FUNDAMENTALS OF WIRELESS

COMMUNICATION ENGINEERING

TECHNOLOGIES

WILEY SERIES ON INFORMATION AND COMMUNICATION TECHNOLOGY **Series Editors: T. Russell Hsing and Vincent K. N. Lau**

The Information and Communication Technology (ICT) book series focuses on creat ing useful connections between advanced communication theories, practical designs, and end-user applications in various next generation networks and broadband access systems, including fiber, cable, satellite, and wireless. The ICT book series examines the difficulties of applying various advanced communication technologies to prac tical systems such as WiFi, WiMax, B3G, etc., and considers how technologies are designed in conjunction with standards, theories, and applications.

The ICT book series also addresses application-oriented topics such as service management and creation and end-user devices, as well as the coupling between end devices and infrastructure.

**T. Russell Hsing, PhD**, is the Executive Director of Emerging Technologies and Services Research at Telcordia Technologies. He manages and leads the applied research and development of information and wireless sensor networking solutions for numerous applications and systems. Email: thsing@telcordia.com

**Vincent K.N. Lau, PhD**, is Associate Professor in the Department of Electrical Engineering at the Hong Kong University of Science and Technology. His current research interest is on delay-sensitive cross-layer optimization with imperfect system state information. Email: eeknlau@ee.ust.hk

*Wireless Internet and Mobile Computing: Interoperability and Performance* Yu-Kwong Ricky Kwok and Vincent K. N. Lau

*RF Circuit Design*

Richard C. Li

*Digital Signal Processing Techniques and Applications in Radar Image Processing* Bu-Chin Wang

*The Fabric of Mobile Services: Software Paradigms and Business Demands* Shoshana Loeb, Benjamin Falchuk, and Euthimios Panagos

*Fundamentals of Wireless Communication Engineering Technologies* K. Daniel Wong

FUNDAMENTALS OF WIRELESS

COMMUNICATION ENGINEERING

TECHNOLOGIES K. Daniel Wong

Copyright © 2012 by John Wiley & Sons, Inc. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.

Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at http://www.wiley.com/go/permission.

Limit of Liability/Disclaimer of Warranty: The publisher and the author make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation warranties of fitness for a particular purpose. No warranty may be created or extended by sales or promotional materials. The advice and strategies contained herein may not be suitable for every situation. This work is sold with the understanding that the publisher is not engaged in rendering legal, accounting, or other professional services. If professional assistance is required, the services of a competent professional person should be sought. Neither the publisher nor the author shall be liable for damages arising herefrom. The fact that an organization or Website is referred to in this work as a citation and/or a potential source of further information does not mean that the author or the publisher endorses the information the organization or Website may provide or recommendations it may make. Further, readers should be aware that Internet Websites listed in this work may have changed or disappeared between when this work was written and when it is read.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at www.wiley.com.

***Library of Congress Cataloging-in-Publication Data:***

Wong, K. Daniel.

Fundamentals of wireless communication engineering technologies / K. Daniel Wong. p. cm. – (Information and communication technology series ; 98)

Includes bibliographical references.

ISBN 978-0-470-56544-5

1. Wireless communication systems. 2. Wireless communication systems–Examinations–Study guides. I. Title.

TK5103.2.W59 2011

384.5–dc23 2011013591

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

To my parents and Almighty God

CONTENTS

FOREWORD **xix** PREFACE **xxi**

I PRELIMINARIES

**1 Introduction 3**

1.1 Notation / 4

1.2 Foundations / 4

1.2.1 Basic Circuits / 5

1.2.2 Capacitors and Inductors / 5

1.2.3 Circuit Analysis Fundamentals / 6

1.2.4 Voltage or Current as Signals / 7

1.2.5 Alternating Current / 9

1.2.6 Phasors / 10

1.2.7 Impedance / 11

1.2.8 Matched Loads / 11

1.3 Signals and Systems / 12

1.3.1 Impulse Response, Convolution, and Filtering / 12

1.3.2 Fourier Analysis / 14

1.3.3 Frequency-Domain Concepts / 17

1.3.4 Bandpass Signals and Related Notions / 19

1.3.5 Random Signals / 20

1.4 Signaling in Communications Systems / 27

1.4.1 Analog Modulation / 28

1.4.2 Digital Modulation / 29

1.4.3 Synchronization / 32

Exercises / 33

References / 33

**vii**

**viii** CONTENTS

II RADIO FREQUENCY, ANTENNAS, AND PROPAGATION

**2 Introduction to Radio Frequency, Antennas,**

**and Propagation 37**

2.1 Mathematical Preliminaries / 37

2.1.1 Multidimensional/Multivariable Analysis / 37

2.2 Electrostatics, Current, and Magnetostatics / 41

2.2.1 Electrostatics in Free Space / 41

2.2.2 Voltage / 42

2.2.3 Electrostatics in the Case of Dielectrics/Insulators / 43

2.2.4 Electrostatics Summary / 44

2.2.5 Currents / 44

2.2.6 Magnetostatics Introduction / 45

2.2.7 Magnetostatics in Free Space / 45

2.2.8 Magnetostatics in the Case of Magnetic Materials / 45

2.2.9 Symbols / 45

2.3 Time-Varying Situations, Electromagnetic Waves,

and Transmission Lines / 46

2.3.1 Maxwell’s Equations / 46

2.3.2 Electromagnetic Waves / 47

2.3.3 Transmission-Line Basics / 48

2.3.4 Standing-Wave Ratios / 51

2.3.5 S-Parameters / 55

2.4 Impedance / 56

2.5 Tests and Measurements / 57

2.5.1 Function Generators / 57

2.5.2 Measurement Instruments / 58

2.5.3 Mobile Phone Test Equipment / 61

Exercises / 62

References / 62

**3 Radio-Frequency Engineering 63**

3.1 Introduction and Preliminaries / 64

3.1.1 Superheterodyne Receiver / 64

3.1.2 RF—Handle with Care! / 66

3.1.3 RF Devices and Systems: Assumptions and Limitations / 67 3.1.4 Effect of Nonlinearities / 67

CONTENTS **ix**

3.2 Noise / 70

3.2.1 Types of Noise / 71

3.2.2 Modeling Thermal Noise / 71

3.2.3 Transferred Thermal Noise Power / 72

3.2.4 Equivalent Noise Source Models / 74

3.2.5 Noise Figure / 77

3.3 System Issues Related to Nonlinearity / 80

3.3.1 Gain Compression / 80

3.3.2 Size of Intermodulation Products / 81

3.3.3 Spur Free Dynamic Range / 83

3.4 Mixing and Related Issues / 85

3.5 Oscillators and Related Issues / 87

3.5.1 Phase Noise / 87

3.6 Amplifiers and Related Issues / 89

3.6.1 Low-Noise Amplifiers / 89

3.6.2 Power Amplifiers / 89

3.7 Other Components / 90

3.7.1 Directional Couplers / 90

3.7.2 Circulators / 91

Exercises / 91

References / 92

**4 Antennas 93**

4.1 Characterization / 94

4.1.1 Basic 3D Geometry / 94

4.1.2 Near Field and Far Field / 95

4.1.3 Polarization / 97

4.1.4 Radiation Intensity, Patterns, and Directivity / 98

4.1.5 Beam Area / 101

4.1.6 Antenna Gain / 101

4.1.7 Aperture / 102

4.1.8 Antenna Gain, Directivity, and Aperture / 102

4.1.9 Isotropic Radiators and EIRP / 103

4.1.10 Friis Formula for Receiver Signal Strength / 103

4.1.11 Bandwidth / 104

4.2 Examples / 105

4.2.1 Dipole Antennas / 105

4.2.2 Grounded Vertical Antennas / 106

**x** CONTENTS

4.2.3 Folded Dipoles / 106

4.2.4 Turnstiles / 107

4.2.5 Loop Antennas / 108

4.2.6 Parabolic Dish Antennas / 108

4.2.7 Mobile Device Antennas / 109

4.3 Antenna Arrays / 111

4.3.1 Linear Arrays / 112

4.3.2 Yagi-Uda Antennas / 114

4.3.3 Log-Periodic Dipole Arrays / 115

4.3.4 Base Station Antennas / 115

4.3.5 Newer Ideas for Using Multiple Antennas / 121

4.4 Practical Issues: Connecting to Antennas, Tuning, and so on / 122 4.4.1 Baluns / 122

4.4.2 Feeder Loss / 122

Exercises / 123

References / 124

**5 Propagation 125**

5.1 Electromagnetic Wave Propagation: Common Effects / 126 5.1.1 Path Loss / 126

5.1.2 Reflection and Refraction / 126

5.1.3 Diffraction / 128

5.1.4 Scattering / 131

5.2 Large-Scale Effects in Cellular Environments / 132

5.2.1 Ground Reflection Model / 133

5.2.2 Okumura Model / 135

5.2.3 Hata Model / 135

5.2.4 Lognormal Fading / 136

5.3 Small-Scale Effects in Cellular Environments / 137

5.3.1 Multipath Delay Spread / 137

5.3.2 Flat Fading / 138

5.3.3 Frequency-Selective Fading / 141

5.3.4 Time Variation: The Doppler Shift / 142

5.3.5 Diversity Combining / 145

5.4 Incorporating Fading Effects in the Link Budget / 148

Exercises / 150

Appendix: Ricean Fading Derivation / 151

References / 154

CONTENTS **xi**

III WIRELESS ACCESS TECHNOLOGIES

**6 Introduction to Wireless Access Technologies 159**

6.1 Review of Digital Signal Processing / 160

6.1.1 Impulse Response and Convolution / 160

6.1.2 Frequency Response / 161

6.1.3 Sampling: A Connection Between Discrete

and Continuous Time / 162

6.1.4 Fourier Analysis / 163

6.1.5 Autocorrelation and Power Spectrum / 164

6.1.6 Designing Digital Filters / 166

6.1.7 Statistical Signal Processing / 166

6.1.8 Orthogonality / 167

6.2 Digital Communications for Wireless Access Systems / 169 6.2.1 Coherent vs. Noncoherent / 169

6.2.2 QPSK and Its Variations / 169

6.2.3 Nonlinear Modulation: MSK / 172

6.3 The Cellular Concept / 173

6.3.1 Relating Frequency Reuse with *S/I* / 175

6.3.2 Capacity Issues / 176

6.4 Spread Spectrum / 177

6.4.1 PN Sequences / 178

6.4.2 Direct Sequence / 182

6.5 OFDM / 185

6.5.1 Spectral Shaping and Guard Subcarriers / 188

6.5.2 Peak-to-Average Power Ratio / 189

Exercises / 191

References / 192

**7 Component Technologies 193**

7.1 Medium Access Control / 193

7.1.1 Distributed-Control MAC Schemes / 194

7.1.2 Central Controlled Multiple Access Schemes / 196

7.1.3 Duplexing / 201

7.1.4 Beyond the Single Cell / 202

7.2 Handoff / 202

7.2.1 What Does It Cost? / 203

7.2.2 Types of Handoff / 203

**xii** CONTENTS

7.2.3 The Challenge of Making Handoff Decisions / 205

7.2.4 Example: Handoff in AMPS / 207

7.2.5 Other Examples / 207

7.3 Power Control / 208

7.3.1 The Near–Far Problem / 208

7.3.2 Uplink vs. Downlink / 208

7.3.3 Open- and Closed-Loop Power Control / 209

7.4 Error Correction Codes / 210

7.4.1 Block Codes / 212

7.4.2 Convolutional Codes / 214

7.4.3 Concatenation / 216

7.4.4 Turbo Codes / 216

7.4.5 LDPC Codes / 217

7.4.6 ARQ / 217

Exercises / 217

References / 218

**8 Examples of Air-Interface Standards: GSM, IS-95, WiFi 219**

8.1 GSM / 220

8.1.1 Access Control / 223

8.1.2 Handoffs and Power Control / 225

8.1.3 Physical Layer Aspects / 226

8.2 IS-95 CDMA / 226

8.2.1 Downlink Separation of Base Stations / 227

8.2.2 Single Base Station Downlink to Multiple

Mobile Stations / 228

8.2.3 Downlink Channels / 229

8.2.4 Uplink Separation of Mobile Stations / 231

8.2.5 Uplink Traffic Channel / 232

8.2.6 Separation of the Multipath / 232

8.2.7 Access Control / 232

8.2.8 Soft Handoffs and Power Control / 234

8.3 IEEE 802.11 WiFi / 235

8.3.1 LAN Concepts / 237

8.3.2 IEEE 802.11 MAC / 238

8.3.3 A Plethora of Physical Layers / 245

Exercises / 246

References / 246

CONTENTS **xiii**

**9 Recent Trends and Developments 249**

9.1 Third-Generation CDMA-Based Systems / 249

9.1.1 WCDMA / 250

9.1.2 cdma2000 / 251

9.1.3 Summary / 253

9.2 Emerging Technologies for Wireless Access / 253

9.2.1 Hybrid ARQ / 254

9.2.2 Multiple-Antenna Techniques / 256

9.3 HSPA and HRPD / 258

9.3.1 HSDPA / 259

9.3.2 HSUPA / 261

9.3.3 1×EV-DO / 261

9.3.4 Continuing Enhancements / 261

9.4 IEEE 802.16 WiMAX / 262

9.4.1 Use of HARQ / 263

9.4.2 Use of OFDMA / 263

9.4.3 Other Aspects / 269

9.5 LTE / 270

9.5.1 Use of HARQ / 270

9.5.2 Use of OFDMA on Downlink / 271

9.5.3 SC-FDMA or DFTS-OFDM on Uplink / 271

9.5.4 Other Aspects / 272

9.6 What’s Next? / 273

Exercises / 273

References / 274

IV NETWORK AND SERVICE ARCHITECTURES

**10 Introduction to Network and Service Architectures 277**

10.1 Review of Fundamental Networking Concepts / 278

10.1.1 Layering / 278

10.1.2 Packet Switching vs. Circuit Switching / 281

10.1.3 Reliability / 283

10.2 Architectures / 285

10.2.1 Network Sizes / 285

10.2.2 Core, Distribution, and Access / 285

10.2.3 Topology / 286

**xiv** CONTENTS

10.2.4 Communication Paradigm / 286

10.2.5 Stupid vs. Intelligent Networks / 287

10.2.6 Layering Revisited / 287

10.2.7 Network Convergence / 288

10.3 IP Networking / 290

10.3.1 Features of IP / 290

10.3.2 Transport Protocols / 292

10.3.3 Related Protocols and Systems / 294

10.3.4 Style / 295

10.3.5 Interactions with Lower Layers / 295

10.3.6 IPv6 / 296

10.4 Teletraffic Analysis / 301

10.4.1 Roots in the Old Phone Network / 301

10.4.2 Queuing Theory Perspective / 303

Exercises / 305

References / 306

**11 GSM and IP: Ingredients of Convergence 307**

11.1 GSM / 308

11.1.1 Some Preliminary Concepts / 308

11.1.2 Network Elements / 309

11.1.3 Procedures / 311

11.1.4 Location Management / 311

11.2 VoIP / 315

11.2.1 Other Parts of the VoIP Solution / 317

11.2.2 Session Control: SIP / 317

11.3 QoS / 323

11.3.1 Frameworks / 324

11.3.2 QoS Mechanisms / 326

11.3.3 Wireless QoS / 330

Exercises / 331

References / 332

**12 Toward an All-IP Core Network 333**

12.1 Making IP Work with Wireless / 333

12.1.1 Mobile IP / 334

12.1.2 Header Compression / 339

CONTENTS **xv**

12.2 GPRS / 341

12.2.1 GPRS Attach and PDP Context Activation / 344

12.2.2 GPRS Mobility Management States / 345

12.3 Evolution from GSM to UMTS up to the Introduction of IMS / 346 12.3.1 First UMTS: Release ’99 / 346

12.3.2 From Release ’99 to Release 4 / 348

12.3.3 From Release 4 to Release 5 / 349

12.3.4 From Release 5 to Release 6 / 351

12.3.5 From Release 6 to Release 7 / 351

12.3.6 From Release 7 to Release 8: LTE / 351

12.3.7 Evolved Packet System of LTE / 352

12.4 IP Multimedia Subsystem / 354

12.4.1 Network Functions / 355

12.4.2 Procedures / 359

12.5 Other Networks / 362

12.5.1 cdma2000 / 362

12.5.2 WiMAX / 364

Exercises / 365

References / 365

**13 Service Architectures, Alternative Architectures,**

**and Looking Ahead 367**

13.1 Services / 367

13.1.1 Examples of Services / 369

13.2 Service Architectures / 371

13.2.1 Examples: Presence / 372

13.2.2 Examples: Messaging / 372

13.2.3 Examples: Location-Based Services / 372

13.2.4 Examples: MBMS / 373

13.2.5 The Rise of the Intelligent Network / 373

13.2.6 Open Service Access / 375

13.2.7 Open Mobile Alliance / 376

13.2.8 Services and IMS / 377

13.3 Mobile Ad Hoc Networks / 379

13.3.1 Example: AODV / 380

13.4 Mesh, Sensor, and Vehicular Networks / 384

13.4.1 Mesh Networks / 385

13.4.2 Sensor Networks / 387

**xvi** CONTENTS

13.4.3 Vehicular Networks / 388

Exercises / 389

References / 390

V MISCELLANEOUS TOPICS

**14 Network Management 393**

14.1 Requirements and Concepts / 393

14.2 Network Management Models / 394

14.3 SNMP / 397

14.3.1 Messages / 398

14.3.2 Managed Objects / 400

14.3.3 MIBs / 402

14.3.4 Security / 409

14.3.5 Traps / 409

14.3.6 Remote Monitoring / 410

14.3.7 Other Issues / 411

14.3.8 Suggested Activities / 412

Exercises / 412

References / 412

**15 Security 415**

15.1 Basic Concepts / 415

15.1.1 Attacks / 417

15.1.2 Defenses / 418

15.2 Cryptography / 419

15.2.1 Symmetric Schemes / 419

15.2.2 Asymmetric Schemes / 420

15.2.3 Key Distribution / 420

15.2.4 Algorithms / 421

15.3 Network Security Protocols / 422

15.3.1 IPSec / 423

15.3.2 Access Control and AAA / 429

15.4 Wireless Security / 432

15.4.1 Cellular Systems / 432

15.4.2 802.11 WLAN / 436

CONTENTS **xvii**

15.4.3 Mobile IP Security / 440

Exercises / 441

References / 442

**16 Facilities Infrastructure 443**

16.1 Communications Towers / 444

16.1.1 Protecting Planes / 446

16.1.2 Other Considerations / 448

16.2 Power Supplies and Protection / 450

16.2.1 Power Consumption / 450

16.2.2 Electrical Protection / 453

16.3 Additional Topics / 462

16.3.1 RF Cables / 462

16.3.2 Building Automation and Control Systems / 463

16.3.3 Physical Security / 463

Exercises / 464

References / 465

**17 Agreements, Standards, Policies, and Regulations 467**

17.1 Agreements / 468

17.1.1 Service-Level Agreements / 468

17.1.2 Roaming Agreements / 469

17.2 Standards / 469

17.2.1 IEEE / 470

17.2.2 Example: Standards Development—IEEE 802.16 / 471 17.2.3 ITU / 471

17.2.4 IETF / 475

17.2.5 3GPP / 475

17.2.6 Revisions, Amendments, Corrections, and Changes / 475 17.2.7 Intellectual Property / 478

17.3 Policies / 478

17.4 Regulations / 479

17.4.1 Licensed vs. Unlicensed Spectrum / 480

17.4.2 Example: Regulatory Process for Ultrawideband / 481

Exercises / 484

References / 484

**xviii** CONTENTS

EXERCISE SOLUTIONS **487** APPENDIX A: SOME FORMULAS AND IDENTITIES **497** APPENDIX B: WCET GLOSSARY EQUATION INDEX **499** APPENDIX C: WCET EXAM TIPS **501** APPENDIX D: SYMBOLS **503** APPENDIX E: ACRONYMS **509** INDEX **519**

FOREWORD

Wireless communications is one of the most advanced and rapidly advancing tech nologies of our time. The modern wireless era has produced an array of technologies, such as mobile phones and WiFi networks, of tremendous economic and social value and almost ubiquitous market penetration. These developments have in turn created a substantial demand for engineers who understand the basic principles underlying wireless technologies, and who can help move the field forward to meet the even greater demands for wireless services and capacity expected in the future. Such an understanding requires knowledge of several distinct fields upon which wireless tech nologies are based: radio frequency physics and devices; communication systems engineering; and communication network architecture.

This book, by a leading advocate of the IEEE Communications Society’s Wireless Communication Engineering Technologies certification program, offers an excellent survey of this very broad set of fundamentals. It further provides a review of basic foundational subjects, such as circuits, signals and systems, as well as coverage of several important overlying topics, such network management, security, and regulatory issues. This combination of breadth and depth of coverage allows the book to serve both as a complete course for students and practicing engineers, and as an entrée to the field for those wishing to undertake more advanced study or do research in a particular aspect of the field. Thus, *Fundamentals of Wireless Communication Engineering Technologies* is a very welcome addition to the pedagogical literature in this important field of technology.

H. VINCENT POOR

*Princeton, New Jersey*

**xix**

PREFACE

This book presents a broad survey of the fundamentals of wireless communication engineering technologies, spanning the field from radio frequency, antennas, and propagation, to wireless access technologies, to network and service architectures, to other topics, such as network management and security, agreements, standards, policies and regulations, and facilities infrastructure.

Every author has to answer two major questions: (1) What is the scope of coverage of the book, in terms of breadth of topics and depth of discussion of each topic, focus and perspective, and assumptions of prior knowledge of the readers? and (2) Who are the intended readers of the book? I am honored to have been a member of the Practice Analysis Task Force convened by IEEE Communications Society to draft the syllabus and examination specifications of IEEE Communication Society’s Wireless Communication Engineering Technologies (WCET) certification program. The scope of coverage of this book has been strongly influenced by the syllabus of the WCET program.

This book is designed to be helpful to three main groups of readers:

• Readers who would like to understand a broad range of topics in practical wire less communications engineering, from fundamentals and theory to practical aspects. For example, wireless engineers with a few years of experience in wire less might find themselves deeply involved with one or two aspects of wireless systems, but not actively keeping up-to-date with other aspects of wireless sys tems. This book might help such engineers to see how their work fits into the bigger picture, and how the specific parts of the overall system on which they work relate to other parts.

• Electrical engineering or computer science students with an interest in wireless communications, who might be interested to see how the seemingly dry, abstract theory they learn in class is actually applied in real-world wireless systems.

• Readers who are considering taking the WCET exam to become Wireless Certified Professionals. This group could include readers who are not sure if they would take the exam but might decide after reviewing the scope of coverage of the exam.

I hope this book can be a helpful resource for all three groups of readers. For the third group of readers, those with an interest in the WCET exam, several appendices

**xxi**

**xxii** PREFACE

may be useful, including a list of where various formulas from the WCET glossary are discussed in the text (Appendix B), and a few exam tips (Appendix C). However, the rest of the book has been written so that it can be read beneficially by any of the aforementioned groups of readers.

The book is divided into four main sections, three of which cover important areas in wireless systems: (1) radio frequency, antennas, and propagation; (2) wireless access technologies; and (3) network and service architectures. The fourth main section includes the remaining topics. The first three main parts of the book each begins with an introductory chapter that provides essential foundational material, followed by three chapters that go more deeply into specific topics. I have strived to arrange the materials so that the three chapters that go deeper into specific topics build on what is covered in the introductory chapter for that area. This is designed to help students who are new to an area, or not so familiar with it, to be able to go far on their own in self-study, through careful reading first of the introductory chapter, and then of the subsequent chapters. Numerous cross-references are sprinkled throughout the text, for example, so that students who are reading about a topic that relies on some foundational knowledge can see where the foundational knowledge is covered in the relevant introductory chapter. Also, references might be from the relevant introductory chapter to places where specific topics are covered in more detail, which may help motivate students to understand the material in the introductory chapter, as they can see how it is applied later.

The amount of technical knowledge that a wireless engineer “should know” is so broad that it is practically impossible to cover everything in one book, much less to cover everything at the depth that might satisfy every reader. In this book we have tried to select important topics that can be pulled together into coherent and engaging stories and development threads, rather than simply to present a succession of topics. For example, the results of some of the examples are used in later sections or chapters of the book. We also develop various notions related to autocorrelation and orthogonality with an eye to how the concepts might be needed later to help explain the fundamentals of CDMA.

Thanks to Diana Gialo, Simone Taylor, Sanchari Sil, Angioline Loredo, Michael Christian, and George Telecki of Wiley for their editorial help and guidance during the preparation of the manuscript, and to series editors Dr. Vincent Lau and Dr. T. Russell Hsing for their support and helpful comments. Thanks are also due to Dr. Wee Lum Tan, Dr. Toong Khuan Chan, Dr. Choi Look Law, Dr. Yuen Chau, HS Wong, Lian Pin Tee, Ir. Imran Mohd Ibrahim, and Jimson Tseng for their insightful and helpful reviews of some chapters in the book.

There is a web site for this book at http://www.danielwireless.com/wcet, where various supplementary materials, including a list of corrections and updates, will be posted.

K. DANIEL WONG

PH.D. (STANFORD), CCNA, CCNP (CISCO), WCP (IEEE)

*Palo Alto, California*

I

PRELIMINARIES

1

INTRODUCTION

In this chapter we provide a brief and concise review of foundational topics that are of broad interest and usefulness in wireless communication engineering technologies. The notation used throughout is introduced in Section 1.1, and the basics of electrical circuits and signals are reviewed in Section 1.2, including fundamentals of circuit analysis, voltage or current as signals, alternating current, phasors, impedance, and matched loads. This provides a basis for our review of signals and systems in Sec tion 1.3, which includes properties of linear time-invariant systems, Fourier analysis and frequency-domain concepts, representations of bandpass signals, and modeling of random signals. Then in Section 1.4, we focus on signals and systems concepts specifically for communications systems. The reader is expected to have come across much of the material in this chapter in a typical undergraduate electrical engineering program. Therefore, this chapter is written in review form; it is not meant for a student who is encountering all this material for the first time.

Similarly, reviews of foundational topics are provided in Chapters 2, 6, and 10 for the following areas:

• *Chapter 2:* review of selected topics in electromagnetics, transmission lines, and testing, as a foundation for radio frequency (RF), antennas, and propagation • *Chapter 6:* review of selected topics in digital signal processing, digital com

muncations over wireless links, the cellular concept, spread spectrum, and othogonal frequency-division multiflexing (OFDM), as a foundation for wireless access technologies

*Fundamentals of Wireless Communication Engineering Technologies*, First Edition. K. Daniel Wong. © 2012 John Wiley & Sons, Inc. Published 2012 by John Wiley & Sons, Inc.

**3**

**4** INTRODUCTION

• *Chapter 10:* review of selected topics in fundamental networking concepts, Internet protocol (IP) networking, and teletraffic analysis, as a foundation for network and service architectures

Compared to the present chapter, the topics in Chapters 2, 6, and 10 are generally more specific to particular areas. Also, we selectively develop some of the topics in those chapters in more detail than we do in this chapter.

**1.1 NOTATION**

In this section we discuss the conventions we use in this book for mathematical notation. A list of symbols is provided in Appendix D.

*R* and *C* represent the real and complex numbers, respectively. Membership in a set is represented by ∈ (e.g., *x* ∈ *R* means that *x* is a real number). For *x* ∈ *C*, we write {*x*} and {*x*} for the real and imaginary parts of *x*, respectively.

log represents base-10 logarithms unless otherwise indicated (e.g., log2 for base-2 logarithms), or where an expression is valid for all bases.

Scalars, which may be real or even complex valued, are generally represented by italic type (e.g., *x*, *y*), whereas vectors and matrices will be represented by bold type (e.g., **G**, **H**). We represent a complex conjugate of a complex number, say an impedance *Z*, by *Z*∗. We represent the magnitude of a complex number *x* by |*x*|. Thus, |*x*|2 = *xx*∗.

For *x* ∈ *R*, *x* is the largest integer *n* such that *n<x*. For example, 5*.*67 = 5 and −1*.*2 = −2.

If **G** is a matrix, **GT** represents its transpose.

When we refer to a matrix, vector, or polynomial as being *over*something (e.g., *over the integers*), we mean that the components (or coefficients, in the case of polynomials) are numbers or objects of that sort.

If *x*(*t*) is a random signal, we use *< x*(*t*) *>* to refer to the time average and *x*(*t*) to refer to the ensemble average.

**1.2 FOUNDATIONS**

Interconnections of electrical elements (resistors, capacitors, inductors, switches, volt age and current sources) are often called a *circuit*. Sometimes, the term *network* is used if we want “circuit” to apply only to the more specific case of where there is a closed loop for current flow. In Section 1.2.1 we review briefly this type of electri cal network or circuit. Note that this use of “network” should not be confused with the very popular usage in the fields of computer science and telecommunications, where we refer to computer networks and telecommunications networks (see Chap ters 9 to 12 for further discussion). In Chapter 2 we will see how transmission lines (Section 2.3.3) can be modeled as circuit elements and so can be part of electrical networks and circuits.

FOUNDATIONS **5**

In *electronic* networks and circuits, we also have components with *gain* and/or *directionality*, such as semiconductor devices, which are known as *active* components (as opposed to *passive* components, which have neither gain nor directionality). These are outside the scope of this book, except for our discussion on RF engineering in Chapter 3. Even there, we don’t discuss the physics of the devices or compare different device technologies. Instead, we take a “signals and systems” perspective on RF, and consider effects such as noise and the implications of nonlinearities in the active components.

**1.2.1 Basic Circuits**

Charge, *Q*, is quantified in coulombs. Current is charge in motion: *I* = *dQdt* amperes (1.1)

The direction of current flow can be indicated by an arrow next to a wire. For conve nience, *I* can take a negative value if current is flowing in the direction opposite from that indicated by the arrow.

Voltage is the difference in electric potential:

*V* = *RI* volts (1.2)

Like current, there is a direction associated with voltage. It is typically denoted by + and −. + is at higher potential than −, and voltage drops going from + to −. For convenience, *V* can take a negative value if a voltage drop is in the direction opposite from that indicated by + and −

• Power:

*P* = *V*2

*R* , *P* = *I*2*R* watts (1.3)

• Resistors in series:

*R* = *R*1 + *R*2 +···+ *Rn* (1.4)

• Resistors in parallel:

*R* = *R*1*R*2 ··· *Rn*

*R*1 + *R*2 +···+ *Rn*(1.5)

**1.2.2 Capacitors and Inductors**

A capacitor may be conceived of in the form of two parallel plates. For a capacitor with capacitance *C* farads, a voltage *V* applied across its plates results in charges +*Q* and −*Q* accumulating on the two plates.

*Q* = *CV* (1.6)

**6** INTRODUCTION

*I* = *dQdt* = *CdVdt* (1.7)

A capacitor acts as an open circuit under direct-current (dc) conditions. • Capacitors in series:

*C* = *C*1*C*2 ···*Cn*

*C*1 + *C*2 +···+ *Cn*(1.8)

• Capacitors in parallel:

*C* = *C*1 + *C*2 +···+ *Cn* (1.9)

An inductor is often in the form of a coil of wire. For an inductor with inductance *L* henries, a change in current of *dI/dt* induces a voltage *V* across the inductor:

*V* = *LdIdt* (1.10)

An inductor acts as a short circuit under dc conditions.

• Inductors in series:

*L* = *L*1 + *L*2 +···+ *Ln* (1.11)

• Inductors in parallel:

*L* = *L*1*L*2 ···*Ln*

*L*1 + *L*2 +···+ *Ln*(1.12)

As hinted at by (1.3), an ideal capacitor or ideal inductor has no resistance and does not dissipate any power as heat. However, a practical model for a real inductor has an ideal resistor in series with an ideal inductor, and they are both in parallel with an ideal capacitor.

**1.2.3 Circuit Analysis Fundamentals**

A *node* in a circuit is any place where two or more circuit elements are connected. A *complete loop* or *closed path* is a continuous path through a circuit that begins and ends at the same node.

*Kirchhoff’s Current Law. The sum of all the currents entering is zero.* This requires at least one current to have a negative sign if one or more of the others is positive. Alternatively, we say that the sum of all the current entering a node is equal to the sum of all the current leaving a node.

FOUNDATIONS **7**

*Kirchhoff’s Voltage Law. The sum of all the voltage drops around any complete loop (or closed path) is zero.* This requires at least one voltage drop to have a negative sign if one or more of the others is positive.

***1.2.3.1 Equivalent Circuits*** Often, a subcircuit is connected to the rest of the circuit through a pair of terminals, and we are interested to know what the voltage and current are across these terminals, not how the subcircuit is actually implemented. Norton and Thevenin equivalent circuits can be used for this purpose, for any circuit ´ comprising linear elements. A *Th´evenin equivalent circuit* comprises a single voltage source, *VT* , in series with a single resistor, *RT* . A *Norton equivalent circuit* comprises a single current source, *IN*, in parallel with a single resistor, *RN*. A Thevenin equivalent ´ circuit can be converted to a Norton equivalent circuit, or vice versa, by a simple source transformation.

**1.2.4 Voltage or Current as Signals**

A voltage or current can be interpreted as a signal (e.g., for communications purposes). We usually write *t* explicitly to emphasize that it is a function of *t* [e.g., *v*(*t*) or *i*(*t*) for a voltage signal or current signal, respectively].

If *x*(*t*) is a signal, we say that *x*(*t*) is

• An *energy signal* if

*x*2(*t*) *dt <* ∞ (1.13)

0 *<*

• A *power signal* if

∞ −∞

0 *<* lim *T*→∞

1

*T*

∞ −∞

*x*2(*t*) *dt <* ∞ (1.14)

A *periodic* signal is a signal for which a *T* ∈ *R* can be found such that *x*(*t*) = *x*(*t* + *T* ) for −∞ *<t<* ∞ (1.15)

and the smallest such *T* is called the *period* of the signal.

The duration of a signal is the time interval from when its begins to be nonnegligible to when its stops being nonnegligible.† Thus, a signal can be of finite duration or of infinite duration.

*Sinusoidal Signals.* Any sinusoid that is a function of a single variable (say, the time variable, *t*; later, in Section 2.1.1.4, we see sinusoids that are functions of both

†We say nonnegligible rather than nonzero to exclude trivial blips outside the duration of the signal.

**8** INTRODUCTION

temporal and spatial variables) can be written as

*A* cos(*ωt* + *φ*) = *A* cos(2*πft* + *φ*) = *A* sin(2*πft* + *φ* + *π/*2) = *A*∠*φ* (1.16)

where *A* is amplitude (*A* ∈ *R*), *ω* is *angular frequency* (radians/second), *f* is *fre quency* (cycles/second, i.e., hertz or s−1), *φ* is *phase angle*, and where the last equality shows that the shorthand notation *A*∠*φ* can be used when *f* and the sinusoidal reference time are known implicitly. The period *T* is

*T* = 1*f* = 2*πω* (1.17)

*Continuous-Wave Modulation Signals.* A continuous-wave modulation signal is a sinusoidal signal that is *modulated* (changed) in a certain way based on the information being communicated. Most communications signals are based on continuous-wave modulation, and we expand on this important topic in Section 1.4.

*Special Signals.* A fundamental building block in continuous-time representation of digital signals is the rectangular pulse signal, a rectangular function given by

(*t*) =

1 for |*t*| ≤ 1*/*2

0 for |*t*| *>* 1*/*2 (1.18)

The triangle signal is also commonly used, but not as frequently. It is denoted by

(*t*) =

1 − |*t*| for |*t*| ≤ 1

0 for |*t*| *>* 1 (1.19)

(*t*) and (*t*) are shown in Figure 1.1.

The sinc signal is given by

(sin *πt*)*/πt* for |*t*| =*/* 0

sinc(*t*) =

1 for *t* = 0 (1.20)

Although it may be described informally as (sin *πt*)*/πt*, (sin *πt*)*/πt* is actually unde fined at *t* = 0, whereas sinc(*t*) is 1 at *t* = 0. The sinc function is commonly seen in communications because it is the Fourier transform of the rectangular pulse signal. Note that in some fields (e.g., mathematics), sinc(*t*) may be defined as (sin *t*)*/t*, but

1

0

2 − 1

Π(*t*)

*t*

1

0

21

Λ(*t*)

−1 1 *t*

**FIGURE 1.1** (*t*) and (*t*) functions.

*x*) sinc(

1

0.8

0.6

0.4

0.2

0

−0.2 −0.4

FOUNDATIONS **9**

−4 −3 −2 −1 0 1 2 3 4

*x*

**FIGURE 1.2** Sinc function.

here we stick with our definition, which is standard for communications and signal processing. The sinc function is shown in Figure 1.2.

*Decibels.* It is sometimes convenient to use a log scale when the range of amplitudes can vary by many orders of magnitude, such as in communications systems where the signals have amplitudes and powers that can vary by many orders of magnitude. The standard way to use a log scale in this case is by the use of decibels, defined for any signal voltage or current signal *x*(*t*) as

10 log *x*2(*t*) = 20 log *x*(*t*) (1.21)

If the signal *s*(*t*) is known to be a power rather than a voltage or current, we don’t have to convert it to a power, so we just take 10 log *s*(*t*). If the power quantity is in watts, it is sometimes written as dBW, whereas if it is in milliwatts, it is written as dBm. This can avoid ambiguity in cases where we just specify a dimensionless quantity *A*, in decibels, as 10 log *A*.

**1.2.5 Alternating Current**

With alternating current (ac) the voltage sources or current sources generate time varying signals. Then (1.3) refers only to the *instantaneous power*, which depends on the instantaneous value of the signal. It is often also helpful, perhaps more so, to consider the *average power*. Let *v*(*t*) = *V*0 cos 2*πft*, where *V*0 is the maximum voltage (and correspondingly, let *I*0 be the maximum current), then the average power *P*av is

*P*av = *V*20

2*R*, *P*av = *I*20*R*2 (1.22)

**10** INTRODUCTION

Equation (1.22) can be obtained either by averaging the instantaneous power directly over one cycle, or through the concept of*rms voltage* and *rms current*. The rms voltage is defined for any periodic signal (not just sinusoidally periodic) as

*V*rms =

1

*T*

*T* 0

*v*2(*t*) *dt* (1.23)

Then we have (again, for any periodic signal, not just sinusoidally periodic) *P*av = *V*2rms

*R* , *P*av = *I*2rms*R* (1.24)

which looks similar to (1.3). For sinusoidally time-varying signals, we have further, *V*rms = *V*0

√2, *I*rms = *I*0

√2 (1.25)

**1.2.6 Phasors**

When working with sinusoidal signals, it is often convenient to work with the *phasor representation* of the signals. Of the three quantities amplitude, phase, and frequency, the phasor representation includes only the amplitude and phase; the frequency is implicit.

Starting from our sinusoid in (1.16) and applying Euler’s identity (A.1), we obtain

*A* cos(2*πft* + *φ*) = *A*

*ej*(2*πft*+*φ*) = *Aej*(2*πft*+*φ*) (1.26)

We just drop the *ej*2*πft* and omit mentioning that we need to take the real part, and we have a phasor,

*Aejφ* (1.27)

Alternatively, we can write the equivalent,

*A*(cos *φ* + *j* sin *φ*) (1.28)

which is also called a phasor. In either case, we see that a phasor is a complex number representation of the original sinusoid, and that it is easy to get back the original sinusoid by multiplying by *ej*2*πft* and taking the real part. A hint of the power and convenience of working with phasor representations can be seen by considering differentiation and integration of phasors. Differentiation and integration with respect to *t* are easily seen to be simple multiplication and division, respectively, by *j*2*πf* .

*Rotating Phasors.* Sometimes it helps to think of a phasor not just as a static point in the complex plane, but as a rotating entity, where the rotation is at frequency *f* revolutions (around the complex plane) per second, or *w* radians per second. This is consistent with the *ej*2*πft* term that is implicit in phasors. The direction of rotation is as illustrated in Figure 1.3.

direction of rotation

FOUNDATIONS **11** *j*θ *Ae*

for positive

*fc*

resultant

*j*θ *Ae* direction of rotation

*fc*

for negative

θ

(a) (b) (c)

**FIGURE 1.3** (a) Phasor in the complex plane; (b) rotating phasors and their direction of rotation; (c) vector addition of phasors.

*Expressing Familiar Relationships in Terms of Phasors.* Returning to familiar relationships such as (1.2) or (1.3), we find no difference if *v*(*t*), *i*(*t*) are in phasor representation; however, for capacitors and inductors we have

*I* = *j*2*πfCV* and *V* = *j*2*πfLI* (1.29)

Thus, if we think in terms of rotating phasors, then from (1.29) we see that with a capacitor, *I* rotates 90◦ ahead of *V*, so it *leads V* (and *V lags I*), whereas with an inductor, *V* leads *I* (*I* lags *V*).

Meanwhile, Kirchhoff’s laws take the same form for phasors as they do for non phasors, so they can continue to be used. Thevenin and Norton equivalent circuits can ´ also be used, generalized to work with impedance, a concept that we discuss next.

**1.2.7 Impedance**

From (1.29) it can be seen that in phasor representation, resistance, inductance, and capacitance all have the same form:

*V* = *ZI* (1.30)

Thus, the concept of *impedance*, *Z*, emerges, where *Z* is *R* for resistance, *j*2*πfL* for inductance, and 1*/j*2*πfC* for capacitance, and *Z* is considered to be in ohms. The complex part of *Z* is also known as *reactance*.

Impedance is a very useful concept. For example, Thevenin’s and Norton’s equiv- ´ alent circuits work in the same way with phasors, except that impedance is substituted for resistance.

**1.2.8 Matched Loads**

For a linear circuit represented by a Thevenin equivalent voltage ´ *VT* and Thevenin ´ equivalent impedance *ZT* , the maximum power is delivered to a load *ZL* when

*ZL* = *Z*∗*T* (1.31)

**12** INTRODUCTION

(NB: It is the complex conjugate of *ZT* , not *ZT* itself, in the equation.) This result can be obtained by writing the expression for power in terms of *ZL* and *ZT* , taking partial derivatives with respect to the load resistance and load reactance, and setting both to 0.

**1.3 SIGNALS AND SYSTEMS**

Similarly, suppose that we have a system (e.g., a circuit) that takes an input *x*(*t*) and produces an output *y*(*t*). Let −→ represent the operation of the system [e.g., *x*(*t*) −→ *y*(*t*)]. Suppose that we have two different inputs, *x*1(*t*) and *x*2(*t*), such that *x*1(*t*) −→ *y*1(*t*) and *x*2(*t*) −→ *y*2(*t*). Let *a*1 and *a*2 be any two scalars. The system is *linear* if and only if

*a*1*x*1(*t*) + *a*2*x*2(*t*) −→ *a*1*y*1(*t*) + *a*2*y*2(*t*) (1.32)

The phenomenon represented by (1.32) can be interpreted as the *superposition* prop erty of linear systems. For example, given knowledge of the response of the system to various sinusoidal inputs, we then know the response of the system to any linear combination of sinusoidal signals. This makes Fourier analysis (Section 1.3.2) very useful.

A system is *time-invariant* if and only if

*x*(*t* − *t*0) −→ *y*(*t* − *t*0) (1.33)

Systems that are both linear and time invariant are known as *LTI* (linear time invariant) *systems*.

A system is *stable* if bounded input signals result in bounded output signals. A system is *causal* if any output does not come before the corresponding input.

**1.3.1 Impulse Response, Convolution, and Filtering**

An impulse (or unit impulse) signal is defined as

1*, t* = 0

0*, t/*= 0 (1.34)

and also where

*δ*(*t*) =

∞ −∞

*δ*(*t*) = 1 (1.35)

Strictly speaking, *δ*(*t*) is not a function, but to be mathematically rigorous requires measure theory or the theory of generalized functions. *δ*(*t*) could also be thought of as

*T*→∞ *T* (*tT* ) (1.36)

lim

LTI system

SIGNALS AND SYSTEMS **13**

Time domain output

*x*(*t*) *h*(*t*) *y*(*t*) = *h*(*t*)∗ *x*(*t*)

*X* ( *f* ) *H*( *f* ) *Y*( *f* ) = *H*( *f* )*X* ( *f* )

Frequency domain

**FIGURE 1.4** Mathematical model of an LTI system.

Thus, we often view it as the limiting case of a narrower and narrower pulse whose area is 1.

All LTI systems can be characterized by their *impulse response*. The impulse response, *h*(*t*), is the output when the input is an impulse signal; that is,

*δ*(*t*) −→ *h*(*t*) (1.37)

Convolution: The output of an LTI system with impulse response *h*(*t*), given an input *x*(*t*), is

*y*(*t*) = *h*(*t*) ∗ *x*(*t*) =

*τ*=∞ *τ*=−∞

*x*(*τ*)*h*(*t* − *τ*) *dτ* =

*τ*=∞ *τ*=−∞

*h*(*τ*)*x*(*t* − *τ*) *dτ* (1.38)

This is shown as the output of the LTI system in Figure 1.4.

With (1.38) in mind, whenever we put a signal *x*(*t*) into an LTI system, we can think in terms of the system as *filtering* the input to produce the output *y*(*t*), and *h*(*t*) may be described as the impulse response of the filter. Although the term *filter* is used in the RF and baseband parts of wireless transmitters and receivers, *h*(*t*) can equally well represent the impulse response of a communications channel (e.g., a wire, or wireless link), in which case we may then call it the *channel response* or simply the *channel*.

***1.3.1.1 Autocorrelation*** It is sometimes useful to quantify the similarity of a signal at one point in time with itself at some other point in time. Autocorrelation is a way to do this. If *x*(*t*) is a complex-valued energy signal (a real-valued signal is a special case of a complex-valued signal, where the imaginary part is identically zero, and the complex conjugate of the signal is equal to the signal itself), we define the autocorrelation function, *Rxx*(*τ*), as

*Rxx*(*τ*) =

∞ −∞

*x*(*t*)*x*∗(*t* + *τ*) *dt* for −∞ *<τ<* ∞ (1.39)

For a complex-valued periodic power signal with period *T*0,

*Rxx*(*τ*) = 1*T*0 *T*0*/*2 −*T*0*/*2

*x*(*t*)*x*∗(*t* + *τ*) *dt* for −∞ *<τ<* ∞ (1.40)

**14** INTRODUCTION

whereas for a complex-valued power signal, in general,

*Rxx*(*τ*) = lim *T*→∞

1

*T*

*T/*2 −*T/*2

*x*(*t*)*x*∗(*t* + *τ*) *dt* for −∞ *<τ<* ∞ (1.41)

**1.3.2 Fourier Analysis**

*Fourier analysis* refers to a collection of related techniques where:

• A signal can be broken down into sinusoidal components *(analysis)* • A signal can be constructed from constituent sinusoidal components *(synthesis)*

This is very useful in the study of linear systems because the effects of such a system on a large class of signals can be studied by considering the effects of the system on sinusoidal inputs using the superposition principle. (NB: The term *analysis* here can be used to refer either to just the breaking down of a signal into sinusoidal components, or in the larger sense to refer to the entire collection of these related techniques.)

Various Fourier *transforms* are used in analysis, and *inverse transforms* are used in synthesis, depending on the types of signals involved. For most practical pur poses, there is a one-to-one relationship between a time-domain signal and its Fourier transform, and thus we can think of the Fourier transform of a signal as being a dif ferent *representation* of the signal. We usually think of there being two domains, the *time domain* and the *frequency domain*. The (forward) transform typically transforms a *time-domain representation* of a signal into a *frequency-domain representation*, whereas the inverse transform transforms a frequency-domain representation of a signal into a time-domain representation.

***1.3.2.1 (Continuous) Fourier Transform*** The (continuous) Fourier trans form of a signal *x*(*t*) is given by

*X*(*f* ) =

∞ −∞

*x*(*t*)*e*−*j*2*πft dt* (1.42)

where *j* = √−1, and the inverse Fourier transform is given by

*x*(*t*) =

∞ −∞

*X*(*f* )*ej*2*πft df* (1.43)

Table 1.1 gives some basic Fourier transforms.

***1.3.2.2 Fourier Series*** For periodic signals *x*(*t*) with period *T* , the Fourier series (exponential form) coefficients are the set {*cn*}, where *n* ranges over all the integers,

SIGNALS AND SYSTEMS **15**

**TABLE 1.1 Fourier Transform Pairs*a***

Time Domain, *x*(*t*) Frequency Domain, *X*(*f* )

*δ*(*t*) 1

1 *δ*(*f* )

*δ*(*t* − *t*0) *e*−*j*2*πft*0

*e*±*j*2*πf*0*t δ*(*f* ∓ *f*0)

cos 2*πf*0*t*12[*δ*(*f* − *f*0) + *δ*(*f* + *f*0)] sin 2*πf*0*t*12*j*[*δ*(*f* − *f*0) − *δ*(*f* + *f*0)]

*u*(*t*) =

1 for *t >* 0 0 for *t <* 0

2*δ*(*f* ) +1

1

*j*2*πf*

*e*−*atu*(*t*), *a >* 01 *a* + *j*2*πf*

*te*−*atu*(*t*), *a >* 01 (*a* + *j*2*πf* )2

*e*−*a*|*t*|, *a >* 02*a*

*t*

*a*2 + (2*πf* )2 *T* sinc *fT*

*B* sinc *Bt fB*  
 *T*

*t T*

*T* sinc2*fT*

∞

*k*=−∞

*δ*(*t* − *kT* )1*T* ∞

*δ*

*n*=−∞

*f* − *nT*

*a* (*t*) and (*t*) are the rectangle and triangle functions defined in Section 1.2.4. ∞*k*=−∞ *δ*(*t* − *kT* ) is also known as an impulse train.

and *cn* is given by

*cn* = 1*T* *T/*2 −*T/*2

*x*(*t*)*e*−*j*2*πf*0*nt dt* (1.44)

where *f*0 = 1*/T* , and the Fourier series representation of *x*(*t*) is given by

*x*(*t*) =  ∞ *n*=−∞

*cnej*2*πf*0*nt* (1.45)

***1.3.2.3 Relationships Between the Transforms*** The (continuous) Fourier transform can be viewed as a limiting case of Fourier series as the period *T* goes

**16** INTRODUCTION

to ∞, and the signal thus becomes aperiodic. Since *f*0 = 1*/T* , let *f* = *nf*0 = *n/T* . Using (1.44), then

*T/*2

*T*→∞ *cnT* = lim

*x*(*t*)*e*−*j*2*πnt/T dt*

lim

=

*T*→∞ ∞

−∞

−*T/*2

*x*(*t*)*e*−*j*2*πft dt*

= *X*(*f* ) (1.46)

Since 1*/T* goes to zero in the limit, we can write 1*/T* as *f* . *f* → 0 as *T* → ∞. Then (1.45) can be written as

*x*(*t*) =  ∞ *n*=−∞

=  ∞

*n*=−∞

=  ∞

*n*=−∞

*T*1*T cnej*2*πf*0*nt*

(*cnT* )*ej*2*πnf*0*t* 1*T*

(*cnT* )*ej*2*πn*( *f* )*t f* (1.47)

*f*→0*x*(*t*) = lim

∞ −∞

*X*(*f* )*ej*2*πft df* (1.48)

where we used the substitution from (1.46) in the last step.

***1.3.2.4 Properties of the Fourier Transform*** Table 1.2 lists some useful properties of Fourier transforms. Combining properties from the table with known Fourier transform pairs from Table 1.1 lets us compute many Fourier transforms and inverse transforms without needing to perform the integrals (1.42) or (1.43).

**TABLE 1.2 Properties of the Fourier Transform**

Concept Time Domain, *x*(*t*) Frequency Domain, *X*(*f* ) Scaling *x*(*at*)1|*a*|*Xfa*

Time shifting *x*(*t* − *t*0) *X*(*f* )*e*−*j*2*πft*0

Frequency shifting *x*(*t*)*ej*2*πf*0*t X*(*f* − *f*0)

Modulation *x*(*t*) cos(*j*2*πf*0*t* + *φ*)12 *X*(*f* − *f*0)*ejφ* + *X*(*f* + *f*0)*e*−*jφ* Differentiation*dnx*

*dtn* (*j*2*πf* )*nX*(*f* )

Convolution *x*(*t*) ∗ *y*(*t*) *X*(*f* )*Y*(*f* )

Multiplication *x*(*t*)*y*(*t*) *X*(*f* ) ∗ *Y*(*f* )

Conjugation *x*∗(*t*) *X*∗(−*f* )

SIGNALS AND SYSTEMS **17**

**1.3.3 Frequency-Domain Concepts**

Some frequency-domain concepts are fundamental for understanding communica tions systems. A miscellany of comments on the frequency domain:

• In the rotating phasor viewpoint, *ej*2*πf*0*t* is a phasor rotating at *f*0 cycles per cycle. But *F*[*ej*2*πf*0*t*] = *δ*(*f* − *f*0). Thus, frequency-domain components of the form *δ*(*f* − *f*0) for any *f*0 can be viewed as rotating phasors.

• Negative frequencies can be viewed as rotating phasors rotating clockwise, whereas positive frequencies rotate counterclockwise.

• For LTI systems, *Y*(*f* ) = *X*(*f* )*H*(*f* ), where *Y*(*f* ), *X*(*f* ), and *H*(*f* ) are the Fourier transforms of the output signal, input signal, and impulse response, respectively. See Figure 1.4.

***1.3.3.1 Power Spectral Density*** *Power spectral density* (PSD) is a way to see how the signal power is distributed in the frequency domain. We have seen that a periodic signal can be written in terms of Fourier series [as in (1.45)]. Similarly, the PSD *Sx*(*f* ) of periodic signals can be expressed in terms of Fourier series:

*Sx*(*f* ) = 1*T* ∞ *n*=−∞

|*cn*|2*δ**t* − *nT*(1.49)

where *cn* are the Fourier series coefficients as given by (1.44).

For nonperiodic power signals *x*(*t*), let *xT* (*t*) be derived from *x*(*t*) by *xT* (*t*) = *x*(*t*)(*t/T* ) (1.50)

Then *xT* (*t*) is an energy signal with a Fourier transform *XT* (*f* ) and an energy spectral density |*XT* (*f* )|2. Then the power spectral density of *x*(*t*) can be defined as

1

*Sx*(*f* ) = lim *T*→∞

*T* |*XT* (*f* )|2 (1.51)

Alternatively, we can apply the *Wiener–Kinchine theorem*, which states that

*Sx*(*f* ) =

∞ −∞

*Rxx*(*τ*)*e*−*j*2*πfτ dτ* (1.52)

In other words, the PSD is simply the Fourier transform of the autocorrelation function. It can be shown that (1.51) and (1.52) and equivalent. Either one can be used to define the PSD and the other can be shown to be equivalent. Whereas (1.51) highlights the connection with the Fourier transform of the signal, (1.52) highlights the connection with its autocorrelation function.

Note that the Wiener–Kinchine theorem applies whether or not *x*(*t*) is periodic. Thus, in the case that *x*(*t*) is periodic with period *T* , clearly also *Rxx*(*τ*) is periodic with the same period. Let *R**xx*(*t*) be equal to *Rxx*(*t*) within one period, 0 ≤ *t* ≤ *T* , and

**18** INTRODUCTION

zero elsewhere, and let *S**x*(*f* ) be the power spectrum of *R**xx*(*t*). Note that

*Rxx*(*t*) =  ∞ *k*=−∞

=  ∞

*k*=−∞

*R**xx*(*t* − *kT* )

*R**xx*(*t*) ∗ *δ*(*t* − *kT* )

*δ*(*t* − *kT* ) (1.53)

Then

= *R**xx*(*t*) ∗  ∞ *k*=−∞

*Sx*(*f* ) = *F* (*Rxx*(*τ*))

= *F*

*R**xx*(*t*) ∗  ∞ *k*=−∞

*δ*(*t* − *kT* )

= *F* *R**xx*(*t*)*F* ∞ *k*=−∞

*δ*(*t* − *kT* )

= *S**x*(*f* )1*T* ∞ *n*=−∞

*δ*

*f* − *nT*(1.54)

*One-Sided vs. Two-Sided PSD.* The PSD that we have been discussing so far is the *two-sided PSD*, which has both positive and negative frequencies. It reflects the fact the a real sinusoid (e.g., a cosine wave) is the sum of two complex sinusoids rotating in opposite directions at the same frequency (thus, at a positive and a negative fre quency). The *one-sided PSD* is a variation that has no negative frequency components and whose positive frequency components are exactly twice those of the two-sided PSD. The one-sided PSD is useful in some cases: for example, for calculations of noise power.

***1.3.3.2 Signal Bandwidth*** Just as in the time domain, we have a notion of duration of a signal (Section 1.2.4), in the frequency domain we have an analogous notion of *bandwidth*. A first-attempt definition of bandwidth might be the interval or range of frequencies from when the signal begins to be nonnegligible to when it stops being nonnegligible (as we sweep from lower to higher frequencies). This is imprecise but can be quantified in various ways, such as:

• *3-dB bandwidth* or *half-power bandwidth*

• *Noise-equivalent bandwidth* (see Section 3.2.3.2)

Often, it is not so much a question of finding the *correct* way of defining bandwidth but of finding a useful way of defining bandwidth for a particular situation.

SIGNALS AND SYSTEMS **19**

Bandwidth is fundamentally related to channel capacity in the following celebrated formula:

*C* = *B* log

1 +*SN*(1.55)

The base of the logarithm determines the units of capacity. In particular, for capacity in bits/second,

*C* = *B* log2

1 +*SN*(1.56)

To obtain capacity in bits/second, we use (1.56) with *B* in hertz and *S/N* on a linear scale (not decibels).

This concept of capacity is known as *Shannon capacity.* Later (e.g., in Section 6.3.2) we will see other concepts of capacity.

**1.3.4 Bandpass Signals and Related Notions**

Because bandpass signals have most of their spectral content around a carrier frequency, say *fc*, they can be written in an envelope-and-phase representation:

*xb*(*t*) = *A*(*t*) cos[2*πfct* + *φ*(*t*)] (1.57)

where *A*(*t*) and *φ*(*t*) are a slowly varying envelope and phase, respectively. Most communications signals while in the communications medium are continuous-wave modulation signals, which tend to be *bandpass* in nature.

***1.3.4.1 In-phase/Quadrature Description*** A bandpass signal *xb*(*t*) can be written in envelope-and-phase form, as we have just seen. We can expand the cosine term using (A.8), and we have

*xb*(*t*) = *A*(*t*)cos(2*πfct*) cos *φ*(*t*) − sin(2*πfct*) sin *φ*(*t*)

= *xi*(*t*) cos(2*πfct*) − *xq*(*t*) sin(2*πfct*) (1.58)

where *xi*(*t*) = *A*(*t*) cos *φ*(*t*) is the *in-phase* component, and *xq*(*t*) = *A*(*t*) sin *φ*(*t*) is the *quadrature* component. Later, in Section 6.1.8.1, we prove that the in-phase and quadrature components are orthogonal, so can be used to transmit independent bits without interfering with each other.

If we let *Xi*(*f* ) = *F*[*xi*(*t*)], *Xq*(*f* ) = *F*[*xq*(*t*)], and *Xb*(*f* ) = *F*[*xb*(*t*)], then *Xb*(*f* ) = 12*Xi*(*f* + *fc*) + *Xi*(*f* − *fc*)− *j*2*Xq*(*f* + *fc*) − *Xq*(*f* − *fc*)(1.59)

***1.3.4.2 Lowpass Equivalents*** There is another useful representation of band pass signals, known as the *lowpass equivalent* or *complex envelope* representation. Going from the envelope-and-phase representation to lowpass equivalent is analogous

**20** INTRODUCTION

to going from a rotating phasor to a (nonrotating) phasor; thus we have *x*lp(*t*) = *A*(*t*)*ejφ*(*t*) (1.60)

which is analogous to (1.27). An alternative definition given in some other books is *x*lp(*t*) = 12*A*(*t*)*ejφ*(*t*) (1.61)

which differs by a factor of 1*/*2. [This is just a matter of convention, and we will stick with (1.60).]

The lowpass equivalent signal is related to the in-phase and quadrature represen tation by

*x*lp(*t*) = *xi*(*t*) + *jxq*(*t*) (1.62)

and we also have

*xb*(*t*) =

*x*lp(*t*)*ej*2*πfct*(1.63)

In the frequency domain, the lowpass equivalent is the positive-frequency part of the bandpass signal, translated down to dc (zero frequency):

*X*lp(*f* ) = *Xi*(*f* ) + *jXq*(*f* )

= 2*Xb*(*f* + *fc*)*u*(*f* + *fc*) (1.64)

where *u*(*f* ) is the step function (0 for *f <* 0, and 1 for *f* ≥ 0).

Interestingly, we can represent filters or transfer functions with lowpass equiva lents, too, so we have

*Y*lp(*f* ) = *H*lp(*f* )*X*lp(*f* ) (1.65)

where

*H*lp(*f* ) = *Hb*(*f* + *fc*)*u*(*f* + *fc*) (1.66)

**1.3.5 Random Signals**

In well-designed communications systems, the signals arriving at a receiver appear random. Thus, it is important to have the tools to analyze random signals. We assume that the reader has knowledge of basic probability theory, including probability dis tribution or density, cumulative distribution function, and expectations [4].

Then a *random variable* can be defined as mapping from a sample space into a range of possible values. A sample space can be thought of as the set of all outcomes of an experiment. We denote the sample space by and let *ω* be a variable that can represent each possible outcome in the sample space. For example, we consider a coin-flipping experiment with outcome either heads or tails, and we define a random

SIGNALS AND SYSTEMS **21**

variable by

*X*(*ω*) =

1 if *ω* = heads

2 if *ω* = tails (1.67)

where the domain of *ω* is the set {heads, tails}. If *P*(heads) = 2*/*3 and *P*(tails) = 1*/*3, then *P*(*X* = 1) = 2*/*3 and *P*(*X* = 2) = 1*/*3. The *average* (also called *mean*, or *expected value*) of *X* is (2*/*3)(1) + (1*/*3)(2) = 4*/*3. Note that when we write just *X*, we have omitted the *ω* for notational simplicity.

***1.3.5.1 Stochastic Processes*** Now we consider cases where instead of just mapping each point in the sample space, *ω*, to a value, we map each *ω* to a function. To emphasize that the mapping is to a function, and that this is therefore not the same as a normal random variable, it is called a *stochastic process* or *random process*. It could also be called a *random function*, but that could be confused with *random variable*, so it may be best to stick with *random variable* in general and *stochastic process* in cases where the mapping is to a function. Depending on the application, we may think of a stochastic process as a *random signal*.

For example, a stochastic process could be defined by a sinusoid with a random phase (e.g., a phase that is uniformly distributed between 0 and 2*π*):

*x*(*t, ω*) = cos(2*πft* + *φ*) (1.68)

where *φ*(*ω*) is a random variable distributed uniformly between 0 and 2*π* (and where we usually omit writing the *ω*, for convenience). Stochastic processes in wireless communications usually involve a time variable, *t*, and/or one or more spatial variables (e.g., *x*, *y*, *z*), so we can write *f* (*x, y, z, t, ω*) or just *f* (*x, y, z, t*) if it is understood to represent a stochastic process.

The entire set of functions, as *ω* varies over the entire sample space, is called an *ensemble*. For any particular outcome, *ω* = *ωi*, *x*(*t*) is a specific *realization* (also known as*sample*) of the random process. For any given fixed *t* = *t*0, *x*(*t*0) is a random variable, *X*0, that represents the ensemble at that point in time (and hence a stochastic process can be viewed as an uncountably infinite set of random variables). Each of these random variables has a density function *fX*0 (*x*0) from which its *first-order*

*statistics* can be obtained. For example, we can obtain the mean *xfX*0 (*x*) *dx*, the variance, and so on. The relationship between random variables associated with two different times *t*0 and *t*1 is often of interest. For example, let their joint distribution be written as *fX*0*,X*1 (*x*0*, x*1); then, if

*fX*0*,X*1 (*x*0*, x*1) = *fX*0 (*x*0)*fX*1 (*x*1) (1.69)

the two random variables are said to be *independent* or *uncorrelated*. The *second order statistics* may be obtained from the joint distribution. This can be extended to the joint distribution of three or more points in time, so we have the *n*th-*order statistics*.

**22** INTRODUCTION

As an example of these ideas, assume that at a radio receiver we have a signal *r*(*t*) that consists of a deterministic signal *s*(*t*) in the presence of additive white Gaussian noise (AWGN), *n*(*t*). If we model the AWGN in the usual way, *r*(*t*) is a stochastic process:

*r*(*t*) = *s*(*t*) + *n*(*t*) (1.70)

Because of the nature of AWGN, *n*(*t*1) and *n*(*t*2) are uncorrelated for any *t*1 =*/ t*2. Furthermore, since AWGN is Gaussian distributed, the first-order statistics depend on only two parameters (i.e., the mean and variance). Since *n*(*t*) = 0 for all *t*, we just need to know the variance, *σ*2(*t*1), *σ*2(*t*2), and so on. Must we have *σ*2(*t*1) = *σ*2(*t*2) for *t*1 =*/ t*2? We discuss this in Section 1.3.5.4. Here, we have just seen that a deterministic communications signal that is corrupted by AWGN can be modeled as a stochastic process.

***1.3.5.2 Time Averaging vs. Ensemble Averaging*** Averages are still useful for many applications, but since in this case we now have multiple variables over which an average may be taken, it often helps to specify to which average we are referring. If we are working with a specific realization of the random signal, we can take the *time average*. For a periodic signal (in time, *t*) with period *T*0,

*< x*(*t*) *>*= 1*T*0 *T*0 0

*x*(*t*) *dt* (1.71)

If it is not a periodic signal, we may still consider a time average as given by

*< x*(*t*) *>*= lim *T*→∞

*T/*2 −*T/*2

*x*(*t*)

*Tdt* (1.72)

Besides the time average, we also have the *ensemble average*, over the entire ensemble, resulting in a function (unlike the time average, which results in a value). For a discrete probability distribution this may be written as

*x*(*t*) = *px,tx* (1.73)

where *px,t* is the probability of event *x*(*t*) at time *t*. The ensemble average for a continuous probability distribution can be written as

*x*(*t*) =

*fXt*(*x*)*x dx* (1.74)

In this book we generally use *<* · *>* to denote time averaging or spatial averaging, and · to denote ensemble averaging.

***1.3.5.3 Autocorrelation*** As we saw in Section 1.3.1.1, for deterministic signals the autocorrelation is a measure of the similarity of a signal with itself.

SIGNALS AND SYSTEMS **23**

The autocorrelation function of a stochastic process *x*(*t*) is

*Rxx*(*t*1*, t*2) = *x*(*t*1)*x*(*t*2) (1.75)

Unlike the case of deterministic signals, this is an ensemble average and in general is a function of two variables representing two moments of time rather than just a time difference. In general, it requires knowledge of the joint distribution of *x*(*t*1) and *x*(*t*2). Soon we will see that these differences go away when *x*(*t*) is an ergodic process.

***1.3.5.4 Stationarity, Ergodicity, and Other Properties*** Going back to example (1.70), we saw that *n*(*t*) was uncorrelated at any two different times. How ever, do the mean and variance have to be constant for all time? Clearly, they do not. In that radio receiver example, suppose that the temperature is rising. To make things simple, we suppose that the temperature is rising monotonically as *t* increases. Then, as we will see in Section 3.2, Johnson–Nyquist noise in the receiver is increasing monotonically with time. Thus,

*σ*2(*t*1) *< σ*2(*t*2) for *t*1 *< t*2

If, instead,

*σ*2(*t*1) = *σ*2(*t*2) for all *t*1 =*/ t*2

there is a sense in which the stochastic process *n*(*t*) is stationary—its variance doesn’t depend on time.

The concept of stationarity has to do with questions of how the statistics of the signal change with time. For example, consider a random signal at *m* time instances, *t*1*, t*2*,...,tm*. Suppose that we consider the joint distribution *fXt*1 *,Xt*2 *,...,Xtm* (*x*1*, x*2*,...,xm*). Then a stochastic process is considered *strict-sense stationary* (SSS) if it is invariant to time translations for all sets *t*1*, t*2*,...,tm*, that is,

*fXt*1+*τ ,Xt*2+*τ ,...,Xtm*+*τ* (*x*1*, x*2*,...,xm*) = *fXt*1 *,Xt*2 *,...,Xtm* (*x*1*, x*2*,...,xm*) (1.76)

A weaker sense of stationarity is often seen in communications applications. A stochastic process is *weak-sense stationary* (WSS) if

1. The mean value is independent of time.

2. The autocorrelation depends only on the time difference *t*2 − *t*1 (i.e., it is a function of *τ* = *t*2 − *t*1), so it may be written as *Rxx*(*τ*) [or *Rx*(*τ*) or simply *R*(*τ*)] to keep this property explicit.

The class of WSS processes is larger than and includes the complete class of SSS processes. Similarly, there is another property, ergodicity, such that the class of SSS processes includes the complete class of ergodic processes. A random process is *ergodic* if it is SSS and if all ensemble averages are equal to the corresponding time averages. In other words, for ergodic processes, time averaging and ensemble averaging are equivalent.

**24** INTRODUCTION

*Autocorrelation Revisited.* For random processes that are WSS (including SSS and ergodic processes), the autocorrelation becomes *R*(*τ*), where *τ* is the time difference. Thus, (1.75) becomes

*Rxx*(*τ*) = *x*(*t*)*x*(*t* + *τ*) (1.77)

which is similar to (1.39).

Furthermore, for ergodic processes, we can even do a time average, so the auto correlation then converges to the case of the autocorrelation of a deterministic signal (in the case of the ergodic process, we just pick any sample function and obtain the autocorrelation from it as though it were a deterministic function).

***1.3.5.5 Worked Example: Random Binary Signal*** Consider a random binary wave, *x*(*t*), where every symbol lasts for *Ts* seconds, and independently of all other symbols, it takes the values *A* or −*A* with equal probability. Let the first symbol transition after *t* = 0 be at *T*trans. Clearly, 0 *< T*trans *< Ts*. We let *T*trans be distributed uniformly between 0 and *Ts*.

The mean at any point in time *t* is

*E*[*x*(*t*)] = *A*(0*.*5) + (−*A*)(0*.*5) = 0 (1.78)

The variance at any point in time *t* is

*σ*2 = *E*[*x*2(*t*)] − (*E*[*x*(*t*)])2 = *A*2 − 0 = *A*2 (1.79)

To figure out if it is WSS, we still need to see if the autocorrelation is dependent only on *τ* = *t*2 − *t*1. We analyze the two autocorrelation cases:

• If |*t*2 − *t*1| *> Ts*, then *Rxx*(*t*1*, t*2) = 0 by the independence of each symbol from every other symbol.

• If |*t*2 − *t*1| *< Ts*, it depends on whether *t*1 and *t*2 lie in the same symbol (in which case we get *σ*2) or in adjacent symbols (in which case we get zero).

What is the probability, *Pa*, that *t*1 and *t*2 lie in adjacent symbols? Let *t*1 = *t*1 − *kTs* and *t*2 = *t*2 − *kTs*, where *k* is the unique integer such that we get both 0 ≤ *t*1 *< Ts* and 0 ≤ *t*2 *< Ts*. Then, *Pa* = *P*(*T*trans lies between *t*1 and *t*2) = |*t*2 − *t*1|*/Ts*.

*E*[*x*(*t*1)*x*(*t*2)] = *A*2 (1 − *Pa*) = *A*2

1 − |*t*2 − *t*1| *Ts*

= *A*2

1 − |*τ*|*Ts*(1.80)

Hence, it is WSS. And using the triangle function notation, we can write the complete autocorrelation function compactly as

*Rxx*(*τ*) = *A*2(*τ/Ts*) (1.81)

This is shown in Figure 1.5.

SIGNALS AND SYSTEMS **25**

*A*2

*t* −*Ts Ts*

**FIGURE 1.5** Autocorrelation function of the random binary signal.

***1.3.5.6 Power Spectral Density of Random Signals*** For a random signal to have a meaningful power spectral density, it should be wide-sense stationary. Each realization of the random signal would have its own power spectral density, different from other realizations of the same random process. It turns out that the (ensemble) average of the power spectral densities of each of the realizations, loosely speaking, is the most useful analog to the power spectral density of a deterministic signal. To be precise, the following procedure can be used on a random signal, *x*(*t*), to estimate its PSD, *Sx*(*f* ). Let us denote the estimate by *S* *x*(*f* ).

1. Observe *x*(*t*) over a period of time, say, 0 to *T* ; let *xT* (*t*) be the truncated version of *x*(*t*), as specified in (1.50), and let *XT* (*f* ) be the Fourier transform of *xT* (*t*). Then its energy spectral density may be computed as |*XT* (*f* )|2.

2. Observe many samples *xT* (*t*) repeatedly, and compute their corresponding Fourier transforms *XT* (*f* ) and energy spectral densities, |*XT* (*f* )|2. 3. Compute *S*˜*x*(*f* ) by computing the ensemble average (1*/T* ) |*XT* (*f* )|2.

One may wonder how to do step 2 in practice. Assuming that *x*(*t*) is ergodic, then (1*/T* ) |*XT* (*f* )|2 is equivalent to time averaging, so we get a better and better estimate *S* *x*(*f* ) by obtaining *xT* (*t*) over many intervals of *T* from the same sample function, and then computing

*S* *x*(*f* ) =

1

*T* |*XT* (*f* )|2

(1.82)

This procedure is based on the following definition of the PSD for random signals: 1

*Sx*(*f* ) = lim

*T*→∞

which is analogous to (1.51).

*T* |*XT* (*f* )|2 (1.83)

Also, as with deterministic signals, the Wiener–Kinchine theorem applies, so

*Sx*(*f* ) =

∞ −∞

*Rxx*(*τ*)*e*−*j*2*πfτ dτ* (1.84)

which can be shown to be equivalent to (1.83).

**26** INTRODUCTION

*S* ( *f* ) *x S* ( *f* ) *y*

⋅

*f*

*f*

2 *H*( *f* )

*f*

**FIGURE 1.6** Filtering and the PSD.

***1.3.5.7 Worked Example: PSD of a Random Binary Signal*** Consider the random binary signal from Section 1.3.5.5. What is the power spectral density of the signal? What happens as *Ts* approaches zero?

We use the autocorrelation function, as in (1.81), and take the Fourier transform to obtain

*Sx*(*f* ) = *A*2*Ts* sinc2(*fTs*) (1.85)

As *Ts* gets smaller and smaller, the autocorrelation function approaches an impulse function. At the same time, the first lobe of the PSD is between −1*/Ts* and 1*/Ts*, so the it becomes very broad and flat, giving it the appearance of the classic “white noise.”

***1.3.5.8 LTI Filtering of WSS Random Signals*** Once we can show that a random signal is WSS, the PSD behaves “like” the PSD of a deterministic signal in some ways; for example, when passing through a filter we have (Figure 1.6)

*Sy*(*f* ) = |*H*(*f* )|2*Sx*(*f* ) (1.86)

where *Sx*(*f* ) and *Sy*(*f* ) are the PSDs of the input and output signals, respectively, and *H*(*f* ) is the LTI system/channel that filters the input signal.

For example, if *Sx*(*f* ) is flat (as with white noise), *Sy*(*f* ) takes on the shape of *H*(*f* ). In communications, a canonical signal might be a “random” signal around a carrier frequency *fc*, with additive white Gaussian noise (AWGN) but with interfering signals at other frequencies, so we pass through a filter (e.g., an RF filter in an RF receiver) to reduce the magnitude of the interferers.

***1.3.5.9 Gaussian Processes*** A *Gaussian process*is one where the distribution *fXt*(*x*) is Gaussian and all the distributions *fXt*1 *,Xt*2 *,...,Xtm* (*x*1*, x*2*,...,xm*) for all sets *t*1*, t*2*,...,tm* are joint Gaussian distributions.

For a Gaussian process, if it is WSS, it is also SSS.

***1.3.5.10 Optimal Detection in Receivers*** An important example of the use of random signals to model communications signals is the model of the signal received

SIGNALING IN COMMUNICATIONS SYSTEMS **27**

Matched filter

*T*

*r*(*t*) *s*(*T* −*t*)

**FIGURE 1.7** Matched filter followed by symbol rate filtering.

at a digital communications receiver. We give examples of modulation schemes used in digital (and analog) systems in Section 1.4. But here we review some fundamental results on optimal detection.

*Matched Filters.* We consider the part of a demodulator after the frequency down translation, such that the signal is at baseband. Here we have a *receiving filter*followed by a sampler, and we want to optimize the receiving filter. For the case of an AWGN channel, we can use facts about random signals [such as (1.86)] to prove that the optimal filter is the *matched filter*. By *optimal* we are referring to the ability of the filter to provide the largest signal-to-noise ratio at the output of the sampler at time *t* = *T* , where the signal waveform is from *t* = 0 to *T* .

**If the signal waveform is** *s*(*t*)**, the matched filter is** *s*(*T* − *t*) **[or more generally, a scalar multiple of** *s*(*T* − *t*)**].**

The proof is outside the scope of this book but can be found in textbooks on digital communications. The matched filter is shown in Figure 1.7, where *r*(*t*) is the received signal, and the sampling after the matched filtering is at the symbol rate, to decide each symbol transmitted.

*Correlation Receivers.* Also known as *correlators*, correlation receivers provide the same decision statistic that matched filters provide (Exercise 1.5 asks you to show this). If *r*(*t*) is the received signal and the transmitted waveform is *s*(*t*), the correlation receiver obtains

*T*

*r*(*t*)*s*(*t*) *dt* (1.87)

0

**1.4 SIGNALING IN COMMUNICATIONS SYSTEMS**

Most communications systems use continuous-wave modulation as a fundamental building block. An exception is certain types of ultrawideband systems, discussed in Section 17.4.2. In continuous-wave modulation, a sinusoid is modified in certain ways to convey information. The unmodulated sinusoid is also known as the *carrier*. The earliest communications systems used analog modulation of the carrier.

These days, with source data so often in digital form (e.g., from a computer), it makes sense to communicate digitally also. Besides, digital communication has advantages over analog communication in how it allows error correction, encryp tion, and other processing to be performed. In dealing with noise and other channel

**28** INTRODUCTION

impairments, digital signals can be recovered (with bit error rates on the order of 10−3 to 10−6, depending on the channel and system design), whereas analog signals are only degraded.

Generally, we would like digital communications with:

• Low bandwidth signals—so that it takes less “space” in the frequency spectrum, allowing more room for other signals

• Low-complexity devices—to reduce costs, power consumption, and so on. • Low probability of errors

The trade-offs between these goals is the focus of much continuing research and development.

If we denote the carrier frequency by *fc* and the bandwidth of the signal by *B*, the design constraints of antennas and amplifiers are such that they work best if *B*  *fc*, so this is usually what we find in communications systems. Furthermore, *fc* needs to be within the allocated frequency band(s) (as allocated by regulators such as Federal Communications Commission in the United States; see Section 17.4) for the particular communication system. The signals at these high frequencies are often called *RF* (radio-frequency) *signals* and must be handled with care with special RF circuits; this is called *RF engineering* (more on this in Chapter 3).

**1.4.1 Analog Modulation**

*Amplitude modulation* (AM) is given by

*Ac*(1 + *μx*(*t*)) cos 2*πfct* (1.88)

where the information signal *x*(*t*) is normalized to |*x*(*t*)| ≤ 1 and *μ* is the *modulation index*. To avoid signal distortion from *overmodulation*, *μ* is often set as *μ <* 1. When *μ <* 1, a simple *envelope detector* can be used to recover *x*(*t*). AM is easy to detect, but has two drawbacks: (1) The unmodulated carrier portion of the signal, *Ac*, represents wasted power that doesn’t convey the signal; and (2) Letting*Bb* and*Bt* be the baseband and transmitted bandwidths, respectively, then for AM, *Bt* = 2*Bb*, so there is wasted bandwidth in a sense. Schemes such as DSB and SSB attempt to reduce wasted power and/or wasted bandwidth.

*Double-sideband modulation* (DSB), also known as *double-sideband suppressed carrier modulation* to contrast it with AM, is AM where the unmodulated carrier is not transmitted, so we just have

*Acx*(*t*) cos 2*πfct* (1.89)

Although DSB is more power efficient than AM, simple envelope detection unfor tunately cannot be used with DSB. As in AM, *Bt* = 2*Bb*.

*Single-sideband modulation* (SSB) achieves *Bt* = *Bb* by removing either the upper or lower sideband of the transmitted signal. Like DSB, it suppresses the carrier to

SIGNALING IN COMMUNICATIONS SYSTEMS **29**

avoid wasting power. Denote the *Hilbert transform* of *x*(*t*) by ˜*x*(*t*); then *x*˜(*t*) = *x*(*t*) ∗1*πt*

and we can write an SSB signal as

*Ac* [*x*(*t*) cos *ωct* ± *x*˜(*t*) sin *ωct*] (1.90)

where the plus or minus sign depends on whether we want the lower sideband or upper sideband.

*Frequency modulation* (FM), unlike linear modulation schemes such as AM, is a nonlinear modulation scheme in which the frequency of the carrier is modulated by the message.

**1.4.2 Digital Modulation**

To transmit digital information, the basic modulation schemes transmit blocks of *k* = log2 *M* bits at a time. Thus, there are *M* = 2*k* different finite-energy waveforms used to represent the *M* possible combinations of the bits. Generally, we want these waveforms to be as “far apart” from each other as possible within certain energy constraints. The *symbol rate* or *signaling rate* is the rate at which new symbols are transmitted, and it is denoted *R*. The data rate is often denoted by *Rb* bits/second (also written bps), and it is also called the *baud rate*. Clearly, *Rb* = *kR*. The symbol period *Ts* is the inverse of the symbol rate, and is the time spent transmitting each symbol before it is time for the next symbol.

A bandlimited channel with bandwidth *B* can support only up to the *Nyquist rate* of signaling, *R*Nyquist = 2*B*. Thus, the signaling rate is constrained by

*R* ≤ *R*Nyquist = 2*B* (1.91)

Digital modulation schemes, especially when the modulation is of the phase or frequency of the carrier, are often referred to as *shift keying* [e.g., amplitude shift key ing (ASK), phase shift keying (PSK), and frequency shift keying (FSK)]. Use of the word *keying* in this context may have come from the concept of Morse code keys for telegraph but is useful for distinguishing digital modulation from analog modulation (e.g., FSK refers to a frequency-modulated digital signal, whereas FM refers to the traditional analog modulation signal that goes by that name). Nevertheless, the dis tinction is not always retained [e.g., a popular family of digital modulation schemes often goes by the name QAM (rather than QASK)].

***1.4.2.1 Pulse Shaping*** A digital modulator takes a simple continuous-time rep resentation of our digital signal and outputs a continuous-time version of our signal, as will be seen in Section 1.4.2.2. How do we prepare our discrete-time digital data to enter a digital modulator? One way of converting from discrete time to continu ous time is to let our data be represented by different baseband pulses for different values. For example, using a basic “rectangle” function, a 1 might be represented as

**30** INTRODUCTION

*p*(*t*) = *π*(*t/Ts*) and a 0 by −*p*(*t*) = −*π*(*t/Ts*) going into the digital modulator; this type of signaling, where one pulse is the exact negative of the other, is called *binary antipodal signaling*.

A problem with using a simple rectangle function in this way is that the spectral occupancy of the signals coming out of the digital modulator would be high—the Fourier transform of the rectangle function is the sinc function, which has relatively large spectral sidebands. Thus, it would be inefficient for use in bandwidth-critical systems such as wireless systems. Thus, it is important to use other *pulse-shaping functions*, *p*(*t*), that can shape the spectral characteristics to use available spectrum more efficiently. However, not just any *p*(*t*) can be used, because it also needs to be chosen to avoid adding *intersymbol interference* unnecessarily between nearby symbols. For example, if we (foolishly) used *p*(*t*) = *π*(*t/*2*Ts*), every symbol would “spill over” into the preceding and/or subsequent symbol (in time) and interfere with them. There is a*Nyquist criterion* for *p*(*t*) to avoid intersymbol interference that can be found in digital communications textbooks. Within the constraints of this criterion, the *raised cosine pulse*, illustrated in Figure 1.8, has emerged as a popular choice for *p*(*t*). The frequency and time domains are shown in the subplots at the top and bottom of the figure, respectively. In the frequency domain we see the raised cosine shape from which the function gets its name. The *roll-off factor*, *α*, is a parameter that determines how sudden or gradual the “roll-off” of the pulse is. In one extreme, *α* = 0, we have a “brick wall” shape in frequency and the familiar sinc function in time (the light solid line on the plots). At the other extreme, *α* = 1, we have the most roll-off, so, the bandwidth expands to twice as much as the *α* = 0 case, as can be seen in the top subplot, with the thick solid line. The case of *α* = 0*.*5 is also plotted in dashed lines in both subplots, and it can be seen to be between the two extremes. For smaller *α*, the signal occupies less bandwidth, but the time sidelobes are higher, potentially resulting in more intersymbol interference and errors in practical receivers. For larger *α*, the signal occupies more bandwidth but has smaller time sidelobes. In practice, to achieve the raised cosine transfer function, a matching pair of *square-root raised cosine filters* are used in the transmitter and receiver, since the receiver would have a matched filter (Section 1.3.5.10). The product of the two square-root raised cosine filters (in the frequency domain) gives the raised cosine shape at the output of the matched filter in the receiver.

***1.4.2.2 Digital Modulation Schemes*** We show examples of common digital modulation schemes.We write examples of these waveforms using lowpass equivalent representation (Section 1.3.4.2) for convenience. In all cases, *p*(*t*) is the *pulse-shaping function*.

*Pulse amplitude modulation* (PAM) uses waveforms of the form *Amp*(*t*) for *m* = 1*,* 2*,...,M* (1.92)

For optimal spacing, the *Am* are arranged in a line with equal spacing between consecutive points.

*H*(*f*)

1

0.8 0.6 0.4 0.2 0

SIGNALING IN COMMUNICATIONS SYSTEMS **31**

α=1

α=0.5

α=0

−3/2*T* −1/*T* −1/2*T* 0 1/2*T* 1/*T* 3/2*T* frequency, *f*

*h*(*t*)

1

0.8 0.6 0.4 0.2 0

α=1

α=0.5 α=0

−0.2

−4*T* −3*T* −2*T* −*T* 0 *T* 2*T* 3*T* 4*T* time

**FIGURE 1.8** Family of raised cosine pulses.

To conserve bandwidth, *SSB PAM* may be used:

*Am*[*p*(*t*) + *jp*˜(*t*)] for *m* = 1*,* 2*,...,M* (1.93)

*Quadrature amplitude modulation* (QAM), where different bits are put in the in-phase (*Ai,m*) and quadrature (*Aq,m*) streams, can be written

(*Ai,m* + *jAq,m*)*p*(*t*) for *m* = 1*,* 2*, . . . , M/*2 (1.94)

Normally, wireless systems would use a form of QAM [e.g., 4-QAM (often just called QAM for short), 16-QAM, 32-QAM, 64-QAM] rather than PAM. Between QAM and PAM, QAM is more efficient because PAM does not exploit the quadrature dimension to transmit information. (For a review of the in-phase and quadrature concept, and to see why different bits can be put in in-phase and quadrature, refer to Sections 1.3.4.1 and 6.1.8.1.) The values *Ai,m* and *Aq,m* for *m* = 1*,* 2*, . . . , M/*2 are chosen to be as

**32** INTRODUCTION

**+ + + +**

**+ + +**

**+**

**+ + +**

**+ +**

**+ +**

**+ +**

**+ +**

**+ + +**

**+**

**+ +**

**+ +** +

**+ +** (a) QAM

**+ +**

(c) QPSK (d) 8PSK

(b) 16-QAM

**FIGURE 1.9** Signal constellations for various digital modulation schemes.

far apart from one another (in signal space) as they can be, given an average power constraint. This is because the farther apart they are, the lower the bit error rates. Examples of 4-QAM and 16-QAM are shown in Figure 1.9.

*Phase shift keying* (PSK) uses waveforms of different phases to represent the different bit combinations:

*ejθm p*(*t*) for *m* = 1*,* 2*,...,M* (1.95)

*Binary PSK* (BPSK) is PSK with *m* = 1, *quadrature PSK* (QPSK) is PSK with *m* = 2, and *8-PSK* is PSK with *m* = 3. QPSK is very popular in wireless systems because it is more efficient than BPSK. 8-PSK is seen in EDGE (Section 8.1.3), for example. QPSK and 8-PSK are shown in Figure 1.9.

***1.4.2.3 Signal Constellations*** A good way to visualize the waveforms in a digital modulation scheme is through the *signal constellation* diagram. We have seen that the (lowpass equivalent of the) *M* possible waveforms in general (except for modulation schemes like PAM) lie in the complex plane. We can therefore plot all the points in the complex plane, and the result is known as the *signal constellation*, some examples of which are shown in Figure 1.9. Notice that the signal constellation of 4-QAM happens to be the same as that of QPSK.

When we discuss wireless access technologies, we elaborate on selected aspects of digital modulation (Section 6.2), especially those having to do with design choices typically encountered in wireless systems.

**1.4.3 Synchronization**

In a digital receiver, two main types of synchronization are needed at the physical layer (there may also be other types of synchronization at higher layers, e.g., frame synchronization, multimedia synchronization, etc.):

• Carrier phase synchronization

• Symbol timing synchronization and recovery

REFERENCES **33**

*Carrier phase synchronization* is about figuring out, and recovering, a carrier signal frequency and phase. *Symbol timing synchronization and recovery* is about figuring out the locations (in time) of the temporal boundaries between symbols. It is also known as *clock recovery*.

**EXERCISES**

**1.1** The form of the Fourier series given in Section 1.3.2 is the exponential form. Show how this is equivalent to the trigonometric form

*x*(*t*) = *a*0 + ∞ *n*=1

*an* cos 2*πf*0*nt* + *bn* sin 2*πf*0*nt* (1.96)

Express *cn* in terms of *an* and *bn*.

**1.2** Instead of the random binary waveform we saw in Section 1.3.5.5, we have a random digital waveform. So it takes not just two values, 1 and −1, but a range of values over a distribution: say, a Gaussian distribution with mean 0 and variance *σ*2. Find the autocorrelation function of the random digital waveform. How does it compare with the autocorrelation function of the random binary waveform given by (1.81)?

**1.3** Suppose we have a signal *x*(*t*) that is multiplied by a sinusoid, resulting in the signal *y*(*t*) = *x*(*t*) cos 2*πft*. Assume that *x*(*t*) is independent of the sinusoid but could otherwise be a (deterministic or random) signal with autocorrelation function *Rxx*(*τ*). Show that the autocorrelation of *y*(*t*) is given by

*Ryy*(*τ*) = *Rxx*(*τ*)

1

2cos 2*πft*

(1.97)

**1.4** Continuing from Exercise 1.3, what is the effect on the power spectral density of multiplication by a sinusoid? In other words, express the power spectral density of *y*(*t*) in terms of the power spectral density of *x*(*t*). This is a fundamental and useful result, since it means that we can up-convert and down-convert signals to and from carrier frequencies, and the autocorrelation function and power spectral density behave in this predictable way.

**1.5** Show that a matched filter followed by sampling at *t* = *T* produces the same output as a correlation receiver.

**REFERENCES**

1. A. B. Carlson. *Communication Systems*, 3rd ed. McGraw-Hill, New York, 1986. 2. L. Couch. *Digital and Analog Communication Systems*, 7th ed. Prentice Hall, Upper Saddle River, NJ, 2007.

**34** INTRODUCTION

3. J. W. Nilsson. *Electric Circuits*, 3rd ed. Addison-Wesley, Reading MA, 1990. 4. A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, New York, 1991.

5. B. Sklar. *Digital Communications: Fundamentals and Applications*, 2nd ed. Prentice Hall, Upper Saddle River, NJ, 2001.

II

RADIO FREQUENCY, ANTENNAS, AND PROPAGATION

2

INTRODUCTION TO RADIO FREQUENCY, ANTENNAS, AND PROPAGATION

In this chapter we review selected topics in electromagnetics that provide foundational support for our coverage of radio frequency (Chapter 3), antennas (Chapter 4), and propagation (Chapter 5). We begin in Section 2.1 with a review of some mathemat ical tools for computing scalar and vector quantities that are typically used in basic electromagnetics. We then review electrostatics and magnetostatics in Section 2.2. Time-varying situations, wave propagation, and transmission lines are examined in Section 2.3. A brief comparison of different notions of impedance is presented in Section 2.4, followed by an introduction to test and measurement equipment in Section 2.5.

**2.1 MATHEMATICAL PRELIMINARIES**

Here we review briefly some mathematical tools for working with scalar and vector functions in three dimensions.

**2.1.1 Multidimensional/Multivariable Analysis**

In Section 1.2.4 the signals concerned were measured at one place in a circuit (e.g., the voltage between two fixed points), where spatial dimensions were not important. Fur thermore, the signals were all scalar functions. Now we extend our signal concepts, as well as concepts of sinusoids and phasors, into one or more spatial dimensions. Fur thermore, the signals may be vector functions, not just scalar functions. For example,

*Fundamentals of Wireless Communication Engineering Technologies*, First Edition. K. Daniel Wong. © 2012 John Wiley & Sons, Inc. Published 2012 by John Wiley & Sons, Inc.

**37**

**38** INTRODUCTION TO RADIO FREQUENCY, ANTENNAS, AND PROPAGATION

we might have a scalar function, *ρ*, of coordinates *x*, *y*, and *z* (and time *t*) *ρ*(*x, y, z, t*), that we may also write as *ρ*(**A***, t*), where *A* is a vector representing a spatial coordi nate [e.g., (*x, y, z*)]. Such a function is also used to represent a *scalar field*. Or we might have a vector function, **H**, that we may write as **H**(*x, y, z, t*) or **H**(**A***, t*). Such a function is used to represent a *vector field*. It has amplitude and direction at every point in space and time within the domain of the function.

***2.1.1.1 Basic Vector Operations*** Let **A** and **B** be vectors, *A* = |**A**|, *B* = |**B**|, *θAB* the angle between **A** and **B**, and **un** the unit vector normal (perpendicular) to **A** and **B**.

• Dot product:

**A** · **B** = *AB* cos *θAB* (2.1)

• Cross product:

**A** × **B** = **un** |*AB* sin *θAB*| (2.2)

***2.1.1.2 Coordinate Systems*** The cylindrical coordinate system, (*r, φ, z*), is shown in Figure 2.1 and the spherical coordinate system, (*R, θ, φ*), in Figure 2.2. For conversion between coordinate systems, see Exercise 2.1. In coordinate systems such as the cylindrical and spherical, we often want to convert a differential change in the

*Z*

*z r*

*y*

**P**

*Y x φ X*

**FIGURE 2.1** Cylindrical coordinates.

MATHEMATICAL PRELIMINARIES **39**

*Z*

**P**

*R*

*θ*

*y*

*φY*

*X*

*z*

*x*

**FIGURE 2.2** Spherical coordinates.

coordinates to a differential change in length. Let us denote the unit vectors by **u** with the appropriate subscripts. Metric coefficients (for length conversions) are:

• Cartesian coordinates:

*dl* = **ux** *dx* + **uy** *dy* + **uz** *dz* (2.3)

• Cylindrical coordinates:

*dl* = **ur** *dr* + **u***φr dφ* + **uz** *dz* (2.4)

• Spherical coordinates:

*dl* = **uR** *dR* + **u***θR dθ* + **u***φR* sin *θ dφ* (2.5)

***2.1.1.3 Gradient, Divergence, and Curl***

• Gradient [for a scalar function of space coordinates, e.g., *V*(*u*1*, u*2*, u*3)]:

∇ ≡

**u1***∂*

*h*1 *∂u*1+ **u2***∂*

*h*2 *∂u*2+ **u3***∂*

*h*3 *∂u*3

(2.6)

**40** INTRODUCTION TO RADIO FREQUENCY, ANTENNAS, AND PROPAGATION In Cartesian coordinates it becomes

∇ ≡

**ux***∂∂x* + **uy***∂∂y* + **uz***∂∂z*(2.7)

and in cylindrical or spherical coordinates, the appropriate metric coefficients *h*1, *h*2, and *h*3 must be applied.

• Divergence [for a vector field, e.g., **A**(*u*1*, u*2*, u*3)]:

*S* **A** · *d***s**

*v* (2.8)

In Cartesian coordinates,

div**A** = lim  *v*→0

div**A** = *∂Ax*

*∂x* +*∂Ay*

*∂y* +*∂Az*

*∂z* (2.9)

Thus, symbolically, div can be written as div **A** =∇· **A**. But this is just symbolic, since it makes sense only in Cartesian coordinates. The real definition is given by (2.8).

• Curl (for a vector field):

curl**A** = lim  *s*→0

1

*s*

**un**

*C*

**A** · *dl*

(2.10)

Just as div can be written symbolically as ∇ · **A**, curl can also be written sym bolically as curl **A** =∇× **A**, which makes sense in Cartesian coordinates.

***2.1.1.4 Sinusoids, Waves, and Phasors*** For a sinusoid that is a function of both position and time, we may expand (1.16) to obtain

*ξ*(*x, t*) = *A* cos(*kx* − *ωt* + *φ*) = *A* cos(*kx* − 2*πft* + *φ*) (2.11)

where *k* is the *spatial frequency*, which is to the spatial dimension what the tempo ral frequency *ω* is to the temporal dimension. With the introduction of the spatial dimension, this sinusoid could be thought of as a wave, although the concept of waves includes more than just a simple sinusoid like this one. It also encompasses the superposition of multiple sinusoids.

Previously, we introduced *T* = 1*/f* = 2*π/ω* as the period of the sinusoid. We see that this relationship remains if we fix *k* and *x* so that *kx* gets absorbed into the phase with *φ*. Thus, at any particular fixed spatial location (fixed *x*), the period remains as *T* = 1*/f* = 2*π/ω*. However, if we now fix *t* and *ω* instead, then *ωt* gets absorbed into the phase with *φ*, and we have, at any particular fixed moment in time, the “spatial period,” much more commonly called the *wavelength*, given by

*λ* = 2*πk* (2.12)

ELECTROSTATICS, CURRENT, AND MAGNETOSTATICS **41**

Analogous to the period in time, the wavelength is the smallest (spatial) distance such that

*ξ*(*x*) = *ξ*(*x* + *λ*) for − ∞ *<x<* ∞ (2.13)

For the case that *ξ*(*x, t*) represents a *traveling wave* (also known as a *propagating wave*), we can relate *λ* and *f* through the velocity of the wave (also known as the *phase velocity*), *v*:

*λ* = *vf* (2.14)

When we want the sinusoidal function to represent phenomena in three spatial dimensions (e.g., a propagating electromagnetic wave), we can replace the scalar function in (2.11) with a vector function. For convenience, we often align the axes so that the direction of propagation is along one of the axes (e.g., the *z*-axis), in which case we could write

*ξ*(*z, t*) = **A** cos(*kz* − *ωt* + *φ*) = **A** cos(*kz* − 2*πft* + *φ*) (2.15)

where **A** = *A***ux**, for example.

In this way, the concept of phasors introduced in Section 1.2.6 could be extended so that:

• Phasors can be both a sinusoidal function of time and a function of space. • Besides having amplitude and phase, they also have a direction. Thus, we go from the scalar phasors of Section 1.2.6 to vector phasors.

For example, we could represent an electric field as

**E**(*x, y, z, t*) =

where **E**(*x, y, z, t*) is a vector phasor.

**E**(*x, y, z*)*ej*2*πft* (2.16)

**2.2 ELECTROSTATICS, CURRENT, AND MAGNETOSTATICS**

In this section we briefly review electrostatics from Sections 2.2.1 to 2.2.4, electric current in Section 2.2.5, and magnetostatics from Sections 2.2.6 to 2.2.8. Section 2.2.9 provides a summary of the various symbols introduced in the section.

**2.2.1 Electrostatics in Free Space**

*Differential Form*

∇ · **E** = *ρ*0(2.17)

**42** INTRODUCTION TO RADIO FREQUENCY, ANTENNAS, AND PROPAGATION

∇ × **E** = 0 (2.18)

where *ρ* is the volume charge density of free charges in C/m3 and 0 is the permittivity of free space (a.k.a. vacuum permittivity; 0 ≈ 1*/*36*π* × 10−9 in F/m).

*Integral Form.* Gauss’s law:

**E** · *d***s** = *Q*0(2.19) *S*

*C*

**E** · *dl* = 0 (2.20)

*Coulomb’s Law.* Force between *q*1 and *q*2 is given by

**F** = **uR***q*1*q*2

4*π*0*R*2 (2.21)

Alternatively, force on *q* is given by

**F** = *q***E** (2.22)

where the unit is newtons.

For a conductor under static conditions, (2.19) and (2.20) can be used to show:

• That the **E** field on the conductor surface is normal to the surface everywhere, and we can write *E*⊥ = *ρs/*0, where *E*⊥ is the normal component and *ρs* is the surface charge density.

• That the tangential component of the **E** field on the conductor surface is zero.

**2.2.2 Voltage**

Since *E* is curl-free, it can be written as the gradient of a scalar field. We define the scalar field to be *electric potential V* and

**E** = −∇*V* (2.23)

Then the units of *E* are volts per meter. We don’t have the space to discuss this further, just to note that electric potential has physical significance, related to the work that needs to be done to move a charge from point to point.

Poisson’s equation in free space is

∇2*V* = − *ρ*0(2.24)

***2.2.2.1 Worked Example: Electric Potential at Distance r from the Spherical Conductor*** As discussed earlier, the **E** field must be normal and hence

ELECTROSTATICS, CURRENT, AND MAGNETOSTATICS **43**

pointing radially outward from the spherical conductor. Since the surface area of a sphere is 4*πr*2, then *ρs* = *Q/*4*πr*2 and using (2.19) gives us

|**E**| = *E*⊥ = *Q*

4*π*0*r*2 (2.25)

Then, taking the point at infinity as the zero-voltage reference point, we have *V* = −

*r* ∞

*E*⊥ *dr* = *Q*

4*π*0*r* (2.26)

***2.2.2.2 Worked Example: Two Connected Spherical Conductors*** Con sider two spherical conductors that are electrically connected by a perfectly conducting wire. Let the radii of the spheres be *r*1 and *r*2, respectively. Assume that the spheres are far enough apart that the charge distribution on each is not influenced by the field caused by the charge distribution on the other; thus, the charge distribution on each is uniform. Let *Q* coulombs of charge be deposited in the spheres. Find (a) the charges on each sphere; (b) the charge density at the surface of each sphere; (c) the electric field intensity at the surface of each sphere.

Let the charge on the spheres be *Q*1 and *Q*2, respectively, so that *Q* = *Q*1 + *Q*2. Since the two spheres are connected by the wire, they are at the same potential, and the potential is given by (2.26), so we have

*Q*1

4*π*0*r*1= *Q*2

4*π*0*r*2(2.27)

So

*Q*1 = *r*1

*r*1 + *r*2*Q* and *Q*2 = *r*2

*r*1 + *r*2*Q* (2.28)

Then the charge densities are

*ρs,*1 = *Q*1

4*πr*21= *Q*

4*πr*1(*r*1 + *r*2)and *ρs,*2 = *Q*2

4*πr*22= *Q*

4*πr*2(*r*1 + *r*2) (2.29)

so

*E*⊥*,*1 = *Q*

4*π*0*r*1(*r*1 + *r*2)and *E*⊥*,*2 = *Q*

4*π*0*r*2(*r*1 + *r*2) (2.30)

Thus, if sphere 1 is bigger than sphere 2, it will have proportionately more charge but a smaller surface charge density and electric field intensity at its surface.

**2.2.3 Electrostatics in the Case of Dielectrics/Insulators**

Dielectrics are also known as *insulators*. When there are dielectrics, there will be polarization charge densities, resulting in a *polarization vector,* **P**. For convenience, we introduce **D**, given by

**D** = 0**E** + **P** (2.31)

**44** INTRODUCTION TO RADIO FREQUENCY, ANTENNAS, AND PROPAGATION

Thus, in considering **D**, we can ignore polarization as it is only affected by free charges (whereas **E** is reduced by the polarization in the dielectric):

∇ · **D** = *ρ* (2.32)

If material is linear and isotropic, **P** = 0*χe***E**, where *χe* is *electric susceptibility*. Then **D** = 0(1 + *χe*)**E** = 0 *r***E** = **E** (2.33)

*r* is known as *dielectric constant* or*relative permittivity, and* is *absolute permittivity* or *permittivity*.

***2.2.3.1 Dielectric Breakdown*** Equation (2.31) holds only when the electric field intensity is below a critical amount, the *dielectric strength* of the material. If the electric field exceeds the dielectric strength, *dielectric breakdown* occurs; then it becomes conducting. When air breaks down at 3 × 106 V/m, sparking or corona discharge occurs.

**2.2.4 Electrostatics Summary**

In summary, for electrostatics we have

∇ · **D** = *ρ* C/m3 (2.34)

∇ × **E** = 0 (2.35)

Further, if material is linear and isotropic,

**D** = **E** (2.36)

**2.2.5 Currents**

There are conduction currents, electrolytic currents, and convection currents. Ohm’s law governs only conduction currents. In circuits, it is *V* = *RI*. Ohm’s law in point form is

**J** = *σ***E** A/m2 (2.37)

where *σ* is *conductivity* in A/V·m or S/m. The reciprocal of *σ* is resistivity. The principle of *conservation of charge* leads to the *equation of continuity*:

∇ · **J** = −*∂ρ∂t*A/m3 (2.38)

ELECTROSTATICS, CURRENT, AND MAGNETOSTATICS **45**

**2.2.6 Magnetostatics Introduction**

In the case of a moving charge, not only is there electric force, as from (2.22), but there is magnetic force. Hence, we have

**F** = *q*(**E** + **u** × **B**) N (2.39)

**2.2.7 Magnetostatics in Free Space**

*B* is magnetic flux density in Wb/m2 or teslas (a weber is a volt-second): ∇ · **B** = 0 (2.40)

∇ × **B** = *μ*0**J** (2.41)

where *μ*0 is the *permeability* of free space (*μ*0 = 4*π* × 10−7 H/m).

**2.2.8 Magnetostatics in the Case of Magnetic Materials**

Just as in dielectrics you had a polarization vector, so in magnetic materials you have a magnetization vector. Let the magnetization vector be **M**. Then

**B** = *μ*0**H** + **M** A/m (2.42)

Thus, we can just deal with the effects of free current, **J**, as in

∇ × **H** = **J** (2.43)

**2.2.9 Symbols**

We recap some of the symbols introduced in prior sections:

*C* is capacitance in F/m

*D* is electric displacement or electric flux density in C/m2

*E* is electric field intensity in V/m

*J* is current density in A/m2

0 is permittivity of free space in F/m

*r* is *dielectric constant* or *relative permittivity* (dimensionless); permittivity rela tive to free space

is *absolute permittivity* or *permittivity* in F/m; how much the medium “permits” some charge *q* to create an electric field

*μ*0 is the *permeability* of free space in H/m

*μ* is *absolute permeability* in H/m

*ρ* is volume charge density of free charges in C/m3

*σ* is conductivity in A/V·m or S/m

**46** INTRODUCTION TO RADIO FREQUENCY, ANTENNAS, AND PROPAGATION

**2.3 TIME-VARYING SITUATIONS, ELECTROMAGNETIC WAVES, AND TRANSMISSION LINES**

We begin in this section with Maxwell’s equations (Section 2.3.1) and proceed on to electromagnetic (EM) waves (Section 2.3.2). Then we discuss transmission lines (Section 2.3.3), standing-wave ratios (Section 2.3.4) and S-parameters (Section 2.3.5).

**2.3.1 Maxwell’s Equations**

In differential form, Maxwell’s equations are

∇ × **E** = −*∂***B***∂t* (2.44)

∇ × **H** = **J** +*∂***D***∂t* (2.45)

∇ · **D** = *ρ* (2.46)

∇ · **B** = 0 (2.47)

In integral form, Maxwell’s equations are

*C*

**E** · *dl* = −*d dt* (2.48)

*∂***D**

*C*

**H** · *dl* = *I* +

*∂t* · *d***s** (2.49) *S*

**D** · *d***s** = *Q* (2.50) *S*

**D** · *d***s** = 0 (2.51) *S*

In linear, isotropic, homogeneous media, Maxwell’s equations can be written as (vector) phasors:

∇ × **E** = −*j*2*πfμ***H** (2.52)

∇ × **H** = **J** + *j*2*πf***E** (2.53)

∇ · **E** = *ρ/* (2.54)

∇ · **H** = 0 (2.55)

TIME-VARYING SITUATIONS, ELECTROMAGNETIC WAVES, AND TRANSMISSION LINES **47**

where we have written the four equations just in terms of *E* and *H* alone for conve nience, because in linear and isotropic media, **D** = **E** and **B** = *μ***H**.

**2.3.2 Electromagnetic Waves**

A *source-free region* is one where *ρ* = 0 and **J** = 0. Assume a source-free region where the medium is linear, isotropic, homogeneous, and nonconducting. Then using (2.44) and (2.45), we have

∇×∇× **E** = −*μ∂∂t*(∇ × **H**) = −*μ∂*2**E**

*∂t*2 (2.56)

But ∇×∇× **E** = ∇(∇ · **E**) − ∇2**E** = −∇2**E**, where *ρ* = 0, so we have ∇2**E** − *μ∂*2**E**

*∂t*2 = 0 (2.57)

Similarly, we can derive

∇2**H** − *μ∂*2**H**

*∂t*2 = 0 (2.58)

These are wave equations, and the speed of the wave is *v* = 1*/*√*μ*. In particular, in free space we have

∇2**E** − 1*c*2*∂*2**E**

*∂t*2 = 0 (2.59)

where

*c* = 1

√*μ*00≈ 3 × 108 m/s (2.60)

NB: Because (2.57) and (2.58) are linear, we can apply the superposition principle and add waves to get the resultant (we do this everywhere, e.g., in adding the transmitted and reflected wave in transmission lines, in adding the contribution of different paths in multipath propagation environments, in analyzing the behavior of an antenna array). The *intrinsic impedance* of a medium is *η* = √*μ/*. For free space, we have

*η*0 =

*μ*0

0≈ 120*π* ≈ 377 (2.61)

Working in phasor notation, (2.57) and (2.58) become

∇2**E** + *k*2**E** = 0 (2.62)

and

∇2**H** + *k*2**H** = 0 (2.63)

**48** INTRODUCTION TO RADIO FREQUENCY, ANTENNAS, AND PROPAGATION

where

*k* = 2*πf*√*μ* = 2*πfv* = 2*πλ* (2.64)

since *λ* = *v/f* .

***2.3.2.1 Flow of Electromagnetic Power and the Poynting Vector*** We define the Poynting vector

*P* = **E** × **H** W/m2 (2.65)

It is a power flux density vector associated with an electromagnetic field. *P* points in the direction of the flow of electromagnetic power and its amplitude is power flux density.

For computing the average power flux density in a propagating wave: *P*av = 12 (**E** × **H**∗) W/m2 (2.66)

which is analogous to the following from circuit theory:

*P*av = 12 (*VI*∗) W (2.67)

Now consider the case of time-harmonic waves, and in particular, a uniform plane wave propagating in a lossy medium in the +*z*-direction, where (in phasor notation)

**E**(*z*) = **u***xE*0*e*−(*α*+*jβ*)*z* (2.68)

Then if the intrinsic impedance of the medium is *η* = |*η*|*ejθη* , we have **H**(*z*) = **u***yE*0

|*η*|*e*−*αze*−*j*(*βz*+*θη*) (2.69)

Thus, in the lossless case, *α* = 0 and we have

*P* = **u***zE*20

|*η*| (2.70)

**2.3.3 Transmission-Line Basics**

For efficient transmission of electromagnetic waves from one point to another point, the electromagnetic waves must be directed or guided. Transmission lines are one way to do that. They are especially useful when signals are at RF and we cannot just use basic circuits (see Section 3.1.2 for further discussion). In this section we introduce transmission lines and work through just enough of the equations to pro vide a foundation to introduce the very important concept of standing-wave ratios in Section 2.3.4.

TIME-VARYING SITUATIONS, ELECTROMAGNETIC WAVES, AND TRANSMISSION LINES **49** Three of the most common transmission-line structures are:

• *Parallel plate*: two parallel conducting plates that are separated by a dielectric slab of uniform thickness (e.g., microstrips in printed circuit board technology). • *Two-wire transmission line*: a pair of parallel conducting wires that are separated

by a uniform distance (e.g., flat lead-in lines connecting TVs and antennas). • *Coaxial transmission line*: inner conducting wire and coaxial outer conducting sheath that are separated by a dielectric medium.

Microstrips are commonly found in microwave integrated circuits. They consist of a conducting strip over a ground plane, with a dielectric material between the strip and the plane. They are small, cheap, and easily produced, but suffer from higher losses and can handle less power than can other transmission lines, such as coaxial cables. Microstrip transmission lines are sometimes also known as striplines, but sometimes they are considered different from striplines. When considered different, the term *striplines* is used to refer specifically to a variant with two ground planes, one on each side of the conducting strip. The ground planes then sandwich the dielectric material, and the conducting strip is embedded in the dielectric material [2].

Shown in Figure 2.3, microstrip transmission lines are closely related to microstrip patch antennas (Section 4.2.7.1). With one set of parameters, the microwave energy is better contained within the structure, and it is used as a transmission line, whereas with another set of parameters, the structure radiates and it is used as an antenna. *W*,  *r*, and *h* (width of the microstrip, dielectric constant of the dielectric, and height of the dielectric) are important parameters, whereas other parameters, such as thickness *t* and conductivity *σ* of the strip, are not as important.

Next, we derive a very useful and convenient model of transmission lines in Section 2.3.3.1. The model is quite accurate for coaxial and two-wire transmission lines. It also works well for parallel-plate transmission lines where the two plates are of equal width with negligible fringing effects. However, it should be used with caution for modeling microstrip transmission lines, since the metal strip might not be very wide. The model provides a reasonable approximation when *h*   
 *W* and for

*W*

*t*

*r* ε

Substrate *h*

Ground Plane

**FIGURE 2.3** Model of a microstrip transmission line.

**50** INTRODUCTION TO RADIO FREQUENCY, ANTENNAS, AND PROPAGATION

lower microwave frequencies. As for higher frequencies such as millimeter-wave, more complicated full-wave analysis might be recommended [1].

***2.3.3.1 Modeling the Behavior of a Transmission Line*** The transmission line can be modeled as broken up into short segments each of length  *x*, as shown in Figure 2.4. Consider the segment between *x* and *x* +  *x*. Let the voltage across the first side (at *x*) be *v*(*x, t*) and the current in be *i*(*x, t*), and the voltage across the other side (at *x* +  *x*) be *v*(*x* +  *x, t*) and the current out be *i*(*x* +  *x, t*). Let *L*, *R*, *C*, and *G* be the inductance, resistance, capacitance, and conductance per unit length. Then the inductance, resistance, capacitance, and conductance in our small segment of circuit can be represented as *L x* and *R x* in series, and *C x* and *G x* in parallel.

Applying Kirchhoff’s current and voltage laws, then dividing by  *x* and taking the limit as  *x* → 0, we have a couple of partial differential equations in *x* and *t*. The steady-state sinusoidally time-varying solutions can be written in phasor notation,

*v*(*x, t*) = Re*i*(*x, t*) = Re

*V*(*x*)*ej*2*πft* (2.71)

*I*(*x*)*ej*2*πft* (2.72)

and it turns out that the solution is a “wave equation,”

*d*2*V*(*x*)

*dx*2 − *γ*2*V*(*x*) = 0 (2.73)

*d*2*I*(*x*)

*dx*2 − *γ*2*I*(*x*) = 0 (2.74)

and the wave propagation constant *γ* is

*γ* = (*R* + *j*2*πfL*) (*G* + *j*2*πfC*) = *α* + *jβ* (2.75)

where *α* is an attenuation constant in nepers per unit length and *β* is a phase constant in radians per unit length.

*i*(*x*,*t*) *i*(*x* + Δ*x*,*t*)

+

+

●

*L*Δ*x R*Δ*x*

●

*C*Δ*x*

*v*(*x*,*t*) *v*(*x* +Δ*x*,*t*) *G*Δ*x*

●

–

–

●

Δ*x*

**FIGURE 2.4** Model of a transmission line.

TIME-VARYING SITUATIONS, ELECTROMAGNETIC WAVES, AND TRANSMISSION LINES **51**

Usually, *R* and *G* are small, and for the *lossless* case, *R* and *G* are zero. Equations (2.73) and (2.74) can be solved by the following functions:

*V*(*x*) = *V* +(*x*) + *V* −(*x*) (2.76)

= *V* +0 *e*−*γx* + *V* −0 *eγx* (2.77)

*I*(*x*) = *I*+(*x*) + *I*−(*x*) (2.78)

= *I*+0 *e*−*γx* + *I*−0 *eγx* (2.79)

where we see the *V* and *I* for a forward (*V* +,*I*+) wave traveling in the +*x* direction, and a backward wave (*V* −,*I*−) traveling in the −*x* direction. The ratio *V* +0 */I*+0 is very important, and we call it the *characteristic impedance* of the transmission line, *Z*0. It can easily be shown that

*Z*0 = *V* +0

*γ* = *γ*

*R* + *j*2*πL*

*I*+0= *R* + *j*2*πL*

**2.3.4 Standing-Wave Ratios**

*G* + *j*2*πC* =

*G* + *j*2*πC* (2.80)

When transmitting, say, to an antenna, along a cable, you have an *incident wave*, also known as a *forward wave*, and a *reflected wave* [as we saw in (2.77) and (2.79)]. A *standing wave* results from the superposition of the incident and reflected waves. The *standing-wave ratio* (SWR) is the ratio of the peak to the trough of the standing wave. It could be a voltage ratio or a current ratio (the numerical ratio should be the same whether we consider voltage or current). Since it could be a voltage ratio, the SWR is also often called the *voltage standing-wave ratio* (VSWR). Calling it VSWR also helps remove ambiguity, as sometimes the *power standing-wave ratio* (PSWR) is also seen, where PSWR is the square of the VSWR. Figure 2.7 shows an example of standing waves in a transmission line.

Although we have seen that the voltage and current, as functions of position, are given by (2.77) and (2.79), we need a bit more theory to understand the ratios of the incident and reflected waves so that we can get a grip on the SWR. We begin by examining the special case where impedances are matched (Section 2.3.4.1) so there is no reflected wave. This will be followed in Section 2.3.4.2 by examining the more general case where the reflected wave exists.

***2.3.4.1 Impedance Matching and the Transmission Line*** A transmission line will have no reflections only if it is infinitely long, *or* if it is connected to a matched load (Figure 2.5). This brings us to the subject of impedance matching in the context of transmission lines. In matching source and loads with transmission lines, it is important to distinguish between:

• The connection from a source to the transmission line on one side. • The connection from the transmission line to a load on the other side.

**52** INTRODUCTION TO RADIO FREQUENCY, ANTENNAS, AND PROPAGATION

+

*ZS*

*i I* +

*LI*

*ZL*

+

*VS*

−

**~** *Vi Zi VL*

−

−

*l*

**FIGURE 2.5** Characterizing a transmission line.

In the former case (i.e., connecting a source to a transmission line), for maximum power transfer the input impedance looking into the transmission line, *Zi* [see (2.81), which we will get to shortly], should equal the complex conjugate of the output impedance of the source (i.e., *Zi* = *Z*∗*s* , where *Zs* is the source impedance). This is what we might expect from basic circuit theory, and coincides with the concept of impedance matching from basic circuit theory. It is also indicative of one way in which we can treat transmission lines as circuit elements in basic circuits. In the latter case, the load connected to a transmission line should have input impedance **equal to the characteristic impedance** of the transmission line for matched load and best efficiency (i.e., *Z*0 = *ZL*). This is *different* from what we might expect from basic circuit theory, so we have to be careful. It is not the complex conjugate of the characteristic impedance but the characteristic impedance itself. This is because this matching is based on a different principle from normal circuit conjugate impedance matching. This matching is based on eliminating the reflected wave, which can result in serious power losses. In fact, for transmission lines, *Z*0 = *ZL* is more important (to reduce power loss from reflected waves) than conjugate matching on the source side. Having said this, we now proceed to discuss *Zi*.

For a transmission line of length *l*, with characteristic *γ* and *Z*0, if it is connected to a load with input impedance *ZL*, the input impedance *Zi* of the transmission line and load combination is given by [5]

*Zi* = *Z*0*ZL* + *Z*0 tanh *γl*

*Z*0 + *ZL* tanh *γl* (2.81)

In the lossless case, *γ* = *jβ* and tanh *jβl* = *j* tan *βl*, so [5]

*Zi* = *Z*0*ZL* + *Z*0*j* tan *βl*

*Z*0 + *ZLj* tan *βl* (2.82)

Notice that when we have a load matched to the transmission line (i.e., *ZL* = *Z*0), the input impedance from (2.81) is*Zi* = *Z*0. Thus, the combination of the transmission line and the load in this special case looks exactly the same (same input impedance, same voltage and current distribution over the line), as if the transmission line is infinitely long, with no termination; and there is no reflected wave.

TIME-VARYING SITUATIONS, ELECTROMAGNETIC WAVES, AND TRANSMISSION LINES **53**

+

*ZS*

*i I*

+

*Vi VS Zi*

**~**

−

−

**FIGURE 2.6** Replacing the transmission line with its input impedance.

From the perspective of the source, the transmission line and load could be replaced with a load of impedance *Zi* [given by (2.81)], as shown in Figure 2.6. This equivalent circuit gives the same input current *Ii* and input voltage *Vi* as if we had the transmission line and load there.

***2.3.4.2 Characterizing Transmission Lines with Reflected Waves*** Define *voltage reflection coefficient* of the load impedance *ZL* (it changes depend ing on the load attached) as the ratio of the complex amplitudes of the reflected and incident voltage waves at the load. It can be shown that

= ||*ejθ* = *ZL* − *Z*0

*ZL* + *Z*0(2.83)

Notice that = 0 when *ZL* = *Z*0. In general, is a complex number with magnitude 1 or less. Meanwhile, the current reflection coefficient is the negative of the voltage reflection coefficient.

Then we can define the SWR (or VSWR) as

*S* = |*V*max|

|*V*min| = 1 + ||

1 − ||= |*I*max|

|*I*min| (2.84)

Meanwhile, the inverse relationship is

|| =*S* − 1

*S* + 1 (2.85)

To help with the visualization of standing waves, we plot the standing waves (in a specific case of resistive termination and lossless line that we describe next) in Figure 2.7 as a function of distance from the load. Both the voltage and current are plotted. As can be seen, the voltage SWR and current SWR are the same, and this is also true in general. VSWR is often expressed as a ratio (e.g., *S* = 1 is expressed as 1 : 1, *S* = 1*.*5 as 1*.*5 : 1, etc.).

**54** INTRODUCTION TO RADIO FREQUENCY, ANTENNAS, AND PROPAGATION

*V* or *I* amplitude of

3.5

|*V*(*y*)|

|*I*(*y*)|

3

2.5

2

1.5

1

0.5

0

0 λ/4 λ/2 3λ/4 λ *y* (distance from load)

**FIGURE 2.7** Standing waves in a transmission line.

*Resistive Termination.* For the case of resistive termination and lossless line, we have real, given by

= *ZL* − *Z*0

*ZL* + *Z*0(2.86)

where both *ZL* and *Z*0 are real. Important special cases include:

• *ZL* = *Z*0 (matched load), = 0, *S* = 1.

• *ZL* = 0 (short circuit), = −1, *S* → ∞.

• *ZL* → ∞ (open circuit), = 1, *S* → ∞.

In general, we try to achieve matched load conditions, or close to matched load conditions, to keep the VSWR low, since the higher the VSWR, the more power is lost. For example, in connecting an RF transmitter/receiver to an antenna with an RF cable, we might try to achieve a VSWR *<*2 : 1.

Figure 2.7 shows an example with resistive termination and lossless line, where *ZL > Z*0.

***2.3.4.3 SWR Summary*** The SWR is a very important value for practical pur poses. A low SWR is desirable to minimize loss of signal power in cables, such as those between RF equipment and antenna. The SWR depends on:

• The cable length

• Impedance matching

• Losses (from the resistance and conductance of the transmission line)

TIME-VARYING SITUATIONS, ELECTROMAGNETIC WAVES, AND TRANSMISSION LINES **55**

As a practical matter, because all lines have losses, it is best to measure SWR near the receiving side (e.g., near the antenna). Loss factors will attenuate the reflected wave so that if SWR is measured near the transmitting side, the reflected wave would be most attenuated and the transmitted wave least attenuated there. Thus, the SWR measured there may be artificially low.

**2.3.5 S-Parameters**

We have seen a number of concepts that can be considered part of a more general, abstract concept. A transmission line, for example, was assumed implicitly to have two interfaces: the input interface and the output interface. These can be called *ports*, and hence the transmission line is an example of a *two-port network*. In general, a port can be said to be a point where current enters or exits an electronic network. Figure 2.8 shows a two-port network.

A *scattering parameter* (or S-parameter, for short), *Smn*, is the ratio of voltage out of port *m* to voltage into port *n*, when all unused ports are attached to matched loads (matched to the system impedance of the network). For a two-port network, *m* and *n* take values 1 or 2. The S-parameters for a two-port network are illustrated in Figure 2.9. Thus, *S*11 is the reflection coefficient under matched load conditions (with port 2 being attached to a matched load). If the network represents an amplifier, port 1 is the input, port 2 is the output, and the ports are attached to the appropriate loads, then *S*21 is the gain. Note that the S-parameters may depend on frequency, temperature, control voltage, and so on, so these should be specified as necessary. It

•

*P*in

*Pr*

Component

*Pt*

•

• • **FIGURE 2.8** Two-port network.

*S*21

•

•

2- Port

*S*11 *S*22

Port 1 Port 2

network

•

•

*S*12

**FIGURE 2.9** S parameters.

**56** INTRODUCTION TO RADIO FREQUENCY, ANTENNAS, AND PROPAGATION

is possible to generalize from two-port networks to four-port networks and networks with other numbers of ports. In those cases, *m* and *n* would range over the number of ports.

**2.4 IMPEDANCE**

In this chapter and Chapter 1, we have seen a variety of notions of impedance:

• (Electrical) impedance

• (Intrinsic) impedance

• (Wave) impedance

• (Characteristic) impedance

• Input and output impedance

They are all measured in ohms, and unfortunately all are called impedance, but they refer to different concepts. Nevertheless, they can all take complex values and are ratios of phasors.

*Electrical impedance* applies in electrical circuits, where we have a voltage across two points and current flow across the same two points. Then electrical impedance is the complex-valued ratio of the phasors *V* and *I*, *Z* = *V/I*.

*Intrinsic impedance* is a property of a medium (e.g., air) being simply related to and *μ* of a medium, as *η* = √*μ/*. It gives us the ratio of the electric and magnetic fields when there is a single wave traveling in one direction in that medium. It is represented most often by *η*, or *η*0 for a specific medium. The intrinsic impedance of air, for example, is about 377 , whereas the electrical impedance between two points separated by air may be on the order of thousands of ohms (assuming that the voltage is large enough to exceed the dielectric breakdown voltage).

*Wave impedance* is the ratio of the transverse component of the electric and mag netic fields, **E** and **H**, respectively. For a single wave traveling in one direction, wave impedance equals the intrinsic impedance everywhere (with the possible exception of a sign; in some variations of the definition, wave impedance may equal *η* for a single wave traveling in the +*z* direction and −*η* for a single wave traveling in the −*z* direction). It can be different from the intrinsic impedance (e.g., when we have two waves traveling in opposite directions). In that case, wave impedance would also be a function of position. Cheng [3] calls it *wave impedance of the total field*, which helps distinguish it from intrinsic impedance.

*Characteristic impedance*, usually written as *Z*0, is a convenient single-parameter characterization of a transmission line. Characteristic impedance should be used only to refer to a transmission line. It is the ratio of *V* to *I* measured at the input of the transmission line if there is no reflected wave. Thus, it differs from intrinsic impedance and wave impedance, which have to do with field strengths. Also, it differs from electrical impedance in that characteristic impedance equals electrical impedance

TESTS AND MEASUREMENTS **57**

(of the transmission line and load combined) only when the load is matched to the transmission line. Otherwise, the relationship between *Z*0 and electrical impedance, *Zi*, is given by (2.81). The electrical impedance *Zi* in this type of context is also called the input impedance.

*Input impedance* of a circuit or device has to do with the electrical impedance seen looking into it (i.e., at its input). More precisely, for a circuit or device, it is the Thevenin equivalent impedance at the input. For a transmission line, it is the ratio of ´ the resultant *V* (from forward and reflected waves) and *I* (from forward and reflected waves), which can be expressed in terms of *Z*0 and *ZL* as given by (2.81).

Unfortunately, there are some cases where these careful distinctions are not made; for example, “characteristic impedance” is used to refer to intrinsic impedance or wave impedance outside transmission lines. Hence, the reader should read everything with caution and use context to help clarify meaning.

**2.5 TESTS AND MEASUREMENTS**

For testing and validation purposes in RF, antennas, and propagation, various devices and tools are available.

An *oscilloscope* allows electrical signals to be viewed in the time domain (e.g., as a plot of voltage against time). Whereas an oscilloscope show a time domain representation of signals, a *spectrum analyzer* shows a corresponding frequency-domain representation of signals. Spectrum analyzers can be used for many types of measurements, making them especially useful for RF testing. Spec trum analyzers are used in RF engineering for harmonic distortion measurement, intermodulation distortion measurement, measurement of modulation sidebands, and so on.

A *network analyzer* measures the characteristics of a device, system, or network. This is in contrast to a spectrum analyzer or an oscilloscope, both of which measure and analyze a *signal*. (Of course, such measurements are not disjoint from what a network analyzer does in the sense that related or similar information about a device, system, or network could be obtained; for example, the frequency response of a device, system, or network could be deduced after measuring an input spectrum and its corresponding output spectrum using a spectrum analyzer.)

A *time-domain reflectometer* (TDR), as the name suggests, sends a short (time domain) pulse into a cable, device, or system and measures the reflections, if any. The amplitude, duration, and shape of the reflected wave gives information about the length of the cable, its characteristic impedance, and so on.

Further details on oscilloscopes, spectrum analyzers, network analyzers, and TDRs are provided in Section 2.5.2.

**2.5.1 Function Generators**

It is often useful in testing to be able to generate various functions (e.g., sine waves, square waves) to be used as system inputs. Basic sine-wave oscillators may

**58** INTRODUCTION TO RADIO FREQUENCY, ANTENNAS, AND PROPAGATION

generate only sine waves (and perhaps square waves, too), whereas a function gen erator might have additional features (e.g., also generating other waveforms, such as triangular waves and modulated waves, frequency sweep, and dc offset adjust ment). With frequency sweep, the instantaneous frequency of the wave will change with time, sweeping through a range of frequencies (e.g., linearly or logarithmically with time).

More sophisticated function generators, sometimes called *arbitrary waveform gen erators*, can generate arbitrary waveforms. Usually, this means that waveforms are generated using direct digital synthesis. Thus, any arbitrary waveform can be stored digitally and the waveform is synthesized using digital-to-analog conversion followed by lowpass filtering and amplification.

While a function generator is general-purpose and has an upper limit in the tens of MHz, other, more specialized generators are also used in testing labs, including pulse generators and RF signal generators. A *pulse generator*specializes in producing pulses and square waves of high precision and high quality. Since such pulses would have a very wide bandwidth, they would not be produced as cleanly by a regular function generator as by a pulse generator. It may range up to 1 GHz.

The signals produced by a general-purpose function generator are too low in fre quency for RF testing, so RF signal generators are used that can produce signals from several kilohertz to several gigahertz. Besides producing sinusoids in the RF range, they may also provide modulated signals, including various digital modulation for mats. Usually, these generators need very accurate and stable frequency generation for obvious reasons (e.g., for testing adjacent channel rejection of receivers, phase noise and inaccurate frequencies from the signal generator will corrupt the results). Moreover, the RF signal generators need to be capable of producing very low side bands. To test receiver sensitivity, the amplitude of the signal produced also needs to be very accurate.

**2.5.2 Measurement Instruments**

Measurement instruments always affect the circuit they are measuring, however slightly. To minimize the impact on the circuit they are measuring, many measure ment instruments are designed with a very high input impedance. This draws as little current as possible (less loading). Furthermore, this maximizes the voltage transfer to the measurement instrument, making readings of open-circuit voltage more accurate.

At RF, components are usually matched. Input and output impedances are both 50 (or 75 in some cases). For such systems, the measurement instruments would also use a matching input impedance of 50 . Some signals can be interpreted as mixed ac and dc signals, with an ac signal being offset by a dc value. An *ac coupled* device only gives the ac portion of a signal, removing the dc offset through the use of a coupling capacitor at the input.

Each instrument has only finite resolution and finite accuracy. Resolution has to do with the granularity of changes in the measured value that can be detected, whereas accuracy has to do with how close to the correct value the measurement is. The resolution and accuracy of the instrument should be considered to see if it is

TESTS AND MEASUREMENTS **59**

**FIGURE 2.10** RF probe. (Courtesy of Aeroflex Inc.)

appropriate for its intended use. The bandwidth and rise time of the measurement instruments should be considered relative to the particular signal to be measured.

***2.5.2.1 Voltmeters, Multimeters, and RF Probes*** Voltmeters can be either dc or ac. Although ac voltmeters normally give readings as root mean square (rms), there are a couple of ways these measurements could be made, so in some ac volt meters, the readings are calibrated correctly only for sinusoids. So-called “true rms” ac voltmeters give correct rms readings for nonsinusoidal waves as well, but the range of frequencies over which the readings are accurate is always finite. As an alternative to the use of ac voltmeters, high-frequency ac measurements can be made with an *RF probe* together with a dc voltmeter. Usually, the RF probe is a peak detector, and so may be calibrated only for sine waves. Figure 2.10 shows an RF probe. *Multimeters* are popular because they combine voltmeter, ammeter, and ohmmeter. They may also include a few other useful features, such as a frequency counter (often limited in bandwidth).

***2.5.2.2 Oscilloscopes*** Oscilloscopes are very versatile. They can be used to display *eye diagrams* of digitally modulated signals. An eye diagram is a visual tool used to detect the impact of intersymbol interference, noise, and so on, on the quality of a digital signal.

***2.5.2.3 Frequency Counters*** The frequency of a periodic signal is how often it repeats itself, so it is natural to expect that measurement of frequency could be accomplished by something as simple as counting. In a frequency counter, a time base controls the opening of a gate, during which time cycles of the signal are counted. Alternatively, especially for low frequencies, the time base and the signal being mea sured could be swapped internally by the frequency counter so that the signal being measured controls the gate, and then the counting will be counting the number of cycles of the timebase that occur during one period of the signal. Hence, the period of

**60** INTRODUCTION TO RADIO FREQUENCY, ANTENNAS, AND PROPAGATION

the signal is measured. While period and frequency are reciprocals of each other, so that obtaining either one is sufficient, more accuracy is obtained at lower frequencies counting the period and at higher frequencies counting the cycles, of the signal being measured. Usually, the time base comes from a crystal oscillator, so it is essential that it be very stable and accurate.

***2.5.2.4 Spectrum Analyzers*** Spectrum analyzers show the amplitude of a sig nal as a function of frequency rather than time. A straightforward first attempt at building a spectrum analyzer might be to use a filter bank of relatively narrow filters. Each of these filters would filter out a different narrowband from all the other filters, where together the filters will span a desired range of frequencies. However, it is not a practical approach, since many filters would be needed for most cases. For example, if the desired range of frequencies is 0 to 2 MHz and each filter is 1 kHz wide, 2000 of these filters would be needed.

Practical spectrum analyzers use methods that do not require thousands of filters. One method is to use the fast Fourier transform (FFT) to obtain the spectrum. To obtain the spectrum of a continuous-time signal, the signal would first have to be sampled. Unlike with the filter-bank approach, this approach is susceptible to aliasing, so the FFT should be preceded by an antialiasing lowpass filter. Another method is to design the spectrum analyzer like a heterodyne radio receiver, with a high-quality fixed IF filter taking input from a mixer that mixes the signal being analyzed with the output of a tunable local oscillator. When such an approach is taken and the tunable local oscillator is swept automatically across a range of desirable frequencies, the spectrum analyzer may be known as a *swept spectrum analyzer*.

A fundamental property of a spectrum analyzer is the resolution bandwidth. If there are multiple distinct spectral components within the resolution bandwidth at a particular band, the spectrum analyzer will not be able to resolve them into dis tinct components. Making the resolution bandwidth narrower allows the spectrum analyzer to resolve components to a finer level and also reduces the noise in the measurements (since the noise equivalent bandwidth would be smaller). However, narrower resolution bandwidth comes with a longer settling time.

Spectrum analyzers come in different shapes and sizes, and there are even hand held spectrum analyzers that can be brought out into the field, for example, to make measurements at wireless base stations. An example of a handheld spectrum analyzer is shown in Figure 2.11.

***2.5.2.5 Network Analyzers and TDRs*** Network analyzers can be used to measure the S-parameters of a two-port network (Section 2.3.5). A *scalar network analyzer* measures amplitude only and not phase, whereas a *vector network analyzer* also measures phase. A network analyzer can generate its own signal to input to the system, or take an external signal for that purpose.

Both TDRs and network analyzers can be used to obtain frequency response. In the case of network analyzers, the inputs are often narrowband (but the frequency may be swept over a range). In the case of the TDR, it sends a very narrow pulse, and the frequency response is obtained by computing the Fourier transform of the reflected

TESTS AND MEASUREMENTS **61**

**FIGURE 2.11** Handheld spectrum analyzer. (Courtesy of Aeroflex Inc.)

signal. The narrow pulse will have a finite width, say, 10 ps, so in the frequency domain the impulse response is multiplied by the Fourier transform of the narrow pulse, which would be on the order of 100 GHz.

***2.5.2.6 Antenna Couplers*** It can be difficult in some cases to measure the RF signal directly from a transmitter to its antenna. For example, the antenna might be integrated into a mobile phone together with the transmitter and receiver. An alternative approach, then, is to use an *antenna coupler*to measure the signal indirectly as it is being radiated from the antenna. The antenna coupler may be a broadband antenna that is placed very close to the device under test.

**2.5.3 Mobile Phone Test Equipment**

So far, we have discussed oscilloscopes, spectrum analyzers, network analyzers, time domain reflectometers, and so on, that have wide applicability to many different application scenarios in electrical engineering and electronics. There are also devices such as SWR meters and antenna couplers that might have a narrower range of appli cations. However, even these are not specific to any wireless system standard. In contrast, there is also an entire range of test and measurement devices that incorpo rate specifics about various wireless systems (e.g., GSM, CDMA, LTE, etc.) to enable more specific tests and measurements to be performed on devices for those particular wireless systems.

These test and measurement devices can be very useful in determining if a mobile phone, for example, is in conformance with specific system specifications on spectrum mask, sensitivity, selectivity, and so on, at the RF level. Some of these devices can also test and measure at other layers and are related to other aspects of the systems, such as BER performance, radio link protocols, and network protocols. Thus, these test and measurement devices can be useful for certifying phones and equipment in the wireless infrastructure and also for repair purposes. An example of one such mobile phone test equipment is shown in Figure 2.12.

**62** INTRODUCTION TO RADIO FREQUENCY, ANTENNAS, AND PROPAGATION

**FIGURE 2.12** Example of mobile phone test equipment. (Courtesy of Aeroflex Inc.) **EXERCISES**

**2.1** Observe the geometry in Figures 2.1 and 2.2. Given a point (*x, y, z*) in Cartesian coordinates, what are the cylindrical coordinates of the same point? And the spherical coordinates?

**2.2** Now work the other way around and convert from a point in cylindrical coor dinates to Cartesian coordinates, and from spherical coordinates to Cartesian coordinates.

**2.3** Consider an electromagnetic wave propagating in a lossless medium. Suppose that at a point P, the **E** field is given by **E** = **ux***E*0 and the **H** field by **H** = **uy***H*0. In what direction is the wave propagating? If *E*0 = 377 mV/m and the medium is air, what is *H*0? What is the Poynting vector? What is the average power flow per unit area at P?

**2.4** In general, what is the range of possible values for SWR, *S*? What would be the corresponding range of values for ||? Referring to Figure 2.7, what is the SWR? What is ?

**REFERENCES**

1. K. Chang. *RF and Microwave Wireless Systems*. Wiley, Hoboken, NJ, 2000. 2. K. Chang, I. Bahl, and V. Nair. *RF and Microwave Circuit and Component Design for Wireless Systems*. Wiley, Hoboken, NJ, 2002.

3. D. K. Cheng. *Fields and Wave Electromagnetics*. Addison-Wesley, Reading, MA, 1990. 4. S. Ramo, J. Whinnery, and T. Van Duzer. *Fields and Waves in Communication Electronics*. Wiley, New York, 1984.

5. M. Sadiku. *Elements of Electromagnetics*. Oxford University Press, New York, 2006.