.

Schottky-barrier diodes can be used at frequencies as high as 100 GHz. They are used as detectors and mixers. The noise figures for mixers using these diodes are as low as 4 dB at 2 GHz and 15 dB at 100 GHz.

17.5 PIN Diodes

Figure 17.2 shows the PIN diode, which consists of a narrow layer of p-type semiconductor separated from an equally narrow layer of n-type material by a somewhat thicker region of intrinsic semiconductor material, thus the term PIN. PIN diodes often use lightly doped n-type semiconductor material rather than intrinsic material. PIN diodes usually are made of Si, although GaAs is sometimes used.

The PIN diode is used for microwave power switching, attenuating, limiting, and modulation. At microwave frequencies, the PIN diode acts as a variable resistance. The simplified equivalent circuits and the resistance variation with reverse and forward bias are shown in Figure 17.3(a, b). With reverse (negative) bias, the resistance to microwave energy typically is about 5,000–10,000 Ω . When the diode is forward biased, the positive-bias resistance to microwave energy is typically 1–10 Ω . If the PIN diode is placed across a waveguide, a 50 Ω coaxial line or other

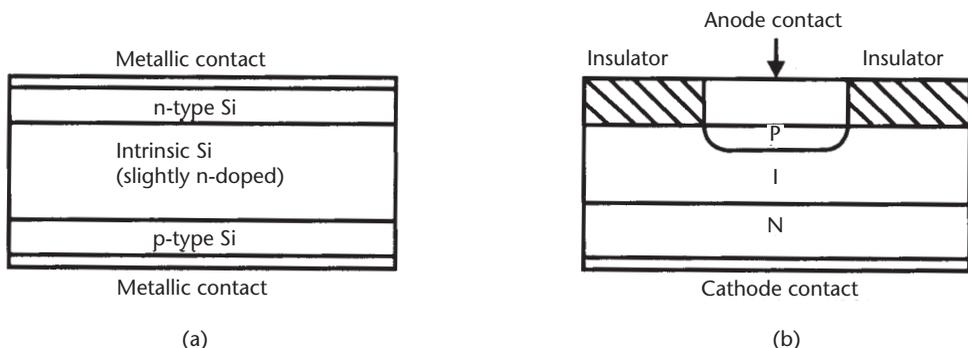


Figure 17.2 PIN diode: (a) schematic diagram and (b) planar PIN diode.

transmission medium, it does not significantly load the line when negatively biased. When positively biased, however, it presents a near short circuit across the line and thus produces reflections on the line.

Figure 17.3(c) is a schematic diagram of a series-mounted PIN diode switch for a coaxial or microstrip line. The dc bias is fed into the diode using an RF choke, and the signal is injected using a coupling capacitor. The output is across an RF choke to ground.

Figure 17.3(d) is a schematic diagram for a shunt-mounted PIN diode switch for a coax line. Again, the dc bias is fed in to the diode using an RF choke, and the signal is fed in using a coupling capacitor. In this case, no output choke is used, and the PIN diode is connected directly to ground. The output is fed out using a coupling capacitor.

PIN diodes can be used in parallel or series, as desired. Individual diodes can handle up to about 200-kW peak or 200-W average. Several diodes in parallel can handle as much a 1-MW peak. Switching times are in the range of 1–40 ns, depending on the power levels used.

17.6 Varactor Diodes

Almost every semiconductor diode has a junction capacitance that varies with the applied reverse bias. If the diode is manufactured to have suitable microwave characteristics, it is called a varactor diode. Varactor diodes are made of Si or GaAs. GaAs has the advantage of higher maximum operating frequency. Figure 17.4

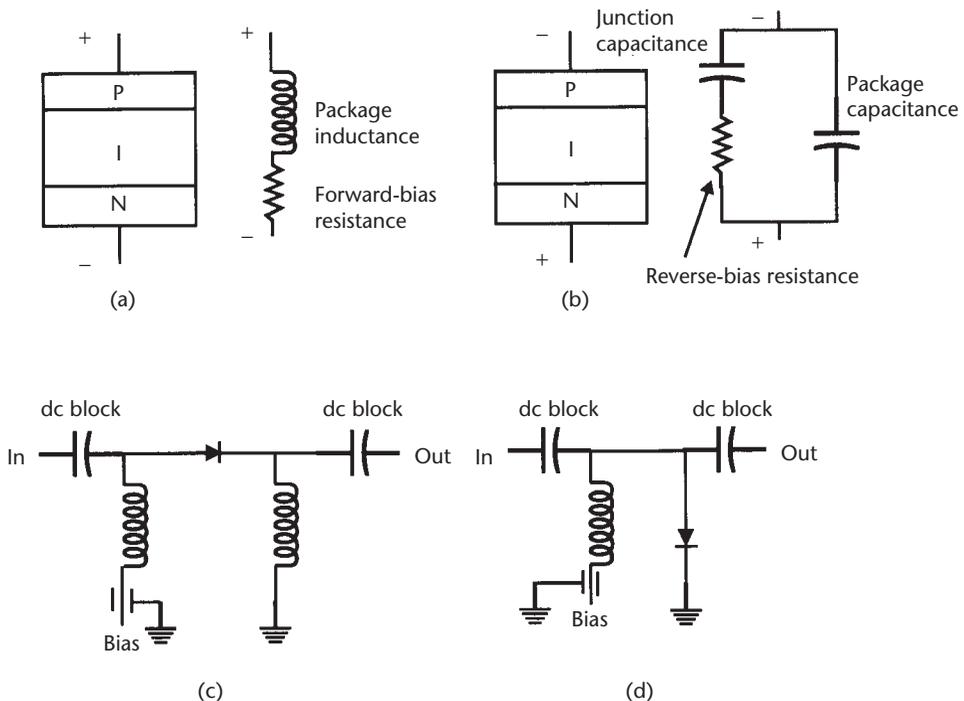


Figure 17.3 PIN diode simplified equivalent circuits and example switches: (a) forward bias condition; (b) reverse bias condition; (c) series-mounted switch; and (d) shunt-mounted switch.

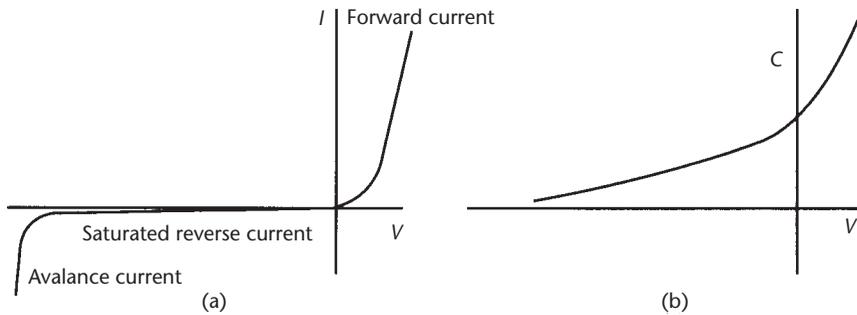


Figure 17.4 Varactor diode characteristics: (a) current versus voltage and (b) capacitance versus voltage. (After: [2].)

shows the characteristics of varactor diodes, including the current versus voltage characteristics and the junction capacitance versus voltage characteristics. The bias voltage region of interest for a varactor diode is between just above the avalanche breakdown point and zero volts. For typical Si varactors, the minimum capacitance is about 1 pF and the maximum capacitance is about 25 pF.

Varactors find application as voltage-variable capacitors for frequency modulation and oscillator tuning. They also are used in frequency multipliers. Because snap-off varactor diodes multiply by high factors with better efficiency than ordinary varactor diodes, they are used where possible. GaAs varactors often are used at the higher frequencies. A varactor multiplier of this type can have an efficiency for a 60-GHz doubler of greater than 50%.

The maximum output power for the varactor diode multipliers ranges from more than 10W at 2 GHz to about 25 mW at 100 GHz. Tripler efficiencies range from 70% at 2 GHz to about 40% at 36 GHz; however, that is with proper design, including idlers to reflect fundamental and second harmonic power back to the input. Otherwise, 33% efficiency is more typical.

One of the current applications for multiplier chains is to provide a low-power signal to phase-lock a Gunn or IMPATT diode oscillator. (These devices are discussed later.)

17.7 Step-Recovery Diodes

A step-recovery diode is a Si or GaAs p-n junction diode with construction similar to that of a varactor diode. It stores charge when conducting in the forward direction. When reverse bias is applied, the diode briefly discharges the stored energy in the form of a sharp pulse, an impulse, that is rich in harmonics. The duration of the pulse typically is only 100 to 1,000 ps, depending on diode design.

Step-recovery diodes are not available for frequencies above about 20 GHz, whereas varactors can be used well above 100 GHz. Step-recovery diodes are available for powers in excess of 50W at 300 MHz, 10W at 2 GHz, and 1W at 10 GHz. Multiplication ratios up to 12 commonly are available. Efficiency can be in excess of 50% for triplers at frequencies up to 1 GHz. The efficiency drops to about 15% for a times-5 multiplier with an output frequency of 12 GHz.

17.8 Microwave Tunnel Diodes and Circuits

Figure 17.5 shows the voltage-current characteristics for a Ge junction tunnel diode. This diode differs from the ordinary junction diode in that the semiconductor material is heavily doped, perhaps 1,000 times that of an ordinary rectifier diode. That permits a depletion layer so thin that tunneling can occur easily.

In the voltage region from A to B, there is a region of negative resistance. That means that this device can be used as an oscillator. The tunnel diode oscillator found use early after its development, but it no longer is used extensively because other negative-resistance semiconductor devices now provide higher output power.

The tunnel diode also can be used as a microwave amplifier when used with a circulator. A circuit of this type is shown in Figure 17.6. The tunnel diode amplifier (TDA) is a low-noise system. Reasons for that are that a tunnel diode is a low-resistance device, and the operating current is low. Tunnel diode amplifiers can be fairly broadband systems. They are small and simple, and they can operate at microwave frequencies with upper limits in excess of 50 GHz.

17.9 Microwave Gunn Diodes and Circuits

Figure 17.7 shows an epitaxial GaAs Gunn diode. A negative resistance is provided by such a device if the voltage gradient across the slice of GaAs is in excess of about 3,300 V/cm.

Oscillations then occur if the slice is connected to a suitably tuned circuit. Proper doping profile is also required.

The Gunn effect is a bulk property of semiconductors and does not depend on either junction or contact properties. It occurs only in n-type materials, so it is associated only with electrons and not holes. GaAs is one of the few materials for which the Gunn effect works. In this material with n-type doping, there is an empty energy band higher in energy than the highest filled or partially filled band. The forbidden energy gap is small. In this diode with its very high voltage gradient, electrons acquire enough energy to be transferred to the higher energy band, in which they are

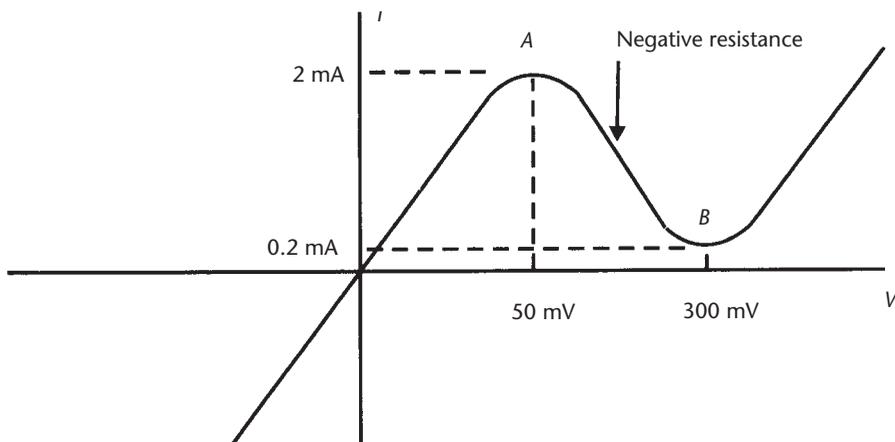


Figure 17.5 Tunnel diode voltage-current characteristics. (After: [2].)

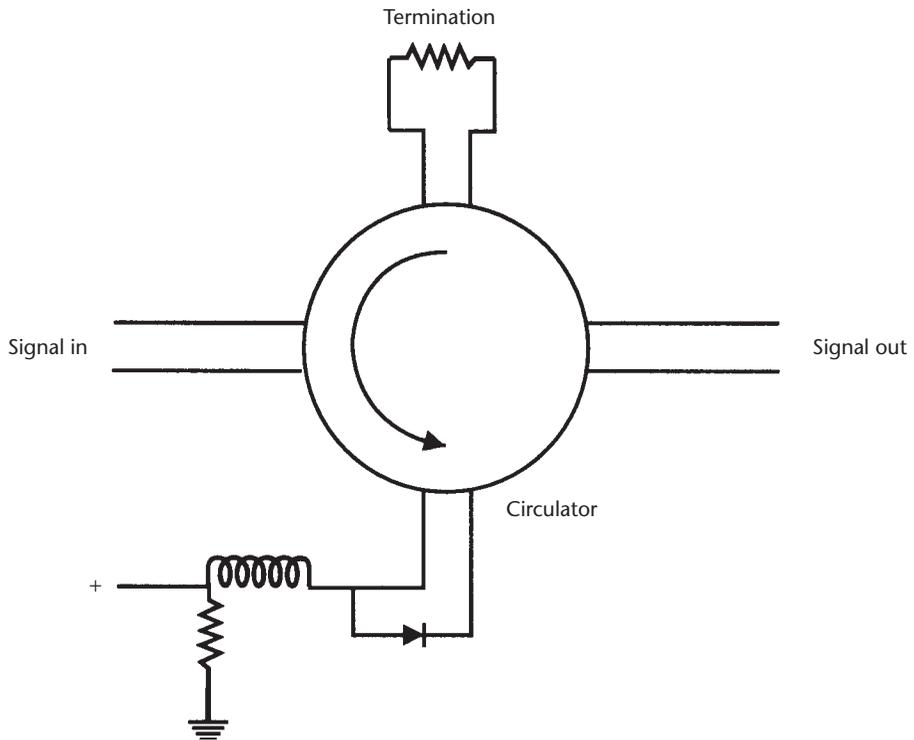


Figure 17.6 Tunnel-diode amplifier with circulator. (After: [2].)

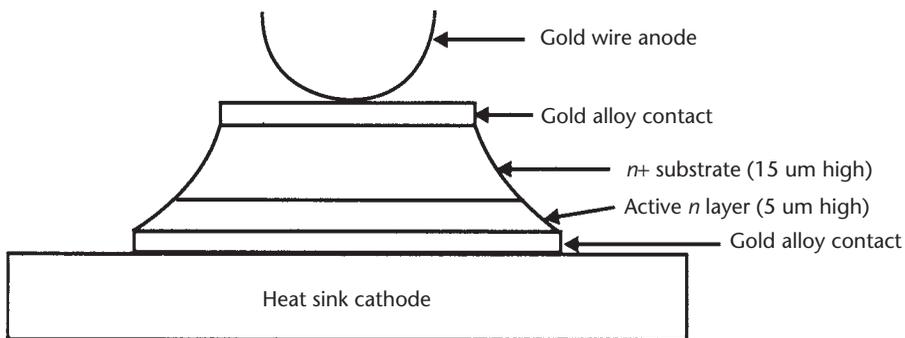


Figure 17.7 Epitaxial GaAs Gunn diode. (After: [2].)

much less mobile. Thus, the current has been reduced as a result of voltage rise. This voltage region therefore is a region of negative resistance. Eventually with increasing voltage, the voltage becomes high enough to remove electrons from the higher energy, lower mobility band so the current will increase with voltage once again.

A second phenomenon with Gunn diodes is important: the formation of bunches of electrons. Negative-resistance domains are formed that are less conductive. These travel toward the positive anode at a speed of about 10^7 cm/s in practice. These traveling domains may be thought of as low-conductivity, high-electron-transfer regions corresponding to a negative pulse of voltage. When they arrive

at the positive end of the GaAs slice, a pulse is received by the associated resonant circuit, and oscillations take place. It is actually this arrival of pulses at the anode, rather than the negative resistance proper, that is responsible for oscillations in Gunn diode oscillators.

A typical Gunn oscillator uses a coaxial cavity. The Gunn diode is located between the end of the center conductor of the cavity and the end of the cavity. The bias for the diode is fed through the other end of the cavity using a bypass capacitor and a bias feed-through capacitor. A sliding short plunger is used to mechanically tune the half-wave cavity. A coaxial output is connected to a capacitive probe that enters the cavity. A tuning screw is used for coupling adjustment. This type of system is shown in Figure 17.8.

Gunn diodes are available for operation in the frequency range from 4 to beyond 100 GHz. A typical X-band Gunn diode oscillator requires a 9V dc bias and an operating current of 950 mA. The output RF power is 300 mW in the frequency range of 8–12.4 GHz. The efficiency, therefore, is only about 3.5%. A typical Gunn diode oscillator operating in the 26.5–40 GHz band produces about 250 mW with an efficiency of 2.5%. Gunn diode oscillators are used as low- and medium-power oscillators in microwave receivers. Higher-power Gunn oscillators are used in a wide variety of frequency-modulated transmitters. Other applications currently include police radar, CW Doppler radar, burglar alarms, and aircraft rate-of-climb indicators.

17.10 Microwave IMPATT Diodes

IMPATT diodes, another important type of microwave oscillator, are also called avalanche diodes. The schematic diagram for this type of diode is shown in Figure 17.9. The IMPATT diode is reverse biased. An extremely high voltage gradient (on the order of 400 kV/cm) is applied to the IMPATT diode. This causes a flow of minority carriers (electrons in this case) across the junction.

Figure 17.10 shows the dc voltage at the avalanche threshold. Let us assume the existence of oscillations and therefore an RF voltage added to the dc voltage. Avalanche takes place over a period of time such that the current pulse maximum at the junction occurs at the instant when the RF voltage across the diode is zero and going

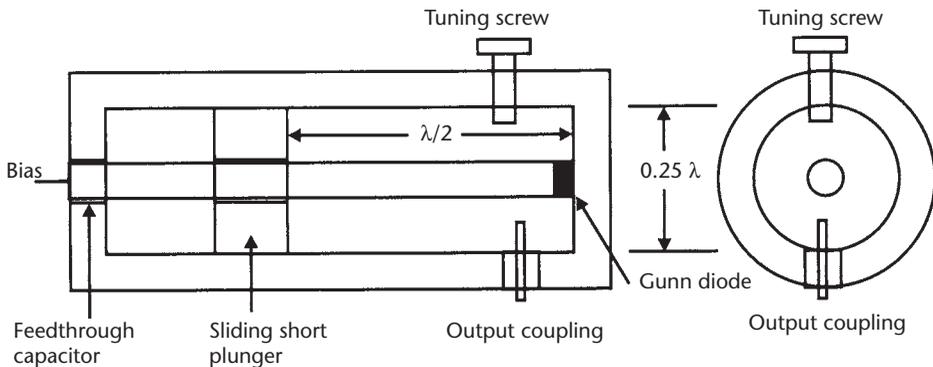


Figure 17.8 Layout of Gunn diode oscillator. (After: [2].)

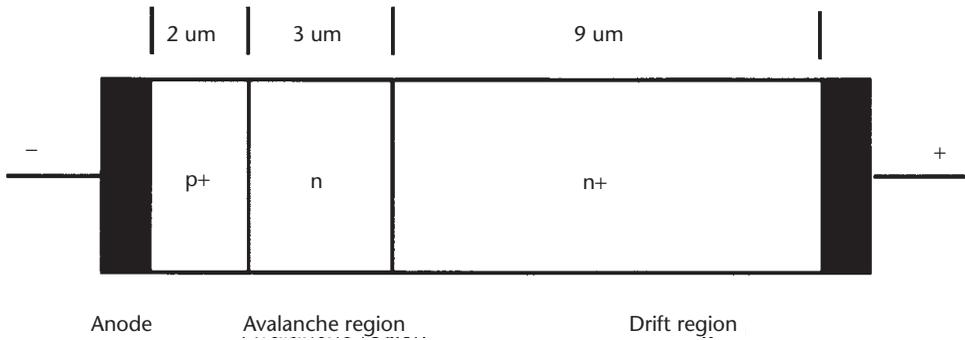


Figure 17.9 IMPATT diode schematic diagram. (After: [2].)

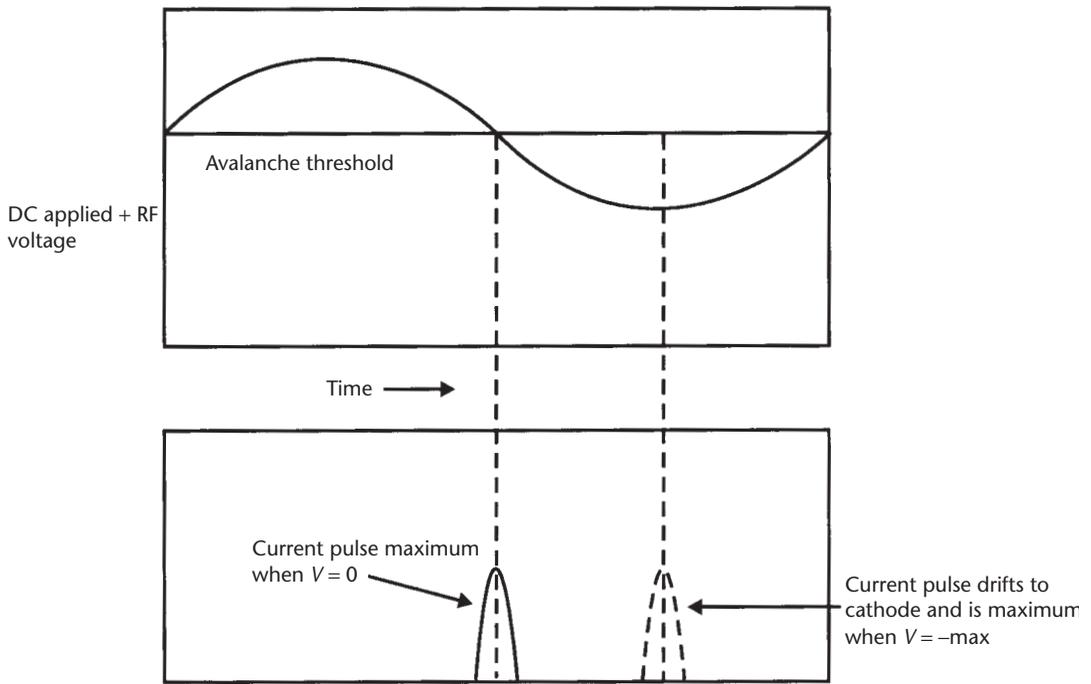


Figure 17.10 Current and voltage characteristics of IMPATT diodes. (After: [2].)

negative. A 90-degree phase difference between voltage and current has thus been obtained, as shown in Figure 7.10. The current pulse then drifts toward the cathode. The thickness of the drift region is selected such that the delay in reaching the cathode adds another 90-degree phase shift and thus a total phase shift between current pulse and voltage of 180 degrees. This is one type of negative resistance.

IMPATT diodes usually are made of Si; they also may be made of GaAs. IMPATT diodes are essentially narrowband devices because the thickness of the drift region is critical to the operation of the device.

Commercial IMPATT diodes currently are produced over the frequency range of about 4–200 GHz. The maximum power per diode varies from nearly 20W near 4 GHz to about 50 mW at the HF end. Above 20 GHz, this type of diode produces higher CW power output than any other semiconductor device.

17.11 Semiconductor IR Laser Diodes

A GaAlAs diode of the type shown in Figure 17.11 is capable of producing laser action. Depending on its precise chemical composition, it is capable of producing an output with wavelength in the range of 0.75 to 0.9 μm with 0.85 μm being typical. This is in the near-infrared region. The attenuation in a fiber optic cable at this wavelength is about 2.5 dB/km.

The GaAlAs laser is forward biased to turn it on. Electrons and holes originating in the GaAlAs layers cross the heterojunctions (junctions between dissimilar semiconductor materials, GaAlAs and GaAs in this case) and give off their excess recombination energy in the form of light. The heterojunctions are opaque, and the active region is constrained by them to the p-layer of GaAs. This layer is only a few micrometers thick. The two ends of the slice are highly polished, so that reinforcing reflection takes place between them, as in other lasers, and a continuous beam is emitted in the direction shown. The laser is capable of powers in excess of 1W.

Figure 17.12 shows an InGaAs phosphide laser. This semiconductor laser, a more recent development than the GaAlAs device, was developed to produce laser outputs at wavelengths longer than those the GaAs laser is capable of producing. The operation of this laser is similar to that of the GaAlAs laser. Output wavelengths in this case are at 1.3 μm or 1.55 μm . That permits the laser to take advantage of low-attenuation windows in the transmission spectrum of optic fibers. The attenuation in a fiber optic cable at 1.3 μm is 0.4 dB/km. The attenuation in a fiber optic cable at 1.55 μm is an even lower 0.25 dB/km.

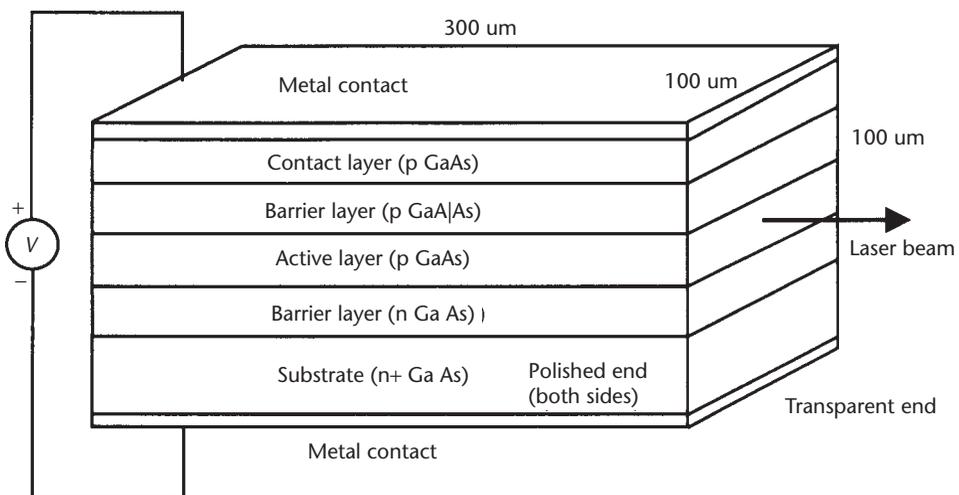


Figure 17.11 GaAlAs laser diode (0.75 to 0.9 μm wavelength). (After: [2].)

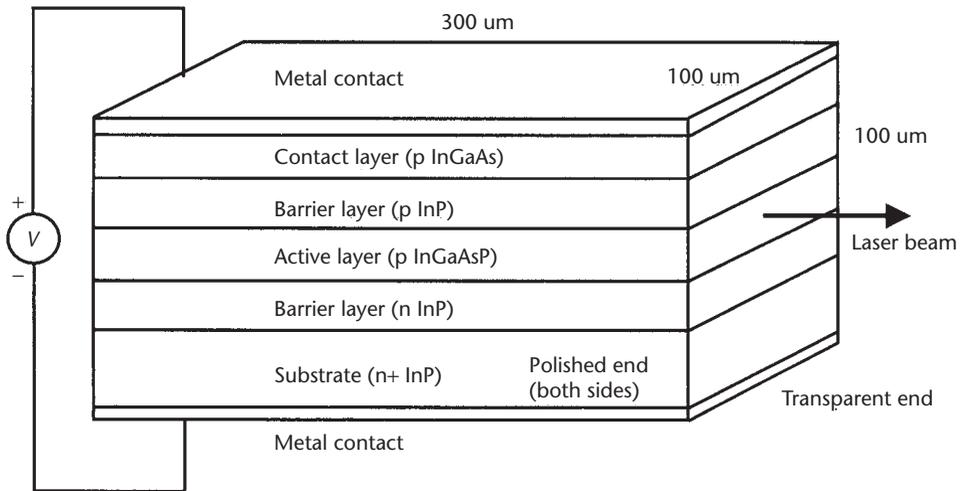


Figure 17.12 InGaAs phosphide laser diode (1.2 to 1.6 μm wavelength). (After: [2].)

17.12 Light-Emitting Diodes

The construction of an LED is similar to that of a laser diode, but the structure is simpler. There are no polished ends, and laser action does not take place. Consequently, the power output is much lower, and a much wider beam of light results. The light is not monochromatic. A small lens is often used with the diode.

In the operation of LEDs, electrons and holes are injected across heterojunctions, and light energy is given off during recombination. The materials used are the same as for the corresponding laser diodes.

17.13 IR Photodiodes

Photodiodes are used in a variety of applications, including optical communications, picosecond-pulse detection, and optical fiber characterization. A number of different types of photodiodes can be used. One type is the GaAlAs/GaAs PIN diode, which has a typical rise time of 50 ps. This photodiode has a 10-dB bandwidth of greater than 7.0 GHz and a quantum efficiency of nearly 65%. It also features extremely low capacitance and extremely low dark current.

The photodiode structure consists of a photosensitive GaAs layer, overlaid by a transparent GaAlAs window layer, grown on a semi-insulating substrate. The use of a semi-insulating GaAs substrate significantly reduces parasitic capacitance. The photodiode chip is passivated with a dielectric coating to ensure high reliability and is protected in the detector housing by a sapphire window. An SMA-type connector is incorporated into the diode for ease in coupling to 50 Ω systems.

Semiconductor types used for PIN-type photodetectors include Ge, Si, GaAlAs, and InGaAs.

One problem with the PIN photodiode is that it is not as sensitive as we would like it to be. No gain takes place in the device. A single photon of light cannot create more than one hole-electron pair. This problem is overcome by the use of the

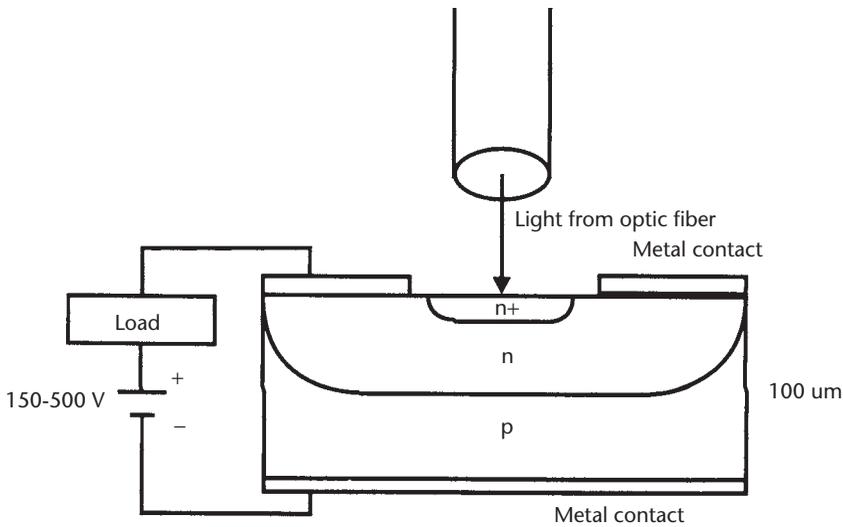


Figure 17.13 APD schematic. (After: [2].)

avalanche photodiode (APD). An APD is reverse biased close to breakdown using a voltage in the range of 100–500V. A light quantum impinging on the diode causes a hole-electron pair to be created. With the high voltage used, avalanche multiplication can take place, as in the IMPATT diode. A typical APD is 10 to 150 times more sensitive than a PIN photodiode. Response time also is much shorter. The materials used for the APD are the same as those used for the PIN diode.

Figure 17.13 is an illustration of an APD.

References

- [1] Streetman, B. G., “Semiconductors and Transistors,” Chap. 18 in *Reference Data for Engineers*, 8th ed., M. E. Van Valkenburg, (ed.), SAMS, Carmel, IN: Prentice Hall Computer Publishing, 1993.
- [2] Kennedy, G., *Electronic Communication Systems*, 3rd ed., New York: McGraw-Hill, 1985, pp. 393–451.

Bipolar and Field-Effect Transistors and Their Circuits

18.1 Bipolar Junction Transistors

Figure 18.1 shows the geometry, representation, and the symbol for an NPN BJT. Figure 18.1(a) shows the cross-section of a planar BJT. This drawing shows a central volume of n-type doped Si called the emitter, a p-type Si layer around the emitter volume called the base, and an n-type volume of doped Si called the collector. Metallic leads are connected to each of these semiconductor sections to form the transistor.

Figure 18.1(b) is another representation of an NPN BJT, showing the n-type emitter, the p-type base, and the n-type collector stacked one above the other, with leads exiting from each region. This representation shows the junctions between the elements.

The symbol for the NPN BJT used in schematic diagrams is shown in Figure 18.1(c). The arrow points in the direction of conventional current flow.

The operation of the Si NPN transistor is as follows. This device acts as a current-controlled valve, controlling the flow of electrons between the emitter and the collector. The collector is biased very positive with respect to the base. The combination of base and collector thus acts like a back-biased pn junction diode.

The base may be biased positive or negative with respect to the emitter, depending on whether the device is turned on or off. The combination of base and emitter also acts as a junction diode. If the base is biased positive with respect to the emitter, electrons are allowed to flow from the emitter to the base by the process of diffusion. When entering the base, the electrons are minority carriers (holes being the majority carriers in p-type material). Some of the carriers exit the base as base current. The largest numbers of them, however, see the very high positive field at the collector-base junction and are swept into the collector region. There they exit the collector through the collector lead. The more positive the base bias, the larger the base current and the larger the collector current. Thus, the device is a current amplifier, since a small change in base current produces a large change in collector current.

Figure 18.2 shows similar illustrations for the PNP-type BJT. Figure 18.2(a) shows the cross-section of a planar BJT. The figure illustrates a central p-type volume of doped Si called the emitter, an n-type Si layer around the emitter volume called the base, and a p-type volume of doped Si called the collector. Metallic leads are connected to each of these semiconductor sections to form the transistor.

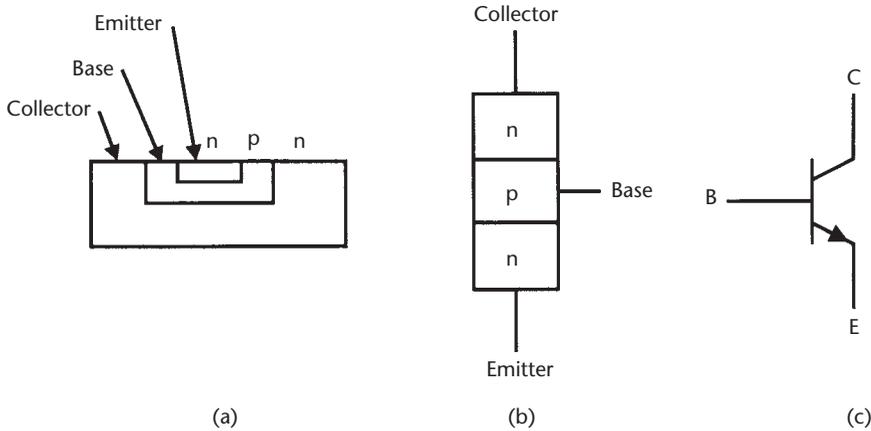


Figure 18.1 NPN BJT: (a) cross-section of a planar NPN BJT; (b) elements of an NPN BJT; and (c) symbol for the NPN BJT.

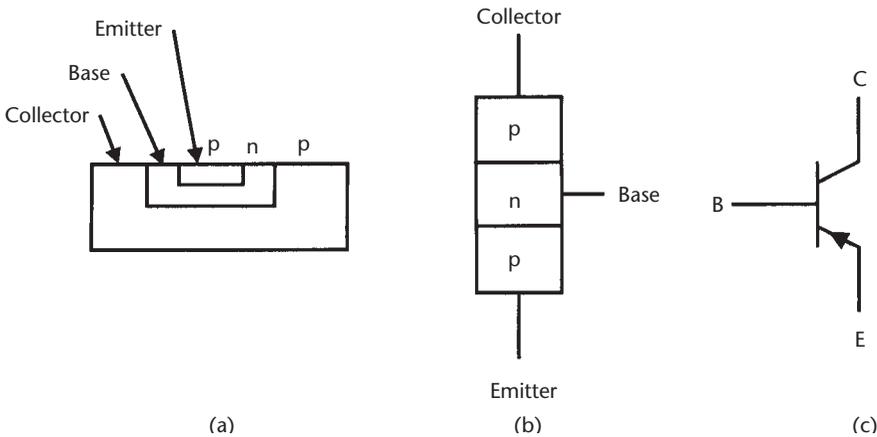


Figure 18.2 PNP BJT: (a) cross-section of a planar PNP BJT; (b) elements of a PNP BJT; and (c) symbol for the PNP BJT.

Another representation of a PNP BJT is shown in Figure 18.2(b). Illustrated are the p-type emitter, the n-type base, and the p-type collector stacked one above the other, with leads exiting each region. The figure shows clearly the junctions between the elements.

Figure 18.2(c) is the symbol for the PNP BJT used in schematic diagrams. As in Figure 18.1, the arrow points in the direction of conventional current flow.

The operation of the Si PNP transistor is as follows. This device acts as a current-controlled valve, controlling the flow of holes between the collector and the emitter. This time, the collector is biased very negative with respect to the base. The combination of base and collector acts like a back-biased pn junction diode.

The base may be biased negative or positive with respect to the emitter, depending on whether the device is turned on or off. The combination of base and emitter also acts as a junction diode. If the base is biased negative with respect to the emitter, holes are allowed to flow from the emitter to the base by the process of diffusion.

When entering the base, the holes are minority carriers (electrons being the majority of carriers in n-type material). Some of these carriers exit the base as base current with electrons from the base combining with the holes. The largest number of the holes, however, sees the very high negative field at the collector-base junction and is swept into the collector region. There they exit the collector through the collector lead (they combine with electrons, so there is a flow of electrons to the collector). The more negative the base bias, the larger the base current and the larger the collector current. Hence, this device is also a current amplifier, since a small change in base current produces a large change in collector current.

Currently both types of BJTs are used in electronic systems. NPN devices are the most frequently used of the two types. UHF and higher frequency devices are of the NPN type because of the higher mobility of electrons as majority carriers, which means improved HF power gain. PNP-type transistors are used primarily in land-mobile communications equipment requiring a positive ground system. The following discussions of amplifier configurations apply only to NPN transistors.

18.2 BJT Amplifier Configurations

There are three main types of BJT amplifier configurations: the common emitter, the common base, and the common collector amplifiers.

18.2.1 Common-Emitter Amplifier

The common-emitter amplifier is the configuration most often used at VHF and lower frequencies. It is also frequently used at UHF and microwave frequencies. Advantages of this configuration over the common-base amplifier include higher voltage gain, higher current gain, higher power gain, and higher input impedance. It usually has the lowest noise figure, good stability, and is relatively easy to match input and output impedances to 50Ω or 75Ω , as applicable. Figure 18.3 shows two schematic diagrams for this type of circuit (both for NPN transistors).

The circuit in Figure 18.3(a) uses resistors between the collector power supply and the collector and between the base bias power supply and the base. Signals are coupled in and out using coupling capacitors. A resistive load is used. This circuit is typical for a wideband, lower frequency, small-signal amplifier.

The circuit in Figure 18.3(b) uses RFCs between the collector power supply and the collector and between the base bias power supply and the base. Signals are coupled in and out using coupling capacitors. A resistive load also is used. This circuit is more typical for a narrowband, HF, small-signal amplifier.

There are other possible circuit variations for the NPN BJT common-emitter amplifier. One important one is the circuit that uses a transformer in place of the RFC or resistor that connects between the collector and the collector power supply. This amplifier configuration is shown in Figure 18.4.

The transformer may be either tuned or untuned. Transformers sometimes are also used as inputs to the base. One end of the transformer secondary is connected to the base bias supply. Such transformers also may be either tuned or untuned, depending on the application. These transformers can be used for impedance

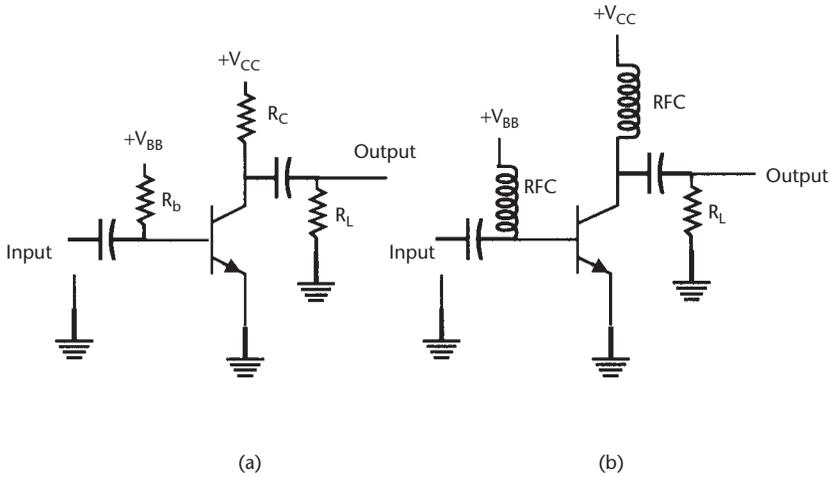


Figure 18.3 BJT common-emitter amplifiers: (a) amplifier using a resistor connected to V_{cc} and (b) amplifier using an RFC connected to V_{cc} .

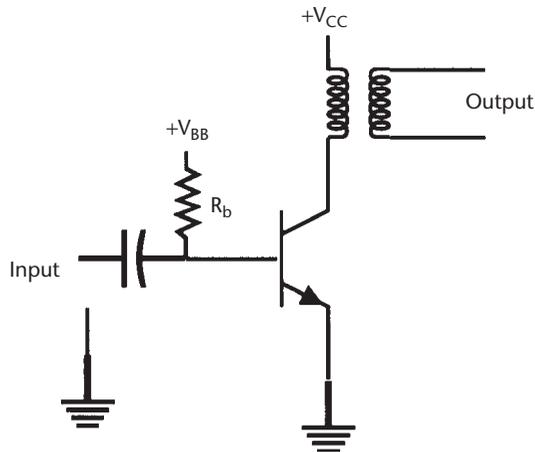


Figure 18.4 BJT common-emitter amplifier using a transformer connected to V_{cc} .

matching and filtering. Other types of impedance matching, filtering, and coupling used with common-emitter amplifiers are discussed later.

Figure 18.5 shows a typical common-emitter characteristics plot for the amplifier in Figure 18.3(a). This plot shows the assumed collector current for the transistor as a function of base current and collector-to-emitter voltage. This plot shows typical dc and ac load lines with a Q point corresponding to the bias conditions. The supply voltage is +8V; the collector resistance, R_c , is 1 k Ω ; the load resistance, R_L , is 1 k Ω ; and the base bias current is 40 Micro MA. The value of h_{FE} (current gain) for this transistor is assumed to be 100, so the collector current is 4 MA, and the dc collector voltage is 4V.

The dc load line is constructed by locating a first point where there is no current through the resistor ($I_c = 0$ and $V_{ce} = V_{cc}$), and a second point where the current is a

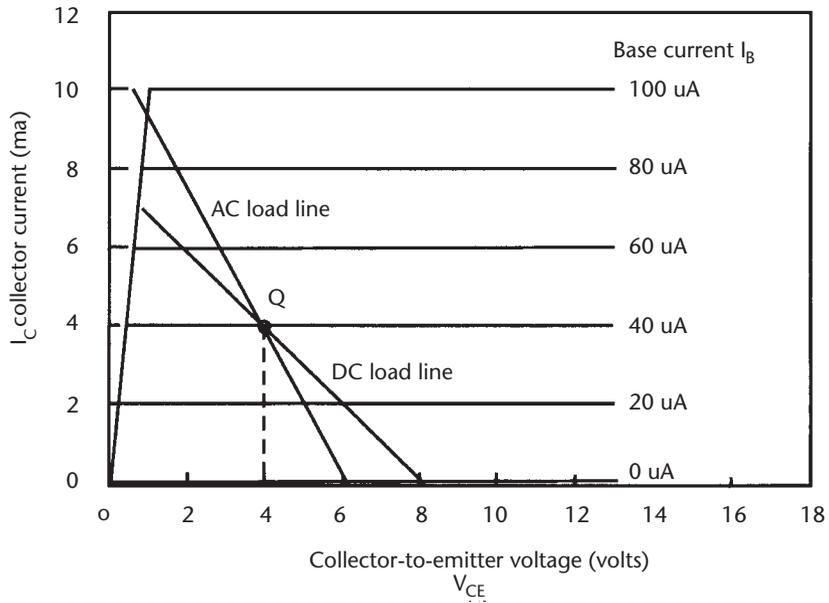


Figure 18.5 Plot of typical common-emitter characteristics (resistor between collector and power supply).

maximum and all the voltage is across the resistor ($I_C = 8 \text{ mA}$ and $V_{CE} = 0$). The dc load line is a straight line drawn between those points.

The ac load line for this case is not the same as the dc load line, because the collector sees R_C in parallel with R_L . This parallel resistance is 500Ω . Thus, we have a 500Ω ac load line passing through the Q point. With no ac signal input, the transistor is biased at point Q . With an ac signal present, the operating point moves up and down the ac load line above and below the Q point. If the input signal goes positive so that the base current increases to $60 \mu\text{A}$, the operating point is moved to $V_{CE} = 3\text{V}$ and $I_C = 6 \text{ mA}$. If the input signal goes negative so that the base current decreases to $20 \mu\text{A}$, the operating point is moved to $V_{CE} = 5\text{V}$ and $I_C = 2 \text{ mA}$.

Figure 18.6 is a plot of the typical common-emitter characteristics plot for the amplifier in Figure 18.3 with an RFC connected to the collector for dc bias current. The collector is capacitively coupled to a $1\text{-k}\Omega$ load resistor. Typical dc and ac load lines are shown, with point Q corresponding to the bias conditions.

The supply voltage is again $+8\text{V}$ and the bias base current is $40 \mu\text{A}$. The assumed value of h_{FE} for this transistor is again 100, so the collector current is 4 mA , and the dc collector voltage is 8V .

The dc load line is constructed by locating a first point where there is no current through the choke ($I_C = 0$ and $V_{CE} = V_{CC}$), and a second point where the current is a maximum and no voltage is across the choke ($I_C = 8 \text{ mA}$ and $V_{CE} = V_{CC}$). The dc load line is a straight line drawn between those points. In this case, the dc load line is a vertical line.

The ac load line is constructed by locating a first point at the Q point and a second point where the current is a maximum and no the voltage is across the transistor ($I_C = 8 \text{ mA}$ and $V_{CE} = 0$). The ac load line is a straight line drawn between those points and extending to the point where the collector current is zero.

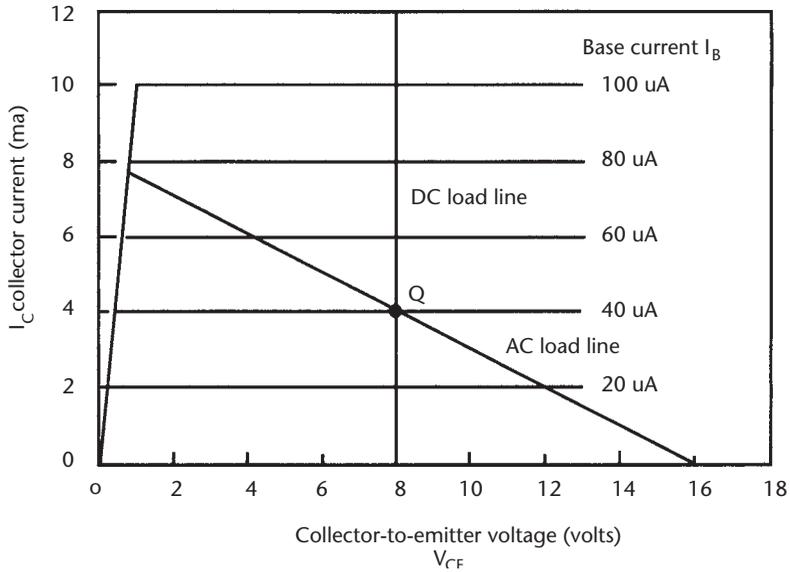


Figure 18.6 Plot of typical common-emitter characteristics (RFC between collector and power supply).

With no input ac signal, we are at point Q . With the presence of an ac signal, movement is along the ac load line above and below the Q point. If the input signal goes positive so that the base current increases to $60 \mu\text{A}$, the operating point is moved to $V_{CE} = 4\text{V}$ and $I_C = 6\text{ mA}$. If the input signal goes negative so that the base current decreases to $20 \mu\text{A}$, the operating point is moved to $V_{CE} = 12\text{V}$ and $I_C = 2\text{ mA}$.

The current gain, β , of the transistor is not constant with frequency but changes, as shown in Figure 18.7. Over much of the frequency range, it decreases inversely with frequency. Key frequencies that are shown are f_{β} , where the power gain is down by 3 dB; f_t , which is known as the gain-bandwidth product; and f_1 , where β reaches a value of 1.0. The value of h_{fe} or β changes widely from device to device of the same type.

Figure 18.8 shows ac equivalent circuits for the BJT common-emitter amplifier. The figure illustrates that the transistor is a fairly complex electrical circuit consisting of a current generator, resistors, and capacitors. Figure 18.8(a) shows the midband equivalent circuit, where no capacitors are involved. The circuit in Figure 18.8(b) shows the HF equivalent circuit, with two capacitors. C_{π} typically is in the range of 30–500 pf. C_{μ} typically is in the range of 1–20 pf.

The following midband design equations are for the BJT common-emitter amplifier.

Current gain:

$$A_i = h_{fe} = \beta = i_C / i_B \quad (18.1)$$

Input resistance at the base:

$$r_i = r_{\pi} = (\beta + 1)(25 \text{ mV} / I_E) \quad (18.2)$$

A small-current symbol (i) indicates ac; a large-current symbol (I) indicates dc.

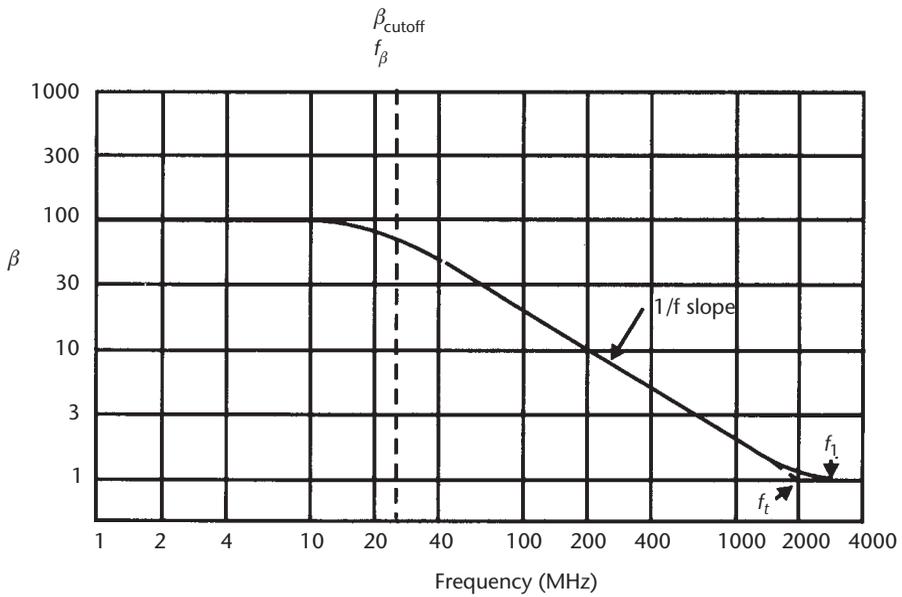


Figure 18.7 Characteristics of an NPN common-emitter amplifier current gain. (After: [1].)

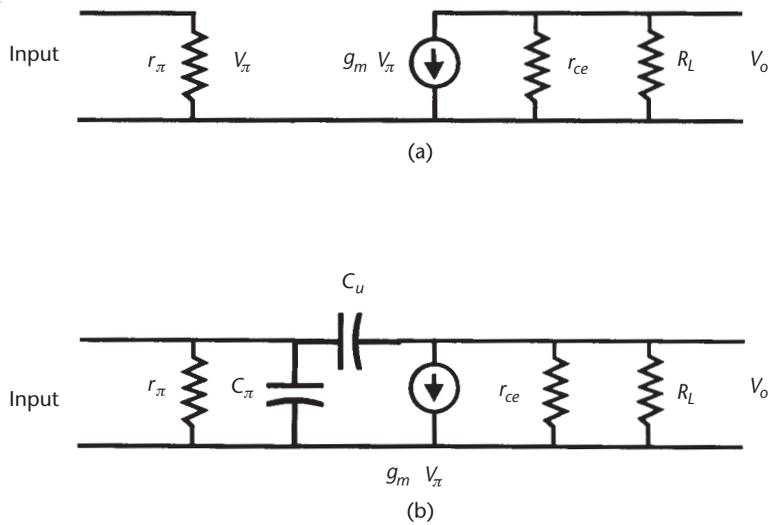


Figure 18.8 Equivalent circuit for BJT common-emitter amplifier: (a) midband equivalent circuit and (b) HF equivalent circuit. (After: [2].)

Transconductance for the transistor:

$$g_m = I_E / (25 \text{ mV}) = 1/r_e \tag{18.3}$$

where I_E = dc emitter current.

Output resistance for the amplifier:

$$r_o = R_L \parallel r_{ce} \tag{18.4}$$

where R_L = load resistance seen by collector, and \parallel indicates in parallel.

Voltage gain for the amplifier:

$$A_v = -g_m r_0 \quad (18.5)$$

The negative sign indicates a 180-degree phase shift between input and output.

Power gain for the amplifier:

$$A_p = g_m^2 r_0 r_\pi \quad (18.6)$$

For an example of the use of (18.1) to (18.6) in the evaluation of a common-emitter amplifier, assume the following:

$$h_{fe} = \beta = 100$$

$$r_{ce} = 10 \text{ k}\Omega$$

$$I_E = 2.5 \text{ mA}$$

$$R_L = 1 \text{ k}\Omega$$

Then,

$$g_m = I_E / (25 \text{ mV}) \quad (18.7)$$

$$g_m = 2.5 \text{ mA} / (25 \text{ mV}) = 0.1\text{S}$$

$$r_i = r_\pi = (\beta + 1)(25 \text{ mV} / I_E) \quad (18.8)$$

$$r_i = r_\pi = (101)(25 \text{ mV} / 2.5 \text{ mA}) = 1010\Omega$$

$$r_0 = R_L \parallel r_{ce} \quad (18.9)$$

$$r_0 = 1,000 \times 10,000 / 11,000 = 909\Omega$$

$$A_v = -g_m r_0 \quad (18.10)$$

$$A_v = -0.1 \times 909 = -90.9$$

$$A_p = g_m^2 r_0 r_\pi \quad (18.11)$$

$$A_p = 0.01 \times 909 \times 1,010 = 9,181 = 39.6 \text{ dB}$$

At the higher frequencies, the equivalent circuit also involves the inductance of the device input and output conductors. It is common practice for higher frequency applications to use S-parameters to characterize the devices.

18.2.2 Common-Base Amplifier

The common-base amplifier finds extensive use at the higher frequencies, such as UHF and microwave. That is because of higher cutoff frequency capability for this configuration and reduced coupling between output and input. Reduced coupling

provides improved stability, thereby reducing the tendency for unwanted oscillations. Figure 18.9 is a schematic diagram of a common-base NPN BJT amplifier.

The base in a common-base amplifier is connected directly to ground or at least to RF ground. The signal input is through a coupling capacitor to an RFC that is connected at one end to the emitter and at the other end to a negative emitter bias power supply. The collector is connected through an RFC to the collector power supply. The signal output is through a coupling capacitor to a resistive load. The common-base configuration is the one used primarily in the design of microwave oscillators.

The following design equations are for the common-base amplifier at midband.
Current gain:

$$A_i = \alpha = i_C / i_E \quad (18.12)$$

Input resistance at the base:

$$r_i = (25 \text{ mV} / I_E) \quad (18.13)$$

Transconductance for the transistor:

$$g_m = I_E / (25 \text{ mV}) \quad (18.14)$$

Output resistance for the amplifier:

$$r_o = R_L \parallel r_{ce} \quad (18.15)$$

Voltage gain for the amplifier:

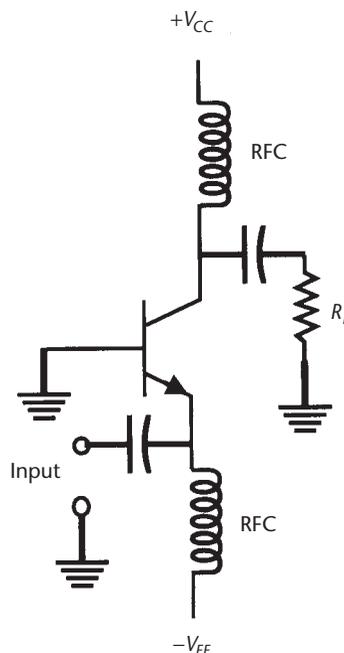


Figure 18.9 Schematic diagram of a common-base NPN BJT amplifier.

$$A_v = g_m r_0 \quad (18.16)$$

Power gain for the amplifier:

$$A_p = \alpha^2 g_m r_0 \quad (18.17)$$

For an example of the use of (18.12) to (18.17) in the evaluation of a common-base amplifier, assume the following:

$$\alpha = 1$$

$$r_{ce} = 10\text{k}\Omega$$

$$I_E = 2.5 \text{ mA}$$

$$R_L = 1\text{k}\Omega$$

Then,

$$g_m = I_E / (25 \text{ mV}) \quad (18.18)$$

$$g_m = 2.5 \text{ mA} / (25 \text{ mV}) = 0.1\text{S}$$

$$r_i = (25 \text{ mV} / I_E) \quad (18.19)$$

$$r_i = (25 \text{ mV} / 2.5 \text{ mA}) = 10\Omega$$

$$r_0 = R_L \parallel r_{ce} \quad (18.20)$$

$$r_0 = 909\Omega$$

$$A_v = \alpha g_m r_0 \quad (18.21)$$

$$A_v = 0.1 \times 909 = 90.9$$

$$A_p = \alpha^2 g_m r_0 \quad (18.22)$$

$$A_p = 0.1 \times 909 = 90.9 = 19.6 \text{ dB}$$

18.2.3 Common-Collector Amplifier

Figure 18.10 is a schematic diagram of a common-collector amplifier.

The circuit in Figure 18.10 is usually called an emitter-follower. In this type of circuit, the output voltage follows the input voltage with an offset of about 0.7V. The emitter-follower is a feedback amplifier with a gain of about 1. The input resistance is high, as indicated by (18.23) and (18.24).

Input resistance:

$$r_\pi = (\beta + 1)25 \text{ mV} / I_E \quad (18.23)$$

$$r_i = r_\pi + (\beta + 1)R_L \quad (18.24)$$

Other design equations are as follows:

Output resistance:

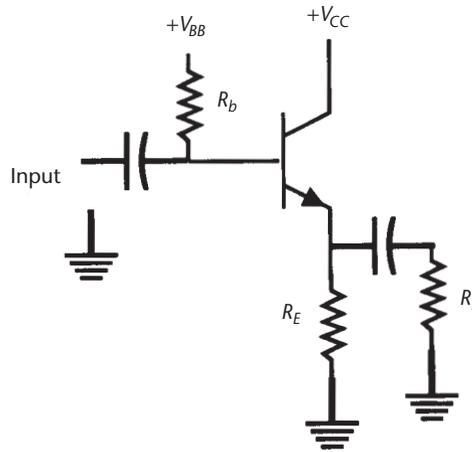


Figure 18.10 A common-collector amplifier.

$$r_o = R_L \parallel \left[\frac{(R_1 + r_\pi)}{(\beta + 1)} \right] \quad (18.25)$$

where R_1 is the signal source resistance.

Voltage gain:

$$A_v = (\beta + 1)R_L / r_i \quad (18.26)$$

Current gain:

$$A_i = \beta + 1 \quad (18.27)$$

For an example of the use of (18.23) to (18.27) in finding the performance characteristics of a common-collector amplifier, assume the following:

$$\beta = 50$$

$$I_E = 10 \text{ mA}$$

$$R_1 = 50 \Omega$$

$$R_L = 200 \Omega$$

Then,

$$r_\pi = (\beta + 1)25 \text{ mV} / I_E \quad (18.28)$$

$$r_\pi = (50 + 1)25 \text{ mV} / 10 \text{ mA} = 127.5 \Omega$$

$$r_i = r_\pi + (\beta + 1)R_L \quad (18.29)$$

$$r_i = 127.5 + (51 \times 200) = 10,327 \Omega$$

$$r_o = R_L \left[\frac{(R_1 + r_\pi)}{(\beta + 1)} \right] \quad (18.30)$$

$$r_o = 200 \left[\frac{(50 + 127.5)}{51} \right] = 3.4 \Omega$$

$$A_v = (\beta + 1)R_L / r_i \quad (18.31)$$

$$A_v = 51 \times 200 / 10,327 = 0.99$$

$$A_i = \beta + 1 \quad (18.32)$$

$$A_i = 50 + 1 = 51$$

18.3 Field Effect Transistors and Circuits

A number of different types of FETs are used in RF systems, including junction FETs (JFETs), metal-semiconductor FETs (MESFETs), and metal-oxide semiconductor FETs (MOSFETs).

18.3.1 Junction Field-Effect Transistors

An n-channel, JFET is illustrated in Figure 18.11. This device is made of Si with n- and p-type doping. The elements include the source, the gates, and the drain. An n-channel JFET has an n+ source, p+ gates, an n+ collector, and an n conducting channel connecting the source to the drain. The transistor acts as a voltage-controlled resistor, with the resistance of the conducting channel depending on the gate-to-source voltage, V_{GS} .

Under all conditions of bias, a depletion region forms near the junction of a JFET. The depletion region effectively reduces the width of the n-channel between source and drain and increases the resistance of the channel. The stronger the reverse bias, the larger the depletion region is and the greater the resistance of the channel. When the field becomes sufficiently strong, the channel is completely pinched off and no current flows.

Figure 18.12 is a schematic diagram of a common-source n-channel JFET amplifier. The gate is connected to a negative voltage by R_G , and the drain is connected to a positive voltage by R_D . The output is to a resistive load by means of a coupling capacitor connected to the drain.

The following midband design equations are for JFET common-source amplifiers:

Current gain:

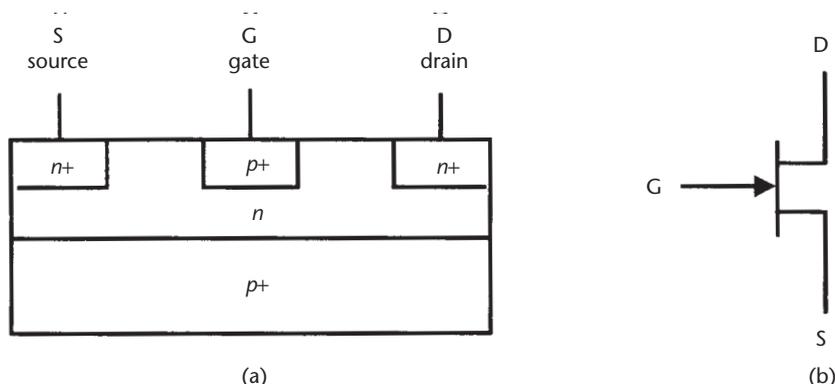


Figure 18.11 JFET: (a) geometry and (b) symbol. (After: [3].)

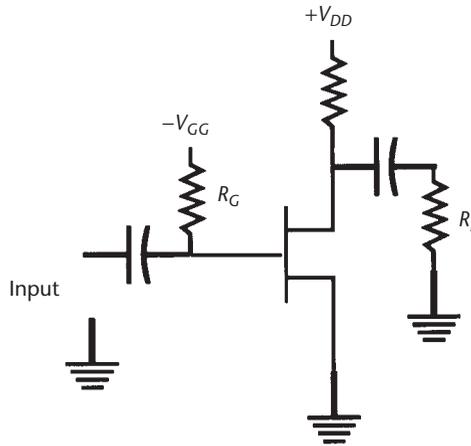


Figure 18.12 Schematic diagram of a common-source JFET amplifier.

$$A_i \Rightarrow 10^7 \quad (18.33)$$

Input resistance at the gate:

$$r_i = r_{gs} \quad (18.34)$$

Transconductance for the transistor: given by the manufacturer's data sheet.

Output resistance for the amplifier:

$$r_o = R_L \parallel r_{ds} \quad (18.35)$$

Voltage gain for the amplifier:

$$A_v = -g_m r_o \quad (18.36)$$

Typical small-signal parameters for JFETs are as follows:

$$g_m = 0.1 \text{ to } 10 \text{ mA/V}$$

$$r_{gs} = 10^9 \Omega$$

$$r_{ds} = 0.01 \text{ to } 1 \text{ M}\Omega$$

$$C_{gs} = 1 \text{ to } 10 \text{ pF}$$

$$C_{gd} = 1 \text{ to } 10 \text{ pF}$$

For an example of the performance of a JFET common-source amplifier, assume the following:

$$V_{DD} = 20 \text{ V} \quad g_m = 0.005 \text{ S}$$

$$I_D = 5 \text{ mA} \quad R_L = 2 \text{ k}\Omega$$

$$R_E = 200 \Omega \quad r_{ds} = 100 \text{ k}\Omega$$

Then,

$$A_v = -g_m r_o = -0.005 \times 1,961 = -9.8 = 19.8 \text{ dB}$$

The voltage gain of JFET amplifiers tends to be less than that of BJT amplifiers.

Figure 18.13 is a schematic diagram of a common-drain JFET amplifier, also known as a source-follower.

The following midband design equations are for JFET common-drain amplifiers:

Current gain:

$$A_i \Rightarrow 10^7 \quad (18.37)$$

Input resistance at the base:

$$r_i = r_{gs} \quad (18.38)$$

The transconductance for the transistor, which depends on the design of the JFET, usually is given by the manufacturer's data sheet. It depends on the design of the JFET.

Output resistance for the amplifier:

$$r_o = R_L \parallel 1/g_m \quad (18.39)$$

Voltage gain for the amplifier:

$$A_v = R_L / (1/g_m + R_L) \quad (18.40)$$

The JFET is not a good HF device, largely because it has the problem of charge storage in the gate and it does not use high-mobility materials. It also is not a high-power device and therefore is used primarily for low-power LF applications, where high input resistance is desirable.

18.3.2 Metal-Semiconductor Field-Effect Transistors

Figure 18.14 is an illustration of a MESFET. The MESFET diagram looks very much like the JFET diagram. The difference is that the PN junction-type diode used as the

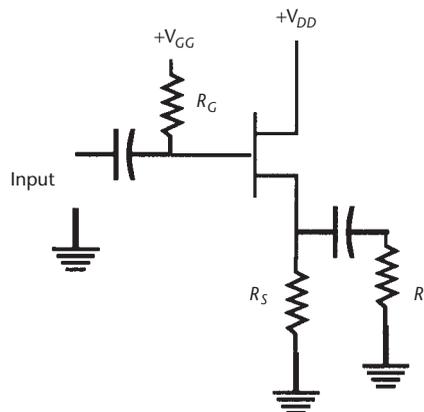


Figure 18.13 Common-drain JFET amplifier (source follower).

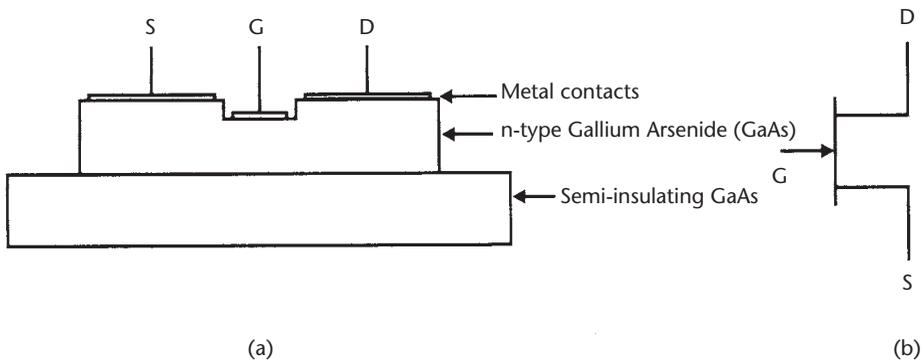


Figure 18.14 MESFET: (a) geometry and (b) symbol. (After: [4].)

gate is replaced by an HF Schottky-barrier diode. That eliminates the problem of charge storage, one of the main limitations for the JFET for HF operation.

The MESFET uses an n-type GaAs semiconductor channel, which connects a source contact region and a drain contact region. Between those regions is a narrow gate consisting of a metallic contact placed on the surface of the semiconductor channel and forming a Schottky-barrier diode. Associated with this interface is a depletion region that has a width controlled by the gate voltage.

GaAs is used with this device rather than Si because of its much higher mobility. That, coupled with the use of a Schottky-barrier gate with a length of only about 1 μm , allows its use as a microwave amplifier with very good operating characteristics.

The operation of the MESFET is similar to that of the JFET. Both devices are depletion-mode devices in which the depletion region cross-section near the gate is determined by the voltage at the gate diode. As the gate-to-source voltage is made more negative, the cross-section of the conducting channel decreases. At some negative voltage, the conducting channel is completely pinched off. A typical pinch-off voltage is about -1.5V . At saturation, the output resistance is about $400\text{--}500\Omega$.

The main type of amplifier configuration used with the MESFET is the common-source amplifier. Common-gate designs also are used. The voltage gain for the MESFET amplifier is approximately equal to $g_m r_o$. Thus, for an output resistance of 400Ω and a g_m of 60 mS , the voltage gain would be 24. There could be a fairly large power gain for the device because of the very high input resistance of the device.

The MESFET has been the workhorse of the microwave industry for many years. It is used as the active device for both low-noise amplifiers and power amplifiers. It is also used for oscillators, mixers, and other microwave circuits. It can be used in monolithic circuitry as well as in single devices. For more information on MESFETs, see [5].

18.3.3 Metal-Oxide Semiconductor Field Effect Transistors

An n-channel enhancement-mode metal-oxide semiconductor field effect transistor (MOSFET) is illustrated in Figure 18.15. The symbol for this transistor shows a broken line between the source and the drain. That indicates that the device is

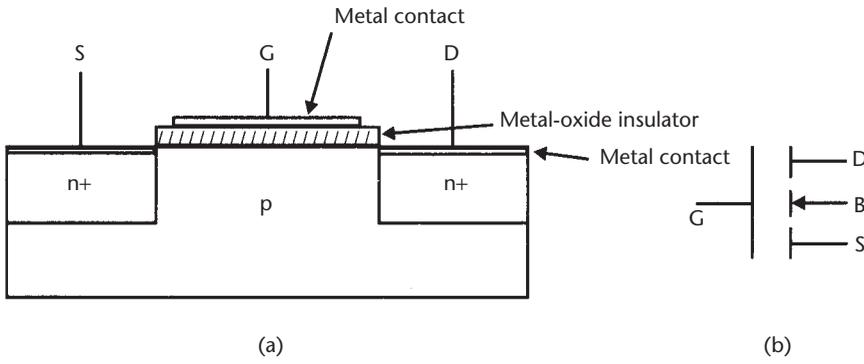


Figure 18.15 N-channel enhancement-mode MOSFET: (a) geometry and (b) symbol. (After: [3].)

normally cut off. (Sometimes the symbol for a MOSFET does not use the broken line, for reasons of simplicity.)

The channel current for a MOSFET is controlled by a voltage at a gate electrode that is isolated from the channel by an insulator. For that reason, the resulting device also is called an insulated-gate FET (IGFET). In the most common configuration, a metal-oxide layer is grown or deposited on the semiconductor surface, and the metal gate electrode is deposited onto that oxide layer.

In an insulated-gate or enhancement-mode MOSFET the source and drain regions may be n+ type material. These regions are separated by p-type material, with the insulated gate bridging over this connecting material. When a positive voltage is applied to the gate, electrons in the p-type material are attracted to the surface below the insulator, forming a connecting near-surface region of n-type material between the source and the drain. The larger or more positive the gate-to-source voltage, the lower the resistance of this region between the source and the drain. It takes some minimum gate voltage, called the threshold voltage, to turn this device on. A typical value may be in the range of 2–4V, depending on the device. Near the threshold voltage, the transistor is very nonlinear. At some higher voltages, the device becomes more nearly linear, and the value of forward transconductance, g_m , becomes larger.

The only amplifier configuration that is used extensively for the MOSFET is the common-source amplifier. This circuit is shown in Figure 18.16. The gate normally is biased a few volts positive with respect to the source, so the device will operate in the more linear and higher gain region of its operating characteristics.

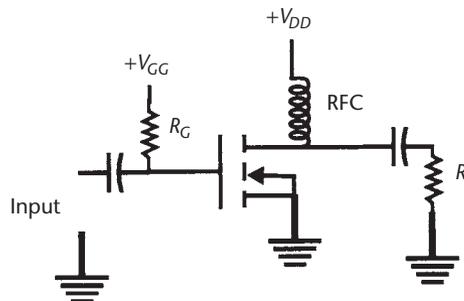


Figure 18.16 Common-source MOSFET amplifier.

An n-channel depletion/enhancement-mode MOSFET is illustrated in Figure 18.17. The solid line in the symbol between the source and the drain indicates that the device normally is conducting.

The channel current is controlled by a voltage at a gate electrode that is isolated from the channel by an insulator. In the most common configuration, a metal-oxide layer is grown or deposited on the semiconductor surface, and the metal gate electrode is deposited onto that oxide layer.

In a depletion/enhancement-mode MOSFET, the source and drain regions may be n+ type material. These regions are separated by n-type material, with the insulated gate bridging over the connecting material. When the gate-to-source voltage is zero, there is a moderate conductance between the source and the drain. When a negative voltage is applied to the gate, the number of electrons in the n-type channel between source and drain are reduced at the surface below the insulator. The resistance of the conduction channel thus is increased. The more negative the gate-to-source voltage, the higher the resistance of the region between the source and the drain. When a positive voltage is applied to the gate, the number of electrons in the n-type channel between source and drain are increased at the surface below the insulator. The resistance of the conduction channel is thus decreased. The more positive the gate-to-source voltage, the lower the resistance of the region between the source and the drain.

Figure 18.18 shows a common-source amplifier using an n-channel depletion/enhancement-mode MOSFET. The gate is connected to ground using a resistor. The drain is connected to a positive drain power supply, using an RFC. The output is to a resistive load via a coupling capacitor.

18.4 Comparison of FET and BJT Amplifiers

As indicated earlier, there are only two types of BJTs: the NPN and the PNP. UHF and higher frequency devices are of the NPN type because of the higher mobility of electrons as majority carriers, which means higher f_t and improved HF power gain. PNP-type transistors are used primarily in land-mobile communications equipment

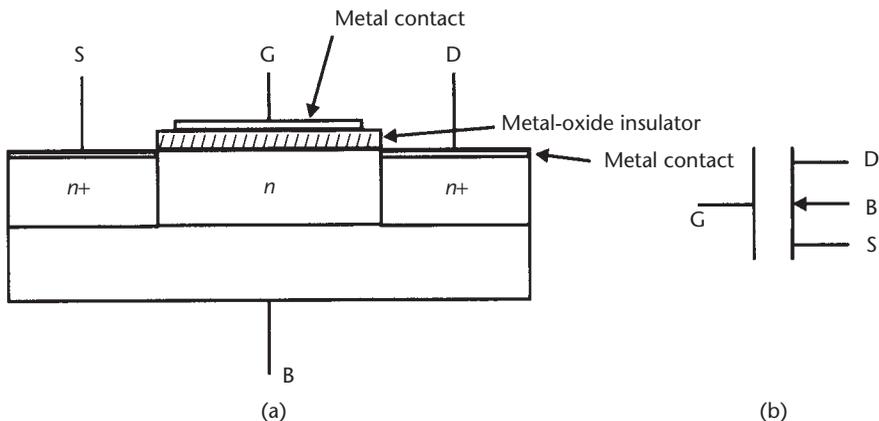


Figure 18.17 N-channel depletion/enhancement-mode MOSFET: (a) geometry and (b) symbol.

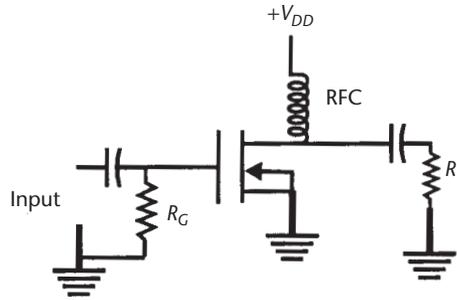


Figure 18.18 Depletion/enhancement type MOSFET amplifier.

requiring a positive ground system. They also may be used in circuits involving complementary symmetry operation, as in a push-pull amplifier.

Far more types of FETs are commercially available for RF amplifier use. Many of these devices have been around only a fairly short time, whereas BJTs have been around quite a while. The first 1-GHz BJTs were Ge, came to market around 1965 from Texas Instruments, and cost about \$300 each. One such device was the TIX3024 transistor. It took five or six of them for a 20-dB gain, 1- to 2-GHz amplifier.

There are applications where BJTs clearly are better than FETs, and there will continue to be many systems that use this type of device. On the other hand, there clearly are many applications where FETs are superior to BJTs. That is particularly true for power applications and for microwave and millimeter-wave small-signal amplifiers. FETs have the advantage over BJTs at the higher frequencies because they are able to use GaAs rather than Si. GaAs has a higher mobility than Si and higher peak electron velocities. These two features result in faster transit time and lower power dissipation. Therefore, they have higher frequency capability, higher gain, lower noise figure, and usually better efficiency.

MESFETs are used at all frequencies, from a few cycles per second to 40 GHz or more. At the lower frequencies, these transistors use Si as the device material. At microwave frequencies, these devices use GaAs. Several watts per transistor are available up to 15 GHz. Power outputs of a few hundred milliwatts are available from single GaAs FETs at 30 GHz. Noise figures below 0.3 dB are attainable at 4 GHz for low-noise GaAs FETs (e.g., the NEC32584 device). They are as low as 1.4 dB at 20 GHz.

MOSFET power transistors have relatively high CW output power at frequencies up to about 1.5 GHz. Output CW power from the M/A-COM type DU28200M transistor is reported to be 200W at 175 MHz. The M/A-COM type UHF28150J transistor is reported to provide an output CW power of 150W at 500 MHz, while the M/A-COM type LF40100 transistor is reported to provide an output CW power of 100W at 1,000 MHz.

The same RF design practices, such as grounding, filtering, bypassing, and creating a good circuitboard layout apply equally to circuits using any of these devices. Precautions must be taken with each type of device to prevent damage or destruction of the device. FETs are highly sensitive to gate rupture, which can be caused by excessive dc potentials or transients between the gate and the source. Of particular concern is static electricity. Technicians handling or testing the devices must be careful to be properly grounded along with the device.

A weak spot with BJTs is the possibility of thermal runaway. The main reason for thermal runaway with BJTs is the increasing h_{FE} . Care must be taken in the design of the bias circuit to make sure that does not happen. Self-biasing techniques often are used that provide negative feedback designed to prevent the problem.

18.5 High Electron Mobility Transistors and Heterojunction Bipolar Transistors

High electron mobility transistors (HEMTs) and heterojunction bipolar transistors (HBTs) are important recent developments in microwave and millimeter-wave transistors. These devices make use of heterojunctions for their operation. The heterojunctions are formed between semiconductors of different compositions and bandgaps, for example, GaAs/AlGaAs and InGaAs/InP. That is unlike conventional transistors, which use junctions between like materials. These relatively new types of devices offer significant improvements for low-noise amplifiers and microwave power amplifiers.

Some of the information presented in this section is adapted from [6].

18.5.1 High Electron Mobility Transistors

Figure 18.19 shows a schematic cross section of an HEMT structure using GaAs and AlGaAs. The conventional HEMT is similar to a GaAs MESFET. As seen in Figure 18.19, the HEMT has two ohmic contacts (source and drain) and a Schottky gate that modulates the flow of current in the channel between the two contacts. The difference between the two types of devices and the key to the HEMT's improved performance is in the underlying semiconductor material. The HEMT has superior electron transport properties and much higher sheet charge density than the MESFET because of a two-dimension electron gas layer that is formed in a thin layer between the AlGaAs and the undoped GaAs layers. This is shown in Figure 18.19.

HEMTs have demonstrated unprecedented noise performance at cryogenic temperatures and good microwave and millimeter-wave noise and power performance at room temperature at frequencies up to 60 GHz. Typical noise figures at 12 GHz for commercially available low-noise HEMTs are about 1.0 dB. The best reported noise figures at the same frequency are about half that, or about 0.5 dB.

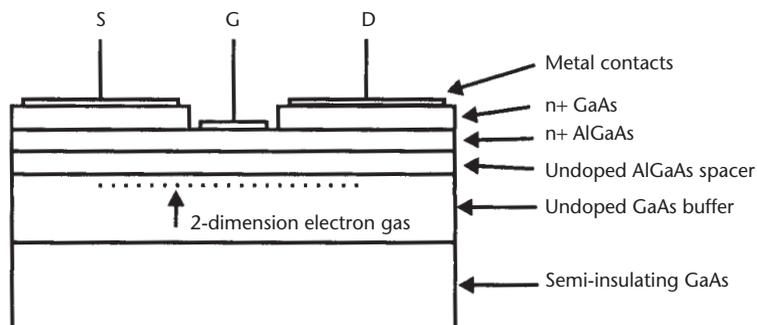


Figure 18.19 An HEMT device. (After: [6].)

Figure 18.20 shows a comparison of HEMT and MESFET room-temperature noise performance up to about 60 GHz. This figure is based on the best-reported noise figures for the two devices and clearly shows the advantages provided by the HEMT device.

In addition to lower noise figure, HEMTs also have several characteristics that make them more attractive for low-noise applications. They are easier to provide impedance matching, and they have a larger gain-bandwidth product.

18.5.2 Heterojunction Bipolar Transistors

The cross-section of a basic n-p-n HBT is shown in Figure 18.21. The n-type emitter is formed in the wideband gap AlGaAs while the p-type base is formed in the lower-band gap GaAs. The n-type collector is also formed in GaAs.

The GaAs MESFET is currently the most widely used microwave device for the amplification of microwave signals. HBTs are also well suited to amplification of large microwave signals. Laboratory results have shown that in the future HBT amplifiers will have a significant advantage over MESFETs for power amplification in terms of output power, compactness, efficiency, and novel circuit usage such as complementary amplifiers. HBTs have been used for frequencies ranging from 3 to 60 GHz. At 3 GHz, up to 1W of CW output power has been obtained with 61% efficiency. The highest output power obtained at microwave frequencies is 5.3W of CW of output power at 8 GHz with 33% efficiency. Improvements are expected in the future.

18.6 DC Bias Circuits for BJT Amplifiers

Good design of the bias circuit for a BJT requires that these two deficiencies of the transistor be overcome: (1) that the transistor is temperature sensitive and (2) that

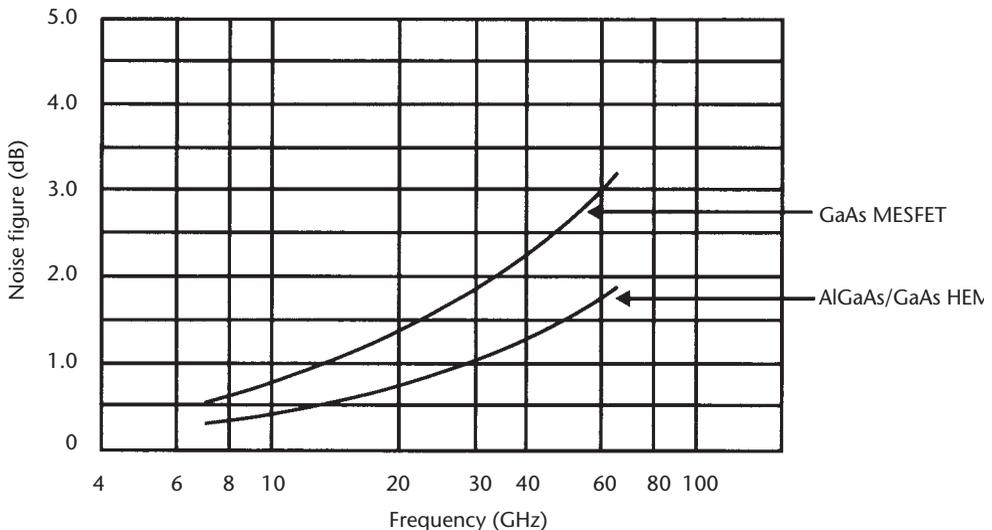


Figure 18.20 Comparison of noise figures for GaAs MESFET and AlGaAs/GaAs HEMT devices. (After: [6].)

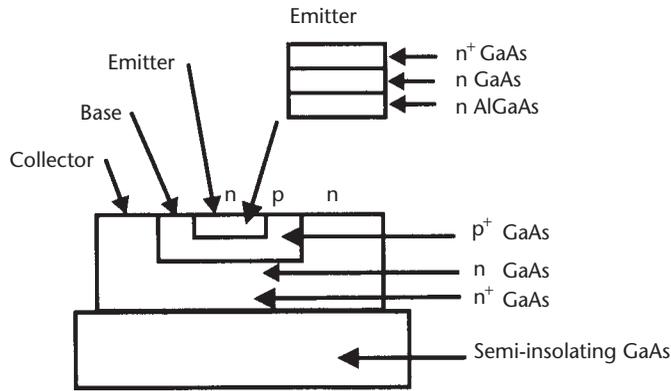


Figure 18.21 An HBT structure. (After: [6].)

the parameters of the transistor are subject to process variations. The bias circuit shown in Figure 18.22 is often used to solve these problems. In the circuit in Figure 18.22, the transistor is shown with a feedback resistor, R_E , and two base resistors, R_1 and R_2 . It is called a self-bias circuit because only one supply voltage is required. This same circuit can also be used for MESFET common-source amplifiers and usually is done that way unless separate gate bias is supplied.

The value of V_{BB} and R_B are given by (18.41) and (18.42).

$$V_{BB} = R_1 V_{CC} / (R_1 + R_2) \tag{18.41}$$

$$R_B = R_1 R_2 / (R_1 + R_2) \tag{18.42}$$

The value of R_E usually is chosen such that there is a 1–3-V drop across it with no RF signal. Thus, if the quiescent operating point for the circuit is 4 ma and a 2-V drop is used, the value of R_E would be 500V.

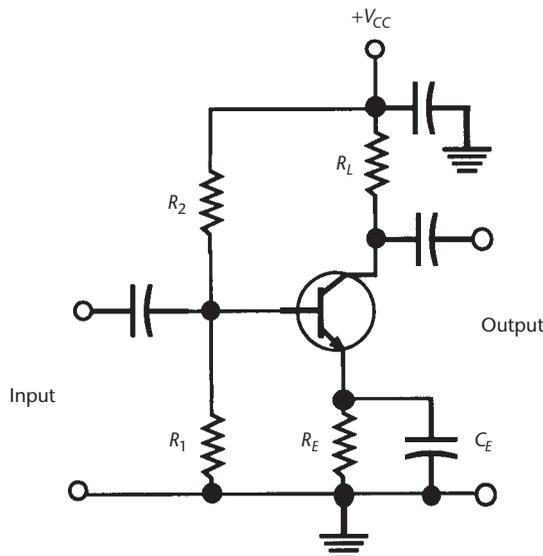


Figure 18.22 Self-bias circuit for common-emitter amplifier.

The value of R_1 is chosen to be less than $\beta_{\min} R_E/5$. Thus, with a transistor minimum β of 50 and $R_E = 500\Omega$, R_1 would be

$$50 \times 500/5 = 5 \text{ K}\Omega$$

The voltage at the base is the voltage drop across $R_E + 0.7\text{V}$. For our example, that would be 2.7V. The current in R_1 thus would be about $2.7/5\text{K} = 0.54 \text{ ma}$. If the value of V_{CC} is 20V, the voltage drop across R_2 is 17.3V.

The current through R_2 is the current through R_1 plus the base current. Assuming a β of 50, the base current would be 0.08 ma. The value of R_2 is thus $17.3/0.62 = 27.9 \text{ K}\Omega$. The value of RB is thus 8.5 K Ω .

With this method of bias, there tends to be only a small change in Q-point current for a large change in β or a large change in temperature. Part or all of the resistance, R_E , can be bypassed with a capacitor to avoid negative feedback for ac signals. That allows the gain of the amplifier to be high and the noise figure to be low. For the example shown, all the emitter resistance is bypassed by C_E for maximum gain.

High-power BJT RF amplifiers sometimes use a clamping diode plus adjustable series resistance to establish the desired base bias voltage for a common-emitter amplifier with the emitter grounded. This dc voltage is fed through an RFC to the base circuit. The diode can be connected mechanically to the transistor heat sink to perform a temperature-compensating function.

More sophisticated bias sources can be made that use integrated circuit voltage regulators. Transistor source followers can be used with these IC regulators to provide the needed current boost and to lower the source impedance.

18.7 Bias Circuits for FET Amplifiers

Low-power LF FET amplifiers can use simple voltage dividers for bias circuits. A circuit of this kind is shown in Figure 18.23 for a small-signal enhancement-mode MOSFET amplifier. The gate in this case is biased positive with respect to the source. A typical threshold voltage would be in the range of 2–4V. For the case of a common-source amplifier with class A operation, the bias voltage would need to be a few volts positive with respect to the threshold voltage.

18.8 Stability with BJT and FET Amplifiers

One of the important problems with high-gain amplifiers is the problem of unwanted oscillations. One approach used to help improve stability and prevent oscillations is to use shielding. It is common practice to place amplifiers in small metal containers with feed-through capacitors and chokes for dc inputs and coaxial connectors and cables for RF inputs and outputs. Radar-absorbing material may be attached to the inside of the cover to reduce radiative coupling effects. A shielding system of this type is shown in Figure 18.24.

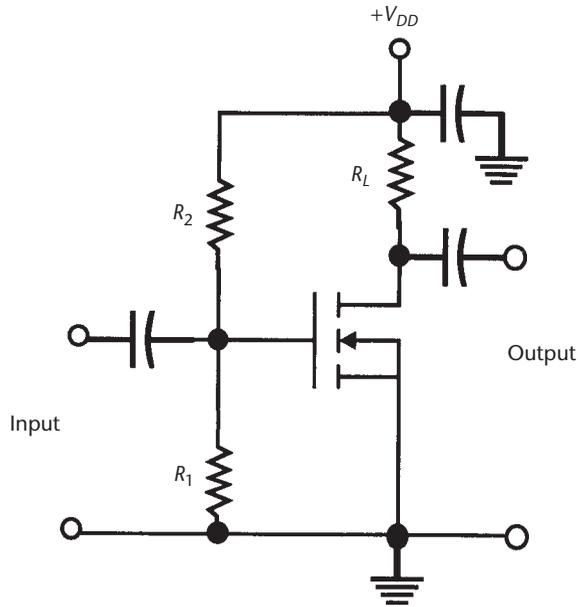


Figure 18.23 A dc bias circuit for enhancement-mode MOSFET.

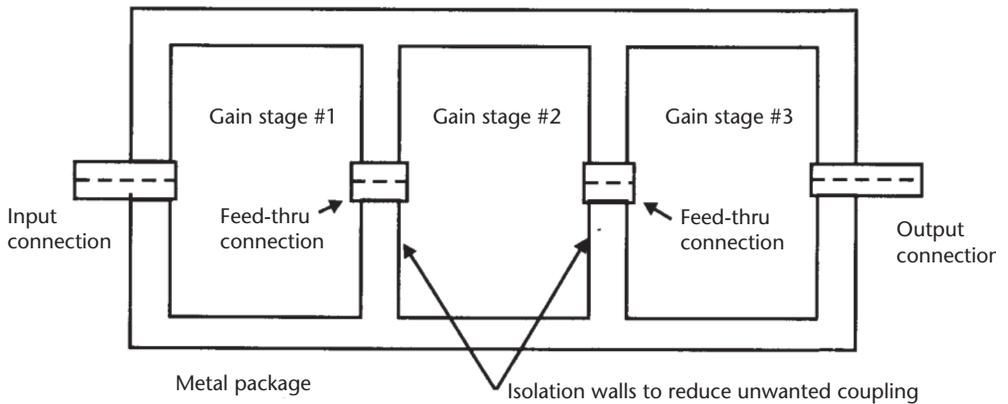


Figure 18.24 Shielding concepts for RF amplifiers.

Another way to improve stability is to use neutralizing circuits. The goal is to provide a negative feedback signal that is equal in amplitude but opposite in phase to the positive feedback signal that is causing the unwanted oscillations.

18.9 Impedance Matching

There are two main methods for providing impedance matching for transistor amplifiers and other RF circuits: with transformers and with L, T, and π LC circuits. The transformer approach is useful mainly at 500 MHz and lower frequencies. The

L, T, and π LC circuits' method is useful at all RF frequencies, including UHF and microwave. The types of components used with the LC circuits are lumped constant circuits for frequencies below about 500 MHz. Above 500 MHz, it is difficult to realize discrete inductors and capacitors. It then becomes necessary to use microstrip or stripline techniques to realize the required matching circuits.

Methods of impedance matching with LC circuits were discussed in detail in Chapter 14. Methods of impedance matching with transformers were discussed in detail in Chapter 15.

18.10 Design Methods with S-Parameters

Some of the following material is quoted or adapted from [7].

Many RF transistors for HF and higher frequencies are characterized using S-parameters. Packaged amplifier and other RF products also may be characterized by S-parameters.

S-parameters can be measured quickly and accurately using test equipment called network analyzers with S-parameter test sets added. Automated S-parameter test equipment is available from Hewlett-Packard Company and Wiltron Corporation and are used by the manufacturers of RF components. Less costly nonautomated S-parameter test equipment nearly always is available in industry for users of the RF components.

S-parameters are simply the coefficients of the incident and reflected voltage waves. S_{11} is the input voltage reflection coefficient and is defined as reflected-wave voltage (b_1) divided by incident-wave voltage (a_1) when incident-wave voltage (a_2) is zero volts. S_{22} is the output voltage reflection coefficient and is defined as the reflected-wave voltage (b_2) divided by incident wave voltage (a_2) when incident-wave voltage (a_1) is zero volts. S_{21} is the forward voltage transmission coefficient and is defined as output-wave voltage (b_2) divided by the input-wave voltage (a_1) voltage when the incident-wave voltage (a_2) is zero volts. S_{12} is the reverse transmission coefficient and is defined as the output-wave voltage (b_1) divided by the incident- or input-wave voltage (a_2) when the incident wave (a_1) is zero volts.

The quantity $|S_{21}|^2$ is the power gain of the transistor at the specified bias conditions and frequency with 50Ω source and load terminations. When S_{21} is given in a data sheet expressed in decibels, it is referring to $10 \log |S_{21}|^2$ or $20 \log |S_{21}|$.

The term G_{tu} is the unilateral, that is, one-way transmission only, transducer power gain with $S_{12} = 0$. It is defined as the power delivered to the load divided by the power available from the source with $S_{12} = 0$. The assumption that S_{12} is close to zero is usually a good one for many high-quality transistors, and it usually is used in calculations of amplifier gain since it greatly simplifies calculations. The value of G_{tu} is given by (18.43).

$$G_{tu} = \frac{[1 - |\Gamma_S|^2]}{|1 - S_{11}\Gamma_S|^2} \times |S_{21}|^2 \times \frac{[1 - |\Gamma_L|^2]}{|1 - S_{22}\Gamma_L|^2} \quad (18.43)$$

where Γ_S is the source reflection coefficient, and Γ_L is the load reflection coefficient.

Equation (18.43) can be broken into three sources of gain. These are:

$$G_s = \left[\frac{1 - |\Gamma_s|^2}{|1 - S_{11} \Gamma_s|^2} \right] \quad (18.44)$$

$$G_o = |S_{21}|^2 \quad (18.45)$$

$$G_L = \left[\frac{1 - |\Gamma_L|^2}{|1 - S_{22} \Gamma_L|^2} \right] \quad (18.46)$$

G_s is the gain contribution achieved through input impedance matching, G_o is the gain contribution of the transistor itself, and G_L is the gain contribution achieved by output impedance matching. The total gain for the three sources expressed in decibels is given by (18.47):

$$G_m = G_s + G_o + G_L \quad (18.47)$$

where the terms are in decibels.

If the circuit design is narrowband and we want maximum power gain, all that is required is to set $\Gamma_s = S_{11}^*$ and $\Gamma_L = S_{22}^*$, where * indicates complex conjugate. If the circuit is broadband and we want a certain amount of gain across a band of frequencies, we can use circuits that compensate for the variations in gain with frequency for the device. That can be done with selective mismatching.

For an example of the use of S-parameters in the design of a small-signal RF amplifier, assume that the source and load impedances are each 50Ω . Also assume that interest is only in maximum gain at 1,000 MHz. The assumed transistor is the Motorola MRF 571. Our bias will be 6V at 50 mA. The manufacturer's data sheet shows that f_T is near its peak at 50 mA, and values of scattering parameters are given for this bias.

S-parameters at the desired frequency and bias points are as follows:

$$S_{11} = 0.60 \text{ at an angle of } +156 \text{ degrees}$$

$$S_{22} = 0.11 \text{ at an angle of } -164 \text{ degrees}$$

$$S_{12} = 0.09 \text{ at an angle of } +70 \text{ degrees}$$

$$S_{21} = 4.40 \text{ at an angle of } +75 \text{ degrees}$$

The required value of $\Gamma_s = S_{11}^* = 0.60$ at an angle of -156 degrees. The required value of $\Gamma_L = S_{22}^* = 0.11$ at an angle of $+164$ degrees.

The gains contributed by each of the gain sources are as follows:

$$G_s = \left[\frac{1 - |\Gamma_s|^2}{|1 - S_{11} \Gamma_s|^2} \right] \quad (18.48)$$

$$G_s = 0.64/0.41 = 156 = 193 \text{ dB}$$

$$G_o = |S_{21}|^2 \quad (18.49)$$

$$G_o = 100 = 2000 \text{ dB}$$

$$G_L = \left[\frac{1 - |\Gamma_L|^2}{|1 - S_{22} \Gamma_L|^2} \right] \quad (18.50)$$

$$G_L = 0.988/0.976 = 1.012 = 0.053 \text{ dB}$$

The total gain with conjugate matching is then

$$\begin{aligned} G_{tu} &= G_S + G_0 + G_L \\ G_{tu} &= 193 + 20.00 + 0.05 = 2198 \text{ dB} \end{aligned} \quad (18.51)$$

From the preceding example, note that there is already a nearly perfect impedance match at the output without adding matching circuits. Therefore, we will select a design that uses only input matching.

Next, select the design of the impedance-matching sections using the Smith chart shown in Figure 18.25.

One approach to finding the impedance-matching components involves starting at the 50Ω input and moving to the complex conjugate value of S_{11}^* . That value is $S_{11}^* = 0.60$ at an angle of -156 degrees.

For the input, start at the center of the Smith chart at point 0 and use the chart as an admittance chart. Travel on the 1.0 conductance circle to point A. The length of that path is 1.63 units long, indicating a normalized capacitive susceptance component of $+j 1.63$ and a real capacitive susceptance of $0.032S$. Next, transfer through the center of the chart to point B, converting to an impedance chart. Then travel to point C, which is the S_{11}^* point. The length of that path is found to be 0.24 unit long, indicating a normalized inductive reactance component of $+j 0.24$. The real inductive reactance is 50 times that value, or $+j 12.0\Omega$.

Let us now find the component values for impedance matching at 1,000 MHz.

$$X_L = 12.0\Omega$$

$$L = X_L / 2\pi f = 12.0 / (2\pi \times 1,000 \times 10^6) = 1.9 \text{ nH}$$

$$B_C = 0.32S$$

$$C_1 = BC / 2\pi f = 0.032 / (2\pi \times 1,000 \times 10^6) = 5.1 \text{ pF}$$

The optimum source reflection coefficient for minimum noise figure for a transistor amplifier can be given by a transistor data sheet. In general, that will be different from the S_{11}^* point for maximum gain. The gain available when the transistor is matched for low-noise figure general will in be less than would be the case if the amplifier using the device had been designed for maximum or near-maximum gain. That gain is termed associated gain and usually is supplied on data sheets for low-noise devices.

18.11 Manufacturers' Data Sheets For Transistors

Manufacturers' data sheets normally are available for the RF transistors and packaged amplifiers of interest. In the data sheets, transistors are characterized by two types of parameters: dc and functional. The dc specifications consist of breakdown voltages, leakage currents, the dc beta (h_{FE}) for BJTs, and capacitances. The functional specifications cover gain, noise figure, Z_{in} , Z_{out} , S-parameters, current gain-bandwidth product, distortion, ruggedness, and so on. The data sheets also provide thermal characteristics for the transistors and packaged amplifiers. In providing those characteristics, the data sheets may provide minimum, typical, and

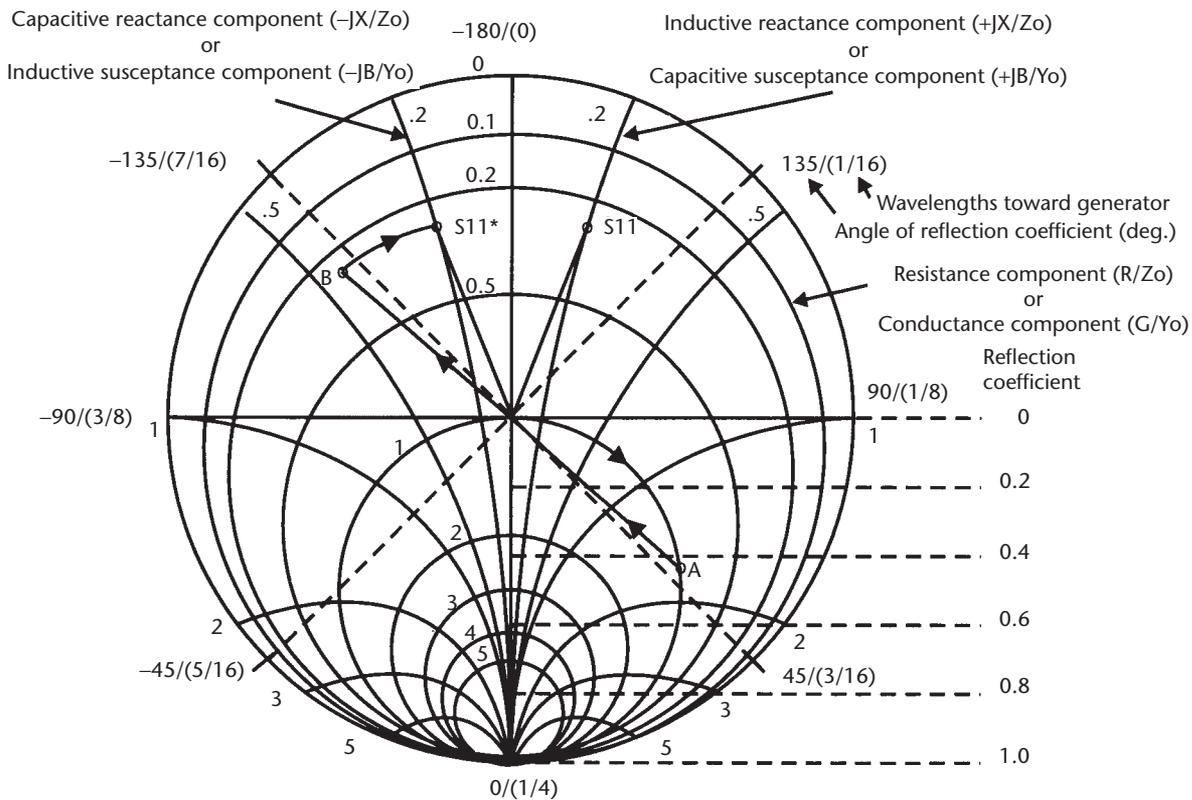


Figure 18.25 Example design of impedance matching circuits for narrowband transistor amplifier using S-parameters and Smith chart.

Table 18.1 DU1260T N-Channel RF Power MOSFET

<i>Features</i>				
DMOS structures				
Lower capacitances for broadband operation				
High saturated output power				
Lower noise figure than bipolar devices				
Specifically designed for 12-V applications				
<i>Absolute Maximum Ratings</i>				
<i>Parameter</i>	<i>Symbol</i>	<i>Rating</i>	<i>Units</i>	
Drain-source voltage	V_{ds}	40	V	
Gate-source voltage	V_{gs}	20	V	
Drain-source current	I_{ds}	24	A	
Total device dissipation at 25°C	P_d	250	W	
Junction temperature	T_{+j}	200	°C	
Storage temperature range	T_{stg}	-55 to +150	°C	
Thermal resistance		0.7	°C/W	
<i>Electrical Characteristics at 25°C</i>				
<i>Parameter</i>	<i>Symbol</i>	<i>Min</i>	<i>Max</i>	<i>Units</i>
Drain-source breakdown voltage	B_{dss}	40		V
(Test condition $V_{gs} = 0V, I_{ds} = 30.0 \text{ mA}$)				
Drain-source leakage current	I_{dss}		6	mA
(Test condition $V_{ds} = 15.0V, V_{gs} = 0V$)				
Gate-source leakage current	I_{gss}		6	μA
(Test condition $V_{gs} = 20.0V, I_{ds} = 0 \text{ mA}$)				
Threshold voltage	$V_{gs}(\text{th})$	2	6	V
(Test condition $V_{ds} = 10.0V, I_d = 600.0 \text{ mA}$)				
Forward transconductance	G_m	3		
(Test condition $V_{ds} = 10.0V, I_d = 6,000.0 \text{ mA}$)				
Input capacitance	C_{iss}		200	pF
(Test condition $V_{ds} = 12.0V, f = 1 \text{ MHz}$)				
Output capacitance	C_{oss}		240	pF
(Test condition $V_{ds} = 12.0V, f = 1 \text{ MHz}$)				
Reverse capacitance	C_{rss}		48	pF
(Test condition $V_{ds} = 12.0V, f = 1 \text{ MHz}$)				
Power gain	P_g	8		dB
(Test condition $V_{ds} = 12.0V, I_{dq} = 600 \text{ mA}, f = 175.0 \text{ MHz}, P_o = 60.0\text{W}$)				
Drain efficiency			60	%
(Test condition $V_{ds} = 12.0V, I_{dq} = 600 \text{ mA}, f = 175.0 \text{ MHz}, P_o = 60.0\text{W}$)				
Load mismatch tolerance	VSWR 30:1			
(Test condition $V_{ds} = 12.0V, I_{dq} = 600 \text{ mA}, f = 175.0 \text{ MHz}, P_o = 60.0\text{W}$)				
<i>Typical Device Impedance</i>				
	<i>Frequency (MHz)</i>	$Z_m(\Omega)$	$Z_{ol}(\Omega)$	
	30	4.5-j8.0	4.6-j7.9	
	100	1.4-j4.0	1.4-j8.0	
	175	1.0-j0.5	1.0-j0.5	

$V_{ds} = 12.0V, I_{dq} = 600 \text{ mA}, P_o = 60.0\text{W}$

maximum ratings for the various characteristics. In other cases, they provide only minimum and maximum values or ratings. Curves often are supplied for the various parameters as a function of frequency. In some cases, schematic diagrams of test fixtures used in measuring the various parameters are provided. Table 18.1 is an example of a typical set of manufacturers' data sheets for an RF power MOSFET. The manufacturer is M/A-COM Power Hybrids Operation, 1742 Crenshaw Blvd., Torrance, CA 90501.

Z_{in} is the series equivalent input impedance of the device from gate to source.

Z_{ol} is the optimum series equivalent load impedance as measured from drain to ground.

Drawings and Performance Curves Presented

1. Typical drain efficiency vs. frequency
2. Typical gain vs. frequency
3. Typical power output vs. power input
4. Typical power output vs. supply voltage
5. Package outline
6. Test fixture

References

- [1] Dye, N., and H. Granberg, *Radio Frequency Transistors: Principles and Practical Applications*, Boston, MA: Butterworth-Heinemann, 1993.
- [2] Van Valkenburg, M. E., (ed.), *Reference Data for Engineers*, 8th ed., SAMS, Carmel, IN: Prentice Hall Computer Publishing, 1993.
- [3] Van Valkenburg, M. E., (ed.), *Reference Data for Engineers*, 8th ed., SAMS, Carmel, IN: Prentice Hall Computer Publishing, 1993, pp. 18-13-18-21.
- [4] Kennedy, G., *Electronic Communication Systems*, 3rd ed., New York: McGraw-Hill, 1985.
- [5] Golio, J. M., *Microwave MESFETs & HEMTs*, Norwood, MA: Artech House, 1991.
- [6] Ali, F., and A. Gupta, (eds.), *HEMTs and HBTs: Devices, Fabrication, and Circuits*, Norwood, MA: Artech House, 1991.
- [7] Dye, N., and H. Granberg, *Radio Frequency Transistors: Principles and Practical Applications*, Boston, MA: Butterworth-Heinemann, 1993, Chap. 13.

High-Power Vacuum Tube Amplifiers and Oscillators

Vacuum tubes are used for nearly all high-power transmitter systems including power grid tubes, klystrons, helix-type TWTs, CCTWTs, crossed-field amplifiers (CFAs), magnetrons, gyrotrons oscillators, and gyrotron amplifiers.

19.1 Grid Tubes

Much of the information in this section is adapted from [1].

19.1.1 Triode, Tetrode, and Pentode Vacuum Tubes

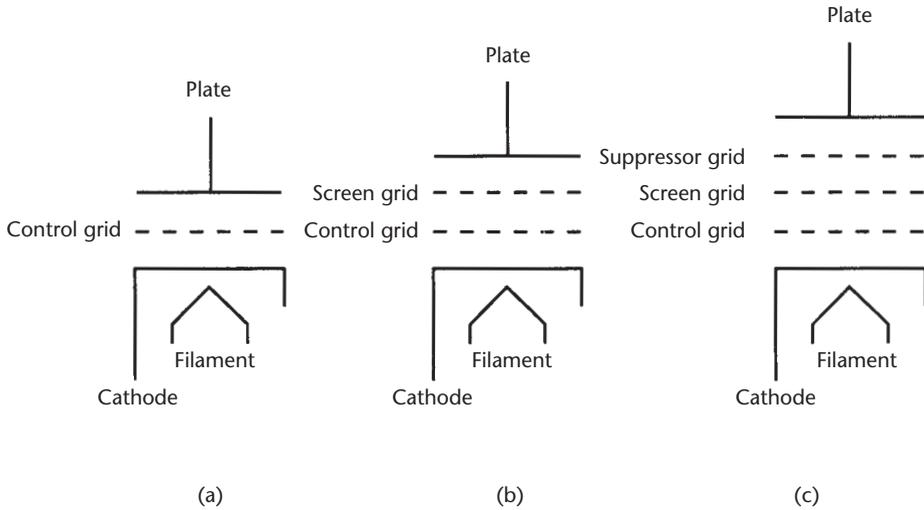
Grid-type vacuum tubes find extensive use as high-power amplifiers and modulators for frequencies below about 4 GHz. Their main application is at VHF and lower frequencies. Types of power grid tubes used include triodes, tetrodes, and pentodes. The triode is shown in Figure 19.1(a). Elements of the triode include the thermionic cathode, which emits electrons; the control grid, which controls the flow of electrons; and the anode, which collects most of the electrons. If the grid of a triode is biased to a sufficiently negative voltage, no current flows. As the grid is made less negative, more current flows to the anode. When the grid becomes positive with respect to the cathode, both the grid and the anode draw current.

In operation as an RF power amplifier, the triode must be either neutralized or operated in the common-grid mode. If that is not done, internal capacitance between the anode and the grid produces positive feedback that may cause oscillation.

Very large triodes may be used at the lower frequencies in high-power applications. As frequency is increased, however, the triodes must become smaller. This limitation is the result of reduced available transit time as frequency is increased. The transit time, the time it takes an electron to move from the cathode to the control grid, must be much less than one-half the period of the RF cycle. With short available transit time, the space between the cathode and the control grid must be made small, and the power handling capability of all elements thus is reduced.

Small, planar triodes are used at UHF and microwave frequencies up to about 4 GHz. The anode supply voltage typically is 1,000V or more.

Small cylindrical triodes are used mainly at VHF and UHF where CW power of a few hundred watts or pulse power of tens of kilowatts is required. Modern triodes



Note: Filament is a heater

Figure 19.1 Power grid tubes: (a) triode; (b) tetrode; and (c) pentode.

use beam-forming cathodes and control-grid geometry to allow simplicity of design and the circuit advantages of a triode with the gain of a tetrode.

The tetrode is illustrated in Figure 19.1(b). It has the same three elements as the triode plus a screen grid between the control grid and the anode. The presence of the screen grid greatly reduces the capacitance between the anode and the control grid and makes neutralization completely unnecessary or at least easy to accomplish.

Tetrode tubes are available in various sizes. Applications are in radio broadcasting (AM and FM), television broadcasting (VHF and UHF), communications, radar, and navigational aids, among others. Uses are in RF power generation, modulation (AM), and switching for pulse service.

The pentode is shown in Figure 19.1(c). It has the same elements as the tetrode plus an additional suppressor grid to control secondary electrons. Modern tetrodes accomplish this control in other ways and have almost completely displaced pentodes.

There is a limit to the power capability of the tubes that is determined by power dissipation of the anode and the grids, which depends on the type of cooling that is provided. The lower-power triodes usually use fins and air cooling. In some cases, this will be forced air cooling. The higher power tubes usually use water cooling.

Some important terms used with grid-type vacuum tubes follow:

μ = amplification factor with current held constant ($\Delta E_b / \Delta E_{c1}$)

R_p = dynamic plate resistance ($\Delta E_b / \Delta I_b$)

G_m = transconductance ($\Delta I_b / \Delta E_{c1}$)

E_b = total instantaneous plate voltage

E_{c1} = total instantaneous control grid voltage

I_b = total instantaneous plate current

The value of μ may be in the range of 20–200, depending on the design of the tube. In a triode, plate current increases with increasing plate voltage. In a tetrode and a pentode, the plate current is nearly independent of plate voltage. Figure 19.2 shows the characteristics of a typical tetrode plate current. Plate current is plotted as a function of plate voltage and control-grid voltage for a fixed screen voltage of 300V.

Figure 19.3 is a drawing of the radial beam power tetrode, type 4CX15000A. The figure shows the dimensions of the tube and identifies the location of the elements. This tube is typical of high-power air-cooled tubes.

19.1.2 Grid-Type Vacuum Tube Amplifiers and Modulated Amplifiers

Figure 19.4 is a schematic diagram of a tetrode RF power amplifier. This circuit is an example of a low-VHF, HF, MF, or lower frequency system that uses lumped constant inductors and capacitors. Higher frequency VHF or UHF amplifier circuits use cavity resonators rather than LC circuits.

The filament or heater circuit and the cathode circuit for the amplifier in Figure 19.4 are shown at the lower part of the diagram. The cathode is ac coupled to ground. It is also dc coupled to ground through the center tap of the filament transformer. The grid bias is injected through an RFC, as is the plate bias current. Bypass capacitors also are incorporated to decouple the RF signals from the power supplies. The RF input to the grid circuit is supplied by a tuned transformer with series tuning on the primary and parallel tuning on the secondary. The RF output is coupled from the plate using a series capacitor followed by an impedance-matching circuit. The circuit also shows the use of a neutralization capacitor. With neutralization, feedback is provided from the plate circuit to the grid circuit in such a way that other feedback signals are canceled. That is done to prevent unwanted oscillations.

Most of the unwanted oscillations in RF power amplifiers using grid-type tubes fall into the following three categories:

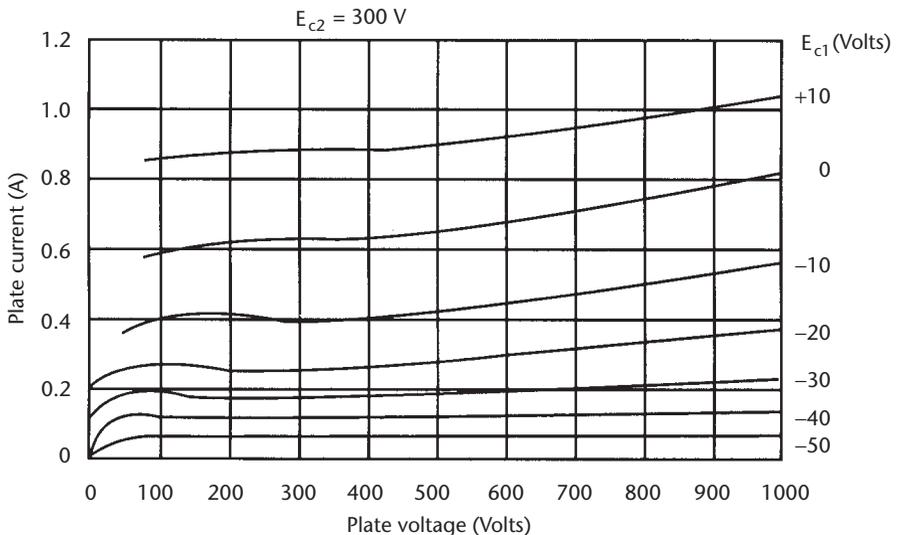


Figure 19.2 Characteristics of a typical tetrode plate current. (After: [1].)

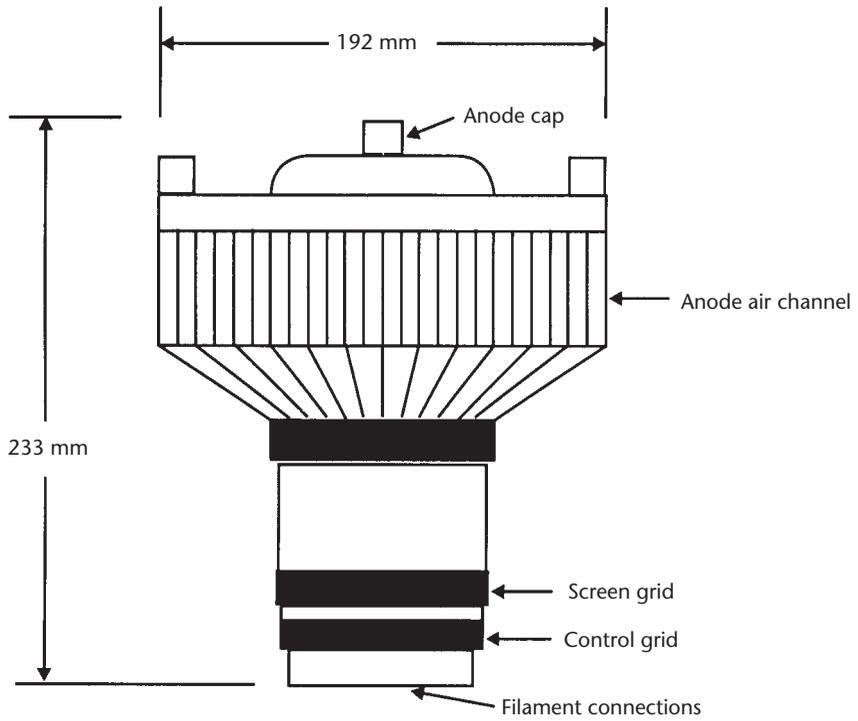


Figure 19.3 Drawing of the type-4CH15000A radial beam power tetrode. (After: [1].)

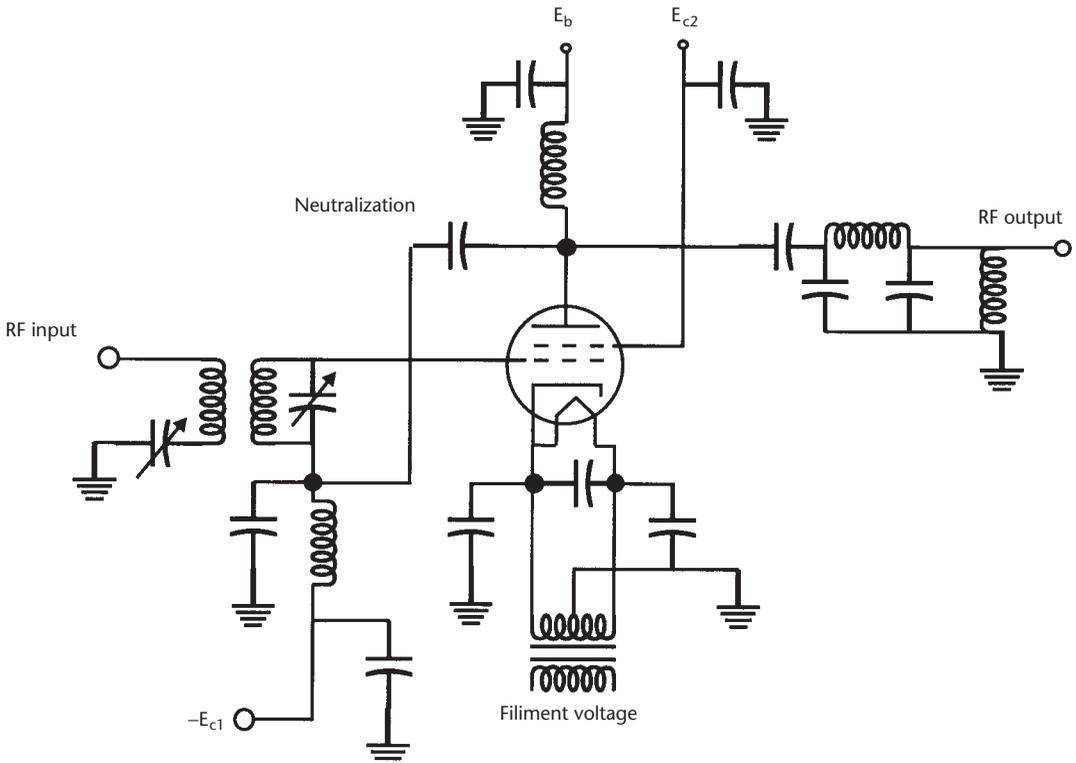


Figure 19.4 Schematic diagram of a tetrode RF power amplifier. (After: [1].)

- Oscillations at VHF from about 40 to 200 MHz, regardless of the operating frequency of the amplifier;
- Self-oscillation on the fundamental frequency of the amplifier;
- Oscillation at a low RF below the normal frequency of the amplifier.

LF oscillations usually involve the RFCs. These may be eliminated by changes in the positions or orientations of the chokes. Oscillations near the fundamental frequency involve the normal resonant circuits. These are eliminated by effective design and the use of the neutralization circuit. Oscillations at VHF usually are referred to as parasitic oscillations. These may not be eliminated by the simple neutralization circuit. Solutions to this problem may require the addition of an RFC or an RFC plus a resistor in parallel that are located between the plate and the bias choke or output circuit.

Figure 19.5 shows one version of the plate-modulated power amplifier that is frequently used for AM applications. In this instance, triodes are used rather than tetrodes. Either type of tube may be used.

The RF signal is fed to the grid of the upper triode using a single-tuned transformer. The tube is operated as a class C amplifier. It has its cathode connected to ground and its grid biased negatively. The output short pulses of current at the RF frequency are fed to a single-tuned RF transformer, which, by the fly-wheel effect, converts the pulse-type signal to a sine wave signal.

The audio or other modulating signal is fed to the grid of the lower triode amplifier. That amplifier modulates the voltage applied to the RF amplifier, thereby amplitude modulating the RF signal.

A more common plate-modulated amplifier is one that uses a modulation transformer in the plate voltage supply and a class B amplifier to drive the transformer.

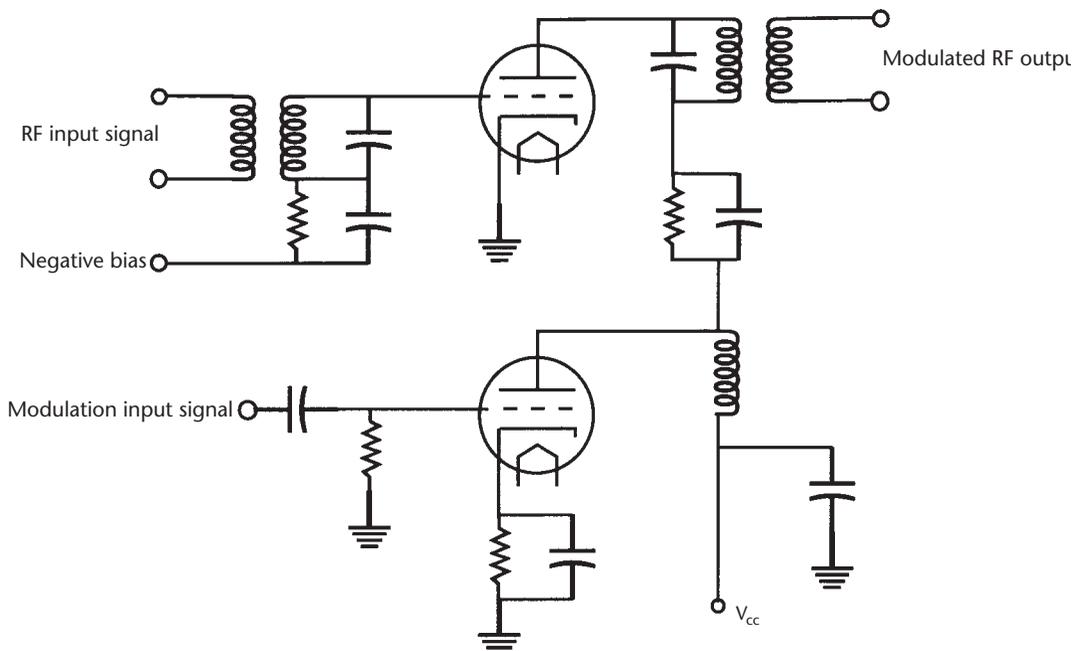


Figure 19.5 Classic plate-modulated class C power amplifier circuit for AM. (After: [1].)

All the sideband power must be supplied by the audio or other LF amplifier that drives the transformer. Thus, the amplifier is a large, high-power system for high-power modulators.

Figure 19.6 shows grid modulation of a class C RF amplifier. In this case, the modulating signal is applied to a transformer that is in series with the input RF carrier signal transformer, and the combination is applied to the grid of the triode power amplifier. Again, a single-tuned transformer is used at the output of the triode amplifier to convert the pulsed signals to modulated sine waves. This type of circuit avoids the problems of supplying high modulating power to the plate circuit.

19.1.3 The Use of Cavities as Resonators for High-Power, High-Frequency Triodes and Tetrodes

Cavities frequently are used as resonators for high-power triode or tetrode amplifiers that operate at VHF and higher frequencies. The type of cavities most often used is coaxial cavities. These may be either quarter-wavelength cavities or half-wavelength cavities.

Figure 19.7 shows the case of the quarter-wave cavity with a tetrode amplifier. An RFC connected to a coupling line brings dc plate bias into the tube. The RF input is to the base terminals. The RF output is coupled to a coax line using a small coupling loop. Tuning is accomplished by means of an adjustable shorting deck. Additional tuning is performed by means of tuning capacitors, which may include a movable tuning-capacitor plate.

In a quarter-wavelength cavity, the short circuit at the end of the cavity is reflected as an open circuit. Thus, it behaves like a parallel resonant circuit at the position of the plate.

The equivalent circuit for the quarter-wavelength cavity amplifier with capacitive coupling is shown in Figure 19.8. The cavity acts as a high-Q parallel resonant circuit.

Half-wavelength power cavities also are used for tube amplifiers. In this case, the transmission line cavity is open circuited at the far end rather than being

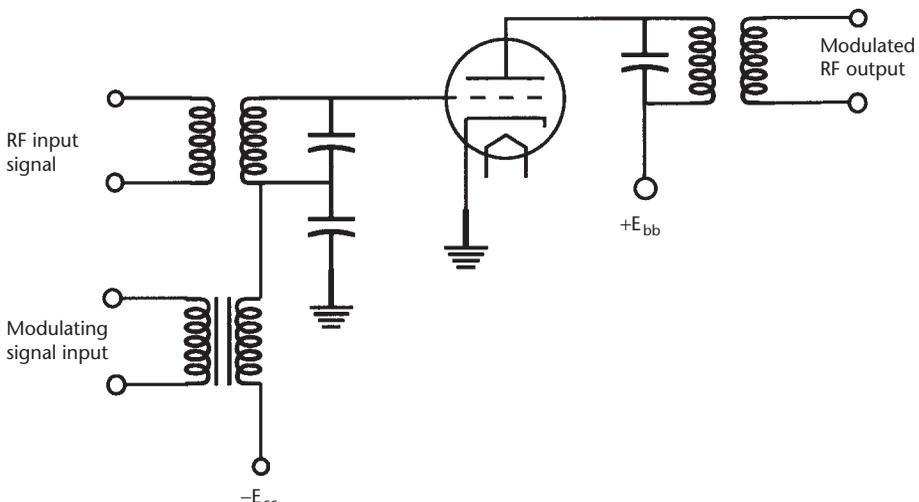


Figure 19.6 Grid-modulated class C amplifier for AM. (After: [1].)

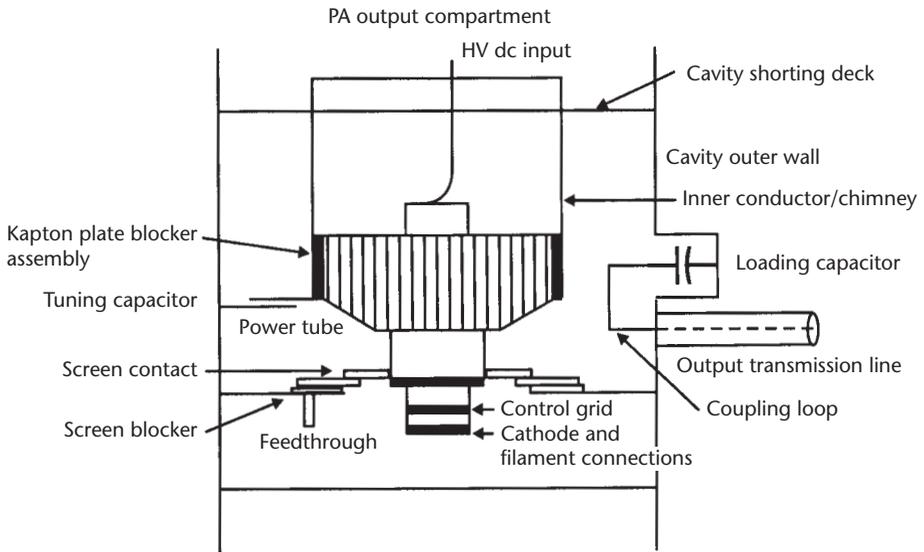


Figure 19.7 Example VHF quarter-wavelength cavity for high-power tetrodes. (After: [1].)

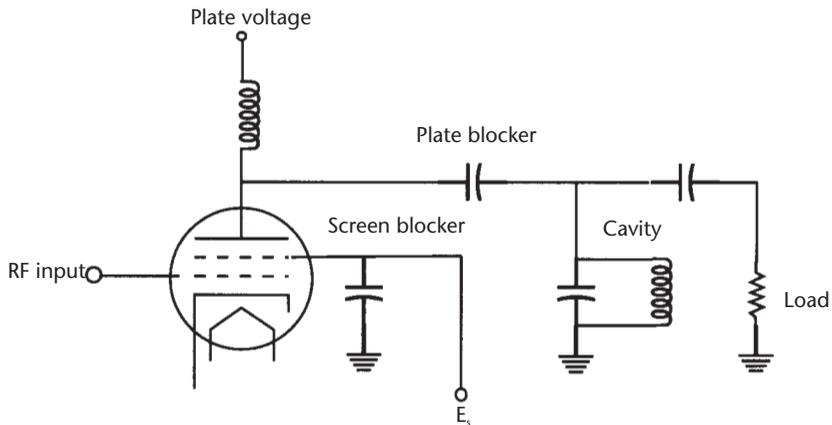


Figure 19.8 Equivalent circuit for a quarter-wavelength cavity amplifier with capacitive coupling. (After: [1].)

short-circuited. The open circuit is reflected as an open circuit at the location of the plate, again acting as a parallel resonant circuit. Needless to say, this type of cavity is large at the lower frequencies. For example, at 88 MHz the length of the cavity above the plate would be 67 inches.

19.2 Microwave Tubes and Circuits

19.2.1 Introduction to Microwave Tubes

Grid-type tubes such as triodes and tetrodes cannot be used at high microwave frequencies because of the problem of transit-time limitations. Transit time is the time

it takes for an electron to travel from the cathode to the control grid and needs to be much less than the time for a half-cycle of the RF signal. Microwave tubes, on the other hand, do not have this problem. As a matter of fact, some microwave tubes, such as the klystron and the TWT, use transit-time effects to an advantage. Instead of using electron-density modulation, as in the case of triodes and tetrodes, they use velocity modulation of the electron beam as a means of producing electron bunching.

Table 19.1, which is adapted from information in [2], is a comparison of the operating characteristics for the four main types of high-power microwave tubes: the klystron, the CCTWT, the magnetron, and the CFA. Other important types of microwave tubes are not shown in the table: lower power, helix-type TWTs; low-power, two-cavity klystron oscillators; reflex klystron oscillators; high-power gyrotron oscillators; extended interaction devices; and fast-wave devices.

The maximum average or CW power capability of microwave tubes is shown in Figure 19.9 as a function of frequency. Also included on the plot are the maximum CW power capabilities for solid-state devices and for triodes and tetrodes.

Table 19.1 Comparison of Operating Characteristic for Medium- to High-Power Microwave Tubes

<i>Linear Beam Tubes (LBT)</i>		
<i>Characteristic</i>	<i>Klystron</i>	<i>CCTWT</i>
Application	Amplifier	Amplifier
Frequency	UHF to K_a -band	UHF to K_s -band
Maximum peak power	5 MW at L-band	240 kW at UHF
Maximum average power	1 MW at L-band > 10 kW at X-band	12 kW at L-band 10 kW at X-band
Cathode voltage for peak power	Up to 125 kV (5 MW, L-band)	Up to 42 kV (0.2 MW, L-band)
Percentage bandwidth	1–10%	5–15% for CCTWT 100% for helix TWT
Gain (dB)	30–65	30–65
Efficiency	Up to 65%	Up to 60%
<i>Crossed-Field Tubes (CFTs)</i>		
	<i>Magnetron</i>	<i>CFA</i>
Application	Oscillator	Amplifier
Frequency	UHF to K_a -band	UHF to K_a -band
Maximum peak power	1 MW at L-band	5 MW at S-band
Maximum average power	1.2 kW at L-band 100W at X-band	1.3 kW at L-band 2 kW at X-band
Cathode voltage for peak power	Up to 60 kV (1 MW, L-band)	Up to 105 kV (5 MW, L-band)
Percentage bandwidth	1–15%	5–15%
Gain (dB)		10–20
Efficiency	Up to 70%	Up to 80%

Source: [2].

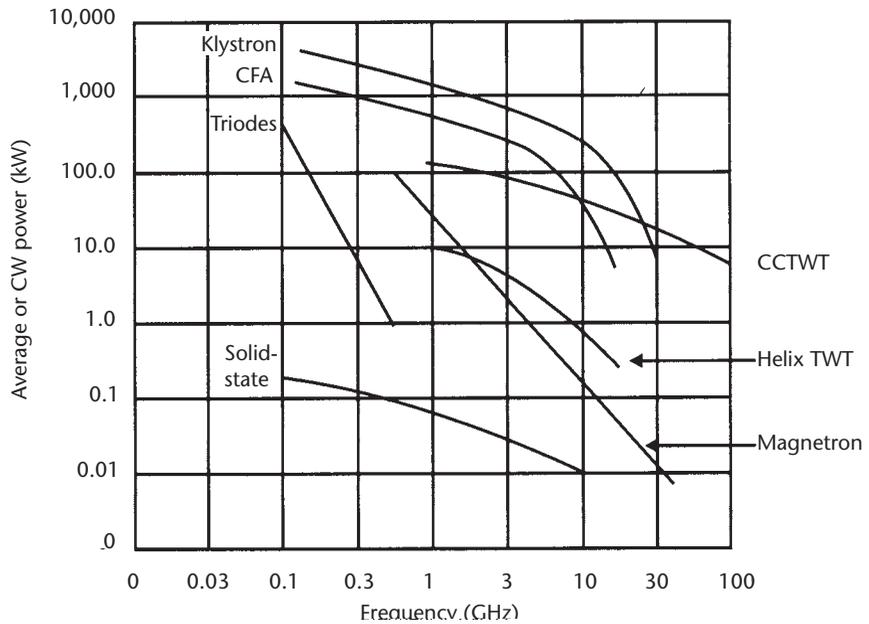


Figure 19.9 Average power for microwave power tubes. (After: [3].)

19.2.2 Multiple-Cavity Klystron Amplifiers

A three-cavity klystron is shown in Figure 19.10. Elements of this tube include an electron gun, which forms a thin pencil-like electron beam; an input resonant cavity with openings to the beam; a similar intermediate resonant cavity tuned to a slightly higher frequency; an output resonant cavity tuned to the same frequency as the input cavity; and a collector for capturing the electrons. Electromagnets are used to keep the beam from spreading. Simple klystron amplifiers may use only two cavities (an input cavity and an output cavity). Very high gain cavities use more than three cavities.

The RF input signal field in the input cavity velocity modulates the electron beam. That causes the beam to become density modulated as the electrons move toward the collector. The intermediate cavity helps to further velocity modulate the beam. As the bunched electrons pass through the output cavity, they couple energy into the cavity. Suitable couplings are used at both the input cavity and the output cavity, so the cavities can be connected to the input and output transmission lines.

Klystrons are used for UHF television transmitters. Varian Associates produces tubes of this kind with output powers up to 65-kW CW. Examples of characteristics of this type of UHF-TV klystron amplifier are listed in Table 19.2. These characteristics are for the Varian VKP-7553S klystron, which is only 1 of 22 tubes of this kind listed by Varian in their sales brochure.

Many modern radar systems need combinations of high peak and average power, high gain, broad bandwidth, and long pulse capability in conjunction with linear phase characteristics and stable performance. These requirements can be satisfied by high power pulsed klystron and twystron amplifiers. Twystrons are a combination of a klystron and a TWT made by Varian. Somewhat larger bandwidths can be obtained with twystrons than with klystrons.

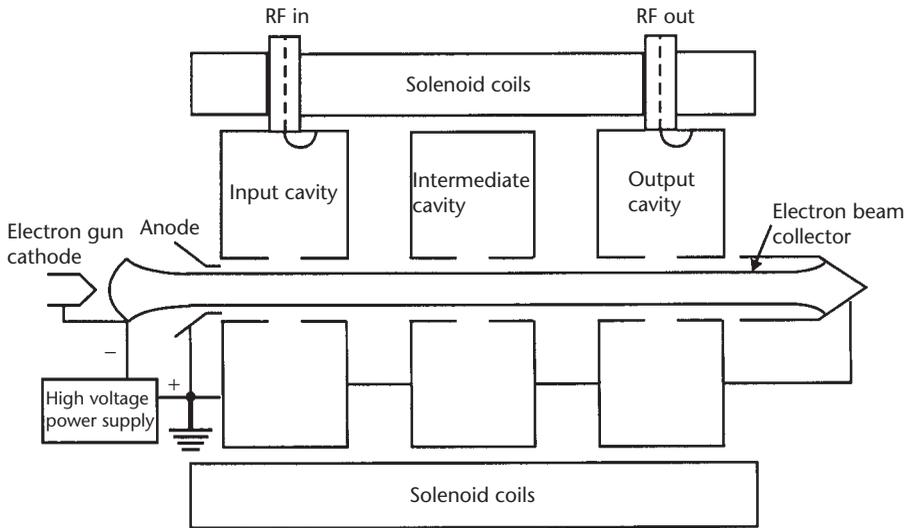


Figure 19.10 Three-cavity klystron amplifier. (After: [4].)

Table 19.2 Characteristics of Varian Associates VKP-7553S Klystron

Characteristics	Values
Operating frequency (MHz)	470–566
Output power (kW)	60
Gain (dB)	31
Tuning range (MHz)	96
1-dB bandwidth (MHz)	8
Number of cavities	5
Beam voltage (kV dc)	24.5
Beam current (A dc)	5.2
Heater voltage (V)	7.0
Heater current (A)	17.0
Weight (pounds)	930 (includes both tube and magnet)
Cooling	Vapor/liquid/forced-air cooled

Table 19.3 lists characteristics of three types of high-power pulse klystrons made by Varian: types VA-963A and VKL-7796 high power; pulsed klystron amplifiers; and type VA-145E high-power, pulsed twystron. Varian also produces lower power klystrons for radars in the frequency range of 0.43–36 GHz. Typical lower power ranges are 50–100 kW.

Note that the VA-145E twystron has a 1.5-dB bandwidth of 200 MHz, whereas the klystron amplifiers have 1-dB bandwidths of only 15 and 70 MHz. The VA-145E twystron is used for applications involving chirp modulation, where FM is used in the form of linear sweep over the duration of a fairly long pulse. That permits the generation of high average power needed for long-range capability and, at the same time, short pulses needed for good range resolution. An example of an

Table 19.3 Characteristics for Varian Types VA-963A and VKL-7796 High-Power, Pulsed Klystron Amplifiers and VA-145E High-Power, Pulsed Twystron

<i>Characteristics</i>	<i>VA-936A</i>	<i>VKL-7796</i>	<i>VA-145E</i>
Frequency (GHz)	1.25–1.35	1.29–1.36	2.9–3.1
Peak output power (MW)	5 MW	4 MW	2.5 MW
Typical gain (dB)	50	45	37
Typical efficiency (%)	40	43	35
1-dB bandwidth (MHz)	15	70	200
Duty cycle	0.002	0.075	0.002
Pulse length (ms)	3	130	7
Tunable in system	Yes	No	No
Beam voltage (kV dc)	130	112	117
Beam current (A dc)	101	85	80
Tube weight (pounds)	150	600	140
Tube length (inches)	60	95	43
Cooling	Liquid	Liquid	Liquid
Tube type	Klystron	Klystron	Twystron

application is the long-range tracking radar at the Pacific Missile Test Center in Vandenberg, California.

19.2.3 Helix-Type Traveling-Wave Tube Amplifiers

The basic helix TWT is illustrated in Figure 19.11. This tube uses an electron gun to form a thin pencil-type beam. Focusing magnets are used to keep the beam from spreading. The electrons in the beam are collected at the far end of the tube by a beam collector. A helix that is capable of propagating a slow wave is placed around the electron beam. The RF wave on this helix travels at about one-sixth the speed of light. The velocity of the electron stream is adjusted to be approximately the same as the axial phase velocity of the wave on the helix so there can be interaction between a signal on the slow-wave structure and the electrons in the beam.

The input signal on the helix causes velocity modulation of the electron beam. That, in turn, causes the electron beam to become density modulated as the electrons move down the tube. The modulated electron beam induces waves on the helix. The mutual interaction continues along the length of the tube with the net result that dc energy is given up by the beam to the circuit as RF energy and the wave is thus amplified. It is necessary to provide a resistance midway in the helix to prevent backward-wave oscillations. The wave on the helix is attenuated by the resistive material that surrounds the helix. The signal is then reintroduced to the helix after the resistor by the density-modulated electron beam. The type of TWT that uses the helix slow-wave structure is characterized by wide bandwidth (about an octave). Gains of 20–50 dB are possible, depending on the TWT design. Some TWT amplifiers are designed to be low-noise, low-power front ends for microwave receivers. They have smaller gains than the higher power amplifiers.

Some helix-type TWTs use one or more grids to switch the electron beam on or off. In a radar, that is done to reduce the noise to the receiver when in the receive mode. Many TWTs use a more complex collector than that shown. By using a

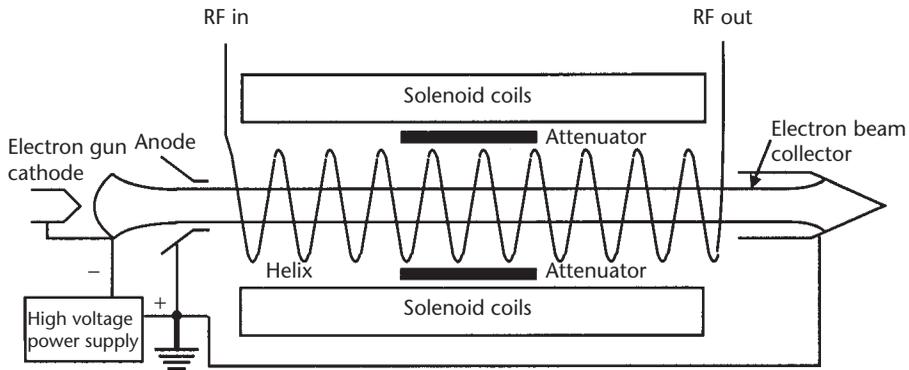


Figure 19.11 Basic helix type TWT. (After: [5].)

so-called depressed collector in which the voltage is more positive than the cathode but less than zero volts, it is possible to still collect the beam, increasing the overall efficiency of the tube.

TWTs are used extensively for microwave communication systems in ground, air, and satellite applications. The devices may be either low or high power. Table 19.4 lists some characteristics for the Varian VTU-6291C1 conduction-cooled, K_u -band TWT used for communications.

Table 19.5 lists some characteristics of a Varian VTF-5121A8 1-kW pulsed TWT that operates in the 2.5–2.8-GHz range.

Table 19.6 lists some Varian millimeter-wave TWTs. These tubes have output powers less than 100W. They are small and lightweight.

TWTs are used in satellite transmitters. This type of transmitter has the advantage of better efficiency than solid-state transmitters of the same power. Typical efficiency for the TWT transmitter is about 60%, compared with about 30% for the solid-state version. Much higher efficiency is important for a satellite system, where prime power must come from solar cells. Additional weight in orbit is costly.

Table 19.4 Characteristics of the Varian VTU-6291C1 Communication-Type TWT

<i>Characteristics</i>	<i>Values</i>
Frequency (GHz)	14–14.5
Minimum output power (W)	125
mGain (dB)	50
Helix voltage (kV)	6.9
Cathode current (mA)	0.15
Helix current with RF (mA)	1.5
Collector voltage (kV)	3.2
Cooling	Conduction
Input connector	SMA coax
Output connector	WR-75 waveguide
Dimensions, L × W × H (inches)	14.5 × 3 × 2.8
Weight (pounds)	5

Table 19.5 Characteristics of the Type VTF-5121A8 Pulsed TWT

<i>Characteristics</i>	<i>Values</i>
Frequency (GHz)	2.5–2.8
Minimum output power (kW)	1.0
Maximum drive power (dBm)	25
Grid pulse (V)	+160
Grid bias (V)	–200
Beam voltage (kV)	9
Collector voltage (kV)	6.4
Grid current (mA)	10
Beam current (Amps)	1.4
Beam duty cycle (%)	4
Maximum pulse width (ms)	125
Modulating element	Unigrid
Dimensions, L × W × H	21.5 × 2.5 × 3
Weight (pounds)	7.5
Input connector	SMA coax
Output connector	TNC coax
Cooling	Conduction

Table 19.6 Characteristics of Several Varian Millimeter-Wave TWTs

<i>Characteristics</i>	<i>VTK-6193D1</i>	<i>VTA-6193M2</i>	<i>VTE-6193G1</i>
Frequency (GHz)	18–26.5	27.5–30.0	50–60
Minimum output power (W)	20	80	20
Gain (dB)	47	49	40
Helix voltage (kV)	12.5	12.5	14.0
Helix current (mA)	2.0	0.2	0.2
Collector voltage (kV)	2	5.6	6.3
Beam current (mA)	95	90	50
Modulating element	Focus electrode	Focus electrode	Focus electrode
Electrode on/off voltage (V)	–5/–700	–5/–700	–5/–700
Anode voltage (V)	None	+500	+500
Input RF connector	SMA	UG-599/U	UG-383U Mod
Output RF connector	SMA coax	UG-599/U	UG-383U Mod
Dimensions, L × W × H (inches)	12.3 × 0.8 × 1.6	16 × 2.8 × 3.4	16 × 2.8 × 3.4
Weight (pounds)	4	7	7
Cooling	Conduction	Conduction	Conduction

The main type of microwave tube used for EW transmitters is the wideband helix-type TWT amplifier. Sophisticated ECM systems require ultrawide bandwidths, high efficiency, multistage depressed collectors, low voltage, and grid-controlled electron guns. TWTs using these features range from MINI-TWT amplifiers to high-power, wide-bandwidth TWTs such as the Varian VTM-6292 family, which is well suited for active phased-array applications, driver amplifiers,

and final output stages. An example of an ECM pod is the AN/ALQ-131 ECM pod, which contains a high-power ECM TWT amplifier. Table 19.7 lists some characteristics of the Varian VTM-6292F4 CW EW TWT.

19.2.4 Coupled-Cavity Traveling-Wave Tube Amplifiers

Helix-type TWTs are limited in power capability by the heat-dissipation capability of the helix. For higher power output capability, it is necessary to use a different type of slow-wave structure. The best one to use is the coupled-cavity slow-wave structure. A TWT of this type is shown in Figure 19.12.

CCTWTs are linear amplifiers. They can be used over the frequency range from UHF to K_a -band. The maximum output peak power at UHF is about 240 kW. The maximum average power output is about 12 kW at L-band and 10 kW at X-band. The amplifier gain is in the range of 30–65 dB. The percentage bandwidth is in the range of 5–15%. That compares with 1–10% for the klystron. The efficiency can be as high as 60%. For the higher frequencies, it tends to be much lower. The cathode voltage for peak power can reach 42 kV.

19.2.5 Conventional Magnetrons

Figure 19.13 shows an important type of conventional magnetron oscillator. It consists of a cathode and an anode with openings to cavities. This tube is a diode, and it operates as an oscillator. It is mounted between the pole pieces of a magnet. The interaction between the magnetic field and the movement of electrons causes the electrons to move in a circular path. The path becomes modified by interaction with the RF fields associated with the cavities. The result is that electron spokes are formed that rotate. Energy is coupled out of one of the cavities to a waveguide or a coax.

Openings are provided from the cavities to the space between the anode and the cathode. An output is taken from one of the cavities. In this example, the output is coupled using a loop coupler and coax line.

Table 19.7 Characteristics of Varian VTM-6292F4 CW

<i>Characteristics</i>	<i>Values</i>
Frequency (GHz)	6.5–18
Minimum output power (W)	200
Minimum gain (dB)	30
Helix voltage (kV)	10.3
Cathode current (mA)	280
Helix current with RF (mA)	9
Collector voltage (kV)	5.1/3.0
Electron gun configuration	Focus electrode
Voltage for beam on/off (V)	–10/–1,200
Input/output connectors	WRD650
Dimensions, L × W × H (inches)	17 × 3 × 3.5
Weight (pounds)	9

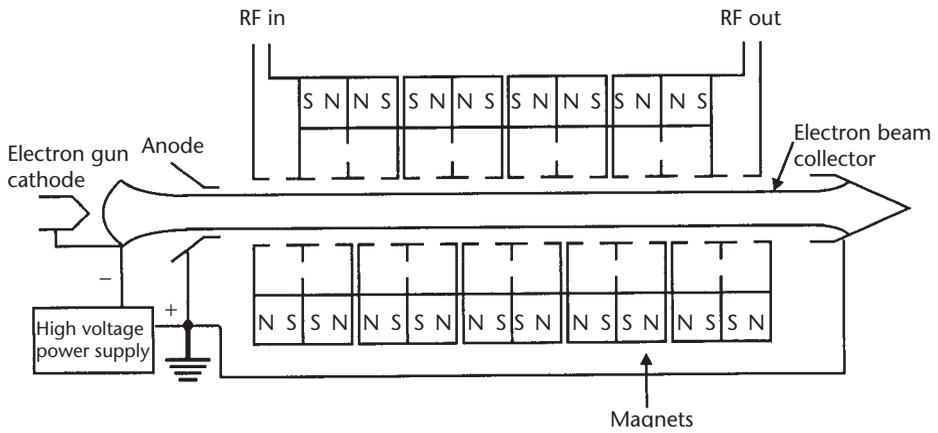


Figure 19.12 CCTWT. (After: [5].)

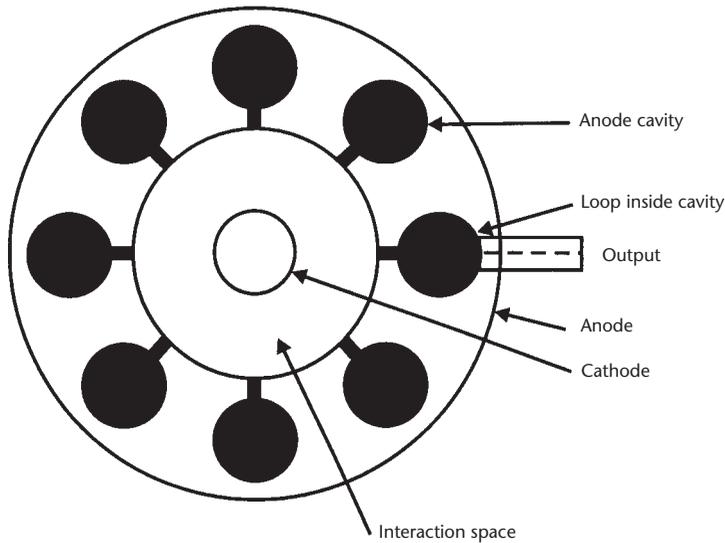


Figure 19.13 Conventional magnetron structure using eight cylindrical cavities in the anode. (After: [6, 7].)

Magnetrons of this type having eight or more coupled cavities usually are strapped in such a way that there are two sets of conductors connected between every other anode section between the cavities. That is done to avoid mode jumping.

Figure 19.14 shows the case of the vane-type magnetron anode structure, which is used with very high frequency magnetrons. The configuration has the advantage that it does not require mode strapping.

19.2.6 Coaxial Cavity Magnetron

A different type of magnetron, a coaxial magnetron, is shown in Figure 19.15. The magnetic field is perpendicular to the page. It is seen that there is an integral coaxial

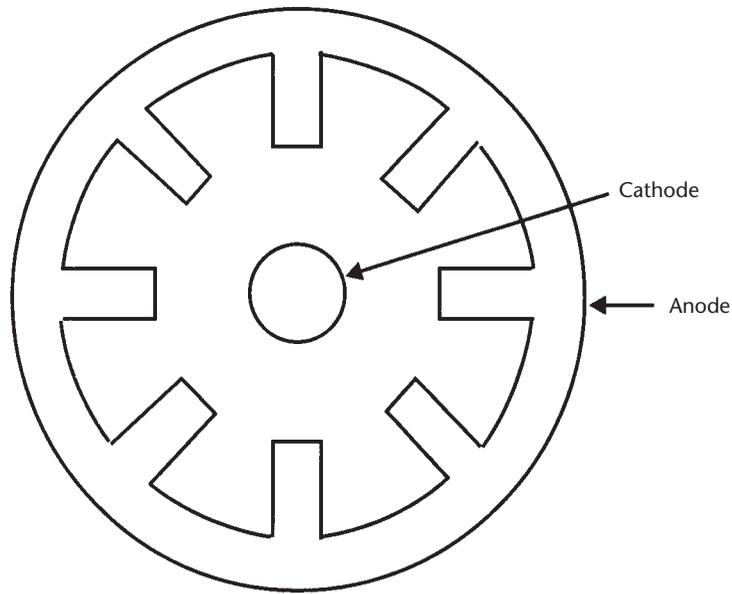


Figure 19.14 Vane-type magnetron anode configuration. (After: [6, 7].)

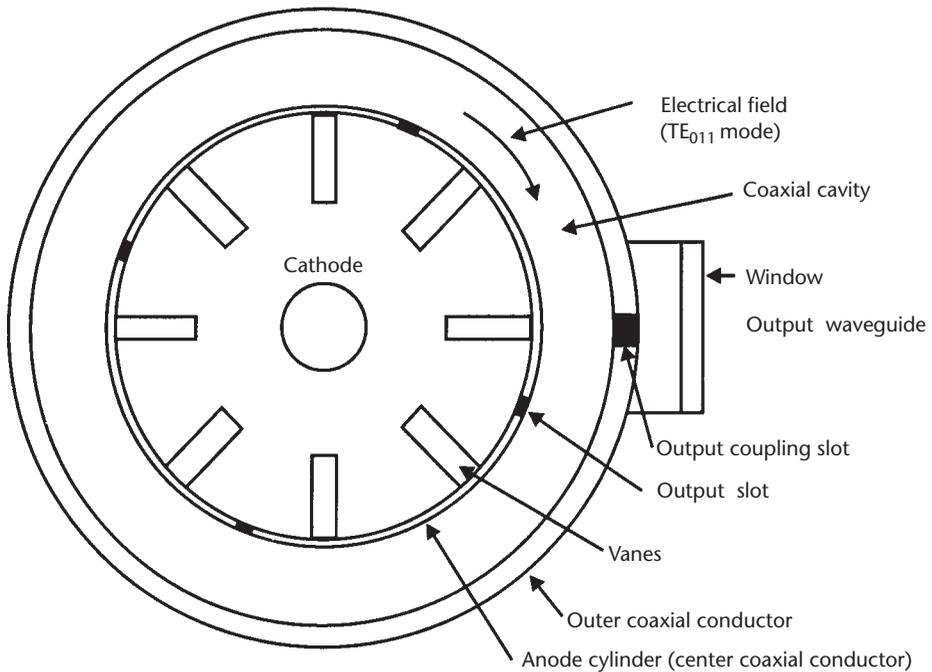


Figure 19.15 Cross-section of a coaxial cavity magnetron. (After: [6, 7].)

cavity present in this magnetron. The Q of the cavity is much higher than the Q s of the various resonators, so it is the coaxial cavity that determines the operating frequency. The output from the cavity is through a waveguide window to a rectangular waveguide operating in the dominant TE_{10} mode.

The performance of the coaxial magnetron is better than that of earlier types of magnetrons. The enlarged anode area, compared with conventional magnetrons, permits better dissipation of heat and consequently smaller size for a given output power. Mean time between failure (MTBF) is considerably longer.

An example of a coaxial magnetron is the SFD-341 mechanically tuned C-band coaxial magnetron. This tube is used extensively for shipboard and ground-based radars. It delivers a peak power of 250 kW with a 0.001 duty cycle. It operates over the frequency range of 5.45–5.825 GHz. The efficiency of the tube is in the range of 40–45%. The operating life of coaxial magnetron tubes can be 5,000 to 10,000 hours. That is a five- to twentyfold improvement over conventional magnetrons.

Since most of the RF energy is stored in the TE_{011} mode cavity rather than in the resonator region, reliable broadband tuning of the magnetron can be accomplished by a noncontacting plunger in the cavity. Both the pushing figure (change in the frequency with a change in anode current) and the pulling figure (change in the frequency with a change in the phase of the load) are much less in a coaxial magnetron than in the conventional configuration.

Magnetron oscillators are low-cost, rugged devices. For those reasons, the magnetrons are used in home microwave ovens. This type of tube also is used for industrial heating and numerous other applications.

19.2.7 Amplitrons

Another crossed-field device of interest is shown in Figure 19.16: the CFA known as the amplitron. In its operation, an amplitron is a cross between a TWT and a magnetron. It uses a magnetic field normal to the surface of the page and a large cathode. It uses a slow-wave structure to provide a continuous interaction between the electron beam and a moving RF field. In practice, a vane-type slow-wave structure usually is used.

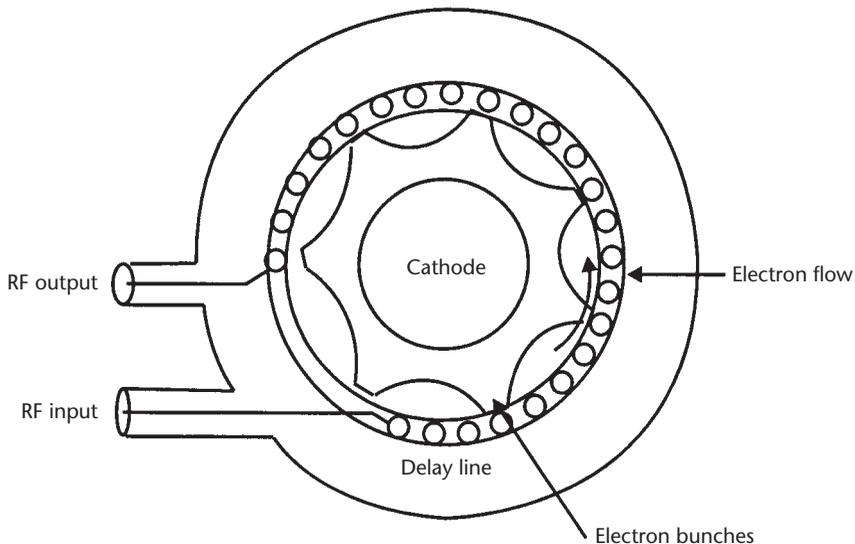


Figure 19.16 Operational concepts of the amplitron. (After: [8].)

CFAs also find application in radar systems. These amplifiers have only limited gain (10–20 dB), so they must be driven by a fairly high-power driver stage, which could be a klystron or a CCTWT. This microwave tube is very much like a magnetron in appearance and operation. The big difference, of course, is that the CFA is a saturated amplifier, whereas the magnetron is an oscillator.

The main advantages of the CFA are its very high power capability and a very high efficiency. The maximum peak power is about 5 MW at S-band. The maximum average power is 1.3 kW at L-band and 2 kW at X-band. The efficiency can be up to 80%. The cathode voltage is up to 15 kV.

The bandwidth is 5–15%. The narrower the bandwidth, the higher the gain.

19.2.8 Gyrotron Oscillators and Amplifiers

The gyrotron oscillator is a microwave vacuum tube that operates on the basis of the interaction between an electron beam and microwave fields where coupling is achieved by the cyclotron resonance condition. This type of coupling allows the beam and microwave circuit dimensions to be large compared to a wavelength. Hence, the gyrotron avoids the power-density problems encountered in conventional klystrons and TWTs at millimeter wavelengths.

The device includes a cathode or electron gun, a beam collector area, an interaction cavity, and a circular waveguide of gradually varying diameter. Electrons are emitted at the cathode with small variations in speed. The electrons are then accelerated by an electric field and guided by a static magnetic field through the device. The nonuniform induction field causes the rotational speed of the electrons to increase. The linear velocity of the electrons, as a result, decreases. The interaction of the microwave field within the waveguide and the rotating (helical) electrons cause bunching similar to the bunching in a klystron. A decompression zone at the end of the device permits decompression and collection of the electrons.

The circuit area in the gyrotron devices is 100 times the circuit area in a klystron or conventional TWT. That permits gyro-devices to have 100 times the output power capability of conventional tubes.

Another interesting feature of the gyrotron is that the applied magnetic field is proportional to frequency. Superconducting magnets are required for operation at 50 GHz and higher.

Initially, gyro-devices were single-cavity oscillators, in which the entire interaction occurred in a single microwave cavity. However, the same basic interaction can be used with different variations, such as amplifiers, using several resonant cavities, or traveling-wave circuits. These variations are called gyro-klystrons and gyro-TWTs, respectively.

Varian gyrotrons have produced over 300 kW at 28 GHz, 200 kW at 60 GHz, 850 kW at 140 GHz, and 12 kW at 250 GHz. A wide range of both oscillators and gyro-TWTs are under development by Varian and others at the time of this writing.

19.2.9 Circuit Configurations for Microwave Tubes

All the vacuum-tube circuits discussed in this chapter need high-voltage power supplies, controls, and protection circuits, which must be tailored to the devices

involved. These systems include sensors for measuring voltage, current, temperature, and standing-wave ratio. Means for automatic shut-off are provided in the event that a problem is detected. A number of different types of controls and displays are provided by these systems. X-ray shielding is an important safety feature required in very high voltage klystrons. Other required safety features are interlock systems that turn the systems off if a cabinet is opened while the systems are turned on.

Figure 19.17 shows two examples of high-power microwave transmitter systems that use TWTs. Figure 19.17(a) is a single-stage TWT transmitter and associated power supply and control system. The RF signal is fed into the TWT at a relatively low power level, since the TWT normally has very high gain. The output is then connected to the antenna circuit, which may consist of a duplexer circuit, monitoring circuits, and the antenna. The TWT in the figure is connected to a high-voltage power supply and a modulator. Inputs to this stage are the prime power and modulation pulses. The modulation pulses are used to turn the electron beam on or off. The third input is a control bus that provides the operating controls for the transmitter. Outputs from this system are monitor and status signals.

Figure 19.17(b) shows a high-power transmitter that uses an amplitron CFA as the output stage and a TWT as the driver for the amplifier. This system might be used for a high-power radar. The RF input signal to the TWT is first amplified to a level 10–15 dB below the final output. The output of the TWT is then fed to the input of the amplitron.

The output of the amplitron is connected to the antenna circuit, which may consist of a duplexer circuit, monitoring circuits, and the antenna. The TWT shown is connected to a high-voltage power supply and a modulator. Inputs to this stage are the prime power and modulation pulses. The modulation pulses are used to turn the

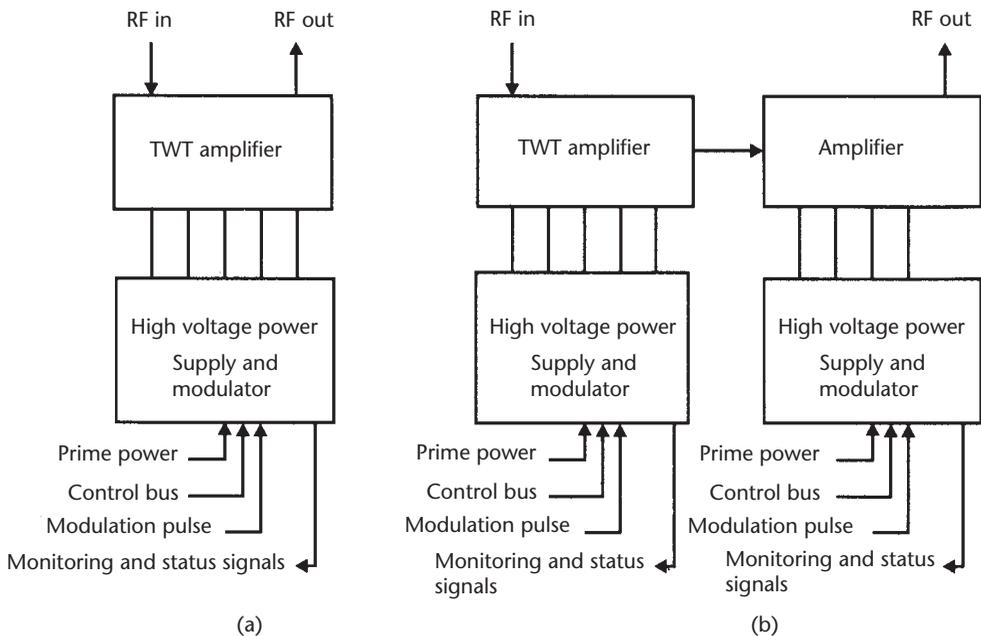


Figure 19.17 High-power microwave transmitters that use TWT amplifiers: (a) single-stage TWT transmitter; and (b) high-power transmitter that uses an amplitron CFA as the output stage and a TWT as the driver stage for the amplitron.

TWT electron beam on or off. The third input is a control bus that provides the operating controls for the transmitter. Outputs from this system are monitor and status signals. A similar high-voltage power supply is used for the ampliftron.

References

- [1] Whitaker, J. C., *Power Vacuum Tubes Handbook*, New York: Van Nostrand Reinhold, 1994.
- [2] Sivan, L., *Microwave Tube Transmitters*, London, England: Chapman and Hall, 1994, p. 15.
- [3] Sivan, L., *Microwave Tube Transmitters*, London, England: Chapman and Hall, 1994, p. 11.
- [4] Sivan, L., *Microwave Tube Transmitters*, London, England: Chapman and Hall, 1994, pp. 125–129.
- [5] Sivan, L., *Microwave Tube Transmitters*, London, England: Chapman and Hall, 1994, pp. 39–41.
- [6] Kennedy, G., *Electronic Communication Systems*, 3rd ed., New York: McGraw-Hill, 1985, pp. 351–385.
- [7] Skolnik, M. I., *Introduction to Radar Systems*, 2nd ed., New York: McGraw-Hill, 1980, pp. 192–216.
- [8] Skolnik, M. I., *Introduction to Radar Systems*, 2nd ed., New York: McGraw-Hill, 1980, p. 196.

About the Author

Ferril A. Losee has been both a university professor and an RF engineer who has successfully practiced in industry. He worked for 22 years in industry and for 18 years at Brigham Young University. He is a professor emeritus.

Mr. Losee received a B.S. in electrical engineering from the University of Utah in 1953 and an M.S. in electrical engineering from the University of Southern California in 1957. He first worked for Hughes Aircraft Company, where he was an RF design engineer and a project leader in areas of HF communications and satellite communications. He then worked for the Aeronutronic Division of Ford on space and satellite communication, missile defense penetration aids, and missile defense systems. He joined the faculty of Brigham Young University (BYU) as the chairman of the Electrical Engineering Department in 1965, a position that he held for 16 years. Mr. Losee took early retirement from the university in 1983, returning to industry to work for SRS Technologies in the areas of radar and special-purpose X-ray systems. His final position in industry was with EG&G Special Projects, where his main area of work was instrumentation radar. Mr. Losee retired in 1993. During much of his time at BYU, he worked as a consultant for the government and for industry in areas such as missile defense and special purpose antennas.

Mr. Losee has been married to Dona Jean Hansen for 51 years. They have 9 children, 31 grandchildren, and 16 great-grandchildren.

Index

2.4-GHz cordless telephones, 6
5.8-GHz DSS cordless telephones, 6–7
8-PSK, 141, 189
16-PSK, 141, 189–90
90-degree coupled-line hybrid coupler, 249
3/4-DQPSK modulation, 189

A

Advanced Automatic Crash Notification System (AACN), 12–13

Advanced Mobile Phone Service (AMPS), 15, 16

Aeronautical mobile satellite service, 25

Aeronautical mobile service, 24–25

Airborne direction finder (ADF), 35–36

Airborne radars

 Doppler navigation, 59–60

 military multiple function, 61

 side-looking, 61–62

 terrain-following, 61

 weather-avoidance, 58–59

See also Radars

Air-core loop antennas, 302–4

Air traffic control radar beacon system (ATCRBS), 39

Amateur service, 25

American digital cellular (ADC) system, 15

Amplifiers, 149–63

 BJT, 417–26

 collector-modulated class C transistor, 178–79

 common-base, 422–24

 common-collector, 424–26

 common-emitter, 417–22

 coupled-cavity traveling-wave tube, 458
 gain, 22

 grid-modulated class C, 178

 grid-type vacuum tube, 447–50

 gyrotron, 462

 helix-type traveling-wave tube, 455–58

 high-power transistor, 160

 high-power vacuum tube, 445–64

 IF, 155–57

 LF, 157

 multiple-cavity klystron, 453–55

 plate-modulated class C, 177–78

 RF, 149–55, 157–63

Amplitrons, 461–62

Amplitude modulation (AM), 122–23

 detectors, 192, 193

 modulators for, 177–79

See also Modulation

Analog-to-digital converter (ADC), 46, 137

Angle measurements, 73–74

 defined, 73

 pointing errors, 73

Antenna arrays, 322–30

 broadside line, 324–25

 end-fire line, 322–24

 frequency-scanned, 328–29

 mechanically scanned, 326–29

 patch, 338

 phase-scanned, 327–28

 planar, 326

 scanning methods, 326–30

 with space feeds, 329–30

Antenna gain, 22

 beam angle relationship, 287

 dipole antennas, 296

 power vs. directional, 287

 Yagi-Uda antenna, 300

Antennas, 83–84, 287–338

 aperture-type, 288

 corner reflector, 330

 dipole, 292–99

 ERP, 288

 flat-plate reflector, 330

 helical, 304–6

 hohorn, 332

 horn, 318–20

 isotropic, 287

- Antennas (continued)
- lens, 320–22
 - log-periodic, 308–13
 - loop, 302–4
 - monopole, 289–92
 - notch, 318
 - parabolic reflector, 330–33
 - patch, 44, 45, 333–38
 - receive, 84, 287
 - sleeve, 301–2
 - slot, 313–17
 - spiral, 306–8
 - transmit, 83–84, 287
 - turnstile, 297–99
 - Yagi-Uda, 299–301
- Archimedean spiral antennas, 307–8
- Atmospheric absorption, 87–88
- Atmospheric noise, 107–8
- Atmospheric refraction, 89–90
- Attenuation, 88–89
- coefficients, 87, 88
 - ice, 89
 - for knife-edge diffraction, 91
 - rain, 89
 - waveguide, 265, 266
- Automatic gain control (AGC), 157
- Avalanche photodiode (APD), 414
- B**
- Baluns, 243–44
- defined, 243
 - examples, 244
 - illustrated, 244
- Bandpass filters, 367–68
- SAW, 395
 - stripline, 251–52
 - YIG, 393
- See also* Filters
- Beam angle, 287
- Beidou system, 50
- Bessel functions, 127
- Binary phase-shift keying (BPSK), 140, 187
- Biphase phase shift keying, 145
- Bipolar junction transistors (BJTs), 415–17
- NPN, 415
 - PNP, 415–16
 - uses, 417
- See also* BJT amplifiers
- BJT amplifiers, 417–26
- clamping diode, 436
 - common-base, 422–24
 - common-collector, 424–26
 - common-emitter, 417–22
 - configurations, 417–26
 - dc bias circuits, 434–36
 - FET amplifier comparison, 431–33
 - high-power, 436
 - stability with, 436–37
 - weak spot, 433
- See also* Bipolar junction transistors (BJTs)
- Bluetooth, 31
- Branch line 90-degree hybrid couplers, 245–46
- Branch-line duplexers, 280–82
- Brewster angle, 93
- Broadbanding, 336
- Broadcasting satellite service, 27
- Broadcasting service, 26–27
- Broadside line antenna arrays, 324–25
- defined, 324
 - grating lobe angle, 325
 - illustrated, 324
- See also* Antenna arrays; Antennas
- Bulk acoustic wave delay lines, 400–402
- configurations, 400, 401
 - defined, 400
 - delay devices, 402
 - delay media, 401
 - at microwave frequencies, 401
 - performance parameters, 402
- See also* Delay lines
- C**
- Capacitive reactance, 348, 349
- Capacitive susceptance, 349
- Capacitors, 346–49, 383
- bypass, 447
 - current, 346
 - elements, 346
 - fixed-value, 348
 - mica dielectric, 348
 - two-plate, 347
- Carrier-phase measurements, 48–49
- Cavity-backed rectangular-slot antennas, 315
- Cavity resonators, 283–84
- cylinder-type, 283, 284
 - defined, 283
 - halfwave, 283, 284
 - reentrant-type, 283–84
- C-band radar, 56
- Cellular telephone systems, 12–19
- AMPS, 15
 - calls, 18

- characteristics, 15
- commercially available, 19
- digital, 15–17
- electronic serial number (ESN), 18
- handoff, 18
- introduction, 12–13
- mobile identification number (MIN), 18
- propagation, 14
- roaming, 18
- spatial frequency reuse, 13–14
- types of, 14–17
- See also* Telephone systems
- Ceramic filters, 387–88
 - characteristics, 387
 - IF amplifier, 389
 - piezoelectric ceramics, 387
 - See also* Filters
- Chebyshev filters, 366
- Chirp pulse modulation, 133
- Circular waveguides, 270–71
 - cutoff wavelengths, 271
 - defined, 270
 - field configurations, 270–71
 - guide wavelength, 271
 - illustrated, 271
 - See also* Waveguides
- Circulators
 - ferrite, 252–54
 - illustrated, 252
 - waveguide, 279
- Citizen band radio, 27–28
- Coaxial cable connectors, 223
- Coaxial cable lines, 4–5
- Coaxial cavity magnetrons, 459–61
 - cross-section, 460
 - defined, 459–60
 - example, 461
 - performance, 461
- Coaxial directional couplers, 242–43
 - loop-type, 242–43
 - two-hole probe-type, 243
 - See also* Directional couplers
- Coaxial electromechanical switches, 254–56
 - illustrated, 255
 - SPDT, 254–55
 - SPMT, 255
 - transfer, 255–56
- Coaxial terminations, 242
- Coaxial transmission lines, 220–22
 - characteristics, 222
 - coupling to waveguides, 269
 - defined, 220
 - illustrated, 220
 - microwave, 222
 - upper frequency limit, 221
 - upper power-handling limit, 221
 - See also* Transmission lines
- Collector-modulated transistor amplifiers, 178–79
- Colpitts crystal oscillator, 164, 385
- Common-base amplifiers, 422–24
 - current gain, 423
 - defined, 422–23
 - power gain, 424
 - schematic diagram, 423
 - voltage gain, 423–24
 - See also* BJT amplifiers
- Common-collector amplifiers, 424–26
 - current gain, 425
 - defined, 424
 - schematic diagram, 425
 - voltage gain, 425
 - See also* BJT amplifiers
- Common-emitter amplifiers, 417–22
 - characteristics plot, 418, 419, 420
 - circuit variations, 417
 - current gain, 421
 - defined, 417
 - equivalent circuits, 420, 421
 - illustrated, 418
 - power gain, 422
 - self-bias circuit, 435
 - transformers, 417–18
 - voltage gain, 422
 - See also* BJT amplifiers
- Communication satellite relay systems, 11
- Communication services, 24–29
 - aeronautical mobile, 24–25
 - aeronautical mobile satellite, 25
 - amateur, 25
 - broadcasting, 26–27
 - broadcasting satellite, 27
 - citizen band radio, 27–28
 - mobile, 29
 - pager systems, 28–29
 - remote control systems, 29
 - VHF/UHF FM two-way radio, 28
 - See also* Wireless communication systems
- Communication systems
 - example block diagrams, 201–16
 - FM broadcast, 205–8
 - HF, with single-sideband modulation, 201–3
 - link analysis, 115–16

- Communication systems (continued)
 - microwave relay, 208–11
 - propagation performance equations, 99–100
 - required SNR, 115
 - satellite relay, 211–14
 - transmitter RF power amplifiers for, 158–61
 - VHF/UHF ground-to-air, 203–5
- Complex resonant circuits, 352–54
- Conical spiral antennas, 308
- Continuous-wave modulation, 134
- Conventional magnetrons, 458–59
- Cordless telephones, 5–7
 - 2.4-GHz, 6
 - 5.8-GHz DSS, 6–7
 - base unit, 6
 - system illustration, 5
 - See also* Telephone systems
- Corner reflector antennas, 330
- Coupled-cavity TWT amplifiers, 458
- Course acquisition (C/A-code), 42–43, 44
- Crossed-field amplifier (CFA) chain, 161
- CW amplifier chain, 158
- CW radar, 53
 - speed measurement, 58
 - velocity measurement with, 70–72
 - See also* Radar
- D**
- Dc bias circuits, 434–37
 - for BJT amplifiers, 434–36
 - for enhancement-mode MOSFET, 437
- Decibels, 22–23
 - amplifier gain, 22
 - antenna gain, 22
 - filter loss, 22
 - power level, 23
 - voltage ratio, 23
- Delay-line cancelers, 77–78
- Delay lines, 393–402
 - bulk acoustic wave, 400–402
 - dispersive, 395, 396–99
 - nondispersive, 393–95
- PSK, 396
- SAW, 393–94
 - surface acoustic wave, 393–99
 - tapped, 395–96
- Demodulators, 191–98
- Demultiplex system (deMUX), 216
- Detectors, 191–98
 - amplitude modulation, 192, 193
 - FM, 194–97
 - I and Q, 194
 - phase, 197–98
 - PLL, 195–97
 - product, 192–94
 - waveguide, 279–80
- Dielectric lens antennas, 321
- Dielectric materials, 224
- Dielectric resonator oscillators, 164, 388–91
 - coupling, 390
 - description, 388
 - diameter, 389
 - dimensions, 389
 - electrical tuning, 391
 - examples, 391
 - mechanical tuning, 391
 - parameters, 388–90
 - See also* Oscillators
- Differential corrections, 49
- Differential GPS (D-GPS), 49
- Diffraction, 90–92
 - with communication system, 91
 - defined, 90
 - knife-edge, 91
- Digital cellular telephone systems, 15–17
- Digital signal processing, 78–79
- Diode mixers, 169–74
 - double-balanced, 171–74
 - single-balanced, 170–71
 - single-ended, 169–70
 - See also* Mixers
- Diodes
 - IR photo, 413–14
 - junction, 403–4
 - light-emitting, 413
 - microwave Gunn, 408–10
 - microwave IMPATT, 410–12
 - microwave tunnel, 408
 - PIN, 405–6
 - Schottky-barrier, 404–5
 - semiconductor IR laser, 412
 - step-recovery, 407
 - varactor, 406–7
 - Zener, 404
- Dipole antennas, 292–99
 - dipole sizes, 299
 - distribution/antenna patterns, 296–97
 - electrically small, 294
 - fan, 297
 - fat, 294, 295
 - folded, 294, 295
 - half-wave, 292–93

- impedance, 294–96
 - peak antenna gain, 296
 - sleeve, 302
 - thin-wire, 292–94
 - turnstile, 297–99
 - See also* Antennas
 - Directional couplers
 - broadside-coupled, 249
 - coaxial, 242–43
 - quarter-wave coupled-line, 248–49
 - single-section quarter-wave, 249
 - waveguide, 276–77
 - Direct sequence spread spectrum, 142–48
 - applications, 147–48
 - biphase modulator, 145
 - jamming margins, 146–47
 - systems, 144–47
 - waveforms, 146
 - Dispersive delay lines, 395, 396–99
 - illustrated, 398
 - performance parameters, 399
 - SAW, 396
 - See also* Delay lines
 - Distance-measuring equipment (DME), 37
 - Double-balanced diode mixer, 171–74
 - frequency spectrum, 171–72
 - illustrated, 171
 - LO output signal, 172
 - Double reflector matching, 242
 - Double-sideband (DSB) suppressed carrier modulation, 123–25
 - frequency spectrum, 124
 - modulators for, 179–80
 - signal production, 123
 - transmitter system, 180
 - waveform, 124
 - Duplexers, 280–83
 - balanced, 282, 283
 - branch-line, 280–82
 - defined, 280
 - gas-tube switches, 280
 - E**
 - Effective radiated power (ERP), 288
 - Electromechanical transfer switches, 255–56
 - Electronic countermeasures (ECM), 117–18
 - jamming signal and, 117–18
 - RF power amplifiers/oscillators for, 161–63
 - Electronic serial number (ESN), 18
 - End-fire line antenna arrays, 322–24
 - defined, 322
 - illustrated, 323
 - maximum element spacing, 322–23
 - See also* Antenna arrays; Antennas
 - Equiangular spiral antennas, 306–7
 - F**
 - Faraday rotation isolators, 277–78
 - Fat dipole antennas, 294, 295
 - FCC frequency allocation, 23–24
 - Ferrite circulators/isolators, 252–54
 - Ferrite-core loop antennas, 304
 - FET mixers, 175
 - Fiber optic cable propagation, 97
 - Fiber optics cable telephone relay systems,
 - 8–10
 - defined, 8–9
 - FTX-180, 10
 - illustrated, 10
 - optical fibers, 9
 - Field-effect transistors (FETs), 426–31
 - amplifier comparison, 431–33
 - amplifier stability, 436–37
 - bias circuits, 436
 - junction (JFETs), 426–28
 - metal-oxide semiconductor (MOSFETs), 429–31
 - metal-semiconductor (MESFETs), 428–29
 - Filters
 - bandpass, 367–68, 393
 - ceramic, 387–88
 - Chebyshev, 366
 - highpass, 366
 - LC, 365–68
 - loss, 22
 - lowpass, 365–66
 - monolithic crystal, 386–87
 - stripline, 249–52
 - YIG, 391–93
- Flat-plate reflector type antennas, 330
- Flat square spiral inductors, 343–44
 - approximate inductance, 344
 - illustrated, 344
 - uses, 343
- Flexible waveguides, 266, 272–73
- FMCW radar
 - modulation, 134, 135
 - range measurements, 69, 71
 - velocity measurements, 72
 - See also* Radar
- Folded dipole antennas, 294, 295
- Foster-Seeley discriminator, 194, 195
 - illustrated, 195
 - operation, 195

Free-space path loss, 87
 Frequency allocation, 23–24
 Frequency bands, 22, 23
 Frequency division multiple access (FDMA), 215
 Frequency-division multiplex modulators, 184
 transmitter system, 136–37
 Frequency modulation, 126–29
 detectors, 194–97
 index, 127
 modulators, 184–87
 narrowband, 128
 phase-shift method, 186
 PLL method, 187
 spectra, 129
 waveform amplitude, 126
 Frequency modulation broadcast systems, 205–8
 example, 206
 illustrated, 206
 receiver, 207
 transmitter, 207
 Frequency multipliers, 166–68
 step-recovery diode, 167–68
 transistor, 168
 varactor diode, 166–67
 Frequency-scanned arrays, 328–29
 beam-pointing direction, 328–29
 illustrated, 328
 planar, 329
 Frequency shift keying (FSK), 140, 187
 Frequency synthesizers, 165–66
 defined, 165
 direct, 165
 indirect, 165, 166
 Front-end low-noise RF amplifiers, 149–55

G

Galactic noise, 108–10
 GALILEO system, 50
 Gas-tube switches, 280
 Global Positioning System (GPS), 12, 40–51
 applications, 50–51
 C/A-code, 42–43
 carrier-phase measurements, 48–49
 constellation, 40
 control segment, 43
 differential (D-GPS), 49
 pseudorange measurements, 47–48
 receiver hardware, 44–47
 relative positioning, 49

 satellite description, 41–43
 satellite generations, 40–41
 space segment, 40
 system architecture, 40–50
 user segment, 43–44
 Global system for Mobile Communications (GSM), 15, 17
 GLONASS satellite system, 49–50
 GPS. *See* Global Positioning System
 Grid-modulated amplifiers, 178
 Grid tubes, 445–51
 pentode, 446
 power capability limit, 446
 terms, 446
 tetrode, 446
 triode, 445–46
 Grid-type vacuum tube amplifiers, 447–50
 Ground-based instrumentation, 60–61
 Ground-based military radars, 62
 Ground noise, 111
 Ground-to-air communication systems, 203–5
 Ground-wave propagation, 95–96
 Gunn diodes, 408–10
 epitaxial GaAs, 408, 409
 frequency range, 410
 phenomenon, 408–9
 Gyrotron oscillators/amplifiers, 462

H

Half-wave dipole antennas, 292–93
 pattern, 293, 294
 with reflectors, 330
 See also Dipole antennas
 Handoff, 18
 Helical antennas, 304–6
 in arrays, 306
 axial-mode, 305
 coaxial input, 304
 half-power beamwidth, 304–5
 input resistance, 305–6
 See also Antennas
 Helix-type TWT amplifiers, 455–58
 characteristics, 456, 457
 defined, 455
 illustrated, 456
 uses, 456
 wideband, 457–58
 Heterojunction bipolar transistors (HBTs), 434
 HF communication systems, 201–3
 HF radar, 55
 HF wideband RF transformers, 374–78
 4:1, 376, 378

9:1, 377
 conventional, 375–76
 one-turn, 375
 transmission-line, 375, 378–80
 uses, 374
See also RF transformers

High electron mobility transistors (HEMTs),
 433–34
 noise performance comparison, 434
 schematic cross section, 433

Highpass filters, 250–51, 366

High-power microwave (HPM) generators,
 132–33

Hoghorn antennas, 332

Horn antennas, 318–20
 illustrated, 319
 optimum, 320
 pyramidal, 318–19
 rectangular, 320
See also Antennas

Hybrid coupler circuits, 381

Hybrid junctions, 273–75
 hybrid ring/rat race, 274–75
 T-junction, 273–74
See also Waveguides

I

ICBM defense radars, 62–63

IEEE 802.11 standard, 30

IEEE 802.11b standard, 30

IEEE 802.11g standard, 31

IF amplifiers, 155–57
 frequencies, 156
 output, 202–3, 205
See also Amplifiers

IMPATT diodes, 410–12
 current/voltage characteristics, 411
 defined, 410
 material, 411
 schematic diagram, 411

IMPATT oscillator, 165

Impedance matching, 437–38
 design with Smith chart, 360–65
 with LC circuits, 358–60
 L-section impedance matching design with,
 362
 pi-section impedance-matching design with,
 364
 with quarter-wave transformer, 236–40
 with short-circuited stub, 240–42
 with Smith chart, 236–42
 for transistor amplifiers, 437–38

T-section impedance matching design with,
 363
 waveguide, 275–76

Inductive reactance, 343

Inductors, 342–46
 flat square spiral, 343–44
 toroidal, 344–45

Input interdigital transducers (IDTs), 394, 395

Instrument landing system (ILS), 37–38
 defined, 37
 illustrated, 38
See also Radionavigation systems

Ionospheric propagation, 93–95
 example reflection paths, 94
 HF communication and, 95
See also Propagation

IR photodiodes, 413–14

IS-136, 16

Isolators, 252–54
 Faraday rotation, 277–78
 resonance absorption, 278
 waveguide, 277–79

J

Jamming margins, 146–47

Japanese digital cellular system (JDS), 15

Joint Tactical Information Distribution System
 (JTIDS), 130

Junction diodes, 403–4

Junction FETs (JFETs), 426–28
 defined, 426
 illustrated, 426
 schematic diagram, 427
 voltage gain, 428

K

K-band radar, 56

K_a -band radar, 57

K_u -band radar, 56

Klystrons, 162, 453, 454

L

Land clutter, 67

L-band radar, 55

LC filters, 365–68
 attenuation vs. frequency, 366
 bandpass, 367–68
 band-reject, 366
 Chebychev, 366
 frequency response, 365
 highpass, 366

- LC filters (continued)
 - lowpass, 365–66
 - multi-stage, 367
 - uses, 365
 - LC oscillators, 163
 - Lens antennas, 320–22
 - dielectric, 321
 - illustrated, 321
 - Luneburg, 321–22
 - metallic-plate, 322
 - LF amplifiers, 157
 - Light-emitting diodes, 412
 - Lightning protection, 257
 - Limited space-charge accumulation (LSA) mode oscillators, 165
 - Link analysis, 115–17
 - communication system, 115–16
 - radar system, 116–17
 - Local loops, 3
 - Log-periodic antennas, 308–13
 - dipole array, 308–11
 - trapezoidal-toothed, 311–13
 - triangular-toothed, 313, 314
 - See also* Antennas
 - Loop antennas, 302–4
 - air-core, 302–4
 - electrically small, 303
 - ferrite-core, 304
 - illustrated, 303
 - pattern, 302, 303
 - See also* Antennas
 - Loop-type directional couplers, 242–43
 - Loran system, 34–35
 - Low-noise RF amplifiers, 149–55
 - bipolar junction transistors, 150
 - classes, 150–51
 - conduction angles, 151
 - example design, 152, 153
 - response curves, 154
 - transistor transfer function, 152, 153
 - See also* Amplifiers; RF amplifiers
 - Lowpass filters, 249–50, 365–66
 - LP dipole array (LPDA) antennas, 308–11
 - active region, 309
 - configuration, 309
 - construction, 308
 - defined, 308
 - gain, 310
 - impedance, 309
 - pattern gain, 309
 - scale factor, 308
 - two-layer construction, 309
 - Luneburg lens antennas, 321–22
- ## M
- Magnetrons, 458–61
 - coaxial cavity, 459–61
 - conventional, 458–59
 - Man-made noise, 111–12
 - Manufacturer's data sheets (transistors), 440–43
 - Master control station (MCS), 43
 - Mechanically scanned arrays, 326–29
 - Metallic-plate lens antennas, 321, 322
 - Metal-oxide semiconductor FETs (MOSFETs), 429–31
 - common-source amplifier, 430
 - depletion/enhancement type, 432
 - N-channel depletion/enhancement-mode, 431
 - N-channel enhancement-mode, 430
 - power transistors, 432
 - Metal-semiconductor FETs (MESFETs), 428–29, 432
 - MF radar, 54–55
 - Microstrip rat race hybrid coupler, 246–47
 - Microstrip resonators, 333, 335
 - Microstrip transmission lines, 224–25
 - characteristic impedance, 225
 - defined, 224
 - dielectric materials, 224
 - illustrated, 224
 - power-handling capability, 225
 - See also* Transmission lines
 - Microwave Gunn diodes, 408–10
 - Microwave IMPATT diodes, 410–12
 - Microwave landing system (MLS), 39
 - Microwave relay systems, 7–8, 208–11
 - block diagram, 210
 - C-band, 208–9
 - elements, 7–8
 - illustrated, 8, 209
 - K_u -band, 208
 - signal-flow diagram, 209
 - uses, 208
 - Microwave tubes, 451–64
 - amplitrons, 461–62
 - average power, 453
 - circuit configurations, 462–64
 - gyrotron oscillator, 462
 - introduction to, 451–53
 - magnetrons, 458–61
 - operating characteristics comparison, 452
 - Microwave tunnel diodes, 408

- Miller oscillator, 385
- Millimeter-wave radar, 57
- Minimum shift keying (MSK), 130
- Mixers, 169–75
 - diode, 169–74
 - FET, 175
 - performance characteristics, 175
 - termination insensitive (TIM), 173–74
 - transistor, 174–75
 - waveguide, 280, 281
- Mobile identification number (MIN), 18
- Mobile service, 29
- Mobile telecommunication switching office (MTSO), 15
- Modular shift register generators (MSRGs), 144
- Modulation, 121–48
 - amplitude, 122–23
 - chirp pulse, 133
 - continuous-wave, 134
 - direct sequence spread spectrum, 142–48
 - double sideband suppressed carrier, 123–25
 - frequency, 126–29
 - frequency-modulation CW, 134, 135
 - phase code modulated pulse, 133–34
 - pulsed CW, 130–32
 - for radar, 130–34
 - for telemetry, 129
 - two-state types for binary signals, 139–40
 - vestigial sideband, 125
- Modulators, 177–91
 - 8-PSK, 189
 - for amplitude modulation, 177–79
 - balance, 182
 - BPSK, 187, 188
 - for double-sided modulation, 179–80
 - for FDM, 184
 - for FM, 184–87
 - for FSK, 187
 - for PCM TDM, 190–91, 192
 - for PSK, 187–90
 - QPSK, 188
 - for single-sideband modulation, 181–84
 - vestigial sideband, 181
- Monolithic crystal filters, 386–87
 - characteristics, 387
 - frequency response, 386
 - See also* Filters
- Monopole antennas, 289–92
 - electrically small, 291–92
 - example, 290
 - illustrated, 289
 - impedance, 290–91
 - large-size, 291
 - quarter-wave, 289
 - reactance, 291
 - resistance, 291
 - sleeve, 301–2
 - thin-wire, 289–90
 - wideband, 290
 - See also* Antennas
- Monopulse tracking radars, 80–81
- Moving-target indicator. *See* MTI radars
- MTI radars, 74–80
 - block diagram, 75, 76
 - defined, 74
 - delay-line cancelers, 77–78
 - digital signal processing, 78–79
 - moving platform, 79–80
 - with power amplifier transmitter, 75
 - with power oscillator transmitter, 76
 - signal processing, 76–79
 - See also* Radars
- Multipath, 92–93
- Multiple-cavity klystron amplifiers, 453–55
- Multiplexing, 3
- N**
- Negative resistance two-terminal oscillators, 165
- Noise
 - atmospheric, 107–8
 - galactic, 108–10
 - ground, 111
 - limitations, 103
 - man-made, 111–12
 - multiple-stage systems with, 106–7
 - power, 104–5
 - RF, 103–4
 - solar, 110–11
 - sources, 104
 - types of, 107–12
- Nondispersive delay lines, 393–95
- Notch antennas, 318
- O**
- Octonary phase-shift keying (OPSK), 141
- Omega system, 33–34
 - defined, 33
 - lane boundaries/ambiguity, 34
 - navigation aid, 34
 - See also* Radionavigation systems
- Open-slot antennas, 313–15
- Open-wire lines, 4

- Oscillators, 163–65
 avalanche diode, 165
 Colpitts, 164, 385
 dielectric resonator, 164, 388–91
 Gunn, 165, 410
 gyrotron, 462
 high-power vacuum tube, 445–64
 JFET, 163
 LC, 163
 LSA, 165
 Miller, 385
 negative resistance two-terminal, 165
 quartz crystal, 163–64, 383–86
 for radars, navigation, ECM, 161–63
 surface acoustic wave delay line, 399–400
 transistor feedback, 163–64
 varactor-tuned, 185–86
 VCO, 163
 YIG resonator, 164
 YIG-tuned, 393
- P**
- Pacific Digital Cellular (PDC), 17
 Pager systems, 28–29
 Parabolic reflector antennas, 330–33
 with Cassegrain feed, 333
 directive gain, 331
 feed systems, 331
 field pattern, 331
 illustrated, 331
 with offset feed, 332
 uses, 330
 wavefronts, 330
See also Antennas
 Parallel resonant RLC circuits, 350–52
 admittance, 351
 elements, 350
 impedance characteristics, 351
 uses, 352
 Patch antennas, 44, 45, 333–38
 arrays, 338
 broadbanding, 336
 circular polarized, 337, 338
 defined, 333
 dual polarized, 337
 feeds, 335–36
 introduction, 333–34
 with multilayer feed structure, 338
 radiation patterns, 334
 SSFIP, 336–37
 types of, 334
See also Antennas
 Pentodes, 446
 Personal Communication Service (PCS), 17
 Phase code modulated pulse modulation,
 133–34
 Phase detectors, 197–98
 Phase-locked loop (PLL)
 detectors, 195–97
 oscillators, 140
 Phase-scanned arrays, 327–28
 Phase-shift keying (PSK)
 $\delta/4$ -DQPSK, 189
 8-PSK, 189
 16-PSK, 189–90
 BPSK, 187
 modulators for, 187–90
 QPSK, 187–88
 PIN diodes, 405–6
 in parallel, 406
 photodiodes, 413
 planar, 405
 schematic, 405
 in series, 406
 switches, 256–57
 uses, 405
See also Diodes
 Planar arrays, 326
 Plate-modulated amplifiers, 177–78
 Position Location and Reporting System
 (PLARS), 130
 Power amplifiers, 158–63
 for radars, navigation, ECM, 161–63
 transmitter RF, 158–61
 TWT, 161
See also Amplifiers
 Power combiners, 381
 Power level, 23
 Probability of intercept (POI), 113
 Product detectors, 192–94
 AM detection, 194
 DSB detection, 193–94
 illustrated, 193
 SSB detection, 193
 Propagation
 antennas, 83–84
 atmospheric absorption, 87–88
 atmospheric attenuation, 88–89
 atmospheric refraction, 89–90
 diffraction, 90–92
 fiber optic cable, 97
 free-space path loss, 87
 ground-wave, 95–96
 ionospheric, 93–95

- multipath, 92–93
 - performance equations for communication systems, 99–100
 - performance equations for radar systems, 100–102
 - RCS and, 97–98
 - scatter, 96–97
 - Pseudorandom noise (PN) generators, 142–44
 - Pseudorandom noise (PRN) codes, 42
 - Pseudorange measurements, 47–48
 - Pulse code modulation (PCM), 53
 - defined, 53
 - time-division multiplex (TDM), 190–91, 192
 - Pulse compression modulation, 53, 396–97
 - Pulsed-continuous-wave signals, 121–22
 - Pulsed CW modulation, 130–32
 - Pulse Doppler radars (PDRs), 74
 - Pulse position modulation (PPM), 129
 - Pulse repetition interval (PRI), 130
 - Pulse repetition rate (PRR), 113
 - Pulse-type radar
 - MTI and, 74–80
 - range measurements, 70
 - velocity measurements, 72–73
 - See also* Radars
- Q**
- Quadrature amplitude modulation (QAM), 8
 - Quarter-wave coupled-line directional couplers, 248–49
 - Quarter-wave monopole antennas, 289–90
 - Quarter-wave transformer, 236–40
 - Quartz crystals
 - CINOX, 385
 - equivalent circuit, 384
 - oscillators, 163–64, 383–86
 - sizes/shapes, 384
 - symbol, 384
 - use, 383
 - vibration, 383
 - Quaternary phase-shift keying (QPSK), 141, 187–88, 198
- R**
- Radar absorbing material (RAM) coatings, 64, 154
 - Radar altimeters, 59
 - Radar clutter, 65–68
 - defined, 65
 - geometry, 66
 - land, 67
 - RCS, 66–67
 - sea, 68
 - Radar cross section (RCS), 53
 - for aircraft, 64
 - for clutter, 66–67
 - defined, 63
 - fluctuations, 65
 - frequency and, 65
 - at microwave frequencies, 65
 - propagation and, 97–98
 - for sphere, 63
 - of targets, 63–65
 - Radar measurements, 63–74
 - angle, 73–74
 - clutter, 65
 - range, 68–70
 - RCS, 63–65
 - velocity, 70–73
 - Radars, 53–82
 - airborne, 61–62
 - airborne Doppler navigation, 59–60
 - airborne weather-avoidance, 58–59
 - C-band, 56
 - concepts, 53–54
 - CW, 53, 58
 - with electronic countermeasures, 117–18
 - frequency bands, 54
 - ground-based instrumentation, 60–61
 - ground-based military, 62
 - HF, 55
 - ICBM defense, 62–63
 - K-band, 56
 - K_a -band, 57
 - K_u -band, 56
 - L-band, 55
 - link analysis, 116–17
 - MF, 54–55
 - millimeter-wave, 57
 - modulation for, 130–34
 - monopulse, 73, 74
 - MTI, 74–80
 - propagation performance calculations, 100–102
 - RF power amplifiers/oscillators for, 161–63
 - S-band, 56
 - ship-based military, 62
 - ship-based search/surveillance, 60
 - shore-based search/surveillance, 60
 - space applications, 60
 - surface-based fire-control, 162
 - surveillance, 57–58, 60

- Radar measurements (continued)
 - tracking, 80–82
 - types, 57–63
 - types, based on frequency, 54–57
 - UHF, 55
 - V-band, 57
 - VHF, 55
 - W-band, 57
 - weather, 162
 - X-band, 56
 - Radar scattering, 64
 - Radio beacons, 35–36
 - Radionavigation systems, 33–39
 - ADF, 35–36
 - ATCRBS, 39
 - DME, 37
 - ILS, 37–38
 - Loran, 34–35
 - MLS, 39
 - Omega system, 33–34
 - TACAN, 38–39
 - Transit system, 39
 - VOR, 36
 - Range gate pull-off (RGPO), 82
 - Range measurements, 68–70
 - concepts, 69–70
 - defined, 69
 - Rat race, 274–75
 - Receive antennas, 84, 287, 288
 - Rectangular waveguides, 260–67
 - attenuations, 265, 266
 - characteristics, 262
 - CW power ratings, 265
 - defined, 260–61
 - fields, 263
 - group velocity, 262
 - higher-order modes, 266–67
 - illustrated, 260
 - length, 261
 - phase velocity, 261–62
 - with slots, 264–65
 - TE₁₀ mode, launching, 268
 - TE₁₁ mode, 266–67
 - TE₂₀ mode, 266, 267
 - wavelength, 262–63
 - See also* Waveguides
 - Reflection coefficients, 227–28, 232
 - Reflective array compressor, 398
 - Remote control systems, 29
 - Resistors, 342
 - Resonance absorption isolators, 278
 - Resonators
 - dielectric, 388–91
 - quartz crystal, 383–86
 - YIG, 391–93
 - RF amplifiers, 149–55, 157–63
 - BJT, 151, 152
 - FET, 151, 152
 - front-end low-noise, 149–55
 - for radars, navigation, ECM, 161–63
 - shielding concepts, 437
 - transmitter, for communication systems, 158–61
 - transmitter chains, 157–58
 - RF chokes, 346, 358
 - RF electrical field waveforms, 86–87
 - RF noise, 103–4
 - RF resistors, 342
 - RF transformers, 371–81
 - with brass tubes, 375
 - for calculations, 372
 - double-tuned, 374
 - ferrites, 372
 - high-frequency wideband, 374–78
 - magnetic core materials, 371–73
 - maximum power transfer, 374
 - transmission-line, 378–80
 - See also* Transformers
 - Ridged waveguides, 270
 - RLC circuits, 349–52
 - parallel resonant, 350–52
 - series resonant, 349–50
 - Roaming, 18
 - Rotary joints, 272
- ## S
- Satellite relay communication systems, 211–14
 - converting to TDM, 214
 - digital transmission, 214
 - evolution, 212–13
 - expansion, 214
 - geostationary satellites, 213
 - illustrated, 212
 - for mobile stations, 213
 - Satellite relay Earth stations, 215–16
 - SAW delay lines, 393–94
 - S-band radar, 56
 - Scatter propagation, 96–97
 - Schottky-barrier diodes, 404–5
 - Sea clutter, 68
 - Self-bias circuits, 435
 - Semiconductor IR laser diodes, 412
 - Semiconductor materials, 403
 - Sensing and diagnostic module (SDM), 12

- Series resonant RLC circuits, 349–50
- Ship-based military radar, 62
- Ship-based search/surveillance radar, 60
- Shore-based search/surveillance radar, 60
- Short-circuited stub impedance matching, 240–42
 - illustrated, 240
 - Smith chart, 241
 - See also* Impedance matching
- Shunt stub dc returns, 245
- Signal-to-noise improvement, 112–13
- Signal-to-noise ratio (SNR), 113–15
 - communication systems, 115
 - probability of detection vs., 114
- Single-balanced diode mixers, 170–71
- Single-channel transmitter system, 134–36
 - block diagram, 135
 - mixer output, 136
- Single-ended diode mixers, 169–70
- Single-sideband (SSB) modulation, 125–26
 - demodulator output, 203
 - efficiency, 125
 - HF communication system with, 201–3
 - modulators for, 181–84
 - phase-shift method, 183
 - reduced carrier frequency spectrum, 126
 - suppressed carrier, 126
 - transmitter block diagram, 182
 - See also* Modulation
- Sixteen amplitude-phase keying, 142
- Sixteen phase-state keying (16-PSK), 141, 189–90
- Skin depth, 341–42
- Sleeve antennas, 301–2
 - dipoles, 302
 - illustrated, 301
 - monopoles, 301–2
 - See also* Antennas
- Sliding correlator, 147
- Slot antennas, 313–17
 - cavity-backed rectangular, 315
 - illustrated, 314
 - open, 313–15
 - waveguide-fed, 315–17
 - See also* Antennas
- Smith chart, 230–36
 - for circuit analysis, 355–57
 - in converting S-parameters, 359
 - defined, 231–32
 - examples using, 232–36
 - illustrated, 231
 - impedance/admittance coordinates, 230–32
 - impedance at point along transmission line, 237
 - impedance matching design with, 360–65
 - impedance matching with, 236–42
 - load admittances plotted on, 235
 - load impedances plotted on, 233
 - L-section impedance matching design with, 362
 - pi-section impedance matching design with, 364
 - reflection coefficients, 232
 - standing-wave circles plotted on, 233
 - T-section impedance matching design with, 363
 - uses, 230
 - voltage standing-wave ratio circles, 232
- Solar noise, 110–11
- Space radar applications, 60
- S-parameters, 357–58
 - defined, 357, 438
 - design methods with, 438–40
 - measurement setups, 357, 358
 - Smith chart in converting, 359
 - use example, 439–40
- Sparkgap switches, 257
- Spatial frequency reuse, 13–14
- Spiral antennas, 306–8
 - Archimedean, 307–8
 - conical, 308
 - equiangular, 306–7
 - See also* Antennas
- Split inline hybrid dividers/combiners, 247–48
 - compensated, 247–48
 - illustrated, 247
 - uncompensated, 247
- Splitters, 381
- Spread spectrum processing gain, 146
- Standing-wave ratio, 228–30
- Step-recovery diodes, 167–68, 407
- Stereo FM multiplex demodulator system, 208
- Stripline bandpass filters, 251–52
 - defined, 251–52
 - design, 252
 - illustrated, 251
- Stripline highpass filters, 250–51
 - defined, 250
 - illustrated, 251
- Stripline lowpass filters, 249–50
 - defined, 249–50
 - illustrated, 250
- Stripline quarter-wave directional couplers, 248–49

- Stripline rat race hybrid coupler, 246–47
- Stripline transmission lines, 225–27
 - characteristic impedance, 226
 - defined, 225
 - illustrated, 225
 - See also* Transmission lines
- Strip-slot-foam-inverted patch (SSFIP), 336–37
- Submarine cable relay systems, 10–11
- Subsidiary communication authorization (SCA)
 - generator, 205
- Superheterodyne receivers, 156
- Surface acoustic wave delay line oscillators, 399–400
- Surface acoustic wave delay lines, 393–99
- Surveillance radars, 57–58
 - CRT PPI display, 58
 - information friend or foe (IFF), 57
 - See also* Radars
- Switches
 - ATR, 280
 - coaxial electromechanical, 254–56
 - ferrite, 254
 - gas-tube, 280
 - PIN diode, 256–57
 - sparkgap, 257
 - TR, 280
 - waveguide, 279
- T**
- TACAN system, 38–39
- Tapped delay lines, 395–96
- TAT systems, 10–11
- Telemetry, modulation for, 129
- Telephone lines, 4–7
 - coaxial cable, 4–5
 - cordless telephones, 5–7
 - open-wire, 4
 - twisted-pair, 4
- Telephone relay systems, 7–11
 - communication satellite, 11
 - fiber optic cable, 8–10
 - microwave, 7–8
 - submarine cable, 10–11
- Telephone systems, 3–19
 - cellular, 12–19
 - cordless, 5
 - elements, 4
 - introduction, 3–4
 - lines, 4–7
 - relay, 7–11
- Termination insensitive mixers (TIMs), 173–74
 - defined, 173
 - schematic, 174
 - See also* Mixers
- Tetrodes, 446
 - high-power/frequency, 450–51
 - illustrated, 446
 - quarter-wave cavity with, 450
 - RF power amplifier schematic, 448
 - See also* Grid tubes
- Thin-wire dipole antennas, 292–94
 - bandwidth, 296
 - patterns, 293
 - resistance/reactance curves, 295
 - See also* Dipole antennas
- Thin-wire monopole antennas, 289–90
- Time assignment speech interpolation (TASI), 11
- Time-division multiple access (TDMA), 191
- Time division multiplex (TDM) transmitter, 137–39
 - block diagram, 139
 - with pulse code modulation, 137–39
 - system illustration, 138
 - wideband channel, 139
- Toroidal inductors, 344–45
 - defined, 344
 - illustrated, 345
 - inductance, 344–45
- Tracking radars, 80–82
 - monopulse, 80–81
 - with sequential gating, 81–82
 - See also* Radars
- Transformers
 - conventional, 369–71
 - hybrid couplers using, 381
 - ideal, 369
 - LF response, 369
 - RF, 371–73, 374–80
 - tuned, 373–74
 - with two windings, 370
- Transistor data sheets, 440–43
- Transistor feedback oscillators, 163–64
- Transistor mixers, 174–75
- Transistor multipliers, 168
- Transit system, 39
- Transmission lines
 - characteristics, 227–30
 - coaxial, 220–22
 - microstrip, 224–25
 - reflection coefficients, 227–28
 - standing-wave ratio, 228–30
 - stripline, 225–27

- TEM, 268, 269
 - two-wire, 219–20
 - wave velocity, 227
 - Transmission-line transformers, 375, 378–80
 - concepts, 379
 - configurations, 380
 - LF performance, 380
 - Transmit antennas, 83–84, 287
 - Transmitter RF amplifier chains, 157–58
 - Transmitter RF power amplifiers, 158–61
 - Transverse electromagnetic (TEM) waves, 85–86
 - defined, 85
 - illustrated, 85
 - power density, 86
 - TRAPATT oscillator, 165
 - Trapezoidal-toothed log-periodic antennas, 311–13
 - bandwidth, 313
 - measurements, 312
 - performance, 313
 - scale factor, 312
 - sheet metal-type, 311, 312, 313
 - wire-type, 311, 312
 - See also* Antennas
 - Traveling-wave tubes (TWTs), 41, 159
 - coupled-cavity (CCTWT), 161, 458
 - helix, 162, 455–58
 - power amplifiers, 161
 - uses, 161
 - Triangular-toothed log-periodic antennas, 313, 314
 - Triodes, 445–46
 - high-power/frequency, 450–51
 - illustrated, 446
 - See also* Grid tubes
 - Tuned transformers, 373–74
 - Tunnel diodes, 408, 409
 - Turnstile antennas, 297–99
 - azimuth pattern, 300
 - configuration, 300
 - defined, 297
 - elevation pattern, 300
 - polarization, 299
 - See also* Dipole antennas
 - Twisted-pair lines, 4
 - Two-hole probe-type directional couplers, 243
 - Two-wire transmission line impedance transformer, 244–45
 - Two-wire transmission lines, 219–20
 - as balanced lines, 219
 - characteristic impedance, 219
 - illustrated, 220
 - See also* Transmission lines
- U**
- UHF radar, 55
- V**
- Varactor diode frequency multipliers, 166–67
 - circuit illustration, 167
 - maximum output power, 167
 - Varactor-tuned crystal oscillators, 185–86
 - V-band radar, 57
 - Velocity measurements, 70–73
 - with CW radar, 70–72
 - with FMCW radar, 72
 - with pulse-type radar, 72–73
 - Very small antenna terminals (VSATs), 213
 - Vestigial-sideband (VSB) modulation, 125
 - block diagram, 181
 - modulators, 181
 - VHF omnidirectional range (VOR) systems, 36
 - VHF radar, 55
 - VHF/UHF FM two-way radio, 28
 - VHF/UHF ground-to-air communication systems, 203–5
 - Voltage-controlled oscillators (VCOs), 163
 - Voltage ratio, 23
 - Voltage standing-wave ratio circles, 232
 - Volume reflectivity, 69
- W**
- Waveguide-fed slot antennas, 315–17
 - illustrated, 317
 - side-wall, 316–17
 - top-wall, 315–16, 317
 - See also* Slot antennas
 - Waveguides, 259–84
 - accessories, 273
 - attenuators, 276
 - behavior, 259–60
 - characteristic wave impedance, 268–69
 - circular, 270–71
 - circulators, 279
 - coupling from coaxial line to, 269
 - defined, 259
 - detectors, 279–80
 - directional couplers, 276–77
 - E-plane T, 276
 - flanges, 271–72
 - flexible, 266, 272–73
 - hardware, 271–73

- Waveguides (continued)
 - hybrid junctions, 273–75
 - impedance matching, 275–76
 - introduction, 259–60
 - irises, 275
 - isolators, 277–79
 - mixers, 280, 281
 - modes, 259
 - rectangular, 260–67
 - resistive loads, 276
 - ridged, 270
 - rotary joints, 272
 - switches, 279
 - tapered transition sections, 272
 - types, 270–71
 - Wave velocity, 227
 - W-band radar, 57
 - Wireless communication systems, 21–31
 - background information, 21–23
 - Bluetooth, 31
 - communication services, 24–29
 - decibel use, 22–23
 - frequency allocation, 23–24
 - frequency bands, 22, 23
 - IEEE 802.11 standard, 30
 - IEEE 802.11b standard, 30
 - IEEE 802.11g standard, 31
 - introduction, 21
 - units/conversion, 21–22
 - WLANs, 29–30
 - Wireless local-area networks (WLANs), 29–30
- X**
- X-band radar, 56
- Y**
- Yagi-Uda antennas, 299–301
 - configuration, 300
 - defined, 299–301
 - gain characteristics, 300
 - patterns, 300, 301
 - See also* Antennas
 - YIG bandpass filters, 393
 - YIG resonators, 264, 391–93
- Z**
- Zener diodes, 404

Recent Titles in the Artech House Microwave Library

Active Filters for Integrated-Circuit Applications, Fred H. Irons

Advanced Techniques in RF Power Amplifier Design, Steve C. Cripps

Automated Smith Chart, Version 4.0: Software and User's Manual,
Leonard M. Schwab

Behavioral Modeling of Nonlinear RF and Microwave Devices,
Thomas R. Turlington

Broadband Microwave Amplifiers, Bal S. Virdee, Avtar S. Virdee, and
Ben Y. Banyamin

Computer-Aided Analysis of Nonlinear Microwave Circuits, Paulo J. C. Rodrigues

Design of FET Frequency Multipliers and Harmonic Oscillators, Edmar Camargo

Design of Linear RF Outphasing Power Amplifiers, Xuejun Zhang,
Lawrence E. Larson, and Peter M. Asbeck

Design of RF and Microwave Amplifiers and Oscillators, Pieter L. D. Abrie

Distortion in RF Power Amplifiers, Joel Vuolevi and Timo Rahkonen

*EMPLAN: Electromagnetic Analysis of Printed Structures in Planarly Layered
Media, Software and User's Manual*, Noyan Kinayman and M. I. Aksun

FAST: Fast Amplifier Synthesis Tool—Software and User's Guide, Dale D. Henkes

Feedforward Linear Power Amplifiers, Nick Potheary

Generalized Filter Design by Computer Optimization, Djuradj Budimir

High-Linearity RF Amplifier Design, Peter B. Kenington

High-Speed Circuit Board Signal Integrity, Stephen C. Thierauf

Intermodulation Distortion in Microwave and Wireless Circuits, José Carlos Pedro
and Nuno Borges Carvalho

Lumped Elements for RF and Microwave Circuits, Inder Bahl

Microwave Circuit Modeling Using Electromagnetic Field Simulation,
Daniel G. Swanson, Jr. and Wolfgang J. R. Hoefer

Microwave Component Mechanics, Harri Eskelinen and Pekka Eskelinen

Microwave Engineers' Handbook, Two Volumes, Theodore Saad, editor

Microwave Filters, Impedance-Matching Networks, and Coupling Structures,
George L. Matthaei, Leo Young, and E.M.T. Jones

Microwave Materials and Fabrication Techniques, Second Edition,
Thomas S. Laverghetta

Microwave Mixers, Second Edition, Stephen A. Maas

Microwave Radio Transmission Design Guide, Trevor Manning

Microwaves and Wireless Simplified, Thomas S. Laverghetta

Modern Microwave Circuits, Noyan Kinayman and M. I. Aksun

Neural Networks for RF and Microwave Design, Q. J. Zhang and K. C. Gupta

Noise in Linear and Nonlinear Circuits, Stephen A. Maas

Nonlinear Microwave and RF Circuits, Second Edition, Stephen A. Maas

QMATCH: Lumped-Element Impedance Matching, Software and User's Guide,
Pieter L. D. Abrie

Practical Analog and Digital Filter Design, Les Thede

*Practical RF Circuit Design for Modern Wireless Systems, Volume I: Passive Circuits
and Systems,* Les Besser and Rowan Gilmore

*Practical RF Circuit Design for Modern Wireless Systems, Volume II: Active Circuits
and Systems,* Rowan Gilmore and Les Besser

*Production Testing of RF and System-on-a-Chip Devices for Wireless
Communications,* Keith B. Schaub and Joe Kelly

Radio Frequency Integrated Circuit Design, John Rogers and Calvin Plett

RF Design Guide: Systems, Circuits, and Equations, Peter Vizmuller

RF Measurements of Die and Packages, Scott A. Wartenberg

The RF and Microwave Circuit Design Handbook, Stephen A. Maas

RF and Microwave Coupled-Line Circuits, Rajesh Mongia, Inder Bahl, and
Prakash Bhartia

RF and Microwave Oscillator Design, Michal Odyniec, editor

RF Power Amplifiers for Wireless Communications, Steve C. Cripps

RF Systems, Components, and Circuits Handbook, Second Edition, Ferril A. Losee

Stability Analysis of Nonlinear Microwave Circuits, Almudena Suárez and
Raymond Quéré

TRAVIS 2.0: Transmission Line Visualization Software and User's Guide, Version 2.0,
Robert G. Kaires and Barton T. Hickman

Understanding Microwave Heating Cavities, Tse V. Chow Ting Chan and
Howard C. Reader

For further information on these and other Artech House titles, including previously considered out-of-print books now available through our In-Print-Forever® (IPF®) program, contact:

Artech House

685 Canton Street

Norwood, MA 02062

Phone: 781-769-9750

Fax: 781-769-6334

e-mail: artech@artechhouse.com

Artech House

46 Gillingham Street

London SW1V 1AH UK

Phone: +44 (0)20 7596-8750

Fax: +44 (0)20 7630 0166

e-mail: artech-uk@artechhouse.com

Find us on the World Wide Web at: www.artechhouse.com
