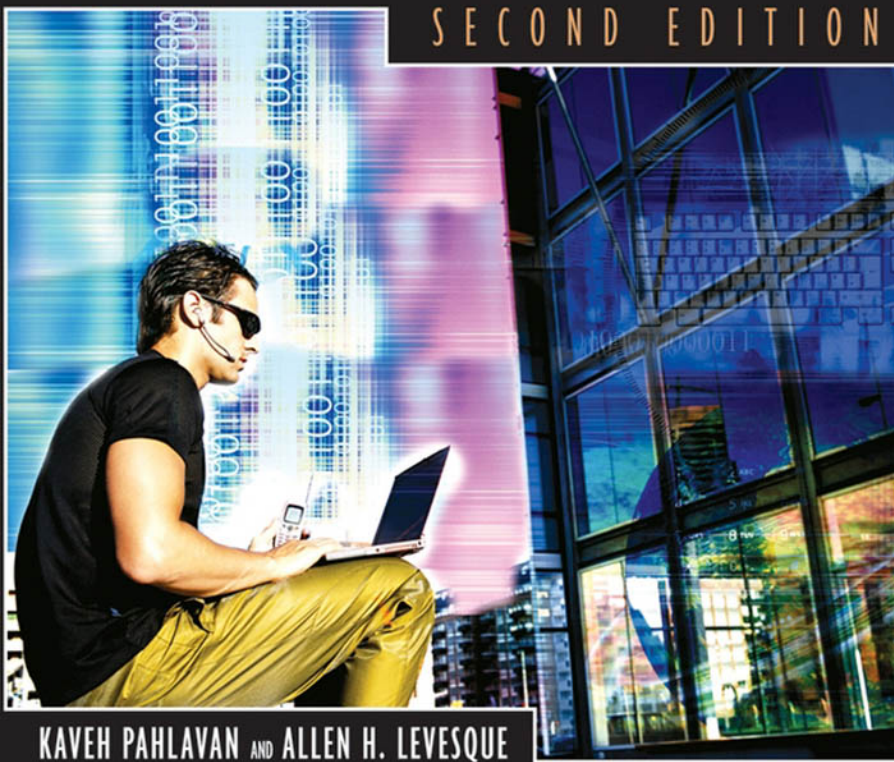


WIRELESS INFORMATION NETWORKS

SECOND EDITION



KAVEH PAHLAVAN AND ALLEN H. LEVESQUE

WIRELESS INFORMATION NETWORKS

WIRELESS INFORMATION NETWORKS

Second Edition

KAVEH PAHLAVAN
ALLEN H. LEVESQUE



A JOHN WILEY & SONS, INC., PUBLICATION

Copyright © 2005 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: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

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:

Pahlavan, Kaveh, 1951–

Wireless information networks / by Kaveh Pahlavan and Allen H. Levesque.—2nd ed.
p. cm.

“A Wiley-Interscience publication.”

ISBN-13 978-0-471-72542-8 (cloth)

ISBN-10 0-471-72542-0 (cloth)

1. Wireless communication systems. I. Levesque, Allen H. II. Title.

TK5 103.2.P34 2005

621.382—dc22

2005041792

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

*To those from whom we learned,
to those we taught, and
to those we love*

CONTENTS

Preface	xi
PART I INTRODUCTION TO WIRELESS NETWORKS	1
1 Overview of Wireless Networks	3
1.1 Introduction, 3	
1.2 Network Architecture and Design Issues, 6	
1.3 Key Trends in Wireless Networking, 20	
1.4 Outline of the Book, 21	
Questions, 22	
2 Evolution of the Wireless Industry	23
2.1 Introduction, 23	
2.2 Three Views of the Wireless Industry, 29	
2.3 Three Generations of Cellular Networks, 32	
2.4 Trends in Wireless Technologies, 43	
Questions, 49	
PART II CHARACTERISTICS OF RADIO PROPAGATION	51
3 Characterization of Radio Propagation	53
3.1 Introduction, 53	
3.2 Multipath Fading and the Distance–Power Relationship, 55	
3.3 Local Movements and Doppler Shift, 64	
3.4 Multipath for Wideband Signals, 66	
3.5 Classical Uncorrelated Scattering Model, 72	
3.6 Indoor and Urban Radio Propagation Modeling, 81	
Questions, 86	
Problems, 87	
Projects, 89	
4 Modeling and Simulation of Narrowband Signal Characteristics	93
4.1 Introduction, 93	

4.2	Modeling Path Loss and Slow Shadow Fading, 96	
4.3	Doppler Spectrum of Fast Envelope Fading, 110	
4.4	Statistical Behavior of Fast Envelope Fading, 122	
4.5	Simulation of Fast Envelope Fading, 126	
	Questions, 133	
	Problems, 134	
	Projects, 137	
5	Measurement of Wideband and UWB Channel Characteristics	149
5.1	Introduction, 149	
5.2	Time-Domain Measurement Techniques, 151	
5.3	Frequency-Domain Measurement Techniques, 171	
5.4	Advances in Frequency-Domain Channel Measurement, 180	
	Questions, 197	
	Problems, 198	
	Project, 200	
6	Modeling of Wideband Radio Channel Characteristics	205
6.1	Introduction, 206	
6.2	Wideband Time-Domain Statistical Modeling, 208	
6.3	Wideband Frequency-Domain Channel Modeling, 234	
6.4	Comparison Between Statistical Models, 243	
6.5	Ray-Tracing Algorithms, 245	
6.6	Direct Solution of Radio Propagation Equations, 261	
6.7	Comparison of Deterministic and Statistical Modeling, 263	
6.8	Site-Specific Statistical Model, 265	
	Appendix 6A: GSM-Recommended Multipath Propagation Models, 270	
	Appendix 6B: Wideband Multipath Propagation Models, 272	
	Questions, 274	
	Problems, 275	
	Projects, 277	
PART III	MODEM DESIGN	279
7	Narrowband Modem Technology	281
7.1	Introduction, 282	
7.2	Basic Modulation Techniques, 284	
7.3	Theoretical Limits and Practical Impairments, 307	
7.4	Traditional Modems for Wide-Area Wireless Networks, 312	
7.5	Other Aspects of Modem Implementation, 328	
	Questions, 335	
	Problems, 336	
	Projects, 338	
8	Fading, Diversity, and Coding	341
8.1	Introduction, 341	

8.2	Radio Communication on Flat Rayleigh Fading Channels, 343	
8.3	Diversity Combining, 347	
8.4	Error-Control Coding for Wireless Channels, 353	
8.5	Space-Time Coding, 363	
8.6	MIMO and STC, 365	
	Questions, 372	
	Problems, 372	
	Projects, 374	
9	Broadband Modem Technologies	377
9.1	Introduction, 378	
9.2	Effects of Frequency-Selective Multipath Fading, 380	
9.3	Discrete Multipath Fading Channel Model, 384	
9.4	Adaptive Discrete Matched Filter, 389	
9.5	Adaptive Equalization, 393	
9.6	Sectorized Antennas, 405	
9.7	Multicarrier, OFDM, and Frequency Diversity, 411	
9.8	Comparison of Traditional Broadband Modems, 421	
9.9	MIMO in Frequency-Selective Fading, 423	
	Appendix 9A: Analysis of the Equalizers, 425	
	Questions, 428	
	Problems, 429	
	Projects, 431	
10	Spread-Spectrum and CDMA Technology	435
10.1	Introduction, 435	
10.2	Principles of Frequency-Hopping Spread Spectrum, 439	
10.3	Principles of Direct-Sequence Spread Spectrum, 444	
10.4	Interference in Spread-Spectrum Systems, 464	
10.5	Performance of CDMA Systems, 476	
	Questions, 494	
	Problems, 495	
PART IV	SYSTEMS ASPECTS	499
11	Topology, Medium Access, and Performance	501
11.1	Introduction, 501	
11.2	Topologies for Local Networks, 503	
11.3	Cellular Topology for Wide-Area Networks, 506	
11.4	Centrally Controlled Assigned Access Methods, 521	
11.5	Distributed Contention-Based Access Control, 537	
	Questions, 572	
	Problems, 573	
	Project, 576	
12	Ultrawideband Communications	581
12.1	Introduction, 581	

12.2	UWB Channel Characteristics, 584	
12.3	Impulse Radio and Time-Hopping Access, 589	
12.4	Direct-Sequence UWB, 595	
12.5	Multiband OFDM, 599	
	Questions, 603	
	Problems, 604	
13	RF Location Sensing	607
13.1	Introduction, 607	
13.2	RF Location-Sensing Techniques, 611	
13.3	Modeling The Behavior of RF Sensors, 619	
13.4	Wireless Positioning Algorithms, 626	
	Questions, 636	
	Problems, 637	
14	Wireless Optical Networks	639
14.1	Introduction, 639	
14.2	Implementation, 641	
14.3	Eye Safety, 643	
14.4	IR Channel Characterization and Data-Rate Limitations, 644	
14.5	Modulation Techniques for Optical Communications, 653	
14.6	Multiple Access and Data Rate, 659	
	Questions, 661	
15	Systems and Standards	663
15.1	Introduction, 663	
15.2	GSM, GPRS, and EDGE, 664	
15.3	CDMA and HDR, 674	
15.4	Other Historical Systems, 679	
15.5	Wireless LANs, 682	
15.6	Speech Coding in Wireless Systems, 685	
	Questions, 687	
	References	689
	Index	713
	About the Authors	721

PREFACE

The first edition of this book, published in 1995, was the first textbook to provide a comprehensive introduction to the field of wireless information networks. That book presented wireless networking as the enabling communications technology of the 1990s and beyond. Now, only a decade later, mobile and portable telephones and wireless data services are a familiar part of our daily lives, as the twenty-first century witnesses widespread deployment of wireless networks, which has revolutionized the concept of communication and information processing for business, professional, and private applications. The field of wireless communications continues to experience unprecedented market growth, as evidenced by over 1.5 billion cellular telephone subscribers worldwide and the rapid increase in the size of the wireless local area network market for office, home, and public access applications. The initial growth in the market for second-generation cellular products and services spurred important new initiatives toward the development and deployment of third-generation cellular networks. More recently, attention has been focused on location-aware broadband ad hoc wireless networks as the foundation for the next generation of wireless networking technology, which is expected to enable systems of geographically dispersed sensors. The emerging wireless sensor and ad hoc networks are expected to interconnect numerous terminals with a variety of data-rate requirements with traditional multimedia Internet networks to create a worldwide communication medium among RFID tags, a variety of sensors, home appliances, and small robotic devices.

These developments were all part of a major paradigm shift in the world of telecommunication, a shift away from nearly exclusive reliance on wired networks to an era of “tetherless” communications based largely on wireless technology, and a shift in the computer industry toward integration of high-performance distributed computing and portable devices in a pervasive mobile computing environment. We adopted the title *Wireless Information Networks* in 1995 as an encompassing name intended to include all applications related to evolving wireless networks in the telecommunication and computer industries, and that book provided comprehensive coverage of the signal processing and system engineering aspects of this field. Given the tremendous growth of the industry in both its signal processing and systems engineering aspects, it becomes increasingly difficult to treat all the important topics in a single volume. With that in mind, in 2002 the lead author published a new book, *Principles of Wireless Networks—A Unified Approach*, coauthored by Prashant Krishnamurthy, which is more focused on systems engineering aspects, and began preparing this second edition of the original book, with more emphasis on signal processing topics. The objective of

Principles of Wireless Networks is to provide a systems engineering treatment that can be taught to both electrical and computer engineering (ECE) and computer science students in undergraduate or first-year graduate courses. This second edition of *Wireless Information Networks* places greater emphasis on signal processing and is more suitable for ECE graduate students and possibly senior undergraduate students.

At the emergence of the wireless industry in late 1980s and early 1990s, all telecommunication enterprises that were involved in traditional wired communications services and product development made major investments in wireless technology. Computer companies invested in wireless communications to add mobile computing and ad hoc networking features to the laptop, handheld, notepad, and other portable computing devices that are coming into increasingly pervasive use. Later, large corporations, as end users, included wireless components in their network infrastructures to extend the accessibility of their networks to their traveling personnel. Military agencies developed location-aware ad hoc sensor networks for use in tactical environments, as well as portable devices that place a large amount of computational power in the hands of the foot soldier operating in urban fighting scenarios. Today, almost all companies in engineering disciplines other than telecommunications are entering the wireless communications business, for applications such as in-vehicle networking or home networking, and are now a part of this wireless revolution. All of this means that there are a great many engineers, computer science specialists, and managers with a variety of interests who are faced with having to educate themselves in this area. This major new emphasis on wireless communications has also spurred a renewed emphasis on the teaching of principles of wireless communications in colleges and universities. This second edition is designed to provide students, engineers, and scientists with an introduction to the major signal processing aspects of wireless networks.

The book is written from a systems engineering perspective, by which we mean that the various technical topics are presented in the context of ongoing development of specific new systems and services, as well as key recent developments in national and international spectrum allocations and standards. Our method of presentation is to organize the myriad of emerging wireless technologies into logical categories that reflect the variety of perspectives that users have toward different networks and services. The book addresses the major segments of wireless technology: first-, second-, and third-generation wide-area cellular networks, wireless local area networks (WLANs), and wireless personal area networks (WPANs), with special attention to the emerging location-aware broadband wireless sensor and ad hoc networks. Although the book covers technology applicable to a wide range of wireless systems, as in the first edition, particular attention is given to indoor wireless communications, an area that is not treated in great depth in most other books.

In writing the book, we have endeavored to bring together treatments of all the major topics to be considered in the design of wireless information networks, but have avoided the presentation of detailed mathematical derivations that are available in other texts. In each instance, we have tried to provide the motivation for various wireless system design choices in the context of overall system considerations. We believe that this is an effective approach to training systems engineers, who should have an overall perspective of an entire system as well as a working knowledge of how to apply the results of specific research to an engineering problem.

The first edition of the book has been used as a graduate-level textbook in universities throughout the world. It has also been used as a reference book for indoor

radio communication research programs at DARPA and the National Science Foundation and for indoor radio channel modeling for WLAN and WPAN standardization activities such as IEEE 802.11n and IEEE 802.15.3. Being the first comprehensive textbook on wireless networks, the first edition has served many of today's leading researchers as a key resource in gaining a comprehensive understanding of the important issues related to wireless communications. As a result, in this second edition we have tried to maintain the comprehensive treatment and have included major technical developments that have emerged since publication of the first edition, such as ultrawideband communications, wireless positioning, space-time coding, multiple-input multiple-output antennas, orthogonal frequency-domain multiplexing, interference cancellation, and multiuser detection. Therefore, we have increased the number of chapters from 12 to 15 to include ultrawideband communications and wireless positioning in two new chapters and to expand the modem design chapters from three to four. Some readers of the first edition indicated that certain problem sets were overly difficult. Thus, in this edition we have added a number of simpler problems and have turned some of the more difficult ones into projects with expanded explanations, to make them easier to understand. As in the first edition, the questionnaire format is used to emphasize the importance of having a general understanding of the overall system at hand and of the rationale behind key engineering design choices. The traditional problem sets are exercises for derivation and understanding of the detailed mathematical analysis of various concepts. Projects provide more detailed exercises, usually involving computer simulations or extensive analysis of data. We have directed these problems and projects toward application-oriented issues. This approach provides students with an understanding of the issues, motivates them to use the computer as a tool in the learning process, and shifts their viewpoint toward real-world engineering problems rather than mathematical drills. We believe that this approach is essential for the proper training of engineers for productive careers in the market-driven telecommunication industry, where simple ideas and added features will often generate greater revenues than will the latest technical inventions.

This edition of the book covers four categories of topics, organized into four parts. In Part I of the book, Chapters 1 and 2, we provide an overview of major categories of wireless communications and outline the user and market perspectives toward various wireless systems and services. Then we review briefly the current state of development of wireless and mobile communications systems, including the important issues of spectrum administration and standards.

In Part II, Chapters 3 through 6, we describe the characteristics of radio propagation, as well as measurement and simulation methods used in evaluating existing third-generation cellular, WLAN, and WPAN systems and emerging location-aware broadband wireless ad hoc networks. We provide a detailed description of time- and frequency-domain statistical channel modeling techniques and their application to popular standards such as GSM, IEEE 802.11, and IEEE 802.15. We also describe the ray-tracing algorithm and give a brief overview of direct solution of the radio propagation equations.

In Part III, Chapters 7 through 10, we discuss wireless modem design technologies. We begin in Chapter 7 with a description of traditional narrowband modem technologies and issues arising in their application to radio channels. In Chapter 8 we address fading, diversity, and coding in relation to the analysis and performance evaluation of wireless modems. In this chapter we also introduce the concepts of

multiple-input multiple-output and space-time coding, which are gaining considerable attention as techniques to be used in modem design. Chapter 9 is devoted to broadband modem technologies, including equalization techniques, smart antenna techniques, and orthogonal frequency-division multiplexing (OFDM). Chapter 10 is devoted to spread-spectrum techniques and code-division multiple-access techniques for direct-sequence and frequency-hopping systems.

Part IV, Chapters 11 through 15, is devoted to network access and system aspects. Chapter 11 treats network access methods and provides a comprehensive description of voice-oriented assigned access and data-oriented random access techniques and discusses performance evaluation methods as well. Chapter 12 is devoted to ultra-wideband (UWB) communications. In this chapter we discuss the detailed behavior of UWB channels and describe impulse radio, multiband OFDM, and direct-sequence UWB techniques being considered by IEEE 802.15 for the next generation of WPAN applications. Chapter 13 is devoted to RF location-sensing techniques, the foundation for wireless positioning and indoor geolocation science. We discuss RF channel behavior in the context of positioning applications and describe the popular received signal strength systems used in WLAN positioning as well as the time-of-arrival techniques used for more accurate positioning. These chapters provide a comprehensive understanding of the emerging technologies for implementation of wireless sensor and ad hoc networks. Chapter 14 is the same as Chapter 10 of the first edition and provides the principles of infrared communications. Chapter 15, devoted to systems and standards, is a revision of the concluding chapter of the first edition.

The book can be used in its entirety for a first- or second-year graduate course in wireless communications networks in electrical and computer engineering curricula. As preparation for such a course, students should have an understanding of the elements of probabilistic signal and system analysis and some background in the principles of modulation and coding. This material is taught at the Worcester Polytechnic Institute (WPI) and the University of Oulu in Finland as a 14-week course meeting three hours per week. The first two chapters are taught in the first week, Chapters 3 to 6 in the next five weeks, Chapters 7 to 10 in four weeks, and Chapters 12 and 13 in two weeks. The remaining two weeks can be spent on other topics selected by the instructor or on student presentations, followed by course exams. Weekly assignments include answering selected questions and solving a few selected problems at the end of the chapter. In addition, students are asked to do three of the projects throughout the course. The material in some chapters is covered completely, whereas material in other chapters is covered with more emphasis on the concepts and less emphasis on the details of mathematical derivations. To cover all chapters of the book in full detail, a two-semester course format is advisable, although most of the material might be covered in a fast-paced one-semester course with selective omission of the more specialized topics. With appropriate selection of topics, the book can also be used at the undergraduate level. An extensive list of references is included, which will be especially helpful to the individual reader using the book for self-study or reference purposes.

Much of the material in the first edition and part of the material in the second edition was drawn from the published work of the lead author and his students in the Center for Wireless Information Network Studies (CWINS) at WPI. We are pleased to acknowledge students' contributions to advancing the understanding of wireless channels and networks. In particular, we thank Dr. Steven Howard, Dr. Rajamani Ganesh, Dr. Ker Zhang, Dr. Ganning Yang, Dr. Thomas Sexton, Dr. Mitch Chase, Timothy

Holt, Dr. Aram Falsafi, Glen Bronson, Joseph Meditz, Dr. Mudhafar Hassan Ali, and Dr. Sheping Li, whose contributions were helpful in the preparation of the first edition of the book. In addition, we would like to express our appreciation to Dr. Prashant Krishnamurthy, Dr. Xinrong Li, Dr. Ali Zahedi, Jeff Feigin, Dr. Aram Falsafi, Dr. Robert Tingley, Hamid Hatami, Bardia Alavi, Emad Zand, Muzafer Kannan, Nayef Alsindi, Mohammad Heidari, Leon Teruo Metreud, and other recent affiliates of the CWINS Laboratory, as well as Dr. Mika Ylianttila and Juha-Pekka Makela of the University of Oulu, who have directly or indirectly helped the lead author to extend his knowledge in this field and shape his thoughts for preparation of the second edition of the book. We owe special thanks to the National Science Foundation, DARPA, and the United States Department of Defense, to TEKES, Nokia, Sonera, and the Finnish Air Force in Finland, and to many other companies and research organizations whose support of the CWINS program at WPI enables graduate students and the staff of CWINS to pursue continuing research in this important field. A substantial part of the new material in this second edition has flowed out of these sponsored research efforts.

Much of the writing of the lead author in this second edition was accomplished while teaching and carrying out research at the University of Oulu, Finland, as well as during his sabbatical at Olin College of Engineering, Needham, Massachusetts. He would like to express his deep appreciations to the University of Oulu, Olin College, and Worcester Polytechnic Institute for providing him with these opportunities. In particular, he thanks Professor Pentti Leppanen of the University of Oulu for his continual encouragement and creative administrative support, and Professor Matti-Latva-aho of the University of Oulu and Dr. Sassan Iraj of Nokia Research Center for fruitful discussions on current developments in spread-spectrum and CDMA technologies. He also thanks David Kerns, provost of Olin College, for providing him the opportunity to spend the fall 2004 semester at Olin. Also, he thanks Professor Fred Looft, head of the WPI ECE Department, and WPI provost John Carney for their support of a sabbatical leave for work on this second edition.

The lead author would also like to express his deep appreciation to Dr. Phillip Bello, Professor John Proakis, and Dr. Jerry Holsinger, through whom he has increased the depth of his understanding of the theory and practice of telecommunications, and to Professor James Matthews for introducing him to the field of radio communications.

His coauthor would like to express appreciation to his many colleagues at GTE (now Verizon) who during his industrial career helped in many ways in his work in mobile and cellular communications. He would also like to thank Professors John Orr and Fred Looft and provost John Carney for providing him with the opportunity for graduate teaching and participation in CWINS research activities at WPI.

Most of all, the authors are indebted to their families for their patience and support throughout this long and challenging project.

K. P.
A. H. L.

PART I

INTRODUCTION TO WIRELESS NETWORKS

Part I consists of two chapters that provide an introduction to wireless network technologies and standards. These chapters cover the market sectors and describe incentives for the use of wireless networks by the telecommunications and computer industries and the military.

Chapter 1: Overview of Wireless Networks

In this chapter we provide an overview of wireless information networks. We describe the basic elements of a wireless network and the key technical issues to be considered in the design of these networks. We also discuss the market sectors that constitute the wireless industry and the trends apparent in voice- and data-oriented networks. In the final section of the chapter we outline the remaining chapters of the book.

Chapter 2: Evolution of the Wireless Industry

In this chapter we consider the evolution of wireless networking technology, which has been built upon developments occurring not only in the telecommunications industry but also in the computer industry, as communications and computer technologies have drawn closer together. Many observers see the wireless industry as one that has integrated radio science with communications and computer technologies. Thus, before delving into details of radio propagation and signal processing—the primary focus of this book—it is useful to consider the wireless industry from the separate viewpoints of the telecommunications and computer industries, which is the approach we take in this chapter. We also discuss briefly the view of an important user community, military ground forces, which have had a long history of reliance on wireless communications networks.

1

OVERVIEW OF WIRELESS NETWORKS

- 1.1 Introduction
 - 1.1.1 Elements of a Wireless Network
 - 1.1.2 Key Technical Issues for Wireless Networks
 - 1.1.3 Four Market Sectors
- 1.2 Network Architecture and Design Issues
 - 1.2.1 Network Architectures
 - 1.2.2 Wireless Versus Wired Networks
 - 1.2.3 Elements of a Wireless Network Architecture
 - 1.2.4 Technical Aspects of a Wireless Infrastructure
 - 1.2.5 Technical Aspects of the Air Interface
- 1.3 Key Trends in Wireless Networking
 - 1.3.1 Voice-Oriented Networks
 - 1.3.2 Data-Oriented Networks
 - 1.3.3 Where Is the Complexity?
- 1.4 Outline of the Book
 - Questions

1.1 INTRODUCTION

The second half of the twentieth century witnessed enormous transformations in electronic communications, including the development of data transmission over legacy telephone networks, the introduction of packet-data networks, the development of high-speed local area networks (LANs), and the development of mobile wireless communications networks, most notably cellular networks, paging systems, and even mobile satellite systems. By the start of the current century, cellular and paging services had come into widespread use in support of business communications and personal communications as well. The early analog cellular networks were rapidly supplanted by digital networks affording increased traffic capacity and capable of supporting an expanding menu of data-oriented services. In this first decade of the new century, we

are seeing rapidly increasing interest in higher-rate forms of wireless data communication, including multimedia transmission to and from portable phones, and wireless access to the Internet from laptop computers. The technology underlying these wireless communications developments is the specific focus of this book.

The worldwide growth of the wireless communications industry has been truly phenomenal. At this writing, there are more than 1 billion cellular telephone users throughout the world, and the aggregate annual revenue of the wireless industry exceeds the revenues of the wired-telephone service industry. About 10 years ago, Internet access began expanding from the business environment to include the home environment, and this soon generated annual revenues comparable to those of traditional telephone service and wireless service. Currently, the *information exchange industry*, defined to include both wired and wireless phone services as well as Internet access, enjoys annual revenues of several trillion dollars and is by far the largest industry in the world.

Underlying this rapid development of all communications services and networks has been the ongoing evolution of digital technology, particularly large-scale integration and microprocessor chip technology. The digital revolution enabled transformation of the core of a traditional telephone network to a digital infrastructure providing greater reliability, increased capacity, and an ever-widening array of services to customers. About 10 years ago, digital technology began to have an impact on mobile wireless services and networks, increasing network capacities and capabilities as well as lowering the cost and increasing the battery life of mobile devices. An interesting and important aspect of the burgeoning worldwide wireless communications industry has been the rebirth of wireless LAN (WLAN) technology, driven by the steadily increasing popularity of laptop computers, the demand for wireless Internet access, and the expanding deployment of wireless access points on campuses and, increasingly, in public commercial venues.

Many of the wireless technology developments of the past decade have focused on improved physical (PHY) layer and medium access control (MAC) layer designs. The technical core of these protocol layers comprises digital signal processing (DSP) techniques and technology, to which most of this book is directed.

In this chapter we provide an overview of wireless information networks. We describe the basic elements of a wireless network and the key technical issues to be considered in the design of these networks. We also discuss the market sectors constituting the wireless industry and the trends apparent in voice- and data-oriented networks. In the final section of the chapter we outline the remaining chapters of the book.

1.1.1 Elements of a Wireless Network

An information network is an infrastructure that interconnects telecommunication devices to provide them with means for exchanging information. Telecommunication devices are terminals that allow users to run applications that communicate with other terminals through the information network infrastructure. The basic elements of an information network infrastructure are switches or routers that are connected by point-to-point links. Switches include fixed- and variable-rate voice-oriented circuit switches, low-speed (X.25) and high-speed (frame relay) data-oriented packet switches (routers), and ATM switches. The point-to-point links include a variety of fiber links, coaxial cables, twisted-wire pairs, and wireless connections.

To support transmission of voice, data, and video, several wired information network infrastructures have evolved throughout the past century. Wireless networks allow a mobile telecommunication terminal to access these wired information network infrastructures. At first glance it may appear that a wireless network is only an antenna site connected to one of the switches in the wired infrastructure, enabling a mobile terminal to be connected to the backbone network. In reality, in addition to the antenna site, a wireless network may also need its own mobility-aware switches and base station control devices in order to support mobility, that is, enabling a mobile terminal to change its point of connection to the network. Thus, a wireless network has a fixed infrastructure with mobility-aware switches and point-to-point connections, similar to other wired infrastructures, as well as antenna sites and mobile terminals.

Important examples of wireless networks are cellular telephone networks and wireless Internet access networks, which we discuss in greater detail in Section 1.2. There, we show how these networks extend the structure and services of existing wired networks to support either voice- or data-oriented wireless services

1.1.2 Key Technical Issues for Wireless Networks

As we can see in the two examples mentioned above, a wireless network includes not only the wireless terminals and radio-frequency (RF) links to fixed antennas, but also network elements and functions needed to support both interoperability with the existing fixed-wired networks and mobility for the wireless user. The set of characteristics of the wireless connection between the mobile terminal and a base station, including all the PHY- and MAC-layer details of access method, modulation, coding, and transmission formats, is commonly referred to as the *air interface*. Thus, we can say that the key technical issues for wireless networks are networking issues and air-interface design issues. Although these two sets of issues are not totally independent of each other, they are largely independent and can be treated separately. As we shall see in subsequent sections of the book, the networking issues relate primarily to interoperability between the wireless and wired infrastructures and to support of user mobility. On the other hand, air-interface issues relate primarily to the quality of service provided to wireless users and to efficiency in the use of available RF bandwidth.

1.1.3 Four Market Sectors

The market for wireless networks has evolved within four different segments that can be divided logically into two classes: the voice-oriented market and the data-oriented market. The *voice-oriented market* has evolved around wireless connection to the public switched telephone network. These services evolved further into local and wide area markets. The local voice-oriented market is based on low-power, low-mobility devices with a higher quality of voice, including cordless telephone, personal communication services (PCSs), wireless private branch exchanges, and wireless Telepoint. The voice-oriented wide-area market evolved around cellular mobile telephone services using terminals with higher power consumption, comprehensive coverage, and lower quality of voice. Figure 1.1a compares several features of these two sectors of the voice-oriented market.

The wireless *data-oriented market* evolved around the Internet and computer-communication network infrastructure. Data-oriented services can be divided into local

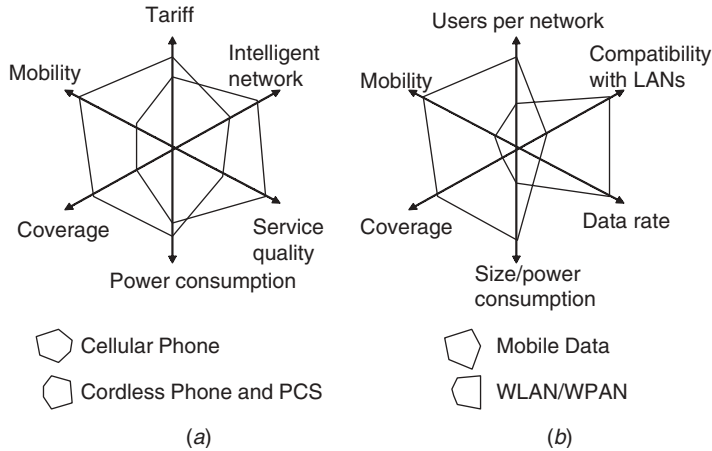


FIGURE 1.1 Wireless market sector comparisons: (a) voice-oriented networks; (b) data-oriented networks.

broadband and ad hoc markets on the one hand, and wide-area mobile data markets on the other. The wide-area wireless data market provides Internet access for mobile users. Local broadband and ad hoc networks include wireless LANs and wireless personal area networks (WPANs) that provide high-speed Internet access. The local and ad hoc networks will also support evolving ad hoc wireless consumer product markets. Figure 1.1b illustrates several differences among the local- and wide-area wireless data networks.

1.2 NETWORK ARCHITECTURE AND DESIGN ISSUES

Next we describe the principal system architectures for wireless networks and outline the key design issues that must be addressed in the design of these networks. These architectures and design issues are dealt with in detail in the remainder of the book.

1.2.1 Network Architectures

Here we consider two prominent examples of wireless networks: cellular telephone and wireless Internet.

Cellular Telephone. Figure 1.2 depicts wireless telephone service as an extension of the familiar *public-switched telephone network* (PSTN). The PSTN, designed to provide wired telephone services, is augmented with a wireless fixed infrastructure to support communication with mobile terminals. The mobile terminals communicate with the wireless fixed infrastructure via RF links to fixed antennas, each antenna connected to or integral with a *base station*. The PSTN infrastructure comprises switches, point-to-point connections, and computers used for operation and maintenance of the network.

The fixed infrastructure of the cellular telephone service has its own mobility-aware switches, point-to-point connections, and other hardware and software elements that

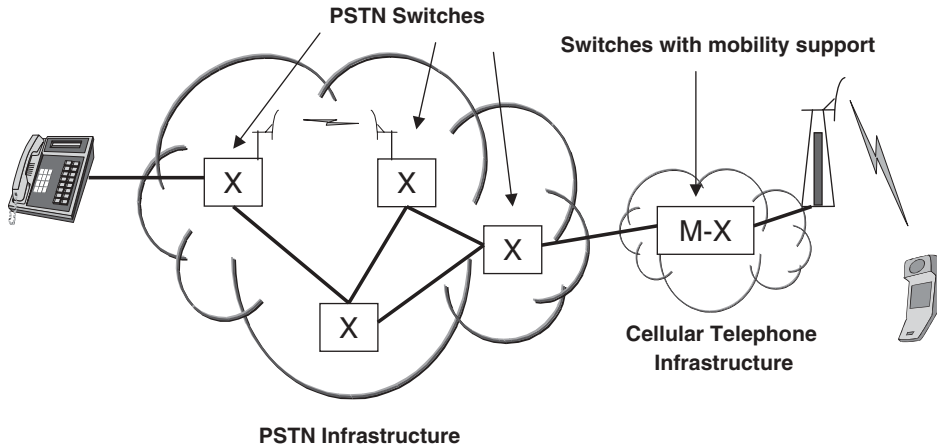


FIGURE 1.2 Cellular telephone infrastructure as an extension of the PSTN.

are needed for operation and maintenance of the mobile network. A wireless telecommunication device (e.g., a cordless telephone) can connect to the PSTN infrastructure by replacing the wire attachment with radio transceivers. However, for the wireless device to change its point of connection, switches in the PSTN must be able to support mobility. Switches in the PSTN infrastructure were not originally designed to support mobility. To solve this problem, a cellular telephone service provider adds its own fixed infrastructure with mobility-aware switches. The fixed infrastructure of the cellular telephone service provider is an interface between the base stations and the PSTN infrastructure that implements the functionality to support mobility. Just as a wired telephone service network needs added infrastructure to allow a mobile telephone to connect to the PSTN, a *wireless data network* needs its own added infrastructure to support wireless Internet access. Consider the next example.

Wireless Internet. Figure 1.3 shows the traditional wired data infrastructure together with an additional wireless data infrastructure that allows wireless connection to the Internet. The traditional data network consists of routers, point-to-point connections, and computers for operation and maintenance. The elements of a wireless network include mobile terminals, access points, mobility-aware routers, and point-to-point connections. This new infrastructure has to implement all the functionalities needed to support mobility.

The difference between the cellular telephone and wireless Internet examples is that the wireless network in Fig. 1.2 is a connection-based voice-oriented network, whereas the wireless network in Fig. 1.3 is a connectionless data-oriented network. A *connection-oriented operation* needs a setup procedure to connect the communicating terminals, and after the connection is established, a certain quality of service (QoS) is guaranteed to the user throughout the communication session. In *connectionless operation* there is no setup procedure and terminals are always connected to the network, in the sense that the communication session remains intact, but the QoS is not guaranteed. Instead, each protocol data unit (e.g., datagram or packet) is communicated between network access points on a best-effort basis. Common examples of connectionless protocols are the *Internet Protocol* (IP) and the *User Datagram Protocol* (UDP), both of

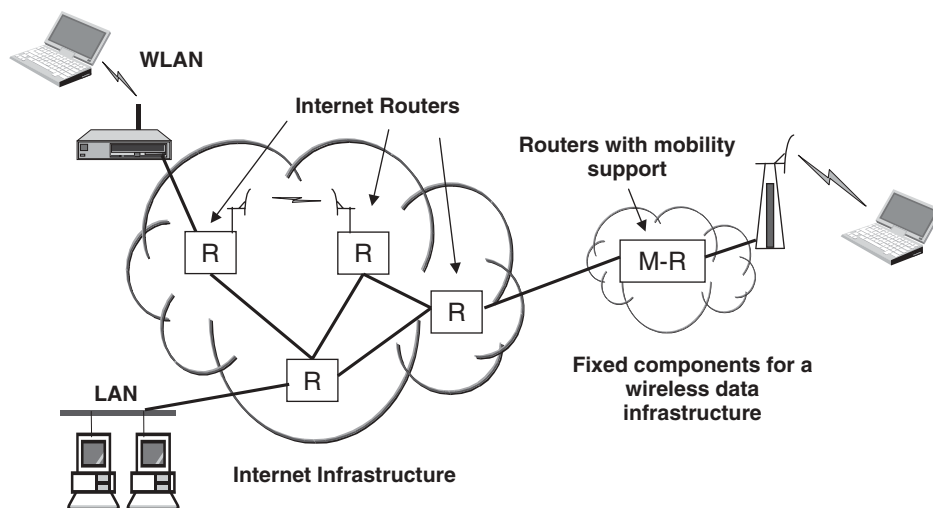


FIGURE 1.3 Wireless connectivity to the Internet.

which are used to support the *Transmission Control Protocol* (TCP), a transport-layer connection-oriented protocol, which in turn supports higher-layer protocols such as *Hypertext Transfer Protocol* (HTTP) and *Simple Mail Transfer Protocol* (SMTP). TCP is connection-oriented because a TCP session is set up and maintained for the full duration of a higher-layer session such as a Web-access session.

1.2.2 Wireless Versus Wired Networks

There are a number of fundamental differences between wired and wireless networks, essentially all stemming from the inherent characteristic of wireless communications (i.e., the replacement of fixed subscriber equipment connections by radio links). This freedom from the wired “tether” provides enormous advantages for customers of communications services but also introduces some new problems not encountered in traditional wired networks.

Perhaps the most important characteristic of wireless networking is that a radio link connecting the user’s device to a wired network infrastructure is inherently less reliable than a fixed wired connection. This characteristic should be obvious and will be familiar to users of cellular phones who have experienced signal breakup and dropped connections on cellular phone calls. This inherent relative unreliability of radio links leads to a need for considerably more complexity in the physical-layer design than is required in traditional wired networks. Also, there is a need for *connection management* techniques as part of the solution to the radio-link reliability problem.

Another important characteristic of wireless communications is the fundamental limitation on the availability of frequency spectrum. For systems that operate in licensed frequency bands (cellular telephone service is the primary example), each service provider operates its network within a fixed band of frequencies, and means must be provided to manage the sharing of allocated bandwidth among a large number of users. Furthermore, as the service provider’s subscriber volume grows, there must be

means for expanding the overall capacity of the service network in an efficient manner to accommodate the growth in demand for service.

The bandwidth limitation problem also gives rise to the need for complexity in the design of source coding techniques (speech coding in the case of voice service, and other compression techniques in the case of multimedia transmission) so as to reduce the amount of bandwidth needed for each user channel signal while maintaining a prescribed level of signal quality as perceived by the user.

A very practical issue for users of mobile wireless devices is the necessary reliance on batteries, with the need for periodic recharging. This issue has led to the clever application of *power management* techniques in the design of mobile devices, so as to extend talk time and recharging cycles.

The inherent advantage of wireless networking—mobility for the user—adds complexity to the network design to manage changing the connection point to the fixed network infrastructure, including changes over both small and large geographical areas. This calls for greater complexity in *registration* and *call routing* techniques than are needed in wired networks, and a need for the use of both permanent and temporary addressing to support mobility.

Finally, the use of wireless transmission creates a vulnerability of the user's communications to eavesdropping and fraudulent intrusion into the network. Because of these problems, considerable attention has been given to providing *security* and *privacy* for wireless communications networks. Security provisions include such techniques as *authentication* to prevent unauthorized access to networks. Privacy provisions include the *encryption* of transmitted digital streams to prevent eavesdropping.

1.2.3 Elements of a Wireless Network Architecture

It is useful to consider the elements of a wireless network in four categories: services, infrastructure, protocols, and network engineering.

Services. From the perspective of the user of the network, the principal aspect of the network is the service or set of services the network is designed to support. In fact, the various industry efforts that lead to interoperability standards invariably begin with agreements among participants as to the array of services to be provided by the intended network standard, and considerable attention is given to the detailed features of those services and the specific ways in which the user's equipment will interact with the network in the operation of the service. Of course, the basic types of services are voice and data services.

In some networks, the voice services might comprise a menu of selectable digital data rates, the higher data rates providing a higher received voice quality at the cost of higher bandwidth requirement, accompanied by an appropriate tariff differential.

Data services may be provided in various forms, the simplest, called *data-bearer service*, being simple transport of data with minimal specification of data format at the mobile data port. Data service offerings might include a choice of transparent (T) or nontransparent (NT) data in either synchronous (clock-driven) or nonsynchronous (start/stop character-driven) formats. Transparent data service will employ forward-error correction coding at a fixed transmission rate in the channel. Nontransparent service will employ error-detection coding and retransmission of faulty data blocks so as to ensure greater accuracy in the delivered user data. Other options might

include circuit-switched (connection-oriented) data versus packet (connectionless) service. Other, application-specific data services, such as Group-3 Facsimile service, will typically be offered with a set of optional data rates.

Short messaging service (SMS) is available in many cellular networks for transmission and reception of short text messages displayed on a small screen. The SMS messages are embedded in the control channels of the cellular network, which enables rapid delivery. A service that is growing rapidly, called *text messaging*, is built on SMS. As the demand for wireless multimedia service grows, data services are being provided at increasingly high data rates.

System Infrastructure. Provisioning of the various services in a wireless network in turn places requirements on hardware and software that must be included in the elements connecting the wireless service customer with the fixed networks. We need to consider two categories of system elements: the mobile terminal and the fixed wireless infrastructure that makes connection with the fixed network.

The mobile terminal is the user's device for sending and receiving signals over a wireless link. For a user requiring only basic voice service, the mobile terminal is the familiar cellular phone, nowadays probably a CDMA digital phone in the United States or a GSM phone in Europe and many other regions of the world. Many cellular phones are designed with a standardized data port for connecting to a portable computer or other data terminal. In supporting data connectivity, the cellular phone is functioning as a wireless modem interfacing baseband data (e.g., ASCII-formatted) with the wireless network. Currently, this wireless modem function is typically implemented in a circuit card to be plugged into a socket on a laptop computer, or even a card already mounted inside the laptop. As wireless networks evolve to support increasing capability for multimedia transmission, a variety of new mobile devices are appearing in the marketplace to support sending and receiving multimedia images.

Signals from the mobile user terminal arrive at an antenna that provides an RF interface to the fixed wireless infrastructure, and that infrastructure in turn provides an interface to the fixed wired infrastructure. In the case of cellular systems, the fixed wired infrastructure will typically be a public-switched telephone network (PSTN) or a public-switched data network (PSDN). In the case of WLAN systems, the fixed network will typically be a wired Ethernet LAN in an office building, office complex, or university campus.

In the case of cellular systems, the fixed wireless infrastructure includes antennas, radio base stations (BSs), mobile switching centers (MSCs), and terrestrial lines (typically, coaxial cable or optical fiber) to make connections among BSs and MSCs as well as between MSCs and the PSTN. The fixed wireless infrastructure will also include computers and a variety of instrumentation needed for operation and maintenance of the cellular network. All of the equipment and software in place, from the antennas to the PSTN connections, will be owned and operated by the cellular service provider. Currently, a cellular service company might have to deploy 50 to 100 BSs to provide satisfactory signal coverage over a major metropolitan area.

Functional partitioning between network equipment elements may vary from one manufacturer's equipment to another's, but in current cellular networks, the BS will typically include not only RF transmission and reception equipment but also speech coder/decoders (*codecs*). In such a configuration, all transmissions between the BS and the PSTN are in digital form. In such a configuration, the BS will also typically

include *interworking functions* (IWFs), also called *modem emulators*, to modulate and demodulate the data streams in support of wireless data services. The MSCs include mobile-aware switches that provide for the setup and routing of call connections to and from mobile terminals and also handle the hand-off of call connections from one BS to another as mobile users move about the cellular service area. The MSCs also include the other hardware and software elements that are needed for mobile network operation, maintenance, and troubleshooting.

Wired Backbones for Wireless Networks. Since wireless networks depend heavily on the wired infrastructures to which they connect, in this section we provide a brief overview of the important wired infrastructures. The most commonly used wired infrastructures for wireless networks are PSTN, Internet, and hybrid fiber coax (HFC), originally designed for voice, data, and cable TV distributions applications, respectively. Figure 1.4 provides an overall picture of these three networks and how they relate to other wired and wireless networks. (A more detailed discussion of this topic can be found in [Pah02a].)

The main sources of information transmitted through telecommunication devices are voice, data, and video. Voice and video are analog in nature, whereas data traffic is digital. The dominant voice application is telephony, that is, a bidirectional symmetric real-time conversation. To support telephony, telephone service providers have developed a network infrastructure that establishes a connection for a telephone call

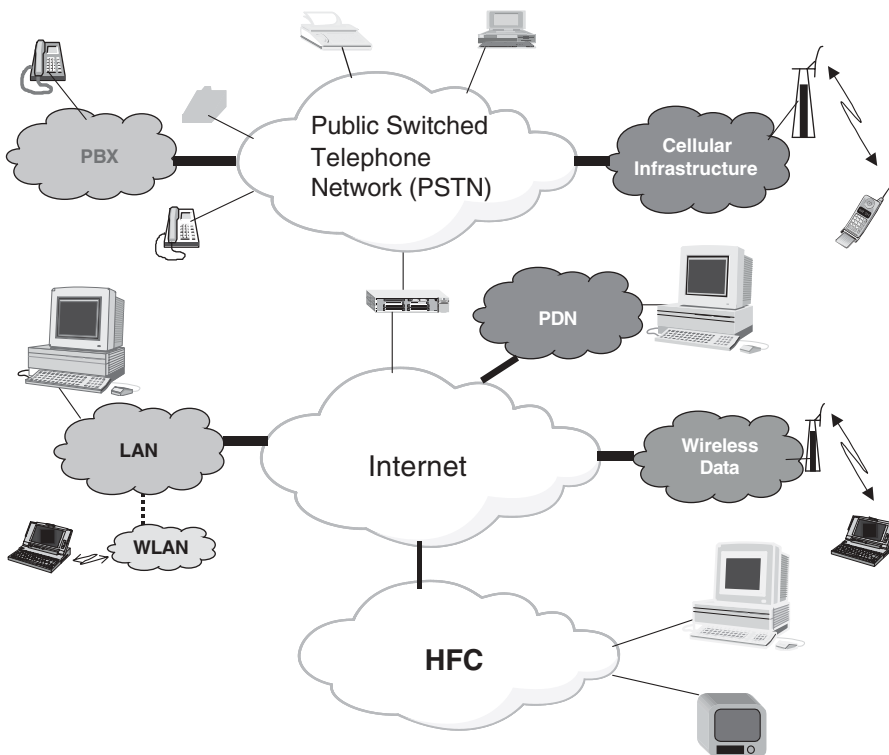


FIGURE 1.4 Interconnection of PSTN, Internet, and HFC.

during the dialing process and disconnects it after completion of the conversation. This network is referred to as the public switched telephone network (PSTN). As shown at the top of Fig. 1.4, the cellular telephone infrastructure provides wireless access to the PSTN. Another network attached to the PSTN is the *private branch exchange* (PBX), a local telephone switch owned privately by a business enterprise. This private switch allows privacy and flexibility in implementing additional services in an office environment. The PSTN physical connection to homes is twisted-pair wiring that is also used for broadband xDSL services. The core of the PSTN is a huge digital transmission system that allocates a 64-kb/s channel for each direction of a telephone conversation. Other network providers often lease the PSTN transmission facilities needed to interconnect their nodes.

The infrastructure developed for video applications is cable television, shown in the lower part of Fig. 1.4. This network broadcasts wideband video signals to residential premises. A cable goes from an end office to a residential neighborhood, and all customers are fed from the same cable. The set-top boxes leased by cable companies provide selectivity of channels, depending on the customer's service subscription. The end offices, where groups of distribution cables arrive, are connected to one another with fiber lines. For this reason, the cable TV network is also called *hybrid fiber coax* (HFC). Nowadays, cable distribution is also used for broadband residential access to Internet.

The data network infrastructure was developed for bursty data applications and evolved into the Internet, which supports Web access, e-mail, FTP, and Telnet applications as well as multimedia (voice, video, and data) sessions with a wide variety of session characteristics. The middle part of Fig. 1.4 shows the Internet and its relation to other data networks. From a user point of view, data-oriented networks are always connected, but they use the transmission resources only when a burst of information is to be transferred. Sessions of popular data communications applications such as Web browsing or FTP are often asymmetric, and a short upstream request burst results in downstream transmission of a large amount of data. Symmetric sessions such as IP telephony over data networks (termed *voice over IP*, or VoIP) are also becoming popular, providing an alternative to traditional telephony. Residential Internet access is a logical access that is physically implemented on other media, such as cable TV wiring or copper telephone lines. Distribution of the Internet in office areas is usually through Ethernet local area networks (LANs). Wireless LANs in offices are usually connected to the Internet through the wired LANs. Nowadays all other private data networks (PDNs), such as those used by banks or airline reservation agencies, are also connected to the Internet. The Internet also serves as the backbone for wireless data services.

Protocol Layering. Wireless communications networks, and cellular networks in particular, must accomplish many complex functions in order to establish call connections to and from mobile users, to implement the services to which each user has subscribed, to manage user authentication, and to provide mobility for wireless user terminals. As we have noted above, these tasks are performed in a number of mobile and fixed elements. At the same time, these networks must provide smooth interoperability among hardware and software elements supplied by a variety of manufacturers. In complex systems such as these, designers have found it beneficial to organize systems designs according to the concepts of *protocol layering*. Perhaps the best known model for

protocol layering is the Open System Interconnect (OSI) seven-layer reference model, adopted as an international standard in 1978. In the OSI model, the lowest layer, layer 1, the physical layer, provides a physical medium for the flow of information across a link. The highest layer in the model, layer 7, the application layer, provides services to users of the network. In the intermediate five layers, the services provided move progressively away from the physical medium toward network- and application-related functions.

The basic concept of protocol layering is to manage the complexity of a network design by segmenting the system functions into a set of *layers*, each layer built on the ones below it. Each protocol layer can be described as performing specific *services* for the higher layers while isolating the higher layers from the details of how the services are actually implemented. The set of rules by which information is processed and formatted in any given layer constitutes a *protocol* for that layer. This assures, for example, that two pieces of equipment performing functions in the same layer can interoperate properly at that layer. A set of layers and their protocols is commonly referred to as a *network architecture*. A list of protocols used in a chosen system, one protocol per layer or sublayer, is referred to as a *protocol stack* [Gar00].

In all of the wireless networks we consider in this book, the system functions are organized according to some version of protocol layering. From one network standard to another, the functional segmentation into layers may be somewhat different. However, the functional segmentation will be the same for all hardware and software elements manufactured to each particular standard. For example, the GSM network architecture consists of five layers: transmission, radio resource management, mobility management, communication management, and operation, administration, and maintenance [Mou92, Meh97, Hei99]. As a second example, the IEEE 802.11 family of standards encompasses two layers: MAC and PHY [O’Ha99].

Traffic Engineering and Deployment. The cost of equipping and deploying a wireless communications network can vary widely depending on the type of network and the application for which it is intended. A WLAN might be installed in a business office or in a university campus building for a few thousand dollars. On the other hand, a cellular telephone network built to serve a metropolitan area might incur costs of tens of millions of dollars. However, regardless of the wireless technology employed or the intended application, principles of sound network engineering apply: The network should be designed to provide good signal coverage to wireless terminals over the intended floor, campus, or geographic service area with a reasonable expenditure of capital for equipment and installation.

In the case of a WLAN installation, *access points* (sometimes called *base stations*) will typically be installed on ceilings or high on walls in locations chosen to provide unobstructed signal coverage for some set of wireless terminals, such as desktop or laptop computers. Multiple access points will be installed to cover the total population of wireless terminals, typically with overlapping coverage areas to avoid gaps in coverage.

In the deployment of a cellular telephone network, the general principle of good network engineering is the same as for a WLAN deployment—install a number of cell sites in such a way as to provide unbroken signal coverage for mobile users over the geographic area in which the cellular company offers service. The cost of equipping and installing a single cell site might well be on the order of \$1 or 2 million, including

acquisition of real estate, and thus it is important that the cell site layout be designed to make optimum use of capital investment.

A key element in the planning of a wireless network is a specification of the traffic the network will be designed to handle. In the case of a WLAN design, we would want to know the number of wireless terminals and some statistics for the amount and type of traffic to be generated by the terminals. We would want to know the profile of short-message traffic, long file transfers, and so on, and the frequency of these transmissions. In other words, we would like to have a *traffic model* as a starting point for planning the network. With a traffic model in hand, and a specification of the traffic capacity of an access point, we can determine the number of access points to be provided. Then specifying the distribution of wireless user terminals will allow us to position the access points appropriately.

In the case of a cellular network deployment, the considerations are much the same, but with the important difference that users are highly mobile and that communication traffic patterns can change significantly from day to day and even from hour to hour. Commuters caught in a traffic jam or in a severe rainstorm will generate an unusually high volume of calls as they try to contact their co-workers or family members to revise their schedules. Fans at a Sunday afternoon football game will generate high volumes of communication traffic in the vicinity of the stadium. Another traffic characteristic specific to the cellular case is the relatively high frequency of call handoffs as users move about the service area. This contrasts with the typical relatively less frequent handoffs experienced in the WLAN environment. Thus, the efficient engineering of a cellular network must take account not only of average statistics of generated traffic but also of the potentially high variability of the traffic. Once again, a traffic model is needed, and the traffic model in the cellular case is likely to be considerably more complex than in the WLAN case. We shall have more to say about wireless network deployment in subsequent sections.

1.2.4 Technical Aspects of a Wireless Infrastructure

Next, we consider important issues that must be addressed in the design and operation of the wireless network infrastructure. These issues are often considered under the heading *network engineering*, as they are issues concerning the design and operation of the network as a whole.

Network Deployment Planning. In the preceding section we briefly discussed traffic engineering as a key element in network planning. With a traffic model, both temporal and geographic, as a starting point, the network engineer can begin to plan the layout of the access points or cell sites that will carry the wireless traffic to and from the fixed wired backbone network. This aspect of network planning is typically performed with the aid of *signal coverage prediction models*.

Signal coverage prediction models, usually based on a combination of radio-wave propagation theory and experimental measurements, provide the designer with a means of estimating the optimum placement of access points or cell sites for covering the intended area of user terminals with acceptable signal quality. A tutorial description of signal coverage in a wireless network with multiple access points or cell sites will typically illustrate signal coverage with a diagram of abutting hexagons or perhaps circles with some overlapping coverage areas. Those are both highly idealized descriptions that do not accurately represent the real world of wireless signal propagation and

coverage. Even in a relatively benign office layout, planning for a WLAN installation must take account of the types and locations of office furniture and equipment, office partitions, walls, doorways, and so on, all of which can affect signal coverage. In a factory setting, even more complex situations might be encountered, with various metal surfaces, manufacturing machinery, and so on, all affecting signal propagation throughout the building. In the case of cellular telephone networks covering large service areas, signal coverage prediction models must take account of a wide range of factors, including terrain type (flat, hilly, mountainous), land use (rural, suburban development, city high-rise, urban canyon), and special situations such as roadways on high bridges and over-water propagation paths.

Some of the more sophisticated cellular network planning tools are elaborate software packages that incorporate time-varying traffic models, population distributions, cellular antenna types, optional propagation models, and call-handoff models in order to make accurate estimates of received signal quality for customers situated in various sectors of a service region. Even very sophisticated network planning tools can provide only an estimate of network performance, and a network engineer may well also conduct *drive tests* in selected regions of the service area in order to verify or refine computer-generated performance estimates.

Mobility and Location Management. An important requirement that users will place on wireless networks is mobility, freedom for the wireless user to maintain a reliable wireless connection while moving about an area that is relevant to the application. In the case of a WLAN system, users may want the capability to move their wireless terminals to different locations in an office building, factory, or campus without having to reregister with the network. Here, users are not likely to move about rapidly, and the problem is a relatively simple one. However, in the case of WANs such as cellular telephone networks, mobility is the *raison d'être* of the technology and is the principal differentiator between traditional wired telephone networks and wireless networks. Users expect to be able to move about freely on foot, by automobile, or even traveling on trains, while enjoying seamless connectivity for their wireless communication. They also expect to be able to migrate from one cellular company's coverage region to another's, placing and receiving calls reliably in any region.

In traditional wired telephone networks, the subscriber's telephone is always wired to the same central office (CO) switch, and the network directs every incoming call to the subscriber's line using his or her telephone number. Outgoing calls are always made through the same local CO to which the subscriber is permanently connected. However, in a cellular network, the cell site to which the user connects when receiving a call depends on the user's physical location at that moment. In order for a subscriber to receive a call, the network must determine the cell in which the user is currently located. This is the essence of the *location management* problem, and this problem has been solved in cellular networks by designing location awareness into both the wireless and wired portions of a wireless network infrastructure.

An important facet of location management is *call handoff*, the process in which a user's call connection is transferred seamlessly from one serving cell to another as the user moves about the service area. This comes under the heading of what is known as *mobility management* in cellular systems. This is accomplished by a combination of signal strength measurements made in the releasing and the receiving cells, and coordination of frequency channels in the two cells, typically done under the control

of a mobile switching center (MSC). Once again, this calls for a considerable amount of complexity in the design of the wired and wireless segments of the wireless network infrastructure.

Related to call handoff is the process of *roaming*, by which a user who has subscribed to particular services in his or her home area can travel to another service provider's region and use the same services. This feature greatly enhances the value of a wireless service to a subscriber by lessening geographic restrictions on his or her access to services. Roaming capabilities in cellular networks have been achieved by cooperation among service providers and among manufacturers, largely in the venue of standards bodies. Roaming requires the adoption of a standardized air interface, standardization of phone-type identification, and cooperation among the operators for transfer of location data between home and visited networks. Cooperation is also required in administrative areas such as transfer of calling charges and subscription information.

Radio Resource and Power Management. A characteristic of any wireless network is that it must operate within a strictly defined spectrum allocation. Radio spectrum is a limited resource, and regulatory agencies set specific spectrum allocations for different services. For example, a cellular network operator has a license for 25 MHz of bandwidth, 12.5 MHz for each direction of full-duplex communication. With a typical cellular reuse factor of 7, about 3.6 MHz of bandwidth is available for two-way traffic in each cell, and this bandwidth must be shared among active users in the cell. The bandwidth available is far less than what would be required if all subscribers in the network were to demand call connections simultaneously. This is in marked contrast to a wired telephone network, in which we may always add new subscribers to the network by installing additional local loops. (To be sure, there is an issue in equipping a wired telephone network in sizing the central office switches and the long-haul switches to ensure enough connections through the switches to meet expected call demand.) Thus, to ensure efficient sharing of the allocated spectrum, RF channels must be assigned and released dynamically, on a per-call basis.

Furthermore, directing a call from the wired network to a specific mobile subscriber is not a trivial procedure. Cellular networks reserve a portion of the allocated bandwidth for *control channels*, which are utilized in establishing and managing call connections. Paging messages are transmitted from cell sites, and a paging/response protocol is used to let the network determine which cell is currently the best one by which to reach the called subscriber. Only when this location determination has been made is an RF channel assigned for the call. All of these functions of assigning and managing the limited number of available RF channels come under the heading of *radio resource management*.

Another important element of radio resource management is *power management*. A cellular network is designed to operate under interference-limited conditions. That is, the dominant source of signal degradation in the network is interference from other active users of the network. With frequency reuse, the signal power radiated from a given cell is held to a sufficiently low level that the same subset of frequencies can be used simultaneously in another cell a reuse-separation distance away. In some cellular networks, power control is performed in both the base stations and the mobile phones. Power control at both ends of the wireless link helps to hold radiated power to a level sufficient to maintain good-quality communication without unduly increasing the overall interference level in the network.

Security. Although the use of wireless communications relieves the user of the wired tether to the public telephone network, with the enormous advantage of freedom of movement, the wireless medium also makes the user's communications vulnerable to eavesdropping and even fraudulent intrusion. In fact, when standards were being developed for digital wireless networks, a major benefit recognized for digital transmission was the facility it provided for the implementation of authentication and encryption techniques. All of the digital wireless interoperability standards have included procedures for authentication of users entering the network. With respect to the privacy problem, WLANs utilize spread-spectrum transmission, which has an inherent resistance to casual eavesdropping. Cellular networks based on CDMA are also using spread spectrum, providing inherently private transmission. In other cellular networks, such as GSM, encryption is provided in some operators' networks as a selectable feature. For communications that are particularly sensitive, some users may employ application-layer end-to-end encryption and use a wireless data service to carry encrypted traffic across the network. In this case, the user does not rely on the wireless network to provide security or privacy.

1.2.5 Technical Aspects of the Air Interface

If we examine the evolution of successive generations of wireless networks of any given type, say WLAN or WAN, the principal differences appearing from one generation to another lie in the details of the air-interface design, encompassing physical (PHY)- and medium access control (MAC)-layer designs. There is good reason for this: Improvements and new developments in PHY- and MAC-layer designs have yielded the most significant enhancements in communication quality and spectrum utilization in these networks. To assess the technical issues underlying the various alternatives for PHY- and MAC-layer designs and why certain choices have been made in the evolution of air-interface standards, it is necessary to begin with a good understanding of the RF medium for each type of wireless system. The fundamental characteristics of an RF transmission medium are summarized next.

Path Loss. As an RF signal radiates from a transmitting antenna to a receiving antenna, there is propagation path loss that attenuates the signal strength, the strength generally decreasing with distance along the path. In an idealized configuration, the loss could be calculated simply as free-space loss, which increases with the square of distance. However, in practical applications, the propagation path is rarely describable by the free-space model, and many other factors must be taken into account. Different factors will be important in different types of systems. For WLAN applications, where the spatial scale is on the order of room and building dimensions, one must account for structural elements and their properties as reflectors and absorbers of signal energy. For WANs such as cellular telephone networks, an even wider array of factors must be considered, and path loss is often difficult to compute theoretically. The path loss will typically be influenced by terrain type, land use, and sometimes by unique topological or architectural elements situated on or near a particular path. In all wireless networks, path loss also depends on the types of antennas used at both the mobile and fixed ends of the wireless link.

Multipath. For wideband signals as used in WLANs and some cellular networks, time dispersion, called *multipath*, must also be accounted for in modeling the propagation

medium. As with path loss, multipath characteristics can vary widely from one type of network to another. Specific characteristics will vary with operating frequency, application setting (indoor versus outdoor), antenna coverage patterns, and structural or topological details of the coverage area. Furthermore, the ways in which multipath characteristics are described and measured will vary depending on the bandwidths of signals employed in each type of network. Given the wide variance among multipath characterizations for different types of networks, designers have tended to develop multipath modeling methods and databases that are specific to the individual types of networks under consideration. In some instances, standards-setting bodies have formalized sets of multipath models that manufacturers and service providers have agreed to use in development, testing, and evaluation of new wireless products. The chief example here is the Joint Technical Committee (JTC) of GSM, which has evolved standardized sets of multipath models, commonly termed *JTC models*, applicable to a variety of cellular propagation environments.

Fading. The third fundamental characteristic of an RF transmission medium is the presence of time variations in the strength of received signals, an effect usually referred to as *fading*. Signal fading can arise from a wide variety of causes, all of them related to dynamics of the propagation medium. Movements of WLAN terminals, or even movements of people, chairs, or doors, in the vicinity of WLAN signal paths, can cause variations in received signal strength. In a busy application environment such as a factory floor, there may be almost-constant movement of workers, vehicles, equipment, and materials, and all of these movements are potential sources of signal fading on WLAN paths.

In a cellular network, an obvious source of signal fading is the mobility of wireless users. When a cell phone is being operated in a moving automobile, or even by a pedestrian walking on a city street, the characteristics of the propagation path to the serving base station are changing constantly. Even if the wireless user is temporarily motionless, the movement of nearby vehicular traffic can result in signal fading.

Signal fading effects are closely related to multipath, discussed above, and in fact, most fading effects are due to the time-varying interaction of multipath signal components. That is, the time-dispersed signal components are typically affected individually by amplitude and phase fluctuations, and when these components are received together, there is an observed variation in the composite signal, commonly termed *multipath fading*. The multipath-fading phenomenon had been well understood for decades prior to the design of modern wireless networks, since it is a fundamental characteristic of long-distance radio propagation in HF, VHF, and UHF frequency bands (3 MHz to 3 GHz). The long-standing interest in utilizing digital transmission techniques in those high-frequency bands led, beginning in the 1950s, to development and refinement of signal-processing techniques such as *diversity combining*, spread-spectrum *RAKE receivers*, and *adaptive equalization* as means of ameliorating the effects of multipath fading in digital communication [Pri58]. Thus, those earlier developments were brought to bear in the PHY-layer designs of modern wireless networks, and there continues to be active research on new and better techniques for dealing with the multipath fading characteristics of wireless channels. Current examples of the fruits of this research include *orthogonal frequency-division multiplexing* (OFDM), *turbo equalization*, and *space-time coding* [Han02].

PHY- and MAC-Layer Alternatives. Most of the research in wireless networking technology in the past two decades has been directed toward advancements in PHY- and MAC-layer designs. Research on these aspects of wireless air interfaces has yielded steady improvements in performance quality and reliability as well as in the efficiency of utilization of spectrum. High quality of service builds customer satisfaction and healthy growth of the industry. Improved spectrum utilization translates directly into increased traffic capacity in the network, enabling greater per-user cost efficiency in the installation and operation of the network. The progression of successive generations of WLAN and WAN standards and products has in fact been characterized largely by advancements in PHY- and MAC-layer techniques.

A key element of PHY-layer design is the modulation technique, and the spectral efficiency afforded by each technique is critical. Thus, we have seen a steady progression of increasingly spectrum-efficient modulations in wireless networks. Access methods used in WANs have evolved from simple *frequency-division multiplexing* (FDM), through embedded digital control channels, to *code-division multiple access* (CDMA). The legacy analog cellular systems had utilized simple *frequency modulation* (FM) with *frequency-division multiplexing* (FDM) of user channels, a design much the same as had been used for decades in *land-mobile radio* (LMR) systems for taxicab fleets and public safety departments. The earliest digital cellular designs (second-generation cellular) saw the introduction of digital modulation techniques such as GMSK and $\pi/4$ -QPSK together with *time-division multiplexing* of multiple digital traffic channels into TDMA frames, with various forms of control information embedded on a per-frame basis. TDMA transmissions were frequency-multiplexed into a FDM-TDMA signal design. The development of IS-95 for U.S. CDMA introduced spread-spectrum PSK modulation with code division as the access method, the combined techniques providing significant capacity enhancements relative to TDMA. Third-generation WAN standards will continue to build on CDMA designs, with higher-bandwidth signals and higher-data-rate services.

Since the 1997 adoption of IEEE 802.11 as the first international standard for WLANs, this standard, together with its subsequent modifications, had dominated the WLAN products industry. The initial 802.11 standard specifies a MAC layer and three PHY-layer options: (1) direct-sequence spread spectrum (DSSS) with both differential binary PSK (DBPSK) and differential quaternary PSK supported, (2) frequency-hopping spread spectrum (FHSS) with Gaussian FSK (GFSK) in either two- or four-level formats, and (3) infrared transmission using pulse position modulation (PPM), with two data rates supported. The 802.11–1997 MAC-layer protocol comprises two sublayers, the lower sublayer providing three access options. The basic access mechanism here is *carrier-sense multiple access with collision avoidance* (CSMA/CA).

Three enhancements of 802.11 have been standardized, designated as IEEE 802.11b and 802.11a, both ratified in 1999, and 802.11g, adopted in 2003. The 802.11a enhancement, which provides data rates up to 54 Mb/s, incorporates *orthogonal frequency-division multiplexing* (OFDM). The OFDM scheme divides the high-data-rate stream into several lower-rate streams. The lower-rate streams are then transmitted simultaneously on multiple subcarriers, producing longer symbol durations in the subchannels, thereby lessening the effect of multipath distortion. The OFDM technique is also employed in the HIPERLAN2 standard.

The IEEE 802.11b enhancement extends the 802.11 PHY layer using high-rate DSSS (HRDSSS). The 802.11b extension provides a modulation mode at 11 Mb/s using *complementary code keying* (CCK). The IEEE 802.11g extension operates at radio frequencies between 2.400 and 2.4835 GHz, the same band as that used by 802.11b. However, the 802.11g specification employs OFDM, the modulation scheme used in 802.11a to obtain higher data speed. Computers or terminals set up for 802.11g can fall back to speeds of 11 Mb/s.

The early mobile data networks ARDIS, Mobitex, and CDPD used PHY-layer modulation techniques that were much like those used in the digital cellular standards then under development. ARDIS used 4-ary FSK, whereas Mobitex and CDPD used GMSK. Since a high demand for those mobile data services never developed, there was little motivation to develop new PHY-layer designs for increased capacity. The MAC-layer protocols used in those original mobile data networks were relatively simple contention-based protocols, including data-sense multiple access (DSMA), slotted ALOHA, and digital-sense multiple access, which is closely related to CSMA with collision detection (CSMA/CD). In time, these mobile data services were subsumed under the cellular data services offered by the major cellular operators.

1.3 KEY TRENDS IN WIRELESS NETWORKING

Now that we have outlined the key characteristics of wireless information networks, it will be useful to summarize the trends to be observed in the continuing evolution of wireless technology. Let us look separately at voice- and data-oriented networks.

1.3.1 Voice-Oriented Networks

From the first-generation starting point of legacy analog cellular networks, second-generation voice-oriented networks introduced a digital air interface, in part to facilitate the introduction of digital data services and other new features and services, but also to expand network capacity over that provided by the legacy analog networks. The introduction of CDMA and third-generation wideband CDMA then provided improved voice quality and system capacity and also met the growing demand for higher data rates. As fourth-generation designs develop and evolve, we will probably see introduction of the use of space-time diversity techniques and multiple-input multiple-output (MIMO) antennas in the air interface, paving the way for further quality and capacity improvements beyond the third-generation systems.

1.3.2 Data-Oriented Networks

The 1997 ratification of the IEEE 802.11 standard was a major milestone in the WLAN industry. The 10-year standardization effort not only produced an international standard assuring product compatibility among WLAN manufacturers, but the project also provided good solutions to some difficult problems that had to be faced in creating wireless extensions to the then-ubiquitous wired LANs. The 802.11 standard dealt with mobility, link reliability management, power management, interference minimization, and security. While the initial standard did not provide data rates as high as then-standard 10-Mb/s Ethernet over wired LANs, the 1- to 2-Mb/s wireless rates met the needs of

many users, who welcomed the mobility afforded by WLAN technology despite the data-rate limitations. The demand for higher data rates was inevitable, of course, and subsequently, the IEEE was able to issue the 802.11a (rates up to 54 Mb/s), 802.11b (rates up to 11 Mb/s), and 802.11g (rates up to 54 Mb/s) enhancements, based on the solid foundation of 802.11, the higher data rates achieved through use of the OFDM and CCK modulation schemes. Now appearing on the horizon is *ultrawide-band* (UWB) pulse transmission, which offers promise of further increases in data rates and coexistence of even larger numbers of simultaneous users.

1.3.3 Where Is the Complexity?

As we indicated earlier, the complexity of radio propagation and its variability from one deployment situation to another pose major challenges in the design of efficient wireless information networks. The fundamental design objective will be to provide good signal coverage and reliable, high-quality service on each link. Thus, the designer must focus on air-interface performance, which in turn places emphasis on PHY- and MAC-layer designs. Because of this, research must continue to address a variety of complex signal-processing techniques, including new time, space and frequency diversity techniques, with the objective of achieving steadily higher data rates, better service quality, and increased spectrum utilization.

Related technologies are being pursued as well, including wireless positioning and optical communications. Wireless positioning is a particular example of new technology that is affected critically by the complexity of the radio propagation environment, and we can expect that increasingly sophisticated propagation modeling techniques will be required as this area of application becomes more important.

1.4 OUTLINE OF THE BOOK

The book is organized into four parts, each focused on a different aspect of wireless information networks.

Part I, Chapters 1 and 2, provides an introduction to wireless information networks. In Chapter 1 we provide an overview of wireless networks, principal design issues, and current trends in the evolution of these networks, and in Chapter 2 we outline the evolution of the wireless industry.

Part II, Chapters 3 to 6, deals with characterization of radio propagation, channel measurement and modeling for narrowband signaling, measurement of wideband channel characteristics, and computer simulation of radio channels. Both indoor and outdoor wireless channels are considered.

Part III, Chapters 7 to 10, deals with modem design, addressing many details of the underlying signal-processing functions employed in wireless networks. Chapter 7 deals with narrowband modem technology, Chapter 8 deals with diversity and coding techniques used to improve communication reliability in wireless channels, Chapter 9 deals with broadband modem technologies, and Chapter 10 deals with spread-spectrum and UWB systems.

Part IV, Chapters 11 to 15, deals with MAC-layer design, ultrawideband communications, geolocation technology, and optical wireless networks, and provides a current summary of important standards and systems.

QUESTIONS

- (a) Consider a digital cellular network giving customers wireless access to the PSTN using their cell phones. What essential functions must be performed by the cellular equipment in carrying a voice signal arriving from the PSTN to a customer's cell phone?
- (b) Why do all digital cellular systems employ methods of speech compression?
- (c) Describe the phenomenon of multipath fading, and describe a few scenarios in which this effect can be expected in a wireless call connection in a wide-area cellular network. Next, describe a few scenarios applicable to indoor communication over a wireless LAN.
- (d) How would you distinguish the traffic characteristics observed in a voice-oriented wireless network from those of a data-oriented network?
- (e) Why is accurate modeling of radio-wave propagation important in the planning of a cellular service network?
- (f) What technical and economic factors are most likely to cause a shift in dominance among competing digital cellular standards?
- (g) Digital cellular standards include specifications of interfaces between major elements of each system. Why are these interface specifications important to (1) cellular network operators, (2) cellular equipment manufacturers, and (3) cellular service customers?
- (h) What advantage do you see in the fact that all versions of the IEEE 802.11 standards for WLANs share the same MAC-layer protocol?
- (i) What in your view is the major factor accounting for the resurgence in popularity of WLAN networking?
- (j) Discuss and compare the issues involved in expanding capacity in the PSTN versus expanding capacity of a cellular network.

2

EVOLUTION OF THE WIRELESS INDUSTRY

- 2.1 Introduction
 - 2.1.1 Evolution of Voice-Oriented Networks
 - 2.1.2 Evolution of Data-Oriented Networks
 - 2.2 Three Views of the Wireless Industry
 - 2.2.1 Telecommunications Industry View
 - 2.2.2 Computer Industry View
 - 2.2.3 Military Sector View
 - 2.3 Three Generations of Cellular Networks
 - 2.3.1 1G Systems and Networks
 - 2.3.2 2G Systems and Networks
 - 2.3.3 3G: W-CDMA for IMT-2000
 - 2.3.4 Beyond 3G and Toward 4G Networks
 - 2.4 Trends in Wireless Technologies
 - 2.4.1 Reemergence of the WLAN Industry
 - 2.4.2 Wireless Home Networking
 - 2.4.3 Home Access Networks
 - 2.4.4 WPANs and Ad Hoc Networking
 - 2.4.5 IEEE 802.15 Working Group on WPAN
 - 2.4.6 HomeRF
 - 2.4.7 Bluetooth
- Questions

2.1 INTRODUCTION

The initial deployment of cellular mobile phone service networks in the early 1980s marked a major milestone in the evolution of the telecommunications industry. There had already been a decades-long history of development in radio communications technology: transoceanic radio telegraphy services beginning at the end of the nineteenth century, land-mobile radio networks for police cars and taxis beginning in the 1920s, military use of *walkie-talkies* in World War II, and AT&T's early *mobile telephone*

service, introduced just after World War II. However, development of the cellular system concept and the widespread deployment of cellular service networks provided an efficient integration of radio technology with the vast public-switched telephone network (PSTN), providing customers with the mobility and convenience of access to at least basic telephone service—*telephony*—over wide geographic areas. The new era of *tetherless access* to public networks had opened up, and the world of communications was changed forever.

First-generation cellular networks, which provided basic voice service using analog FM transmission, were followed in the early 1990s by second-generation networks employing digital voice coding and digital transmission. The second-generation networks provided increased traffic-handling capacity, improved voice quality in poor signal environments, and provided a means for wireless data transmission as well as various enhanced features and services that were already being offered in the PSTN. The cellular telephone service industry grew very rapidly as cellular network coverage and service quality steadily improved, as costs to consumers declined due to advancements in integrated-circuit technology, and as service prices fell in response to industry competition. At this writing, the number of cellular service subscribers worldwide is above 1 billion, representing some 15% market penetration. In some countries the penetration rates are significantly higher. For example, in Finland, where citizens have embraced cellular technology enthusiastically, the penetration rate has reached nearly 75%. Today, the aggregate worldwide income of the wireless industry has already surpassed the income of the wireline telephone industry, and this income is dominated by the income in the cellular telephone industry (see Fig. 2.1).

Although the impressive growth of the wireless industry has been based primarily on the enormous popularity and widespread adoption of cellular voice service, other elements of the industry have been growing as well, including wireless data, paging, text messaging, and the use of wireless local area networks (WLANs). The growth in use of data services and WLANs can be attributed to the ongoing parallel growth in use of the Internet and the Worldwide Web (WWW).

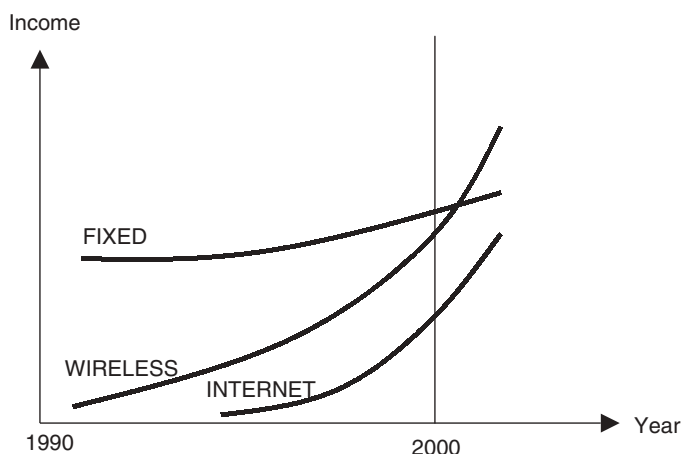


FIGURE 2.1 Relative income growth of the fixed telephone network, wireless, and Internet industries.

Clearly, the wireless industry has become multifaceted, and new developments are emerging constantly, driven by advancing technology and the never-ending demand for new and better services and functionality by a world of users who have learned that wireless communications can transform the ways in which they carry out their personal and professional lives. The evolution of wireless networking technology has been built on developments occurring not only in the telecommunications industry but also in the computer industry, as communications and computer technologies have drawn closer together. In fact, many observers see the wireless industry as one that has integrated radio science with communications and computer technologies. Thus, before delving into details of radio propagation and signal processing—the primary focus of this book—it is useful to consider the wireless industry from the separate viewpoints of the telecommunications industry and the computer industry. We also discuss briefly the view of an important user community, military ground forces, who have had a long history of reliance on wireless communications networks.

2.1.1 Evolution of Voice-Oriented Networks

Table 2.1 is a brief chronology of the evolution of voice-oriented wireless networks. The technology for FDMA analog cellular systems was developed at AT&T Bell Laboratories in the early 1970s. However, the first deployment of these systems took place in the Nordic countries under the Nordic Mobile Telephone (NMT) initiative about a year earlier than the deployment of the Advanced Mobile Phone Service (AMPS) in the United States. In the United States the frequency administration process was slower, resulting in later deployment. The digital cellular networks were first developed in Nordic countries with formation of the GSM standardization group. The GSM group was originally formed to address international roaming, a serious problem for cellular operation in the European Union (EU) countries, where a number of different analog systems were being used and were not interoperable. The standardization group soon decided to standardize on a new digital TDMA technology so as to allow integration

TABLE 2.1 Chronology of Voice-Oriented Wireless Networks

Exploration of first-generation mobile radio at Bell Labs: early 1970s
First-generation cordless phones: late 1970s
Exploration for second-generation digital cordless CT-2: 1982
Deployment of first-generation Nordic analog NMT: 1982
Deployment of U.S. AMPS: 1983
Exploration of the second-generation digital cellular GSM: 1983
Exploration of wireless PBX; DECT: 1985
Initiation for GSM development: 1988
Initiation for IS-54 digital cellular: 1988
Exploration of the Qualcomm CDMA technology: 1988
Deployment of GSM: 1991
Deployment of PHS/PHP and DCS-1800: 1993
Initiation for IS-95 standard for CDMA: 1993
PCS band auction by FCC: 1995
PACS finalized: 1995
3G standardization started: 1998

of other services, thus expanding the horizon of wireless applications [Hau94]. In the United States, however, the motivation for migration to digital cellular was that the growth in analog cellular traffic was predicted to consume the entire capacity of the analog systems in major metropolitan areas such as New York and Los Angeles, and there was a need for increasing system capacity within the constraints of the existing allocated cellular bands. Although Nordic countries, led by Finland, have always had the world's highest rate of cellular penetration, in the early days of this industry the U.S. market was by far the largest. By 1994, there were 41 million subscribers worldwide, 25 million of them in the United States. The need for higher capacity motivated the study of CDMA, which was originally projected to provide capacity at least an order of magnitude higher than other proposed approaches, such as analog band splitting or digital TDMA.

While the debate between proponents of TDMA and CDMA was in progress in the United States, deployment of the GSM technology began in the EU in the early 1990s. At the same time, developing countries began planning for cellular telephone networks, and most of them adopted the GSM digital cellular technology over the legacy analog cellular technology. Soon thereafter, GSM had penetrated into more than 100 countries. An interesting phenomenon in the evolution of the cellular telephone industry was the unexpectedly rapid expansion of this industry in developing countries. In these countries the growth of the infrastructure for wired plain old telephone service (POTS) was slower than the growth in demand for new subscriptions, and a subscriber typically experienced a long waiting time before acquiring a telephone line. As a result, in most of these countries, telephone subscriptions were sold in black markets at highly inflated prices. Penetration of cellular telephone in these countries grew rapidly because subscribers were already accustomed to paying high prices for telephone service. Furthermore, the cellular networks could be built out much more rapidly than could the legacy wired networks.

In the beginning of the race between TDMA and CDMA, the CDMA technology was deployed in only a few countries. Also, on-air experiments had shown that the capacity improvement factor for CDMA was smaller than originally expected. In the mid-1990s, when the first deployments of CDMA technology began in the United States, most cellular service companies were subsidizing the cost of mobile terminals in order to stay in the race with the TDMA and analog alternatives. However, from the start of deployment, the voice quality experienced with CDMA was superior to that of TDMA systems installed in the United States. As a result, CDMA service providers, under such banners as "you cannot believe your ears," began marketing this technology in the United States, and it soon became very popular with users. Meanwhile, given the huge success of digital cellular service, manufacturers worldwide began working on developments for the next-generation IMT-2000 wireless networks. Most of these manufacturers adopted wideband cdma2000 as the technology of choice for IMT-2000, on the premise that CDMA eases integration of services, provides better voice quality, and supports higher capacity than those of proposed alternatives.

The local voice-oriented wireless applications began with the introduction into the market of cordless telephone products in the late 1970s. A cordless telephone provides a wireless connection to replace the wire between a handset and a telephone set. The technology for implementation of a cordless telephone was similar to the

technology used in walkie-talkies, which had been in use in World War II. As soon as the cordless telephone was introduced to the market, it became a major commercial success, selling on the order of tens of millions of phones and generating revenues exceeding several billion dollars. The success of the cordless telephone encouraged further developments in this field. The first digital cordless telephone was CT-2, a standard developed in the UK in the early 1980s. The next generation of cordless telephones was wireless PBX using the Digital European Cordless Telephone (DECT) standard. Both CT-2 and DECT required minimal network infrastructures beyond what was required for the simple cordless telephone, and each covered a larger area and supported multiple applications. However, despite the huge success of the cordless telephone, neither CT-2 nor DECT has yet been considered a great commercial success. These local systems soon evolved into personal communication systems (PCSs), each a complete system with its own infrastructure, very similar to the cellular mobile telephone system.

In the technical communities of the early 1990s, PCS systems were differentiated from cellular systems, as indicated in Fig. 1.1. A PCS service was considered the next generation of cordless telephone designed for residential areas, providing a variety of services beyond those supported by the cordless telephone. The first real deployment of PCS systems was the Personal Handy Phone (PHP), later renamed the Personal Handy System (PHS), introduced in Japan in 1993. At that time, the technical differentiator for PCS relative to cellular was perceived to be smaller cell size, better speech quality, lower tariff, lower power consumption, and lower mobility. However, from a user's point of view, the mobile terminals and services for PCS and cellular looked very similar, and the only significant difference was marketing strategy and the way that they were introduced to the market. For instance, at around the same time that PCS was being introduced in the United States, DCS-1800 service was introduced in the UK as a PCS service. The DCS-1800 systems used GSM technology at a higher frequency of 1800 MHz but were marketed with a different strategy. The last PCS standard was PACS in the United States, finalized in 1995. All together, none of the PCS standards became a major commercial success competing with cellular services.

In 1995 the FCC in the United States auctioned off the frequency bands around 2 GHz as *PCS bands*, but PCS-specific standards were not adopted for these frequencies. Eventually, the name *PCS* started to appear only as a marketing identity used by some service providers for digital cellular services, and in some cases the services offered did not even operate in PCS bands. Whereas the more advanced and complex PCS services evolving from simple cordless telephone application did not succeed and merged into the cellular telephone industry, the simple cordless telephone industry itself remains active. In more recent years the frequency of operation of cordless telephone products has shifted into unlicensed ISM bands rather than licensed PCS bands. Cordless telephones operating in the ISM bands can provide a more reliable wireless connection since they use spread-spectrum technology.

2.1.2 Evolution of Data-Oriented Networks

Table 2.2 outlines the chronology of development of data-oriented networks. As discussed in Chapter 1, data-oriented wireless networks can be divided into wide-area

TABLE 2.2 History of Data-Oriented Wireless Networks
Diffused infrared: 1979 (IBM Rueschlikon Labs, Switzerland)
Spread spectrum using SAW devices: 1980 (HP Labs, California)
Wireless modems: early 1980s (Data Radio)
ARDIS: 1983 (Motorola/IBM)
ISM bands for commercial spread-spectrum applications: 1985
Mobitex: 1986 (Swedish Telecom and Ericsson)
IEEE 802.11 for wireless LAN standards: 1990
Announcement of wireless LAN products: 1990
RAM mobile: 1991 (Mobitex)
Formation of WINForum: 1992
ETSI and HIPERLAN in Europe: 1992
Release of 2.4-, 5.2-, and 17.1- to 17.3-GHz bands in EC: 1993
PCS licensed and unlicensed bands for PCS: 1994
CDPD: 1993 (IBM and nine operating companies)
Wireless ATM Forum started: 1996
U-NII bands released, IEEE 802.11 completed, GPRS started: 1997
IEEE 802.11b and Bluetooth announcement: 1998
IEEE 802.11a/HIPERLAN2 started: 1999

TABLE 2.3 Properties of ISM Bands
Frequencies of operation: 902 to 928 MHz; 2.4 to 2.4835 GHz; 5.725 to 5.875 GHz
Transmitter power limitation of 1 W for DSSS and FHSS
Low power with any modulation

wireless data and local broadband and ad hoc networks. Wireless local networks support higher data rates and ad hoc operation for lower numbers of users. Broadband wireless local networks are usually referred to as WLANs, and ad hoc local networks as WPANs. The concept of WLAN was introduced around 1980. However, the first WLAN products did not emerge until about 10 years later. Today, a key feature of local broadband and ad hoc networks is operation in unlicensed bands. The first unlicensed bands were the *industrial, scientific, and medical* (ISM) bands made available in the United States in 1985. Table 2.3 provides a summary of the ISM bands. Later, in 1994 and 1997, the PCS and U-NII unlicensed bands were also designated in the United States.

The major WLAN standardization activity is IEEE 802.11, begun in the late 1980s and completed in 1997. The IEEE 802.11 and 802.11b operate in the ISM bands, and IEEE 802.11a operates in the U-NII bands. Another extension of 802.11, IEEE 802.11g, ratified in mid-2003, provides data rates and performance comparable to 802.11a but operates in the 2.4-GHz band. The competing European standard for WLAN is HIPER-LAN, developed by the Broadband Radio Access Networks (BRAN) division of ETSI.

The HIPERLAN1 standard was completed in 1997. Its successor, HIPERLAN2, is similar to IEEE 802.11a but operates in the 5-GHz band. However, by the time that the HIPERLAN2 standard was settled, adoption of the IEEE 802 standards was spreading widely, and the 802.11b and 802.11a standards now dominate the WLAN marketplace.

In 1996 the wireless ATM Forum was formed to merge ATM technology with wide-band local access. More recently, following the announcement of Bluetooth technology in 1988, WPANs have attracted tremendous attention. WPANs exhibit more restricted coverage than do traditional WLANs, and they are intended to provide a better ad hoc environment for interconnecting personal equipment such as laptop, cell phone, and headset. Today, IEEE 802.11-based products generate most of the income for this industry, currently about half a billion dollars per year. In the past several years, major investments have been made in WLAN and WPAN chip-set developments all over the world. These investments are being made with the expectation of sizable sales volumes generated by integration of WLANs with cellular systems as well as a large WPAN market for consumer products and home-networking systems.

Mobile data services were introduced in 1983 with the ARDIS project, a collaboration between Motorola and IBM. The purpose of this network was to allow IBM field crews to operate their portable computers on customer premises. In 1986, Ericsson introduced Mobitex technology, which was an open-architecture implementation of the ARDIS system. In 1993, IBM and nine operating companies in the United States initiated the Cellular Digital Packet Data (CDPD) project, expecting a huge market by the year 2000. In late 1990s, GPRS data services were integrated into the successful GSM systems and can support an order-of-magnitude-higher data rates than those of previous technologies, attracting considerable attention. These higher data rates are perceived to be essential for wireless Internet access, thus far the most popular wireless data application. The third-generation cellular systems are projected to provide a mobile data service up to 2 Mb/s, substantially higher than the GPRS data rates. The third-generation data rates would not, however, have the comprehensive geographic coverage of GPRS. The early mobile data networks, ARDIS and Mobitex, were independent networks owning their infrastructure. In contrast, CDPD service used infrastructure overlaid onto AMPS systems, and GPRS was actually integrated into the GSM infrastructure. Thus, we have seen the gradual assimilation of the mobile data industry into the cellular telephone industry, and this will be completed in the next generation of cellular systems.

With integration of the PCS and mobile data industries into the next generation of cellular systems, we see the emergence of two industries: next-generation wide-area cellular systems operating in licensed bands, and local broadband and ad hoc networks operating in unlicensed bands.

2.2 THREE VIEWS OF THE WIRELESS INDUSTRY

In this section we consider the wireless industry from three different viewpoints: those of the telecommunications industry, the computer industry, and the military sector.

2.2.1 Telecommunications Industry View

The telecommunications industry has its origins in Alexander Graham Bell's 1870 invention of the telephone, and this rapidly evolved into the U.S.-wide public telephone network developed by scientists and engineers working at the Bell Telephone

Laboratories. Today, of course, the telephone network is massive, serving and connecting all the developed nations of the world and enabling almost instantaneous voice communications between any two parties subscribing to telephone service. In the years immediately following World War II, the United States was rapidly rebuilding the civilian economy and there was robust demand for new products and services, including, in the telephone service industry, a demand by some subscribers for a mobile telephone service affording radio access to the public-switched telephone network (PSTN). Simple one-way dispatch radio systems had been in use by police departments, taxicab fleets, and so on, since the 1920s, and the wartime use of Army walkie-talkie radios was already well known to much of the general public. Consequently, mobile telephone service, the interconnection of mobile users with the PSTN, was introduced in 1946, when the FCC granted AT&T a license to operate such a service in St. Louis. In less than a year, mobile telephone service was being offered in more than 25 cities in the United States. Demand for mobile telephone service grew rapidly over the next several decades, and finally, in 1975, the FCC allocated spectrum for cellular mobile telephone service and related mobile wireless services.

The first commercial cellular system in the United States, called *Advanced Mobile Phone Service* (AMPS), went into operation in 1983 in Chicago, and the demand for cellular telephone service grew rapidly, in part stimulated by the growing popularity of cordless telephones. Later in the 1980s, digital cellular technology was introduced with the pan-European GSM system and the U.S. TDMA system. Then CDMA cellular technology was introduced with the IS-95 system. Thus, we see the evolution of successive cellular system designs as meeting the steadily growing demands of telephone service subscribers wanting to have mobile access to the PSTN, and the succeeding demands for greater system capacity and for new features and services, including data services.

2.2.2 Computer Industry View

Just as cordless telephones and cellular telephone services met the demands of subscribers for untethered wireless access to the PSTN, early wireless local area networks (WLANs) met the demand for a wireless replacement of the wired attachment of office computers to the installed wired Ethernet LAN. WLAN technology provided the convenience of being able to relocate computers without incurring the costs of rearranging cabling. Also, WLANs provided a logical solution for buildings in which wiring installations would be difficult or especially expensive. Thus, the initial motivation for WLAN technology was simply cable replacement, with resulting savings in LAN-installation and LAN-maintenance budgets. As the computer industry designed and produced portable, laptop, and palm-top computers in steadily more compact and lightweight configurations, the computer users' demands for wireless connectivity became steadily more sophisticated. The laptop computer became a standard piece of equipment carried by the business road warrior, as essential to doing business as the cell phone or pager. Users wanted not only wireless connectivity to the home-office wired LAN, but also wanted wide-area wireless access to the PSTN and to wired data networks and eventually, to the Internet. Along with these user demands came demands for steadily higher data rates on wireless connections from laptop computers. These demands stimulated the development of wireless campus-area networks (W-CANs) and broadband Internet access. Thus, we are seeing the evolution of the laptop computer into a computing-and-communicating device.

2.2.3 Military Sector View

Radio communications has been an important part of the operations of military forces for many decades. A prominent example is the *walkie-talkie radio*, developed by Motorola and used extensively by U.S. soldiers beginning in World War II. The walkie-talkie evolved into a more modern system called a *net-broadcast radio system*, in which a user can talk with any other individual user in a defined group or talk simultaneously to a number of users tuned to the same frequency channel. As communications technology advanced, the military saw a need for *antijamming* (A/J) communications capability, to thwart various forms of deliberate interference with transmitted signals. Beginning in the 1950s, the defense communications industry invested a great amount of effort in research and development on concepts and systems for A/J communications.

A key technique central to these developments was spread-spectrum communications, employing various forms of *direct-sequence* or *frequency-hopped transmission*. A closely related development was that of *code-division multiple access* (CDMA), allowing multiple users to have simultaneous access to a communication channel with a controllable amount of user-to-user interference. For many years, the spread-spectrum and CDMA technologies were considered to be too expensive for use in commercial communications systems, but the ongoing advances in microelectronics and the economies of scale achieved in commercial product manufacturing eventually brought these technologies down to price points that made them feasible for use in commercial communications systems. Given this background, in the early 1990s Qualcomm Corporation developed a digital cellular system based on spread-spectrum signaling and CDMA and proposed this design as a new standard for digital cellular telephone communications. The design was subsequently adopted as the IS-95 standard and is now the predominant cellular telephone technology in the United States.

Along with A/J capability, military forces have always had a critical need for information security in all modes of communication, including wireless networking. This motivated research over many decades in cryptographic techniques for protecting transmitted information. Commercial versions of military cryptographic techniques have thus been adopted in digital cellular networks and into WLAN products as well.

Other wireless communications techniques developed to meet the needs of military operations are also finding their way into commercial system developments. One example is *ad hoc networking*, aimed at meeting the military need for reliable communication connectivity with steadily increasing degrees of mobility in combat operations. Along with the evolution of military connectivity and mobility requirements, we see requirements for integration of geolocation awareness into military communications systems. Another system concept developed for the military in recent years is that of the *body LAN* or *wearable LAN*, in which wireless LAN devices are integrated into a soldier's uniform and into the equipment the soldier carries in a combat environment. Wireless communication between these embedded devices and between the devices and a fixed network can be used to support tactical operations as well as medical operations when soldiers are wounded in battle.

Another important development is the use of *space-time coding*, offering further advances in reliability and capacity in wireless communications networks. Thus, we can expect to see a continuation of new wireless systems concepts and techniques, motivated by stringent military communications requirements, eventually migrating into affordable commercial network technologies.

2.3 THREE GENERATIONS OF CELLULAR NETWORKS

The evolution of successive versions of cellular telephone technology led cellular telephone manufacturers and service providers to categorize wireless communication systems into defined *generations*, and this terminology has been extended to cover all categories of wireless systems, not only cellular systems. *First-generation* (1G) systems comprise voice-oriented analog cellular and cordless telephones. *Second-generation* (2G) voice-oriented wireless systems comprise digital cellular and PCS systems and data-oriented wireless wide- and local-area networks as well. *Third-generation* (3G) networks integrate cellular and PCS voice services with a variety of packet-switched data services in a unified network. Simultaneously with the unified 3G standardization efforts, standardization also proceeded on broadband local area and ad hoc networks, as these networking technologies attracted considerable attention. One of the major current differences between these two sets of technologies is that 3G systems operate in licensed bands, whereas broadband local area and ad hoc networks operate in unlicensed bands. The potential for integrating broadband local access in unlicensed bands with the 3G standards is a key issue to be addressed in standardizing forthcoming generations of wireless networks.

The industry's vision for the fourth generation (4G) is still being formulated, but many observers expect that 4G will include integration of wide-area networks operating in licensed bands with WLANs operating in unlicensed bands, improved voice service quality, and higher rates for data services, as well as incorporation of emerging technologies such as location-aware services and ad hoc networking.

2.3.1 1G Systems and Networks

Table 2.4 summarizes the worldwide 1G analog cellular systems. All these systems use two separate frequency bands for forward (from base station to mobile) and reverse (from mobile to base) channels, a scheme referred to as *frequency-division duplex* (FDD) *transmission*. The typical allocated overall band in each direction (e.g., for AMPS, TACS, and NMT-900) was 25 MHz in each direction. The dominant frequencies of operation for these systems were in the 800- and 900-MHz bands. In an ideal situation, all countries would use the same cellular standard and the same frequency bands, ensuring compatibility and interoperability among all mobile devices and base stations. However, in practice, as shown in Table 2.4, a variety of frequencies and standards were adopted in various countries and regions of the world. The reason for the differences in frequencies of operation was that the frequency administration agencies in the various countries had to abide by earlier frequency allocation rulings that restricted the assignment choices. The reason for adopting different standards was that cellular providers then assumed that services would be used solely within one country and did not envision an eventual universal service. The channel spacing or bandwidth allocated to each user was either 30 or 25 kHz or a fraction of either. The 25-kHz channel spacing was used previously for mobile satellite services, but the 30-kHz channel spacing was a new allocation for cellular telephone application.

All of the 1G cellular systems used analog frequency modulation (FM), for which the transmission power requirement depends on the transmission bandwidth. On the other hand, power is also related to signal coverage and to the size of mobile radios.

TABLE 2.4 Worldwide First-Generation Cellular Systems

Standard	Forward Band (MHz)	Reverse Band (MHz)	Channel Spacing (kHz)	Region	Comments
AMPS	824–849	869–894	30	United States	Also in Australia, SE Asia, Africa
TACS	890–915	935–960	25	EC	Later, bands were allocated to GSM
ETACS	872–905	917–950	25	UK	
NMT 450	453–457.5	463–467.5	25	EC	
NMT 900	890–915	935–960	12.5	EC	Frequency overlapping, also in Africa and SE Asia
C-450	450–455.74	460–465.74	10	Germany, Portugal	
RTMS	450–455	460–465	25	Italy	
Radiocom 2000	192.5–199.5 215.5–233.5 165.2–168.4 414.8–418	200.5–207.5 207.5–215.5 169.8–173 424.8–428	12.5	France	
NTT	925–940 915–918.5 922–925	870–885 860–863.5 867–870	25/6.25 6.25 6.25	Japan	First band is nationwide, others regional
JTACS/NTACS	915–925 898–901 918.5–922	860–870 843–846 863.5–867	25/12.5 25/12.5 12.5	Japan	All are regional

Therefore, one can compensate for the reduction in transmission bandwidth per user by reducing the size of a cell in a cellular network. Reduction in cell size increases the number of cells and the cost of installation of the infrastructure. By way of example, the AMPS system in North America uses 30-kHz channel spacing, whereas C-450 in Germany uses 10-kHz spacing, one-third the AMPS channel spacing. Therefore, one expects a denser infrastructure for deployment of C-450. As another example, Japan has several systems using full- and split-band operation, with 25 and 12.5 kHz being used in different systems. The cell sizes for split-band operation are smaller than for full-band operation.

A technique called *band splitting* can be utilized to support increased traffic capacity in a service network without having to increase the number of base stations. However, this technique incurs the need for increased investment in network infrastructure [Pah02a].

In the wireless industry, 1G often refers only to analog cellular technology because it is the only system implemented based on popular standards such as AMPS or NMT. However, we can generalize the designation *1G systems* to include other types of wireless services and products. The analog cordless telephone, which appeared in the market in the 1980s, can be considered as a 1G cordless telephone product. Paging services, which were deployed at around the same time as analog cellular systems and cordless telephones, can be referred to as 1G mobile data services providing one-way transmission of short data messages. In the early 1980s, before release of the ISM bands and the start of the WLAN industry, a few small companies in Canada and the United States developed low-speed connectionless wireless local area networks using voiceband modem chip sets and commercially available walkie-talkies. These products operated at the speed of voiceband modems (<9600 bits/s) but used the medium access control techniques then found in data-oriented LANs. Although because of their low data rates, they do not comply with the IEEE 802 community definition for LANs, one may refer to them as 1G wireless LAN products.

2.3.2 2G Systems and Networks

The 2G systems evolved as soon as the wireless industry perceived that the demand for cellular services was growing rapidly and that the analog networks in major market areas would quickly reach saturation. The industry also recognized that customer demand was growing for wide-area wireless data services. Consequently, 2G systems were designed to support a complete set of standards for all four sectors of the wireless information network industry. As we discussed in reviewing the evolution of voice- and data-oriented networks, there are a number of digital cellular, PCS, mobile data, and wireless LAN standards and products that can be classified as 2G systems. In the remainder of this section we cover each of these four categories of 2G systems in a separate subsection.

2G Digital Cellular Systems. Table 2.5 summarizes the major 2G digital cellular standards. There are four standards in this category: (1) GSM, the pan-European

TABLE 2.5 Second-Generation Digital Cellular Standards

System	GSM	IS-54	JDC	IS-95
Region	Europe, Asia	United States	Japan	United States, Asia
Access method	TDMA/FDD	TDMA/FDD	TDMA/FDD	CDMA/FDD
Modulation scheme	GMSK	$\pi/4$ -DQPSK	$\pi/4$ -DQPSK	SQPSK/QPSK
Frequency band (MHz)	935–960 890–915	869–894 824–849	810–826	869–894
			940–956	824–849
			1477–1489	
			1429–1441	
			1501–1513	
Carrier spacing (kHz)	200	30	1453–1465	
			25	1250
			3	Variable
			42	1228.8
			8	1–8 (variable)
Speech coding (kb/s)	13	8	8	1–8 (variable)
Frame duration (ms)	4.615	40	20	20

digital cellular standard; (2) IS-54, which evolved into IS-136 on the North American continent; (3) JDC in Japan; and (4) IS-95 on the North American continent. The first three of these standards all use TDMA technology; the fourth, IS-95, uses CDMA technology.

As in 1G analog systems, 2G systems all utilize FDD transmission and operate in the bands from 800 to 900 MHz. The channel spacing in IS-54 and JDC is the same as channel spacing in 1G analog systems in their respective regions, although GSM and IS-95 use the bandwidth of multiple analog channels to form one digital channel. GSM supports eight users in a 200-kHz digital channel; IS-95 and JDC support three users in 30 and 25 kHz, respectively. As we explain in Chapter 11, where we discuss *access methods*, the number of users that can be served by a CDMA system depends on the acceptable quality of service, and therefore the number of users in a 1250-kHz CDMA channel cannot be fixed theoretically. However, this number is high enough that considering the superior voice quality achieved with CDMA, the CDMA technology has been dominant in the planning for next-generation 3G standards.

In examining the spectrum utilization numbers for these 2G systems, one might come to the conclusion that GSM uses 25 kHz of bandwidth for each caller, whereas IS-95 typically uses about 10 kHz per caller, and therefore GSM supports 2.5 time fewer calls in a given bandwidth. However, the reader should be aware that this is an illusory conclusion, because when the network is deployed, the quality of service delivered also depends on the *frequency reuse factor* and signal-to-noise interference requirements, which will change these calculations significantly. These issues are addressed in Chapter 11.

The channel bit rate in the GSM standard is 270 kb/s, whereas IS-54 and JDC use 48 and 42 kb/s, respectively. The higher channel bit rate in a digital cellular system allows convenient implementation of higher data rates for data services. By assigning several voice slots to one user on a single carrier, one can easily increase the maximum supportable data rate for a data service offered by the network. The higher channel rate of GSM, which utilizes eight voice slots, allows support of higher data rates, as we discuss in Chapter 15, where we treat GPRS and EDGE mobile data services. Using a similar argument, one may notice that the 1228.8-kb/s channel bit rate of IS-95 provides a good framework for integration of higher data rates into the IS-95 standards. This fact has been exploited in 3G wideband CDMA systems to support data rates up to 2 Mb/s.

Cellular standards were developed with an expectation of large cell sizes and a large number of users per cell, which necessitates lower speech coding rates. Thus, the speech coding techniques used in 2G systems all operate at around 10 kb/s. On the other hand, those standards were developed initially assuming installation of mobile phones in automobiles, where power consumption and battery life were not an issue. The peak transmission power of mobile terminals in these standards ranges from several hundred milliwatts up to 1 W [Pah95], and on the average they consume around 100 mW. All these systems employ central power control, which reduces battery consumption and helps in controlling the overall interference level in the network. In digital communication, information is transmitted in packets. The duration of a packet frame in the transmission channel should be short enough so that the channel does not change significantly during the transmission, and long enough that the required guard-time gap between packets is much smaller than the length of the packet. A frame length of around $\frac{1}{2}$ to several tens of milliseconds is typically used in voice-oriented digital cellular networks.

2G PCS Systems. As we discussed in reviewing the development history of wireless voice-oriented networks, 2G PCS standards evolved out of the 1G analog cordless telephone industry and later merged into 3G cellular systems. Table 2.6 illustrates a quantitative comparison of PCS and cellular industries that at its time was used to justify the existence of two separate voice-oriented standards. The basic philosophy was that PCS was intended for residential applications, and small cell sizes, zonal coverage, and antennas installed on existing structures (such as utility poles). Since PCS was not intended for high-mobility use, the complexity of the handsets and base stations was low. These standards incorporated 32-kb/s speech coding to provide customers with voice quality comparable to that of wireline service. Furthermore, in the interest of achieving simpler implementation, PCS systems shared the same spectrum in different zones, and most systems used time-division-duplex (TDD) and noncoherent modulation techniques.

Table 2.7 provides a summary of specifications for the four major PCS standards. CT-2 and CT-2+ were the earliest digital cordless telephone standards; PHS, which

TABLE 2.6 Quantitative Comparison of PCS and Cellular Characteristics

System Aspect	PCS	Cellular
Cell size	5–500 m	0.5–30 km
Coverage	Zonal	Comprehensive
Antenna height (m)	<15	>15
Vehicle speed (km/h)	<5	<200
Handset complexity	Low	Moderate
Base station complexity	Low	High
Spectrum access	Shared	Exclusive
Average handset power (mW)	5–10	100–600
Speech coding	32-kb/s ADPCM	7- to 13-kb/s vocoder
Duplexing	Usually TDD	FDD
Detection	Noncoherent	Coherent

TABLE 2.7 Second-Generation PCS Standards

System	CT2+	DECT	PHS	PACS
Region	Europe, Canada	Europe	Japan	United States
Access method	TDMA/TDD	TDMA/TDD	TDMA/TDD	TDMA/FDD
Frequency band (MHz)	864–868 944–948	1880–1900	1895–1918	1850–1910 1930–1990
Carrier spacing (kHz)	100	1728	300	300, 300
Bearer channels/carrier	1	12	4	8 per pair
Channel bit rate (kb/s)	72	1152	384	384
Modulation	GFSK	GFSK	$\pi/4$ -DQPSK	$\pi/4$ -DQPSK
Speech coding (kb/s)	32	32	32	32
Handset Tx power (mW)				
Average	5	10	10	25
Peak	10	250	80	200
Frame duration (ms)	2	10	5	2.5

later became PHP, was the first and the only one of these systems to be deployed nationwide; and PACS is the last standard developed with this philosophy. Except for CT2+, all of these standards were designed for operation in the 1.8- and 1.9-GHz frequency bands, which are commonly referred to as *PCS bands*; all use TDMA/TDD except PACS, which adopted frequency-division duplex (FDD) for two-way transmission.

To support voice quality comparable to that of wireline service, speech coding at 32 kb/s is used in all of these standards. This rate is about three times higher than the speech-coding rate used in digital cellular systems. The channel bandwidth (1.728 MHz) in DECT is even higher than that in GSM (200 kHz), which had the highest channel bandwidth of the TDMA digital cellular systems. This channel bandwidth is even higher than in IS-95 (1.2288 MHz), the 2G CDMA standard. This feature provides an advantage to DECT in supporting high-speed data connections for Internet access.

Power consumption in PCS systems is almost one order of magnitude lower than the power consumption in digital cellular systems because PCS systems are designed for smaller cells. If digital cellular systems were deployed with the same cell sizes, the average power consumption could be comparable to that of PCS systems. The modulation techniques used for PCS standards, GFSK and DQPSK, are less bandwidth efficient and more power efficient than are the modulation techniques used in digital cellular systems. These modulation techniques can be implemented with simpler noncoherent receivers. The shorter propagation time for the short-distance PCS standards allows shorter packet frames, benefiting the voice quality despite the presence of wireless channel impairments.

Mobile Data Services. Mobile data services provide wide-area access to packet-switched data networks at moderate data rates. Following the success of the paging industry, mobile data networks emerged to provide two-way transmission for longer messages. Table 2.8 provides a comparison among a number of important mobile data services. ARDIS and Mobitex use their own frequency bands in the region 800 to 900 MHz; Terrestrial European Trunked Radio (TETRA) uses its own band at 300 MHz; CDPD shares the AMPS bands and site infrastructure; and GPRS shares GSM's complete radio system.

TABLE 2.8 Mobile Data Services

System	ARDIS	Mobitex	CDPD	TETRA	GPRS
Frequency band (MHz)	800 bands, 45-kHz separation	935–940 896–961	869–894 824–849	380–383 390–393	890–915 935–960
Channel bit rate (kb/s)	19.2	8.0	19.2	36	200
RF channel spacing (kHz)	25	12.5	30	25	200
Channel access/multiuser access	FDMA/DSMA	FDMA/dynamic S-Aloha	FDMA/DSMA	FDMA/DSMA	FDMA/TDMA/reservation
Modulation technique	4-FSK	GMSK	GMSK	$\pi/4$ -QPSK	GMSK

The early systems, ARDIS, Mobitex, and CDPD, were developed before the growth in popularity of the Internet, and the dominant design criteria were coverage and cost rather than data rate. These systems provided a wireless replacement for voiceband modems operating at data rates up to 19.2 kb/s, which was the achievable rate of these modems at that time. TETRA is designed for pan European civil service application and has its own features for that purpose. GPRS supports data rates more suitable for Internet access. The advantage of GPRS is that it is incorporated into the popular GSM digital services, with a large number of terminals deployed all over the world. Thus, the early mobile data systems have largely been overtaken by data services integrated into the GSM and CDMA cellular networks.

Channel spacing used in mobile data service networks is based on the channel spacing of cellular telephone networks, with 30- or 25-kHz bands or a fraction (12.5 kHz) or a multiple (200 kHz) of them. These services are designed to use multiple carriers in an FDMA format and use different versions of random access techniques such as DSMA, BTMA, or ALOHA, discussed in Chapter 11, which deals with access methods. Modulation techniques used in these systems are like those in digital cellular and PCS systems.

Wireless LANs. Wireless LANs provide high-data-rate (minimum of 1 Mb/s) access in a local area (<100 m) to wired LANs and the Internet. Today, all successful wireless LAN products operate in the unlicensed bands. Each new product design must undergo FCC approval, but the owner of the WLAN equipment may deploy the equipment at will, and its operation requires no license and is not subject to further regulation.

Considering that the PCS bands had been auctioned off at very high prices, in the past several years users have given renewed attention to the use of wireless LANs. Table 2.9 summarizes the IEEE 802.11 family of standards for wireless LAN products. The IEEE standards include 802.11 and 802.11b operating at 2.4 GHz, 802.11a operating at 5 GHz, and 802.11g operating at 2.4 GHz. Another extension of the IEEE family, IEEE 802.11n, intended for even higher data rates, is still under development, and completion is expected by late 2006.

Table 2.10 summarizes the HIPERLAN standards. Both HIPERLAN1 and 2, developed under ETSI, operate at 5 GHz. The standardization initiatives for WLANs operating in the 5-GHz bands led the FCC in 1997 to release the Unlicensed National Information Infrastructure (U-NII) bands, summarized in Table 2.11.

TABLE 2.9 IEEE 802.11 Specifications

Parameter	802.11b	802.11a	802.11g
Standard approved	July 1999	July 1999	June 2003
Maximum data rate (Mb/s)	11	54	54
Modulation	CCK	OFDM	OFDM and CCK
Data rates (Mb/s)	1, 2, 5.5, 11	6, 9, 12, 24, 48, 54	CCK: 1, 2, 5.5, 11 OFDM: 6, 9, 12, 24, 36, 48, 54
Frequencies (GHz)	2.4–2.497	5.15–5.35 5.425–5.675 5.725–5.875	2.4–2.497

TABLE 2.10 HIPERLAN Standards

Parameter	HIPERLAN2	HIPERLAN1
Frequency band (GHz)	5	5
PHY layer, modulation	OFDM	GMSK
Data rate (Mb/s)	6, 9, 12, 18, 24, 36, 54	23.5
Access method	Central control, reservation- based access	Active contention resolution, priority signaling

TABLE 2.11 Properties of U-NII bands

Band of Operation (GHz)	Maximum Tx Power (mW)	Maximum Power with Antenna Gain of 6 dBi (mW)	Maximum PSD (mW/MHz)	Applications: Suggested and/or Mandated	Other Remarks
5.15–5.25	50	200	2.5	Restricted to indoor applica- tions	Antenna must be an integral part of the device
5.25–5.35	250	1000	12.5	Campus LANs	Compatible with HIPERLAN
5.725–5.825	1000	4000	50	Community networks	Longer range in low-interference (rural) environs

The 2.4-GHz products operate in ISM bands using spread-spectrum technology to support data rates ranging from 1 to 11 Mb/s. HIPERLAN1 uses GMSK modulation with decision feedback equalization (DFE) signal processing at the receiver and supports rates up to 23.5 Mb/s. The IEEE 802.11a and HIPERLAN2 standards use an OFDM physical layer to support data rates up to 54 Mb/s. The access method for all IEEE 802.11 standards is the same and includes CSMA/CA, PCF, and RTS/CTS, which are described in Chapter 11. The access method of HIPERLAN1 is similar to that of 802.11, but the access method for HIPERLAN2 is a voice-oriented access technique suitable for integration of voice and data services. The IEEE 802.11 and HIPERLAN standards can be considered as 2G wireless LANs. The 3G wireless LANs use OFDM modulation. The IEEE 802.11g standard, approved in June 2003, operates in the 2.4-GHz band, using DSSS and OFDM, providing data rates up to 54 Mb/s. We describe these systems in further detail in Chapter 9, under the topic of broadband modem technologies.

2.3.3 3G: W-CDMA for IMT-2000

The motivation for migrating to 3G technologies was to develop an international standard combining and gradually replacing 2G digital cellular, PCS, and mobile data

services. The 3G systems were expected to improve voice quality, expand network capacity, and increase the data rates of wireless data services. The primary standard for 3G systems is referred to as *International Mobile Telecommunications Beyond the Year 2000* (IMT-2000).

Among the several *radio transmission technology* (RTT) proposals submitted to the International Telecommunications Union (ITU), most were based on the use of CDMA. Given the experience gained with 2G cellular systems, it was recognized that CDMA systems provide voice quality superior to that of other systems. Furthermore, CDMA provides a very flexible air-interface design amenable to customization for higher-rate multimedia applications.

Two Approaches for IMT-2000. In the deliberations on RTT proposals for IMT-2000, there were two major overall approaches: (1) wideband CDMA (W-CDMA) based on the UMTS Terrestrial Radio Access (UTRA) FDD and TDD proposals, and (2) the cdma2000 proposal, which is backwardly compatible with the IS-95 standard. The first approach is intended to build on the success of the installed GSM infrastructure (UMTS/IMT-2000); the second approach is intended to build on the experience with cdmaOne (cdma2000/IMT-2000). The distinctions between these two overall approaches lie mainly in chip-rate selection, synchronous-versus-asynchronous base station operation, and pilot structure [Zen00].

3G Perspective on Wireless Access Methods. Figure 2.2 provides an overview of current wireless access methods. The horizontal axis shows user bit rates in Mb/s, and the vertical axis gives an indication of relative user mobility for various categories of networks. For wide-area networks, in the low-bit-rate region, we see 2G cellular systems with bit rates limited to about 50 kb/s but with a wide range of mobility options. The 3G cellular systems provide the same mobility characteristics and offer data rates up to about 1 Mb/s. WLANs provide even higher data rates, 10 Mb/s and higher, but with more restricted user mobility. Wireless personal area networks (WPANs) provide

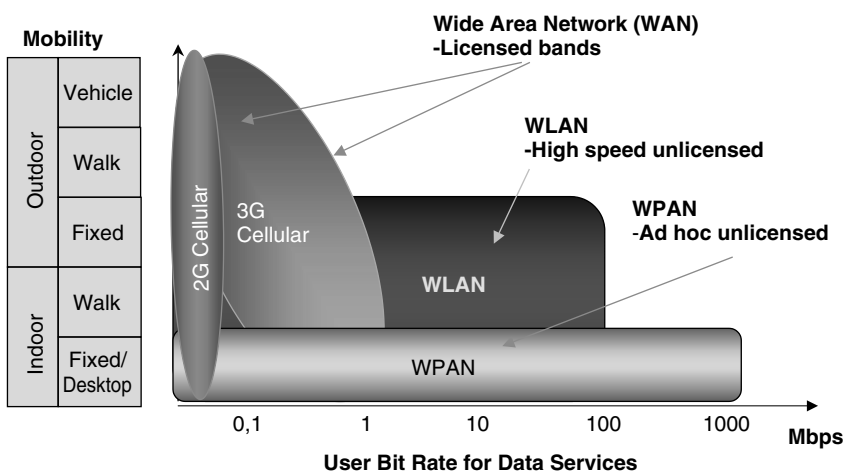


FIGURE 2.2 Comparison of second- and third-generation cellular, local broadband, and ad hoc networks relative to mobility and data rate.

data rates comparable to those of 3G cellular systems but are designed to enable connectivity between wireless devices over relatively short distances. Of course, a major distinction between the WANs, on the one hand, and WLANs and WPANs, on the other, is licensed versus unlicensed operation, respectively.

Just as was the case for the 2G market, it is useful to distinguish two sectors of the 3G market, but here the two sectors must be defined somewhat differently to reflect the changes that have taken place in the evolution from 2G to 3G. First, we can identify a 3G market sector characterized as a voice-oriented cellular market that integrates cellular, PCS, and mobile data systems and services. The 3G IMT-2000 standards harmonization process will provide the underlying access methods in these systems. These systems will all operate in licensed bands subject to frequency administration and regulation.

Second, we have a data-oriented market sector characterized by broadband and ad hoc wireless systems. This sector includes traditional WLAN products operating in unlicensed bands providing wireless Internet access as well as newer WPANs and emerging ad hoc networks providing wireless connectivity between consumer devices.

2.3.4 Beyond 3G and Toward 4G Networks

In the last few decades, the telecommunications industry has become especially responsive to market demands for new services and capabilities. This has been particularly true of the wireless segment of the industry, which has seen vigorous growth from cordless phones and first-generation analog cellular networks through 2G digital networks, low-speed mobile data networks, paging systems, and now 3G technologies that provide improved voice quality and integration of data and voice services. It has always been difficult to predict the future of the wireless communications industry, but there are certain trends that one can discern and try to project.

WAN and WLAN Integration. With respect to the critical issue of spectrum allocation and administration, we see 3G systems operating in licensed bands, where service providers must make large investments to secure access to those licenses. On the other hand, WLANs and WPANs operate in unlicensed bands, where one does not need to purchase spectrum and where the user is unencumbered by regulatory rules and regulations. However, there is also no regulatory control of signal interference in the unlicensed bands, and thus connectivity and link performance can often be problematic. It would not be wise to predict that all wireless communications will migrate to unlicensed bands, but it is accurate to say that the last several years have witnessed a renewed interest and vigorous growth in the use of unlicensed-band systems. One possible migration path is the eventual integration of WANs with WLANs in unlicensed bands.

Ad Hoc Networking. Another important evolving technology is ad hoc networking, which uses a distributed network topology (see Chapter 11) and has the capability for network reconfiguration without the need for a geographically fixed infrastructure. This technology was developed for military networking requirements but has found some application in commercial voice and data services. The ad hoc networking topology is suitable, as an example, for rapid deployment of any wireless network in a mobile or fixed environment.

UWB and S-T Coding. It is clear that CDMA is emerging as the preferred transmission technology for 3G systems, providing enhanced voice quality and increased network capacity relative to 2G systems, while OFDM has been adopted in WLANs operating at 5 GHz. It is safe to project that OFDM will continue to play an important role in the future of broadband wireless access. Other important emerging technologies include *ultrawideband (UWB) communication* and *space-time (S-T) coding*. The UWB concept (see Chapter 12) uses transmission of narrow noiselike pulses with spectrum extending over several gigahertz and offers promise of supporting very large numbers of simultaneous users [Sch00]. The S-T coding concept was devised to improve performance and increase spectrum utilization efficiency on bandlimited wireless channels by combining channel coding, modulation, transmitter diversity, and optional receiver antenna diversity.

Location Awareness. Another evolving technology is position location, and there is particular interest now in indoor applications (see Chapter 13). Examples of how this technology can be beneficial include location of patients, medical professionals, and instrumentation in a hospital; location tracking of merchandise in a large warehouse; and tracking of systems and components in a large factory. Other potential applications include personnel location in military, firefighting, and disaster-recovery situations. It is expected that this technology will become an integral part of future wireless networks. In the United States, the FCC has already mandated the integration of position location systems with cellular networks (e.g., for E-911 services), although the extent and method of integration is not yet clear.

Infrastructure-Based and Ad Hoc Access. It is useful to distinguish between two aspects of evolving broadband access technology: infrastructure-based access technology and ad hoc access technology, distinctions based on network topology. In infrastructure-based broadband access, the network includes a fixed (wired) infrastructure that supports communication between mobile terminals and between mobile and fixed terminals. A typical example is a WLAN employing one or multiple *access points* (APs), with APs connected by a wired (typically cabled) backbone. Two mobile stations in the same AP coverage area will communicate through that AP, and wider-area connectivity is supported by AP-to-AP communication over the wired backbone. A common example of infrastructure-based broadband access is a WLAN based on the popular IEEE 802.11b standard, operating in the 2.4- to 2.497-GHz ISM band, providing broadband access to the Internet at data rates of 1, 2, 5.5, and 11 Mb/s.

Since adoption of the 802.11b specification in July 1999, 802.11a and 802.11g have been developed to provide steadily increasing data-rate options. Currently, 802.11n is under development to provide the next step in available data rates. Some manufacturers are proposing the use of 40 MHz of bandwidth, up from 22 MHz in the 802.11b/g specifications, for the new specification.

In Europe, the high-performance radio LAN (HIPERLAN) standards evolved out of the earlier wireless ATM initiative of the ATM Forum [Ray92]. The HIPERLAN1 specification, completed in 1997, uses the same modulation technique, OFDM, as IEEE 802.11a, both standards operating at 5 GHz. The HIPERLAN1 standard was not widely adopted by manufacturers and is generally considered an unsuccessful standard. Subsequent ETSI efforts led to the HIPERLAN2 standard, which bears many similarities to IEEE 802.11a and provides a series of data rates up to 155 Mb/s, rates

approaching the capabilities of wired LANs. Unlike 802.11a, HIPERLAN2 includes features better suited to supporting not only data traffic but also time-critical services such as packetized voice and multimedia service. These aspects of the HIPERLAN2 specification lend themselves to integration of data, voice, and multimedia services. A key objective in the HIPERLAN2 standardization effort was to provide seamless interoperability of different wireless networks, including 3G networks. However, it appears that the HIPERLAN standards have been overtaken by the IEEE 802.11 standards.

In ad hoc networking, the network is reconfigurable and can operate without the need for a fixed infrastructure. This is sometimes referred to as *distributed-network topology*. Such networks are used primarily in military communications, but have also found application in some commercial networks for voice and data transmission. Ad hoc networks may employ either single-hop (peer-to-peer) or multihop connectivity. By way of example, the 802.11 WLAN standards support single-hop peer-to-peer ad hoc networking. When an 802.11 terminal is powered up, it first searches for a beacon signal transmitted by an access point or another terminal announcing the existence of an ad hoc network. If no beacon is detected, the terminal takes the responsibility of announcing the existence of an ad hoc network. Also, several other wireless technologies, such as the Personal Handyphone System (PHS) and the NEXTEL satellite network, utilize peer-to-peer push-to-talk communication to establish connection between pairs of voice terminals.

Important emerging areas for application of ad hoc networking technology include *wireless personal-area networks* (WPANs). At present, the wireless industry differentiates WPANs from WLANs by their smaller signal coverage area, ad-hoc-only topology, low power consumption, plug-and-play architecture, and support of both voice and data devices. The earliest WPANs were *BodyLANs*, developed by the U.S. Department of Defense to connect sensors and communications devices carried by a soldier or attached to a soldier's clothing. Commercial applications of the same technology can provide connectivity among laptops, notepads, and cellular phones carried by the business traveler. Motivated by the BodyLAN project, the IEEE in 1997 formed the WPAN study group as part of the 802.11 standardization activity. In 1998 the WPAN group was expanded by the inclusion of two related initiatives, HomeRF and Bluetooth. Also in 1998, a special Bluetooth group was formed within the WPAN group [Sie00]. In March 1999 the 802.15 group was formed as a separate group within the IEEE 802 structure to handle WPAN standardization. Subsequently, Bluetooth was selected as the base specification for IEEE 802.15. In Section 2.4 we provide further details on the role of WPAN, HomeRF, and Bluetooth in the evolution of the WLAN industry.

2.4 TRENDS IN WIRELESS TECHNOLOGIES

During the past two decades, as the vision of the WLAN industry evolved, WLANs were implemented based on a variety of innovative technologies, and at times industry expectations were high for development of a sizable market. Today, the major differentiator of WLANs from wide-area cellular services is the method of delivery of data to users, data-rate limitations, and frequency band regulation. Cellular data services are delivered by operating companies, whereas WLANs are owned by enterprises that utilize them in conducting their own businesses. At a time when 3G cellular industries are striving to provide 2-Mb/s packet data services, the WLAN industry is implementing standards that provide data rates up to 54 Mb/s, and new WLAN standards under

development will begin to compete with wired LAN data rates. Another differentiation with other radio networks is that today almost all WLANs operate in the unlicensed bands, where frequency regulations are lenient and there is no fee or waiting time to obtain access to a band. Next, we summarize briefly some of the important ongoing trends in wireless networking technologies.

2.4.1 Reemergence of the WLAN Industry

In the closing years of the past century a resurgence of interest has reenergized the WLAN industry. This industry, which had an almost exclusively North American market with an income equal to only a fraction of the cellular industry, suddenly began to attract widespread attention in Japan and the EC as well as renewed interest in the United States. In Japan, the restricted size of office spaces led to increased use of laptop computers as replacements for desktop PCs. Logically, WLAN technology provided the natural networking solution for laptop users. In the EC, the highly successful cellular industry started considering WLANs as a part of their next generation of high-speed packet data services. The motivation is twofold: (1) WLANs provide a practical answer to the demand for high-speed data transfer, and (2) they operate in unlicensed bands unburdened by the steadily increasing cost of acquiring licensed spectrum.

In 1999, the Wireless Fidelity (Wi-Fi) Alliance was formed as a nonprofit international association with the role of certifying the interoperability of WLAN products based on IEEE 802.11 specifications. At this writing, the Wi-Fi Alliance comprises over 200 member companies worldwide, and more than 1500 products have received Wi-Fi certification. An especially significant recent development in this arena is the emergence of *public wireless LANs* (PWLANS) serving *hot spots* in many areas of the world. The Wi-Fi certification initiative has been instrumental in stimulating growth in the use of wireless networking with laptop computers. Today, essentially all laptop computers are being manufactured with built-in 802.11-compliant wireless interfaces, and the use of public hot spots for mobile computing is growing rapidly. This segment of the wireless industry is evolving rapidly as a variety of business models for provision of public wireless connectivity are being tried.

In North America the successful growth of residential broadband Internet access has opened a new window for a sizable market in *home networking*. These trends have been catalyzed further by the emergence of new low-power personal-area ad hoc wireless networking technologies such as Bluetooth and ultrawideband (UWB) for local distribution, LMDS for home access, and indoor positioning for a variety of applications. Availability of low-power, low-cost wireless chip sets started a new revolution in consumer product development, raising hopes of sales exceeding hundreds of millions of these chip sets per year. All together, these hopes initiated a boom in chip manufacturing for WLAN and WPAN applications that continues. As far as technical directions in this industry are concerned, they continue to be toward providing higher data rates, comprehensive coverage, reduced interference, and lower cost. Further discussion of these trends can be found in [Pah02a].

2.4.2 Wireless Home Networking

Figure 2.3 illustrates the typical networking connections found in many residences. The residence is connected to the PSTN for telephone services, the Internet for Web access,

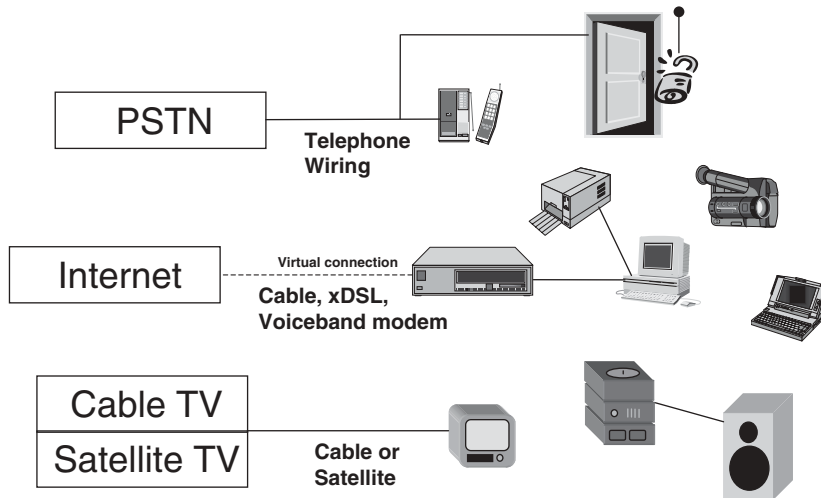


FIGURE 2.3 Today's fragmented networks.

and a cable network for TV services. Within the home, computers and printers are connected to the Internet through voiceband modems, xDSL services, or cable modems. The telephone services and security systems are connected through PSTN wiring. The TV is connected to multichannel services through hybrid fibercoax (HFC) cables or satellite dishes. The audio and video entertainment equipment, such as videocameras and stereo systems, and other computing systems, such as laptops, are either isolated or have proprietary wired connections. This fragmented networking environment has prompted a number of recent initiatives to create a unified home network. The home networking industry started within the last few years with the design of *home* or *residential gateways* for connecting the increasing number of information appliances through a single Internet connection to the home.

Many observers project rapid growth for the home networking market. The number of home networks in United States is expected to nearly double each year for the foreseeable future. As shown in Figure 2.4, this industry has two distinct segments, home access and home distribution. *Home access technology* employs different wireless and wired alternatives to secure broadband Internet access to the home gateway, that access to be extended to the user's information appliances. *Home distribution technology* or the *home area network* (HAN) interconnects all home appliances and connects them to the Internet through the home gateway.

2.4.3 Home Access Networks

Early home access technology was based on voiceband modems. Today, broadband home access (with data rates on the order of 10 Mb/s) is provided through cable modems and xDSL services over telephone lines. Cable modems operate on the cable TV wiring. The bandpass channel allocated to the TV channels is used by the cable modem operating with QAM modulation to provide a high rate of data transmission. The cable distribution plant in residential neighborhoods has a bus topology that is optimally designed for one-way TV signal distribution. The bus carries all the channels to the neighborhood, and the cable TV box selects particular TV channels for

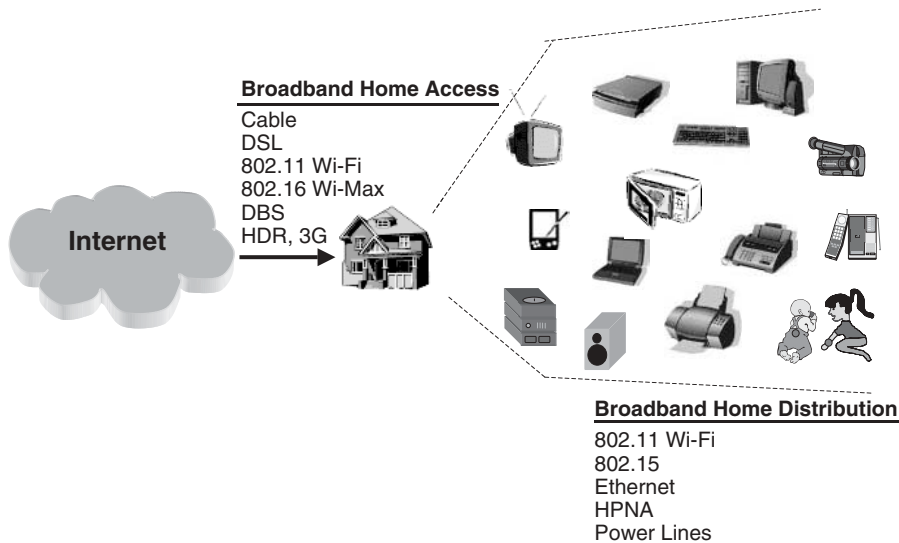


FIGURE 2.4 Two technology segments for home networking and future directions.

the residential customer. The set-top box also descrambles the signals for premium channels. To control set-top boxes, a reverse channel is available in all modern cable wiring. Broadband cable services use one of the video channels and the reverse channel to establish two-way communication and access to the Internet. The xDSL services use the frequency range 25 kHz to 1.1 MHz on the telephone line, and multisymbol QAM modulation, to support high data rates to the user. The topology of the telephone line is a star topology that connects every user directly to an end office where the xDSL data are directed to the Internet through a router.

Higher-speed wireless home access uses LMDS or even existing WLAN interLAN bridges to provide the service. The advantage of using a fixed-wireless solution is that it does not involve installing wiring in the streets. If no wiring is available in the neighborhood, a wireless solution is certainly preferred. The IEEE 802.16 group in the United States and the HIPERACCESS program in the EC are studying the next generation of networks in this category. Other wireless alternatives are direct satellite TV broadcasting and third-generation wireless networks. Direct broadcast suffers from the lack of a reverse channel and high latency, which challenges the implementation of broadband services on this medium. High-speed third-generation wireless packet data services are expected to provide rates up to 2 Mb/s, rates suitable for Internet access. The data rates on these systems are lower and they are using licensed bands that ultimately may be expensive. For further discussion of these evolving home access technologies, see [Pah02a].

2.4.4 WPANs and Ad Hoc Networking

It is useful to describe wideband wireless local access technologies as falling into two categories, WLANs and WPANs. Each of these technology categories has been the subject of considerable standardization effort. In this section we provide an overview of WPAN activities. In recent years, WPANs have been differentiated from WLANs

by their smaller area of coverage, ad-hoc-only topology, plug-and-play architecture, support of voice and data devices, and low power consumption. WPANs originated as BodyLANs that connect sensors and information devices attached to the body or to clothing. In military applications, wireless connectivity is provided to other personnel or to data collection stations. In commercial applications, WPANs can provide interconnection among personal electronic devices such as laptops, notepads, and cell phones.

The very first personal area network to be announced was the BodyLAN, which emerged from a DARPA-sponsored project in the mid-1990s. BodyLAN was a low-power, small, inexpensive wireless PAN with modest bandwidth that could connect personal devices in many collocated systems within a range of about 5 ft [Den96]. Motivated by the BodyLAN project, a WPAN group was started in June 1997 as part of the IEEE 802.11 standardization activity. In January 1998 the WPAN group published the original functionality requirement. In May 1998 the study group invited participation from WATM, Bluetooth, HomeRF, BRAN (HIPERLAN), IrDA (IR short-range access), IETF (Internet standardization), and WLANA (a marketing alliance for WLAN companies in the United States). Only the HomeRF and Bluetooth groups responded to the invitations. In March 1998, the Home RF group was formed. In May 1998 the Bluetooth development was announced and the Bluetooth special group was formed within the WPAN group [Sie00]. In March 1999, the IEEE 802.15 group was approved as a separate initiative in the 802 community to handle WPAN standardization. Currently, IEEE 802.15 WPAN has four task groups: Bluetooth (TG1), coexistence (TG2), high data rate (TG3), and low data rate (TG4). Bluetooth has been selected as the base specification for IEEE 802.15. In the remainder of this chapter we provide an overview of the WPAN, HomeRF, and Bluetooth activities.

2.4.5 IEEE 802.15 Working Group on WPAN

The IEEE 802.15 WPAN group is focused on development of short-distance wireless networks used for networking of portable and mobile computing devices such as PCs, personal digital assistants (PDAs), cell phones, printers, speakers, microphones, and other consumer electronics devices. The WPAN group intends to publish standards that allow these devices to coexist and interoperate with one another and other wireless and wired networks in an internationally acceptable frequency of operation.

The original functional requirement published in January 1998 was based on the BodyLAN project and specified devices with specified characteristics [Hei98]:

- Power management: low current consumption
- Range: 0 to 10 m
- Speed: 19.2 to 100 kb/s (actual)
- Small size (e.g., ~ 0.5 cubic inch) with no antenna
- Low cost (i.e., relative to target device)
- Allowance for overlap of multiple networks in the same area
- Networking support for a minimum of 16 devices

As we will see later in this chapter, these specifications fit the Bluetooth specification that was announced after this premier announcement. The initial activities in the WPAN group included HomeRF and Bluetooth. IEEE 802.15 WPAN has four task

groups. Task group 1 is based on Bluetooth and defines PHY- and MAC-layer specifications for wireless connectivity with fixed, portable, and moving devices within or entering a *personal operating space* (POS). A POS is the space about a person or object that typically extends up to 10 m in all directions and envelops the person whether stationary or in motion. The project addresses quality of service to support a variety of traffic classes.

Task group 2 is focused on coexistence of WPAN and 802.11 WLANs. This group has developed a coexistence model to quantify the mutual interference and a mechanism to facilitate coexistence of an IEEE 802.11 WLAN and an IEEE 802.15 WPAN device. In 2003 the task group approved IEEE 802.15.2, a recommended-practice document describing methods for enhancing the coexistence of IEEE 802.15 and 802.11 networks.

Task group 3 of IEEE P802.15 works on PHY and MAC layers for high-rate (HR) WPANs that operate at data rates higher than 20 Mb/s. In August 2003 the task group released specifications providing connectivity in the 2.4-GHz unlicensed band among fixed and portable devices. The specifications provide raw data rates ranging from 11 to 55 Mb/s, with throughput rates up to about 45 Mb/s. Devices implemented according to the 802.15.3 specifications connect in an ad hoc manner and communicate peer to peer. These specifications assure coexistence with the Bluetooth and 802.11 specifications.

Task group 4 is charged with investigating ultralow-complexity, ultralow-power consuming, ultralow-cost PHY and MAC layers for data rates up to 200 kb/s. Potential applications are sensors, interactive toys, smart badges, remote controls, and home automation. The project may also address the location-tracking capabilities required to support the use of smart tags and badges.

2.4.6 HomeRF

The mission of the HomeRF Working Group (HRFWG), formed in 1998, was to provide the foundation for a broad range of interoperable consumer devices by establishing an open industry specification for wireless digital communication between PCs and consumer electronic devices anywhere in and around the home [Hom00]. The working group eventually drafted a HomeRF specification, but industry support for the initiative declined as the popularity of the IEEE 802.11b specification grew. The rapid growth in adoption of 802.11-based products served to lower price points dramatically, and in addition, 802.11 products offered better performance than HomeRF. The HomeRF Working Group disbanded in January 2003. For a good summary of the objectives and activities of the HomeRF group, see [Pah02a].

2.4.7 Bluetooth

Bluetooth is an open specification for short-range wireless voice and data communications that was originally developed for cable replacement in personal-area networking and intended for worldwide use. In 1994 the initial study for development of this technology began at Ericsson, Sweden. In 1998, Ericsson, Nokia, IBM, Toshiba, and Intel formed a special interest group (SIG) to expand on the concept and to develop a standard under IEEE 802.15 WPAN. In 1999, the first specification, v1.0b, was released and then accepted as the IEEE 802.15 WPAN standard for 1-Mb/s networks. At the time of this writing, over 2000 companies participate as members of the Bluetooth SIG, and a number of companies all over the world are developing Bluetooth chip

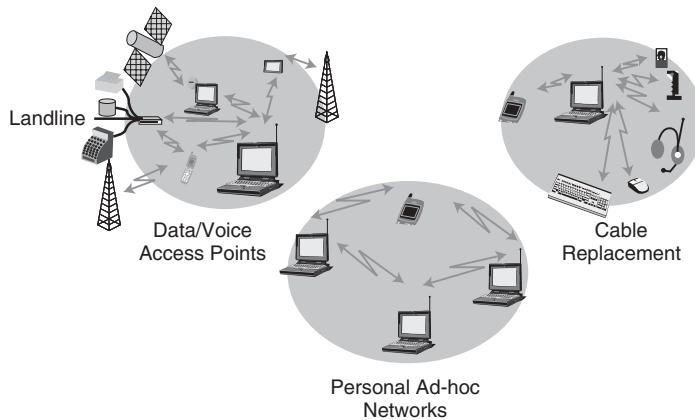


FIGURE 2.5 Three application scenarios considered by Bluetooth. (From [Blu00].)

sets. Marketing forecasts indicate penetration of Bluetooth in more than 100 million cellular phones and several millions of other consumer devices. As noted in earlier paragraphs, the IEEE 802.15 group is also studying coexistence among and interference between Bluetooth and IEEE 802.11 products operating at 2.4 GHz.

Bluetooth is the first popular technology for short-range ad hoc networking that is designed for integrated voice and data applications. Compared with WLANs, Bluetooth has a lower data rate, but it has an embedded mechanism to support voice applications. Unlike 3G cellular systems, Bluetooth is an inexpensive personal area ad hoc network operating in unlicensed bands and owned by the user.

The Bluetooth SIG considers three application-based scenarios that are shown in Fig. 2.5. The first scenario is the wire replacement for connecting a personal computer or laptop to its keyboard, mouse, microphone, and notepad. As the name of the scenario indicates, it avoids multiple short-range wiring surrounding today's personal computing devices. The second scenario is ad hoc networking of several different users within very short range of each other, such as in a conference room. As we saw earlier in the chapter, WLAN standards and products also commonly address this scenario. The third scenario is use as an access point to the wide-area voice and data services provided by cellular networks, wired connection, or satellite links. The 802.11 community also considers this overall concept of the access point. However, the Bluetooth access point is used in an integrated manner to connect to both voice and data backbone infrastructures. A more detailed discussion of Bluetooth, its protocol structure, and its relationship to IEEE 802.11 may be found in [Pah02a].

At this writing, Bluetooth-enabled devices have not yet established a position in the mainstream wireless market, but they are poised to take that next step. Bluetooth technology may well find its first significant market in the automobile industry, where hands-free regulations are helping to drive the movement toward the safety and convenience of cordless headsets.

QUESTIONS

- (a) Explain why the early mobile data services ARDIS, Mobitex, and CDPD have lost their following in the wireless services marketplace.

- (b) Describe the differences between circuit- and packet-switched services, and discuss the importance of these differences in a wireless service network.
- (c) Consider the possibility that voice over IP (VoIP) may become a very popular service in digital cellular networks. What are the implications of this relative to the traffic-handling capabilities of wireless networks?
- (d) Explain the rationale for the choice of 30-kHz RF channels in the USDC standards.
- (e) Conduct a Web search to identify all the ratified air-interface standards for IMT-2000. Choose one TDMA-based standard and one CDMA-based standard, and summarize the technical features of each. Then evaluate each with respect to ease of migration from today's standards-based cellular networks and with respect to international harmonization of standards.
- (f) Today, many large cities are installing public hot spots, which provide free WLAN connectivity to the Internet. How would you characterize the principal differences between this form of service and that of a high-rate cellular-based data service? Under the assumption that this trend will continue and expand to most U.S. cities, assess the possible impact on the market for cellular-based high-data-rate services.
- (g) Make a list of 10 potential applications for wireless-based location technology related to health care and public safety organizations. Make a second list of 10 applications that could be useful in commercial and industrial settings.
- (h) Go to the WWW and gather information on the main standardization programs being conducted under the 3GPP initiative.
- (i) Go to the WWW and gather information on countries and regions of the world that have adopted GSM or CDMA for digital cellular service.
- (j) Make a list of 10 applications for Bluetooth technology. Select five that you think might gain the widest adoption, and explain your answers.

PART II

CHARACTERISTICS OF RADIO PROPAGATION

Understanding the behavior of the wireless medium is essential for appreciating the reasoning behind specific designs for wireless communication protocols. In particular, physical-layer and medium-access protocol designs are influenced heavily by the behavior of the channel, which varies substantially in different locations. This part of the book is devoted to a detailed treatment of radio propagation in both indoor and outdoor settings, the latter including urban and suburban areas.

Chapter 3: Characterization of Radio Propagation

In this chapter we begin by using simple models to familiarize readers with the basic radio propagation parameters used in the design, analysis, and installation of wireless information networks. The most important issues for the design of a wireless communication system are the achievable signal coverage, the maximum data rate supportable on the channel, and the rate of fluctuations in the channel. For a given transmission power the achievable coverage determines the size of the cells in a cellular system and the range of operation for a system operating with a single base station. The maximum data rate is more important for data communications, where users require a high transmission speed for efficient transfer of long messages or data files. The maximum rate of fluctuations in the channel is important in the design of the adaptive parts of the receiver, such as timing and phase recovery circuits and power control algorithms.

Chapter 4: Modeling and Simulation of Narrowband Signal Characteristics

In this chapter we describe measurement and modeling techniques used to determine the narrowband characteristics of radio propagation and present results obtained in such measurements. The results of narrowband measurement and modeling allow us

to model and simulate the behavior of the received signal strength. As the distance between a transmitter and a receiver increases, the received signal power will have short- and long-distance fluctuations, referred to as multipath fading and shadow fading, respectively. We introduce a number of path-loss models for calculation of the received signal strength, and we discuss modeling and simulation of both multipath and shadow fading.

Chapter 5: Measurement of Wideband and UWB Channel Characteristics

In this chapter we describe measurement techniques used to determine the wideband and ultrawideband (UWB) characteristics of radio propagation in indoor and urban areas. We introduce the traditional time-domain measurement systems, which include direct pulse transmission techniques and measurement using direct-sequence spread-spectrum technology and a sliding correlator. Then we introduce a popular and simple frequency-domain measurement technique and discuss its application to measurement of time of arrival for indoor geolocation applications, angle of arrival for the design of multiple-input multiple-output systems, and UWB signals used for wireless personal area network applications.

Chapter 6: Modeling of Wideband Radio Channel Characteristics

In this chapter we describe models for computer simulation of wideband radio propagation characteristics. Channel models for wideband characteristics of wireless channels are divided into statistical models and building-specific deterministic models. The statistical models are divided further into time- and frequency-domain models. Most of the detailed emphasis of this chapter is on time-domain statistical models, which are the most popular among standardization organizations and are recommended by the GSM, JTC, IEEE 802.11, and IEEE 802.15 standardization committees. Frequency-domain statistical models provide models developed from empirical data to be used for generating the frequency response of the channel. Building-specific deterministic modeling techniques are divided into ray-tracing and finite-difference time-domain (FDTD) techniques. The ray-tracing technique provides an approximate ray-optics solution to Maxwell's equations. Simple ray-tracing algorithms using direct transmission and reflections were given in Chapter 3. In this chapter we provide more detailed equations and include transmission through objects, as well as diffraction and scattering. Finally, we discuss briefly the FDTD approach for direct numerical calculation of Maxwell's equations.

3

CHARACTERIZATION OF RADIO PROPAGATION

- 3.1 Introduction
- 3.2 Multipath Fading and the Distance–Power Relationship
 - 3.2.1 Narrowband Signals in Free Space
 - 3.2.2 Multipath Fading and Narrowband Signals
- 3.3 Local Movements and Doppler Shift
- 3.4 Multipath for Wideband Signals
 - 3.4.1 Multipath Delay Spread
- 3.5 Classical Uncorrelated Scattering Model
 - 3.5.1 Correlation Properties in the Delay Variable
 - 3.5.2 Multipath Delay Characteristics in the Frequency Domain
 - 3.5.3 Correlation Properties in the Time Variable
 - 3.5.4 Scattering Function
- 3.6 Indoor and Urban Radio Propagation Modeling
 - 3.6.1 Physical Operating Environments
 - 3.6.2 Traditional Methods for Modeling
 - 3.6.3 Modeling for MIMO, UWB, and Positioning
- Questions
- Problems
- Projects
 - Project 1: Two-Path Outdoor Propagation
 - Project 2: Three-Path Indoor Propagation
 - Project 3: Circular Scattering Model

3.1 INTRODUCTION

The effective design, assessment, and installation of a radio network require accurate characterization of the channel. The channel characteristics vary from one environment to another, and the particular characteristics determine the feasibility of using a proposed communication technique in a given operating environment. Having an accurate channel characterization for each frequency band, including key parameters

and a detailed mathematical model of the channel, enables the designer or user of a wireless system to predict signal coverage, achievable data rate, and the specific performance attributes of alternative signaling and reception schemes. Channel models are also used to determine the optimum location for installation of antennas and to analyze interference between different systems.

The wireless networks that we consider in this book operate at frequencies ranging from a cellular mobile telephone networks at a few hundred megahertz to WLANs and WPANs at a few gigahertz. As described in Chapter 2, cellular telephone networks use licensed bands, most second-generation systems operate at 800 to 900 MHz, PCS systems were targeted at bands around 2 GHz, and third-generation systems use both of these frequency bands. In addition, IEEE 802.11-based WLANs operate in the 2.4- and 5.2-GHz unlicensed ISM bands, and IEEE 802.15-based WPANs use the UWB unlicensed band at 3.1 to 10.6 GHz. Given the importance of cellular networks and the growing importance and convenience of unlicensed bands, we give much of our attention to this region of the spectrum.

Frequencies in the region of a few gigahertz have several attractive features for use in wireless information networks. At these frequencies a transmitter with power of less than 1 W can provide coverage for several floors within a building, and if used outdoors, can cover distances on the order of a few miles, as needed for cellular urban radio communications. Furthermore, at these frequencies the size of an efficient antenna can be on the order of an inch, and antenna separations as small as several inches can provide uncorrelated received signals suitable for achieving diversity in the received signal. At lower frequencies, bandwidth is less plentiful, longer antennas and wider antenna separations are required, and there are higher levels of human-made noise interference from ignition systems. Higher frequencies provide more ample bandwidth, but they suffer greater attenuation in transmission through walls. For frequencies in the region of a few tens of gigahertz, signal propagation is largely confined by the walls of a room, and this restricts the applications for some systems. From the standpoint of security, however, confinement in a room can be an attractive feature of these frequencies. Signal coverage can be extended throughout a building using a leaky cable antenna [Sal87a], and leaky cables are used for communication in tunnels and for paging systems in hospital buildings.

Radio propagation in both indoor and outdoor environments is complicated by the fact that the shortest direct path between transmitter and receiver is usually blocked by walls, ceilings, or other objects in an interior space, or by buildings and terrain features outdoors. Thus, the signal power is typically carried from the transmitter to the receiver by a multiplicity of paths with various strengths. The arrival times of signals on various paths are proportional to the lengths of the paths, which are in turn affected by the size and architecture of the environment and locations of objects around the transmitter and receiver. The strengths of such paths depend on the attenuation caused by passage of the signal through, or reflection of the signal by, various objects in the path. The deterministic analysis of propagation mechanisms in such an environment is limited to simpler cases. For more complex cases, statistical analysis is more useful and indeed more typically used. In statistical modeling, the statistics of channel parameters are collected from actual measurements at various locations of the transmitter and receiver.

The unpredictability of existing paths between transmitter and receiver in an indoor environment is very similar to the situation with outdoor channels, and in fact the work that has been done in characterization of mobile radio channels provides a useful

guideline for modeling indoor channels. In an indoor environment the multipath is caused by reflection from the walls, ceiling, floor, and objects within an office; in mobile radio channels, multipath is caused by the ground as well as the buildings and vehicles in the vicinity of the mobile terminal. Because the distances in an office environment are shorter, the delays between arriving paths are smaller, resulting in a smaller *multipath spread* of the received signal.

We are generally interested in different channel parameters for narrowband and wideband signaling. For narrowband communication applications, such as cordless telephone or low-speed data, we are concerned mainly with the statistics of the received power, whereas for high data rates or inherently wideband transmission, such as spread spectrum, the multipath characteristics of the channel are also important.

In this chapter we begin by using simple models to familiarize the reader with the basic radio propagation parameters used in the design, analysis, and installation of wireless information networks. The most important issues for the design of a wireless communication system are the achievable signal coverage, the maximum data rate supportable on the channel, and the rate of fluctuations in the channel. For a given transmission power the achievable coverage determines the size of the cells in a cellular system and the range of operation for a system operating with a single base station. The maximum data rate is more important for data communications, where one desires high transmission speed for efficient transfer of long messages or data files. The maximum rate of fluctuations in the channel is important in the design of the adaptive parts of the receiver, such as timing and phase recovery circuits or power control algorithms. To determine the coverage of a system, the distance-power relationship and the statistics of the power fluctuations at a given distance are needed. The data-rate limitations are determined by the multipath structure of the channel. The rapidity of variations in the channel is determined by analyzing the *Doppler spread* of the channel. These concepts are discussed in more detail in Sections 3.2 to 3.4.

In Section 3.5 we describe a more general way of modeling radio channels, one that has been widely accepted and applied in the analysis of a variety of radio systems operating in many different frequency bands. The modeling approach is based on a statistical treatment of time-varying channels, and it yields some useful insights into the key channel characteristics that affect the way signaling schemes should be designed for such channels. In Section 3.6 we lay the groundwork for a more detailed examination of channel measurement and modeling techniques, which are described in Chapters 4, 5, and 6.

3.2 MULTIPATH FADING AND THE DISTANCE-POWER RELATIONSHIP

In most radio channels the transmitted signal arrives at the receiver from various directions over a multiplicity of paths. Figure 3.1 provides several examples of multipath fading radio channels. Figure 3.1a represents a troposcatter radio communication link used in military applications for communication at long distances. The transmitted signal is directed toward the troposphere layer of the atmosphere, the incident wave is scattered, and some of the scattered signal energy reaches the receiver. Communication between the transmitter and the receiver can be modeled with several paths. Figure 3.1b represents a line-of-sight (LOS) microwave radio link, as is widely used in nationwide networks for terrestrial communications. At installation, the antennas

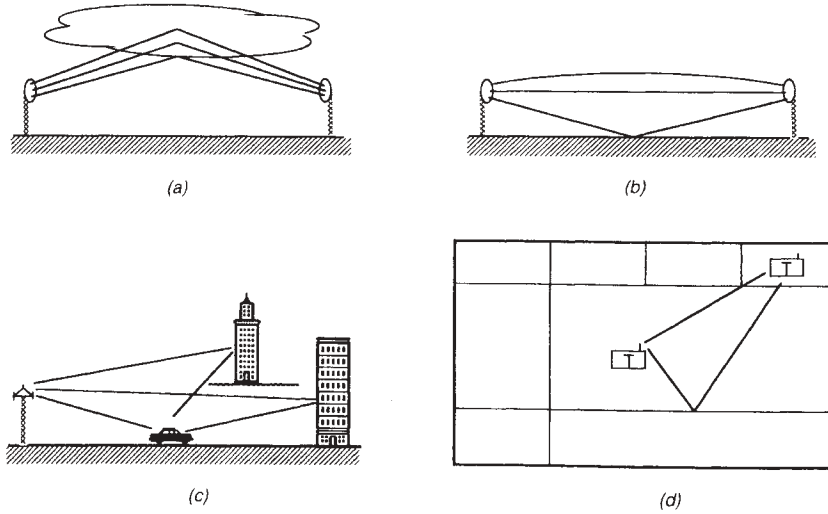


FIGURE 3.1 Examples of multipath in different radio channels. (a) Troposcatter, (b) microwave LOS, (c) mobile radio, (d) indoor radio.

are aligned to provide LOS communication. However, for occasional short periods of time, atmospheric conditions can affect radio propagation in such a way that signal components reflected from the ground and the atmosphere become comparable to the LOS component, creating a multipath condition. Figure 3.1c represents a mobile radio scenario where the received signal arrives by several paths: bounced from large objects such as buildings and local paths scattered from objects close to the receiver, such as ground or trees. Figure 3.1d represents a simple multipath condition for an indoor area.

The phase and amplitude of the signal arriving on each path are related to path length and conditions; this results in considerable amplitude fluctuation of the composite received signal. An exact analysis of multipath propagation can be done by solving Maxwell's equations with boundary conditions representing the physical properties and architecture of the environment. This method is computationally burdensome; even with today's most sophisticated computers, only the simplest structures can be treated. A simpler analytical approach is to approximate the radio-wave propagation with optical-wave propagation and to determine the directions of the arriving paths through the rules of geometric optics. This method is commonly referred to as the *ray-tracing method*. The transmitting and receiving antennas are assumed to be radiating points, and each path is modeled as a ray. A ray is the path of an ideal bullet traveling in a straight line and reflecting from the objects according to the rules of geometric optics. Figure 3.2

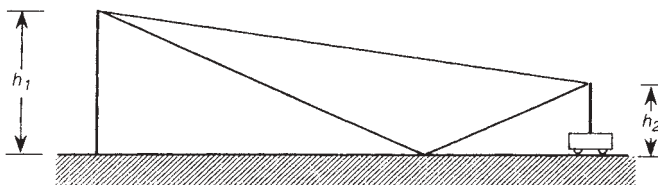


FIGURE 3.2 Setup for a mobile radio environment.

represents a mobile radio environment where the received signal arrives from two paths: (1) the direct LOS connection between the transmitter and the receiver, and (2) the path arriving after reflection from the ground. A more complete ray-tracing algorithm includes the mechanism of transmission through walls, reflection from the walls, and diffraction at the edges of buildings. Further details of the direct solution to Maxwell's equations and the ray-tracing algorithm are provided in Chapter 6. In this chapter we use a simple ray-tracing technique to familiarize the reader with the principles of radio propagation modeling for communication systems applications.

3.2.1 Narrowband Signals in Free Space

Free space provides an ideal environment for single-path communication. To analyze the multipath condition, we start with a simplified description of radio propagation in a single-path free-space channel. In free space, the relationship between transmitted power P_t and received power P_r at frequency f is given by

$$\frac{P_r}{P_t} = G_t G_r \left(\frac{\lambda}{4\pi d} \right)^2 \quad (3.2.1)$$

where G_t and G_r are the transmitter and receiver antenna gains, respectively, d is the distance between the transmitter and the receiver, $\lambda = c/f$ is the wavelength of the transmitted signal, and c is the velocity of radio-wave propagation in free space, which is equal to the speed of light. Defining $P_0 = P_t G_t G_r (\lambda/4\pi)^2$ as the normalized received power at a distance of 1 m, Eq. (3.2.1) reduces to

$$P_r = \frac{P_0}{d^2}$$

Over a single path, the received signal power decreases with the square of distance. In logarithmic form (decibel scale) we have

$$10 \log_{10} P_r = 10 \log_{10} P_0 - 20 \log_{10} d$$

which reveals the 20-dB/decade (or 6-dB/octave) loss of signal power as a function of distance in free space. The transmission delay is $\tau = d/c \simeq 3d$ ns, or a 3-ns delay per meter.

Example 3.1: Transmission Power Loss and Delay For a 1-GHz center frequency [$\lambda = c/f = (3 \times 10^8)/10^9 = 0.3$ m] and dipole antennas with $G_t = G_r = 1.6$, the received power calculated from Eq. (3.2.1) at a distance of $d = 1$ m from the transmitter is 28.4 dB below the transmitted power. The received powers at distances of 10 and 100 m are 48.4 and 68.4 dB below the transmitted power, respectively. The transmission delays associated with the 10- and 100-m distances are $\tau = d/c = 10/(3 \times 10^8) = 33$ and 333 ns, respectively.

3.2.2 Multipath Fading and Narrowband Signals

Let us assume that a single cosine with amplitude A_t and frequency f , $\text{Re}(A_t e^{j2\pi f t})$, is transmitted in free space with only the LOS path between the transmitter and the

receiver. In practice, achieving LOS transmission usually requires a very wide open area or a very narrow transmitter antenna pattern. The received signal is $\text{Re}[A_r e^{j2\pi f(t-\tau)}] = A_r e^{j\phi_r} e^{j2\pi f t}$, where A_r is the amplitude of the signal and $\phi_r = -2\pi f \tau = -2\pi f d/c = -2\pi d/\lambda$ is the phase of the signal. Because the power decreases with the square of the distance, the amplitude of the received signal decreases linearly with distance between the transmitter and the receiver. Therefore, the received amplitude of the signal at a distance d is $A_r = A_0/d$, where $A_0 = \sqrt{P_0}$ is the amplitude of the received signal 1 m from the transmitter.

In a multipath environment, the composite received signal is the sum of the signals arriving along different paths. Except for the LOS path, all paths are going through at least one order of reflection, transmission, or diffraction before arriving at the receiver. At this stage, let us consider only the reflections. Upon each reflection of a path from a surface, a certain fraction of the power is absorbed by the surface, and the remainder of the power in that path carries beyond the reflection. If the path has been reflected K_i times before arriving at the receiver, and at each reflection the reflection coefficient is a_{ij} , the overall reflection factor is

$$a_i = \prod_{j=1}^{K_i} a_{ij}$$

where a_{ij} is the reflection coefficient for the j th reflection of the i th path. Therefore, the amplitudes of the signals received from paths other than the LOS path are subject to reflection loss as well as the standard distance-attenuation factor.

If we have L paths and the distance traveled by the i th path is d_i , the amplitude and the phase of the received signal are given by

$$A_r e^{j\phi_r} = A_0 \sum_{i=1}^L \frac{a_i}{d_i} e^{j\phi_i}$$

where $\phi_i = -2\pi d_i/\lambda$. Figure 3.3 shows a phasor diagram representing the signals arriving from different paths as well as the received signal amplitude and phase. The received power is given by

$$P_r = P_0 \left| \sum_{i=1}^L \frac{a_i}{d_i} e^{j\phi_i} \right|^2 \quad (3.2.2)$$

The right-hand side of the equation shows the magnitude squared of the vector sum of all paths. If the phase of the first path is used as the reference and the vector sum is taken with the phase of all paths relative to the first path, the result remains the same. For a mobile user, the amplitude of the path changes slowly, but the phase changes rapidly at a rate of $2\pi/\lambda$ radians per meter. This means that for a mobile with a carrier frequency of 1 GHz every $\lambda = \frac{1}{3}$ m (see Example 3.1), we have a 360° change in the phase. Therefore, if we desire to visualize the received signal strength in a multipath environment, we should consider Fig. 3.3 when all the amplitudes and phases are changing randomly, amplitudes slowly and phases rapidly. The received signal amplitude, A_r , is the vector sum of all path amplitudes and phases. When all paths are in line, they add up and make a very strong amplitude, and when they oppose one another, they result in very small amplitudes. Therefore, as a mobile moves, it

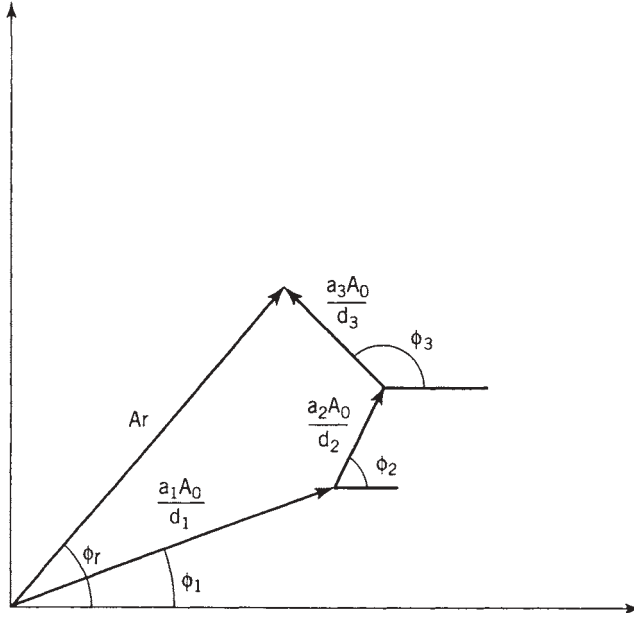


FIGURE 3.3 Phasor diagram for narrowband signaling on a multipath channel.

observes extensive fluctuations in its amplitude. Let us consider three examples that use the results of the preceding discussion to show the effects of multipath on the average and instantaneous received signal power.

Example 3.2: Distance–Power Relationship in Mobile Radio Channels Figure 3.2 depicts a mobile radio environment in which the height of the base station and mobile station antennas are h_1 and h_2 , respectively. The distance d between the transmitter and receiver is assumed to be much larger than either antenna height, and it is assumed that there are two signal paths, one the direct LOS path and the other reflected from the ground with a reflection coefficient of $a_1 = -1$, which means that the ground acts as an ideally lossless reflector.

Using Eq. (3.2.2) with the foregoing assumptions, the lengths of the two paths can be assumed to be approximately the same, d . Then the received power is given by

$$P_r \simeq \frac{P_0}{d^2} |1 - e^{j\Delta\phi}|^2$$

where $\Delta\phi = 2\pi f \Delta d/c = (2\pi/\lambda) \Delta d$ is the phase difference between the two paths, with Δd being the difference between the two path lengths. The lengths of the two paths are given by

$$d_1 = \sqrt{(h_1 + h_2)^2 + d^2} \simeq d + \frac{(h_1 + h_2)^2}{2d}$$

and

$$d_2 = \sqrt{(h_1 - h_2)^2 + d^2} \simeq d + \frac{(h_1 - h_2)^2}{2d}$$

Therefore,

$$\Delta d = d_1 - d_2 \approx \frac{2h_1h_2}{d}$$

and

$$\Delta\phi \approx \frac{2\pi}{\lambda} \frac{2h_1h_2}{d}$$

For small values of $\Delta\phi$, we have

$$|1 - e^{j\Delta\phi}| \simeq |1 - (1 + j\Delta\phi)| \simeq |\Delta\phi|$$

Then the received power is given by

$$P_r = \frac{P_0}{d^2} |\Delta\phi|^2 = \frac{P_0}{d^2} \left(\frac{2\pi}{\lambda} \right)^2 \frac{4h_1^2h_2^2}{d^2} = P_t G_t G_r \frac{h_1^2h_2^2}{d^4}$$

Note that the gradient of the distance–power relationship is increased to 4. Thus, the power will decrease 40 dB/decade of distance, in contrast with the 20 dB/decade found for LOS transmission in free space.

The first conclusion to be drawn from this example is that the multipath changes the distance–power relationship. The second conclusion is that for mobile radio communications, when the cause of multipath is reflection from the ground, 40 dB/decade is a reasonable model for the path-loss characteristic.

Example 3.3: Fading Caused by Multipath In this example we consider the hypothetical indoor environment shown in Fig. 3.4. We assume that we have a very large space (such as an automobile assembly line area) and that two mobile units [e.g., mobile robots or automatic guided vehicles (AGVs)] are communicating over a wireless link. Furthermore, we assume that the antennas are polarized horizontally so that the electromagnetic fields radiated toward the ceiling and the floor are the same as those along the LOS path, and the power of signals reflected from the walls is negligible with respect to the signal arriving along the LOS path or reflected from the ceiling or floor. In this example the ceiling height is assumed to be 5 m and the antennas are 1.5 m above the floor. The received power in this case is given by

$$P_r = P_0 \left| \sum_{i=1}^e \frac{a_i e^{j\phi_i}}{d_i} \right|^2$$

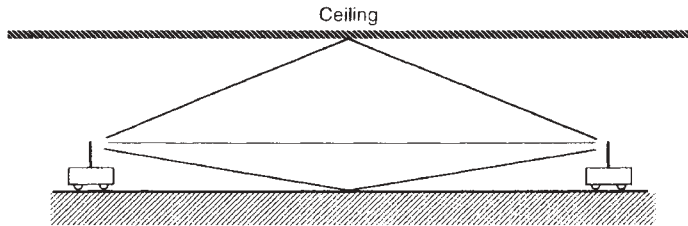


FIGURE 3.4 A hypothetical large indoor environment.

in which the reflection coefficients are assumed to be $a_1 = +1$ (the LOS path) and $a_2 = a_3 = -0.7$ and where the path distances are related by

$$d_2 = 2\sqrt{\frac{d_1^2}{4} + (1.5)^2}$$

and

$$d_3 = 2\sqrt{\frac{d_1^2}{4} + (3.5)^2}$$

Figure 3.5 shows the normalized received power versus distance calculated for distances ranging from 1 to 100 m using a scientific computer tool similar to MatLAB. The plot shows power in decibels, and distance on a linear scale.

It can be seen from Fig. 3.5 that while the average power decreases with distance, the power also fluctuates as much as 20 to 30 dB. The reason for the fluctuation is that the relative phases of the arriving paths are changing as we move from one location to another. Therefore, there is a randomness in the summation of these paths. At certain locations all the paths are essentially in phase alignment, producing relatively large received power, and in other locations the paths nearly cancel each other, producing a drastic reduction in the received power. These fluctuations constitute distance-dependent fading observed by mobile users. Example 3.2 did not exhibit these power fluctuations because the assumption there was that the distance between the terminals was much greater than the height of the antenna. As a result, the phase

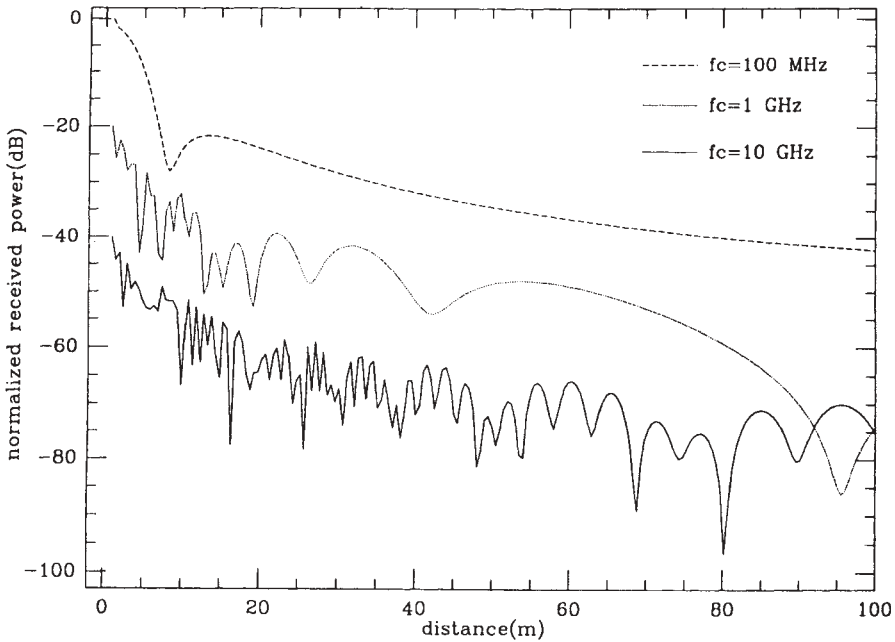


FIGURE 3.5 The normalized received power versus distance for distances between 1 and 100 m.

difference between the two paths always remained small but nonzero, preventing complete cancellation of the two paths and a resulting deep fade in the received signal.

From Example 3.3 we can conclude that the multipath causes extensive power fluctuation in the received signal, producing deep fades at particular locations.

Example 3.4: Two-Dimensional Ray Tracing Inside a Room In this example we consider a more complicated situation, in order to study the effects of multiorder reflected paths. Here the transmitter and receiver both have vertically polarized omnidirectional antennas, and the antenna pattern prevents strong components from being reflected from the ceiling or the floor. As a result, only paths reflected from the walls will contribute significantly to the received signal. To find all the paths under these circumstances, we need only trace the paths in two dimensions. Figure 3.6 depicts a two-dimensional map of the inside walls of a room with examples of LOS and first-, second-, and third-order reflected paths. The propagation paths between the transmitter and the receiver are determined by simple rules of geometric optics. The walls are assumed to be dark mirrors reflecting a portion of the signal energy and absorbing the remainder.

Figure 3.7 shows the received power in decibels versus distance, where the receiver is located at the center of a $50\text{-} \times 50\text{-m}$ room and the transmitter is moved along a straight line 2 to 20 m from the receiver. Figure 3.7a gives results obtained for the LOS path, and Fig. 3.7b–d include all the first-, second-, and third-order reflections. The second- and third-order reflections contribute very little to the received power. In this example, tracing the first-order paths is adequate to show the distance-dependent fading caused by the multipath. The gradient of the distance–power relationship remains very nearly the same as that for free space (1.90 versus 2).

Figure 3.8 shows the received power in one-fourth of a $30\text{-} \times 30\text{-m}$ room for different locations of a receiver when the transmitter is fixed at the center of the room. The results shown in this figure assume a reflection coefficient of -0.7 , with reflections of up to third order being considered.

With four walls there are four first-order, 12 second-order, and 48 third-order reflections. These components arrive with different amplitudes and phases, and invoking the

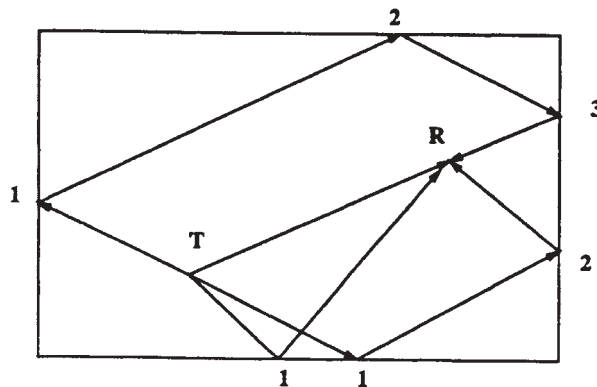


FIGURE 3.6 Reflections for ray tracing in a rectangular room.

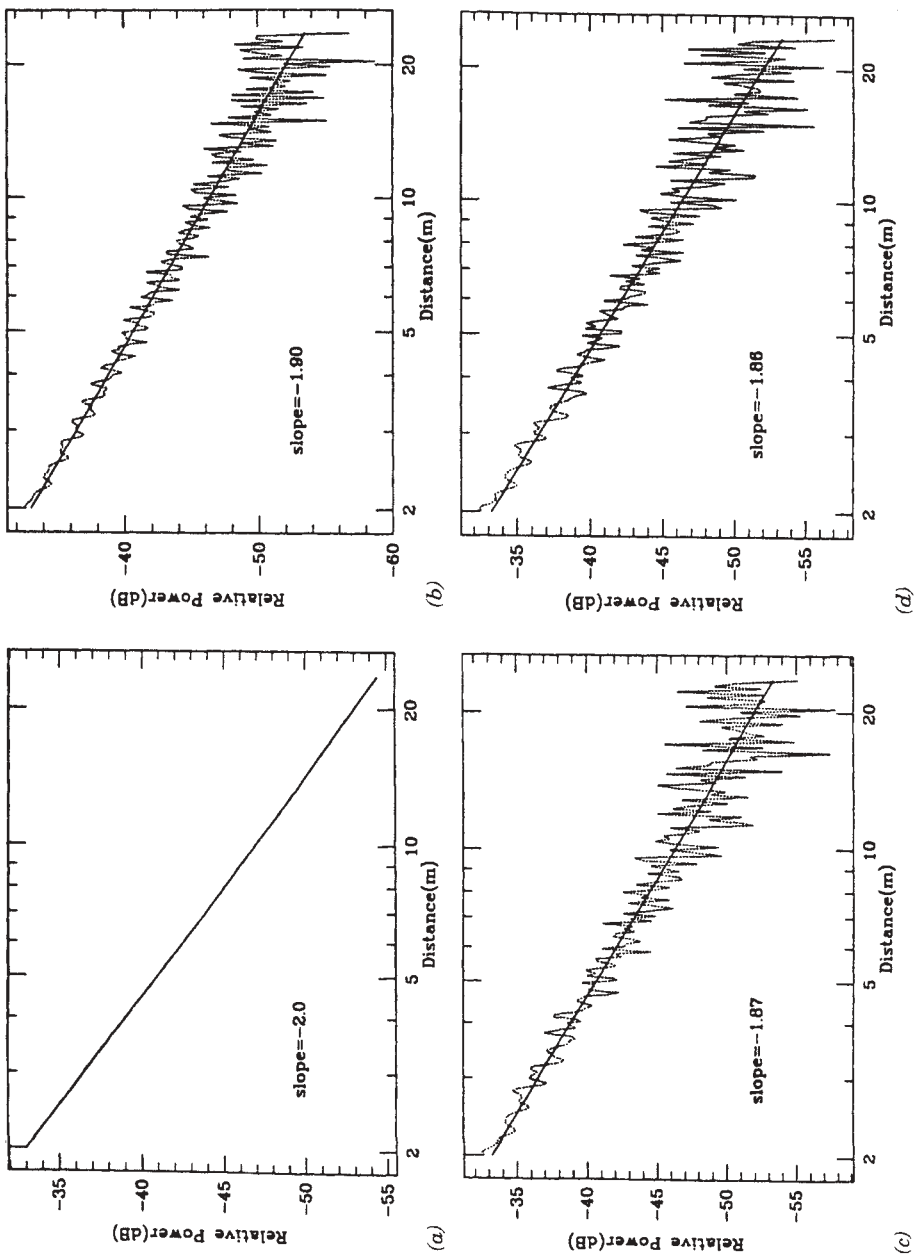


FIGURE 3.7 Received narrowband power obtained from two-dimensional ray tracing in a room. (a) LOS path, (b) first-order reflection, (c) second-order reflection, (d) third-order reflection.

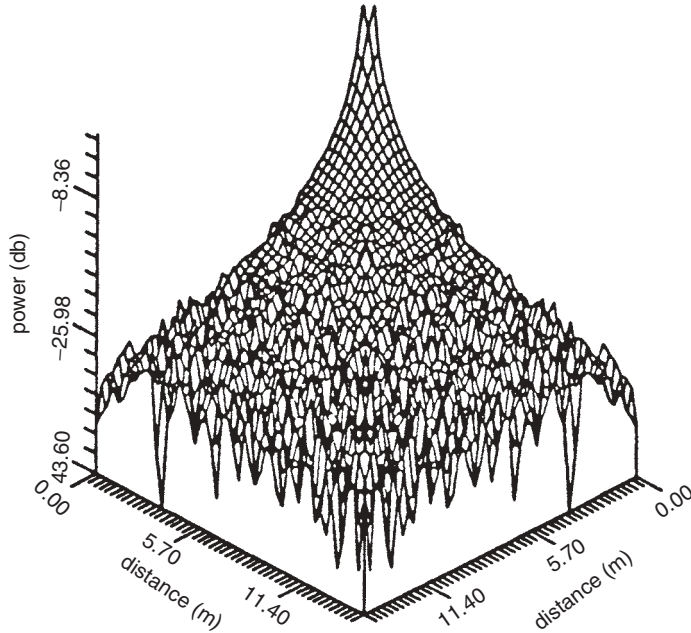


FIGURE 3.8 The received power in one-quarter of a $30\text{-} \times 30\text{-m}$ room for different locations of a receiver when the transmitter is fixed in the center of the room. Optical reflections of up to third order are considered with a reflection coefficient of 0.7.

central limit theorem, the summation of these signal components should approximate a zero-mean complex Gaussian random variable. The LOS path always exists and adds a nonzero mean to the complex Gaussian variable. The amplitude of the complex Gaussian variable in general obeys a Rician distribution, which reduces to a Rayleigh distribution when the mean is zero. Based on these considerations, it is typically assumed in the literature that the received amplitude in the absence of an LOS signal component is Rayleigh, whereas in LOS environments the received signal is assumed to be Rician.

3.3 LOCAL MOVEMENTS AND DOPPLER SHIFT

In Section 3.2 we developed a simple description of radio-wave propagation by analyzing the reception of a narrowband signal (a sine wave) transmitted over a multipath channel. In this section we examine the behavior of the signal in the frequency domain to show the effects of movements on the characteristics of the received signal.

It is well known from the fundamentals of physics that whenever a transmitter and a receiver are in relative motion, the received carrier frequency is shifted relative to the transmitted carrier frequency. This shifting of frequency is the Doppler effect of wave propagation between nonstationary points. We shall now show how the Doppler effect constitutes a source of signal fading in a multipath environment.

Figure 3.9 shows a typical example in which a fixed and a portable terminal are communicating over a radio link. The distance between the transmitter and the receiver

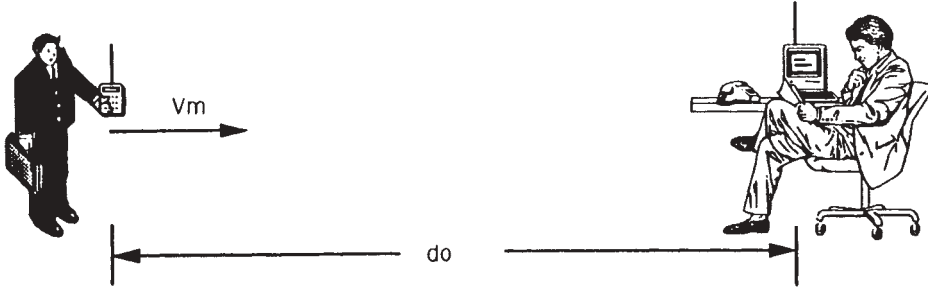


FIGURE 3.9 A typical example in which a fixed and a portable terminal are communicating over a radio link. The distance between the transmitter and the receiver is d_0 , and the portable terminal is moving with speed V_m toward the fixed terminal 2.

is d_0 and the portable terminal is moving with speed v_m toward the fixed terminal. Let us assume that the portable terminal is transmitting a tone at frequency f_c and the amplitude of the received signal is A_r . If the transmitter is stationary, the received signal is represented by $r(t) = \text{Re}[A_r e^{j2\pi f_c(t-\tau_0)}]$, where $\tau_0 = d_0/c$ is the time required for the radio wave to propagate from the transmitter to the receiver with velocity c .

As the transmitter moves toward the receiver, the propagation time will change with time as

$$\tau(t) = \frac{d(t)}{c} = \frac{d_0 - v_m t}{c} = \tau_0 - \frac{v_m}{c} t$$

The received signal is then given by

$$r(t) = A_r e^{j2\pi f_c[t-\tau(t)]} = A_r e^{j[2\pi(f_c + f_d)t - \phi]}$$

where $\phi = 2\pi f_c \tau_0$ is a constant phase shift and

$$f_d = \frac{v_m}{c} f_c$$

is a shift in the frequency observed at the receiver, commonly referred to as the *Doppler frequency shift*. The Doppler frequency shift is either positive or negative, depending on whether the transmitter is moving toward or away from the receiver. In practical situations in a wireless mobile application the direction of movement and the LOS connection between transmitter and receiver are two independent parameters. If we represent the angle between the direction of the movements of the mobile and the LOS connection between the transmitter and the receiver by, θ , the speed in which the mobile is getting close to the fixed station is $v = v_m \cos \theta$ and the associated Doppler shift would be given by $f_d = (v_m \cos \theta / c) f_c$. The maximum value of Doppler shift is then attained when the mobile is moving toward or away from the fixed terminal along the LOS path connecting the terminals. This maximum value is

$$f_M = \frac{v_m}{c} f_c \quad (3.3.1)$$

Deviations of the Doppler shift values are then bounded between $\pm f_M$, and the bandwidth of all deviations is $B_D = 2f_M$.

Example 3.5: Doppler Shift for Pedestrians and Cars If Eq. (3.3.1) is applied to a typical indoor environment, a person walking at 3 miles/hr (1.33 m/s) will cause a maximum Doppler shift of $f_M = [1.33/(3 \times 10^8)](910 \times 10^6) = 4$ Hz for a carrier frequency of 910 MHz. For a mobile user with a speed of 60 miles/hr (26.6 m/s), the associated Doppler shift for the same frequency is ± 80 Hz.

In a realistic indoor environment, the received signal arrives from several reflected paths with different path distances, and the velocity of movement in the direction of each arriving path is generally different from that of another path. Thus, a transmitted sinusoid, instead of being subjected to a simple Doppler shift, is received as a spectrum, referred to as the *Doppler spectrum*. This effect, which can be viewed as a spreading of the transmitted signal frequency, is referred to in a general way as the *Doppler spread* of the channel. Doppler spread also occurs with a fixed transmitter and receiver when a person or an object moves within the propagation path, producing time-variant multipath characteristics. In indoor and outdoor communication applications, as the terminals move about, or other objects move around the terminals, the received signal level fluctuates. The width of the Doppler spread in the frequency domain is closely related to the rate of fluctuations in the observed signal. The adaptation time of algorithms used in receivers (e.g., for automatic gain control or adaptive equalization) must be faster than the Doppler spread of the channel in order to accurately track fluctuations in the received signal. Classical modeling of the Doppler spread is explained in Section 3.5, and the results of Doppler spread measurements in the indoor and outdoor radio channels are presented in Chapter 4.

3.4 MULTIPATH FOR WIDEBAND SIGNALS

In the preceding two sections we developed a simple description of radio-wave propagation by analyzing the reception of a narrowband signal (a sine wave) transmitted over a fading multipath indoor radio channel. We also showed that the relative movement of the terminals while transmitting a sine wave causes Doppler shifts of the various multipath signal components, and this results in Doppler spread of the received signal. In this section we extend our analysis to the case of a wideband signal. If we regard a sinusoid as an ideal narrowband signal, the analogous ideal signal for the wideband case is an impulse function, which has infinite bandwidth. We analyze some simple cases of transmission of an impulse on an indoor radio channel to provide some insight into the effect of multipath on wideband communications.

Given the same multipath situation that we examined earlier, a transmitted impulse $\delta(t)$ will arrive at the receiver as the sum of several impulses with different magnitudes and phases. The composite impulse response for given locations of the transmitter and receiver is then represented by

$$h(\tau, t) = A_0 \sum_{i=1}^L \frac{a_i}{d_i} e^{j\phi_i} \delta(t - \tau_i) \quad (3.4.1)$$

where τ_i and ϕ_i are determined in the same way as they were for narrowband signaling, and $A_0 = \sqrt{P_0}$. If we define $\beta_i = A_0 a_i / d_i$, we have

$$h(\tau, t) = \sum_{i=1}^L \beta_i e^{j\phi_i} \delta(t - \tau_i) \quad (3.4.2)$$

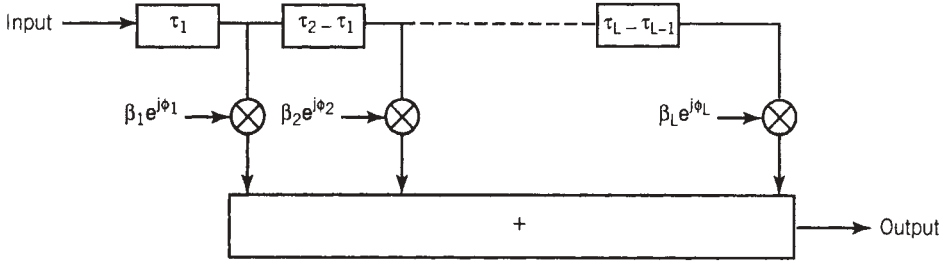


FIGURE 3.10 Block diagram for the discrete delay channel model.

where β_i and ϕ_i represent the amplitude and phase of the i th path arriving at delay τ_i . Equation (3.4.2) is widely used for statistical modeling of both indoor and outdoor radio propagation. Figure 3.10 shows a block diagram that is helpful for computer simulation of the wideband characteristics of the channel. For ideal wideband communication, the paths are isolated and independent of one another, and therefore the phase differences between arriving paths do not change the amplitude characteristics of the channel. In other words, impulses arriving at different times do not interact with each other. The received power in this case is given by

$$P_r = P_0 \sum_{i=1}^L \left| \frac{a_i}{d_i} \right|^2 = \sum_{i=1}^L |\beta_i|^2 \quad (3.4.3)$$

Here the received signal power is the sum of squares of all path amplitudes. In the case of narrowband signaling, Eq. (3.2.2), the amplitudes were added vectorially and the overall power was the square of the resulting vector magnitude. As a result, the normalized received power of a narrowband signal is less than or equal to that of a wideband signal. In simple terms, for wideband transmitted signals, the received paths are in effect isolated by the correlation properties of the signal, and the powers from different paths add algebraically. With narrowband signaling the paths are added together vectorially in accordance with their individual phases, and this interaction among the paths reduces the normalized received power relative to the wideband case.

In practice, the bandwidth of the channel is finite, and realistic impulsive signals are represented by pulses of very short but nonzero duration. Figure 3.11a shows a sample of a ray-traced impulse response in a typical square room discussed in our examples, with 2-ns pulses replacing the ideal impulses. We see that the multipath channel has spread the transmitted signal in the time domain, just as multipath with motion had spread the transmitted sinusoid in the frequency domain in the narrowband case examined earlier. Figure 3.11b and c represent the response when the transmitted impulse function is replaced by a narrow pulse of width of 5 and 10 ns, respectively. Figure 3.12 represents a sample wideband indoor radio channel measured in both the time and frequency domains. The resolution in the time domain is 5 ns, which accounts for a transmission bandwidth of 200 MHz. Note that the frequency response varies by as much as 40 dB from one frequency to another.

Example 3.6: Received Power for Wideband Versus Narrowband Signals The environment of this example is the same as that of Example 3.4 for narrowband signaling. Figure 3.13 shows received wideband power versus distance from 1 to 25 m for a

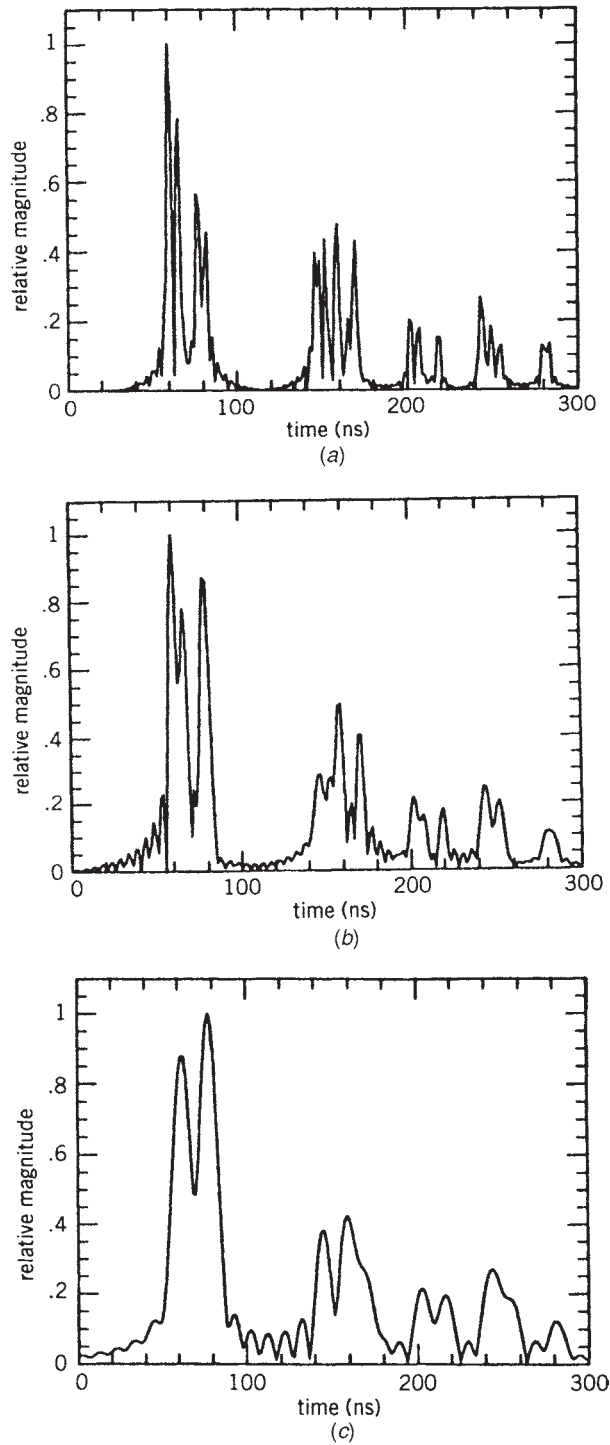


FIGURE 3.11 (a) The ideal impulse response of the channel for a given location of the transmitter and the receiver. (b, c) The response if a narrow pulse with width of 5 or 10 nsec, respectively, is used.

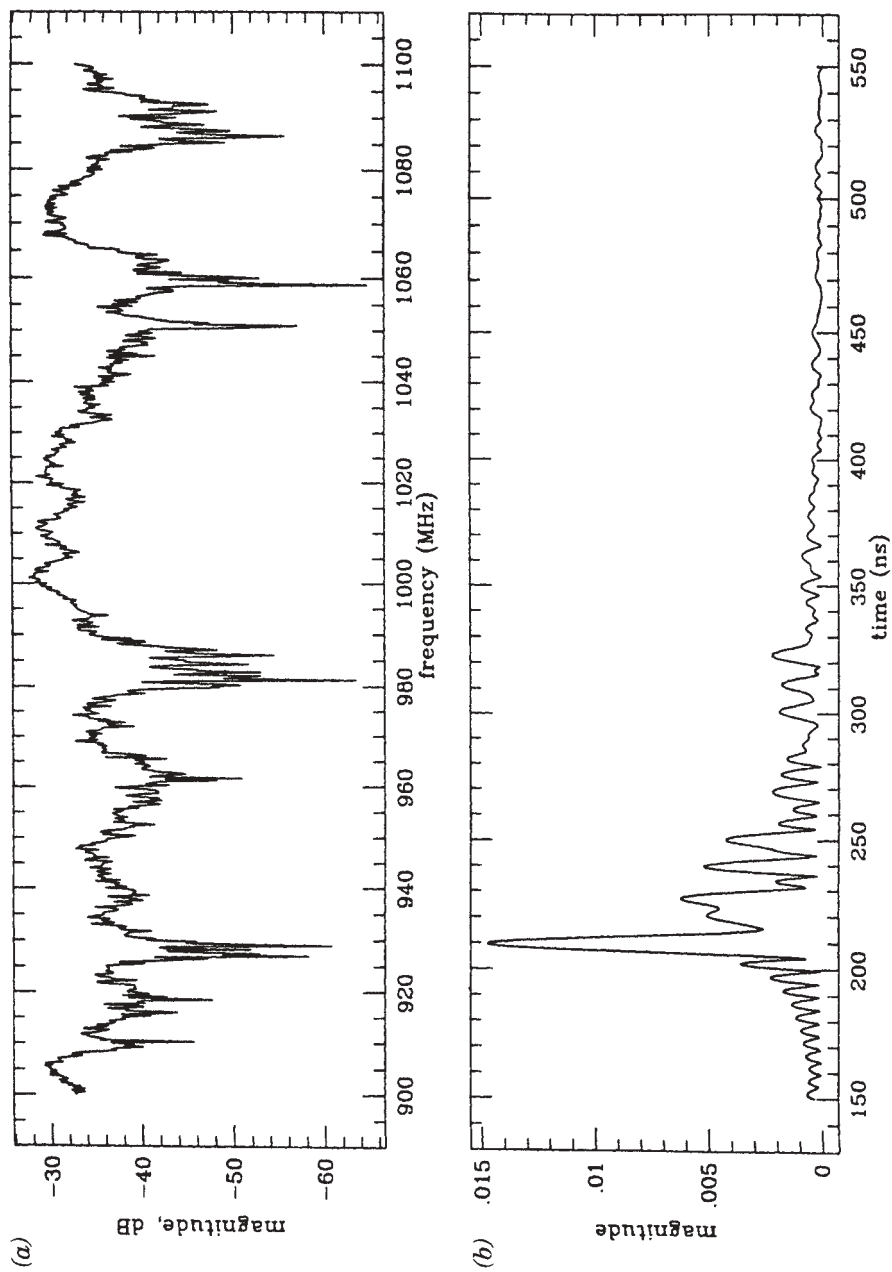


FIGURE 3.12 Sampled measured time (a) and frequency (b) response in an indoor area.

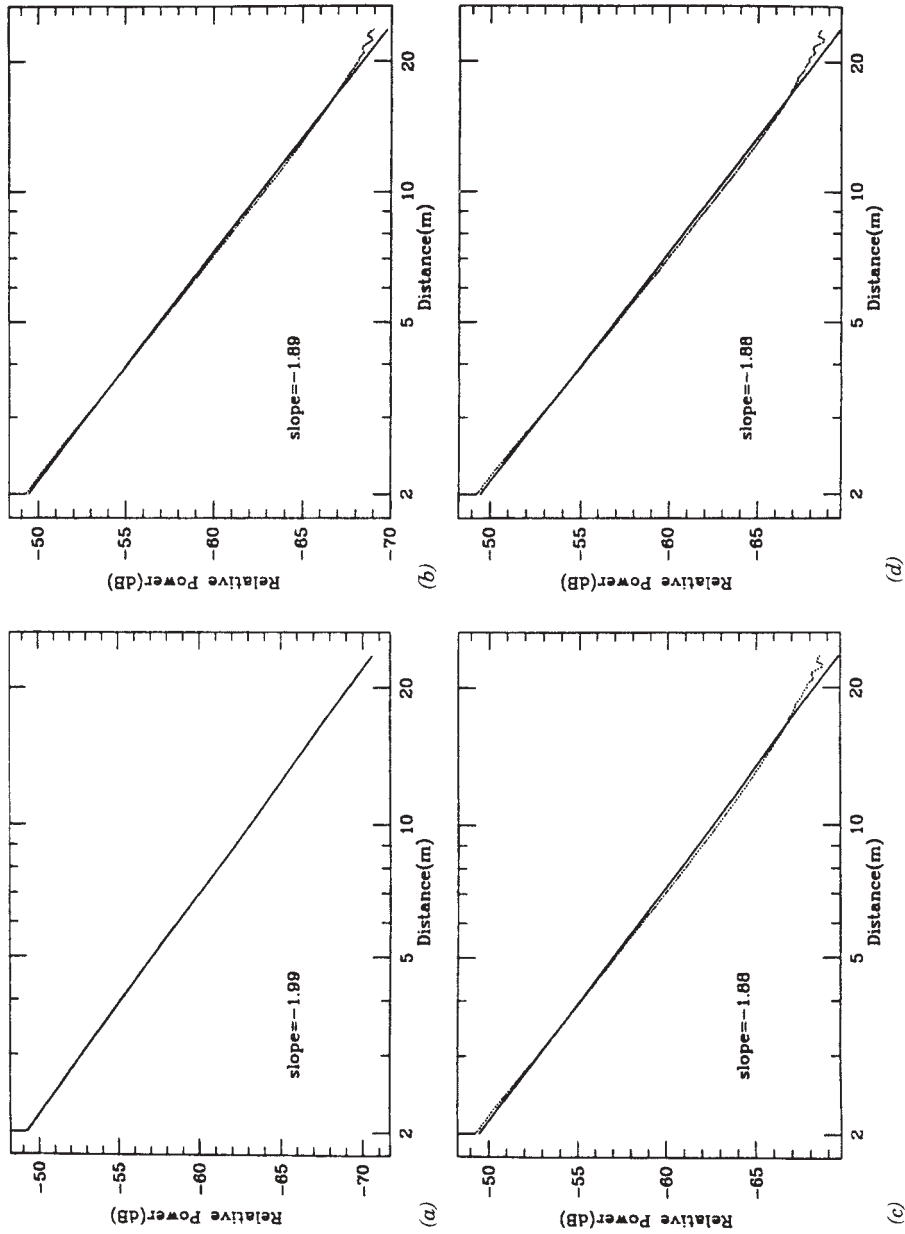


FIGURE 3.13 Received wideband power versus distance for different numbers of reflections used in the two-dimensional ray-tracing algorithm. (a) LOS path, (b) first-order reflection, (c) second-order reflection, (d) third-order reflection.

50- × 50-m room for different numbers of reflections, together with the best-fit line to the calculated signal power. This figure should be compared with Fig. 3.7 for narrowband signaling. In general, the best-fit line and the gradient of the distance–power relationship are nearly the same for wideband as for narrowband signals. If none of the paths with reflections are considered, the LOS path provides the same power for both cases as well. If we include the reflected paths, fluctuations in power for the narrowband signal are significantly more than those for the wideband signal. This characteristic is due to the fact that in wideband signaling the phase of the received signal does not play a role in calculation of the power, whereas the received power in narrowband signaling is the result of phasor summation of several vectors, which is very sensitive to the phases of the arriving paths.

3.4.1 Multipath Delay Spread

To be able to assess the performance capabilities of various wireless systems, we want to have a convenient numerical measure of the time dispersion or *multipath delay spread* of the channel. The simplest measure of multipath delay spread is the overall span of path delays (i.e., earliest arrival to latest arrival), sometimes referred to as the *excess delay spread*. However, this is not necessarily the best indicator of how any system would perform on the channel. This is because different channels with the same excess delay spread can exhibit very different profiles of signal intensity over the delay span, and different intensity-delay profiles will have greater or lesser impact on the performance of any given system. Thus, a better measure of delay spread is the *root mean square (rms) delay spread*, τ_{rms} , which is the second central moment of the channel impulse response. It is given mathematically by

$$\tau_{\text{rms}} = \sqrt{\overline{\tau^2} - (\overline{\tau})^2} \quad (3.4.4)$$

where given L propagation paths,

$$\overline{\tau^n} \equiv \frac{\sum_{i=1}^L \tau_i^n |\beta_i|^2}{\sum_{i=1}^L |\beta_i|^2} \quad n = 1, 2 \quad (3.4.5)$$

Example 3.7: Calculation of the RMS Delay Spread in Discrete Form A channel has two discrete paths. The first path is identified at an excess delay of zero ($\tau_1 = 0$) with an amplitudes of 0 dBm ($|\beta_1|^2 = 1$). The second path has an excess delay of arrival of 50 ns ($\tau_2 = 50$) and an amplitude of −10 dBm ($|\beta_2|^2 = 0.1$). The first and second moments of delay spread are calculated from Eq. (3.4.5):

$$\begin{aligned} \overline{\tau} &= \frac{0 \times 1 + 50 \times 0.1}{1 + 0.1} = 4.55 \text{ ns} \\ \overline{\tau^2} &= \frac{0 \times 1 + 2500 \times 0.1}{1 + 0.1} = 227.27 \text{ ns} \end{aligned}$$

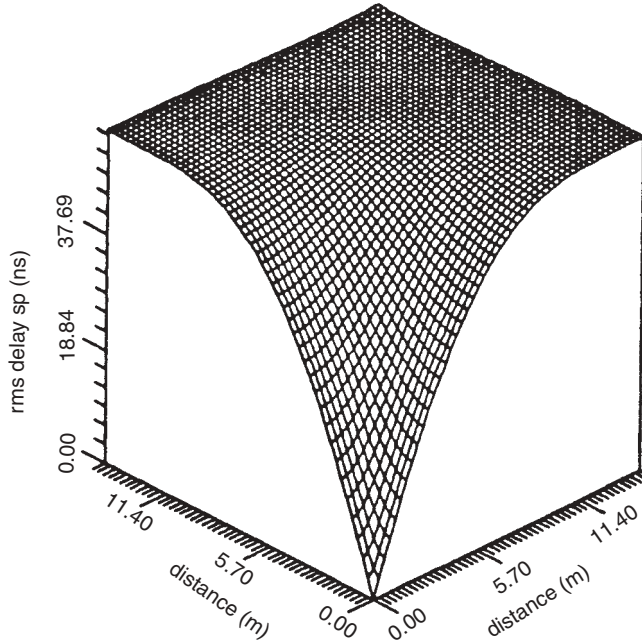


FIGURE 3.14 The rms delay spread in a 30- × 30-m room.

Then the rms delay spread is calculated from Eq. (3.4.4):

$$\tau_{\text{rms}} = \sqrt{\overline{\tau^2} - (\overline{\tau})^2} = \sqrt{227.27 - (4.55)^2} = 14.37 \text{ ns}$$

Example 3.8: Distribution of RMS Delay Spread Inside a Room Figure 3.14 shows the rms multipath spread as a function of location in one-fourth of a 30- × 30-m room. Similar to the last part of Example 3.4 shown in Fig. 3.8, the simple ray-tracing model illustrated in Fig. 3.6 is used to generate a set of discrete channel impulse responses on a grid covering the area. The impulse responses in different locations are then used to determine the magnitudes and delays of each path in a profile. These amplitudes and delays are used in Eqs. (3.4.4) and (2.4.5) to calculate the rms delay spreads in different locations on the grid. One can see from the figure that a maximum rms multipath delay spread is at the corner of the room, where the distance between the transmitter, located in the center, and the receiver is maximum. In this location, in addition to the direct LOS path between the transmitter and the receiver, we have a number of reflected paths with comparable amplitudes. The minimum rms delay spread is in the center of the room, where the direct path is much stronger than reflected paths.

3.5 CLASSICAL UNCORRELATED SCATTERING MODEL

In the preceding discussion we showed how changes in the relative phases among multiple reflected signal paths cause fluctuations in the power of the received composite signal. On indoor and urban radio channels these changes are caused either by

movement of the transmitter or receiver or by movement of people or vehicles near the transmitter or receiver. Without such movements, given fixed locations of the transmitter and receiver, the channel impulse response remains constant. As the location of the transmitter or receiver is changed or some object moves close to the transmitter or receiver, the impulse response of the channel will change. The rate of change depends on the speed of the movements. The classical model, which we describe next, provides us with a certain mathematical structure that relates to one another the key parameters that we described earlier in this chapter.

The classical method of channel modeling was developed to describe signal transmission over a variety of radio channels having randomly time-varying impulse responses. In communication over such channels, even when the transmitter and receiver are stationary, signals are subject to time dispersion and random fluctuations, caused by the constantly changing characteristics of the transmission media. Common examples are long-distance ionospheric communications in the 3- to 30-MHz high-frequency (HF) band and beyond-the-horizon tropospheric scatter communications in the 300- to 3000-MHz ultrahigh-frequency (UHF) and 3000- to 30,000-MHz superhigh-frequency (SHF) bands.

In the case of HF communications, long-distance propagation is achieved by refraction of the transmitted signal at various layers of the ionosphere. The heights, thicknesses, and ion densities of the ionospheric layers, together with the constant random motion of the ions within each layer, cause time dispersion and random amplitude and phase fluctuations in the signal as it is bent back to Earth. In the case of the tropospheric scatter (troposcatter) channel, it is more accurate to describe the received signal as consisting of a continuum of multipath components. These components are created by the physical characteristics of the troposphere, such as meteorological effects and the constantly changing interaction among multipath components, producing random fading in the received signal.

In other frequency bands, the details of the propagation mechanisms might be different, but in each case the overall effect is some combination of time dispersion and apparently random amplitude and phase fluctuations in the received signal, a set of characteristics commonly termed *multipath fading*. To assess the effectiveness of some signal design and the corresponding performance of a receiving system operating on a given multipath fading channel, it is important to be able to characterize the behavior of the channel mathematically. Because, to the observer, variations in the received signal are not predictable, but apparently random, the variations are best described in statistical terms. In particular, we want to characterize a multipath fading channel in terms of *correlation functions* and *power spectral density functions*.

3.5.1 Correlation Properties in the Delay Variable

We begin by assuming that the effects of the transmission medium are sufficiently random, and the number of multipath signal components sufficiently large, that we can invoke the central limit theorem. We can then assume that the overall impulse response of the channel is represented accurately by a complex Gaussian process $h(\tau, t)$, where inclusion of the variable t in the argument indicates that in general the channel impulse response is time varying. The channel impulse response for indoor and outdoor applications defined in Eq. (3.4.2) was a discrete function of the delay variable τ , whereas here the impulse response is a continuous function of τ . For a transmitted waveform

with complex envelope $p(t)$, the complex envelope of the received signal in the case of the continuous delay function is given by

$$r(t) = \int_{-\infty}^{\infty} h(\tau, t) p(t - \tau) d\tau$$

If we were to use the discrete channel model of Eq. (3.4.2), the received signal would be

$$r(t) = \sum_{i=1}^L \beta_i e^{j\phi_i} p(t - \tau_i) \quad (3.5.1)$$

The block diagram of Fig. 3.10, adapted for the continuous delay channel, is shown in Fig. 3.15. The transmitted signal is passed through a tapped delay line with delay values of $d\tau$ and with tap gains of $h(\tau, t)d\tau$.

As a way of modeling such a channel, Bello [Bel63a] suggested the assumption of *wide-sense stationary uncorrelated scattering* (WSSUS). This assumption leads to several interesting and useful conclusions. The physical meaning of the assumption, which is valid for most radio transmission channels, is that the signal variations on paths arriving at different delays are uncorrelated and the correlation properties of the channel are stationary; that is, they do not change with time. In mathematical terms, the assumption results in the following simplification. Autocorrelation of the observed impulse response at two different delays and two different times is given by

$$R_{hh}(\tau_1, \tau_2; t_1, t_2) = E\{h^*(\tau_1; t_1)h(\tau_2; t_2)\} = R_{hh}(\tau_1; \Delta t)\delta(\tau_1 - \tau_2)$$

Given the assumption of *uncorrelated scattering*, the only nonzero value of the correlation is observed when the delays are the same; given stationarity, the correlation values depend only on the difference in time of occurrence of the two impulse responses, not the time of occurrence of each event. For $\Delta t = 0$, this function is represented by $Q(\tau)$ and is referred to as the *delay power spectrum* of the channel:

$$Q(\tau) = R_{hh}(\tau; 0) \quad (3.5.2)$$

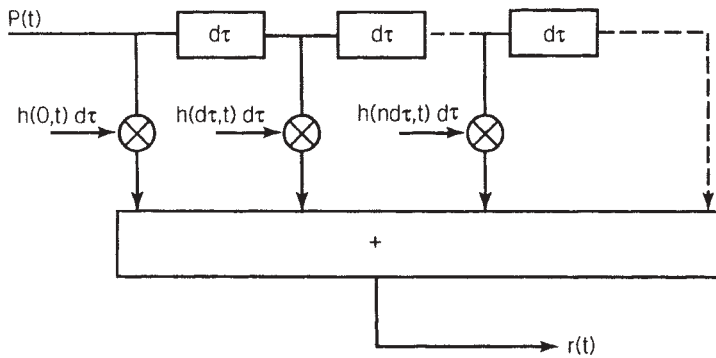


FIGURE 3.15 Block diagram for the continuous delay channel impulse response.

The delay power spectrum represents the received power as a function of time delay, given that an impulse function is transmitted. It represents the received power at different delays averaged over time. The overall range of values of τ for which $R_{hh}(\tau; 0)$ has significant nonzero value is referred to as the *excess delay spread* or simply the *delay spread* of the channel. The second central moment of this function is referred to as the *rms delay spread* and is defined as

$$\tau_{\text{rms}}^2 = \frac{\int_{-\infty}^{\infty} (\tau - \bar{\tau})^2 Q(\tau) d\tau}{\int_{-\infty}^{\infty} Q(\tau) d\tau} \quad (3.5.3)$$

where

$$\bar{\tau} = \frac{\int_{-\infty}^{\infty} \tau Q(\tau) d\tau}{\int_{-\infty}^{\infty} Q(\tau) d\tau}$$

The rms delay spread represents the effective value of the time dispersion of a transmitted signal, as caused by the multipath in the channel. For reliable digital communication over the channel, the time duration of each transmitted symbol should be much longer than this value in order to minimize the distortion of the symbol shape observed at the receiver. Because the duration of a transmitted symbol is inversely proportional to the data rate, the inverse of the rms delay spread can be taken as a measure of the data-rate limitations of a fading multipath channel. If we refer to the inverse of the rms delay spread as the *coherence bandwidth* of the channel, we can state that the rate of transmitted symbols should be much smaller than the coherence bandwidth of the channel in order to minimize the distortion of the transmitted pulse shapes. (*Note:* Some authors choose to define the coherence bandwidth as the inverse of the overall delay spread of the channel [Pro89].) There are systems that operate reliably over radio channels with symbol durations near the rms delay spread (signal bandwidth close to the coherence bandwidth), but these systems require the use of *adaptive equalization* or other anti-multipath techniques, such as orthogonal frequency-division multiplexing (OFDM) or multiple-input multiple-output (MIMO) antenna system to compensate for the distortions introduced by multipath and fading. We say more about these techniques in Chapters 8, 9, and 10, when we discuss the design of communication techniques over the radio channel.

Accurate measurement of the delay power spectrum is possible only when the channel impulse response varies slowly with time. For such a slowly varying channel the correlation properties of the channel remain the same during the measurement time Δt and we have

$$R_{hh}(\tau; \Delta t) \simeq R_{hh}(\tau; 0) = Q(\tau) \quad (3.5.4)$$

Example 3.9: Delay Power Spectrum on a Troposcatter Channel Figure 3.16 shows an example of delay power spectra derived analytically for a troposcatter channel. Figure 3.17 shows the experimentally measured values of this function for a real troposcatter link.

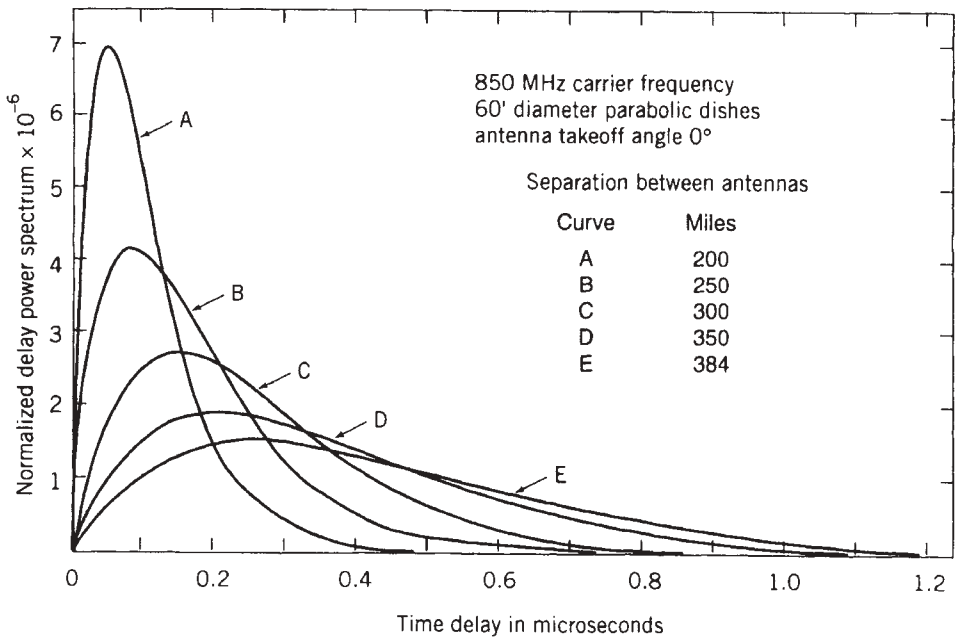


FIGURE 3.16 Analytically predicted delay power spectra for a troposcatter link. (From [Bel69] © IEEE.)

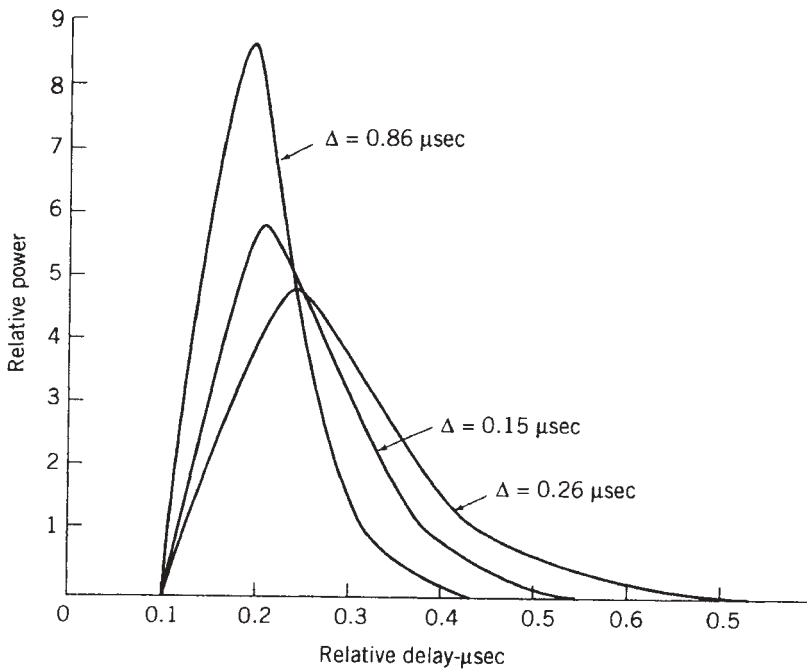


FIGURE 3.17 Experimentally measured delay power spectra for a troposcatter link. (From [She75] © IEEE.)

When modeling indoor and outdoor radio channels with the discrete channel-impulse response of Eq. (3.4.2), it is often assumed that the channel does not change with time, in which case the average of the channel impulse response is the same as the impulse response itself. The delay power spectrum in this case is simply the square of the magnitude of the channel impulse response. The rms multipath delay spread is then given by Eq. (3.4.4).

Example 3.10: Calculation of the RMS Delay Spread in Continuous Form For the two-path channel model described in Example 3.7, the delay power spectrum is given by

$$Q(\tau) = \delta(\tau) + 0.1\delta(\tau - 50)$$

Using Eq. (3.5.3) gives us

$$\begin{aligned}\bar{\tau} &= \frac{0 \times 1 + 50 \times 0.1}{1 + 0.1} = 4.55 \text{ ns} \\ \tau_{\text{rms}}^2 &= \frac{(0 - 4.55)^2 \times 1 + (50 - 4.55)^2 \times 0.1}{1 + 0.1} = 206.61 \text{ ns}^2\end{aligned}$$

Therefore, $\tau_{\text{rms}} = 14.37 \text{ ns}$, and as expected, it is the same as the results of Example 3.7. Equations (3.5.3) and (3.4.4) are two different formulations for calculation of the second central moment of the delay power spectrum.

3.5.2 Multipath Delay Characteristics in the Frequency Domain

Using the WSSUS assumption, we can derive several useful and mathematically interesting properties of the channel correlation function in the frequency domain. Given a channel with impulse response $h(\tau; t)$, the frequency response is defined as the Fourier transform of this function on the argument τ , which is written as

$$H(f; t) = \int_{-\infty}^{\infty} h(\tau; t) e^{-j\omega\tau} d\tau$$

For ideal measurement of the impulse response, an impulse function is transmitted through the channel, whereas for ideal measurement of the frequency response, sinusoids at different frequencies should be transmitted. Now, given the assumptions of the WSSUS model, the channel impulse response $h(\tau; t)$ is a wide-sense stationary zero-mean Gaussian process in the time variable t . Therefore, the frequency response $H(f; t)$, being obtained as a linear operation on $h(\tau; t)$, is also a wide-sense stationary zero-mean Gaussian process in t . Figure 3.12 shows channel time- and frequency-domain responses measured on a channel in a typical indoor area. The time-domain response shows the arrival of the multiple paths, and the frequency response exhibits amplitude variations from one frequency to another. The cause of these variations is the multipath structure of the channel, which causes constructive interference and signal enhancement at certain frequencies but causes destructive interference and deep fades at other frequencies. This channel characteristic is referred to as *frequency-selective multipath fading*.

If we were to show additional frequency responses measured at various points in time, we would see that the positions of the highs and lows in the frequency

response vary randomly from one measurement to another. To characterize these variations statistically, we can compute the correlation between values of the frequency response taken at various frequency spacings. The correlation in the frequency domain is defined as

$$\begin{aligned}
 R_{Hh}(f_1, f_2; \Delta t) &= E\{H^*(f_1; t)H(f_2; t + \Delta t)\} \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E\{h^*(\tau_1; t)h(\tau_2; t + \Delta t)\} e^{j2\pi(f_1\tau_1 - f_2\tau_2)} d\tau_1 d\tau_2 \\
 &= \int_{-\infty}^{\infty} R_{hh}(\tau_1; \Delta t) e^{j2\pi\Delta f\tau_1} d\tau_1 = R_{Hh}(\Delta f; \Delta t) \quad (3.5.5)
 \end{aligned}$$

where $\Delta f = f_1 - f_2$ and the channel is assumed to be WSSUS. The new function $R_{Hh}(\Delta f; \Delta t)$ is referred to as the *spaced-time, spaced-frequency correlation function* of the channel. As shown above, this function is the Fourier transform of the spaced-time correlation function $R_{hh}(\tau; \Delta t)$ on the delay variable. Equation (3.5.5) shows that this process is wide-sense stationary over both time and frequency variables.

For a slowly time-varying channel, the value of $R_{Hh}(\Delta f; \Delta t)$ calculated with observation times separated by Δt is the same as that found with no time separation, and thus we have

$$R_{Hh}(\Delta f; \Delta t) \simeq R_{Hh}(\Delta f; 0) = R_{Hh}(\Delta f)$$

which can be measured by transmitting two frequencies Δf apart and determining the correlation between the received signals. The inverse Fourier transform of this function is the delay power spectrum $Q(\tau)$.

3.5.3 Correlation Properties in the Time Variable

In the preceding paragraphs we discussed the correlation properties in the delay variable of the channel impulse response. We introduced the delay power spectrum and its Fourier transform, which is the spaced-frequency spaced-time autocorrelation function of the channel, and we showed how these functions are related to channel measurements in the time and frequency domains. We also discussed the special case of a slowly time-varying channel, where the time variable has no effect on the derivation of the correlation functions and power spectral density functions or, as a practical matter, on the measurements of these channel characteristics. In the following discussion we further analyze the WSSUS channel model with attention to fluctuations in time.

We first take the Fourier transform of the spaced-time spaced-frequency correlation function on the time variable, which yields

$$R_{HH}(\Delta f; \lambda) = \int_{-\infty}^{\infty} R_{Hh}(\Delta f; \Delta t) e^{-j2\pi\lambda\Delta t} d(\Delta t)$$

Now, for $\Delta f = 0$ we have $R_{Hh}(0; \Delta t)$ under the integral and the transform gives

$$D(\lambda) = R_{HH}(0; \lambda) \quad (3.5.6)$$

which is called the *Doppler power spectrum* of the channel. The Doppler power spectrum is a symmetric function and its first moment is zero. In addition, the spectrum

is always limited by $\pm f_M = (v_m/c)f_c$, in which v_m is the velocity of the mobile. Therefore, the rms Doppler spread, $B_{D,\text{rms}}$, is given by

$$B_{D,\text{rms}}^2 = \frac{\int_{-\infty}^{\infty} \lambda^2 D(\lambda) d\lambda}{\int_{-\infty}^{\infty} D(\lambda) d\lambda} = \frac{\int_{-f_M}^{f_M} \lambda^2 D(\lambda) d\lambda}{\int_{-f_M}^{f_M} D(\lambda) d\lambda} \quad (3.5.7)$$

The rms Doppler spread is a measure of variation of the channel. In Section 4.5 we use this parameter to calculate the rate of fades crossing a given threshold and the duration of a fade.

Example 3.11: RMS Value of a Uniform Doppler Spectrum The Doppler power spectrum of an indoor radio channel is sometime modeled as a uniformly distributed function given by

$$D(\lambda) = \frac{1}{2f_M} \quad |\lambda| < f_M$$

in which f_M is the maximum Doppler shift of the channel for the maximum velocity of movement of a person or a mobile vehicle inside a building. The rms Doppler spread for this channel is then calculated as

$$B_{D,\text{rms}}^2 = \frac{\int_{-f_M}^{f_M} \lambda^2 (1/2f_M) d\lambda}{\int_{-f_M}^{f_M} (1/2f_M) d\lambda} = \frac{f_M^2}{3}$$

If we use Eq. (3.3.1) for calculation of maximum Doppler shift, we have

$$B_{D,\text{rms}} = \frac{f_M}{\sqrt{3}} = \frac{v_m f_c}{\sqrt{3} c}$$

That relates the rms Doppler spread to the maximum velocity of the mobile, the carrier frequency of the system, and the velocity of light.

The Doppler power spectrum represents the strength of the Doppler shift at various frequencies caused by movements of the terminals or the objects close to them. To measure $R_{Hh}(0; \Delta t)$, we can transmit a single sinusoid ($\Delta f = 0$) and determine the autocorrelation function of the received signal. The Doppler power spectrum is the Fourier transform of this autocorrelation function. On the other hand, we know that the Fourier transform of the autocorrelation function of a time series is the magnitude squared of the Fourier transform of the original time series. Therefore, we may simply transmit a sinusoid and use Fourier analysis to generate the power spectrum of the received signal amplitude; this power spectrum is the Doppler power spectrum of the channel. The width of the Doppler power spectrum is referred to as the *Doppler spread* of the channel and provides a measure of the *fading rate* of the channel. We might regard the Doppler power spectrum as the frequency-domain dual of the delay power spectrum, which we discussed near the beginning of Section 3.5.1. In a manner

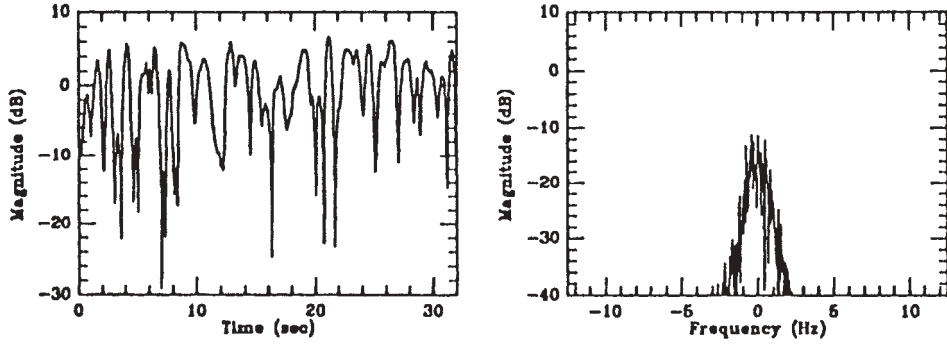


FIGURE 3.18 A sample of measured amplitude fluctuation on an indoor radio channel and its Fourier transform. The maximum Doppler spread in this sample is around 4 Hz.

similar to the treatment of delay spread, the second central moment of the Doppler spread function, the *rms Doppler spread*, is sometimes used as a measure of the fading rate in a channel. However, in the design of communication receivers, the maximum rate of variations of the channel is important, and therefore the more commonly used parameter is the overall Doppler spread rather than rms Doppler spread.

The reciprocal of the Doppler spread, called the *coherence time* of the channel, is a measure of the time interval over which a transmitted symbol will be relatively undisturbed by channel fluctuations. For slowly time-varying channels, the long coherence time is beneficial to accurate measurement of the channel characteristics, as one can use long observation times during which the measured channel response does not change significantly.

Example 3.12: RMS Doppler Spread from Measurements Figure 3.18 shows an example of amplitude fluctuations measured on an indoor radio channel, together with the calculated Fourier transform. The maximum Doppler spread in this example is about 4 Hz.

3.5.4 Scattering Function

The inverse Fourier transform of $R_{HH}(\Delta f, \lambda)$ on the Δf variable, which is the Fourier transform of $R_{hh}(\tau, \Delta t)$, taken over Δt , is called the *scattering function*:

$$S(\tau, \lambda) = R_{hh}(\tau; \lambda)$$

It represents the rate of variations of the channel at different delays. To measure the scattering function, the received signal in individual taps of a tapped delay line is analyzed in the frequency domain. In practice, it is usually assumed that the time and frequency components of the scattering function are independent. With this assumption the scattering function is decomposed into the delay and Doppler power spectra:

$$S(\tau, \lambda) = Q(\tau)D(\lambda) \quad (3.5.8)$$

Figure 3.19 shows a three-dimensional description of this function measured in an urban radio environment. Figure 3.20 summarizes all the correlation functions we have discussed here and shows the relationships among them.

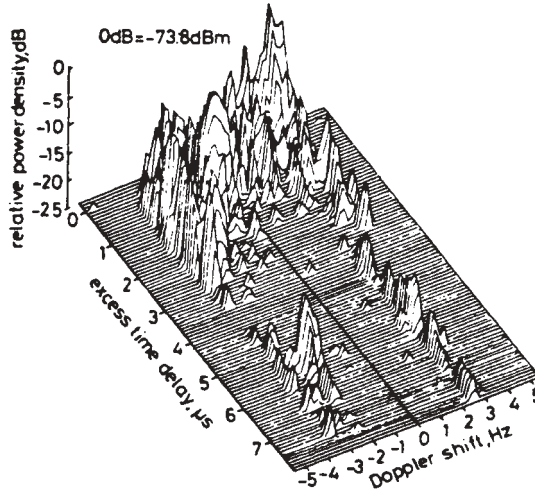


FIGURE 3.19 The measured scattering function on a troposcatter channel. (From [Par89], © Blackie, with permission.)

3.6 INDOOR AND URBAN RADIO PROPAGATION MODELING

The first step in constructing a channel model is to classify the physical characteristics of the channel. Then, the models for narrowband and wideband signaling are developed for different environments. In narrowband modeling we are interested only in the measurement and modeling of the received power. In each physical environment we relate the path loss to the distance between the transmitter and receiver. In wideband applications, we are interested in modeling the multipath structure or frequency-selective behavior of the channel in different physical environments. The modeling is based either on the statistics of the measured channel profiles or on the direct solution of radio propagation equations. The approaches to modeling indoor and outdoor radio propagation environments are quite different, as we discuss in the following subsections.

3.6.1 Physical Operating Environments

Wireless networks operate in a variety of different environments requiring attention to various aspects of radio propagation. WLANs were originally designed (around the 1990s) for use in offices and commercial buildings. Around the year 2000 they became popular in residential areas and hotspot applications. Cordless telephones and PCS systems were to operate in and around buildings in residential as well as office areas. WPANs are emerging for short-distance communications in residential, commercial, and office buildings. Cellular telephone systems were originally designed in the 1980s for use in all outdoor areas, particularly in moving vehicles. Mobile data services around the 1990s were intended to cover metropolitan areas, and most applications require in-building penetration. In the late 1990s, cellular and mobile data services were integrated in third-generation cellular networks, intending to provide comprehensive coverage all the time and everywhere. Table 3.1 shows a classification

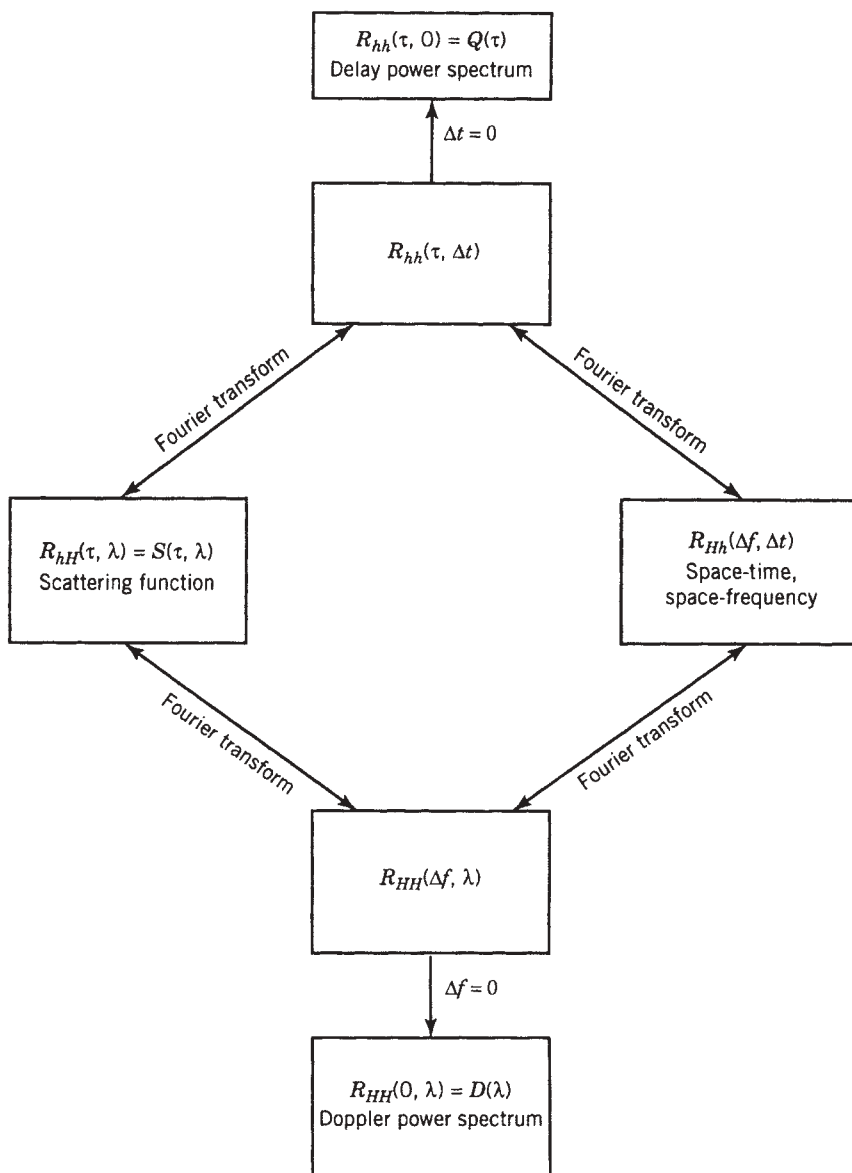


FIGURE 3.20 Summary of the correlation functions in classical modeling.

of the physical operating environments for various wireless networks. This table provides a guideline for development of channel models needed for various applications. Generally, environments are either indoor or outdoor areas. The indoor areas include residential, office, and commercial buildings. The outdoor areas are categorized as urban high-rise, urban/suburban low-rise, and residential areas. As we discussed earlier in the chapter, characteristics of the radio channel change when terminals move or when other objects are moving in the vicinity of the transmitter or receiver. The rapidity of change is proportional to the speed of the movements. Basically, there are

TABLE 3.1 Physical Operating Environments for Wireless Information Networks

General Environment	Specific Environment	Doppler Shift Range (Hz)	
		Minimum	Maximum
Indoor	Residential	2	10
	Office	2	10
	Commercial	2	10
Outdoor			
Pedestrian	Urban high-rise	2	10
	Urban/suburban low-rise	2	10
	Residential	2	10
Vehicular	Urban high-rise	5	150
	Urban/suburban low-rise	5	200
	Residential	5	100

two ranges of speed of primary interest for wireless networks: pedestrian speeds of around 3 miles/hr (1.34 m/s) and vehicular speeds of about 55 miles/hr (24.6 m/s). For networks in indoor areas, we are concerned only with pedestrian movements, whereas in outdoor environments we must deal with either pedestrian or vehicular movements.

Indoor residential areas are typified by wooden-frame single-family houses of one or two stories. The interior walls are typically covered with a thin layer of plaster inside cardboard (gypsum board). The exterior frame is filled with insulation and covered by plywood and then wooden siding or brick. With a typical residence, having many windows, there is significant radio penetration from outside the structure. Indoor office areas typically consist of large spaces partitioned into cubicles. In each cubicle there are several metallic objects, such as bookshelves and desks. The frame of the building is usually constructed with metallic studs and sometimes concrete frames, while the insulation and the exterior walls can be similar to those in residential construction. The ceilings and floors are usually heavier than in residential construction and they include significant amounts of metal and concrete, presenting a stronger barrier to radio-wave penetration from one floor to another. The indoor commercial areas include large open spaces such as manufacturing floors, shopping malls, storage areas, and transportation stations. These areas usually have high ceilings, thick layers of concrete, and heavy metallic framing.

The outdoor urban high-rise area is typified by the downtown area in any large city, often referred to as an “urban canyon.” Thick layers of concrete and heavy metallic frames in the exteriors of buildings restrict radio-wave propagation into and through the buildings. The rooftops are high, and therefore signal propagation is aided very little by diffraction. The radio waves are guided through the streets by the mechanism of reflection, with a significant power loss. The large number of moving vehicles causes continual changes in the channel characteristics. The urban/suburban low-rise areas typically include wide streets bordered by low-rise buildings. Here, propagation is aided by diffraction from the roofs of buildings. Vehicle speeds are typically much higher than in downtown high-rise areas. The outdoor residential areas are the streets of the indoor residential areas. The roads are usually two lanes wide, cars are parked alongside the streets, and the volume of vehicular traffic is usually low. In this type

of environment, the trees along the street can also influence the radio propagation characteristics.

3.6.2 Traditional Methods for Modeling

There exists an extensive body of literature on radio propagation prediction and modeling, beginning with papers published as early as the mid-1930s. Many researchers have developed a variety of experimentally or theoretically based models to predict radio propagation in various frequency bands and for various physical characteristics of the transmission path. A number of prediction models have been developed that take into account antenna height, path length, Earth curvature, terrain irregularity, foliage, urban streets and buildings, tunnels, and so on. Widely used propagation models include those of Bullington [Bul47, Bul77], Longley and Rice [Lon68], Okumura [Oku68], and Lee [Lee82]. Some of the key papers in this field are reprinted in [Bod84, IEE88a], where many other pertinent references can be found as well. In addition, a book by Jakes [Jak74] provides an extensive treatment of radio propagation for mobile radio systems, with emphasis on the frequency range from 450 MHz up to 10 or 20 MHz. The Jakes text includes the description of a simulation model for signal fading in the mobile environment, a model that is used extensively in the mobile communications industry. Narrowband models for mobile radio propagation are also treated in [Lee82, Lee89]. Broadly speaking, the modeling work we have cited in this paragraph essentially addresses the narrowband communication case. That is, the associated models provide predictions or simulations of received signal strength but do not provide detailed information regarding the time dispersion imposed on the signal by multipath effects. Work by Turin [Tur72], which we refer to next, does address time-domain characteristics.

The most commonly used statistical models for indoor radio propagation are the time-domain statistical models. These models, originally suggested by Turin [Tur72] for modeling urban radio channels, assume that the channel impulse response is in the form of Eq. (3.4.2); and based on measured data they provide statistics for the amplitudes, delays, and phases of the arriving paths. Details of the analysis for urban radio channels are available in [Suz77, Has79, Par89]. Various methods of regenerating the time-domain response of indoor radio measurements are described in [Sal87b, Gan89, Gan91a, Gan91b, Rap91b, Yeg91, Gan92, Gan93, Has93a, Has93b].

Another approach to reproducing the measured channel responses is to use the frequency response of the channel for statistical modeling. The frequency response of the channel shown in Fig. 3.12 is assumed to be an autoregressive process. The poles of the process at different locations are calculated from the sample measurement of the channel frequency response in different locations. The statistics of the locations of the poles over a set of measurements represent the model. The poles are then used in a filter driven by complex Gaussian noise. The output of the filter is used as the frequency response, and its inverse Fourier transform is used as the impulse response of the channel [Pah90b, How90b, How91, How92, Mor92]. The relationship between the arriving paths and locations of the poles is more complex than in the time-domain approach. However, evaluation of the parameters for the autoregressive model is simpler, and it requires fewer statistical parameters to represent the channel.

Cellular and PCS standards recommend a simplified version of the statistical time-domain models for design and performance evaluation of second- and third-generation

systems. In Chapters 4 and 6 we provide statistical models for amplitude and delay of arrival of paths as well as examples of more simplified versions recommended by standard organizations. In Chapter 6 we also describe principles of frequency-domain modeling of the wideband radio propagations. More recent work on time- and frequency-domain channel modeling for UWB applications is described in Chapter 12.

Statistical models cannot relate radio propagation characteristics to the exact locations of the transmitter and receiver; rather, they provide only a collection of possible channel profiles. Deterministic radio propagation modeling relates the radio propagation to the physical layout of a building by solving the radio propagation equations. The statistical models are based on actual measurements in specific buildings. The deterministic models are based on a simplified layout of a building, omitting the details of furniture and the exact properties of the structural materials. Deterministic models are much more demanding of computational power than are statistical models. A relatively simple approximate solution to indoor radio propagation is obtained by the ray-tracing algorithm [Des72, Gla89]. In this method, walls, ceilings, and floors are assumed to be dark mirrors. The paths between a transmitter and receiver are determined through transmission, reflection, and diffraction mechanisms. Computational time with the ray tracing algorithm grows exponentially with the complexity of the building. For applications in which directions of the arriving paths are important, such as analysis of systems using sectorized antennas, ray tracing provides a more reasonable model for the channel. Several groups of investigators are developing ray-tracing techniques for indoor radio propagation, as reported in [McK91, Rus91, Hol92a, Hol92b, Hol92c, Hon92, Law92, Rap92, Bro93, Yan93a, Yan93b, Ho94].

Using numerical analysis methods, one can also carry out direct solutions of Maxwell's equations. In particular, the *finite-difference time-domain* (FDTD) *method* can be used to solve the equations. The advantage of the FDTD method is that it provides a complete solution for all points in a map simultaneously. This is very important when signal coverage throughout an area is to be determined. The FDTD method solves the equations over the area with a grid on the order of magnitude of the wavelength. As a result, memory requirements increase with the increase in frequency of operation and area size. Some results for indoor radio propagation using the FDTD method are available in [Yan93b].

3.6.3 Modeling for MIMO, UWB, and Positioning

With the emergence of third-generation cellular networks and the success of the WLAN industry in the early 2000s, location-aware broadband ad hoc networks attracted considerable attention as the focus of research for next-generation wireless networks. The cellular industry considered using MIMO for next-generation location-aware cellular networks, the IEEE 802.11 standards considered MIMO for WLANs with data rates exceeding hundreds of Mb/s, and the IEEE 802.15 standards community worked on UWB networks for ad hoc WPANs. The enabling technologies for implementation of location-aware broadband networks are MIMO, UWB, and positioning technologies. Since the existing channel models developed for traditional wireless networks were not suited for analysis of the behavior of these technologies, a new wave of channel modeling research started around the year 2000.

A channel model suitable for the analysis of MIMO systems needs to provide for the angle of arrival of each path. Therefore, existing models for indoor and urban areas

that were using Eq. (3.4.2) to model the arrival time, amplitude, and phase of the paths were modified to

$$h(\tau, t, \theta) = \sum_{i=1}^L \beta_i e^{j\phi_i} \delta(t - \tau_i) \delta(\theta - \theta_i)$$

in which θ_i is the angle of arrival of the individual paths. Measurement and modeling of the angle of arrival need another layer of complexity. Measurement systems and preliminary models for the behavior of the angle of arrival of paths are available in [Spe00, Tin00, Tin01].

Channel models used for traditional broadband communication systems were based on measurements systems with a maximum bandwidth of about 100 MHz. UWB systems are expected to use gigahertz bandwidths, which are one order of magnitude wider. UWB signals attenuate differently and resolve more paths. Measurements at gigahertz bandwidths resolve more paths that were not originally observed in narrower bandwidths. Therefore, the earlier models had to be refined for UWB systems to include new path-loss models and adjust for the arrival of the paths. The early work on time-domain measurement and modeling for UWB is available in [Cas02], and more recently these models have been considered for IEEE 802.15 standards [Foe02]. For frequency-domain measurement and modeling of the UWB signals, the reader can refer to [Gha03a,b]. We discuss these models in Chapter 12.

Location-aware networks use the received signal strength (RSS), angle of arrival (AOA), and time of arrival (TOA) of the direct path (DP) to estimate the location of a mobile. Design and performance evaluation of these systems demands models for the RSS, AOA, and TOA of the received signal. The path-loss models developed for telecommunication applications can be applied to design and performance evaluation of the RSS algorithm. MIMO models providing angle of arrival are also useful for AOA positioning systems. TOA positioning systems operating in indoor and urban areas with extensive multipath arrivals suffer from extensive distance measurement error. However, traditional telecommunication systems do not pay attention to the accuracy of the TOA of the DP and always assume that it is accurate. To analyze this situation and design algorithms to remedy the large errors caused by multipath, we need channel models with particular emphasis on TOA of the DP. Pioneering work in this area is available in [Pah98, Kri99a,b, Pah02, Ala03a,b]. We address channel modeling for indoor geolocation applications in Chapter 13.

In the next three chapters we delve further into radio propagation for wireless networks. In Chapter 4 we describe results of measurement and modeling for narrow-band signals. In Chapter 5 we analyze wideband measurement systems and the results of measurements in indoor and outdoor areas. In Chapter 6 we provide the details of statistical and building-specific methods of modeling and simulating radio propagation.

QUESTIONS

- (a) Name two major classes of wireless applications in which a channel model is needed either for design or for performance evaluation of a system.
- (b) Describe the power loss in decibels per octave of increase in distance as a function of the distance–power gradient α .

- (c) What causes signal fading?
- (d) Why do signal arrivals from different paths cause a narrowband signal to fade?
- (e) Why is the multipath spread greater in outdoor areas than in indoor areas?
- (f) Explain why the Doppler spread is greater on a mobile radio channel than it is in indoor areas.
- (g) Why are the power fluctuations for wideband signals smaller than for narrowband signals?
- (h) Is the distance–power gradient for narrowband signals the same as that for wideband signals? Explain.
- (i) List the basic assumptions underlying the WSSUS channel model.
- (j) What are the principal methods used for indoor and outdoor radio propagation modeling?
- (k) What parameter is commonly used to represent the multipath delay spread on a radio channel?
- (l) What parameter is commonly used to represent the Doppler phenomenon?

PROBLEMS

1. IEEE 802.11 WLANs operate at a maximum transmission power of 100 mW (20 dBm) using multiple channels with different carrier frequencies. IEEE 802.11g uses 2.402 to 2.480-GHz bands, and IEEE 802.11a uses 5.150 to 5.825-GHz bands. Both standards use OFDM modulation with a bandwidth of 20 MHz.
 - (a) Calculate received signal strength in dBm at a 1-m distance of an IEEE 802.11g access point for the smallest and largest possible carrier frequencies in the band. Assume that transmitter and receiver antenna gains are 1 and within a 1-m distance, signal propagation follows the free-space propagation rules.
 - (b) Repeat part (a) for the IEEE 802.11a WLANs.
 - (c) Compare the received signal strengths at a 1-m distance from the IEEE 802.11g and IEEE 802.11a devices. Use the middle of the allocated band for each standard as the carrier frequency in your calculations.
 - (d) Compare the rate of the received signal fluctuations, due to the change in frequency of operation, for IEEE 802.11g and IEEE 802.11a. Use the middle of the allocated band for each standard as the carrier frequency in your calculations.
2. The IEEE 802.15 community is considering the unlicensed bands between 3.4 and 10.6 GHz for ultrawideband (UWB) WPAN devices. One of the leading proposals for this standard uses multiband OFDM technology, for which each channel occupies 512 MHz. Repeat Problem 1 including multiband OFDM in a comparison of 802.11 options.
3. A multipath channel has three paths at 0, 50, and 100 ns with the relative strengths of 0, -10 , and -15 dBm, respectively.
 - (a) What is the multipath spread of the channel?

- (b) Calculate the rms multipath spread of the channel.
 - (c) What would be the difference between multipath spreads and rms multipath spreads of this three-path channel and a two-path channel formed by the first and third paths of this profile?
 - (d) What would be the difference between multipath spreads and rms multipath spreads of this three-path channel and a two-path channel formed by the first and second paths of this profile?
4. For a sinusoidal signal of the form $x(t) = A \cos \theta(t)$, the frequency is the rate of variations or derivative of the phase:

$$f(t) = \frac{1}{2\pi} \frac{d}{dt} \theta(t) \quad (\text{P3.1})$$

- (a) If in the communication scenario depicted in Fig. 3.9, the mobile terminal transmits a sinusoid $x(t) = A \cos \omega_c t$, determine the phase of the received signal, $\theta(t)$, in terms of the distance between the transmitter and the receiver, d_0 , and the velocity of the mobile, v_m .
 - (b) Determine the frequency of the transmitted and received signals using Eq. (P3.4).
 - (c) Explain the physical meaning of the Doppler shift defined in Eq. (3.3.1) based on the results of part (b).
 - (d) Determine the phase, $\theta(t)$, and frequency, $f(t)$, of the received signal if the mobile was moving with the same velocity, v_m , but its path had an angle α with the direct path between the transmitter and receiver.
5. Starting with Eq. (3.5.3), show that if the channel impulse response is represented by Eq. (3.4.2), the rms delay spread is given by Eq. (3.4.4).
6. Consider a fading channel with scattering function

$$S(\tau, \lambda) = Q(\tau)D(\lambda)$$

where $Q(\tau)$ and $D(\lambda)$ are uniformly distributed functions within the ranges $0 < \tau < 100$ ms and $|\lambda| < 10$ Hz.

- (a) What are the multipath spread and rms multipath spread of the channel?
 - (b) What are the maximum Doppler spread and rms Doppler spread of the channel?
 - (c) What is the coherence bandwidth of the channel?
 - (d) What is the coherence time of the channel?
7. Repeat Problem 6 for

$$Q(\tau) = e^{-\tau/T} \quad \tau \geq 0$$

and

$$D(\lambda) = \frac{1}{f_m} \left[1 - \left(\frac{\lambda}{f_m} \right)^2 \right]^{-1/2} \quad |\lambda| < f_m$$

Assume that $T = 10$ ns and $f_m = 10$ Hz.

8. Repeat Problem 6 for

$$Q(\tau) = 0.7\delta(\tau) + 0.3\delta(\tau - 20 \times 10^{-9})$$

and

$$D(\lambda) = \frac{0.8}{f_m} \left[1 - \left(\frac{\lambda}{f_m} \right)^2 \right]^{-1/2} + 0.2\delta(\lambda) \quad |\lambda| < f_m$$

9. Rederive Eq. (3.5.5) and explain the details of the derivation.

PROJECTS

Project 1: Two-Path Outdoor Propagation

Part I: Calculation of the Received Signal Strength (RSS). In Example 3.2 we used algebraic manipulations to find an approximation to the narrowband RSS in a wide open area when the received signal arrives from two paths. In that calculation we assumed both paths have the same length but different phases. This is a good approximation for long distances. For the two-path model considered in that example, we can calculate the RSS for any distance using exact values of the distance and phase in Eq. (3.2.2). As the first step in this project, give an equation for exact calculation of P_r in Example 3.2 for any distance between the transmitter and the receiver. Then do the following:

- Assuming that $h_1 = 100$ m, $h_2 = 3$ m, $P_0 = 0$ dBm, and $f_c = 800$ MHz, sketch and label the exact value of P_r in decibels versus d in logarithmic form (similar to Fig. 3.7) for $10 < d < 100$ m. Use 100 points for your plot, and use MatLAB or an alternative computational tool for calculations and plots.
- Compare the results of part (a) with the approximated results of Example 3.2.
- Repeat parts (a) and (b) for $100 < d < 1000$ m.
- Repeat parts (a) and (b) for $1000 < d < 10,000$ m.

Part II: Calculation of the Delay Spread Characteristics. Multipath spread is often characterized either by rms multipath spread or the excess delay spread that is the difference between the arrival delay of the first and the last arriving paths. Give an equation for calculation of τ_{rms} , the rms value of the delay spread and $\Delta\tau$, the delay between the arrival of the two paths in Fig. 3.2 in terms of the distance between the mobile and base stations and the height of the antennas in each side.

- Sketch and label $\Delta\tau$ versus d on a logarithmic scale for $10 < d < 100$ m. Use 100 points for your plot, and use MatLAB for calculations and plots.
- Repeat part (a) for $100 < d < 1000$ m.
- Repeat parts (a) and (b) with τ_{rms} replacing $\Delta\tau$.
- If the maximum data rate of a modem, R , is related to the rms multipath spread of the channel by $R \approx 0.1/\tau_{\text{rms}}$, sketch the maximum data rate versus distance for $10 < d < 100$ m.

Project 2: Three-Path Indoor Propagation

Part I: Calculation of the RSS. In Example 3.3 we showed the results of simulation for the RSS of a three-path large and open indoor area. In the first part of this project we repeat Example 3.3 to get familiar with the details of this simple ray tracing technique. In the second part we expand our analysis to include the behavior of the multipath characterization parameters in the same area.

- Using MatLAB or an alternative computation tool, repeat Example 3.3 (the three-path indoor model) and plot the narrowband received power in decibels versus distance ($1 < d < 100$ m) on a log scale for the center frequency of 100 MHz. Assume that $P_0 = 0$ dBm.
- Determine the distance–power gradient by finding the slope of the best-fit line to the received power plot.
- Repeat parts (a) and (b) for a center frequency of 1 GHz.
- Repeat parts (a) and (b) for a center frequency of 10 GHz.
- What is the difference in the distance–power gradient obtained in parts (b), (c), and (d)? Explain.

Part II: Multipath Characteristics

- For Example 3.3, sketch the rms multipath delay spread in nanoseconds versus the distance on a log scale for $1 < d < 100$ m. Assume that $f_c = 1$ GHz. Do you see any relationship between the distance and the rms multipath delay spread? Can you generalize your conclusion?
- Repeat part (a) for a center frequency of 10 GHz. Is there any difference in the rms multipath due to this change in the center frequency?

Project 3: Circular Scattering Model

Part I: Principle of Operation. The circular scattering model for a mobile radio channel assumes that the paths from a mobile transmitter are scattered from a uniform circle around the transmitter before they arrive at the base station antenna. Figure P3.1 shows a typical situation and relevant parameters used in the model. Associated with each path in this model we have several parameters: the distance between the transmitter and the receiver d , the radius of the scattering circle $R \ll d$, the angle of arrival of a path θ , and the rate of variation of the distance (velocity) in the direction of arrival of a path v_θ .

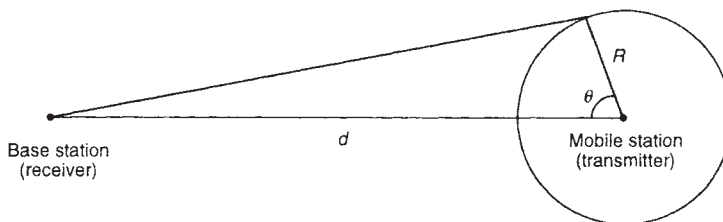


FIGURE P3.1

- (a) Show that if the transmitter moves toward the receiver with a velocity of v , the velocity in the direction of the path with angle θ is given by

$$v_\theta = v \cos \theta \quad \text{m/s}$$

- (b) Give $f(\theta)$, the Doppler shift of the path with the angle of arrival of θ , in terms of the velocity of the vehicle, the frequency of operation, and the angle of arrival. Sketch $f(\theta)$ as a function of θ for $0 < \theta < 2\pi$. What are the angles that provide the minimum and maximum Doppler shifts?
- (c) Because $d \gg R$, all the path lengths are approximately R and the path loss associated with all paths is the same. As a result, the amplitudes of all arriving paths are the same. However, because the rate of variation of the distance for different paths is not the same, the Doppler shift associated with each path is different. The difference in the Doppler shift for different paths causes a phase difference among the arriving paths, and the phasor representing the complex envelope of each received path is given by

$$x_\theta(t) = A e^{j2\pi f(\theta)t}$$

where A is the fixed path amplitude. The complex envelope of the received signal from all paths is then given by

$$r(t) = A \int_0^{2\pi} x_\theta(t) d\theta$$

Plot the envelope of the received signal (magnitude of the received phasor) in decibels as a function of time, and use 100 samples at sampling intervals of 0.1 ms. Assume that $v = 80$ km/h, $f = 800$ MHz, and $A = 1$.

- (d) Plot the probability density function of the linear magnitude of the 100 amplitude samples of the received signal. Name a distribution function that fits the observed samples and determine its mean and variance.

Part II: Simulation of the Model. Use the circular scattering model of Part I to generate 200 samples of the magnitude and phase of the channel. Using MatLAB or an alternative programmable computation tool, calculate and plot in decibels the magnitude of the Fourier transform of the samples generated at sampling intervals of 0.1 ms. Assume that $v = 80$ km/h, $f = 800$ MHz, and $A = 1$.

4

MODELING AND SIMULATION OF NARROWBAND SIGNAL CHARACTERISTICS

- 4.1 Introduction
- 4.2 Modeling Path Loss and Slow Shadow Fading
 - 4.2.1 Basic Path-Loss Model for Indoor Areas
 - 4.2.2 Wall-Partitioned Path-Loss Models for Indoor Areas
 - 4.2.3 Distance-Partitioned Models for Indoor and Microcellular Areas
 - 4.2.4 JTC Path-Loss Model for Microcells and Macrocells
 - 4.2.5 Okumura–Hata and COST-231 Models for Macrocells and Microcells
- 4.3 Doppler Spectrum of Fast Envelope Fading
 - 4.3.1 Clarke Model for a Doppler Spectrum in Urban Areas
 - 4.3.2 Measurement of a Doppler Spectrum in Indoor Areas
 - 4.3.3 Fading Rate and Fade Duration
- 4.4 Statistical Behavior of Fast Envelope Fading
 - 4.4.1 Distribution Functions for Envelope Fading
 - 4.4.2 Measurement of Envelope Fading Statistics
- 4.5 Simulation of Fast Envelope Fading
 - 4.5.1 Filtered Gaussian Noise for Simulation of a Mobile Radio Channel
 - 4.5.2 The Clarke–Jakes Model for Simulation of a Mobile Radio Channel
 - 4.5.3 Envelope-Fading Simulation for a Flat Spectrum in Indoor Areas
- Questions
- Problems
- Projects
 - Project 1: Deployment of IEEE 802.11b and g WLANs
 - Project 2: Simulation Techniques for Fast Envelope Fading
 - Project 3: Simulation of Shadow Fading and Handoff

4.1 INTRODUCTION

In Chapter 3 we showed that due to the constructive and destructive interference of multipath components received at different locations, multipath propagation causes

substantial variations in the amplitude of a received radio signal. We also showed that the Doppler shifts imparted to the various multipath signals due to movement of the terminals or movement of people or objects around the transmitter and the receiver cause spectral spreading of the received signal. Then we discussed how the multipath and Doppler effects place limitations on the rate of signaling achievable over the channel, and we showed these effects to be related to three parameters:

1. The distance–power gradient (α)
2. The root mean square (rms) delay spread (τ_{rms}) of the channel
3. The Doppler spread of the channel (f_d)

The distance–power gradient is used for the determination of power decrease as a function of distance from the transmitter. As a simple rule, 10α is the average attenuation per decade of increase in the distance. The Doppler spread is related to the aggregate of Doppler shifts of multipath components; each shift is approximated by v_m/λ , where v_m is the effective closing velocity of the path and λ is the wavelength of the carrier frequency. The rms multipath delay spread limits the symbol transmission rate R of a simple modulation technique to an approximate value $R \simeq 0.1/\tau_{\text{rms}}$. In general, measurements are performed using either narrowband or wideband techniques and equipment, and the results are used to develop narrowband or wideband models, respectively. Narrowband measurements can provide parameters α and f_d , and τ_{rms} can be determined from the results of wideband measurements.

In this chapter we describe more detailed measurement and modeling techniques used to determine the narrowband characteristics of radio propagation and present some results obtained in such measurements. Narrowband measurements are made when the transmission rate of the intended application is well below the coherence bandwidth of the channel. As an example, as we will see later, the coherence bandwidth of the indoor radio channel for distances less than 100 m between the transmitter and the receiver is around a few megahertz, which means that transmission rates on the order of several hundred kilobits per second are considered to be narrowband. For digital cordless applications the transmission rates are always below these values. As a result, cordless telephone applications provided the main motivation for pioneering narrowband measurements and modeling in indoor areas in the early 1980s [Ale82, Ale83].

In measurement and modeling for narrowband signaling applications we are mainly interested in the behavior of the received signal strength. As we discussed in Sections 3.2 and 3.3, the received power in a multipath environment always varies with small local changes, on the order of the wavelength of the carrier frequency, in the location of the transmitter and receiver or the movement of the objects around them. However, the average received power over a small area is related to the distance from the transmitter to the center of the receiving area. Therefore, as the distance between a transmitter and a receiver increases, the received signal power will have *short- and long-distance fluctuations*, referred to as *multipath fading* and *shadow fading*, respectively. Figure 4.1 illustrates variations of the received signal with respect to distance caused by multipath and shadow fading. Multipath fading is the rapid instantaneous changes in the received signal power caused by fast changes in the phase of the received signal from different paths due to small movements. Shadow fading is the long-term average changes in the received signal strength caused by changes in the relative position of large objects, such as buildings in urban areas, between the

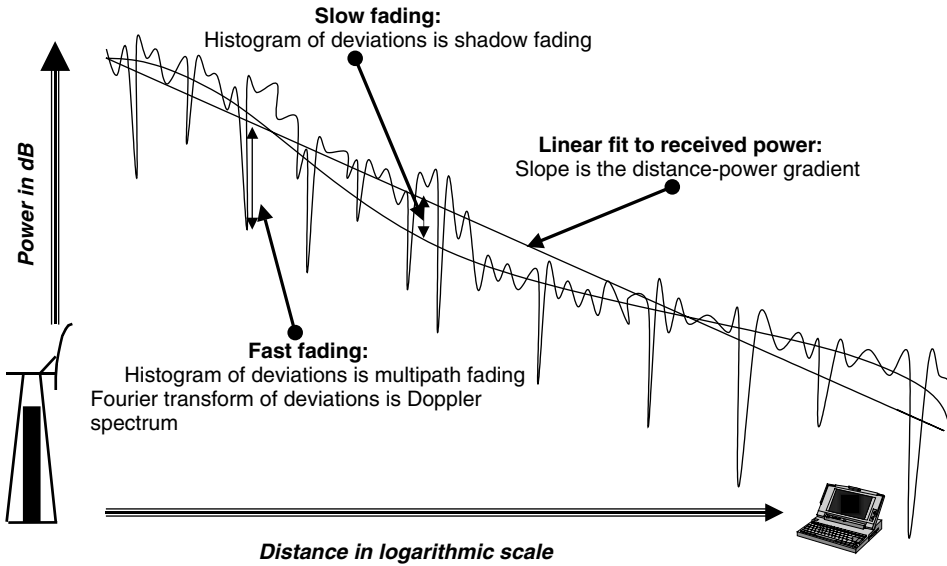


FIGURE 4.1 Received power versus distance between a mobile terminal and a base station, linear fit with a fixed slope, multipath fading, and shadow fading.

transmitter and the receiver. If we know all the paths between the transmitter and the receiver, similar to Examples 3.3 and 3.4, we can use Eq. (3.2.2) to calculate the short- and long-term variation of the received signal deterministically using a ray-tracing computer program. Since ray-tracing software is site specific and computationally intensive, much simpler statistical models, discussed in this chapter, based on empirical measurements are the most popular methods for narrowband signal modeling.

When we have the empirical results, similar to Fig. 4.1, we plot the received signal strength versus distance to illustrate the temporal multipath fading. Fluctuations of the average received signal strength over a short window of time show the effects of shadow fading, which is the difference between the average received signal strength and the best-fit linear line to the data. The channel characteristics extracted from these narrowband channel measurements are (1) the relationship between distance and the average received power; (2) the statistics of the fluctuations in received signal power in local and extended areas; and (3) the Doppler spread, which provides a measure of the rate of fading in the channel. Considering Fig. 4.1, the slope of the best-fit line to the observed data is the *distance–power gradient*, representing the exponential rate of variation of power with the distance. The statistics of the temporal fast multipath fading are characterized by the probability density function (PDF) of the sampled values of the fast variations of the channel. As we will see later in this chapter, the most popular distribution for this variation is the Rayleigh distribution, and for that reason this type of fading is sometimes referred to as *Rayleigh fading*. The Fourier transform of the samples of the variation of the signal is the Doppler spectrum of the channel. The probability density function of the variations of the average amplitude of the fade gives the shadow fading characteristics of the channel.

Narrowband modems are designed to operate with certain tolerance to fluctuations in the power of the received signal. The range of operation of the receiver

and, consequently, the size of the cells in a cellular architecture, depend on the distance–power relationship. This relationship in indoor areas is related to the layout of the building and the materials used in its construction. As we show in Chapter 8, the statistics of the amplitude fluctuations provide information for the calculation of probability of error and probability of outage for different modulation techniques. The Doppler spread is helpful in the specification and design of adaptive algorithms such as automatic gain control and timing- or phase-recovery circuits. In this chapter we discuss measurement and modeling of the envelope of the received signal. These models divide the received signal envelope fluctuations into slow- and fast-fading variations, shown in Fig. 4.1. Section 4.2 is devoted to modeling the slow fluctuation of the average received signal envelope. Sections 4.2 and 4.3 provide models for the spectrum and statistics of fast multipath fading, respectively. Section 4.5 is devoted to simulation of the received signal envelope.

4.2 MODELING PATH LOSS AND SLOW SHADOW FADING

In this section we present models for the relationship between the average received power and the distance between an access point or a base station and a mobile terminal. The average received power changes slowly with the distance between the transmitter and the receiver and the architectural setting of the objects, such as walls or buildings, in the area in which wireless communication is taking place. As shown in Fig. 4.1, we separate the changes in the slow average received signal strengths into a linear component represented by a distance–power gradient and random deviations from the linear fit, which we refer to as slow shadow fading. In mobile telephone communication settings, shadow fading is caused by slow appearance of large objects, such as buildings or walls, between the transmitter and the receiver. In quasistationary wireless data applications for WLANs or WPANs, movement of people close to the transmitter or receiver antennas causes shadow fading. In this section we present several models for average received power in indoor areas, suitable for coverage calculation of the WLANs and WPANs, and some popular models for micro- and macrocellular systems applicable to the deployment of cellular systems.

4.2.1 Basic Path-Loss Model for Indoor Areas

The simplest method of relating the received signal power to the distance is to state that the received signal power P_r is proportional to the distance between transmitter and receiver d , raised to a certain exponent, which is referred to as the *distance–power gradient*; that is,

$$P_r = \frac{P_0}{d^\alpha}$$

where P_0 is the received power 1 m from the transmitter. For a free-space path, $\alpha = 2$; and for the simplified two-path model of an urban radio channel given as Example 3.2, $\alpha = 4$. For indoor and urban radio channels, the distance–power relationship will change with building and street layouts, as well as with construction materials, density, and height of buildings in the area. Generally, variations in the value of the distance–power gradient in different outdoor areas are smaller than variations observed in indoor and high-rise urban areas. The results of indoor radio propagation

TABLE 4.1 Coverage Areas and Distance–Power Relationships in Several Buildings

Building	Construction	Distance–Power			
		1 MW Distance ^a (m)	Relationship (Gradient)	Power Correlation ^b	Spread ± dB ^c
1. Offices	Brick	17	3.9	0.97	8
2. Offices First floor ^d	Brick	12	3.9	0.86	10
3. Offices	Brick	12	3.9	0.96	6
	Brick/block/ plasterboard, reinforced concrete shell	25	6.1	0.89	16
4. Offices	Brick/plasterboard	>Floor	5.3	0.99	1
Ground floor ^d	Brick/plasterboard	16	4.3	0.94	12
First floor	Brick/plasterboard	12	4.8	0.95	8
Second floor	Brick/plasterboard				
Through floors 1–5	Reinforced concrete floors	10	5.1	0.98	3
5. Offices First floor ^d	Plasterboard with	27	6.2	0.95	9
Ground floor	metal support studding	8	3.1	0.93	6
6. Laboratory	Block plus some metal-faced partitioning	20	6.5	0.96	8
7. Offices	Plasterboard	30 within, 60 outside	2.8	0.75	16
8. Offices	Plasterboard	32	3.7	0.96	7
9. Offices	Steel	10	5.7	0.92	10
10. House	Brick/breeze/ plasterboard	>Building	1.4	0.54	7
11. House	Brick/breeze block	>Building	4.0	0.76	7
12. House	Brick/breeze block	>Building	2.2	0.70	12
13. Workshop	Open plan	60	2.5	0.97	4
14. Hangar	Open plan	>Building	1.2	0.99	1

Source: [Ale83a].

^aRadial extent of coverage area with 1-mW source power.

^bCorrelation is the degree of fit of the best-fitting straight line computed by linear regression.

^cSpread is the maximum scatter of the points about the line.

^dBase receiver remains on this floor when measuring other floors in this building.

studies show values of α smaller than 2 in corridors or large open indoor areas and values as high as 6 in metal buildings (see Table 4.1). Similar large variations in the distance–power gradient are observed in urban canyons with dense high-rise buildings.

The distance–power relationship (in decibels) is given by

$$10 \log_{10} P_r = 10 \log_{10} P_0 - 10\alpha \log_{10} d$$

where $10 \log_{10} P_r$ and $10 \log_{10} P_0$ represent transmitted and received power at 1 m in decibels, respectively. The last term on the right-hand side of the equation represents the power loss in decibels with respect to the received power at 1 m, and it indicates

that for a 1-decade increase in distance, the power loss is 10α dB, and for a 1-octave increase in distance, it is 3α dB.

Example 4.1: Power Loss in a Decade and an Octave of Distance For a free-space path, $\alpha = 2$ and the power loss is 20 dB/decade or 6 dB/octave of distance. An additional 60 dB of power, provided by larger transmitters and more directed antennas with high antenna gains, increases the distance three orders of magnitude. This means that with 60 dB of additional power a transmitter that covers 1 km (the length of several blocks of buildings) can cover 1000 km (something close to twice the distance between Boston and Washington, DC). Another 60 dB can take us into distances useful for space communications. In urban areas, given the two-ray approximation discussed in Example 3.2, attenuation is 40 dB/decade or 12 dB/octave of distance. Therefore, we need an additional 120 dB to extend the coverage three orders of magnitude.

Example 4.1 illustrates the convenience of establishing an intuitive relation between power and distance when power is described in decibels. Another convenient move is to resort to path loss rather than the power to separate the transmitted power from the characteristics of the channel. If we define the path loss in decibels at a distance of 1 m, as $L_0 = 10 \log_{10} P_t - 10 \log_{10} P_0$, where P_t is the transmitted signal power, the total path loss L_p in decibels is given by

$$L_p = L_0 + 10\alpha \log_{10} d \quad (4.2.1)$$

which represents the total path loss as the path loss in the first meter plus the power loss relative to the power received at 1 m. The received power in decibels is the transmitted power in decibels minus the total path loss L_p . This normalized equation is occasionally used in the literature to represent the distance–power relationship. Using Eq. (3.2.1) path loss at 1 m is given by

$$L_0 = 10 \log_{10} P_t - 10 \log_{10} P_0 = 10 \log_{10} \frac{P_t}{P_0} = -10 \log_{10} G_t G_r \left(\frac{\lambda}{4\pi} \right)^2$$

Using this equation to calculate the path loss in the first meter and then using Eq. (4.2.1), one can simply find the coverage of a system in a given environment.

Example 4.2: Coverage Calculation for IEEE 802.11b and g in an Office Area IEEE 802.11b and g operate at 2.4 GHz. Assuming that transmitter and receiver antenna gains are 1, the path loss in 1 m is

$$L_0 = 10 \log_{10} G_t G_r \left(\frac{f/c}{4\pi} \right)^2 = 10 \log_{10} \left(\frac{3 \times 10^8 / 2.4 \times 10^9}{4\pi} \right)^2 = 40.04 \text{ dB}$$

The maximum transmitted power is 100 mW (20 dBm) and the minimum sensitivity of the receiver is around -90 dBm, which allows a maximum path loss of $20 \text{ dBm} - (-90 \text{ dBm}) = 110 \text{ dB}$. In a semiopen office area with a distance–power gradient of 3, coverage of the system can be determined from Eq. (4.2.1); $110 = 40.04 + 10 \times 3 \log_{10} d$. That leads to $d = 10^{69.96/30} = 215 \text{ m}$.

Distance–Power Gradient Measurement. The empirical data are collected from the received signal strength at different distances. Accurate measurement of the signal strength can be performed by a measurement system transmitting a continuous sinusoid and measuring variations in the received signal strength with time. We discuss one such system using a network analyzer in Section 4.3.2. Another simple and inexpensive method of measuring the signal strength, often used in cellular and WLAN industry, is to take advantage of the infrastructure of the wireless network. All base station and access points send periodic synchronization packets to announce the network identification and provide a signal to synchronize the mobile terminals with the base station or the access point. There is commercially available software that reads the received signal strength of these packets and provides for measurement of the signal strength.

Example 4.3: Measurement of Received Power for IEEE 802.11b and g IEEE 802.11 access points send a special packet called a *beacon* almost every 100 ms. This packet provides information as to the access point address and how to synchronize to the system. There are a number of commercially available protocol analyzer software packages that take advantage of the signal strength of the beacon packets to measure the received signal strength in a mobile terminal such as a laptop or a notepad computer. Figure 4.2 shows three sample measurements of the received signal strength from three different access points installed in the Atwater Kent Laboratories

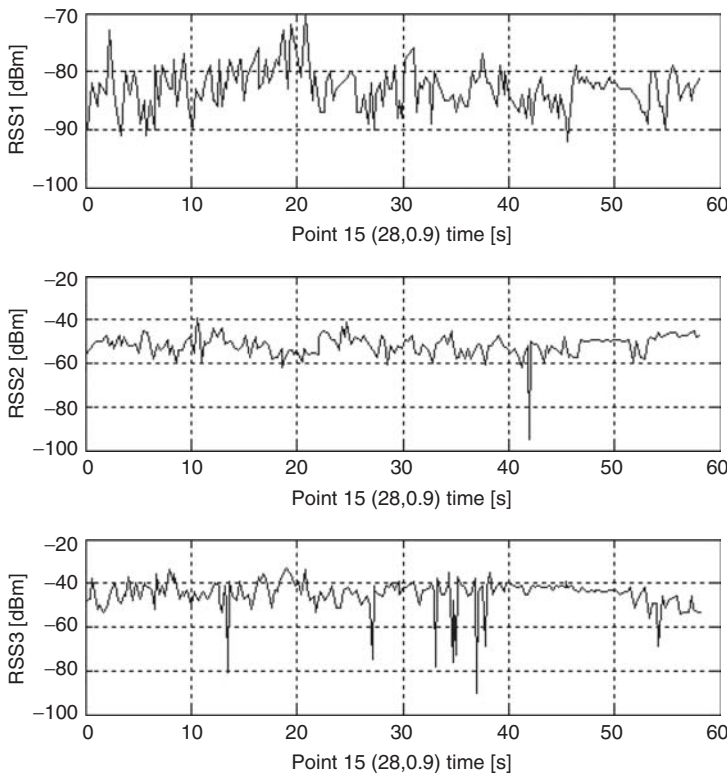


FIGURE 4.2 Measurement of the received signal strength from an IEEE 802.11b and g access point.

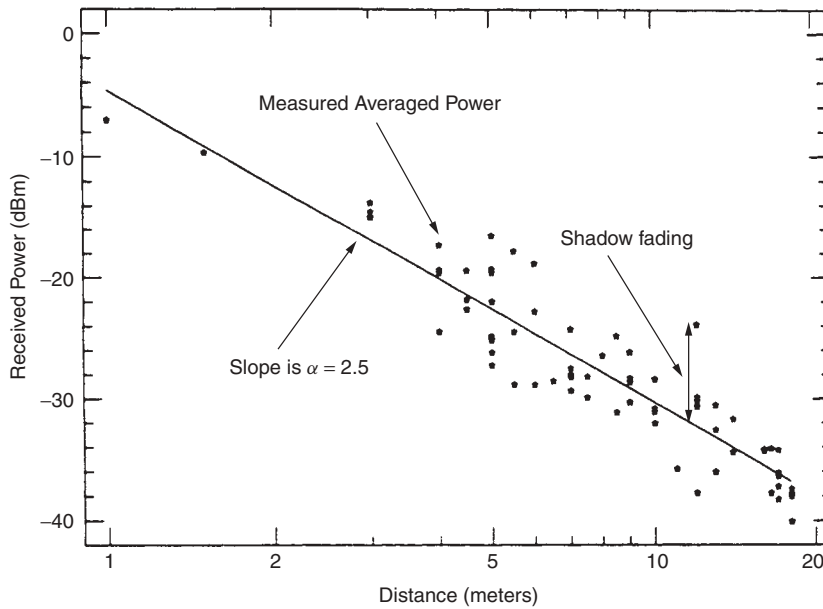


FIGURE 4.3 Scatter plot of power (dBm) versus distance on a logarithmic scale for a wide-band indoor radio measurement experiment.

at Worcester Polytechnic Institute. Measurements are taken for 60 s in a fixed location in the third floor, and variations in the amplitude are due to multipath fading caused by local movements. The location is somewhere between access points 2 and 3, and access point 1 is in the lower floor of the building. As a result, the average of the RSS3 (the closest access point) is around -40 dBm, the average of RSS2 is around -50 dBm, and the average of the RSS1 is less than -80 dBm. Another independent and yet interesting observation is that using these RSS measurements, one can have an idea about the location of the mobile terminal. This principle is used in indoor positioning and is discussed in Chapter 13.

As we described earlier in this section, one way to measure the distance–power gradient is to move away from the antenna and record the power as a function of increasing distance. Then we plot the power in decibels for distances in logarithmic values and determine the distance power gradient as the slope of the best-fit line to the measurements, the way it is described in Fig. 4.3. This method is suitable for large open areas, where the distance–power relation remains the same in all directions, and in urban areas, where the setting of streets forces the mobiles to move in straight lines. However, in most indoor areas, when we change the direction of movement away from a base station or access point, the architectural setting between the transmitter and the receiver and, consequently, the distance–power gradient change. Therefore, to measure the gradient of the distance–power relationship in a given indoor area, the transmitter is fixed at one location and the receiver is placed at a number of locations in different directions with different distances between the transmitter and receiver. Either average received power or average path loss is plotted in decibels against the distance on a logarithmic scale. The slope of the best-fit line through the measurements is taken as the gradient of the distance–power relationship.

Example 4.4: Measurement of the Distance–Power Gradient in Indoor Areas As we saw in Figs. 3.7 and 3.13, the distance–power gradient obtained from narrowband and wideband measurements have the same values. Calculation of power from the result of wideband measurements provides an average received power in a local area, resulting in smaller deviations from the best-fit line caused only by shadow fading. Figure 4.3 shows a set of scattered wideband measurements of averaged received power taken in an indoor area at distances from 1 to 20 m, together with the best-fit line through the measurements. The bandwidth of the measurement system is 200 MHz (details of wideband measurement systems are presented in Chapter 5), which we assume is wide enough to average the fast variations of received signal strength due to multipath fading. The distance–power gradient (i.e., the slope of the best-fitline) in this area is $\alpha = 2.5$, which is slightly more than the free-space propagations. Deviations of each sample measurement from the best-fit line are interpreted as sample values of shadow fading. The statistics of these samples can be used to model the behavior of shadow fading.

The earliest statistical measurements of the distance–power gradient and power fluctuations in an office environment were reported by Alexander of British Telecom for a cordless telephone application [Ale82]. The measurements were made by fixing the transmitter while moving the receiver to various locations in a multiple-room office. The measurements were made using a small handheld 30-mW transmitter operating at 941 MHz, with a vertically polarized quarter-wave dipole antenna. The receiver used a half-wave vertically polarized dipole antenna and had a dynamic range of 60 dB.

The first experiment of a series was performed in a building with steel partitioning [Ale82]. The receiver was fixed in one room, and the transmitter was moved among 13 other rooms. The distance–power gradient measured in this experiment was 5.7. That experiment was followed by a set of measurements made in buildings constructed with various materials [Ale83]. Table 4.1 shows the results of measurements made in various buildings. For each building or area within a building, the table shows the coverage for 1 mW of transmitted power, the distance–power gradient, the correlation coefficient,¹ and the decibel spread from the best-fit line. The maximum values of the distance–power gradient are around 6, which corresponds to buildings with concrete and metal structures, with communication among floors. Values of about 2 or smaller are shown for open areas and plasterboard partitions, where the gradient is close to or even better than that of free space. Gradients lower than those for free space are observed in areas such as hallways, where the structure acts as a waveguide between transmitter and receiver. Measurements in brick buildings give intermediate values of about 4. The maximum spread of power is ± 16 dB from the best-fit line. The correlation has its highest value where the building architecture provides a linear arrangement of rooms; minimum correlation is observed in houses with brick or breeze-block construction.

¹Correlation between two sequences $\{x(n)\}$ and $\{y(n)\}$, of complex variables each of length N is given by

$$R_{xy} = \frac{\sum_{n=1}^N x^*(n)y(n)}{\sqrt{\sum_{n=1}^N |x(n)|^2 \sum_{n=1}^N |y(n)|^2}}$$

In our case the first sequence is the measured signal strengths, and the second sequence is the estimate of these measurements from the best-fit linear line. This provides a number for the quality of the fitted curve.

After those earliest measurements in the early 1980s, many researchers performed narrowband measurements within buildings for a few years, primarily to determine the distance–power relationship and the statistics of the received signal envelope. Arnold et al. [Arn89] reported copolarized attenuation measurements made at 815 MHz within two office buildings. The measurements were made using a modified handheld 815-MHz transceiver with an integral vertical half-wavelength coaxial sleeve dipole as the signal source. The received signal was captured with a similar antenna attached to a 9.5-ft mast connected to a modified 815-MHz FM communications receiver and a digital storage oscilloscope. Narrowband radio-wave propagation measurements into and within a building were also reported by Barry and Williamson at Auckland University [Bar86]. In this work, measurements were made using a 17-W 927-MHz transmitter feeding a half-wavelength dipole antenna. The mobile receiver had a similar antenna mounted 1.6 m above floor level and connected to field-strength measuring equipment. More measurements in frequencies from 914 MHz to 4 GHz by Anderson et al. [And94] in residential home, offices, and manufacturing floors report distant power gradients of as low as 2.2 up to 3.3.

The simple model provided by Eq. (4.2.1) is popular for coverage on a single floor of a building. To extend this model for application to multifloor buildings, additional signal attenuation by the floors in the building is included as a constant independent of the distance [Mot88b]. The path loss in this case is given by

$$L_p = L_0 + L_f(n) + 10\alpha \log_{10} d \quad (4.2.2)$$

where $L_f(n)$ represents the signal attenuation provided by each floor and n is the number of floors through which the signal passes. The simplest model for the path loss per floor is $L_f(n) = nF$, which assumes that path loss per floor, F , is a fixed value. To determine F , the received power is plotted versus distance, and the best-fit line is determined for each different value of F . The value of F that provides the minimum mean-squared error between the line and the data is taken as the value of F for the experiment. For indoor radio measurements at 900 MHz and 1.7 GHz, Motley and Keenan have reported values of $F = 10$ and 16 dB, respectively [Mot88a]. In work presented in [Mot88a], the relationship between path loss and the number of floors is linear. However, the results of measurements reported in [Ake88, Rap92, Sei92] do not agree with this assumption. Toward the end of this section we present the JTC model, in which path loss associated with different floors is not uniform.

Shadow Fading and Fading Margin. Equation (4.2.2) suggests an exact relationship between path loss and distance. But in general, buildings are not symmetric and the furnishing is not the same in all directions, and therefore we expect to find somewhat different path losses in different directions. A deterministic model for this variation is not feasible, and therefore we usually resort to statistical models. The cause of this power loss, obstruction by other objects around the receiver, is usually referred to as *shadow fading* or *large-scale fading*. To determine the statistics of shadow fading, the results of Eq. (4.2.2) are compared with the measured average path loss in a large area. The distribution of the error between the results of measurement of the average path loss and the prediction given by Eq. (4.2.2) provides the model for shadow fading. The result of measurements on indoor [Mot88a, Mot88b, Gan91b, How91] and urban [Jak74, Lee89] radio channels show that a lognormal distribution best fits the large-scale variations of the signal amplitude. In [Mot88b], variations of the mean value

of the signal in indoor areas were found to be lognormal with a variance of 4 dB. To add shadow fading to our basic model for path loss, we may improve upon Eq. (4.2.2) with

$$L_p = L_0 + L_f(n) + 10\alpha \log_{10} d + l \quad (4.2.3)$$

where l is a zero-mean normally distributed random variable with standard deviation σ , representing the shadow fading.

Example 4.5: JTC Path-Loss Model for PCS Bands Here we describe briefly a path-loss model for indoor areas, recommended by a technical working group of the TIA/ANSI Joint Technical Committee (JTC) for 1900-MHz PCS bands [JTC94]. Table 4.2 gives a set of suggested parameters in decibels for the path-loss calculation using Eq. (4.2.3). The rows of the table provide the path loss in the first meter, the gradient of the distance–power relationship, the equation for calculation of multifloor path loss, and the variance in the lognormal shadow fading parameter. It is assumed that the base and portable stations are inside the same building. The parameters are provided for three classes of indoor areas: residential, offices, and commercial buildings.

When, similar to Example 4.2, we use Eq. (4.2.1) or (4.2.2) for calculation of the coverage, the terminals located at the edge of the coverage will be in outage (have no coverage) 50% of the time. This is due to the fact that according to the effects of shadow fading reflected by random variable, l , in Eq. (4.2.3) we have an additional normally distributed random variable that produces a positive shadow fading element to the path loss 50% of the time. To compensate for the poor performance at the edges of the coverage caused by shadow fading, in cellular networks the transmitted power is increased by a fixed value, F_M , referred to as *fade margin*. This increase in power will reduce the probability of outage at the edges from 50% to

$$P_{\text{out}} = \int_{F_M}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-l^2/2\sigma^2} dl = \frac{1}{2} \text{erfc} \left(\frac{F_M}{\sqrt{2}\sigma} \right)$$

The complementary error function for normal distribution, erfc ,² is available in standard tables and MatLAB. Considering the effects of shadow fading and fading margin, the meaning and calculation of the coverage would slightly change.

TABLE 4.2 Recommended Parameters for Path-Loss Calculations in PCS Indoor Radio Environments

Parameter	Environment		
	Residential	Office	Commercial
L_0 (dB)	38	38	38
10α	28	30	22
$L_g(n)$ (dB)	$4n$	$15 + 4(n - 1)$	$6 + 3(n - 1)$
Lognormal fading (standard deviation dB)	8	10	10

Source: [JTC94].

² $\text{erfc}(x) = (2/\sqrt{\pi}) \int_x^{\infty} e^{-t^2} dt$.

Example 4.6: Coverage Calculations with a Fading Margin Considering the outage effects, the $d = 10^{69.96/30} = 215$ m coverage of IEEE 802.11b and g, calculated in Example 4.2, was for 50% outage at the edges of the coverage area. Assuming that the standard deviation of the lognormal shadow fading is 8 dB, for an outage rate of 5% at the edges of the coverage area, the fading margin, F_M , is calculated from $0.05 = 0.5\text{erfc}[F_M/(\sqrt{2} \times 8)]$. Using MatLAB or a table, one calculates this margin to be $F_M = 13.16$ dB. Since maximum transmitted power is fixed at 20 dBm, the maximum acceptable path loss to maintain 5% outage at the edges would be $20 - (-90) - 13.16 = 96.84$ dBm. The coverage is then calculated from $96.84 = 40.04 + 10 \times 3 \log_{10} d$, which leads to $d = 10^{56.80/30} = 78$ m.

The basic single-gradient model introduced in this section differentiates buildings on the basis of their distance–power gradient. In practice, as we saw in the case of the JTC model, we divide buildings into several categories, three in the case of JTC, and we assign typical values of distance–power gradient, path loss per floor, and variance of the shadow fading to each of these buildings. In the next two sections we introduce two other models for the indoor radio communications.

4.2.2 Wall-Partitioned Path-Loss Models for Indoor Areas

Another simple path-loss model for indoor areas is the wall-partitioned model described in [Sei92, Rap02]. This model fixes the path loss at $\alpha = 2$ and introduces path loss per wall for all walls between the transmitter and the receiver. Therefore, the path-loss model in this case is given by

$$L_p = L_0 + 20 \log_{10} d + \sum_i L_{w_i} \quad (4.2.4)$$

in which L_{w_i} is the path loss for the i th path between the transmitter and the receiver. Table 4.3 shows some decibel loss values measured by Harris Semiconductor at 2.4 GHz for various types of partitions [Pah02a]. More detailed measurements of loss per wall at various frequencies and number of walls are available in [Rap02]. In a multifloor environment, we can easily add the path loss per floor as another wall. To complete this model we need to add shadow fading in the same way as we did in Section 4.2.1. The advantage of this model is that it includes the building material in the calculation of path loss. The weakness of this model is that accurate measurement of the path loss is difficult and we need long tables to accommodate all situations.

**TABLE 4.3 Partition-Dependent Losses:
Attenuation per Wall at 2.4 GHz (dB)**

Window in brick wall	2
Metal frame, glass wall into building	6
Office wall	6
Metal door in office wall	6
Cinder wall	4
Metal door in brick wall	12.4
Brick wall next to metal door	3

4.2.3 Distance-Partitioned Models for Indoor and Microcellular Areas

The area between the transmitter and receiver is often not homogeneous, with a single distance–power gradient. In these cases the power loss should be described with multiple distance–power gradients, each associated with a segment of the path between the transmitter and receiver. Results of wideband measurements in a partitioned indoor area show significant differences among the values of the distance–power gradient in different parts of a building [Gan91a].

Example 4.7: Multiple Gradients in Indoor Areas Figure 4.4 depicts the middle part of the third floor of the Atwater Kent Laboratories at Worcester Polytechnic Institute. The receiver is located at the center of Room 317, and the transmitter is moved to different locations in various rooms for measurements of the received signal strength. The area is divided into three segments: the interior of a small laboratory (Room 317), corridors around the laboratory, and offices on the opposite side of the corridor. The three gradients 1.76, 2.05, and 4.21 were calculated from the results of the measurements made in the three subareas. Inside the small laboratory, all the locations provide a strong line-of-sight (LOS) connection and the gradient is 1.76, which is less than the free-space gradient. This is consistent with the results of two-dimensional ray tracing in a similar environment presented in Examples 3.4 and 3.6. In the corridors there is at least one plaster wall with metal studs between the transmitter and receiver, and the gradient is close to that of free-space propagation. The third subarea, with a gradient of 4.21, includes at least two walls, one of which contains a number of metal doors. Also, inside the rooms are several metal shelves, cabinets, and desks.

In [Ake88], based on measurements in a multistory building, the path loss was modeled with four different gradients. In these measurements the transmitter was fixed in the middle of a corridor and the receiver was moved away from the transmitter to other corridors and rooms. The model developed from the measurements suggests a gradient $\alpha = 2$ for the distances $1 < d < 10$ m, a value of 3 for $10 < d < 20$ m, a value of 6 for $20 < d < 40$ m, and a value of 12 for $d > 40$ m. This leads to the following equations for path-loss calculations, which include the effects of shadow fading:

$$L_p = L_0 + \begin{cases} 20 \log_{10} d, & 1 < d < 10 \text{ m} \\ 20 + 30 \log_{10} \frac{d}{10}, & 10 < d < 20 \text{ m} \\ 29 + 60 \log_{10} \frac{d}{20}, & 20 < d < 40 \text{ m} \\ 47 + 120 \log_{10} \frac{d}{40}, & d > 40 \text{ m} \end{cases}$$

Example 4.8: WLAN Coverage Using a Distance-Partitioned Model We use the equations above for calculation of IEEE 802.11b and g with a maximum path loss of 110 dB and a path loss in the first meter of 40.4 dB, as we used in Examples 4.2 and 4.6. At 40 m the path loss is $40.4 + 47 < 110$ dB; therefore, the coverage must be more than 40 m, and we shall use the last line of the equation. Then we have $110 = 40.4 + 47 + 120 \log_{10}(d/40)$, from which we have $d = 40 \times 10^{22.6/120} = 61.72$ m. This result provides a more pessimistic estimate of the 78-m coverage calculated in

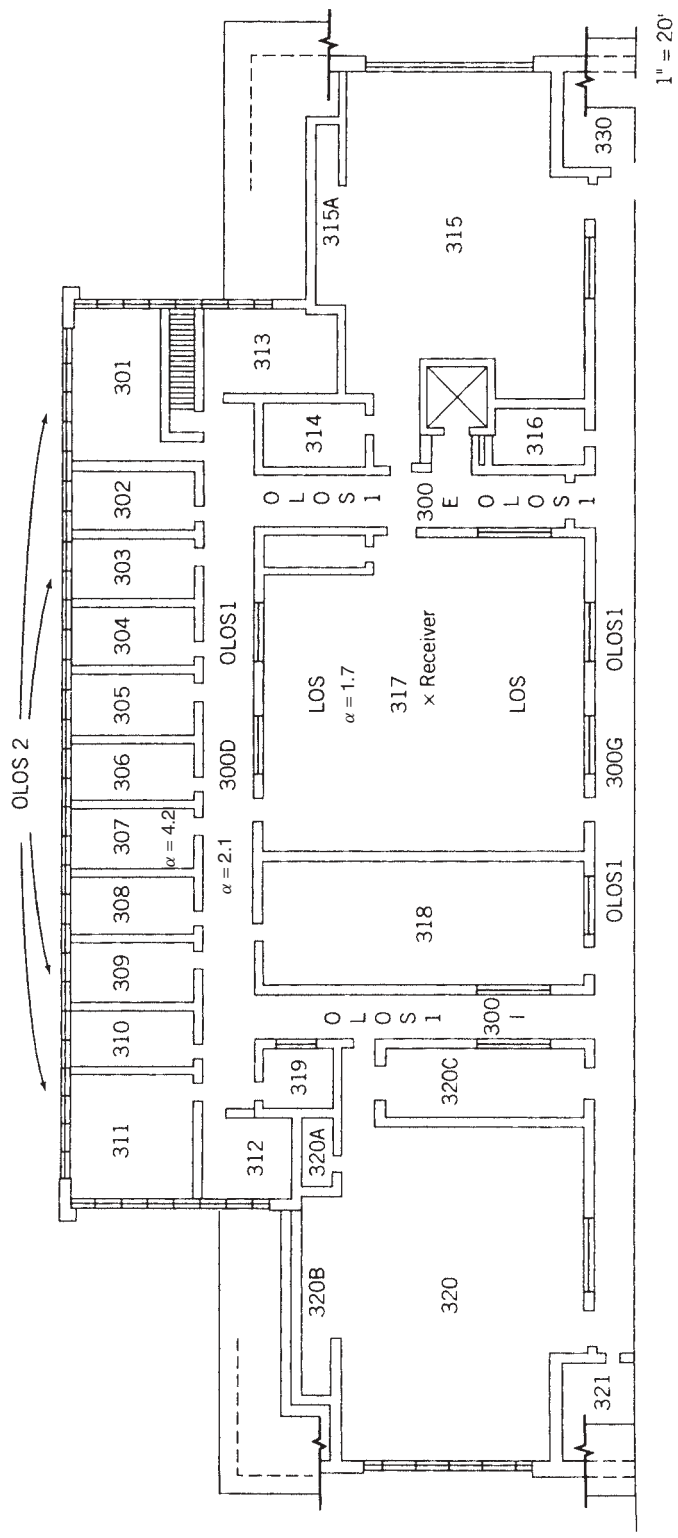


FIGURE 4.4 Layout of the third floor of the Atwater Kent Laboratories at Worcester Polytechnic Institute, used for partitioned measurements [Gan91a].

Example 4.6. The reader should note that statistical models for the coverage provide a statistical approximation for the actual coverage, so the results obtained from different models are not always the same.

More recently, path-loss models for UWB applications use distance-partitioned models. We provide more details of these models in Chapter 12.

4.2.4 JTC Path-Loss Model for Microcells and Macrocells

In Table 4.2 we described the JTC's path-loss model for indoor areas. Here we describe briefly path-loss models recommended by the same TIA/ANSI Joint Technical Committee (JTC) group for micro- and macrocellular path-loss modeling in 1900-MHz PCS bands [JTC94]. These models assume the distance between the base and mobile stations to be less than 1 km. We first provide two models for path loss in a microcell environment, which assumes further that the base station antenna height is below rooftop level. One of the two models assumes that the physical geometry of the microcell is known and the other model is recommended for the situation where the geometry is not available. For microcells with the known physical geometry of typical blocks of streets, as shown in Fig. 4.5, the model divides the distances into two line-of-sight (LOS) and one obstructed-line-of-sight (OLOS) regions. The first LOS region is inside the Fresnel zone defined by the breakpoint distance [Jen65]:

$$d_{bp} = \frac{4h_b h_m}{\lambda}$$

where h_b and h_m are the heights of the base and mobile station antennas, respectively, and λ is the wavelength of the carrier frequency. In this region the power received from the LOS path dominates the total power of the other paths and the propagation loss is the same as for free-space propagation. The second LOS region starts at d_{bp} and continues to d_{cor} , where the mobile unit turns a corner and loses the LOS path. In this region the gradient is assumed to be 4, to include the direct LOS path as well as the path reflected from the ground. The third region starts from d_{cor} , where the mobile loses the LOS path. The gradient in this region is assumed to be 5, and an additional

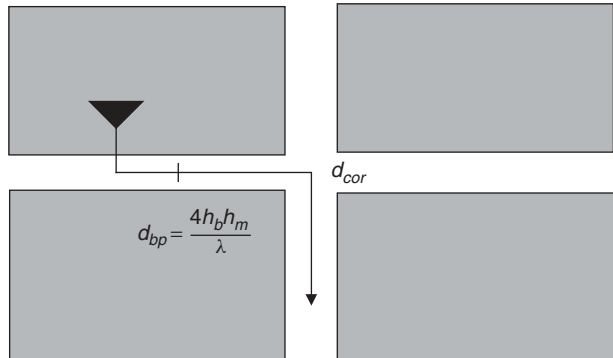


FIGURE 4.5 Geometric JTC path-loss model for microcells.

path loss of $L_{\text{cor}} = 20$ dB is added to compensate for the immediate power drop after turning the corner. The formula for calculation of the path loss is then given by

$$L_p = 38.1 + \begin{cases} 20 \log_{10} d, & d < d_{\text{bp}} \\ 20 \log_{10} d_{\text{bp}} + 40 \log_{10} \frac{d}{d_{\text{bp}}}, & d_{\text{bp}} < d < d_{\text{cor}} \\ L_{\text{cor}} + 20 \log_{10} d_{\text{bp}} + 40 \log_{10} \frac{d_{\text{cor}}}{d_{\text{bp}}} + 50 \log_{10} \frac{d}{d_{\text{cor}}}, & d > d_{\text{cor}} \end{cases}$$

where the path loss in the first meter of distance from the base station is $L_0 = 38.1$ dB, which is the path loss at 1900 MHz in a 1-m distance with antenna gains set to 1.

For cases where a detailed description of the microcellular environment is not available, the JTC document recommends the following general path-loss model:

$$L_p = 38.1 + \begin{cases} 25 \log_{10} d, & d < d_{\text{bp}} \\ 25 \log_{10} d_{\text{bp}} + 45 \log_{10} \frac{d}{d_{\text{bp}}}, & d > d_{\text{bp}} \end{cases}$$

In this model we have two regions: the inside and outside of the Fresnel zone. Inside the zone the distance power gradient is assumed to be 2.5, which is slightly more than the free-space path-loss gradient. Outside the zone the gradient is assumed to be 4.5, which is the average of the gradients in the second and third zones of the model with specific geometry.

The JTC recommendation for the macrocell environment, where the antenna height is above the rooftop level, is given by

$$L_p = \max(A + B \log_{10} d, 38.1 + 20 \log_{10} d)$$

where $A = 88 - 13.82 \log_{10} h_b + C$ and $B = 49 - 6.55 \log_{10} h_b$. Table 4.4 provides the clutter correction factor C used in the foregoing equation as well as the building penetration loss and standard deviation of the lognormal shadowing parameters, as suggested by the JTC working group. The building penetration losses and shadow fading statistics in this table can also be applied to models.

TABLE 4.4 Recommended Parameters for Path-Loss Calculations in PCS Outdoor Radio Environments

Parameter	Environment			
	Urban High-Rise	Urban/Suburban Low-Rise	Residential	Rural
Clutter correction factor C (dB)	0	−6	−12	−18
Building penetration loss (dB)	15	15	10	10
Lognormal shadowing	10	10	10	10

Source: [JTC94].

4.2.5 Okumura–Hata and COST-231 Models for Macrocells and Microcells

The JTC models for macro- and microcellular systems provided in Section 4.2.4 work for cell overages of less than 1 km. The coverage of the cellular systems are an order of magnitude higher than this. For such operations the height of the transmitter and the receiver antenna can be substantially different because the transmitter antennas are usually installed on high-altitude locations such as the tops of the hills, and the mobile terminal operates in a large area that may include low valleys. The most commonly used path-loss model for these areas is the model originally developed by Okumura et al. [Oku68], based on extensive radio propagation studies made in Tokyo. This model was adapted for computer simulation by Hata [Hat80]. The path loss in the Okumura–Hata model is given by the expression

$$L_p = 69.55 + 26.16 \log_{10} f - 13.82 \log_{10} h_b - A(h_m) \\ + (44.9 - 6.55 \log_{10} h_b) \log_{10} d$$

where the frequency of operation, f , is in megahertz, the height of the base station, h_b , and mobile, h_m , antennas are in meters, and the distance, d , is expressed in kilometers. In this model the range of frequency is $150 < f < 1500$ MHz, the range of the height of the base station antenna is $30 < h_b < 300$ m, and the distance range is given by $1 < d < 20$ km. The function $A(h_m)$ in decibels for a small or medium-sized city is

$$A(h_m) = (1.1 \log_{10} f - 0.7)h_m - (1.56 \log_{10} f - 0.8)$$

with $1 < h_m < 10$ m. For a large city we have

$$A(h_m) = 3.2[\log_{10}(11.75h_m)]^2 - 4.97, \quad d \geq 400 \text{ MHz}$$

Over the restricted range of parameters, Hata's equations provide a simple but very accurate approximation to Okumura's method. These equations have evolved out of experimental results by taking into account various parameters causing attenuation. More detailed treatments of models for path loss in urban radio channels are available in [Jak74, Lee89, Par89].

Example 4.9: WLAN Coverage Using the Okumura–Hata Model The path loss of a 900-MHz cellular system operating in a large city from a base station with a height of 100 m and a mobile station installed in a vehicle with an antenna height of 2 m, at a distance between the mobile and the base station of 4 km is

$$A(h_m) = 3.2[\log(11.75h_m)]^2 - 4.97 = 1.045 \text{ dB} \\ L_p = 69.55 + 26.16 \log f_c - 13.82 \log h_b - A(h_m) \\ + (44.9 - 6.55 \log h_b) \log d = 137.3 \text{ dB}$$

To extend the Okumura–Hata model for PCS frequency bands operating at 1500 to 2000 MHz, the European Co-operative for Scientific and Technical Research (COST) came up with the COST-231 model for urban radio propagation at 1900 MHz. This modified Okumura–Hata model is defined as

$$L_p = 46.3 + 33.9 \log f_c - 13.82 \log h_b - A(h_m) + (44.9 - 6.55 \log h_b) \log d + C_M$$

in which C_M is 0 dB for large cities and 3 dB for medium-sized cities and suburban areas. The COST-231 model is good for distances between 1 and 20 km with the height of the mobile antenna and base station ranging between 1 to 20 and 30 to 200 m, respectively.

The path loss modeling methods described in this section are based on generalizations of results obtained in certain specific measurement programs. However, there is no universally accepted model for path loss, in particular for indoor and urban canyon areas. One important limitation of these modeling methods for indoor and urban canyon areas is that they do not include the specification of building characteristics. As a consequence, much attention is being given to building-specific radio propagation models such as ray tracing, and these techniques may emerge as the leading techniques for the future. However, there are drawbacks in the use of building-specific radio propagation models: the complexity of computation, the need for large amounts of computer memory, and the enormous cost of creating a detailed electronic map. With the growing availability of electronic maps and the continuing increase in the computational power and memory capacity of computers, it is expected that increasingly accurate building-specific radio propagation models will evolve. This subject is discussed further in Chapter 6.

4.3 DOPPLER SPECTRUM OF FAST ENVELOPE FADING

In Section 4.2 we modeled the average received power over a large area with a deterministic and a random component. The deterministic component was a function of distance, a function that changed from one physical environment to another. The random component, shadow fading, was modeled using a lognormal-distributed random variable with a variance that can be slightly different in different environments. This slow-fading component represents the difference in the overall characteristics of the environment. It remains essentially invariant in areas with dimensions an order of magnitude larger than the wavelength of the carrier frequency, and it changes as the receiver moves from an area to another. We use the models developed in Chapter 3 to determine the average received power at and around a location for the receiver. However, in and around each location, as the receiver moves on the order of a wavelength or other objects move close to the transmitter or the receiver, the received narrowband signal power has fast fluctuations, due to *multipath fading*. As we showed in Chapter 3, multipath fading causes power fluctuations on the order of 30 to 40 dB. The statistical fluctuation of the amplitude of the received power is the superposition of fast local multipath fading over slow shadow fading, which is illustrated in Fig. 4.3. The slow shadow fading component causes changes in the mean value of the received power as the terminal moves from one area to another. The fast-fading component changes rapidly as the transmitter or the receiver moves slightly or other objects are moved in the vicinity of the transmitter or the receiver. To model the multipath fading characteristics of the channel, we analyze the statistics of the temporal and local amplitude variations as well as the spectrum of the variations, termed the *Doppler spectrum*, obtained from Fourier transform of the received signal fluctuations.

As we discussed in Chapter 3, the complex envelope of the received narrowband signal on indoor and urban radio channels is represented by the complex addition of individual phasors representing the magnitudes and the phases of the individual paths

(see Fig. 3.3). Small movements of the transmitter and receiver or the movement of objects around them will cause random changes in the magnitude and phase of the individual paths; and according to the central limit theorem, the sum of all paths will form a complex Gaussian random variable. In the absence of a dominant LOS path, the Gaussian process has zero mean, and in the presence of a dominant path it will have a non-zero-mean value. The magnitude of a complex Gaussian random process obeys a Rayleigh distribution if the mean of the process is zero, and obeys a Rician distribution otherwise. The phase of a complex Gaussian process always has a uniform distribution. As a result, for both indoor and urban radio channels we may assume that multipath fading is generally Rayleigh unless there exists a strong LOS component, in which case multipath fading is Rician. To examine the accuracy of this model, in Section 4.4 we examine the results of a few indoor radio propagation experiments. In the rest of this section we introduce the analytical Clarke model for calculation of the Doppler spectrum of mobile radio channels, we show some experimental measurement of the Doppler spectrum caused by traffic or antenna movements in an indoor area, and we show how the Doppler spectrum is used to calculate the rate and duration of fades.

4.3.1 Clarke Model for a Doppler Spectrum in Urban Areas

A simple and useful model for mobile radio channels has been proposed by Clarke [Cla68]. This model assumes a dense array of randomly oriented scattering objects located around the mobile unit. In the definition of this model, Clarke makes the simplifying assumption that all the scatter components arrive with the same amplitude (termed *isotropic scattering*) but that the components are distinguished from one another by the angles of arrival and the phases of the components. The angles of arrival and the phases of the received signals are both assumed to be distributed uniformly, and the arrival angle and phase of each component are assumed to be statistically independent of each other. With a uniform fixed amplitude of the signal components, the addition of phasors with uniformly distributed phase angles will result in a Rayleigh distribution for the magnitude of the complex sum of all the paths. This will change to a Rician distribution in the presence of a strong LOS path. In [Cla68] Clarke shows that the scatter-only model provides an accurate representation of mobile radio signals in heavily built-up areas such as New York City when the signal energy propagates from transmitter to receiver largely by way of scattering, either by reflection from the sides of buildings or by diffraction around buildings or other human-made or natural obstacles. In suburban areas, the received signals are often a combination of a scattered signal and a direct plane-wave signal, a condition represented by the Rician model.

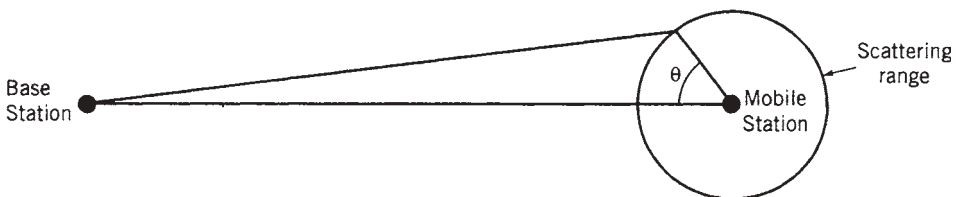


FIGURE 4.6 Ring scattering model for mobile radio communications.

Figure 4.6 provides a simplified description of the isotropic scattering model for mobile radio and is very useful for analyzing the Doppler spectrum of the channel. Assuming that a mobile terminal is moving toward the base station at a constant velocity v_m , the Doppler shifts associated with different paths are not the same because the velocity at which each path length is reduced depends on the angle of arrival of that path. The direct path from the transmitter to the receiver has the maximum positive Doppler shift of $f_m = v_m/\lambda$ associated with angle of arrival of zero, whereas the path arriving from a reflection behind the mobile terminal has the maximum negative Doppler shift of $-f_m$, associated with the angle of arrival of π . Other paths have Doppler shift values between these two limits, given by $f(\theta) = f_m \cos \theta$, where θ represents the angle of arrival of the path. Therefore, in adding the signal strength from all paths, the received signal associated with the path with angle of arrival θ is represented by a phasor of the form $A_\theta \exp[j2\pi f(\theta)t]$, where A_θ is the magnitude of the path associated with the arriving angle θ . With the assumption of equal magnitudes, $A_\theta = A$ for the arriving paths, and taking the model to the case of a continuum of arriving signal components, we can define the *azimuthal distribution of signal power* as

$$Z(\theta) = \frac{1}{2\pi}, \quad -\pi \leq \theta \leq \pi$$

where the total arriving signal power is normalized as 1.0. As described in Project 3 of Chapter 3, if the distance between the two stations is much larger than the radius of the scattering circle, the Doppler frequency is related to the azimuth angle by the relationship $f_d = f_m \cos \theta$, or $\theta = \cos^{-1}(f_d/f_m)$, and we can derive the Doppler spectrum as

$$D(f) = R_{HH}(0, f) = Z(\theta) \left| \frac{d\theta}{df} \right|$$

where we use f in place of f_d . Now, because

$$\frac{d}{dx} \cos^{-1} y = -\frac{1}{\sqrt{1-y^2}} \frac{dy}{dx}$$

we can write the Doppler spectrum as

$$D(f) = \frac{1}{2\pi f_m} \left[1 - \left(\frac{f}{f_m} \right)^2 \right]^{-1/2}, \quad |f| \leq f_m \quad (4.3.1)$$

Figure 4.7a shows the typical spectrum, in which for values of f close to $\pm f_m$, the height of the Doppler component rises to two high peaks at the edges of the spectrum. In the presence of a strong component with Rician-distributed envelope fading, as shown in Fig. 4.7b, the spectrum has an additional impulse, representing the shift associated with the strong component. The spectra described above have been shown to match experimental data gathered for mobile radio channels [Jak74]. For an indoor radio channel, the assumptions of equal component amplitudes and uniform distribution of the angles of arrival do not hold; and as we will see next, the Doppler spectra have shapes different from those of the mobile radio case. The JTC channel model for indoor areas assumes $D(f)$ to be a flat spectrum [JTC94].

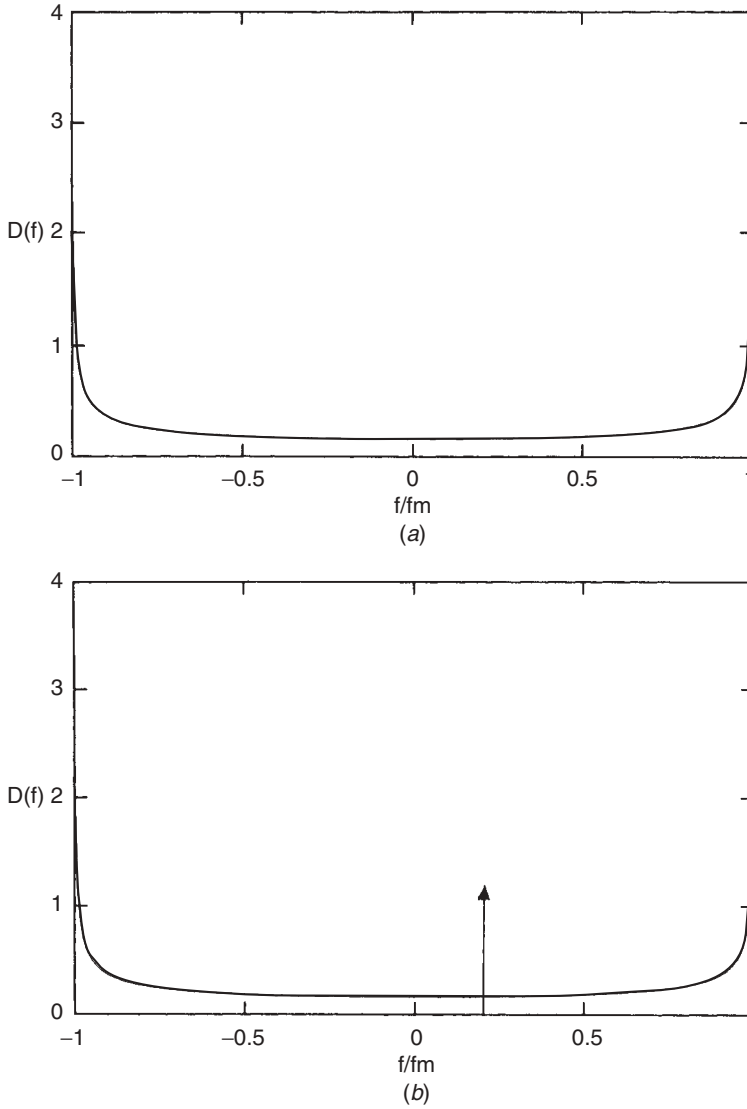


FIGURE 4.7 Doppler spectra for Rayleigh (a) and Rician (b) mobile radio channels.

4.3.2 Measurement of a Doppler Spectrum in Indoor Areas

We now present some results of measurements of Doppler spreading on indoor radio channels, as caused by traffic and local movements of the communication terminals. These were controlled experiments in which the only movements were those for which we were trying to determine the resulting Doppler spread [How90a]. As we saw earlier, the characteristics of the channel are influenced by the existence of a strong LOS path, so we consider both LOS and OLOS experiments. The measurements reported here were made on the third floor of the three-story Atwater Kent Laboratories at Worcester Polytechnic Institute. The layout of this floor is shown in Fig. 4.4.

To determine the Doppler spread of a radio channel, we require a system capable of measuring the short-term variations of the channel. The system should be capable of sampling the amplitude of the received signal at the Nyquist rate associated with the highest Doppler shift caused by movement of the equipment or neighboring objects.

A simple and accurate method of measuring short-term variations in the narrow-band characteristics of the indoor radio channel is to use a network analyzer in the experimental configuration shown in Fig. 4.8. Here a 910-MHz signal generated by the network analyzer is power-split and used as both the reference input to the network analyzer and, after passing through 100 ft of coaxial cable, the input to a transmitter RF amplifier having 45-dB gain. The output of the RF power amplifier is radiated by the transmitter dipole antenna. The signal from the receiver dipole antenna is passed through an attenuator and a series of amplifiers, with an overall gain of 60 dB. The output of the amplifier chain is returned to the network analyzer, where the time variations of the channel relative to the fixed reference input are measured. The measurement data files are then read and stored in a PC controller for subsequent analysis. In a 32-s interval, the network analyzer samples the received amplitude and phase at the rate of 25 samples/s. Therefore, the maximum Doppler shift measurable is 12.5 Hz and the resolution is 0.03125 Hz. As we will show later, the same system, with a different network analyzer configuration, is used for wideband measurements of indoor radio channel characteristics.

Whereas for the case of mobile radio the Doppler spectrum has the relatively regular shape shown in Fig. 4.7, Doppler spectra for indoor wireless applications have a variety of shapes. The indoor WLAN or WPAN user in a small room may observe a stationary channel with no Doppler spread. However, the same user may observe a Doppler spectrum associated with the movements of people around the transmitter and receiver if the system operates in a more populated, larger indoor area, such as a manufacturing floor or an office building. A cordless telephone user observes a Doppler spectrum associated with the random motions of the device as a person speaks on the phone. Figure 4.9 shows four time-domain plots of received signal amplitude variations, along with the four corresponding Fourier transforms $|H(f_c; t)|$, measured using the system shown in Fig. 4.8. Plots are shown for two LOS channels and two OLOS channels. Figure 4.9a shows data from an LOS experiment in which the environment was kept constant and the distance between the transmitter and the receiver was 1 m. There was no time-domain variation of the received signal, and consequently, the computed Fourier transform shows an impulse at zero frequency with no Doppler spread. In Fig. 4.9b the transmitter was moved randomly around a fixed point 12 m away from the receiver. The time-domain measurements show a maximum power deviation, P_{dev} , of 35 dB, and the corresponding Fourier transform has a bell shape with a Doppler spread width of $B_D = 4.9$ Hz. Figure 4.9c shows data from an OLOS experiment in which the transmitter and the receiver were at fixed locations 4 m apart, and there was random pedestrian traffic close to the transmitter. The maximum power deviation is seen to be 20 dB, and the Doppler spread has a two-sided exponential shape with $B_D = 5.1$ Hz. Figure 4.9d shows the results of cyclic motion of the transmitter to change the orientation and, consequently, the polarization of the antenna. The results show a maximum power fluctuation of 10 dB with a $B_D = 5.2$ Hz. The spectrum shows two strong components representing the rate of the cyclic motion. Generally, the shape of the Doppler spectrum is related to the nature of the movement, but the maximum Doppler shift is related to the fastest

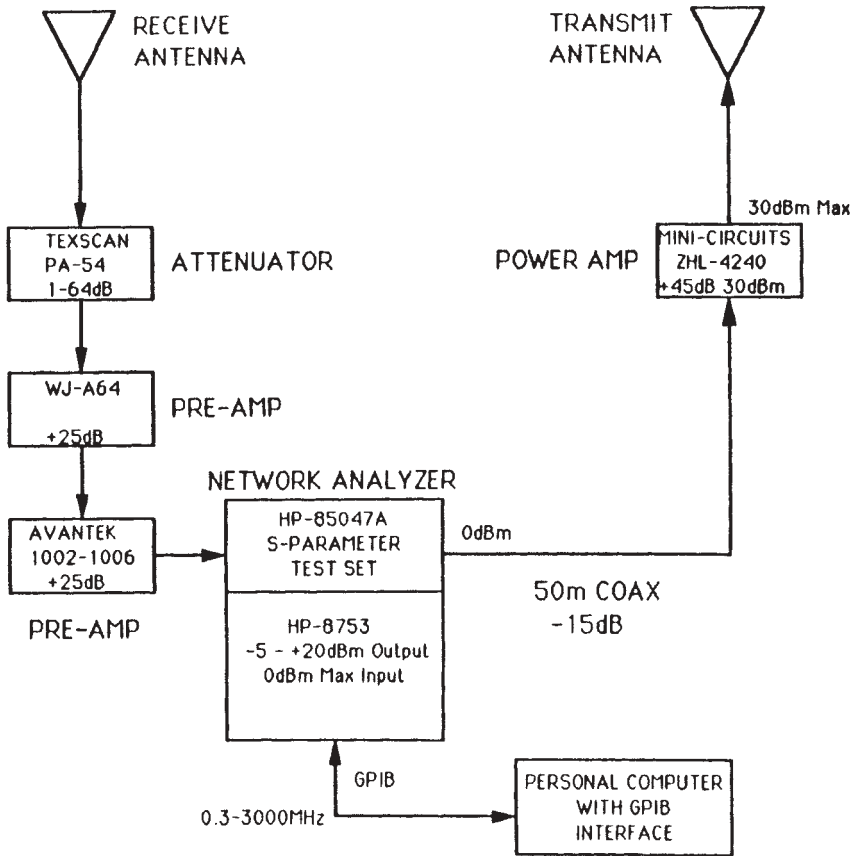


FIGURE 4.8 Measurement system used for narrowband indoor radio propagation measurements.

motion of the human hand and body, which remained almost the same throughout all experiments.

Table 4.5 summarizes the parameter values found in the 11 LOS measurements, which were made in an electronics laboratory. The results are ordered by increasing the distance between transmitter and receiver. To provide a reference, measurement 1 was taken at 1 m with no movements. Three measurements were taken at each of the three distances 3, 6, and 12 m. Measurements 2, 5, and 8 were taken with people moving in the path between the transmitter and receiver. Measurements 3, 6, and 9 were taken with pedestrian traffic close to the transmitter and receiver. Measurements 4, 7, and 10 were taken with small cyclic movement of the transmitter. Measurement 11 was taken at 12-m separation with random motion of the transmitter, to simulate the typical movements of a cordless phone user.

For the OLOS measurements, the receiver was placed in Room 320C of Fig. 4.4. A total of 10 measurements, summarized in Table 4.5, were taken in three different locations of the transmitter. Location 1 was in the corridor next to Room 320c, with one wall separating the transmitter and receiver. Location 2 was in the Room 318 at the other side of the corridor, with two walls separating the transmitter and receiver.

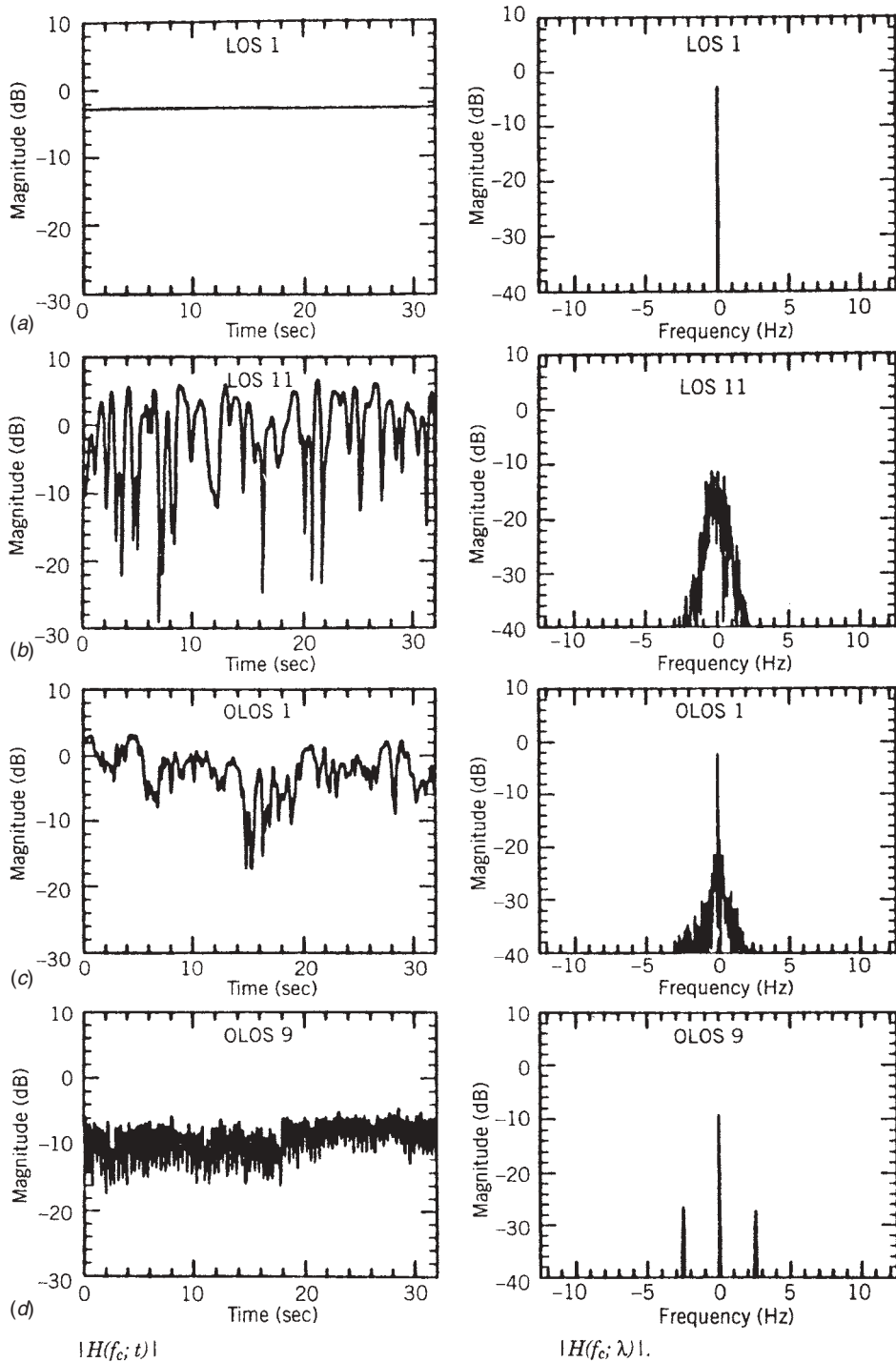


FIGURE 4.9 Sample results of Doppler spectrum measurements on an indoor radio channel: (a) LOS 1: $B_D = 0$, no movement; (b) LOS 11: $P_{\text{dev}} = 35$ dB, $B_D = 4.9$ Hz, random transmitter motion; (c) OLOS 1: $P_{\text{dev}} = 20$ dB, $B_D = 5.7$ Hz, traffic close to transmitter; (d) OLOS 9: $P_{\text{dev}} = 10$ dB, $B_D = 5.2$ Hz, cyclic motion of transmitter [How90a].

TABLE 4.5 Results of Measurement of the Doppler Spread and Its rms Values in Several LOS and OLOS Indoor Areas

Number	LOS				OLOS			
	Distance (m)	$B_{D,rms}$ (Hz)	B_D (Hz)	P_{dev} (dB)	Distance (m)	$B_{D,rms}$ (Hz)	B_D (Hz)	P_{dev} (dB)
1	1	0.016	0.0	0	4	0.373	5.7	20
2	3	0.610	6.1	30	4	0.190	5.1	10
3	3	0.424	4.8	35	4	0.199	4.7	4
4	3	0.092	0.4	4	8	0.873	4.9	30
5	6	0.665	1.9	25	8	0.559	3.6	35
6	6	0.424	3.3	30	8	0.761	4.8	8
7	6	0.236	0.3	3	13	0.461	4.4	25
8	12	0.217	2.0	15	13	0.257	3.0	10
9	12	0.247	3.9	20	13	0.649	5.2	10
10	12	0.130	4.9	4	13	0.288	1.0	8
11	12	0.531	4.9	35				

Source: [How90a].

Location 3 was in Room 317, adjacent to the Room 318, with three walls separating the transmitter and receiver. Measurements 1, 4, and 7 were taken with pedestrian traffic close to the transmitter. Measurements 2, 5, and 8 were taken with traffic close to the receiver. Measurements 3, 6, and 9 were taken with small cyclic movement of the transmitter. Measurement 10 was taken at the third location, with traffic between the transmitter and receiver, but not in the same room as either the transmitter or the receiver. As we discussed earlier, the maximum Doppler frequency shift imparted to an unmodulated carrier is related to the velocity of movement v_m and the wavelength of the carrier λ by $f_M = 0.5B_D = v_m/\lambda$. Therefore, using $v_m = 0.5B_D\lambda$, we can determine the velocity of movements from the measurements of B_D .

Example 4.10: Measurement of Human Body Movements For the frequency of operation of the measurements provided in Fig. 4.9, 910 MHz, the wavelength is $\lambda = 3 \times 10^8 / (910 \times 10^6) = 0.33$ m. Therefore, the velocity of hand for random movements of the antennas in Fig. 4.9b is $v_m = 0.5 \times 4.9 \times 0.33 = 0.8$ m/s (1.8 miles/h). Similarly, we can measure the velocity of the movements of people around the antenna in Fig. 4.9c and cyclic movements of hands in Fig. 4.9c. If we use this equation as an approximation to the Doppler spread, the B_D values are very close, all about 5 Hz. This principle is used in Doppler radars to measure the velocity of vehicles. In this approach Doppler bandwidth depends on the threshold level; therefore, we need to calibrate the system threshold based on the known speed of a vehicle.

To relate the results of these measurements to the classical wide-sense stationary uncorrelated scattering (WSSUS) model described in Section 3.5, we note that our time-domain measurements represent $H(f_c; t)$. The complex autocorrelation function of $H(f_c; t)$ is $R_{Hh}(0; \Delta t)$, and the Doppler power spectrum defined in Eq. (3.5.6) is given by

$$D(\lambda) = R_{HH}(0; \lambda) = \int_{-\infty}^{\infty} R_{Hh}(0; \Delta t) e^{-j2\pi\lambda\Delta t} d(\Delta t) = |H(f_c; \lambda)|^2$$

where

$$H(f_c; \lambda) = \int_{-\infty}^{\infty} H(f_c; t) e^{-j2\pi\lambda t} dt$$

In summary, to determine the Doppler power spectrum, $D(\lambda)$, we transmit a single tone at frequency f_c and we measure the amplitude fluctuations of the received signal in time, $H(f_c; t)$. The magnitude square of the Fourier transform of the measured $H(f_c; t)$ provides us with the $D(\lambda)$. The Doppler spread B_D is the range of frequency λ over which the Doppler power spectrum $D(\lambda) = |H(f_c; \lambda)|^2$ is nonzero. In practice, $|H(f_c; \lambda)|$ is never zero and a threshold is applied to $|H(f_c; \lambda)|$ to determine B_D . The threshold applied for the experiments described in this section is -40 dB [How90a]. A more specific measure of the Doppler spread is the rms Doppler spread, defined by Eq. (3.5.7), which is a weighted measure of the spectral distribution of signal power rather than simply the overall width of the spectrum. The values of rms Doppler spread measured in the LOS and OLOS experiments are also included in Table 4.5.

Other measurements in similar configurations are Bultitude's measurements [Bul87], which include periods of no movement as well as periods of local movements. His analysis of the movement data, which he extracts from the overall sequence of measurements, shows the channel to be wide-sense stationary for time periods of at least 3.4 s. Rappaport [Rap89] measured temporal fading of the received signal envelope over a 100-s period during the normal working hours in a factory. His analysis of the temporal fading data showed the dynamic range to be about 10 dB. Both researchers compared the temporal fading data with the Rayleigh and Rician distributions and showed that the Rician distribution fits the data well.

4.3.3 Fading Rate and Fade Duration

In the mobile environment the received amplitude of the signal fluctuates extensively according to the fast-fading characteristics of the channel. As the received signal strength decreases below the threshold for acceptable performance of a receiver, the signal fades and stays in the fade until the received signal level increases above the fading threshold. The statistics of fading (threshold crossing) rate and the duration of the fade are two important parameters for the designers of wireless networks. Fading rate and duration of fade are functions of average received signal strength, threshold for fading, and the statistics of the fading behavior. Figure 4.10 shows the basic concept and parameters related to the fading rate and duration.

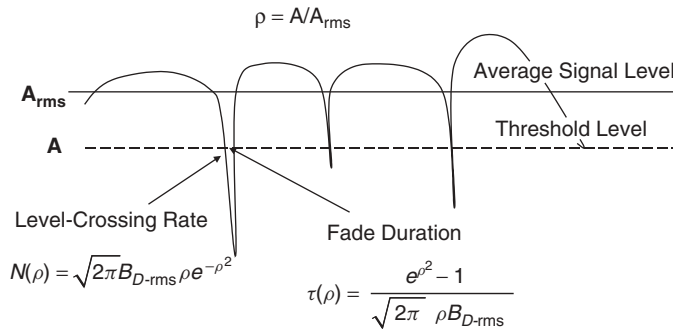


FIGURE 4.10 Level-crossing and fade-duration statistics.

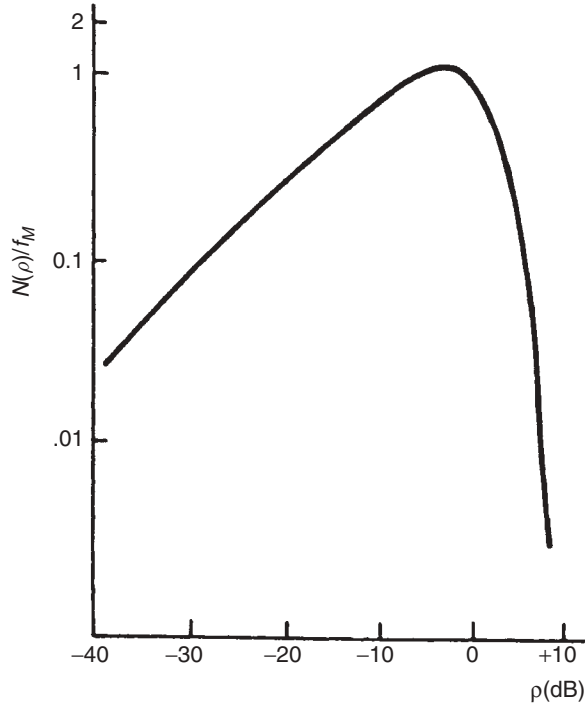


FIGURE 4.11 Normalized level-crossing rate versus normalized threshold for Rayleigh envelope fading.

For a Rayleigh fading envelope distribution, the average number of downward crossings of a level A per second, $N(\rho)$, is given by

$$N(\rho) = \sqrt{2\pi} f_M \rho e^{-\rho^2} \quad (4.3.2)$$

where $\rho = A/A_{\text{rms}}$ is the ratio of the threshold level to the rms amplitude of the fading envelope, and $f_M = 0.5B_D$ is the maximum Doppler spread of the signal defined in Eq. (3.3.1). Figure 4.11 represents the plot of the average number of fades normalized to the Doppler shift, $N(\rho)/f_M$, versus ρ in decibels. For low values of ρ , we hardly have a peak of a deep fade crossing the threshold. As the normalized threshold ρ increases, the number of fades crossing the threshold increases until the threshold gets close to the rms value of the envelope of the signal at which the threshold crossing rate becomes slightly higher than the Doppler spread of the channel. After the peak, when we increase the threshold, the number of zero crossings starts to decrease rapidly because most of the time the signal level is below the threshold, and finally, we reach a point where the entire signal is under the threshold and there are no crossings [Ric48].

Example 4.11: Measurement of Maximum Doppler Shift As we showed in Example 4.10, we can calculate the velocity using measurements of maximum Doppler shift. To find the maximum Doppler shift using this approach we need to take the Fourier transform of the signal, which requires considerable computation, and we need to find the threshold for calculation of the bandwidth experimentally. A simpler

and more accurate alternative for calculation of the maximum Doppler shift is to use Eq. (4.3.2) with the experimental results of level threshold crossing. From Eq. (4.3.2) we have

$$f_M = \frac{N(\rho)}{\sqrt{2\pi} \rho e^{-\rho^2}}$$

In measurements of signal fluctuation using random hand movements, shown in Fig. 4.6b, we observe that the fluctuations are very similar to those observed in mobile radio channels, and the average received signal strength is about 0 dB. If we consider the -10 -dB threshold, the fading signal crosses this threshold 16 times downward during the 30-s measurement time, which means that $N(\rho) = 16/30 = 0.54$. Then the Doppler spectrum is

$$f_M = \frac{N(\rho)}{\sqrt{2\pi} \rho e^{-\rho^2}} = \frac{16/30}{\sqrt{2\pi} \times 0.1 \times e^{-0.01}} = 2.15 \text{ Hz}$$

which can be compared with the $4.9/2 = 2.45$ Hz measured using Fourier transform techniques. If we use this number for calculation of speed, the speed would be

$$v_m = \frac{c}{f_c} f_M = \frac{3 \times 10^8}{910 \times 10^6} \times 2.15 = 0.71 \text{ m/s}$$

compared with the 0.8 m/s calculated in Example 4.10. We cannot apply this technique for calculation of maximum Doppler shift and velocity of movements to Fig. 4.6c or d because fluctuations of the signal are not random enough to form a Rayleigh distribution.

Another important parameter representing the envelope fading characteristics is the duration of fade. The average fade duration for Rayleigh fading and a given threshold ρ is given by [Ric48]

$$\tau(\rho) = \frac{\text{Prob}[\alpha < \rho]}{N(\rho)} = \frac{e^{\rho^2} - 1}{\sqrt{2\pi} \rho f_M} \quad (4.3.3)$$

Figure 4.12 is a plot of the average fade duration normalized to the Doppler shift versus ρ in decibels. At low values of threshold we have very few fades with very short durations. As threshold gets close to its peak values near the rms value of the signal envelope, the duration of the fade increases rapidly until threshold goes above all values of the signal envelope and the fade duration becomes infinity. Using average crossing rate and average fade duration, we can calculate the percentage of time when a received fading signal is above a useful operational threshold.

Example 4.12: Packet Loss Rate Due to Fading Hits A mobile radio installed in a car streams a radio channel through the cellular network infrastructure. The mobile radio operates at 910 MHz and the vehicle drives at 60 miles/hr (26.8 m/s). The maximum Doppler spread for the car is

$$f_M = \frac{v_m}{c} f_c = \frac{26.8}{3 \times 10^8} \times 910 \times 10^6 = 81.3 \text{ Hz}$$

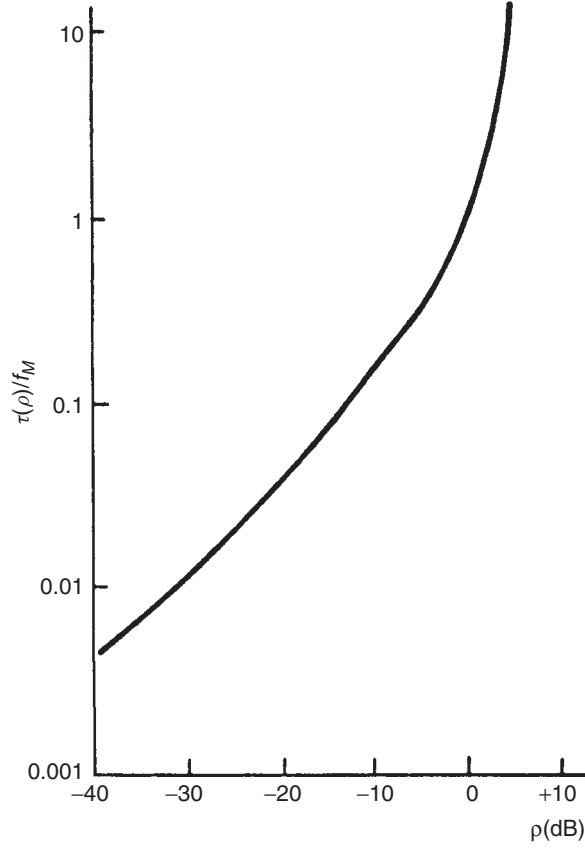


FIGURE 4.12 Normalized average duration of fade versus normalized threshold for Rayleigh envelope fading.

Let's assume that the average power is controlled to be fixed and when the signal level due to multipath fading goes 10 dB below the average received signal level, the received information packets are erroneous and become discarded. Then, on average, the signal goes under the threshold at a rate of

$$N(\rho) = \sqrt{2\pi} f_M \rho e^{-\rho^2} = \sqrt{2\pi} \times 81.3 \times 0.1 e^{-0.01} = 21.18$$

and each time signal goes under the threshold, it takes

$$\tau(\rho) = \frac{e^{\rho^2} - 1}{\rho B_D \sqrt{2\pi}} = 49.3 \text{ ms}$$

On the average, this system goes 21.18 times per second into fade, and each time it stays in the fade for 49.3 ms. Therefore, the percentage of time that the system can send information is

$$S = 1 - 21.18 \times (49.3 \times 10^{-3}) = 89.6\%$$

In packet communications this percentage of useful time for communication is referred to as the *utility* or *throughput* of the channel. In the streaming application this is the percentage of packets that survive channel fading. The effective throughput of this data network can be approximated by the percentage of time that system is above the fade threshold.

4.4 STATISTICAL BEHAVIOR OF FAST ENVELOPE FADING

The statistics for slow shadow fading and distance–power gradient are used for coverage studies. The statistics for fast envelope fading are needed for calculation of the average error rate for different transmission techniques over a fading wireless channel. As we show in Chapter 8, different transmission techniques have different power requirements to support a specific bit error rate required by an application. To calculate the error rate for a given transmitted power we need to consider the error-rate behavior over the statistics for fast envelope fading of the wireless channel. Therefore, we need a statistical model for the behavior of amplitude fading. Models for envelope fading of the channel are used for analytical and simulation-based comparative performance evaluation of the wireless modems. In selecting a model for the behavior of the channel, we need to make a compromise between the accuracy of the model to fit the empirical data and the complexity of the model to be used in mathematical derivations and computer simulations for calculation of average error rates. Analytical calculations of the performance of the modems use the linear statistics of the fast-fading behavior, while the empirical data are collected from measurements of the received signal strength in decibels. In this section we introduce the linear form of the density function of popular models for the characterization of fast fading, and then we discuss validation using empirical data.

4.4.1 Distribution Functions for Envelope Fading

The most popular distribution functions used for fitting to envelope fading data are the Rayleigh, Rician, lognormal, Suzuki, Weibull, and Nakagami distributions [Has93a]. The *Rayleigh distribution* is the most popular distribution function used for statistical modeling of envelope fading of radio signals. The Rayleigh probability density function for the amplitude a is given by

$$f_A(a) = \frac{a}{\Gamma} e^{-a^2/2\Gamma}, \quad a \geq 0$$

which is described by a single parameter, Γ . The mean and variance of the Rayleigh-distributed random variable are given by $\sqrt{\Gamma\pi/2}$ and $(2 - \pi/2)\Gamma$, respectively.

The square of the magnitude of a Rayleigh-distributed random variable, representing the signal power, has an exponential distribution, the chi-square distribution with two degrees of freedom. If we let $\gamma = a^2$, the probability density function of γ is

$$f_\Gamma(\gamma) = \frac{1}{\bar{\gamma}} e^{-\gamma/\bar{\gamma}}, \quad \gamma \geq 0$$

where $\bar{\gamma}$ is the average received power. We use this distribution in several instances in this book for the calculation of error rates of modulation techniques and throughput

of contention-based protocols operating over Rayleigh fading channels. The Rayleigh distribution is also used for calculation of fading characteristics such as the fading rate and average fading duration, as given by Eqs. (4.5.2) and (4.5.3).

The *Rician distribution*, commonly used to model amplitude variation in the presence of a strong dominant LOS path, is described by

$$f_A(a) = \frac{a}{\sigma^2} e^{-(a^2 - v^2)/2\sigma^2} I_0\left(\frac{av}{\sigma^2}\right), \quad v, a \geq 0$$

where σ^2 represents the variance of the random component, v is the amplitude of the fixed component, and $I_0(\cdot)$ ³ is a modified Bessel function of the first kind. The parameter $k = v^2/\sigma^2$ is the ratio of the deterministic to the random component of the process. Usually, k and σ are used as the parameters identifying the Rician distribution function.⁴ Values of k of about 6 dB are typical in modeling indoor radio channel amplitude fluctuations [Bul87].

The *lognormal probability distribution function* is used to model large-scale variations in the received power in indoor and urban radio channels. The model suggests that the decibel value of the average received power over a large area forms a normal (Gaussian) distribution function. The probability density function of a lognormal distributed random variable is given by

$$f_A(a) = \frac{1}{\sqrt{2\pi} \sigma a} e^{-(\ln a - \mu)^2/2\sigma^2}, \quad a \geq 0$$

where μ and σ are the mean and standard deviation of the random variable, respectively. With this distribution for the random variable, α ensures that the $\log_{10} \alpha$, and consequently the decibel value of α has a normal distribution. In indoor and urban radio applications, the mean of the lognormal random variable is assumed to be zero and the variance is the only parameter needed to describe the distribution function. In practice, as in the calculation of fading margin in Section 4.2.1 or in specification in standards recommendations of the statistics for shadow fading, we always refer to the decibel version of the random variable with Gaussian distribution.

In portable and mobile radio channels, the local distribution of the signal amplitude in areas with dimensions on the order of the wavelength is Rayleigh, and the wide-area coverage is represented by a lognormal distribution. The overall distribution of the received signal amplitude is then represented by the integral of the Rayleigh distribution over all possible values of σ represented by the lognormal distribution. This new distribution, suggested by Suzuki, is named for him [Suz77]. The *Suzuki random variable* is defined by the following probability density function:

$$f_A(a) = \int_0^\infty \frac{a}{\sigma^2} e^{-a^2/2\sigma^2} \frac{1}{\sqrt{2\pi} \sigma \lambda} e^{-(\ln \sigma - \mu)^2/2\lambda^2} d\sigma, \quad a \geq 0$$

³ $I_0(x) = (1/2\pi) \int_0^{2\pi} e^{x \cos y} dy$.

⁴For $k \gg 1$, the Rician distribution become very similar to Gaussian distribution, with the same mean and variance.

where λ^2 is the variance of the lognormal distribution. This distribution has a very clear physical interpretation, but the complicated mathematical form has limited its practical applications.

Two other distributions used for modeling of envelope fading in indoor and urban radio channels are those of Weibull and Nakagami [Nak60, Lor79]. These distributions form a superset of other distribution functions. The *Weibull probability density function* is given by

$$f_A(a) = \frac{sr}{\sigma} \left(\frac{ra}{\sigma} \right)^{s-1} e^{-(ra/\sigma)^s}, \quad a \geq 0$$

where s is the shaping parameter, σ is the rms value of the random variable, and $r = [(2/s)\Gamma(2/s)]^{1/2}$ is a normalization factor [Has93a] based on the gamma function. For $s = 2$ the Weibull distribution function reduces to the Rayleigh, and for $s = 1$ it reduces to the exponential distribution.

The *Nakagami distribution* is defined as

$$f_A(a) = \frac{2m^m a^{2m-1}}{\Gamma(m)\Omega^m} e^{-(ma^2/\Omega)}, \quad a \geq 0$$

where Ω is the mean-squared value of the random variable and $m = \Omega^2/\text{Var}[a^2]$, which is constrained to be equal to or larger than $\frac{1}{2}$. For $m = 1$ the Nakagami distribution reduces to Rayleigh, and for $m = \frac{1}{2}$ it is a one-sided Gaussian distribution. With proper adjustment of the parameters it can also fit Rician and lognormal distributions very tightly.

In calculating the average probability of error for digital signaling over a fading channel, the probability distribution of the fading envelope is used to average the conditional probability of error at each envelope level [Ale82, Pro89]. In Chapter 8 we provide a number of useful analytical results related to the performance analysis of modulation techniques over conventional Rayleigh fading channels. Samples of performance evaluation using other distributions may be found in [Liu92, Fed94, Zha96, Alo01]. Educational aspects of simulating these channels using MatLAB is available in [Pra02].

4.4.2 Measurement of Envelope Fading Statistics

A systematic approach to determining the distribution function of received amplitudes is to compare the results of amplitude measurements with a few candidate distributions. Each candidate distribution is represented by a function with a few parameters. The functions are selected to have relevance to the physical environment, and the parameters are determined from the measured data. The cumulative distribution function (CDF) of the resulting curves with the parameters obtained from the measurements is compared with the CDF of the empirical data to find the best fit to the empirical data.

The probability density functions are defined for linear values of the amplitudes, while the measurements are usually documented on a decibel scale. Transferring the logarithmically measured data to a linear scale and performing curve fitting with linear data will reduce the accuracy of the curve-fitting operation. Therefore, for greater accuracy, we should find the CDF in decibels, which involves an algebraically tedious

change of variable, an operation not analytically tractable for all distribution functions. The CDFs of Rayleigh, Weibull, Nakagami, lognormal, and Suzuki distribution functions in decibel scale and the relationship between the parameters of the function and the mean and variance of the empirical data are available in [Lor79, Gan91a]. An analytical solution for the Rician distribution is not available. Therefore, curve fitting with Rician distribution involves transferring the empirical data to a linear scale, which reduces the reliability of the conclusions drawn from the curve-fitting operation. We conclude this section with an example of fitting envelope fading models to empirical data collected in an indoor area.

Example 4.13: Measurement of Envelope Fading in an Indoor Area In this example we describe the results of experimental curve fitting for an OLOS narrowband indoor radio measurement experiment that was part of the more comprehensive study reported in [How90c]. We compare the CDF of 801 samples of received signal envelope $|H(f_c; t)|$ with the lognormal, Weibull, and Rayleigh distributions. Figure 4.13 shows the CDF of measurement and the theoretical CDFs for lognormal, Weibull, and Rayleigh distributions. The Rayleigh distribution is a poor fit, but the Weibull distribution gives a very close fit to the measured data. The lognormal distribution is a much better fit than the Rayleigh and is a marginal fit compared with the Weibull distribution. As we mentioned before, the probability distribution of the fading envelope is used to average the conditional probability of error at each envelope level. It is the near-zero values of the signal envelope, where the conditional probability of error approaches

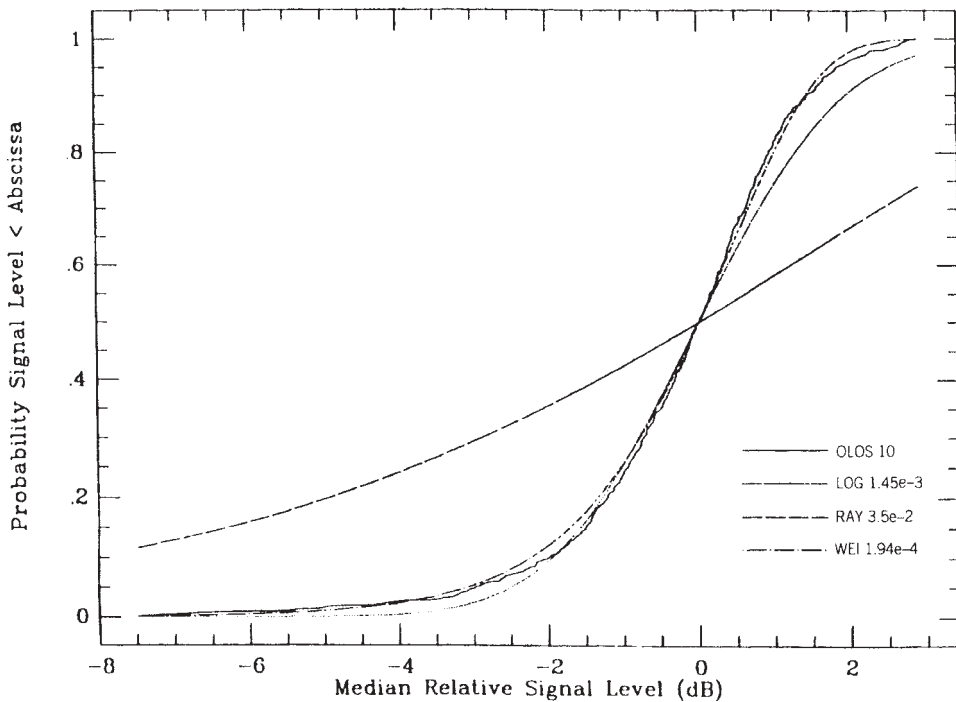


FIGURE 4.13 Sample result of curve fitting for the evaluation of envelope fading statistics.

0.5, that most heavily influence the average probability of error and therefore highlight the region where the assumed theoretical distribution should best fit the measurements. If we consider this argument, for the lower tails of the curves in Fig. 4.13, both the lognormal and Weibull distributions provide close fits to the empirical data. Therefore, the average error rate calculated from the two distributions should provide very close results, and the one that is more convenient in calculations is the best choice for modeling.

4.5 SIMULATION OF FAST ENVELOPE FADING

Simulation of envelope fading is very important for design and performance evaluation of wireless modems because often we cannot find closed-form solutions to compare performance of various modulation and coding techniques over wireless channels. Once we know how to simulate the channel for a narrowband signal, as we will see in Chapter 6, simulation for the wideband signal is only an extension of that.⁵ To simulate a narrowband channel, we need to generate a random process with a specific envelope fading density function and a specific Doppler spectrum. The channel simulation software in these cases should generate the random variables with the distribution function of the envelope fading and shape the Doppler spectrum of the signal using a signal-processing technique. The complexity of the simulation depends on the simulation platform. All computer programming languages used for the development of computer simulations for telecommunication applications, such as C, at least have uniform and Gaussian random number generators. More modern scientific software tools such as MatLAB provide more random variables and filtering functions that further simplify simulations of the channel. If the simulation platform has limited subroutines for generating random variables, we can generate a new random variable from an old random variable using an appropriate mapping with specific rules. Methods for computer simulation of Rayleigh, Rician, lognormal, Suzuki, Weibull, and Nakagami random variables are described in [Pre91, Jer92, Med93].

After generating a random variable with the distribution function of envelope fading, passing the random variable through a filter with a specific spectral shape, resembling the Doppler spectrum of the channel, can do the spectral shaping. If it is inconvenient to develop the needed spectrum by filtering, one may, instead, generate a series of oscillators with different frequencies and add the outputs to form the specific spectrum. The first approach has been used extensively in simulation of a variety of fading channels. The second approach is often used in simulation of mobile radio channels, based on the Clarke assumption of isotropic scattering [Cla68]. We discuss these two approaches for simulation of the envelope fading characteristics in the rest of this section. The envelope fading considered for these simulations is primarily Rayleigh fading, which can be modified to Rician by adding a constant value to the random variable. The Doppler spectrum of specific interest is the double-ear spectrum of Clarke's model for mobile radio channels, shown in Fig. 4.7, and the flat spectrum used in JTC model to model the channel Doppler spectrum in indoor areas.

⁵A wideband simulator is a group of narrowband simulators with different gains connected together through a tapped delay line; for examples, see JTC models in Appendix 6A.

4.5.1 Filtered Gaussian Noise for Simulation of a Mobile Radio Channel

A widely used approach to simulation of fading radio channels is to construct a fading signal from in-phase and quadrature Gaussian noise sources. Because the envelope of a complex Gaussian noise process has a Rayleigh probability density function (PDF), the output of such a simulator will simulate Rayleigh fading accurately. In this approach, applying the appropriate filtering to the Gaussian noise sources provides the Doppler spectrum of the channel of interest. This technique, originally designed for analog hardware simulation of the RF channel, is very popular in digital software and hardware simulations of the envelope fading. Figure 4.14 shows a block diagram of the basic technique for simulating Rayleigh fading as an RF signal using two filtered Gaussian noise processes. In some applications of this technique, a detailed specification of the channel Doppler spectrum is not available.

Example 4.14: Hardware Envelope Fading Simulator Using Filters Arredondo et al. [Arr73] describe a device designed to simulate Rayleigh fading characteristics of mobile radio channels. The hardware simulator used two Gaussian noise sources with identical shaping filters to generate the quadrature components of a Rayleigh

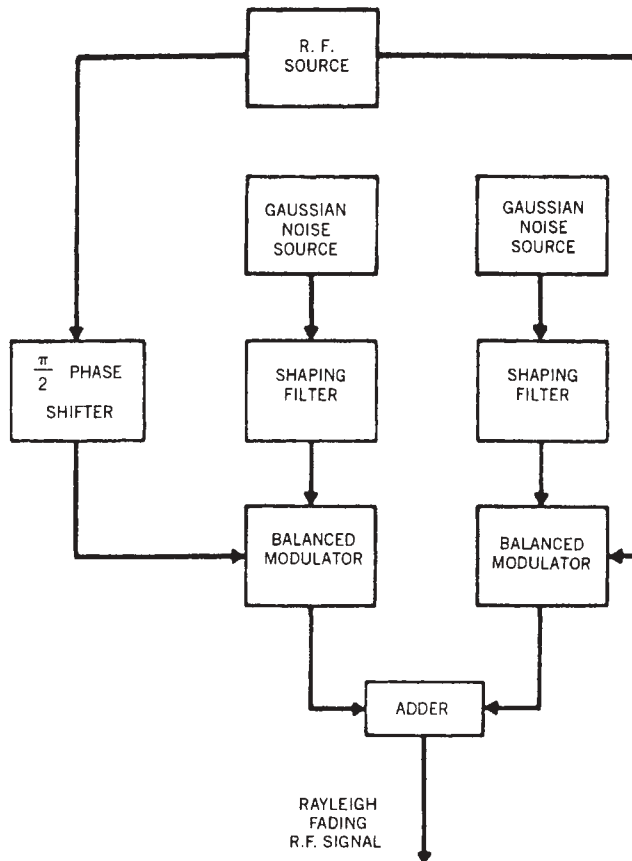


FIGURE 4.14 Block diagram of a filtered Gaussian noise-fading simulator.

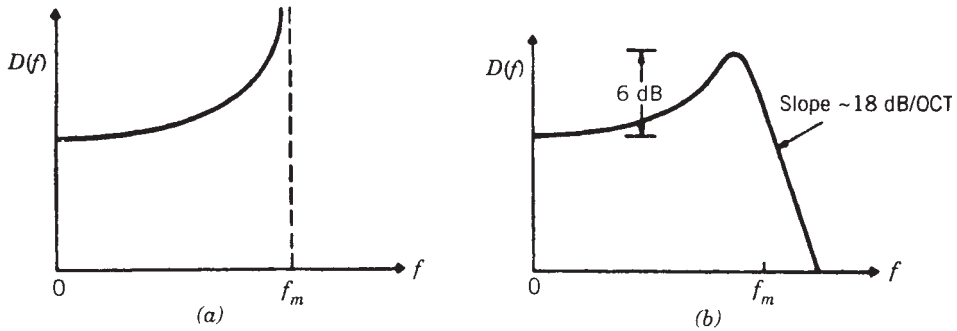


FIGURE 4.15 Theoretical mobile radio spectrum and a simulator shaping filter: (a) theoretical spectral density; (b) shaping filter frequency response. (From [Arr73] © IEEE.)

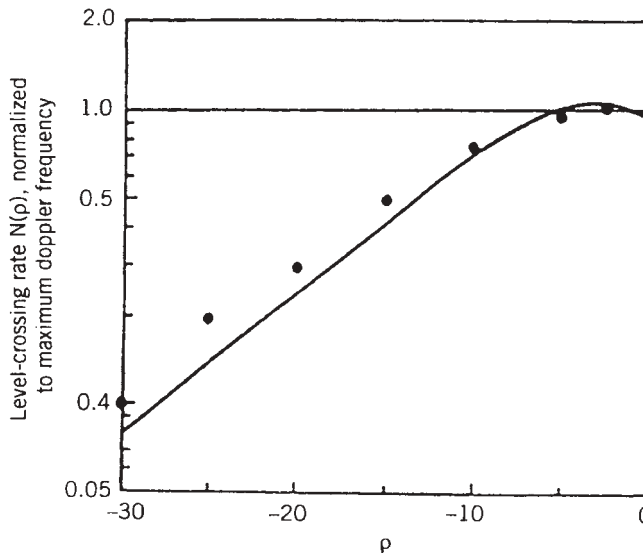


FIGURE 4.16 Simulated and theoretical level-crossing rates. The level-crossing rate is normalized to $f_m = 1$ Hz. (From [Arr73] © IEEE.)

fading signal. The shaping filters were designed to approximate the theoretical spectrum of the mobile radio channel. The theoretical spectrum, given in Eq. (4.3.1), is shown here in Fig. 4.15a. The frequency response of each shaping filter is shown in Fig. 4.15b. The shaping filter consisted of two active filters in cascade: a low-pass filter and a peaking amplifier. The filter was designed to the desired fading rate by equating the second moments of the theoretical and simulated spectra. The degree to which the level-crossing rate compared with theory is shown in Fig. 4.16. In the figure, the measured level-crossing rate $N(\rho)$ in reciprocal seconds (s^{-1}) (normalized to $B_{D \text{ rms}} = 1$ Hz) is plotted against the crossing level ρ in decibels relative to the rms envelope level. The theoretical curve in the figure is given by Eq. (4.5.2). It can be seen from Fig. 4.16 that the measured level-crossing rate agrees with theory to within 3 dB over a range extending down to 30 dB below rms.

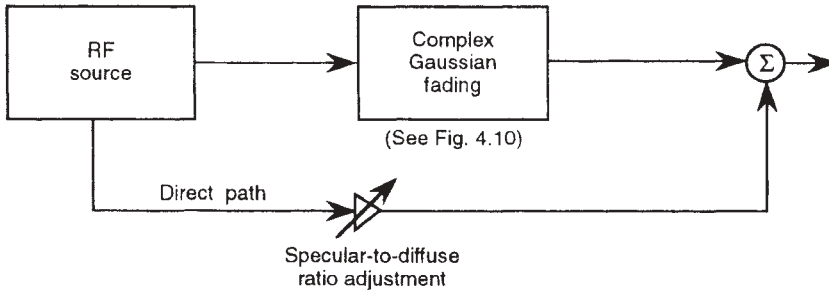


FIGURE 4.17 Simulation of a combined specular and fading channel.

For some applications, the radio channel is characterized by a combination of a fading signal component and one or more nonfading or *specular* components. Figure 4.17 shows a simulation model incorporating a fading component and a single specular component. Such a model is appropriate for simulation of cases such as LOS microwave channels, where there is a nonfading signal arriving on a direct path as well as a fading signal produced, for example, by atmospheric effects. The inclusion of a specular component in this simulation model is equivalent to adding a nonzero mean value to each of the quadrature Gaussian noise sources, and therefore the simulator produces Rician rather than Rayleigh fading. Several other variations of these basic filtered-noise simulation models are discussed in [Jer92]. This technique was also recommended by the JTC standardization committee for the simulation of channel fading [JTC94].

4.5.2 The Clarke–Jakes Model for Simulation of a Mobile Radio Channel

As an alternative to RF modeling with filtered complex Gaussian noise, one may instead approximate the Rayleigh fading process by summing a set of complex sinusoids. The number of sinusoids in the set must be sufficiently large that the power density function of the resulting envelope provides an acceptably accurate approximation to the Rayleigh PDF. With this modeling method, the sinusoids are weighted so as to produce an accurate approximation of the desired channel Doppler spectrum. One technique of this type is that proposed by William Jakes of Bell Laboratories for the simulation of fading mobile radio channels [Jak74]. This simulation technique, based on the isotropic scattering model studied earlier by Clarke [Cla68], has come to be known as the *Clarke–Jakes model* and is used widely in the mobile communications industry. The technique was originally developed for a hardware simulator implementation, but it is often implemented in software as well. Software realizations of the Clarke–Jakes model have been adopted by standards groups for use in testing candidate speech-coding schemes as well as radio-link error-control protocols [Ses91, Lev93].

Jakes [Jak74] shows that the theoretical Doppler spectrum for the isotropic scattering mobile radio channel, given in Eq. (4.3.1), can be well approximated by a summation of a relatively small number of sinusoids, with the frequencies and relative phases of the sinusoids set according to a specific formulation. Following our notation in Section 3.3, the maximum Doppler shift frequency is $f_m = v_m/\lambda$, where v_m is the velocity of the mobile and λ is the wavelength of the carrier frequency. In the model described by

Jakes, the ideal isotropic continuum of arriving scatter components is approximated by N plane waves arriving at uniformly spaced azimuthal angles. The model restricts $N/2$ to be an odd integer and defines another integer $N_0 = \frac{1}{2}(N/2 - 1)$. This leads to a simulation model having one complex frequency oscillator with frequency $\omega_m = 2\pi f_m$ plus a summation of N_0 complex lower-frequency oscillators with frequencies equal to the Doppler shifts $\omega_m \cos \theta_n$, where θ_n is the arrival angle for the n th plane wave (see Fig. 4.6) and where $n = 1, 2, \dots, N_0$. Each oscillator has an initial phase, and these phases are to be chosen as part of initializing the simulation. We can express the complex envelope $T(t)$ of the fading signal in the form

$$T(t) = \frac{E_0}{\sqrt{2N_0 + 1}}(x_c + jx_s)$$

where

$$x_c(t) = 2 \sum_{n=1}^{N_0} \cos \phi_n \cos \omega_n t + \sqrt{2} \cos \phi_N \cos \omega_m t$$

$$x_s(t) = 2 \sum_{n=1}^{N_0} \sin \phi_n \cos \omega_n t + \sqrt{2} \sin \phi_N \cos \omega_m t$$

and where $\omega_n = \omega_m \cos(2\pi n/N)$, $n = 1, 2, \dots, N_0$. In the equations above, ϕ_N is the initial phase of the maximum Doppler frequency sinusoid, and ϕ_n is the initial phase of the n th Doppler-shifted sinusoid. The quantities x_c and x_s are the in-phase and quadrature components, respectively, of the model output. Note that the amplitudes of all the components are made equal to unity except for the one at maximum Doppler frequency ω_m , which is set to $1/\sqrt{2}$.

In a hardware implementation of the simulator intended to operate with RF equipment, the outputs of the individual oscillators, with the appropriate gain factors, are first summed to produce x_c and x_s , which are then multiplied by in-phase and quadrature signal carrier components, respectively, and then summed to produce the final output signal, as acted upon by fading. In a software realization of the model, one would apply x_c and x_s to the in-phase and quadrature components of a baseband signal representation. In a software simulation, one might generate the trigonometric functions using look-up tables [Cas90].

In using this simulation method, one must choose the initial phases (ϕ_n and ϕ_N) of the Doppler-shifted components in such a way that the phase of the resulting fading process will exhibit a distribution as close as possible to uniform. This is discussed in some detail in [Jak74], where rules are given for initializing the phases of the sinusoids. The number of Doppler-shifted sinusoids is chosen large enough that $T(t)$ provides a good approximation to a complex Gaussian process (via central limit theorem), and therefore the envelope $|T(t)|$ is approximately Rayleigh. Jakes suggests that $N_0 = 8$ provides an acceptably accurate approximation to the ideal case of Rayleigh fading.

Example 4.15: Software Envelope-Fading Simulation Using the Clarke–Jakes Model

Consider communication at a carrier frequency of 900 MHz (cellular band) and a vehicle closing velocity of 100 km/h (27.8 m/s). For these conditions the maximum Doppler frequency is $f_m = v_m/\lambda = 83.3$ Hz. Therefore, in simulation with the Clarke–Jakes

model, the highest-frequency sinusoid has frequency $f_m = 83.3$ Hz, and the frequencies of the N_0 remaining sinusoids are $83.3 \cos(2\pi n/N)$, $n = 1, 2, \dots, N_0$. In [Jak74], Jakes suggested two methods for setting the initial phases of the f_m sinusoid and the N_0 lower-frequency sinusoids. Here we use case 2, $\phi_N = 0$ and $\phi_n = \pi n/(N_0 + 1)$, where $n = 1, 2, \dots, N_0$. Figure 4.18 shows samples of the output of the fading process produced with a Clarke–Jakes model simulation using $N_0 = 8$. In that simulation the carrier frequency was 900 MHz, and a vehicle speed of 100 km/h was assumed. Figure 4.18a shows the measured spectrum, Fig. 4.18b shows a histogram of samples of the output envelope, and Fig. 4.18c shows the measured histogram of phases of the fading signal output. Using this simulation, the bit-error performance for $\pi/4$ -QDPSK modulation was evaluated and was found to agree with theoretical performance in flat Rayleigh fading to within about 0.3 dB at BER = 10^{-3} [Lev93].

Filtered Gaussian noise and Clarke–Jakes model simulations of envelope fading using MatLAB are given as a project at the end of the chapter.

4.5.3 Envelope-Fading Simulation for a Flat Spectrum in Indoor Areas

Another model often used to simulate fading channels is a very simple model called the *flat spectrum fading model*. The model is based on an assumption of scatterers having a uniform distribution in three dimensions. As the name implies, the Doppler spectrum defining this model is flat over a range of Doppler shifts symmetric about the carrier frequency:

$$D(f) = \frac{1}{2\pi f_m}, \quad |f| \leq f_m$$

where f_m is the maximum Doppler frequency. This model is often used in applications where the multipath fading results from random movements of scattering elements in the area of the communication path, from random movements of the transmitter or receiver or from both causes. A good example of this type of application would be a WLAN operating in an office environment, or on a factory floor, where there is a good deal of random pedestrian traffic. In such an application the maximum Doppler shift f_m would be set in accordance with an estimate of maximum pedestrian walking speed. For example, for the RF channel model being proposed for use in design of the air-interface specification for PCS (services to operate at 2 GHz), it is recommended that most of the indoor pedestrian communication environments be modeled with the flat-spectrum model, using a maximum Doppler frequency of 9.6 Hz [JTC94]. Channel modeling for PCS comes under the category of wideband channel characterization, which we discuss in Chapter 6.

As with the mobile radio channel model, this model can be implemented using either the filtered complex Gaussian noise method or the sum-of-sinusoids method. With the Gaussian noise method, the low-pass shaping filter would be chosen to have a relatively flat amplitude function and an rms bandwidth approximately equal to f_m . With the sum-of-sinusoids method, the complex envelope of the fading signal is simulated as a summation of uniformly spaced sinusoids, with the maximum frequency set equal to f_m . The number of sinusoids is chosen to provide an acceptably accurate approximation to Rayleigh fading, and the initial phases of the sinusoids are chosen to provide an approximately uniform distribution of the fading signal phase over $(0, 2\pi)$.

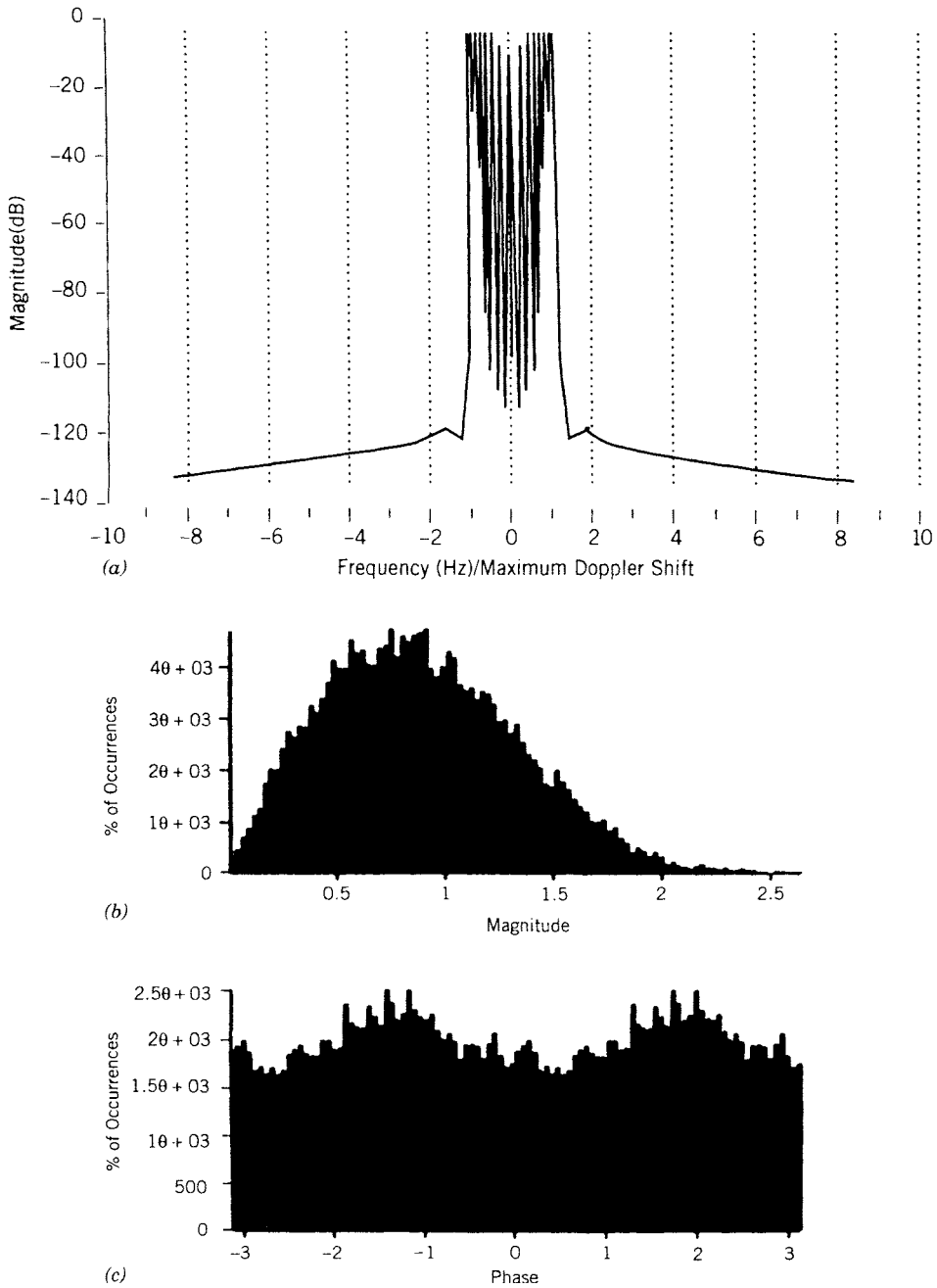


FIGURE 4.18 Example of cellular mobile radio channel simulation using the Clarke–Jakes model. $V = 100$ km/h, $f_c = 900$ MHz, $N_0 = 8$. (a) Measured power spectral density; (b) envelope histogram; (c) phase histogram.

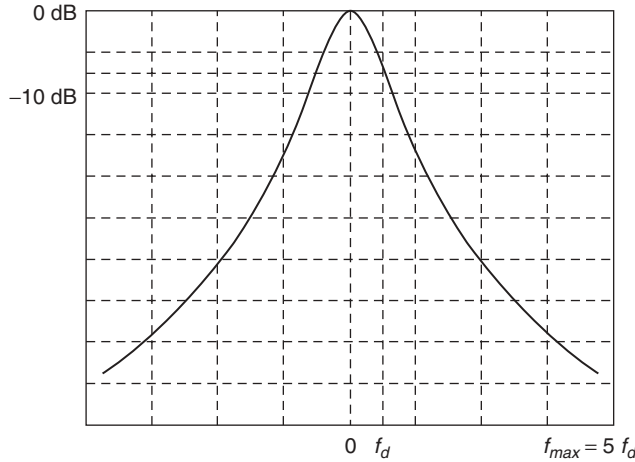


FIGURE 4.19 Bell-shaped spectrum for the Doppler spectra, recommended by the IEEE 802.11 [Erc03].

Another model for Doppler spectrum of envelope fading in indoor areas, recommended by the IEEE 802.11 community [Erc03], is the *bell-shaped spectrum*,

$$D(f) = \frac{1}{1 + Af^2}$$

in which A is a constant selected so that at a given frequency f_d , $(D(f))|_{f=f_d} = 0.1$ and can be calculated from $A = 9/f_d^2$. The shape of the spectrum and definition of f_d as the frequency at which spectrum is dropped 10% of its peak value is shown in Fig. 4.19. This spectral shape is consistent with Doppler spectrum measurements of Fig. 4.9b for random movements in an indoor area. Another parameter in this model is $f_{\max}$, the maximum frequency component of the Doppler spectrum, which limits the range of frequencies to an upper bound and is set arbitrarily to $f_{\max} = 5f_d$. The values for f_d , determined experimentally by the standardization committee in indoor environments, were found to be up to 6 Hz at 5.25 GHz center frequency and up to 3 Hz at 2.4 GHz center frequency. The difference between this model and the JTC model for the indoor areas is that the IEEE 802.11 is for WLAN applications in which transmitter and receiver are stationary and people are moving in between, while JTC model is for PCS cellular phone applications where the user terminal is often moving through the environment.

QUESTIONS

- How do we measure the distance–power gradient on a radio link?
- How do we measure the Doppler spread?
- What does the lognormal element of a path-loss model represent?
- If the transmitter and the receiver are fixed, can the channel have any Doppler spread? Explain.

- (e) Explain the difference between shadow fading and multipath fading. What is the typical envelope distribution associated with each of the two forms of fading?
- (f) What is fading margin, and how does it affect calculation of coverage?
- (g) What is the breakpoint distance for the Fresnel zone, and how do radio waves propagate in this zone?
- (h) What is a typical value of power loss observed when a mobile turns a corner and loses the LOS connection with the base station?
- (i) What are the typical values of the measured distance–power gradient in indoor areas as reported in this chapter, and how do they compare with the JTC recommendations?
- (j) Discuss the difference between the radio propagation characteristics for 800- to 900-MHz cellular and 1800- to 1900-MHz PCS bands.
- (k) What are the typical values for Doppler spread in indoor areas?
- (l) What is the maximum Doppler shift due to measurement of the local short-term variations, as reported in this chapter? What type of movement has caused this maximum fluctuation?
- (m) What are the shapes of the measured Doppler spectra in indoor areas as reported in this chapter, and how do they compare with the JTC recommendation?
- (n) What are typical values of Doppler spread for mobile radio applications?
- (o) What are the fading rate and fading duration?
- (p) Describe the difference between the shapes of the Doppler spectra for indoor and outdoor areas.
- (q) What methods are used for simulation of the narrowband signal fluctuations in mobile radio channels?
- (r) Describe the difference between the Rayleigh and Rician Doppler spectra in mobile radio channels.
- (s) What distribution functions are typically used for modeling the amplitude fluctuations in portable and mobile radio applications?
- (t) Which distributions are represented by the Suzuki distribution?

PROBLEMS

1. Most mobile data applications are expected to operate inside buildings. Consider a mobile data network in which the minimum required received SNR for proper operation is 10 dB, the background noise level in the band is -120 dBm, and the in-building penetration loss is 15 dB. If the transmitter and receiver antenna gains are 2, the frequency of operation is 910 MHz, the height of the base station and mobile station antennas are 100 and 1.4 m, respectively, and the maximum transmitted power is 10 W, determine the coverage of base station using the following:
 - (a) The free-space propagation equation given by Eq. (3.2.1).

- (b) The simple two-path model analyzed in Example 3.2.
 - (c) The Okumura–Hata model for a medium-sized city.
 - (d) The Okumura–Hata model for a large city.
 - (e) How much difference exists among various approaches, and how can it be explained?
2. Consider the floor plan of Fig. 4.4. We want to predict path loss using different path-loss models. Assume that the transmitter and the receiver are located in the center of Rooms 311 and 317, respectively, the transmitter and receiver antenna gains are 1.6, and the frequency of operation is 2.4 GHz. Calculate the path loss using the following:
- (a) The experimental partitioned model based on this floor plan (with three partition gradients of 1.76, 2.05, and 4.21), which was described in Example 4.7.
 - (b) Akerberg’s generalized partitioned model for the indoor propagation loss described at the end of Section 4.2.3.
 - (c) The JTC model described in Example 4.5, with the building classified as an office area.
3. A 100-mW transmitter operates at 1.9 GHz with a receiver having a sensitivity of -90 dBm. We want to determine the coverage in various environments using the JTC model and without any fade margin.
- (a) What is the coverage in indoor residential, office, and commercial environments?
 - (b) What is the coverage in an outdoor microcell with a distance of 10 m between the transmitter and the corner where the LOS connection is lost? Assume that the height of the transmitter and the receiver antennas are 12 and 2 m, respectively.
 - (c) What is the coverage in a microcellular environment without a detailed description of the environment?
 - (d) For your calculations in part (a), what is the probability of having an acceptable signal level at the maximum calculated distance from the transmitter? What are the needed fading margin for part (a) if we want to increase the probability of having an acceptable signal level at the edge of coverage to 90%? Repeat part (a) when you include the fading margin for 90% coverage into your calculation.
4. A 10-W transmitter operates with a receiver having a sensitivity of -90 dBm. We want to determine the coverage in a macrocellular environment.
- (a) What is the coverage if we use the macrocellular JTC transmission-loss model with transmitter and receiver antenna heights of 100 and 2 m, respectively? Use all four macrocellular environments shown in Table 4.4.
 - (b) Compare the results of part (a) with the predicted value from Okumura’s model within a medium-sized city and a center frequency of 1.5 GHz.
5. The Doppler power spectrum $D(\lambda)$ of the indoor radio channel is often assumed to be uniformly distributed with a maximum Doppler spread of 10 Hz.
- (a) Determine the average number of fades per second and the average fade duration of a Rayleigh fading channel for which the fading threshold is 10 dB below the average rms value of the signal.

- (b) A digital communication system operating in the environment described in part (a) loses all its data when the signal goes under the threshold (i.e., with a signal under the threshold, the probability of error is 50%) and has no error when the signal is above the threshold (i.e., the probability of error is zero). What is the overall average probability of error for this system?
 - (c) Repeat parts (a) and (b) for the threshold levels of 3 and 20 dB below average rms, respectively. Discuss the relation between the threshold and the average error rate of the system.
6. (a) The original IEEE 802.11 standard at 2.4 GHz specifies a maximum transmitted power of 100 mW and a minimum receiver sensitivity of -80 dBm. Calculate the fading margin for an outage rate of 10% and a standard deviation of 10 dB for the shadow fading.
- (b) Determine the coverage of the system on one floor of a large residential apartment building so that the terminals at the edge of coverage have acceptable signal 90% of the time. Use the single-gradient model of Eq. (4.2.3) with $\alpha = 3$.
 - (c) Repeat part (b) using the JTC model. To apply the JTC model for IEEE 802.11 devices, you need to adjust the path loss in the first meter so that it confirms with operation at 2.4 GHz rather than 1900 MHz.
 - (d) If you have a 15×15 -m apartment and you install an access point in the center of the apartment, how close should one get to your apartment to connect to your access point from outside?
7. (a) Show that a Rayleigh-distributed random variable, β , can be generated from two independent Gaussian-distributed random variables x and y from the relation $\beta = \sqrt{x^2 + y^2}$.
- (b) Simulate 100 samples of a Rayleigh-distributed random variable with variance 1 using MatLAB or an alternative computational tool. Create the probability density function (PDF) and the cumulative density function (CDF) of the 100 simulated samples, and compare the results with the theoretical PDF and CDF of the Rayleigh distribution.
8. (a) Show that the square of a Rayleigh-distributed random variable, β , forms an exponential distribution. In this way an exponentially distributed random variable, p , can be generated from two independent Gaussian-distributed random variables x and y from the relation $p = x^2 + y^2$.
- (b) Simulate 100 samples of an exponentially distributed random variable with variance 1 using MatLAB or an alternative computational tool. Create the PDF and CDF of the 100 simulated samples, and compare the results with the theoretical PDF and CDF values of the exponential distribution.
9. (a) Give a transformation that generates an exponentially distributed random variable from a uniformly distributed random variable.
- (b) Simulate 100 samples of an exponentially distributed random variable with variance 1 using MatLAB or an alternative computational tool. Create the PDF and CDF of the 100 simulated samples, and compare the results with the theoretical PDF and CDF values of the exponential distribution.

10. Generate a Rician-distributed random variable, and check its CDF against the theoretical CDF of the Rician distribution. Assume that the mean and variance of the random variable are both normalized to 1.

PROJECTS

Project 1: Deployment of IEEE 802.11b and g WLANs

Part I: Modeling of the RSS. To develop a model for coverage of IEEE 802.11b and g WLANs, a group of undergraduate students at Worcester Polytechnic Institute [Bha03] measured the received signal strength (RSS) in six locations on the third floor of the Atwater Kent Laboratories at Worcester Polytechnic Institute, shown in Fig. P4.1. After subtracting the RSS from the transmitted power recommended by the manufacturers, they calculated the path loss for all the points that are shown in Table P4.1. To develop a model for the coverage of the WLANs, they used the simple distance–power gradient model

$$L_p = L_0 + 10\alpha \log_{10}(d)$$

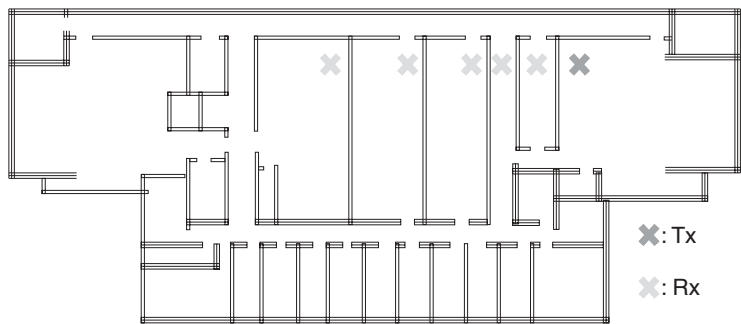


FIGURE P4.1 Location of the transmitter and first five locations of the receiver used for calculation of the RSS and path loss.

TABLE P4.1 Distance Between the Transmitter and the Receiver and the Associated Path Loss for the Experiment

Distance (m)	Number of Walls	L_p (dB)
3	1	62.7
6.6	2	70
9.5	3	72.75
15	4	82.75
22.5	5	90
28.8	6	93

in which d is the distance between the transmitter and the receiver, L_p is the path loss between the transmitter and the receiver, L_0 is the path loss 1 m from the transmitter, and α is the distance–power gradient. One way to determine L_0 and α from the results of measurements is to plot the measured L_p and $\log_{10}(d)$ and find the best-fit line to the results of measurements (similar to Fig. 4.3 for RSS).

- (a) Use the results of measurements by students to determine the distance–power gradient, α , and path loss in the 1-m distance from the transmitter, L_0 . In your report, provide the MatLAB code and the plot of the results and the best-fit curve.
- (b) Assuming that the antenna gain for the transmitter and the receiver are the same and the center frequency for the measurements is 2.41 GHz, use Eq. (3.2.1) to calculate the antenna gains of the transmitter and receiver.
- (c) Manufacturers often provide similar measurement tables for typical indoor environments. Table P4.2 shows the RSS at different distances for open areas (an area without a wall), semiopen areas (typical office areas), and closed areas (harsher indoor environments) provided by Proxim, a manufacturer of WLAN products. Use the results of measurements from the manufacturer and repeat part (a) for the three areas used by the manufacturer. Which of the measurements areas used by the manufacturer resembles the third floor of the Atwater Kent Laboratories used by the students? Assume that the transmitted power used for these measurements was 20 dBm. In your report give the curves used for calculations of the distance–power gradient in different locations.

Part II: Coverage Study. IEEE 802.11b and g WLANs support multiple data rates. As the distance between the transmitter and receiver increases, the WLAN reduces its data rate to expand its coverage. The IEEE 802.11b and g standards recommend a set of data rates for the WLAN. The first column of Table P4.2 shows the four data rates supported by the IEEE 802.11b standard, and the last column represents the RSS required to support these data rates. Table P4.3 shows the data rates and RSS for IEEE 802.11g.

- (a) Plot the data rate versus coverage (staircase functions) for IEEE 802.11b WLANs for closed, open, and semiopen areas using Table P4.2.
- (b) Plot the data rate versus coverage plots (staircase functions) for an IEEE 802.11b WLAN operating on the third floor of the Atwater Kent Laboratories (AKL) using α and L_0 values found for the third floor.

TABLE P4.2 Data Rate, Distance in Various Areas, and the RSS for IEEE 802.11b

Data Rate (Mb/s)	Closed Area (m)	Semiopen Area (m)	Open Area (m)	Signal Level (dBm)
11	25	50	160	–82
5.5	35	70	270	–87
2	40	90	400	–91
1	50	115	550	–94

Source: Proxim.

TABLE P4.3 Data Rates and the RSS for IEEE 802.11g

Data Rate (Mb/s)	RSS (dBm)
54	−72
48	−72
36	−73
24	−77
18	−80
12	−82
9	−84
6	−90

Source: Cisco.

- (c) Plot results of parts (a) and (b) in one figure and find the area suggested in Table P4.2 that resembles measurements from the third floor of AKL.
- (d) Plot the data rate versus coverage (staircase functions) for IEEE 802.11b and g WLANs operating on the third floor of the AKL. Discuss the differences.
- (e) Repeat part (d) for the open areas described in Table P4.2. How does the coverage of IEEE 802.11b and g in AKL compare with the coverage of the open area?

Project 2: Simulation Techniques for Fast Envelope Fading

In this project we examine two techniques to simulate fluctuations of the radio channel seen by a moving vehicle. Simulation of the fluctuations of the radio channel can be used in larger programs to evaluate the optimum system design parameters, such as code lengths for error recovery and training times for adaptive equalizers to combat the harsh nature of the radio channel. To simulate the fluctuations of the radio channel, we need to generate a random process with specific envelope-fading density function and a specific Doppler spectrum. The random variable is to be complex, where the magnitude follows a Rayleigh fading distribution while the phase follows a uniform distribution. The power spectral density or spectrum of the random variable should follow the classic Doppler spectrum, given by

$$D(f) = \frac{1}{2\pi f_m} \frac{1}{\sqrt{1 - (f/f_m)^2}}, \quad |f| \leq f_m$$

where f_m is the maximum Doppler frequency. In the next section we describe how such a random variable with the proper attributes can be constructed.

In Section 4.5 we discussed the simulation of fast envelope fading. In this project we want to simulate the channel and observe its statistics and spectrum. You can select either Clarke–Jakes model or the JTC model to implement the project. You will receive extra credits if you do both approaches.

Part I: The Clarke–Jakes Model. In this model, the fluctuations of the radio channel are obtained by using a signal composed of a combination of discrete sinusoids $T(t)$ provided in the equations below. A set of values of $T(t)$ for different values of t has

a magnitude that approximates a Rayleigh fading distribution and a phase that approximates a uniform distribution. Furthermore, the power spectrum of $T(t)$ approximates the shape of the classic Doppler.

$$T(t) = \frac{E_0}{\sqrt{2N_0 + 1}}(x_c + jx_s)$$

$$x_c = \sqrt{2} \cos \phi_N \cos w_m t + 2 \sum_{n=1}^{N_0} \cos \phi_n \cos w_n t$$

$$x_s = \sqrt{2} \sin \phi_N \cos w_m t + 2 \sum_{n=1}^{N_0} \sin \phi_n \cos w_n t$$

$$\phi_n = \frac{n\pi}{N_0 + 1}, \quad n = 1, 2, \dots, N_0 \quad \text{with} \quad \phi_N = 0$$

$$w_n = w_m \cos \left(\frac{2\pi n}{N} \right), \quad n = 1, 2, \dots, N_0 \quad \text{with} \quad N_0 = \frac{1}{2} \left(\frac{N}{2} - 1 \right)$$

and $w_m = 2\pi f_m$

where f_m is the maximum Doppler frequency.

- (a) Using MATLAB, compute $T(t)$ for 5120 points if the speed of the vehicle is 100 km/h, $E_0 = 1$, $N_0 = 8$, and carrier frequency is 900 MHz (cellular band). For all questions, assume $T(t)$ is sampled at 4 times the maximum Doppler frequency f_m .
- (b) Provide a plot of the magnitude of $T(t)$ (in dB) and the phase of $T(t)$ (in degrees) as a function of t .
- (c) Provide a plot of the histogram of the magnitude of $T(t)$ (in linear form), and the histogram of the phase of $T(t)$ (in degrees). Examine whether they fit the expected Rayleigh and uniform distributions. What is the difference between the histogram and the probability density function?
- (d) Provide the plot of the power spectral density, $|T(f)|^2$, as a function of normalized frequency f/f_m , where $T(f)$ is the Fourier transform of the $T(t)$. Compare the results of your simulation with the expected spectrum shown in Fig. P4.2.
- (e) Provide a plot of simulated and theoretical normalized downward level-crossing rate (all Section 4.5) versus the normalized threshold for Rayleigh envelope fading, similar to the one in Fig. 4.12. (*Hint*: Use the function `trapz` in MATLAB to compute integral)

Part II: JTC Model. In this model, two independent Gaussian (normal) random variables are filtered using a thirty-second-order IIR filter that approximates the classic Doppler spectrum and added using (in-phase and quadrature) configurations. If y_i and y_q are two independent Gaussian-distributed variables, $s = \sqrt{y_i^2 + y_q^2}$ will be a Rayleigh-distributed random variable. Therefore, the magnitude of the output signal of the system shown below will follow a Rayleigh distribution, and the power spectrum approximates the shape of the classic Doppler.

Using MATLAB, generate two sequences of independent Gaussian random numbers x_i and x_q of the length 5120 with function using `randn(1, 5120)`. Compute filtered

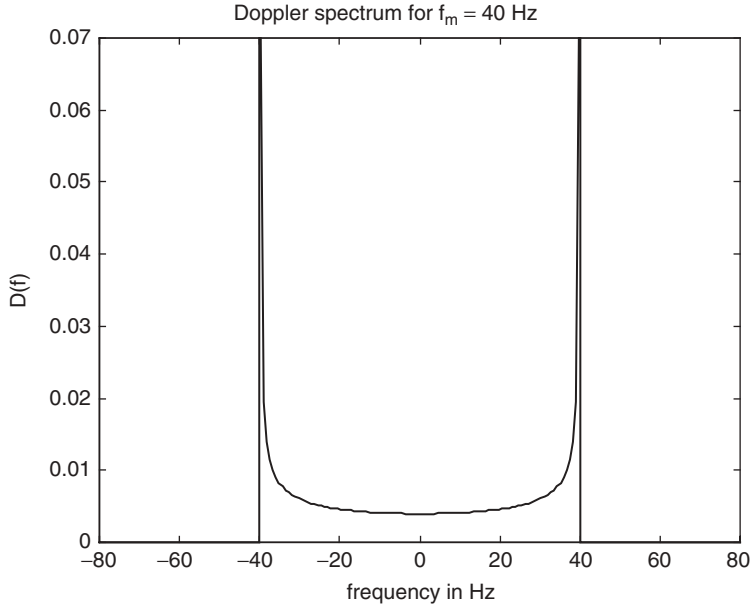


FIGURE P4.2 Example of Doppler spectrum for a maximum Doppler frequency of $f_M = 40$ Hz.

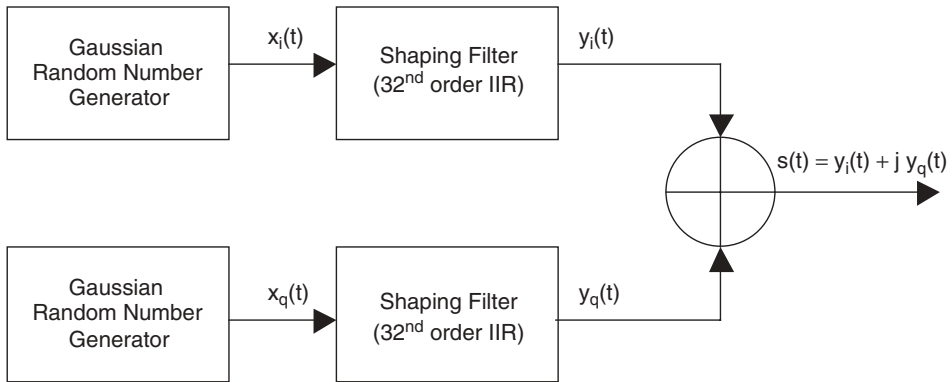


FIGURE P4.3 Simulation of channel fluctuations using a filtered Gaussian noise.

Gaussian vectors y_i and y_q , shown in Fig. P4.3, as the output of the filters using `filtfilt` function in MATLAB with the coefficients of the IIR filter in Table P4.4. For all questions, assume that $s(t)$ is sampled at four times the maximum Doppler frequency f_m .

- Provide a plot of magnitude (in dB) and phase (in degrees) of $|s(t)|$ as a function of time.
- Provide histograms of the magnitude (in linear form, not in dB) and phase (in degrees) of $s(t)$. Examine whether they fit the expected Rayleigh and uniform distributions.

TABLE P4.4 Classic Spectrum IIR Filter Coefficients for the JTC Model

Denominator Coefficients	Numerator Coefficients
1.000000000000000e+00	6.5248059900135200e-02
-1.2584602815172037e+01	-5.6908289014580038e-01
8.3781249094641240e+01	2.7480451166883220e+00
-3.8798703729842964e+02	-9.4773135180288293e+00
1.3927662726637102e+03	2.5786482996126544e+01
-4.1039030305379210e+03	-5.8241097311312117e+01
1.0278517997545167e+04	1.1247173657687033e+02
-2.2393748634049065e+04	-1.8904842233132774e+02
4.3133809439790406e+04	2.7936237305345003e+02
-7.4319282567554124e+04	-3.6418631194112885e+02
1.1554604041649372e+05	4.1715604202981109e+02
-1.6315680006218722e+05	-4.1320604132753033e+02
2.1026268214607492e+05	3.3901659663025242e+02
-2.4818342600838441e+05	-2.0059287960205506e+02
2.6898038693500403e+05	2.3734545818966293e+01
-2.6809721585952450e+05	1.5363912802007360e+02
2.4593366073473063e+05	-2.9424154728837402e+02
-2.0763108908648306e+05	3.7359596060374486e+02
1.6120527209223103e+05	-3.8642988435890055e+02
-1.1492103434104947e+05	3.4521505714177903e+02
7.5041686769138993e+04	-2.7265055759799253e+02
-4.4731841330872761e+04	1.9230535924562764e+02
2.4231115205405174e+04	-1.2153980630698008e+02
-1.1857508216082340e+04	6.8773930574859179e+01
5.2013837692697152e+03	-3.4696126060493945e+01
-2.0246855591971096e+03	1.5489134454590417e+01
6.9005516614518956e+02	-6.0495383196143626e+00
-2.0220131802145625e+02	2.0332679679817174e+00
4.9649188538197400e+01	-5.7404157101686004e-01
-9.8333304002079363e+00	1.3121847123296254e-01
1.4770279039919996e+00	-2.2867487042024594e-02
-1.5005452926258436e-01	2.7118486134987282e-03
7.7628588864503741e-03	-1.6371291227220021e-04

- (c) Provide a plot of the power spectral density of $S(f)^2$ as a function of normalized frequency f/f_m , where $S(f)$ is the Fourier transform of the $s(t)$. Compare the results of your simulation with the expected spectrum shown in Fig. P4.2.
- (d) Provide a plot of simulated and theoretical normalized downward level-crossing rate (see Section 4.5) versus the normalized threshold for Rayleigh envelope fading similar to the one in Fig. 4.12. (*Hint:* Use the function `trapz` in MATLAB to compute integral.)

Project 3: Simulation of Shadow Fading and Handoff

In this project we simulate the fluctuations of average received signal strength due to shadow fading in a microcellular network, and we use that to analyze a simple handoff algorithm. The scenario of operation is shown in Fig. P4.4. Four base stations, BS_i , $i = 1, 2, 3$, and 4, are located in four street crossings in a microcellular

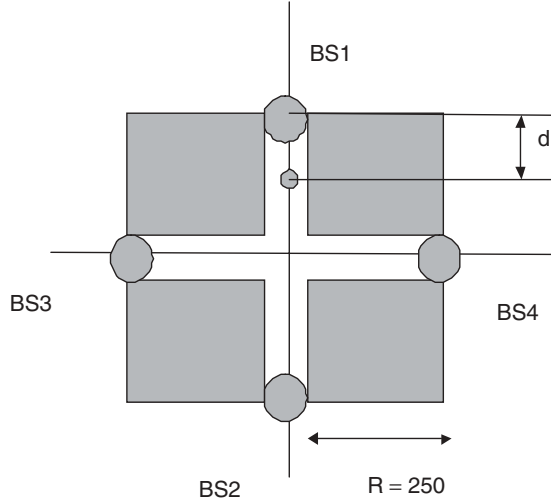


FIGURE P4.4 Four-base-station scenario for a microcellular operation.

network. The mobile host (MH), moving from BS1 toward BS2 in the figure, is communicating through BS1. As MH moves away, the received signal strength (RSS) from BS1 decreases and the RSS from BS2 to BS4 increases. At certain points, the received power from BS-1 becomes weak and MH starts to search for another BS that can provide a stronger signal and selects that base station as its point of connection. This change of base stations is referred to as *handoff*. In an ideal system we would expect the handoff decision only once in the middle of the path between BS1 and BS2. In practice, depending on the handoff algorithm, we may have several handoffs in different locations. In this project we consider the simplest and the most obvious algorithm that simply connects the MH to the BS with the strongest average RSS value. To analyze the situation, we use a channel simulation model to simulate the average RSS from all base stations, and we observe the number and location of handoffs.

We use a distance-partitioned model with two slopes to simulate the channel. In this model the path loss increases with a slope of 2 to a breakpoint at a distance of 150 m; then the slope is increased to 3. With this model, the RSS values at a distance d in the LOS paths associated with BS1 and BS2 are given by

$$\text{RSS}(d) = P_t - P_0 - \begin{cases} 20 \log_{10}(d), & d \leq 150 \\ 20 \log_{10}(150) + 30 \log_{10}(d/150), & d > 150 \end{cases} + l(d)$$

in which d is the distance between the mobile host and BS1 in meters, $P_t = 20$ dBm is the transmitted power of the base stations, $P_0 = 38$ dB is the path loss in the first meter calculated for 1.9-GHz PCS bands, and $l(d)$ is the lognormal shadow fading with variance of 8. For the OLOS propagation associated with the RSS from BS3 and BS4, a LOS propagation is assumed up to the street corner, and after the corner the propagation path loss is calculated by placing an imaginary transmitter at the corner with the transmit power equal to power received at the corner from the LOS

base station. As a result, RSS at a distance $d + R$ from the OLOS base station is given by

$$\text{RSS}(d) = P'_t - \begin{cases} 20 \log_{10}(d), & d \leq 150 \\ 20 \log_{10}(150) + 30 \log_{10}(d/150), & d > 150 \end{cases} + l(d)$$

where P'_t is the RSS in LOS at the middle cross section:

$$P'_t = P_t - P_0 - 20 \log_{10}(150) + 30 \log_{10}(250/150)$$

This model assumes that all power arriving in the cross section is diffracted in other directions. For the simulation of the lognormal fading we assume that a random Gaussian noise $N(0, 1)$ with zero mean and a variance of 1 is passed through a low-pass filter characterized by the transfer function

$$H(z) = \frac{\sigma_2}{1 - \alpha z^{-1}}$$

where α , designating the location of the pole of the filter, is a number very close to 1, to keep the bandwidth very low. Then, samples of shadow fading effects can be simulated from

$$\begin{aligned} \alpha &= e^{1/85} \\ \sigma_1^2 &= 8 \\ \sigma_2^2 &= \sigma_1^2(1 - \alpha^2) \\ s(1) &= \sigma_1 N(0, 1) \\ s(i) &= \alpha s(i - 1) + \sigma_2 N(0, 1) \end{aligned}$$

In this simulation the first point of the simulation must be at $d = \sqrt{g} = \sqrt{150}$, and the last point should be at $d = 2R - \sqrt{g} = 500 - \sqrt{150}$.

The following complementary code facilitates the simulations. This code generates one simulation of RSS from the four base stations when the MH goes from BS1 to BS2, and both gradients are assumed to be 2.

```
% Declare the various variables used for distances
R = 250;
L = 2 * R;
speed = 1;
sample_time = 0.1;
step_distance = speed * sample_time;
g = 150;
min_distance = sqrt(g);
max_distance = L - sqrt(g);
d1 = [min_distance:step_distance:max_distance];
d2 = L - d1;
d3 = abs(R - d1);
```

```

d4 = abs(R - d1);
Ns = length(d1);

% Declare variables and compute RSS
% Part 1: Computations independent of the random variable
%       for shadow fading
Pt = 20;
Po = 38;
grad1 = 2;
grad2 = 2;
alpha = exp(-1/85);
sigma1 = sqrt(8);
sigma2 = sqrt(sigma1^2 * (1 - alpha^2));
RSS01 = Pt - Po - (10 * grad1 * log10(d1) + 10 * grad2 *
    log10(d1/g));
RSS02 = Pt - Po - (10 * grad1 * log10(d2) + 10 * grad2 *
    log10(d2/g));
RSS_corner = Pt - Po - (10 * grad1 * log10(R) + 10 * grad2 *
    * log10(R/g));
RSS03 = RSS_corner - (10 * grad1 * log10(d3) + 10 * grad2 *
    log10(d3/g));
RSS04 = RSS_corner - (10 * grad1 * log10(d4) + 10 * grad2 *
    log10(d4/g));

for i=1:Ns
    if d3(i) < min_distance
        RSS03(i) = RSS_corner;
    end;
    if d4(i) < min_distance
        RSS04(i) = RSS_corner;
    end;
end;

% Part 2: Adding the random variable for shadow fading
s1(1) = sigma1 * randn(1);
s2(1) = sigma1 * randn(1);
s3(1) = sigma1 * randn(1);
s4(1) = sigma1 * randn(1);

for i=2:Ns
    s1(i) = alpha * s1(i-1) + sigma2 * randn(1);
    s2(i) = alpha * s2(i-1) + sigma2 * randn(1);
    s3(i) = alpha * s3(i-1) + sigma2 * randn(1);
    s4(i) = alpha * s4(i-1) + sigma2 * randn(1);
end;

RSS1 = RSS01 + s1;
RSS2 = RSS02 + s2;

```

```

RSS3 = RSS03 + s3;
RSS4 = RSS04 + s4;

% Plot the RSS values obtained
figure(1)
plot(d1, RSS1, 'r')
hold on
plot(d1, RSS2, 'b')
hold on
plot(d1, RSS3, 'g')
hold on
plot(d1, RSS4, 'c')
title('RSS versus distance along route')
xlabel('distance from BS1 in meters');
ylabel('dBm');

```

Figure P4.5 provides a sample result expected from this simulation if both gradients are fixed at two.

Using the above discussion, do the following:

- (a) Write your simulation for the RSS from four base stations with two different gradients described earlier and plot a sample result similar to Fig. P4.5. Assume

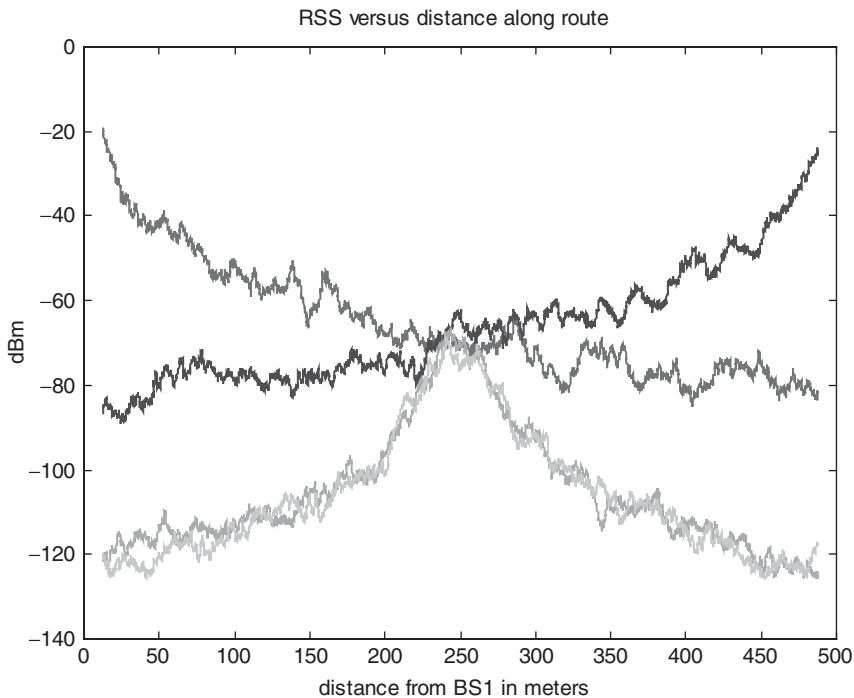


FIGURE P4.5 Sample output of the MATLAB code.

that MH walks from the vicinity of BS1 toward BS2 with a constant speed of 1 m/s and that the distance of a block shown in Fig. P4.4 is $R = 250$ m. The sampling frequency is 10 Hz, which means that MH measures RSS in every 0.1 s.

- (b) Assuming that MH always connects to the BS with the strongest RSS, expand your program to record the location and number of handoffs in each experiment of moving from BS1 to BS2. Run the program three times to get three random trials. Give the number and location of handoffs for all three trials. Suggest some rational modifications to the simple algorithm to reduce the number of handoffs.

5

MEASUREMENT OF WIDEBAND AND UWB CHANNEL CHARACTERISTICS

- 5.1 Introduction
- 5.2 Time-Domain Measurement Techniques
 - 5.2.1 Measurements Using Direct Pulse Transmission
 - 5.2.2 Measurements Using Spread-Spectrum Signals
 - 5.2.3 Results of Time-Domain Wideband Measurements
- 5.3 Frequency-Domain Measurement Techniques
 - 5.3.1 Measurement Using a Network Analyzer
 - 5.3.2 Comparison Between Measurement Systems
- 5.4 Advances in Frequency-Domain Channel Measurement
 - 5.4.1 Frequency-Domain Measurement for TOA Measurements
 - 5.4.2 Superresolution Algorithms for Frequency-Domain Measurements
 - 5.4.3 Frequency-Domain Measurement of the Angle of Arrival
 - 5.4.4 Frequency-Domain Measurement for UWB Measurements
- Questions
- Problems
- Project
 - Project 1: Analysis of Measured Data Using a Network Analyzer

5.1 INTRODUCTION

In narrowband measurements, we analyze the response of the channel at or around a single frequency, and from these measurements we are able to extract the power fluctuations caused by the signal arriving from a number of different paths. Narrowband measurements do not provide any information on the magnitude or the time delay of any individual path. Rather, they reflect the vector addition of the complex amplitudes of the arriving paths as observed in the power fluctuations in the received narrowband signal. Wideband measurements, in contrast, provide information on the multipath delay spread and structure of individual paths as well as the frequency selectivity of the channel. Stated in simple terms, if we assume that the channel is fixed during a measurement

interval, narrowband measurements resemble measurements of the channel response to a single frequency, whereas wideband measurements resemble measurements of the impulse response or overall frequency response of the channel.

Wideband measurements can be performed either in the time domain by direct measurement of the impulse response of the channel, or in the frequency domain by direct measurement of the frequency response of the channel. In theory, using Fourier transform techniques, the measured time and frequency responses should provide identical results. However, as we will see later, there are some shortcomings in using the Fourier transform of the results of measurements, particularly if the measurement system does not provide both the magnitude and phase of the measured characteristics.

In this chapter we describe measurement techniques used to determine the wideband and UWB characteristics of radio propagation and present some results obtained in such measurements. Systematic measurements for wireless applications are done in several ways:

1. *Spatial or large-scale measurements* in which one of the terminals is held fixed and the other terminal is moved to different locations, spaced at least several wavelengths apart.
2. *Local or small-scale measurements* in which the transmitter or receiver is moved about in an area surrounding a specific location, to collect a number of measurements.
3. *Traffic-effect or temporal measurements* in which the transmitter and receiver are held fixed and measurements are made with traffic moving between or around the terminals.
4. *Partitioned measurements* in which the effects of dividing walls on the characteristics of the channel are studied. The overall measurement area is partitioned, and characteristics and parameters of the channel in the smaller areas are measured and compared.
5. *Frequency-dependence measurements* in which characteristics measured at different frequencies are compared.
6. *Measurements of angle of arrival* in which characteristics of multipath components arriving from different angles of arrival into the receiver antenna are considered for MIMO and positioning applications.
7. *Measurement of direct path, TOA* in which the time of arrival (TOA) of the direct path between the transmitter and receiver is measured to be used in TOA-based geolocation systems.

In this chapter we introduce the traditional time-domain measurement systems used for broadband urban and indoor areas. These techniques were the first wideband techniques, used since 1970s for the measurement of a variety of wideband radio channels [Cox72, Tur72, She75]. Then we discuss the frequency-domain measurement techniques that were introduced in the 1990s [How90c, Pah90b], which have become very popular for measurements of MIMO angle of arrival [Spe00, Tin00], UWB measurements [Cas02, Gha03], and TOA for indoor geolocation applications [Ben99].

5.2 TIME-DOMAIN MEASUREMENT TECHNIQUES

The objective of traditional time-domain measurement systems is to measure the impulse response of the channel,

$$h(\tau, t) = \sum_{i=1}^L \beta_i(t) \delta(t - \tau_i(t)) e^{j\phi_i(t)} \quad (5.2.1)$$

by direct measurements of β_i , τ_i , and ϕ , representing the magnitude, arrival time, and phase, characterizing individual paths between the transmitter and receiver. The wireless channels considered in this book are slowly time varying, allowing windows of time for measurement of the individual path parameters. In principle, in a slowly time-varying channel, if we transmit a narrow RF pulse with envelope $p(t)$, resembling an impulse, at time $t = 0$ we can capture the received signal,

$$h(\tau) = h(\tau, 0) = \sum_{i=1}^L \beta_i p(\tau - \tau_1 - \tau_i) e^{j\phi_i} \quad (5.2.2)$$

in which τ_1 is the TOA of the direct path between the transmitter and the receiver. We can then calculate the path parameters from the measured impulse response. The arrival of the first path is $\tau_1 = d/c$, where d is the distance between the transmitter and receiver and c is the velocity of wave propagation, which is only important when we are interested in modeling the TOA for geolocation applications. For telecommunication applications we are interested in the relative arrival and strength of the paths, so we assume that $\tau_1 = 0$ and we treat the measured channel impulse response as

$$h(\tau) = h(\tau, 0) = \sum_{i=1}^L \beta_i p(\tau - \tau_i) e^{j\phi_i} \quad (5.2.3)$$

from which we can extract the amplitude, arrival time, and phase of the individual paths.

In practice, the impulse response of the channel is measured either by transmitting a wideband spread-spectrum signal and correlating the received signal with the transmitted sequence, or by direct transmission of a short radio-frequency (RF) pulse and observing the received signal arriving from different paths. As we show in Section 5.2.2, the spread-spectrum technique is a virtual method for implementation of a pulse transmission technique. In both cases, time resolution of the measurements is inversely proportional to the bandwidth of the measurement system. The spread-spectrum method sends a steady stream of bits, and the ratio of peak to average transmitted power is unity. With the pulse transmission method, an RF pulse is transmitted periodically with a low duty cycle, and the ratio of peak to average power is very high. As a result, given amplifiers designed for identical peak power operation, we can achieve greater coverage with the spread-spectrum approach. In practical implementations of the two systems described in this chapter, we will achieve better coverage with the spread-spectrum technique and better

resolution and acquisition time using the direct pulse-sounding technique. Consequently, for areas of less than 100 m in radius, the pulse-sounding technique is more popular, and for larger areas the spread-spectrum technique is used more typically. For most indoor applications such as WLANs or WPANs, path distances of interest are typically up to a few tens of meters, and thus the pulse-sounding technique has been applied extensively. For mobile radio and PCS applications used in outdoor areas, path distances are longer, and the spread-spectrum technique is used more typically. We provide the details of implementation of these two techniques in the next two subsections. We start with the simpler direct pulse transmission technique, and then describe the more complex spread-spectrum technique.

5.2.1 Measurements Using Direct Pulse Transmission

An obvious way to measure the impulse response of a channel is to transmit a very short RF pulse and measure the impulse response of the channel from the received signal. This method was originally used for urban radio channel measurements in the early 1970s [Tur72]. In the late 1980s it attracted another wave of attention for indoor radio propagation studies [Sal87b, Pah89, Rap89]. If an RF pulse with a complex envelope $p(t)$ is transmitted periodically with a period of T_s , the received signal would be periodic multiple received pulses:

$$r(t, \tau) = \sum_n \sum_{i=1}^L \beta_i(t) p(t - nT_s - \tau_i(t)) e^{j\phi_i(t)} \quad (5.2.4)$$

Assuming a slowly time-varying channel, the channel parameters during several periods remain constant. If we capture one period or average the signal over a few periods, the resulting captured signal is a close approximation to the impulse response of the channel:

$$h(\tau) = \sum_{i=1}^L \beta_i p(\tau - \tau_i) e^{j\phi_i} \approx \sum_{i=1}^L \beta_i \delta(\tau - \tau_i) e^{j\phi_i} \quad (5.2.5)$$

Measurement of the phase needs a coherent receiver with two branches for in-phase and quadrature-phase signal detection. Therefore we need an additional circuit to provide for a stable reference carrier phase. A simpler receiver is an envelope detector that detects the envelope of the received RF signal with a single branch without need of a reference carrier frequency.¹ The square of one period of the noncoherently detected received signal for an envelope-detected receiver is given by

$$Q(\tau) = |h(\tau)|^2 = \sum_{i=1}^L \beta_i^2 p^2(\tau - \tau_i) \quad (5.2.6)$$

in which $Q(\tau)$ is the classical delay power spectrum defined in Chapter 3. In Eq. (5.2.6) the envelope detection process has eliminated the information related to the phase. If we detect the peak of each individually arriving path in the stored profiles, the square

¹More details on quadrature techniques and coherent versus noncoherent modulation are discussed in Chapter 7.

root of its magnitude represents the amplitude β_i and its occurrence time, the arrival time τ_i . In general, the phase of the arriving paths is often assumed to be a uniformly distributed random variable, and the measurement of its statistics is not that important. However, if we measure the phase, we can reproduce the exact frequency response of the channel.

The design parameter for pulse transmission systems is the bandwidth of the transmitted pulses and the period for the transmissions of periodic pulses. The bandwidth of the transmitted RF pulses with complex envelope $p(t)$ should be wide enough to produce narrow pulses capable of resolving multiple arriving paths. Wider bandwidths allows us to detect a larger number of paths, revealing more details about the channel behavior. However, in practice we perform measurements and modeling for particular system applications and a measurement system that has a bandwidth with the same order of bandwidth of the system. The repetition period, T_s , of the transmitted pulses should be longer than multipath delay spread $\tau_1 + \tau_L$ and shorter than coherence time (inverse of the Doppler spread), defined in Section 3.5.3. As we discussed earlier, for telecommunication applications we are only interested in the multipath profile and multipath spread. In this case, the TOA of the first path associated with correct measurement of the arrival time of the direct path is not important and we can assume that $\tau_1 = 0$.

Example 5.1: Measurement Parameters for an Indoor Measurement System The popular IEEE 802.11b and g WLANs are operating in 2.4-GHz ISM bands that have 84 MHz of bandwidth available. The multipath spread of the indoor radio channel is at most around several hundred nanoseconds. The Doppler spectrum of this channel is around $f_M = 10$ Hz. RF pulses with 84 MHz of bandwidth and a repetition time of $T_s = 500$ ns are well suited for this environment. The 84-MHz bandwidth covers the entire spectrum. The 500-ns repetition time is longer than the multipath spread of the channel and shorter than coherence time of $1/f_M = 100$ ms. Since the coherence time is much longer than the repetition period, we can easily average a number of channel profiles to reduce the background noise and still keep the measurement time below fractions of the coherence time. With a fixed transmitter power, a reduction of the background measurement noise increases the range of the measurement system. To bring the averaging into calculation of coverage using path-loss equations, we can simply increase the maximum path loss with the 10 log of the number of cycles used for averaging. Therefore, in an environment with a distance–power gradient of 2, a 100-time averaging will increase the coverage by an order of magnitude.

Next we provide some examples for practical measurement systems using direct pulse transmission. The example pulse measurement systems discussed in this section are very similar in design. All incorporate noncoherent receivers, so that a power-delay profile is measured rather than an impulse response. Because the phase is not available, it is not convenient to find the exact frequency response of the channel by using the Fourier transform. The resolution and measurement time for these systems are better than for the systems described in the next section, but the dynamic range of measurements is more restricted.

In [Sal87b], a 1.5-GHz CW signal was modulated by a train of 10-ns pulses with a 600-ns repetition period. A vertically polarized omnidirectional discone antenna was used to transmit this signal. At the receiver, a similar antenna was followed by an amplifier chain and a coherent square-law detector. A computer-controlled oscilloscope was then used to collect the received power-delay profile. A coaxial cable was

used to trigger the oscilloscope from the transmitter's pulse generator to guarantee a stable timing reference. Using limited measurements made with this system in one office building, rms delay spreads and power–distance relationships were calculated, and a statistical model for indoor multipath propagation was developed. For the measurements and the model, the phases of the multipath components are assumed a priori to be statistically independent uniform random variables over $(0, 2\pi)$. This model, introduced in 1997, has attracted a considerable amount of attention in recent years. A modified version of this model is considered by the IEEE 802.11 community for MIMO channel modeling, and another version is considered by IEEE 802.15 for UWB channel modeling. Details of the basic Saleh–Valenzuela model are presented in Section 6.2.

In [Rap89] a similar wideband measurement system was used to collect propagation data in factory settings. A 1.3-GHz carrier was modulated by a train of 10-ns pulses with a 500-ns repetition period. Discone antennas were used at both the transmitter and receiver. The receiving oscilloscope was triggered internally by the first received pulse in each power–delay profile measurement. Using measurements made with this system in five factory environments, rms delay spreads and power–distance relationships were calculated, and a statistical model for indoor multipath propagation was proposed. Comparing these measurements and some narrowband measurements made using the same measurement system [Rap89], an argument is made for the phases of the multipath components to be statistically independent uniform random variables over $(0, 2\pi)$.

Example 5.2: Details of a Simple Direct Pulse Measurement System We now describe the simple measurement system used in [Pah89] for wideband indoor radio propagation studies. Figure 5.1 shows a schematic diagram of the measurement setup. The setup operates at a 910-MHz carrier frequency. The modulated carrier is fed to a 45-dB amplifier, and the output is transmitted with an omnidirectional quarter-wave dipole antenna placed about 1.5 m above the floor level. The stationary receiver uses the same type of antenna at the same height, which is approximately the height of an antenna mounted on top of a desktop PC. The antenna is followed by a step attenuator and a low-noise, high-gain (~ 60 dB) amplifier chain. The signal is then detected using a noncoherent square-law envelope detector whose output is displayed on a digital storage oscilloscope coupled to a PC with a GPIB instrument bus. A coaxial cable is used to trigger the oscilloscope from the pulse generator of the transmitter, to guarantee a stable timing reference. Typically, 64 repetition periods are averaged by the oscilloscope to form a power–delay profile. The base width of the pulses in the received profiles in this noncoherent system is 5 ns, which has better resolution than the apparatus used in [Sal87b] and [Rap89b]. Figure 5.2 shows three samples of measurements taken with this system in different locations. Figure 5.3 shows a three-dimensional plot of 20 multipath profiles in one line-of-sight (LOS) location. The system is capable of measuring and storing up to 20 complete profiles in 1 s, which is adequate for observation of the effects of indoor Doppler spread in the wideband signal.

5.2.2 Measurements Using Spread-Spectrum Signals

The traditional method for wideband measurement of the multipath spread in radio channels has been the use of principles of spread-spectrum technology. This method

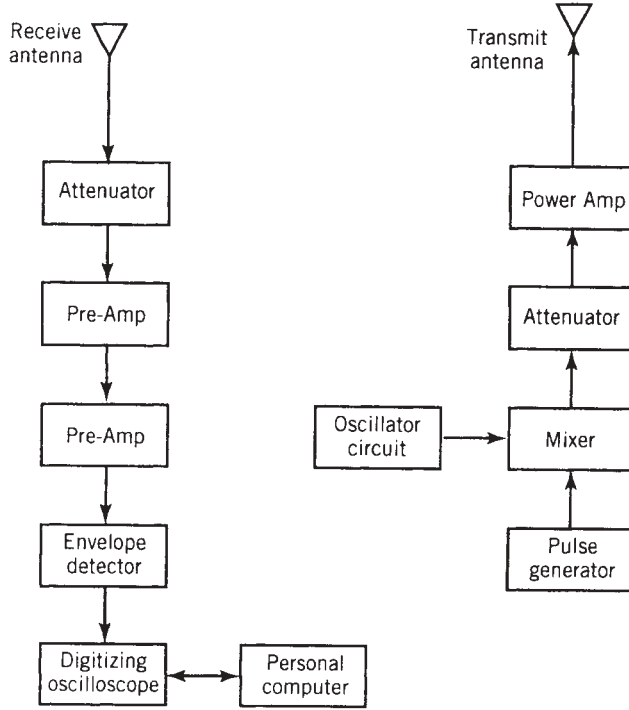


FIGURE 5.1 Simple pulse transmission measurement system used for wideband time-domain measurements of indoor radio propagation.

was used for wideband measurement of the mobile radio channel [Cox72, Par89] as well as other radio channels, such as troposcatter [She75]. The earliest wideband measurements of multipath spread in building environments [Dev84] were made using a spread-spectrum receiver adapted from a measurement system used for the mobile radio channel [Cox72]. The same approach was used in [Bul89] to study and compare indoor radio propagation characteristics at 910 MHz and 1.75 GHz. In this section we outline the basic principles of direct-sequence spread-spectrum (DSSS) communications as applied to wideband channel measurement, and we describe the implementation of DSSS using a *sliding correlator*. Further details of spread-spectrum technology and its applications to wireless information networks are provided in Chapter 10. We start our discussions by showing that a DSSS system is actually a virtual pulse transmission technique.

DSSS as a Virtual Pulse Transmission Technique. To understand the principles of this measurement technique, assume that we have a symbol shape $f(t)$ of duration T_s consisting of a sequence of N narrower pulses $p(t)$, called *chips*, with binary amplitudes $b_i = \pm 1$ and duration $T_c = T_s/N$:

$$f(t) = \sum_{i=1}^N b_i p(t - iT_c) \quad (5.2.7)$$

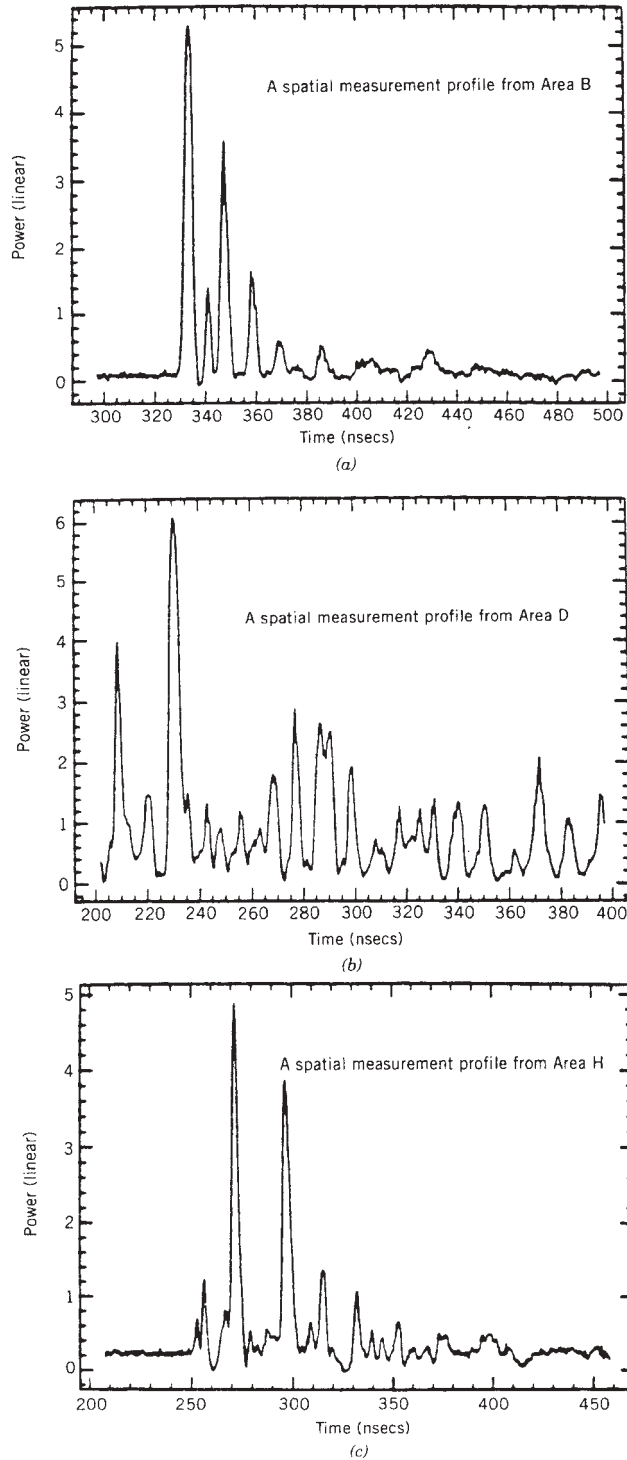


FIGURE 5.2 Samples of indoor multipath delay profiles measured using the time pulse transmission technique: (a) area B; (b) area D; (c) area H.

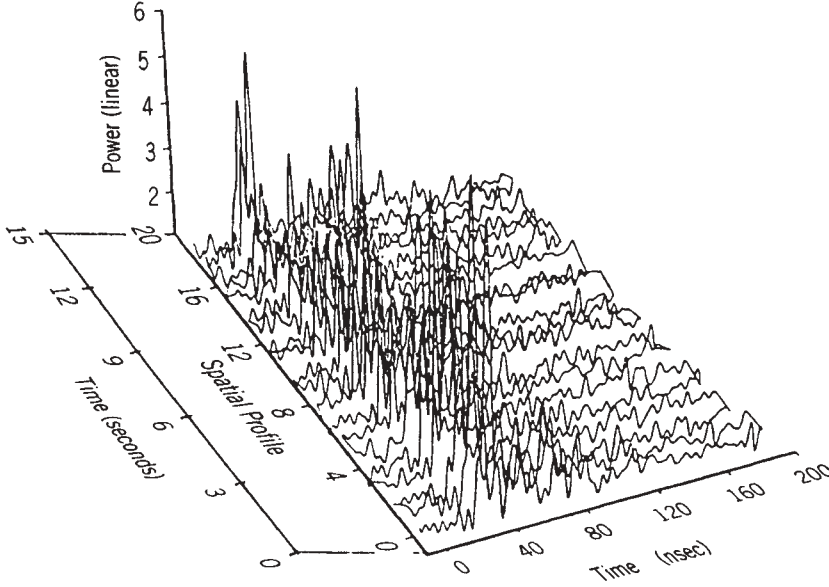


FIGURE 5.3 Plot of 20 multipath delay profiles measured to represent the short-time variations of the channel in one location [Gan93].

Ideally, the pattern of the periodic sequence $\{b_i\}$ of length N is selected so that it is orthogonal to any circularly shifted version of itself.² Because all elements of the sequence are ± 1 , the sum of squares of the sequence is N . Sequences with the orthogonality property, referred to as *pseudonoise (PN) sequences*, are treated extensively in the spread-spectrum communication literature [Sim85] and are discussed in more detail in Chapter 10.

Furthermore, assume that $x(t)$ is the periodic form of $f(t)$ repeated every T_s seconds:

$$x(t) = \sum_n f(t - nT_s) \quad (5.2.8)$$

The function $x(t)$ is a periodic function, and therefore its autocorrelation function (ACF) is also periodic, with the same period T_s . With the orthogonality condition in the sequence $\{b_i\}$, the periodic ACF of $x(t)$ is given by

$$R_{xx}(\tau) = \frac{1}{T_s} \int_0^{T_s} x(t)x(t - \tau) dt = \frac{N}{T_s} \sum_n R_{pp}(\tau - nT_s) \quad (5.2.9)$$

²If you visualize $\{b_i\}$ as a random sequence of N binary digits repeating itself every N digits, orthogonal means that

$$R(k) = \sum_{i=1}^N b_i b_{i-k} = \begin{cases} N, & i = j \\ 0, & i \neq j \end{cases}$$

For an specific example of PN sequences, see Example 10.5.

where

$$R_{pp}(\tau) = \int_0^{T_c} p(t)p(t - \tau) d\tau \quad (5.2.10)$$

where $R_{pp}(\tau)$ is the ACF of the pulse $p(t)$.

Example 5.3: Periodic ACF of a PN Sequence with Rectangular Pulses Figure 5.4 shows an example of $x(t)$, given by Eq. (5.2.8), and its correlation function for the case of rectangular $p(t)$ pulses. The transmitted pulse, $p(t)$, is a rectangular waveform with duration T_c . The transmitted symbol, $f(t)$, defined in Eq. (5.2.7), is a sequence of N rectangular pulses with random amplitudes of ± 1 . The transmitted signal, $x(t)$, is the repetition of $f(t)$ every $T_s = NT_c$ seconds. The ACF of a rectangular pulse $R_{pp}(\tau)$, defined in Eq. (5.2.10), is a triangular function with duration $2T_c$. Therefore, the periodic ACF of the PN sequence, $R_{pp}(\tau)$, is the sequence of triangular pulses repeated each T_s seconds. As shown in Eq. (5.2.9), the peak of the triangles is N/T_s times the peak of $R_{pp}(\tau)$.

Let $x(t)$ be the complex envelope of a transmitted signal and assume that it passes through a multipath channel with equivalent baseband channel impulse response given by Eq. (5.2.3). Then the complex signal envelope at the front end of the receiver is given by

$$r(t) = \sum_{i=1}^L \beta_i x(t - \tau_i) e^{j\phi_i} \quad (5.2.11)$$

that is, a periodic signal with the same period T_s as $x(t)$. If the complex envelope of the received signal, given by Eq. (5.2.11), is cross-correlated with the transmitted periodic signal $x(t)$, the resulting cross-correlation function is also a periodic function

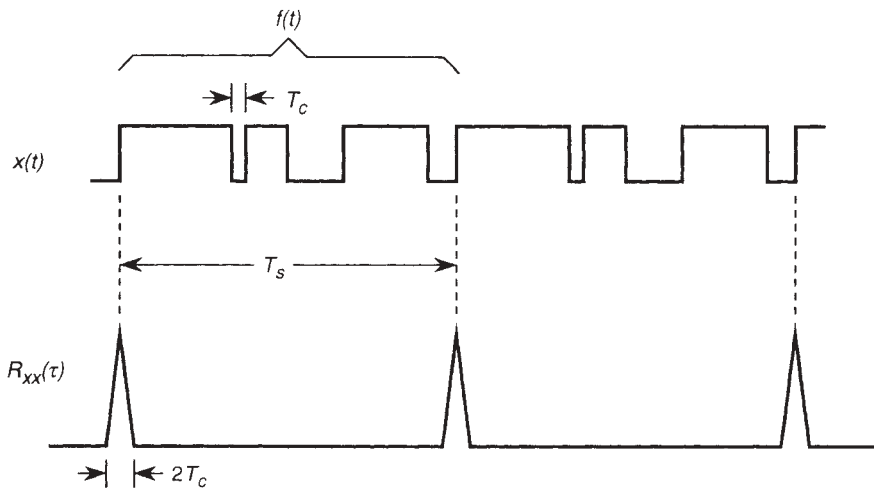


FIGURE 5.4 Periodic PN sequence with rectangular pulses and its periodic correlation function.

with period T_s given by

$$\begin{aligned}
 R_{xr}(\tau) &= \frac{1}{T_s} \int_0^{T_s} \sum_{i=1}^L \beta_i x(t - \tau_i) e^{j\phi_i} x(t - \tau) dt \\
 &= \sum_{i=1}^L \beta_i R_{xx}(\tau - \tau_i) e^{j\phi_i} \\
 &= \frac{N}{T_s} \sum_n \sum_{i=1}^L \beta_i R_{pp}(\tau - nT_s - \tau_i) e^{j\phi_i}
 \end{aligned} \tag{5.2.12}$$

in which $R_{xx}(\tau)$ is the ACF of the transmitted signal defined by Eq. (5.2.9) and $R_{pp}(\tau)$ is the ACF of the chip pulse defined by Eq. (5.2.10). If we assume that $2T_c$, the width of the correlation function $R_{pp}(\tau)$, is narrow enough to resolve all paths, and the multipath delay spread $\tau_L - \tau_0$ is less than T_s , one period of the received signal is identical to the channel impulse response, with impulses replaced by $R_{pp}(\tau)$ and a normalization factor N/T_s included in the result:

$$h(\tau) = \frac{N}{T_s} \sum_{i=1}^L \beta_i R_{pp}(\tau - \tau_i) e^{j\phi_i} \tag{5.2.13}$$

If we replace $NR_{pp}(\tau)$ in Eqs. (5.2.12) and (5.2.13) with $p(t)$, these equations become identical to Eqs. (5.2.4) and (5.2.5) for pulse transmission technique. Therefore, if we consider the received signal after the correlator, the DSSS measurement system is virtually a pulse transmission technique with transmitted pulse shape replaced by the ACF of the chip waveform and a gain of N .³ In other words, a spread-spectrum measurement system with rectangular chip shapes is equivalent to direct pulse transmission techniques sending triangular pulses.

Implementation of DSSS Using Sliding Correlator. Efficient implementation of the cross-correlator defined in Eq. (5.2.12) is one of the most important design issues in DSSS systems. Digital implementation of this correlation function requires a very high sampling rate to accommodate the wide transmission bandwidth and capture the waveform narrow pulses in time. The correlation function is a parametric convolution integral; direct calculation of this integral involves numerical integration of the integral for different delay values or the use of Fourier transform techniques, which are both very computationally extensive at high sampling rates. A relatively simple analog implementation of a cross-correlator that is used extensively in measurement systems is the *sliding correlator*.

Figure 5.5 shows the basic block diagram for implementation of a sliding correlator wideband channel measurement system. The PN sequence of length N with chip rate $R_c = 1/T_c$ is repeated every $T_s = N/R_c$ second to form the transmitted baseband signal $x(t)$. The baseband signal is then modulated onto a carrier at frequency f_c , and the modulated signal after power amplification is fed to the transmitting antenna. The receiver consists of a *sliding correlator* and a demodulator. The received signal is

³In the spread-spectrum literature, this gain is referred to as the *processing gain*.

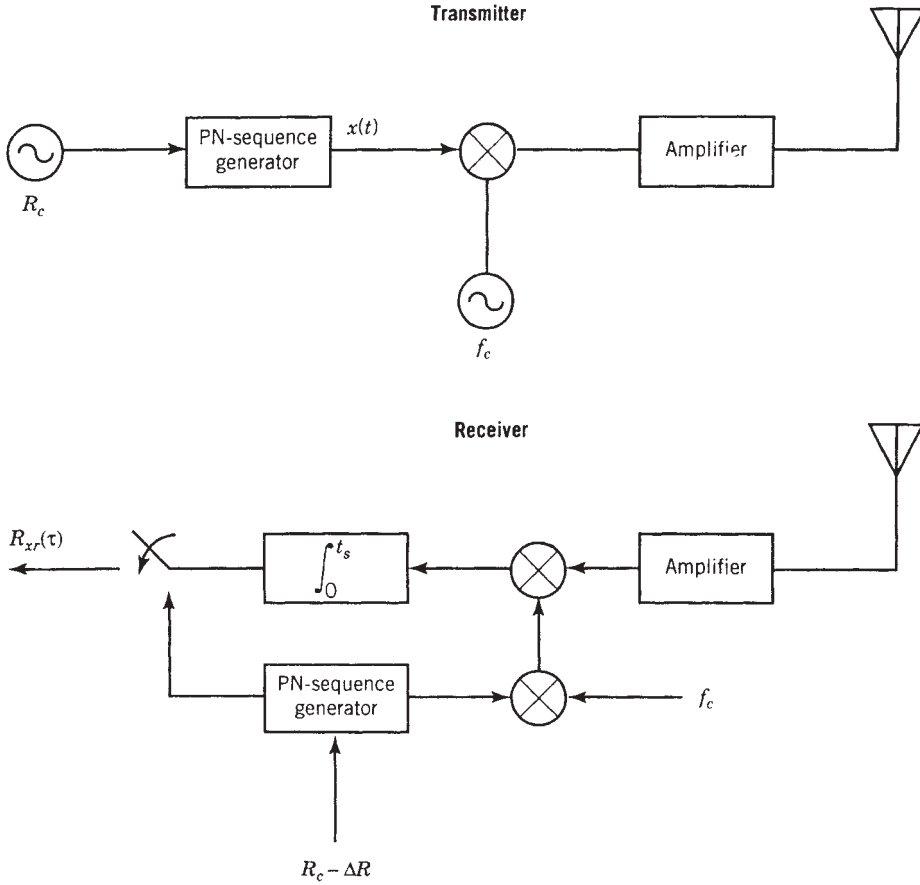


FIGURE 5.5 Block diagram of a spread-spectrum wideband measurement system.

multiplied by a replica of the transmitted sequence, $\tilde{x}(t)$, at a slower rate $R_c - \Delta R$ and integrated over $t_s = N/(R_c - \Delta R)$ to generate samples of the cross-correlation function between the received signal and a close approximation of the transmitted signal. In the following discussions we assume that t_s has a value very close to T_s .

To understand how the sliding correlator works, assume that we have rectangular pulses for the chip waveform, we put the transmitter and the receiver of the Fig. 5.5 in a back-to-back connection, and we turn off the carrier frequency so that we operate at baseband. This way we have a simple one-path channel with no delay and the transmitter and receiver codes start at the same reference time. Inserting these values for the channel in Eq. (5.2.12), we expect the output of the correlator to be $R_{xr}(\tau) = R_{xx}(\tau)$. Then with the rectangular chip pulses, the transmitted and received signals are expected to be the same as those shown in Fig. 5.4.

When we start the system at time zero, at the end of the reception of the first symbol, the sampled output of the correlator is given by

$$R_{xr}(0) = \frac{1}{t_s} \int_0^{t_s} x(t)\tilde{x}(t) dt \approx \frac{1}{T_s} \int_0^{T_s} x(t)\tilde{x}(t) dt = R_{xx}(0)$$

which is what we were expecting. After calculation of this first value of the output signal, since the chip rate at the receiver is ΔR slower than the chip rate of the transmitter, at the start of the second integral, $x(t)$ at the transmitter and its approximation $\tilde{x}(t)$ at the receiver are

$$t_s - T_s = \frac{N}{R_c - \Delta R} - \frac{N}{R_c} = \frac{\Delta R t_s}{R_c}$$

seconds apart, and the second integral results in a sample value of

$$\begin{aligned} R_{xr} \left(\frac{\Delta R t_s}{R_c} \right) &= \frac{1}{t_s} \int_0^{t_s} x \left(t - \frac{\Delta R t_s}{R_c} \right) \tilde{x}(t) dt \\ &\approx \frac{1}{T_s} \int_0^{T_s} x \left(t - \frac{\Delta R t_s}{R_c} \right) \tilde{x}(t) dt = R_{xx} \left(\frac{\Delta R t_s}{R_c} \right) \end{aligned}$$

The next sampled output of the signal represents the next sample of $R_{xr}(\tau)$ evaluated $(\Delta R t_s)/R_c$ seconds later. Following the same pattern, the sampled outputs of the cross-sliding correlator at the receiver taken every t_s seconds in real time will represent samples of $R_{xr}(\tau)$ taken at effective sampling delays of $\tau_s = (\Delta R t_s)/R_c$. In other words, the samples of the sliding correlator output taken every t_s second represent computed samples of $R_{xr}(\tau)$ taken every τ_s seconds in the delay variable. Because $R_{xr}(\tau)$ is a periodic function with period T_s in the delay variable, for every

$$k = \frac{T_s}{\tau_s} = \frac{R_c}{\Delta R} \frac{T_s}{t_s} \quad (5.2.14)$$

samples taken at the output of the sliding correlator, the delay between the receiver and the transmitter is increased by T_s seconds in the delay variable, the codes return to their initial time alignment, and one full sample profile of the channel impulse response is measured. In real time it requires t_s seconds to integrate and generate one sample value of $R_{xr}(\tau)$, and we need k samples to construct one period of $R_{xr}(\tau)$. Therefore, the measurement time for calculation of k samples of a profile is

$$T_m = k t_s = \frac{R_c}{\Delta R} T_s \quad (5.2.15)$$

seconds. Comparing this measurement time with the measurement time of the direct pulse transmission technique, the reader should realize when using the pulse transmission technique that if we do not include averaging, the measurement time is T_s . When we take averages, that measurement time increases linearly with the number of averaged samples. In practice, the larger the measurement time, the more accurate is the measurement result. The physical limitation of measurement time is the rate of variations of the channel reflected in the coherence time, which is the inverse of the maximum Doppler shift. In practice, to avoid measurements of the channel impulse response being affected by changes in the channel due to movements of the measurement terminal or of nearby objects or people, the measurement time is kept as a fraction of the coherence time of the channel.

In the remainder of this section, we provide some practical sliding correlator implementations used as reported in the literature.

Example 5.4: Sliding Correlator Parameters for Indoor Measurements The earliest use of a sliding correlator for indoor radio measurements at 850 MHz was described by Devasirvatham [Dev84]. This system has been used for a variety of multipath profile measurements in and around buildings. In this system, the transmitter and receiver antennas were sleeve dipoles, and the power into the transmit antenna was 26 dBm. The highest ratio of the output signal to the correlation noise level of the PN sequence was 40 dB. The PN sequence was applied to the carrier using biphase modulation, resulting in triangular autocorrelation pulses. The basic parameters of the sliding correlator used in this system were $R_c = 40$ MHz, $\Delta R = 4$ kHz, and $N = 1023$. Using these parameters, resolution (the width of the ACF) is $2/R_c = 50$ ns and the symbol duration $T_s = N/R_c = 25.6$ μ s. The measurement time calculated from Eq. (5.2.15) is

$$T_m = \frac{R_c}{\Delta R} T_s = \frac{40,000}{4} \times 25.6 = 256 \text{ ms}$$

Considering that T_s is approximately the same as t_s , the approximate number of samples used for each measurement of the channel profile, calculated from Eq. (5.2.14), is

$$k = \frac{R_c}{\Delta R} \frac{T_s}{t_s} \approx \frac{R_c}{\Delta R} = 10,000$$

and the distance between two samples of measurement is

$$\frac{\Delta R t_s}{R_c} = \frac{4 \times 25.6}{40,000} = 2.56 \text{ ns}$$

so it requires close to 20 samples for each triangular cross-correlation pulse generated by the sliding correlator.

In practical implementations of the sliding correlator, the integration time t_s is an arbitrary parameter that can be as low as a few fractions of T_s or as high as several T_s . We start with a desirable number of samples per basic ACF function and for the profile calculated, and from there, given the chip rate of the transmitter and the offset of the receiver, we calculate t_s . In the following example, all the parameters are the same as in Example 5.4. However, we want to have eight samples per duration of each triangular ACF pulse. Since pulses are 50 ns wide, we have a sampling interval of 6.25 ns. $t_s = (6.25 \times 40,000)/4 = 62.5$ μ s and the sampling rate is 16 kHz.

Example 5.5: Practical Implementation of a Sliding Correlator The same basic parameters as in [Dev84] were used by Bultitude et al. [Bul89] for extensive indoor radio measurements at 910 MHz and 1.75 GHz. Figure 5.6 shows the details of equipment used in the 910-MHz coherent measurements. At the transmitter, the HP 5065A provided the reference signal for the transmitter and receiver. The 40-MHz clock from the Rockland 5600 was used with the HP 3760A data generator to generate a PN sequence, which in turn modulated the 70-MHz IF signal provided by the Fluke 6160B. The output of the Fluke was also passed through a 12-times frequency multiplier followed by an amplifier to generate an 840-MHz carrier. The carrier was mixed with the 70-MHz modulated IF signal and passed through an amplifier and a filter with an 80-MHz bandwidth centered at 910 MHz. The modulated signal at 910 MHz

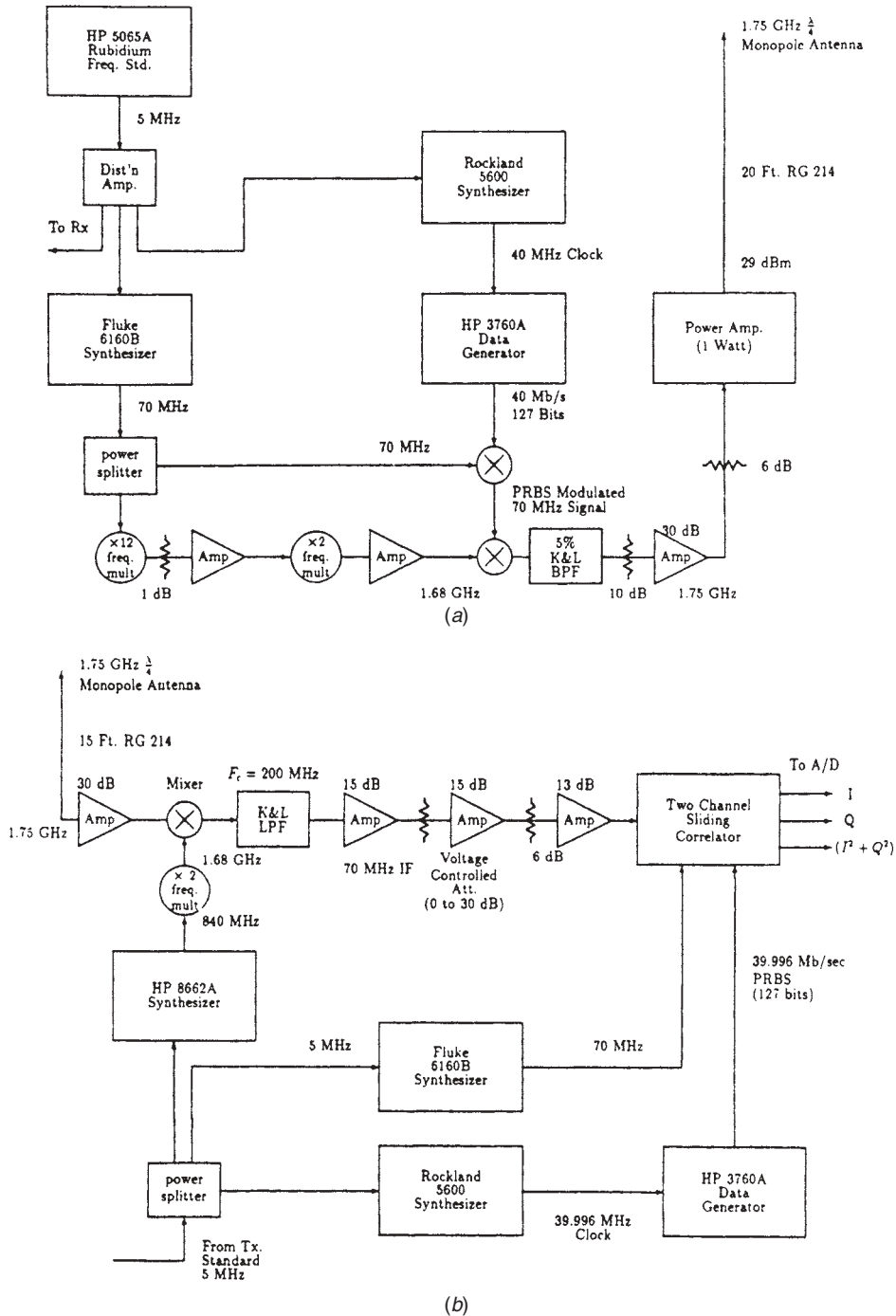


FIGURE 5.6 Details of a spread-spectrum measurement system used for indoor radio propagation studies: (a) transmitter; (b) receiver. (From [Bul89] © IEEE.)

was then amplified and fed through a monopole antenna. Using the reference 5-MHz signal from the transmitter and using circuits similar to those in the transmitter, the reference 840-MHz carrier, a 70-MHz IF reference, and a PN sequence at the rate of $40 \text{ MHz} - 4 \text{ kHz} = 39.996 \text{ MHz}$ were generated. The 840-MHz carrier was mixed with the amplified arriving signal, and the resulting signal was passed through a low-pass filter (LPF) followed by a chain of amplifiers with voltage-controlled gain to generate a 70-MHz IF modulated signal. At the two-channel sliding correlator, the PN sequence was modulated onto in-phase (I) and quadrature (Q) 70-MHz carriers to provide the correlation reference. The received IF signal was mixed with the I and Q references and passed through the LPF integrators. These filters were single-pole RC filters having 3-dB cutoff frequencies of 4 kHz. The output provided the I and Q components and the squared envelope of the received demodulated signal. The I and Q signals were sampled at 16 kHz with a 12-bit A/D converter, and the sampled signal was stored in a digital computer. As we discussed earlier, this sampling rate provided an effective sampling interval of 6.25 ns for the recorded channel impulse responses. Figure 5.7 shows samples of measurements taken with this system.

For mobile radio channel measurements, the differences between the path lengths are larger, and therefore longer excess delay has to be measured, but coarser resolution is acceptable. Typical values of $R_c = 10 \text{ MHz}$, resulting in a resolution of around

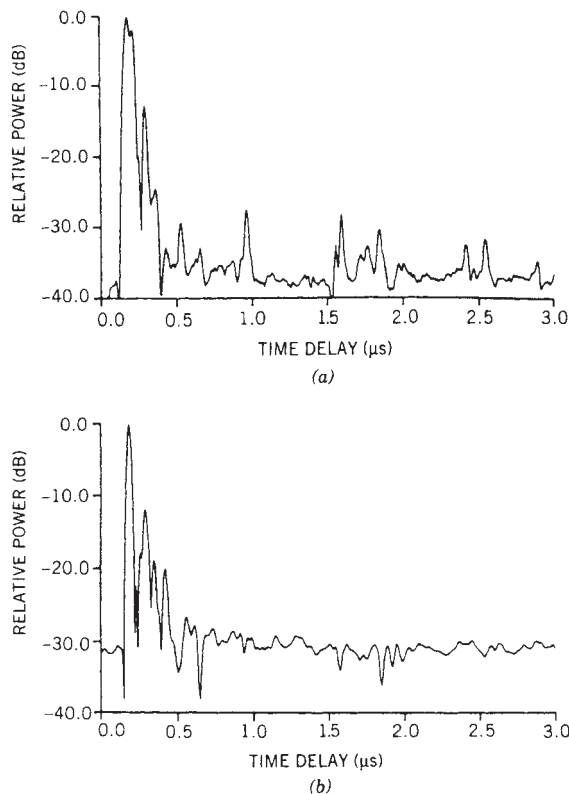


FIGURE 5.7 Samples of measured channel delay profiles using a spread-spectrum system: (a) 900-Hz band; (b) 1.7-GHz band. (From [Bul89] © IEEE.)

0.2 μs , have been used to measure the excess delay of up to $T_s = 102.3 \mu\text{s}$ for a PN-sequence length of $N = 1023$ [Par89]. The ΔR of 5 kHz provides $k = 2000$ samples per profile taken in $T_m = 204.6 \text{ ms}$. With this rate we can capture accurate channel impulse responses only for Doppler shifts up to about 2.5 Hz.

5.2.3 Results of Time-Domain Wideband Measurements

Traditionally, the three parameters of interest extracted from the results of time-domain wideband measurements, given by Eq. (5.2.3), are the received power, mean excess delay or delay spread, and the rms multipath delay spread. For the calculation of these parameters, we need only the magnitude and arrival time of the paths, which can be found from the results of coherent or noncoherent measurements. To calculate these parameters from the magnitude and delay of the arriving paths, we use Eq. (3.4.3) for the received power, $\tau_L - \tau_1$ for the excess delay, and Eq. (3.4.4) for the rms delay spread.

In this section we examine some results of wideband measurements in order to demonstrate their usefulness in comparing the wideband characteristics of various building structures and analyzing the effects of movement and building partitioning. During the early 1990s, several researchers performed wideband measurements of indoor radio propagation at different frequencies. Here we provide an overview of these measurements, with particular emphasis on measurements performed at Worcester Polytechnic Institute. For similar results in outdoor areas, for mobile radio applications, the reader may refer to [Tur72, Par89], and for a survey of other research in indoor radio propagation, the reader may refer to [Has93a].

The behavior of radio propagation in outdoor and indoor areas follows the same general pattern. The mobile radio channel is characterized by a higher multipath spread because of the longer distances involved in mobile communications. The mobile channel also exhibits higher values of Doppler spread because the terminals are intended to operate at vehicular rather than pedestrian speeds. Furthermore, path losses in different buildings exhibit wider variations than those observed in typical outdoor areas.

The measurements performed in various environments are broadly categorized as spatial, local, and mixed measurements. In a *spatial experiment*, measurements are taken at points distributed throughout the test area, such as the floor of a building. This is done by fixing the receiver in a central location and moving the transmitter to various locations. The locations are selected based on the placement or probable placement of communication equipment for planned wireless networks. The surrounding environment is kept stationary during the measurement acquisition time by preventing movements close to the transmitter and receiver. The objective of spatial experiments is to determine the effect of location on the propagation parameters. The objective of *local experiments* is to determine the effect of local movements on the propagation parameters. Local movements are either traffic in the vicinity of the measurement equipment or movements (on the order of a wavelength) of the transmitter or receiver between measurements. A *mixed experiment* is a combination of a spatial and a local experiment; that is, multiple measurements are taken around each location in a spatial distribution of locations.

Samples of Large-Scale Spatial Measurements. In this section we compare the results of spatial measurements taken in five areas in three manufacturing floors as well as in an office environment partitioned into three areas. We compare the measurements

in terms of the distance–power gradient and the characteristics of the rms multipath delay spread. In a spatial measurement, the receiver was fixed in a central location. The transmitter was moved to various locations at each site, such as the probable positions of planned wireless terminals throughout the floor of a building. The received multipath profiles were measured and stored in a computer using the direct pulse transmission system described in Fig. 5.1. Each stored profile is a time average of 64 profiles collected over 15 to 20 s at one location. During the measurement time, care was taken to prohibit movement in the vicinity of the transmitter and receiver. The distance between the transmitter and the receiver varied between 1 and 100 ft. A total of 526 profiles were collected from these measurements.

A manufacturing floor environment is typically characterized by large open areas containing various items of machinery and equipment of different sizes. There are usually no walls between the transmitter and receiver, and a “direct” path is available for most locations. To indicate representative statistical characteristics of such an environment, we show the results of measurements made in five different manufacturing areas [Pah89]. Area A (Infinet Inc., North Andover, Massachusetts) is a typical electronics shop floor having a wide open area containing circuit board design equipment as well as soldering and chip mounting stations. Area B (also at Infinet Inc.) includes test equipment and storage areas for common electronic equipment, partitioned by metallic screens. Area C (Norton Company, Worcester, Massachusetts) is a large open area containing grinding machines, huge ovens, transformers, and other heavy localized machinery. Area D (General Motors, Framingham, Massachusetts) is a car assembly line “jungle” floor, having a dense array of welding and body shop equipment of all kinds. The environment, as presented to radio waves, results in many obstructions by, and signal reflections from, the various objects. Area E (also in the General Motors plant) is a vast open area used for final inspection of new cars coming off the assembly line. This area has many LOS paths between the transmitter and the receiver. The numbers of averaged profiles collected from the five areas were 54, 48, 75, 45, and 66, respectively, resulting in a total of 288 profiles representing the manufacturing environments. Table 5.1 provides short summary descriptions of the five manufacturing areas.

TABLE 5.1 Short Description of the Results of Wideband Time-Domain Measurements in Five Manufacturing Areas and Three Offices

Measurement Area	Number of Gradient Locations	Distance–Power Delay Spread, α	Maximum RMS Delay Spread (ns)	Median RMS Delay Spread (ns)	Mean RMS Fluctuations (ns)	Range of Power (dB)
A	54	2.348	40	15.29	16.64	30.34
B	48	3.329	60	31.62	29.03	39.85
C	75	2.185	152	48.90	52.38	35.50
D	45	2.196	150	52.57	73.13	28.02
E	66	1.398	146	19.37	33.13	24.97
F	54	1.76	48	12.40	15.75	18.0
G	96	2.05	55	44.19	39.53	24.50
H	88	4.21	146	50.3	55.19	28.53

The typical office environment has less open space, and for most sites the “direct” path is obstructed by the presence of one or more walls. The environment, as presented to radio waves, has many reflections from the walls and ceilings. To represent the statistical characteristics of such an environment, three different office areas are considered. The office areas (areas F to H) discussed in this set of spatial measurements are located on the third floor of the Atwater Kent Laboratories at Worcester Polytechnic Institute. The floor plan was shown in Fig. 4.2. For these measurements, the receiver is located in the center of Room 317, an electronics laboratory comprising typical equipment such as oscilloscopes, voltmeters, and power supplies on wooden benches. Area F is inside this laboratory, and hence all test locations in this area have a direct LOS to the receiver. Area G is the corridor (300I, D, G, and E) around this electronics laboratory, separated in most parts by a sheetrock wall with metal studs and some glass windows. Area H consists of all the office rooms, 301–311, on the other side of the corridor, each having typical modern office equipment, including a personal computer and a printer. This area is separated from the receiver by at least two walls of sheetrock and some glass windows in each wall. All the rooms in this area are very similar in structure and size. A total of 234 profiles collected from these areas included 84 in the offices, 96 in the corridor, and 54 inside the electronics laboratory.

Figures 5.8 and 5.9 show the cumulative distribution functions of the rms delay spread for the manufacturing floor areas (A–E) and the college building areas (F–H), respectively. Table 5.1 lists the maximum, median, and mean values of the rms delay spread measured in all the areas. Area A, which is a very open space, has the lowest mean and maximum rms delay spread among all the manufacturing floor areas. Area B exhibits higher values of the mean and the maximum rms delay spread. This is because, in most instances, the direct LOS path is obstructed by metallic objects, and therefore the received signal is composed of several reflections. Areas C and D have numerous pieces of machinery in a small localized area, and thus have higher values for the rms delay spread. Areas E and F have very few metallic structures in the large open area and thus lower median values for the rms delay spread. Areas G and H are, for the most part, obstructed by one or two walls and metal doors, which results in higher values of the rms delay spread. Because of dense local reflections from the body shop and the welding equipment, area D has the highest average rms delay spread.

The average value of the rms delay spread is thus dependent on (1) the availability of a “direct” LOS path between the transmitter and the receiver, (2) the size of the site, (3) materials used for the walls and the ceiling of the building, and (4) the objects in the area surrounding the transmitter and the receiver, locally and globally. For example, in Figs. 5.8 and 5.9, for the points to the left of the dashed “horizontal” line segments, an unobstructed “direct” path between the transmitter and the receiver was available. On the other hand, the LOS path between the transmitter and receiver was blocked by metallic objects for the points to the right of the dashed “horizontal” line segments, leading to higher values of the rms delay spread. Another interesting fact regarding areas F and G is that there is an increase in the average delay spread value due to signal propagation through one wall. When the signal had to propagate through two walls, the average rms delay spread increased further. This increase in the rms delay spread, observed when signal propagation occurs through one or more walls, is useful in assigning data rates to each cell when a cellular indoor radio system is designed.

Table 5.1 also lists the values of the distance–power gradient α obtained for all the areas. The line-fitting method described earlier for narrowband measurements is used

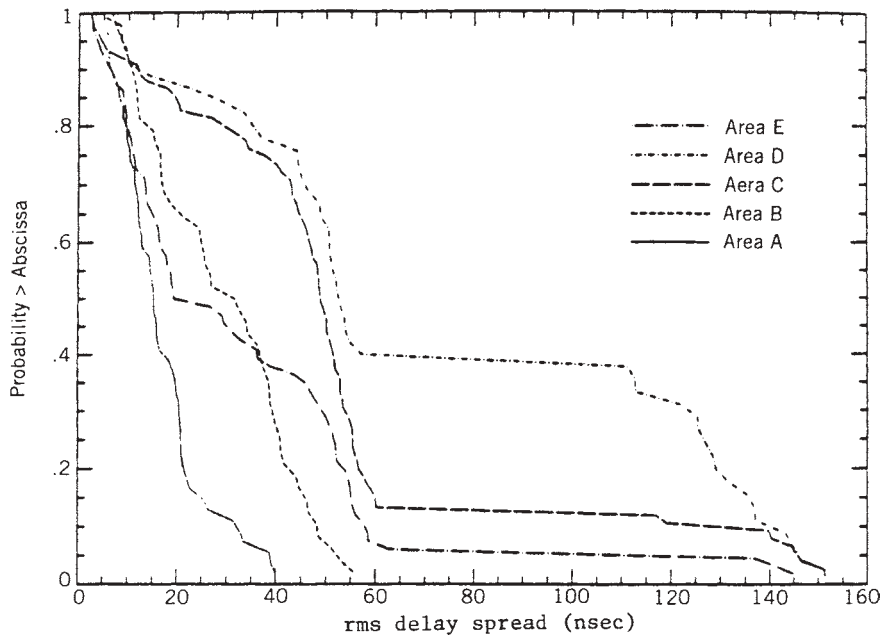


FIGURE 5.8 CDF of the rms delay spread of the measurements in five manufacturing areas [Gan93].

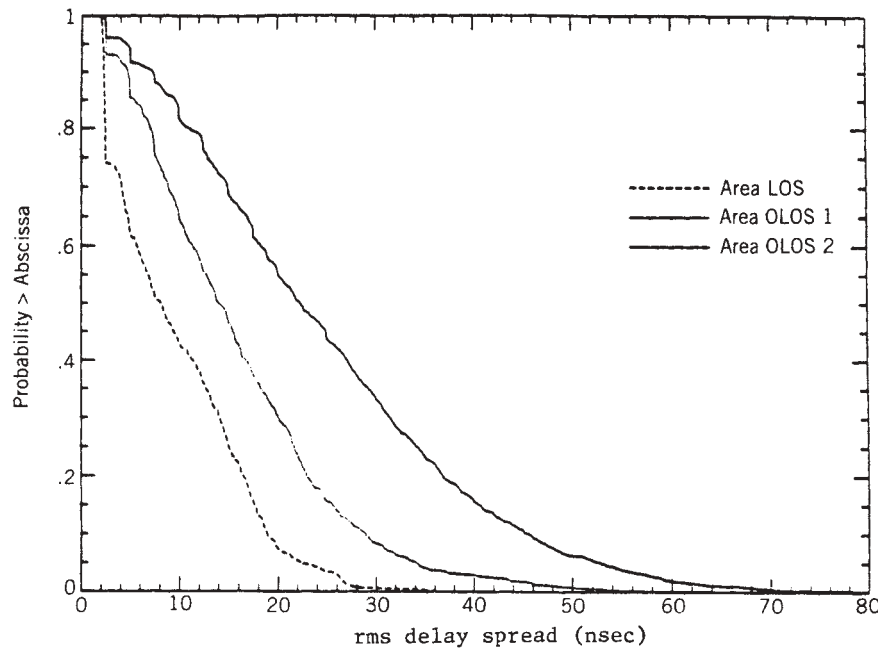


FIGURE 5.9 CDF of the rms delay spread of the measurements in three office areas [Gan93].

again here for the wideband power data. As described in Chapter 3, the distance–power gradients obtained from narrowband and wideband measurements are expected to be the same. Areas A, C, D, and G exhibited values of α between 2 and 2.5. This is due to the open areas available for unobstructed signal propagation and relatively few local surrounding objects taller than the antennas. Areas E and F had vast open areas, and they exhibited an α value less than that of the theoretical α for free space. These open areas had very few objects in the local vicinity of the transmitter and receiver and generally afforded LOS paths between the transmitter and receiver. Area B included partitioning by metallic screens, and the transmitted signal was scattered locally by numerous pieces of equipment in the vicinity of the transmitter. This contributed to a higher value of α . Area H was separated from the receiver by two walls, and there were many local reflections by the walls and the ceiling inside the rooms. The highest value of α was observed in area H. In manufacturing floor environments, although signal propagation through the walls is very unlikely, the presence of a significantly large number of local reflecting objects may cause the value of α to rise. On the other hand, in office environments, the presence of a large open space is less likely, but signal propagation through the walls greatly influences the value of α .

Most wideband indoor radio propagation studies in various buildings report maximum rms multipath delay spreads of around 100 ns [Sal87b, Dev91, Has93b]; higher values are also reported in [Dev87, Rap89]. The rms delay spreads reported for mobile radio channels are on the order of microseconds without distant reflectors such as hills [Cox72, Cox93] and are on the order of several tens of microseconds if there are distant reflectors [Rap90]. The excess delay spread for an indoor radio channel is usually on the order of several hundred nanoseconds, and in urban areas it is typically on the order of several microseconds [Par89, Tur72] without distant reflectors, and around 100 μ s with distant reflectors [Rap90].

Sample Measurement of Temporal Variations in Wideband Characteristics. Local measurements are performed to determine the channel variations observed over a short time period at a fixed location of the terminal. Such variations are induced experimentally by having people moving about in the vicinity of the fixed transmitter/receiver antenna or by manually shaking the antenna on its stand. The objective of these experiments is to compute and compare the statistics of rms delay spread and received wideband power observed in the multipath profiles for these variations. We now describe two sets of experiments performed to induce such variations, in one LOS and one obstructed LOS (OLOS) environment [Gan91a, Gan91b, Gan93]. The first set involved two persons walking briskly around the transmitter and receiver, labeled as experiments A (LOS) and C (OLOS). The second set involved a person “shaking” and “wiggling” the transmitter antenna on its base, labeled as experiments B (LOS) and D (OLOS). These experiments were made with both the transmitter and receiver stationary on the third floor of the Atwater Kent Laboratories. For the LOS experiments, both the transmitter and receiver were located in the central electronics laboratory. For the OLOS experiments, the receiver was located in the communications research lab, comprising typical office furniture and computers as well as radio communication equipment; the transmitter was located in a computer laboratory separated from the receiver by two walls having glass windows. The walls were made of plasterboard with metal studs.

For all four sets of data, the distance between the transmitter and the receiver was fixed at 10 m. A total of 400 profiles were collected from the four experiments. An

adequate sampling rate of 20 samples/s was used to properly sample the short-time variations. While the profiles were being stored during the experiment, care was taken to prohibit any other kind of activity or movement in the vicinity of the experiment.

The rms delay spread and the received power versus time were computed for each of the 100 profiles obtained within each set of experiments (A, B, C, and D). Figure 5.10 shows the variations in the rms delay spread for the four experiments. For the LOS experiments A and B, the variations are about 40 ns, whereas for the OLOS experiments C and D, they are about 20 ns. The standard deviations of the rms delay spread for LOS sets A and B are 9.2 and 12.8 ns; for the OLOS sets C and D they are 3.7 and 5.7 ns. Thus, local variations of the rms delay spread in LOS channels, caused by pedestrian traffic, are greater than those observed in OLOS channels. Also, on average, variations in the delay spread caused by local pedestrian traffic near the antennas were smaller than those observed for movements of the antenna.

Figure 5.11 shows the temporal fluctuations of the received multipath power for the four experiments. The range of short-time fluctuations in the multipath power was 7 to 9 dB in LOS and 5 to 6 dB in OLOS experiments. The standard deviations of fluctuations in multipath power for all the data sets were around 1 dB. These variations are far below the variations observed for similar experiments for narrowband communications discussed in Section 4.3.2. This conclusion is consistent with the observation made from the results of two-dimensional ray tracing in Examples 3.4 and 3.6. Generally, local and temporal variations of the power are affected by the bandwidth of the communication system. As the bandwidth increases, local and temporal variations decrease. The CDF of the multipath power from each of the foregoing experiments was compared with the lognormal and Rayleigh distributions, and the results were shown to fit the lognormal distribution [Gan93]. The results of temporal and local variations of wideband signals did not exhibit Rayleigh characteristics, however. The multipath

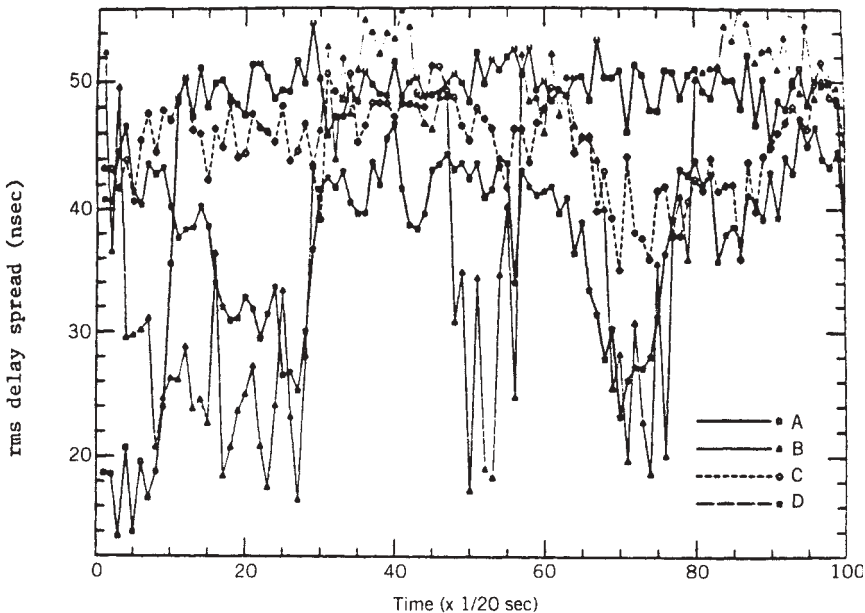


FIGURE 5.10 Temporal variations of the rms delay spread measured during four experiments.

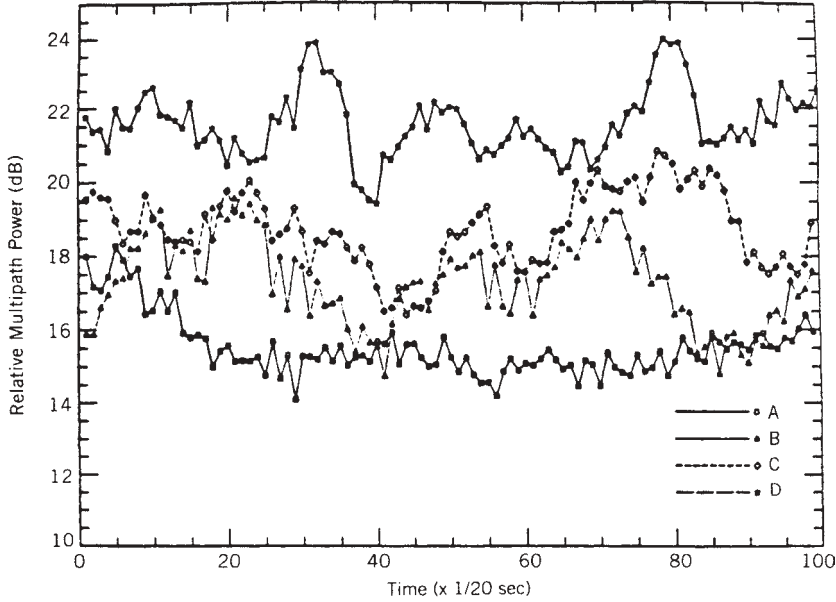


FIGURE 5.11 Temporal variations of the wideband received power measured during four experiments.

power at each location is the sum of squared magnitudes of the path amplitudes, which is independent of the phases of the paths. In narrowband signaling, the phase differences among the arriving paths produce Rayleigh-distributed multipath fading. The power in the wideband signals is averaged over faded and unfaded frequencies, and thus the frequency-selective fading is averaged over the entire band.

5.3 FREQUENCY-DOMAIN MEASUREMENT TECHNIQUES

In frequency-domain measurements of radio propagation characteristics, the frequency response of the channel is measured directly. If we consider Eq. (5.2.1) as the objective of time-domain measurements, in the frequency domain we measure the Fourier transform of this function on the delay variable, given by

$$H(f, t) = \int_{-\infty}^{\infty} h(\tau, t) e^{-j\omega\tau} d\tau = \sum_{i=1}^L \beta_i(t) e^{-j\omega\tau_i(t)} e^{-j\phi_i(t)}$$

In a slowly time-varying channel, the multipath parameters of the channel remain constant during fractions of the coherence time of the channel, and during that period we can measure

$$H(f) = H(f, 0) = \sum_{i=1}^L \beta_i e^{-j\omega\tau_i} e^{-j\phi_i} \quad (5.3.1)$$

which is the Fourier transform of the complex envelope of the ideal channel impulse response. In practice, however, the measurement systems are bandlimited, and what we

measure is actually a windowed frequency characteristics of the channel, defined as

$$H(f) = H(f, 0) = W(f) \sum_{i=1}^L \beta_i e^{-j\omega\tau_i} e^{-j\phi_i}$$

in which $W(f)$ represents the frequency-domain characteristics of the RF filter used in the measurement system.

In practice, frequency-domain measurement is performed by using a network analyzer that sweeps a transmitted RF frequency with fixed power over the bandwidth of interest in the channel and measures the amplitude and phase of the received signal passed through the radio channel. This frequency response is then used for calculation of the time-domain response and multipath parameters using an inverse Fourier transform technique.

5.3.1 Measurement Using a Network Analyzer

Figure 5.12 shows the basic principles of operation of a frequency measurement system using a network analyzer and the details of how the measured signal is translated into time-domain channel profiles. The network analyzer sweeps the channel at discrete frequencies $f_k = f_0 + k \Delta f$, $0 \leq k < N$, with equal increments of Δf , at the transmitter antenna and measures N complex samples of the frequency response $H(k)$, $0 \leq k < N$. If we assume that $f_0 = 0$, the samples are the baseband complex frequency response of the channel, given by Eq. (5.3.1), and we have

$$\begin{aligned} H(k) &= H(f)|_{f=k\Delta f} = \sum_{i=1}^L \beta_i e^{-j\omega\tau_i} e^{-j\phi_i}|_{f=k\Delta f} \\ &= \sum_{i=1}^L \beta_i e^{-j2\pi k\Delta f\tau_i} e^{-j\phi_i}, \quad 0 \leq k < N \end{aligned} \quad (5.3.2)$$

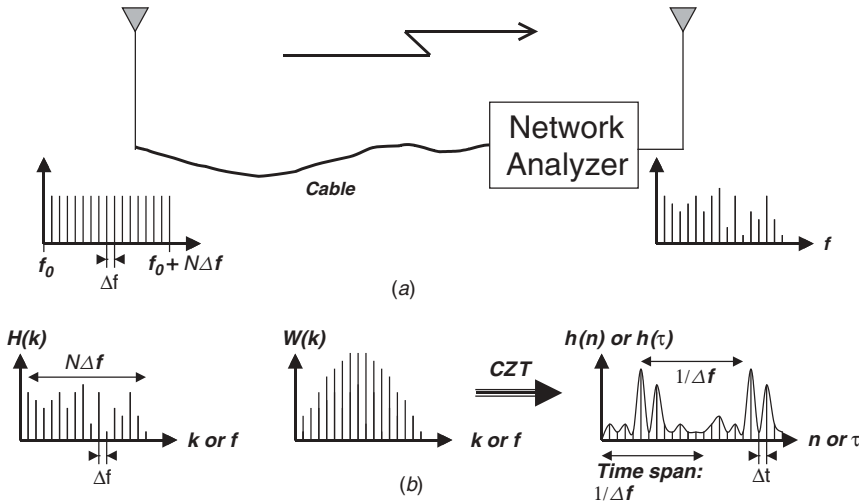


FIGURE 5.12 (a) Basic block diagram of frequency-response measurements using a network analyzer measurement system; (b) postprocessing of data for the calculation of periodic impulse response.

Therefore, the inverse discrete Fourier transform of these measured samples is the coherent baseband impulse response or complex envelope of the channel, from which we can measure the multipath parameters. The impulse response of the channel determined from the inverse discrete-time Fourier transform (DTFT) of the sequence $H(k)$, $0 < k \leq L$, including the effects of filtering in the frequency domain $W(k)$, is given by

$$h(\tau) = \frac{1}{N} \sum_{k=0}^{N-1} W(k)H(k)e^{j2\pi\tau \times k\Delta f} \quad (5.3.3)$$

which is a periodic function of t with a period of $T_m = 1/\Delta f$, that is, the span of time that one can measure with a frequency-domain measurement system. In practice, one has to select Δf so that T_m is larger than the multipath spread of the channel, $\tau_1 + \tau_L$, plus other delays introduced by the cables connected to the antennas. Otherwise, signal aliases appear, degrading the quality of the measurements. If Δf is too small, however, for computation of samples of Eq. (5.3.3) we can calculate only for a region of interest. In practice, as shown in Fig. 5.12b, we calculate the discrete-time impulse response of the channel given by

$$h(n) = h(\tau)|_{\tau=n\Delta t} = \frac{1}{N} \sum_{k=0}^{N-1} W(k)H(k)e^{j2\pi n\Delta t \times k\Delta f}, \quad T_{\text{start}} \leq \tau < T_{\text{start}} + M\Delta\tau \quad (5.3.4)$$

at arbitrary sampling intervals of Δt and M samples. In the digital signal processing literature, the arbitrary sampling of the DTFT in Eq. (5.3.4) is sometimes referred to as the *chirp Z-transform* (CZT).⁴

Example 5.6: Parameters for Measurement in an Indoor Area Assume that we want to design the parameters for a 100-MHz frequency-domain measurement system for telecommunication applications that can cover up to 50 m and needs a time resolution of 1 ns between samples. For the distance of 50 m the propagation time that is the arrival time of the first path is $\tau_1 = 50/(3 \times 10^8) = 166.7$ ns. If we use a 60-m cable and assume that the propagation speed in the cable is two-thirds of the free-space propagation, the delay associated with the cable is $\tau_c = 60/(2 \times 10^8) = 200$ ns. If we assume that the maximum delay spread (the difference between the delay of the first and the last paths) in indoor areas of our interest is $\tau_L = 500$ ns, the minimum time span is $T_m = \tau_c + \tau_1 + \tau_L = 166.7 + 200 + 500 = 866.7$ ns and the sampling rate frequency is $\Delta f < 10^9/866.7 = 1.15$ MHz. Taking 100 samples at $\Delta f = 1$ ns will be adequate for frequency-domain parameters. In the time domain, since we have a telecommunication application, we are only keen on the multipath spread. For a 100-MHz bandwidth, the width of pulses in the time domain is around 10 ns. Therefore, if we start our time sampling 10 ns before the start of the profile, we need around 510 samples in time to calculate the multipath profile that covers the 500 ns of interest at a resolution of 1 ns.

⁴If we take N samples of Eq. (5.3.4) at $\Delta t = T_m/N$, we will have the traditional discrete Fourier transform (DFT), which is usually calculated using fast Fourier transform (FFT) algorithms commonly available in MatLAB or other scientific programming software packages.

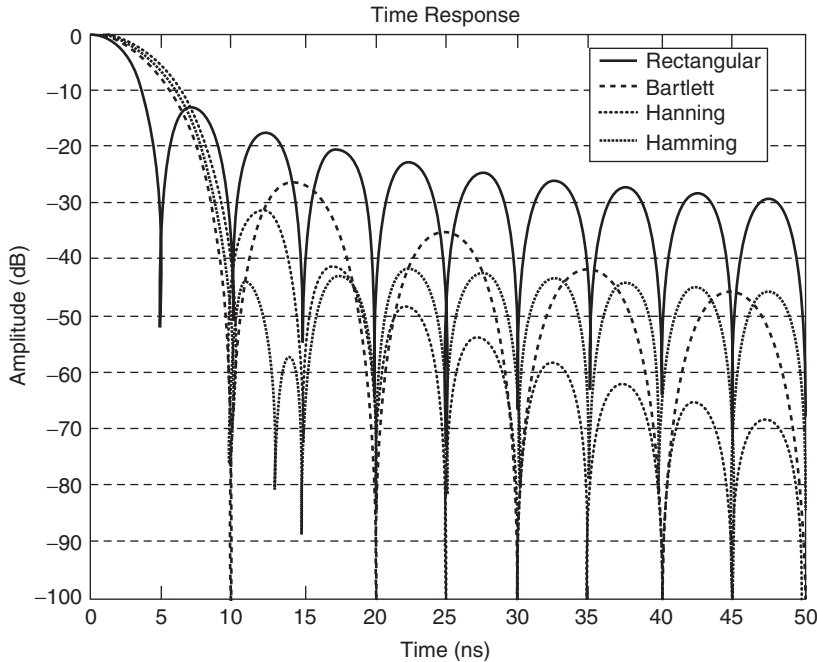


FIGURE 5.13 Frequency response of the filters used for soothing the results of frequency-domain measurements.

Figure 5.13 shows the time response of rectangular, Bartlett (triangular), Hanning, and Hamming windows available in commercial network analyzers. As in any filtering application, the choice among alternatives relies on compromise between the time resolution (base span of a pulse) and the height of the sidelobes. For lower resolutions we can detect more neighboring paths, and for low sidelobes we can protect weaker paths from being buried under the sidelobes of the stronger paths. A rectangular window provides the best resolution in time and the highest sidelobes, about 13 dB below the main peak. The Hamming window provides the best sidelobes, -44 dB below the main peak, with a resolution that is almost twice that of the rectangular window. The following example provides a complete description of the details of a representative system and analysis of the results of measurements.

Example 5.7: Measurement Results in an Indoor Area The measurement system shown in Fig. 4.5, used for the narrowband Doppler-spread measurements discussed in Section 4.4, is used with a different network analyzer setup for wideband frequency-domain measurements [How90b, How90c]. The main component of the measurement system, the HP 8753B network analyzer with a Fourier analysis option, measures the frequency response of the channel and takes its inverse Fourier transform to determine the channel impulse response. The transmitting portion of this frequency-domain measurement system consists of the network-analyzer-synthesized HP 8753B source. The built-in synthesized source of the HP 8753B produces a -15 - to 20 -dBm swept RF signal in the range 900 MHz to 1.1 GHz. The output signal is passed through a cable to be used as the input of a Mini-Circuits ZHL-4240 power amplifier that amplifies

the signal to 30 dBm (1 W). The output of the power amplifier is transmitted using a dipole antenna. The receiving portion of the frequency-domain measurement system consists of the receiving antenna, an attenuator pair, a low-noise amplifier, a high-gain amplifier, and the receiver portion of the network analyzer. The receiving antenna is of the same design as that used for the transmitter. The network analyzer is controlled by a PC with a general-purpose instrumentation bus (GPIB) board. The PC initializes the network analyzer preceding each measurement, and it collects the data at the completion of the measurement. The magnitude and phase of the measured frequency response are the results typically stored for each measurement.

The measurements were made at a set of evenly spaced frequencies, $f_k = f_0 + k \Delta f$, where $f_0 = 900$ MHz is the lowest frequency in the band of interest and $\Delta f = 0.25$ MHz is the frequency-sample spacing. Each frequency sample is measured with the network analyzer dwelling for 0.5 ms at the sample frequency selected. The IF filter bandwidth of the network analyzer is set at 3 kHz. The frequency response consists of 801 complex samples, which require a collection time of 400 ms. From this frequency response a time response of 4000-ns duration is derived using the Fourier transform option. The time response is truncated to show only the portion with significant energy. The 200-MHz bandwidth gives an equivalent resolution of 5 ns in the time domain. This time resolution is comparable to the best resolution obtainable with the time-domain measurement systems reported in [Dev87, Sal87b, Bul89, Pah89, Rap89].

Figure 5.14 shows a plot of the magnitude and phase of a typical frequency response and the corresponding magnitudes of the time-domain response obtained using an

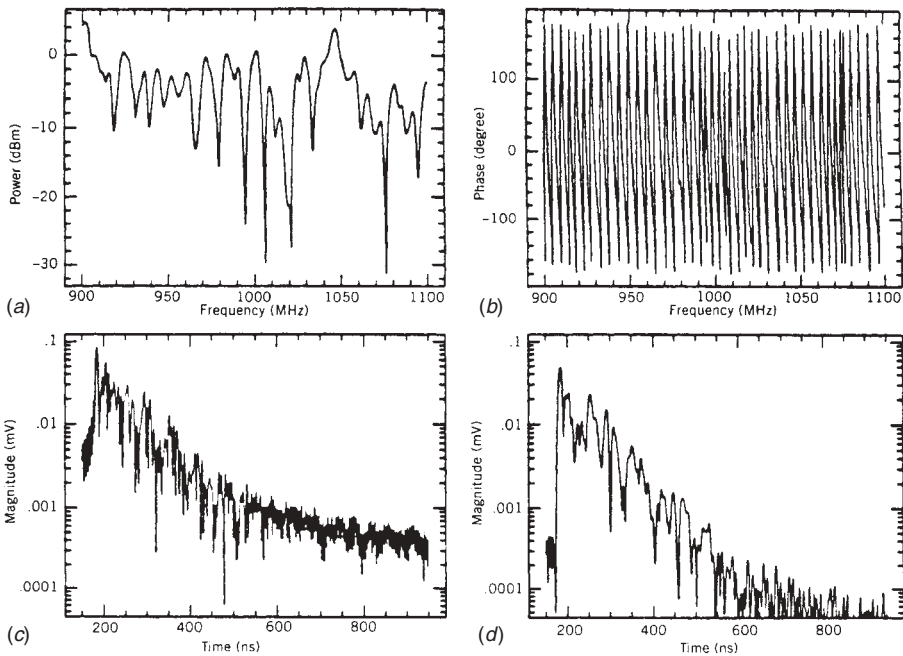


FIGURE 5.14 Results of frequency-domain measurements using a network analyzer: (a) magnitude of the frequency response; (b) phase of the frequency response; (c) multipath delay profile obtained from rectangular windowed frequency response; (d) multipath delay profile obtained from Hamming windowed frequency response.

inverse Fourier transform. The magnitude of the frequency response is shown in decibels, the phase of the frequency response is shown in degrees, and the magnitude of the time response is plotted on a logarithmic scale. In obtaining the time response from the results of frequency-domain measurements, the noise floor (height of the sidelobes) and the resolution of the pulses can be controlled by the type of window applied in the frequency domain before taking the inverse Fourier transform. Figure 5.14c and d show the time responses obtained from a frequency response that has been multiplied by a rectangular window and a Hamming window, respectively. The rectangular window provides the best resolution in the time response (5 ns), but the detection of paths is more difficult because of the -13 -dB sidelobes. The smoother Hamming window provides a time response with sidelobes that are 40 dB down, but it provides a poorer resolution (around 8 ns).

Samples of Frequency-Domain Measurement Experiments. Here we provide the results of some limited spatial, local, and mixed frequency-domain measurements using the system described in Example 5.7 [How90c]. After these experiments a number of other researchers have used variations of this system for indoor radio propagation studies [Win98, Ben99, Spe00, Tin00, Gha03a], and we describe some of these results later. The spatial measurements in [How90c] were taken in 128 locations in two buildings. The area covered in each building is around 50×50 m. The first set of spatial measurements (G1) was made in an IBM office, located on the sixteenth floor of the 32-story Shawmut Bank building in downtown Worcester, Massachusetts. The office consists of a central open area surrounded by adjoining smaller offices. The receiver was placed in a central location, and the transmitter was moved to 70 different locations for the various frequency-response measurements. The second spatial measurement experiment (G2) was performed on the second floor of the three-story Atwater Kent Laboratories at Worcester Polytechnic Institute, shown in Fig. 6.36. The receiver was placed in the central computer terminal room. Measurements were taken from 58 locations near computer terminals in the same room, the adjacent laboratories, a laboratory across a hallway, and offices across another hallway.

For the local-measurement database, two experiments, comprising a total of 60 frequency responses, were performed. The first local experiment (L1) consisted of 28 measurements taken at one location. Measurements were made on the third floor of the Atwater Kent Laboratories, shown in Figure 4.2, with the receiver residing in the Communication Research Laboratory and the transmitter placed in the Electronics Laboratory. The shortest path between the transmitter and receiver passed through three walls. The measurements were taken with people moving in the vicinity of the transmitter or receiver so as to create the maximum amount of variation in the received time response. The second local experiment (L2) consisted of 32 measurements taken with the transmitter in the Electronics Laboratory and the receiver in the Communication Research Laboratory. For the first 16 local measurements the transmitter was fixed, and the receiver was moved over the 16 vertexes of a 67.5×67.5 -cm square with 22.5-cm grid spacing. For the second 16 measurements, the transmitter was moved on a similar grid, while the receiver was fixed.

A mixed experiment (M1) consisting of 621 measurements, nine measurements at 69 spatially distributed locations, was performed on the same floor as was experiment G2. The floor plan and locations for measurements are shown in Fig. 6.36. The receiver was placed in room 1 and the transmitter was moved to 69 different locations, some in the

same room as the transmitter but most in the surrounding rooms. At each location, nine measurements were taken at the positions defined by the nine vertices of a 2×2 -ft square with a 1-ft grid spacing.

The average power for each measurement was calculated by taking the received power at each frequency and averaging over the 801 measurement frequencies. The received wideband power in decibels versus the logarithmic distance was then fitted to a line to determine the distance–power gradient. The areas covered by G1, G2, and M1 give distance–power gradients of 2.6, 2.5, and 2.2, respectively. These results are comparable to results obtained from wideband time-domain measurements in similar areas.

To determine the cumulative distribution function (CDF) of the received power in different experiments, the results of all measurements in an experiment were compared with the theoretical CDFs for lognormal, Weibull, and Rayleigh probability distributions. The lognormal distribution provided the best fit for all the experiments. This observation were consistent with the wideband time-domain measurements reported in [Gan91a, Gan91b]. The spatial power fluctuations due to shadow fading were on the order of 30 dB, which corresponds well with the distances covered. As expected, for the local wideband measurements, the power fluctuations due to multipath fading were small.

Figure 5.15 shows the CDF of the rms delay spread for the five experiments. The means and variances of the rms delay spreads in different experiments are shown in Table 5.2. Using a rectangular window, the mean rms delay spreads obtained for various areas are between 19 and 47 ns. These results agree closely with those of time-domain measurements reported for similar buildings [Sal87b, Bul89, Pah89, Gan91a, Gan91b].

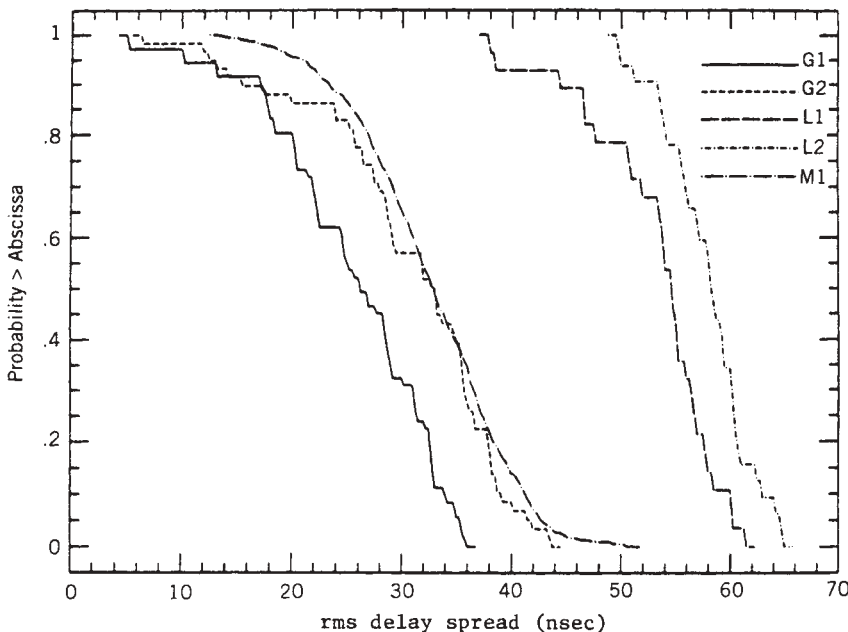


FIGURE 5.15 CDF of the rms delay spread obtained from the results of five experiments using the frequency-domain measurement system.

TABLE 5.2 Summary of the Results of Two Global, Two Local, and One Mixed Measurement Experiment Using the Wideband Frequency-Domain Measurement System

Experiment	Number of Measurements	3-dB Width Median (MHz)	3-dB Width Mean (MHz)	3-dB Width S.D. (MHz)	RMS Delay Spread Mean (ns)	RMS Delay Spread S.D. (ns)
G1	70	5.75	7.74	6.51	19.11	7.56
G2	58	5.25	7.41	7.47	21.03	8.06
L1	28	2.75	3.11	1.13	39.86	9.1
L2	32	2.5	2.56	0.22	46.92	4.74
M1	621	4.75	5.57	2.84	23.1	7.46

With the Hamming window, the mean rms delay spreads increase by 7 to 10 ns for each area; however, the choice of windowing shows no effect on the standard deviations or the received power. The sidelobes of the Hamming window are considerably lower than those of the rectangular window, allowing more paths to be detected. These paths have small amplitude and large delays. Therefore, they can affect the rms delay spread significantly while having a negligible effect on power.

Agreement between the CDFs for experiments G2 and M1 is consistent with the overlap of the measurement areas for those experiments. The average of the rms delay spreads for experiments L1 and L2 is higher than for experiments G1, G2, or M1. In all measurements in the L1 and L2 experiments, there were three walls between the transmitter and receiver, and the measurements were purposely taken in a location with a large τ_{rms} value.

The frequency-domain measurements also provide for experimental measurement of the coherence bandwidth of the channel. The following example provides a statistical approach for measuring the coherence bandwidth.

Frequency-Domain Measurement of the Coherence Bandwidth. The measured samples of the frequency response $H(f_n, t)$ can be interpreted as a random process. The autocorrelation function of this process,

$$R_{Hh}(k, 0) = \frac{1}{N} \sum_{i=1}^{N-K} H^*(f_i, t) H(f_{i-k}, t), \quad k \geq 0$$

is an important function in autoregressive modeling of the indoor radio channel that is discussed in Chapter 6. It is also related to the rms delay spread of the channel. The 3-dB width of $|R_{Hh}(k, t)|$ is a measure of the similarity or coherence of the channel in the frequency domain, which is inversely proportional to the delay spread of the channel and is sometimes referred to as the coherence bandwidth of the channel.

Figure 5.16 shows the CDF of the 3-dB width B_c for each of the five experiments. Table 5.2 gives the median, mean, and standard deviation of the 3-dB width for each experiment. Because it is only one location that is measured repeatedly for each local experiment, variations of 3-dB widths are small compared with those found in spatial experiments. We also note that the local movements of the transmitter in experiment L2 yield the smallest variation.

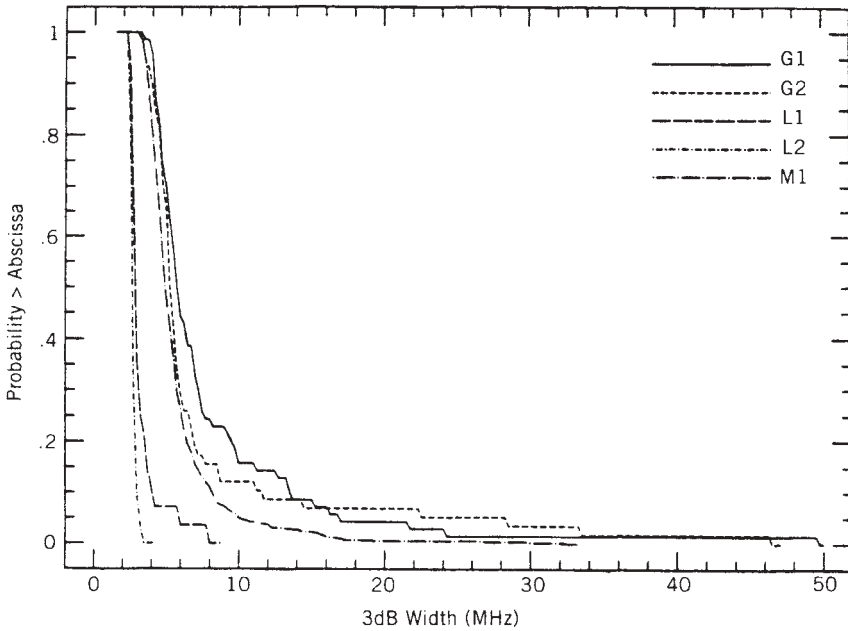


FIGURE 5.16 CDF of the 3-dB width of the frequency correlation function obtained from the results of five experiments using the frequency-domain measurement system.

Using the inverse relationship between width in the frequency domain and duration in the time domain, a relationship of the form $B_c = C\tau_{\text{rms}}^{-\beta}$ between the 3-dB width (MHz) of the frequency correlation function and the rms delay spread (ns) of the channel is determined from a linear regression (on logarithmic scales). Figure 5.17 shows the scatter plot, on a log–log scale, of the 3-dB width of the frequency correlation function versus the corresponding rms delay spread for all experiments. It can be seen that for the spatial experiments, the β values are around 1. Generally, the inverse relationship between the multipath delay spread and the width of the frequency correlation function indicates the direct relationship between the coherent bandwidth of the channel and the 3-dB width of the frequency correlation function.

5.3.2 Comparison Between Measurement Systems

Frequency-domain measurement using a network analyzer is the simplest method of channel measurement for use in small areas. For short distances on the order of several meters, the network analyzer and two antennas can be used, without the need for any additional amplifiers, to measure wideband indoor radio propagation. As the distances increase, we need to add transmitter and receiver amplifiers to the measurement system and increase the length of the cable. The noncoherent pulse transmission technique with envelope detection is also easy to implement. In addition to the amplifier, we need a pulse generator, a mixer, and an oscillator at the transmitter, and an envelope detector at the receiver. The most expensive component in this setup is a digital high-frequency scope used to display and store the channel impulse response. If phase information is needed, the in-phase and quadrature-phase circuitry are also needed, which adds to the

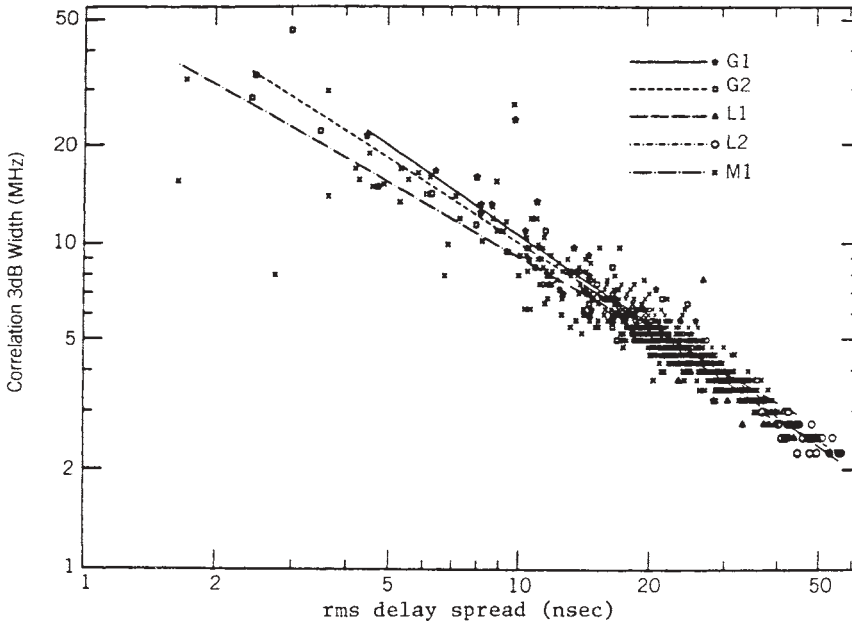


FIGURE 5.17 Width (3 dB) of the frequency correlation function versus the associated rms delay spread and the best-fit line.

complexity of the system. Without phase information the Fourier transform does not provide the correct frequency response of the channel. The spread-spectrum system requires all the circuitry used for the pulse transmission setup, with additional circuits to spread the signal at the transmitter and perform correlation at the receiver.

The ratio of the peak to average transmission power observed with the pulse measurement technique is much higher than with the spread-spectrum and frequency-domain measurement systems. As a result, with the same transmitter and receiver amplifiers, the spread-spectrum and frequency-domain measurement systems have better coverage than the pulse transmission technique. The acquisition time of the pulse transmission technique in the systems mentioned in this chapter was the lowest among the three, and it was the only method with which we could measure temporal variations in the wideband indoor radio channel. The frequency-domain measurement system could measure only the temporal variations of the narrowband signals.

5.4 ADVANCES IN FREQUENCY-DOMAIN CHANNEL MEASUREMENT

Since introduction of frequency-domain measurement techniques in the early 1990s [How90c], frequency-domain measurement techniques have been used in a number of measurement techniques for WLAN and WPAN applications. Frequency-domain measurement systems are easy to design; their bandwidth can easily be adjusted from a narrowband signal up to ultrawideband signals by choosing appropriate antennas, and they can provide onboard filter and Fourier transform capability that would take some time if we have to write our own codes. As a result, these systems are now being used for measurement of the TOA [Ben99], examination of the effectiveness of a

superresolution algorithm, measurements of the angle of arrival of the paths, and measurements of the UWB characteristic of the indoor radio channel. These measurements are needed for the design and performance evaluation of the emerging location-aware broadband ad hoc networks. In the rest of this section we provide an overview of these advances in measurements of radio channel characteristics using frequency-domain channel measurement techniques.

5.4.1 Frequency-Domain Measurement for TOA Measurements

All wideband measurements introduced earlier in this chapter were taken for telecommunication applications, where we were interested in multipath profile and parameters associated with the relative arrival and strength of the individual paths in a profile assuming that the first detected path in the profile is the direct LOS (DLOS) path. In measurement of the TOA for geolocation applications, we are interested in estimation of the arrival time of the DLOS, $\hat{\tau}_1$, from the profile and the difference between this value and the expected arrival time of the DLOS path, $\tau_1 = d/c$, in which d is the distance between the transmitter and the receiver. In practice, the estimated distance measured from the profile is $\hat{d} = c\hat{\tau}_1$, and the difference, $\varepsilon = d - \hat{d}$, is the distance measurement error that is the parameter of interest for geolocation applications. There are two sources for distance measurement error, the error caused by the measurement system and the error caused by the multipath structure of the channel. To have meaningful measurements of the error caused by the channel, the system measurement errors must be very small. In this section we discuss the relation between the distance measurement error and the parameters of a frequency-domain channel measurement system. In Chapter 13 we show multipath conditions causing distance measurement errors, we model them, and we show how we can design algorithms to mitigate these errors.

Figure 5.18a shows a sample channel impulse response calculated from the results of a frequency-domain channel measurement and the measurement parameters that are important for TOA measurement. To create this channel profile using the frequency-domain channel measurement system described in Example 5.7, the frequency response of the channel is measured, and then using the CZT, the discrete channel impulse response is created. To determine the TOA of the DLOS path, using a simple peak detection algorithm the first peak of the channel profile that occurs above the peak detection threshold is determined. The cable delay is subtracted from the arrival time of the first peak to calculate $\hat{\tau}_1$, the estimate of the TOA of the DLOS path. The expected TOA of the DLOS path, $\tau_1 = d/c$, calculated from physical measurement of the distance, is also shown in the figure. This particular figure shows a case in which the transmitter and receiver are in the line of sight (LOS), for which the distance measurement error is very small. Figure 5.19 represents the CDF of distance measurement errors for a number of LOS measurements in the vicinity of the transmitter, where DLOS is dominant and we have minimal effects from the multipath. Therefore, these measurements reflect an upper bound for the errors caused by the measurement system, and they appear to be reasonable. The largest measurement errors occur when the DLOS path is buried under the threshold level, shown in Fig. 5.18, and the receiver interprets the detected first path, erroneously, as the DLOS path. We refer to this situations as undetected direct path (UDP) conditions. As we will see in Chapter 13, although we can reduce the distance measurement error by increasing the bandwidth and using UWB transmissions, the error caused by UDP conditions remains independent of the

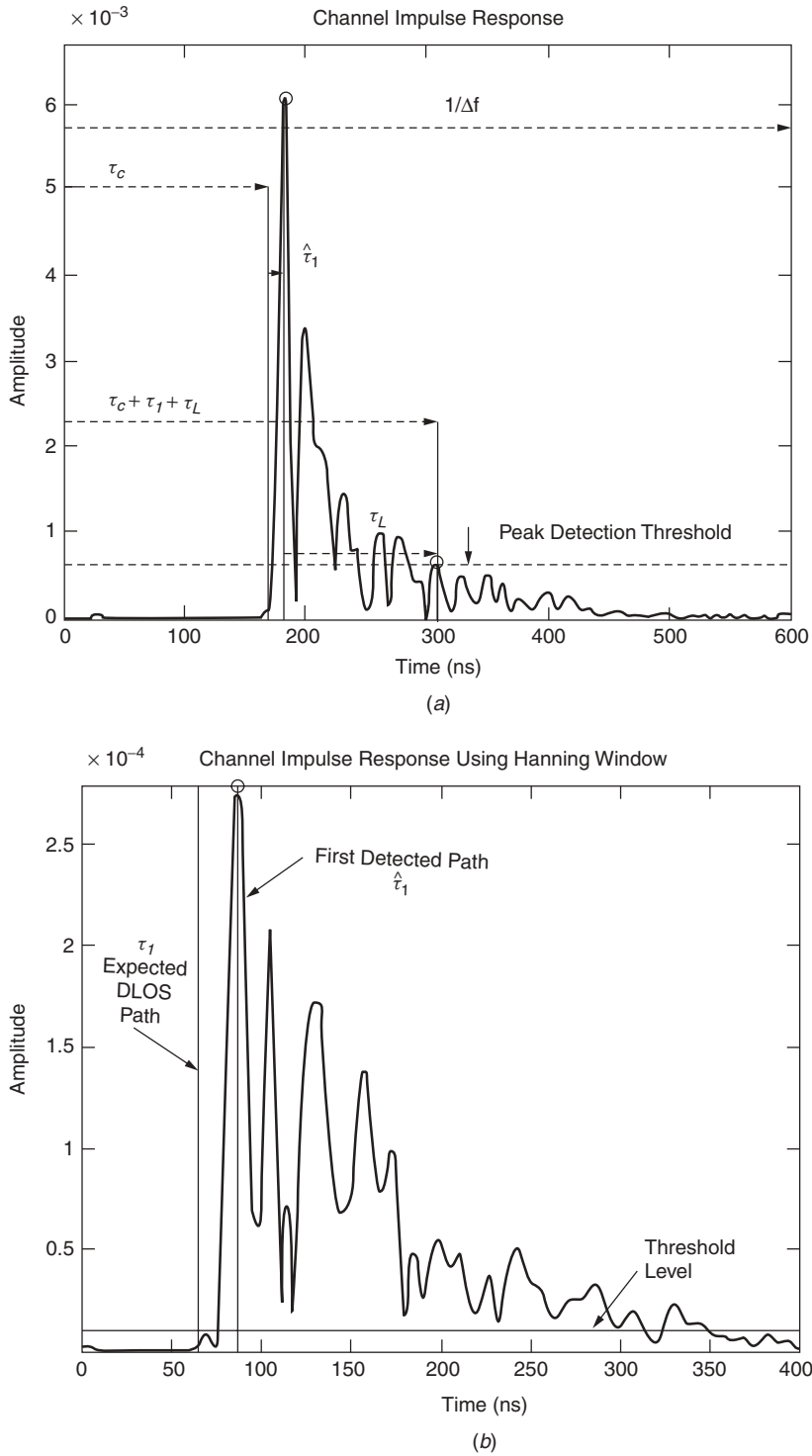


FIGURE 5.18 Parameters for measurement of the TOA in (a) detected direct path condition; (b) undetected direct path condition.

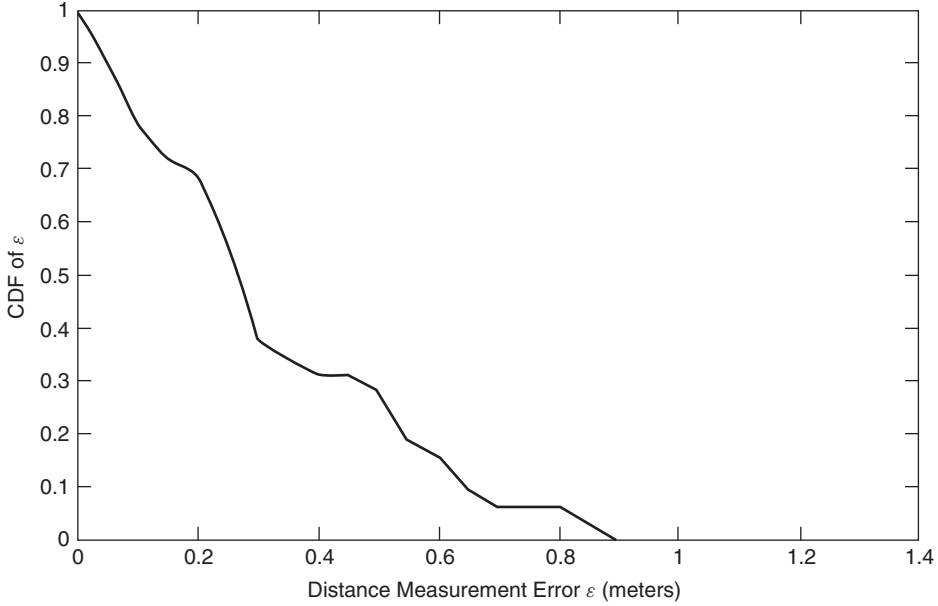


FIGURE 5.19 CDF of distance measurement errors in an LOS area [Zan03a,b].

bandwidth of the measurement system, and for these conditions we may have distance measurement errors as high as a few tens of meters in distances of less than 50 m.

If we neglect the multipath, the only source of the distance measurement error is the fact that our channel profile is calculated using a CZT in which the impulse response is sampled at an arbitrary starting point. Therefore, we have a maximum of $\Delta t/2$ measurement errors for the peak value of the samples from the actual value. For a 1-ns sampling interval, we have a maximum of $(c\Delta t)/2 = (3 \times 10^8)(0.5 \times 10^{-9}) = 0.15 \text{ m} = 15 \text{ cm}$ error. But we cannot eliminate the multipath thoroughly, and even with close distances and LOS connections, we always have some small multipath components that combine with the DLOS path and shift the peak of the first path in the profile slightly. This type of error is caused by limitations in the bandwidth, and they change slightly with the shape of the frequency-domain filter used before CZT operation (see Fig. 5.12). Statistical analysis of the effect of filters on measurement of the TOA in LOS and OLOS areas presented in [Zan03a, Zan03b] reveals that Hamming, Hanning, and Bartlett filters have a similar effect, however. Rectangular windows are not good choices.

5.4.2 Superresolution Algorithms for Frequency-Domain Measurements

The rich body of literature in parametric and nonparametric spectral estimation, developed for analysis of the Fourier transform of a received sampled stochastic signal in the time domain [Man00], can be exploited for the advancement of frequency-domain channel measurement techniques. As shown in Fig. 5.20a, in frequency-domain channel measurements we measure a stochastic process representing the channel response in different frequencies, and we are interested in measurement and modeling of its Fourier transform to determine the occurrence of a number of paths, arriving at different

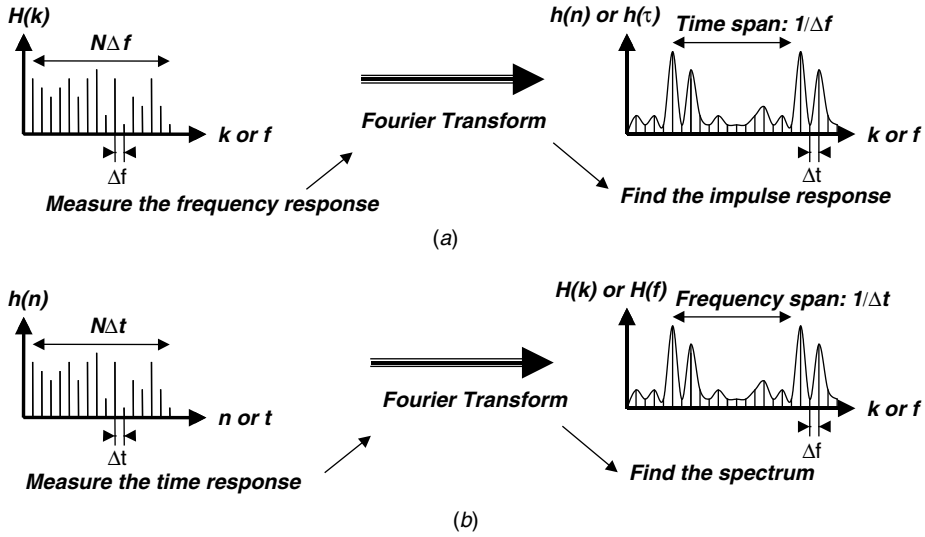


FIGURE 5.20 Similarity between (a) frequency-domain measurement and (b) spectral estimation techniques.

delays. In spectral estimation, as shown in Fig. 5.20b, we measure a stochastic process representing variations in the received signal level in time, and we are interested in measurement and modeling of its Fourier transform to determine the occurrence of a number of frequency components at different frequencies. Naturally, one expects to find useful applications for which the existing body of literature in spectral estimation helps in solving some of the practical problems in the frequency-domain measurement and modeling technique. One of the earliest attempts in this direction was the use of very popular parametric autoregressive modeling in spectral estimation in the indoor radio channel modeling [How90b, Pah90b] discussed in Chapter 6. Another approach was to use superresolution algorithms to refine the time resolution of the measurements of multipath profiles discussed in this section.

Recently, a number of researchers have applied superresolution spectral estimation techniques for multipath radio propagation studies. In [Yam91], the superresolution technique was employed in the frequency domain to estimate multipath time-dispersion parameters, such as mean excess delay and root-mean-square delay spread. A similar method was used in [Lo94] to model indoor radio propagation channels with parametric harmonic signal models. In channel modeling for telecommunication applications, we are interested in the rms delay spread, which does not change drastically with additional resolution in the time-domain signal. Therefore, superresolution algorithms have scientific value. In the estimation of TOA for geolocation applications, however, using superresolution algorithms can improve the accuracy of the TOA estimate. In this section we address the application of superresolution techniques to accurate TOA estimation for indoor geolocation.

In the literature, the time-delay estimation problem has been studied with a variety of superresolution techniques, such as minimum norm [Pal91], root multiple signal classification (MUSIC) [Dum94], and total least-squares-estimation of signal parameters via rotational invariance techniques (TLS-ESPRIT) [Saa97]. Although superresolution

techniques can increase time-domain resolution, they also increase the complexity of system implementation. We present an investigation of frequency-domain superresolution TOA estimation techniques for indoor geolocation in the following subsection.

EV/FBCM Superresolution Algorithms for Indoor Geolocation. Here we summarize a form of the MUSIC algorithm, EV/FBCM (eigenvector forward–backward correction matrix), used as a superresolution technique in TOA estimation for indoor geolocation [Li03a]. The impulse response of a radio channel in the time domain is given by Eq. (5.2.5) and its Fourier transform with Eq. (5.3.1). In frequency-domain measurements we measure samples of frequency response of the channel given by Eq.(5.3.1), and we intend to determine the time of arrival, magnitude, and phase of the individual paths given in Eq. (5.2.5). In the spectral estimation literature, one common approach to solving this problem is to assume a harmonic model for the signal in the time domain. A harmonic signal model can be created by exchanging the role of the time and frequency variables in Eq. (5.3.1), which yields

$$H(\tau) = \sum_{i=1}^L \beta_i e^{-j2\pi f_i \tau} e^{-j\phi_i}$$

This model is well known in the spectral estimation field [Man00]. Therefore, any spectral estimation techniques that are suitable for the harmonic model can be applied to the frequency response of a multipath indoor radio channel to perform time-domain analysis. In [Li03a,b], the MUSIC algorithm is used as a spectral estimation technique to convert the frequency-domain data into the time-domain profile needed for determining the DLOS path and TOA. The discrete measurement data are obtained by sampling the channel frequency response $H(f)$ at L equally spaced frequencies. Considering additive white noise in the measurement, the sampled discrete frequency-domain channel response is given by

$$x(l) = H(f_l) + w(l) = \sum_{k=1}^L \beta_k e^{-j2\pi(f_0 + l\Delta f)\tau_k} e^{j\phi_k} + w(l)$$

where $l = 1, \dots, L$ and $w(l)$ denotes additive white measurement noise with zero mean and variance $(\sigma_w)^2$. The signal model in vector form is

$$\mathbf{x} = \mathbf{H} + \mathbf{w} = \mathbf{V}\mathbf{a} + \mathbf{w} \quad (5.4.1)$$

where $\mathbf{V} = [\mathbf{v}(\tau_1) \quad \mathbf{v}(\tau_2) \quad \dots \quad \mathbf{v}(\tau_{L_p-1})]^T$ and $\mathbf{v} = [1 \quad e^{-j2\pi\Delta f\tau k} \quad \dots \quad e^{-j2\pi(L-1)\Delta f\tau k}]^T$. The MUSIC superresolution algorithm is based on eigendecomposition of the autocorrelation matrix of the signal model in (5.4.1). The autocorrelation matrix is given by

$$\mathbf{R}_{xx} = E\{\mathbf{x}\mathbf{x}^H\} = \mathbf{V}\mathbf{A}\mathbf{V}^H + \sigma_w^2 \mathbf{I}$$

where $\mathbf{A} = E\{\mathbf{a}\mathbf{a}^H\}$ and the superscript H represents the Hermitian, conjugate transpose of a matrix. Therefore, the L -dimensional subspace that contains the signal vector $\mathbf{x}$ is split into two orthogonal subspaces, known as the *signal subspace* and *noise*

subspace, by the signal eigenvectors and noise eigenvectors, respectively. Since the vector $\mathbf{v}(\tau_k)$, $1 \leq k \leq L$, must lie in the signal subspace, we have

$$\mathbf{P}_w \mathbf{v}(\tau_k) = 0$$

where $\mathbf{P}_w \mathbf{v}(\tau_k)$ is the projection matrix of the noise subspace. Thus, the multipath delays τ_k , $1 \leq k \leq L$, can be determined by finding the delay values at which the following MUSIC pseudospectrum achieves its maximum value,

$$S_{\text{MUSIC}}(\tau) = \frac{1}{\|\mathbf{P}_w \mathbf{v}(\tau)\|^2} = \frac{1}{\sum_{k=L_p}^{L-1} |\mathbf{q}_k \mathbf{v}(\tau)|^2}$$

where $\mathbf{q}_k$ are the noise eigenvectors. In practical implementation, when only one snapshot of length N is available, the data sequence is divided into M consecutive segments of length L , and the estimate of the correlation matrix is further improved using the *forward-backward correlation matrix* described in [Pal91]. In the analysis provided in [Li03a], a slight variation on the MUSIC algorithm is used, which is known as the *eigenvector (EV) method*. The pseudospectrum is defined as

$$S_{\text{EV/FBCM}}(\tau) = \frac{1}{\sum_{k=L_p}^{L-1} (1/\lambda_k) |\mathbf{q}_w^H \mathbf{v}(\tau)|^2} \quad (5.4.2)$$

where λ_k , $0 \leq k \leq L$, are the noise eigenvalues. Effectively, the pseudospectrum of each eigenvector is normalized by its corresponding eigenvalue. The performance of the EV method is less sensitive to inaccurate estimation of the parameter L , which is highly desirable in practical implementation.

Figure 5.21 represents two sample frequency-domain measurements of the indoor radio channel with three different postprocessing techniques. The first postprocessing algorithm uses the traditional inverse Fourier transform (IFT) using a Hanning window. The second algorithm simulates a DSSS system by using a raised-cosine pulse with a roll-off factor 0.25 rather than a Hanning window. As shown in Chapter 7, raised-cosine pulses are common in the implementation of RF filters used in radio modems. The third algorithm is EV/FBCM, given by Eq. (5.4.2). Figure 5.21a represents a case in which the DLOS path is detectable, and Fig. 5.21b represents an UDP condition. In both cases, the superresolution algorithm resolves many more paths, allowing a more accurate estimate of the TOA of the DLOS path. The discussion above closely follows [Als03, Als04], in which more extensive results of measurements and post-processing are available. More details of EV/FBCM are available in [Li03b]. More detailed analysis of the results of frequency-domain measurement and superresolution algorithms is provided in Chapter 13.

5.4.3 Frequency-Domain Measurement of the Angle of Arrival

The measurement techniques described earlier in this chapter do not resolve the angles from which signal components arrive at the receiving antenna. Knowledge of the angle of arrival is required for evaluation of evolving MIMO systems that employ smart

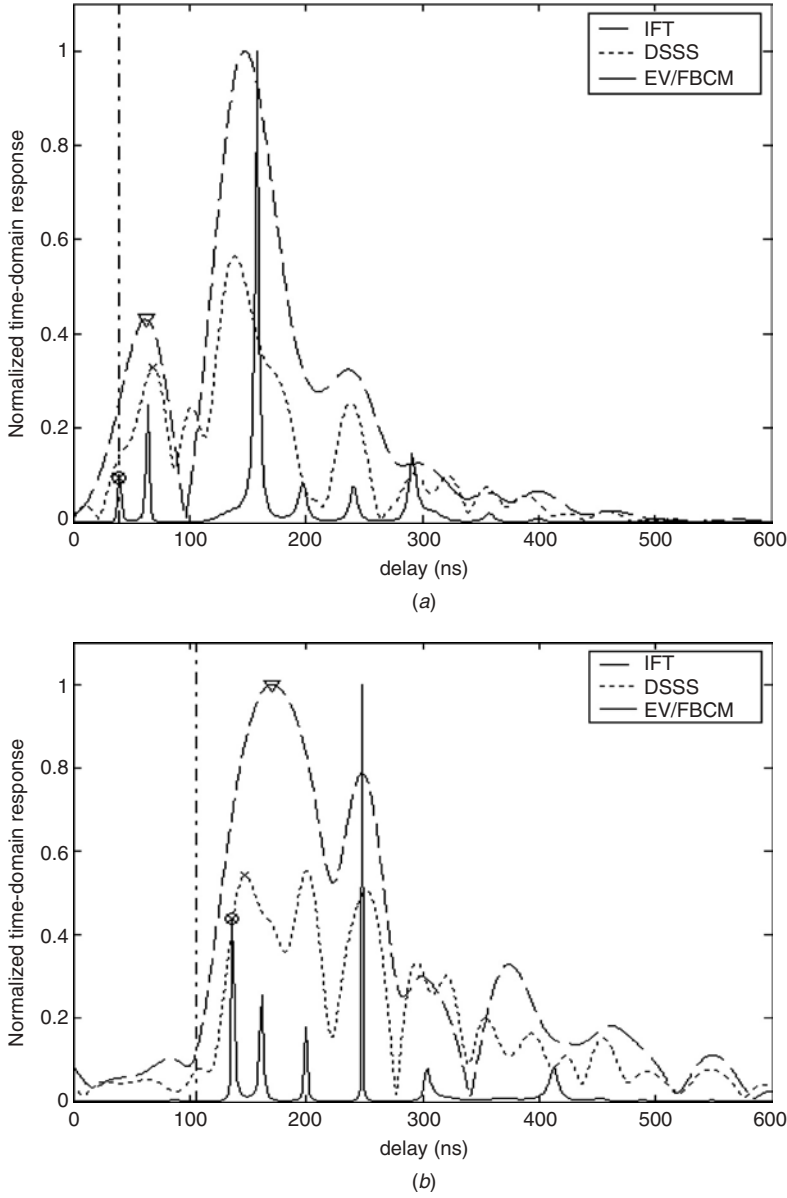


FIGURE 5.21 Effectiveness of the EV/FBCM superresolution algorithm: (a) DLOS path is detectable; (b) DLOS path is buried under the noise.

antennas and space-time coding techniques to fully exploit the capacity of existing bands for wireless communications. If we consider the angles of arrival of different paths, as we discussed in Section 3.6.3, the overall impulse response of the channel, given by Eq. (5.2.5), will be modified to

$$h(\tau, \theta) = \sum_{i=1}^L \beta_i e^{j\phi_i} \delta(\tau - \tau_i) \delta(\theta - \theta_i) \quad (5.4.3)$$

in which θ_i is the angle of arrival of the i th path with arrival delay, amplitude, and phase of τ_i , β_i , and ϕ_i , respectively. Then the measurement system has to be capable of associating arriving paths with the angles at which they arrive. Obviously, omnidirectional antennas used in the measurement systems discussed earlier in the chapter cannot differentiate the angles of arrival of the paths, and we need a directional antenna system.

The first wideband indoor radio measurements to include angle-of-arrival information were reported by Lo and Litva in 1992 [Lo92]. They employed a measurement system wherein a 950-MHz carrier was phase-modulated by a 40-Mb/s pseudonoise sequence. Each data set was acquired by moving the receiving element through a succession of four points on the radius of a circle, as well as to the center. Bearing information was calculated from each data set by considering the five points to constitute a simple array and applying standard beamforming techniques.

This array measurement system was improved in [Ros97], with the addition of a radial armature that facilitates automated movement of the receiving element through the various positions. This system also developed bearing estimates using standard beamforming techniques. However, since the measurements in each data set are collected in one increment rather than 90 increments, the system features a much finer angular resolution of 10° . A significantly higher-performance space-time characterization system presented by Spencer et al. [Spe00], shown in Fig. 5.22, may be viewed as an extension of the frequency-domain channel sounder of Howard and Pahlavan [How90c] described in Example 5.7. The system replaces the omnidirectional receiving element employed in [How90c] with a 60-cm parabolic dish mounted atop a computer-controlled turntable. Each data set is acquired by sweeping the network analyzer, followed by dish rotation through 2° , and repeating this sequence until one cycle has been completed. The authors report achieving spatial resolution of 6° , and temporal resolution of 3 ns. Since each data set requires approximately 20 minutes to acquire, due to the setup, dish rotation, and large number of constituent measurements, it may prove difficult to ensure that all constituents are derived from the same underlying propagation scenario. Perhaps more important, since the beamwidth of a parabolic antenna varies inversely with the frequency of operation, an antenna nearly

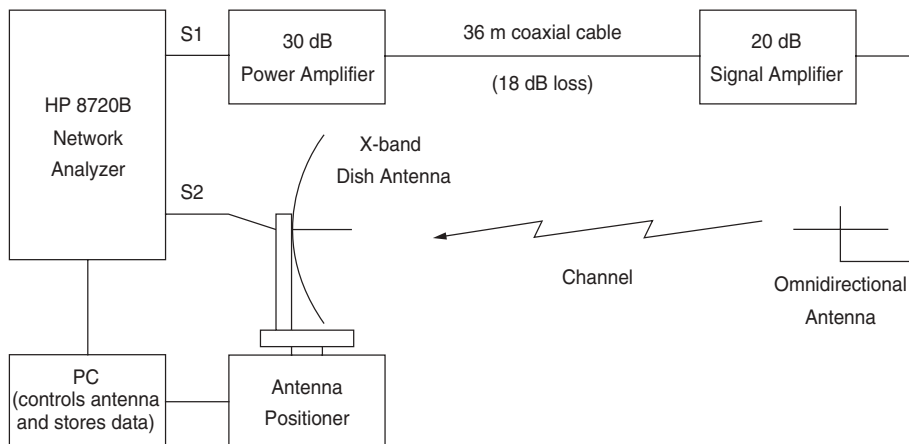


FIGURE 5.22 Frequency-domain measurement system with a directional antenna used in [Spe00].

2 m in diameter is required to maintain the same resolution at 2.4 GHz. Spencer et al. use the results of their measurement to modify the Saleh–Valenzuela model [Sal87b] to include the angle of arrival. A modified version of their model is used by the IEEE 802.11 community to develop a channel model for MIMO systems that is discussed in further detail in Chapter 6.

Another approach to measuring the angle of arrival is to use antenna arrays, which can select the channel electronically rather than mechanically, as implemented in Fig. 5.22. Figure 5.23 shows the basic principles of a frequency-domain measurement system with an eight-element antenna system used by Tingley [Tin00] for the measurement of angle of arrival that we describe in the remainder of this section. This system again employs frequency-domain measurements using a network analyzer, almost identical to the one described in Example 5.7, with an eight-element array antenna to measure eight correlated impulse responses at a frequency range of 2.35 to 2.55 GHz. The eight sequentially measured channel impulse responses are postprocessed with a parameter estimation algorithm to determine the magnitude, phase, and angle of arrival of various paths. The array used in [Tin00], shown in Fig. 5.24a, consists of eight nominally identical quarter-wave monopole elements mounted at a constant radius and separated by one-third wavelength. Figure 5.24b shows eight sample measurements of the frequency response in an indoor area. When configured for 101 points and a predetection bandwidth of 3 kHz, a total of 750 ms is required to measure all eight channel profiles and store the corresponding data to disk. The total data acquisition time may be abbreviated further by employing a larger predetection bandwidth, measuring fewer points, or utilizing a network analyzer with faster internal signal processing. In [Tin00], two algorithms, a spatial filter periodogram (SFP) and discrete maximum likelihood (DML), are introduced for postprocessing of the measured data to estimate all multipath parameters, including the angle of arrival.

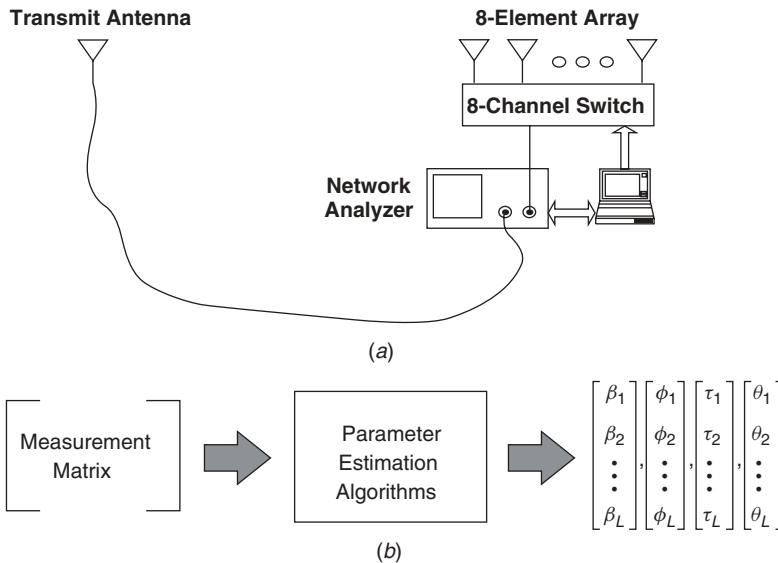
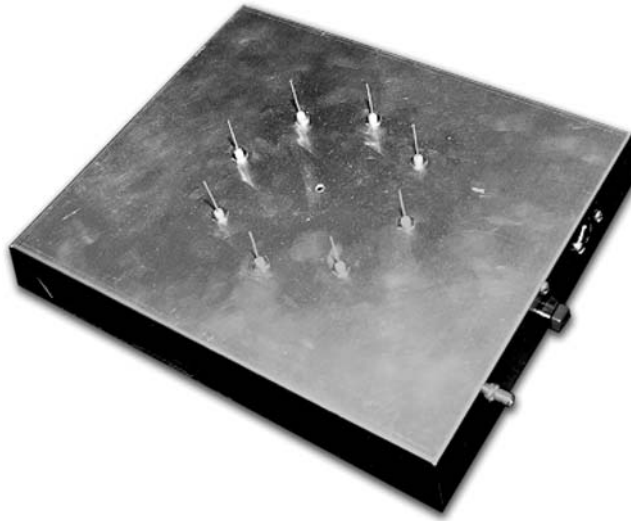
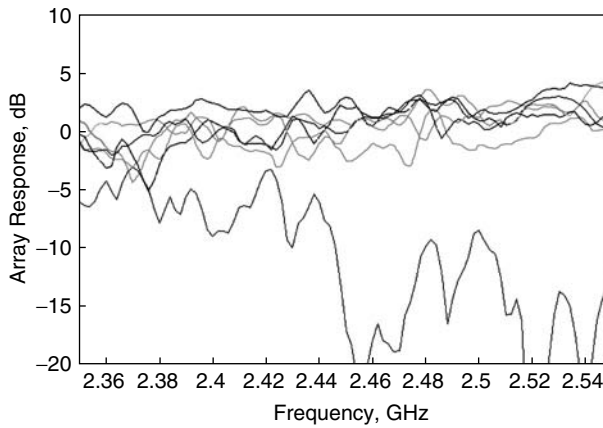


FIGURE 5.23 (a) Frequency-domain measurement system with an eight-element receiver antenna array used for angle-of-arrival measurements in [Tin00]; (b) postprocessing of the measurements of channel profiles from various antennas.



(a)



(b)

FIGURE 5.24 (a) Eight-element antenna used for frequency-domain measurements of the angle of arrival; (b) a sample of eight-measurements at a distance of 3 m [Tin00].

In the approach described in [Tin00], we do not need to move the antennas during the measurement process, and we use only mechanical movement during the calibration process. The purpose of the calibration process is to generate a reference for relating the measured set of channel profiles to the angle of each antenna in the circular array and the DLOS between the transmitter and the receiver. During the calibration process in [Tin00], the transmitting antenna and receiving array are assembled in a small indoor anechoic range. The antennas are mounted atop turntables and separated by an arbitrary close distance of 3 m. In this configuration, the DLOS path, with an angle of arrival of zero, is the dominant path. A series of eight-channel frequency-domain measurements for the eight-elements of the receiving antenna are conducted for each

set of measurements used for calibration. Between each data set, the array is rotated for an arbitrary small angle of 5.625° in azimuth relative to its previous orientation. This measurement and rotation sequence is performed for one complete revolution of the array, which resulted in $360 \div 5.625 = 64$ distinct data sets for 5.625° , given as

$$\mathbf{U}_m = \begin{bmatrix} H_{1,m}(0) & H_{2,m}(0) & \cdots & H_{8,m}(0) \\ H_{1,m}(1) & H_{2,m}(1) & & \\ \vdots & & \ddots & \\ H_{1,m}(101) & & & H_{8,m}(101) \end{bmatrix} = \begin{bmatrix} u_{1,1,m} & u_{1,2,m} & \cdots & u_{1,8,m} \\ u_{2,1,m} & u_{2,2,m} & & \\ \vdots & & \ddots & \\ u_{101,1,m} & & & u_{101,8,m} \end{bmatrix} \quad (5.4.4)$$

where $m = 1, 2, \dots, 64$. The indexes are chosen such that $m = 1$ corresponds to 0° , $m = 2$ corresponds to 5.625° , and so on. The n th column of $\mathbf{U}_m$ represents the 101 samples of the frequency response between the transmitting antenna and the n th elements of the receiving array. The magnitude of all eight impulse responses in $\mathbf{U}_0$ is presented in Fig. 5.24b. The figure displays a flat spectrum for seven antennas but a much higher attenuation at the fifth element, which is obstructed by the first element in direct LOS with the transmitter. The database of all 64 angles forms a basic reference for the expected behavior of the channel for different angles between the transmitter and the receiver. Tingley [Tin00] uses this database during postprocessing to detect the angle and delay of arrival associated with individual paths in a measured set of channel frequency-domain measurements. He introduces two postprocessing techniques, SFP and DML, described in the remainder of this section.

Spatial Filter Periodogram Algorithm. In the SPF, rather than employing a conventional beamforming algorithm, a collection of calibration matrixes is used to design a least-squares spatial two-dimensional filter that shapes the spectrum for the optimal best angle of arrival estimation. We define this spatial filter by

$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1M} \\ w_{21} & w_{22} & & \\ \vdots & & \ddots & \\ w_{81} & & & w_{8M} \end{bmatrix} \quad (5.4.5)$$

in which the m th column provides eight taps, which serve to steer the array in the desired direction while minimizing the energy collected from all other directions. In other words, this filter is a transformation that applies to the eight measurements of the channel frequency responses received from the eight-element antenna:

$$\mathbf{V} = \begin{bmatrix} H_1(0) & H_1(1) & \cdots & H_1(N-1) \\ H_2(0) & H_2(1) & & \\ \vdots & & \ddots & \\ H_M(0) & & & H_M(N-1) \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1N} \\ v_{21} & v_{22} & & \\ \vdots & & \ddots & \\ v_{M1} & & & v_{MN} \end{bmatrix} \quad (5.4.6)$$

and produces M new channel frequency responses for an M angle of arrivals. For the time being, let us defer description of the method used for the design of this filter using calibration data to the end of this section and focus our attention on how to use

this filter. Given an arbitrary measurement matrix, provided in Eq. (4.6), the space-time impulse response representing the behavior in delay and angle of arrival is estimated as

$$\hat{\mathbf{H}} = \mathbf{F}\mathbf{X}\mathbf{V}\mathbf{W} \quad (5.4.7)$$

where $\mathbf{F}$ is the inverse Fourier transform matrix and $\mathbf{X}$ serves as the standard smoothing window function, defined in Eq. (5.3.3) for one-dimensional measurements, which minimizes the sidelobes present in the equivalent time response.

Figure 5.25 represents two sampled two-dimensional measurements of the channel impulse response in LOS and OLOS indoor areas. In the OLOS condition the delay spread is roughly three times longer than the LOS, the arrivals are no longer tightly clustered about the source bearing, and the first arrival is weaker than many subsequent arrivals. Spatial and temporal resolution of the SFP algorithm is investigated in detail

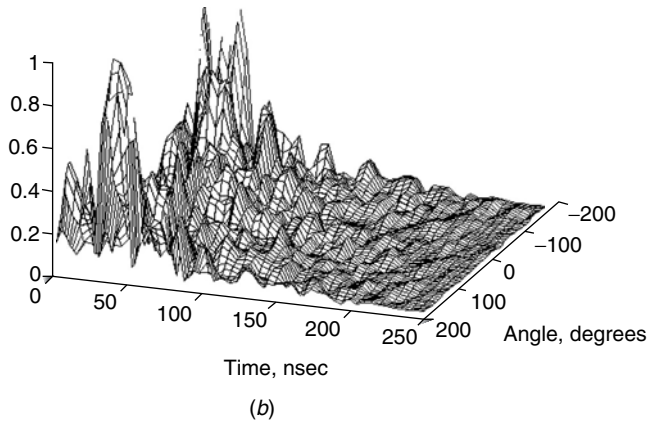
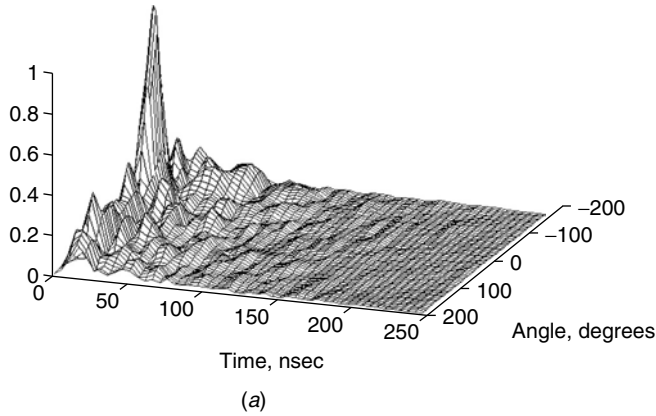


FIGURE 5.25 Samples of two-dimensional indoor channel impulse response measurements using SFP filters and a frequency-domain channel measurement system with eight-element antennas: (a) LOS; (b) OLOS.

in [Tin99a] using computer simulation in which the chamber measurements are combined to form known test matrixes. It is shown that the average spatial resolution is approximately 40° , and the average time resolution is 7 ns. In Figure 5.25a we present the magnitude of $\mathbf{H}$ for a single arrival as measured in the anechoic chamber. Prominently displayed are the spatial sidelobes that result from use of the SFP formulation.

To design the spatial filter we use the calibration data described in Eq. (5.4.4). In [Tin00] there are a total of 64 sets of channel frequency responses collected in the anechoic range, each containing eight-element frequency response with 101 samples of the frequency-domain signal. To determine the coefficients of the filter, we form the cost function

$$f(\mathbf{w}_m) = \sum_{j \neq m} \sum_{l=1}^{101} \left(\sum_{n=1}^8 u_{l,n,j} w_{n,m} \right) \left(\sum_{n=1}^8 u_{l,n,j}^* w_{n,m}^* \right) \quad (5.4.8a)$$

which is minimized, subject to the additional constraint

$$c(\mathbf{w}_m) = \sum_{l=1}^{101} \left(\sum_{n=1}^8 u_{l,n,m} w_{n,m} \right) \left(\sum_{n=1}^8 u_{l,n,m}^* w_{n,m}^* \right) - 1 = 0 \quad (5.4.8b)$$

The cost function serves to minimize the array's sensitivity to energy from all directions except that desired, while the constraint guarantees a solution with a fixed, constant gain in the desired direction. This approach is very similar to the well-known minimum-variance distortionless response method used in two-dimensional adaptive filtering [Hay91a], in which the cost function and constraint are satisfied jointly using the method of complex Lagrange multipliers. The first step in this method is to combine the cost and constraint equations, defined in Eq. (5.4.8), to form the *adjoint equation*

$$\frac{\partial f(\mathbf{w}_m)}{\partial \mathbf{w}_m^*} + \lambda \frac{\partial c(\mathbf{w}_m)}{\partial \mathbf{w}_m^*} = 0 \quad (5.4.9)$$

Performing the partial differentiation over each element of the complex-valued tap-weight vector given by Eq. (5.4.9) produces a system of equations of the form

$$F_k(\mathbf{w}_m, \lambda) = \sum_{j \neq m} \sum_{l=1}^{101} u_{l,j,k} \sum_{n=1}^8 u_{l,j,n}^* w_{n,m} + \lambda \sum_{l=1}^{101} u_{l,m,k}^* \sum_{n=1}^8 u_{l,m,n}^* w_{n,m} = 0 \quad (5.4.10a)$$

where $k = 1, 2, \dots, 8$. Since the constraint equation must also be satisfied, we have

$$F_9(\mathbf{w}_m, \lambda) = \sum_{l=1}^{101} \left(\sum_{n=1}^8 u_{l,n,m} w_{n,m} \right) \left(\sum_{n=1}^8 u_{l,n,m}^* w_{n,m}^* \right) - 1 = 0 \quad (5.4.10b)$$

Taking Eqs. (5.3.10a) and (5.4.10b) together, a solution to the nine equations provides optimal taps for the array processor as well as the Lagrange multiplier, λ . Although the eight equations described by Eq. (5.4.10a) are linear in the independent variables, the ninth equation is quadratic in the array weights. As a result, standard linear algebraic

techniques cannot be used to compute the solution immediately. An iterative solution technique for solving this set of nonlinear equations is described in [Tin99a].

Discrete Maximum Likelihood Algorithm. Although the SFP algorithm provides a simple method to estimate the space–time impulse response from a measurement matrix, it suffers from several limitations. The spatial resolution is strictly limited by the number and configuration of array elements, and in the current design stands at 40° . Similarly, temporal resolution is given by the reciprocal of the sweep bandwidth and is approximately 7 ns. Both of these limitations are independent of the signal-to-noise ratio of the data acquisition system. A third limitation is that the algorithm tends to introduce bias into the channel parameter estimates as a pair of closely spaced arrivals appears [Tin00].

These restrictions can be removed by developing an algorithm based on the maximum likelihood (ML) criteria [Joh93]. The basic principle behind ML estimation is very simple: We consider a set of possible values for channel parameters and we reconstruct a channel frequency response for that set. Then we find the error between the reconstructed response and the measured response, and finally, we do an exhaustive search to find the parameters of the channel that minimizes the error between the measurement and the reconstructed response. Since this is a very computationally extensive approach, the main obstacle is to find an efficient algorithm to search for an optimal solution. In [Tin00], two simple serial and parallel algorithms are provided first, and since serial search is inaccurate and parallel search is computationally impractical, a new algorithm called *recursive serial search* is introduced.

The recursive serial algorithm uses the measured calibration matrices $\mathbf{U}_m$, $m = 1, 2, \dots, M$, defined by Eq. (5.4.4) as a fundamental basis functions. Given the overall two-dimensional channel impulse response defined by Eq. (5.4.3), an arbitrary measurement $\mathbf{V}$ can be modeled as

$$\hat{\mathbf{V}} = \sum_{l=1}^L \beta_l e^{j\phi_l} \mathbf{D}(\tau_l) \mathbf{U}_l \quad (5.4.11)$$

where L is the number of discrete paths, $\beta_l e^{j\phi_l}$ is the complex weight of the l th path, and $\mathbf{D}(\tau_l)$ is the 101×101 diagonal time-delay matrix, defined as

$$\mathbf{D}(\tau) = \begin{bmatrix} e^{-j2\pi f_1 \tau} & 0 & \dots & 0 \\ 0 & e^{-j2\pi f_2 \tau} & & \\ \vdots & & \ddots & \\ 0 & & & e^{-j2\pi f_N \tau} \end{bmatrix}$$

where f_i is the i th sample in the 101th set of a measured channel frequency response. The values of L and $\tau_l, \theta_\lambda, \beta_\lambda$, and ϕ_λ , where $l = \{1, 2, \dots, L\}$ are selected to minimize

$$J_{\text{DML}} = \sum_{k=1}^{101} \sum_{n=1}^8 |v_{kn} - \hat{v}_{kn}|^2 \quad (5.4.12)$$

Using the delay operator matrix $\mathbf{D}(\tau)$ in Eq. (5.3.15), $D - 1$ new versions of each calibration matrix are constructed, with the first is delayed τ_0 seconds, the second $2\tau_0$ seconds, and so on. This procedure yields the basis matrixes $\mathbf{C}_k$, where $k = 1, 2, \dots, DM$.

The algorithm then begins with the assumption that a single path is to be found. A search is conducted over all values of k until that basis is identified that minimizes J_{DML} . The index k_1 is retained in the path history for use in subsequent searches. Once the index of the first path has been identified, the model order is incremented to consider a measurement composed of two discrete arrivals. The first arrival, found in the preceding iteration, is assumed to be fixed at the index k_1 . The second arrival is found by another search over the remaining $DM - 1$ values of k . Once found, the second arrival is assigned the index k_2 .

At each stage of iteration, the minimum value attained by the cost function is given as

$$J_{\text{DML}_{\min}} = m - \mathbf{p}^H \mathbf{R}^{-1} \mathbf{p} \quad (5.4.13a)$$

where $\mathbf{p}$ represents the cross correlation between each of the basis functions and the measurement matrix, and $\mathbf{R}$ represents the correlation between the individual basis functions [Tin99b]. Also,

$$m = \sum_{k=1}^{101} \sum_{n=1}^8 |v_{kn}|^2 \quad (5.4.13b)$$

is the total energy contained in the measurement. The process of incrementing the model order, followed by finding an additional path, is repeated until the desired number of paths have been identified. Since the sequential algorithm seeks to minimize J_{DML} at each stage, paths are identified in order of decreasing energy. This property provides a convenient means to exit the search by monitoring the optimal value of J_{DML} achieved at each iteration. Once the residual drops below, say, 10% of m , we may assume that the model has accounted for most of the energy contained in the measurement, and the algorithm may be terminated. The recursive search requires DML evaluations of the cost function to identify L paths. However, in the recursive algorithm, these evaluations assume the order of the number of paths found. For example, searching for the L th path requires evaluation of DM systems of L linear equations. Although far more computation is required than for a serial search, the algorithm is still quite practical. In most cases, parameter extraction from a measurement consumes from 1 to 3 minutes of CPU time on a contemporary Pentium-based personal computer. The DML algorithm is considered in more detail in [Tin99b], where the spatial and temporal resolutions are determined to be 2° and 1 ns, respectively, using sample measurements in the anechoic chamber.

Figures 5.26a and b show the result of a typical simulation run for a case consisting of 20 paths. In Fig. 5.26, the known path parameters are represented by an open circle and estimates using SFP or DML are represented by crosshatching. The SFP technique finds many of the paths exactly but encounters difficulties for paths that are coincident in time or angle of arrival. By contrast, the DML algorithm identifies all paths exactly. This performance is to be expected as long as a high-measurement signal to noise ratio is maintained. More details of the statistical study may be found in [Tin99b].

5.4.4 Frequency-Domain Measurement for UWB Measurements

Frequency-domain measurement systems are ideal for ultrawideband (UWB) measurements for WPAN applications because they are simple and they can easily be implemented for bandwidths of several gigahertz. The network analyzer used for

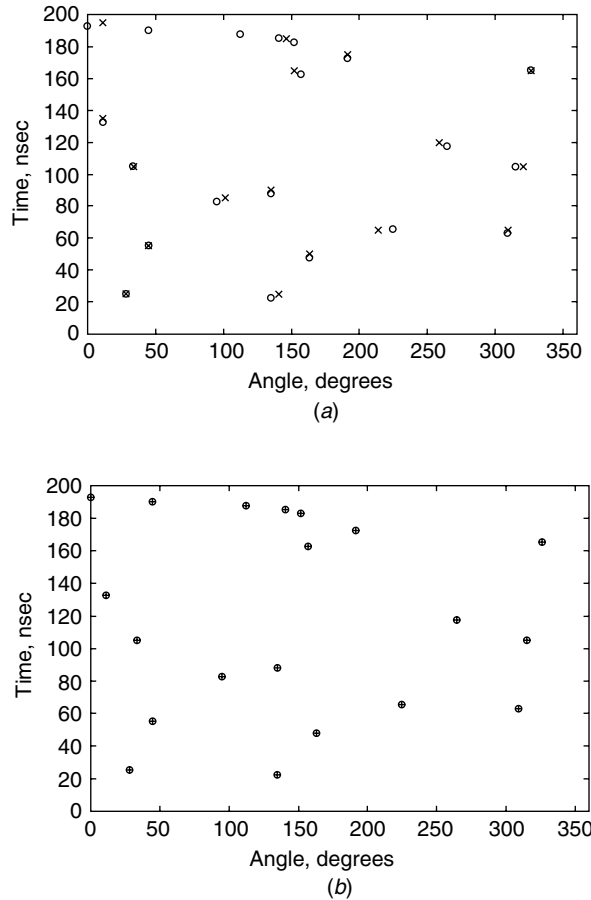


FIGURE 5.26 Results of simulation of a 20-path channel and estimation of its location using (a) SFP and (b) DML [Tin00].

the frequency-domain measurement system described in Example 5.7 can use its s -parameter test to measure up to 6 GHz. The more advanced network analyzer can measure up to several tens of gigahertz, which covers all frequency bands considered for UWB developments. As a result, a number of researchers use the setup in Example 5.7 with UWB antennas to perform UWB channel measurements, and use these measurements for the characterization and modeling of UWB signals. The main challenge for the design of a UWB system was the antenna design. Interested readers will find an overview of these antennas in [Sha04]. The two most popular UWB antennas are the bowtie and cone antennas, shown in Fig. 5.27*a* and *b*, respectively. The more directional bowtie antenna shown in this figure has dimensions of 2.5 by 5.25 inches, and the omnidirectional cone antenna is even smaller. Figure 5.27*c* and 5.27*d* show two sample measurements of the UWB channel characteristics using the cone antenna for detected and undetected direct paths in an indoor area. Comparing with the 200-MHz band measurements shown in Fig. 5.14, UWB measurements resolve a much large number of multipath components. These measurements also average the

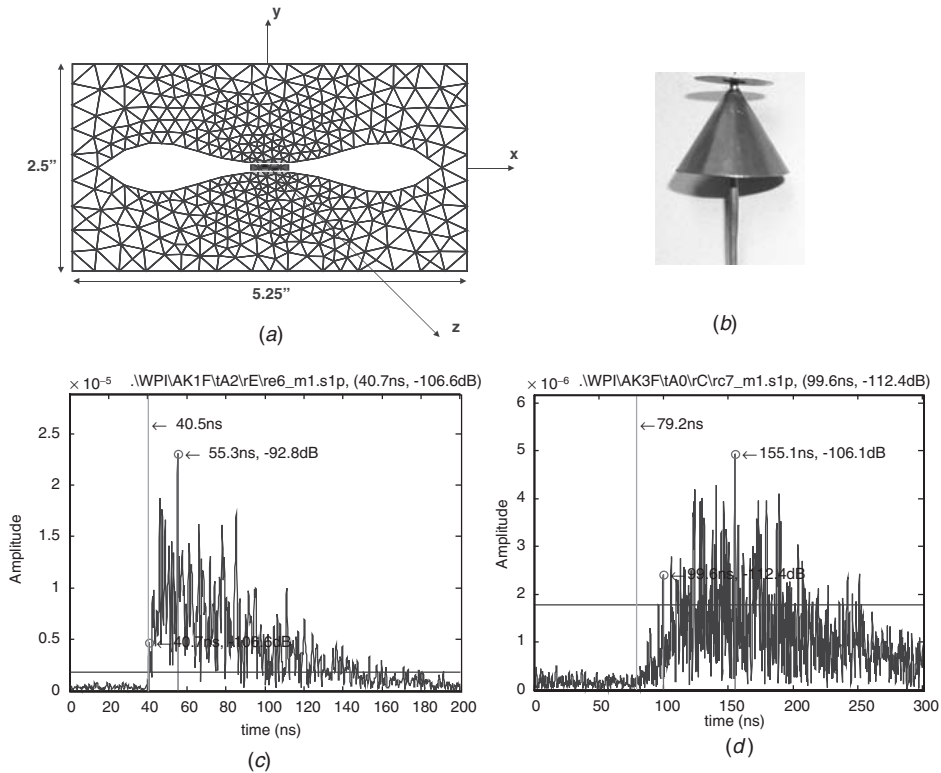


FIGURE 5.27 UWB antennas: (a) bowtie and (b) cone; and measurements from 3 to 6 GHz: (c) detected direct path; (d) undetected direct path.

power over a wider bandwidth, providing a more stable signal with respect to multipath fading. More measurement and modeling results related to UWB systems are provided in Chapter 12.

QUESTIONS

- How many different techniques can you identify for measurements of multipath delay profiles?
- What is the simplest system for wideband measurement of indoor radio propagation characteristics?
- Which measurement technique is most popular for wideband measurement of radio channel characteristics in urban areas?
- Considering the results of indoor radio propagation measurements reported in this chapter, what is a reasonable number for the maximum rms delay spread in indoor areas? How does it compare with the delay spreads adopted for the JTC model for indoor areas?

- (e) Explain how a sliding correlator works. Clarify why it is popular for channel measurement and synchronization but not for data transmission.
- (f) What is the maximum variation in rms delay spread due to measurement of local short-term variations as reported in this chapter? What type of movement has caused this maximum fluctuation?
- (g) Explain how we use the results of time-domain measurements, made with the pulse transmission method depicted in Fig. 5.1, together with the FFT, to produce an accurate representation of the channel frequency response. Consider the question again using the time-domain measurement approach depicted in Fig. 5.6.
- (h) Discuss the resolution capability of the three measurement systems reported in this chapter. Which of the three measurement techniques is most difficult to implement, and which is the simplest?
- (i) Given equal transmission powers, which measurement system provides the largest coverage area?
- (j) In a spread-spectrum measurement system, if we keep the bandwidth fixed, how can we increase the resolution?
- (k) How does the frequency correlation function relate to the data-rate limitations of the channel?
- (l) What are the advantages and disadvantages of the spread-spectrum time-domain measurement method compared with the pulse transmission method?
- (m) Compare various techniques used for measurement of the angle of arrival for different paths.
- (n) What is the difference between measurement techniques used for telecommunication and those used for geolocation applications?
- (o) How does the superresolution algorithm help in measurements of wideband channel characteristics?
- (p) What is the main challenge in UWB measurement using network analyzers?

PROBLEMS

1. We want to use a pulse transmission technique for measurement of the multipath characteristics of in an indoor office building. Assume that the maximum delay spread to be measured is 800 ns, the required resolution of the measurement system (the base of the autocorrelation function of the spreading code) is 10 ns, and the maximum Doppler shift of the channel is 10 Hz. The system is transmitting with 1 W of power at 1900 MHz and we assume that the pulses are ideal triangular pulses.
 - (a) Prepare a block diagram for implementation of this system that uses a 100-MHz digital scope at the receiver.
 - (b) Sketch the ideal frequency response of the transmitted pulse and specify the required 6-dB transmission bandwidth of the system.

- (c) Give the period for repetition of pulses.
 - (d) What is the minimum time it takes to measure one sample of the channel impulse response?
 - (e) If we average the received pulses over the measurement period to reduce the background noise, what is the maximum number of periods that can be used for averaging without significant changes in the channel characteristics? Assume that the channel characteristics are stationary during the coherence time of the channel.
 - (f) What is the coverage of the system with and without averaging? Use the JTC model to calculate the coverage.
2. We want to use the spread-spectrum technique with a sliding correlator for measurement of the multipath characteristics of the urban radio channel. Assume that the maximum delay spread to be measured is $10\ \mu\text{s}$, the resolution of the measurement system (the base of the autocorrelation function of the spreading code) is $100\ \text{ns}$, the maximum Doppler shift of the channel is $100\ \text{Hz}$, and the transmitted power is $10\ \text{W}$. If the system is transmitting at $1900\ \text{MHz}$, the pulses are ideal rectangles, and the chip values are independent of one another:
- (a) Prepare a block diagram for the implementation of the system, and specify parameters for components of the system.
 - (b) Specify the bandwidth expansion factor N , the chip rate R_c , and ΔR , the difference between the chip rates at the transmitter and receiver.
 - (c) Sketch and label a sample ideal received periodic signal after correlation.
 - (d) What are the required transmission bandwidth of the system and the measurement time for a channel profile?
 - (e) What is the coverage of the system if it is used in a macro-cellular environment (use the JTC model) with a transmitter height of $20\ \text{m}$ and a receiver height of $2\ \text{m}$?
3. We want to use a frequency-domain measurement system for measurement of the UWB multipath characteristics of a residential indoor area. In this environment the maximum delay spread to be measured is $200\ \text{ns}$, the maximum Doppler shift is $10\ \text{Hz}$, the transmitted power is $100\ \text{mW}$, and the bandwidth of interest is 4 to $5\ \text{GHz}$.
- (a) Prepare a block diagram for implementation of the system and specify the type of antenna. Give your reasoning for selection of the antenna.
 - (b) Prepare a block diagram of the signal processing of the frequency-domain signal received needed to produce the time-domain channel impulse response.
 - (c) What is the minimum number of samples in the frequency domain? What type of filter in the frequency domain should we use to keep the sidelobes under $30\ \text{dB}$ below the peak value?
 - (d) How many samples are needed in the time domain to ensure that the distance measurement error caused by sampling in time is below $1\ \text{m}$ and we can capture the full time-domain multipath characteristics of the received signal?
4. How can we modify the block diagram in Fig. 5.6 so that the system can be used for direct pulse transmission measurement? In this case, only one pulse with a duration of one chip is transmitted every N chip durations. How do the average transmitted

power and transmitted bandwidth differ from those of the original spread-spectrum measurement system?

5. Prepare a block diagram and detailed equipment specifications for a frequency-domain measurement system that duplicates the measurements taken by the spread-spectrum system described by Fig. 5.6.
6. The frequency-domain measurements can be made for shorter distances using only a network analyzer and two antennas. Simple antennas used in portable or cellular phones can be used for this purpose. These antennas are usually omnidirectional monopoles with antenna gains of around 2 (3 dB). The typical maximum transmitted power from the network analyzer is -20 dBm, and the noise level at the receiver end of the analyzer is -90 dBm. If we use this simple setup for measurement, what are the typical distances that will enable us to measure the channel characteristics at 1 GHz? (If a network analyzer is available, set up this simple experiment and confirm the accuracy of your calculations.)

PROJECT

Project 1: Analysis of Measured Data Using a Network Analyzer

In this project we analyze part of the data collected in the measurement campaign reported in [Ben99] to understand the details of the simple wideband frequency-domain radio channel measurement process. These measurements were taken at a commercial site, the Norton Company, Worcester, Massachusetts, shown in Fig. P5.1. The measurements were collected using the network analyzer measurement system described in Section 5.3.1. In the setup used in this campaign, 401 samples of the magnitude and



FIGURE P5.1 Measurement site at the Norton Company.

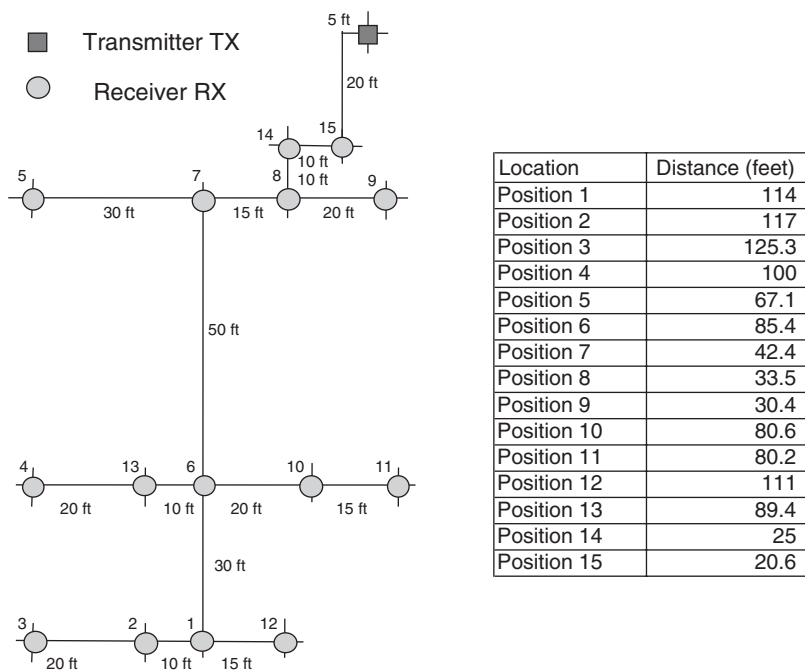


FIGURE P5.2 Measurement layout with receiver and transmitter locations and Tx–Rx distance for measurement at the Norton Company.

phase of the frequency response are collected for each measurement in the frequency band 900 to 1100 MHz. Using this system, the transmitter Tx was fixed at one location and the receiver Rx was moved to 15 different positions. The map of the measurement positions and the distance between the Tx and Rx locations is provided in Fig. P5.2.

The time response is computed using an IFFT technique along with windowing the frequency response with a Hanning window. If no windowing technique is used, the time response will show many sidelobes associated with the $(\sin x)/x$ function (inverse Fourier transform of the square window), which makes it difficult to estimate a genuine path in the multipath profile. It is also desired that all channel impulse responses (CIRs) be computed using the same time intervals. After examination of the data, most time responses die out after 800 ns. Therefore, all CIRs are to be provided in the time interval 0 to 800 ns. To be able to zoom into any portion of the IFFT, the CHIRPZ MATLAB subroutine is used to compute the CIRs. This subroutine allows us to compute the IFFT in the interval 0 to 800 ns with as many points as desired. We chose 1601 points in the time domain, leading to a time increment of 0.5 ns between samples.

To evaluate the characteristics of the radio channel, we first need to extract the magnitude and phase of the arriving paths for each profile. To complete this task, a peak detection algorithm is used which associates a path magnitude and delay to each peak in the CIR. To avoid associating a path with noise or other unreliable data, only paths above a given threshold are kept. Figure P5.3 shows a sample channel impulse response and delays and amplitudes detected for the paths. From the extracted taps, one can estimate the total power and the rms delay spread of each CIR. If β_k represents

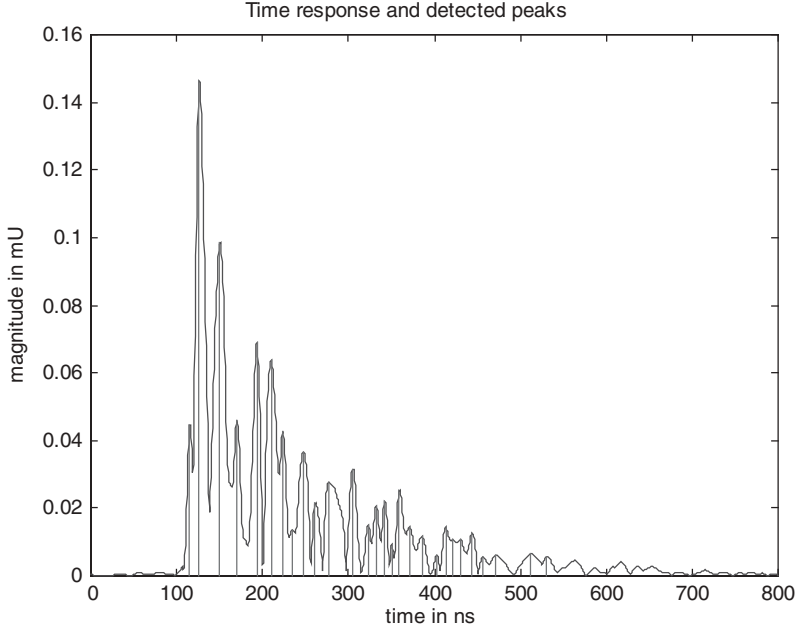


FIGURE P5.3 Sample channel impulse response along with the extracted delays and magnitude of the paths obtained using a peak detection algorithm.

the magnitude of a path and τ_k the time of arrival of that path, as shown in Eq. (3.4.3), the total power is computed as

$$P_r = \sum_{k=1}^N \beta_k^2$$

For the set of measurement data provided in this project, the transmitted power is $P_t = 0$ dBm, leading to path loss $L_p = -P_r$. To calculate the rms delay spread, we combine Eqs. (3.4.4) and (3.4.5) in a single equation:

$$\tau_{\text{rms}} = \sqrt{\frac{\sum_{k=1}^N \tau_k^2 \alpha_k^2}{\sum_{k=1}^N \alpha_k^2} - \left(\frac{\sum_{k=1}^N \tau_k \alpha_k^2}{\sum_{k=1}^N \alpha_k^2} \right)^2}$$

The measured data and a supporting MatLAB code for this project are provided at http://ece.wpi.edu/~kaveh/Book_JW/Measurements/. This site contains 15 *pos_k.raw* measurement data files k from Rx position 1 to Rx position 15. A *pos_k.raw* file is such that the first column is the frequency in Hz, the second column is the frequency-response magnitude in dB, and the third column is the frequency-response phase in degrees. The MatLAB script *ms_extract.m*, which reads in the data file *pos₁.raw*, computes the CIR using the CHIRPZ IFFT technique, and extracts the taps using a peak detection algorithm.

- (a) Examine the script first and expand the script to compute the total power the rms delay spread of the CIR. Plot the magnitude and phase of the frequency response. Also, plot the time-domain CIR magnitude and extracted path parameters. For data file *pos1.raw*, provide plots of frequency magnitude response, frequency phase response, and time-domain CIR magnitude response with extracted paths, similar to Fig. P5.3.
- (b) Provide a four-column table in which each row gives the location number, distance between Tx and Rx, path loss in dB, and rms delay spread in (ns). The distance values are given in Fig. P5.2.
- (c) Using MatLAB, determine L_0 and α such that $L_p = L_0 + 10\alpha \log_{10} d$ best fits the path loss data that you collected in part (b).
- (d) Provide a plot of the cumulative density function (CDF) of the rms delay spread of all measurements. (The x axis should be in ns, and the y axis should be the probability of the rms delay spread $>$ abscissa). What is the value of the rms delay spread for which greater than 80% of the time our measured rms value is under that threshold?

6

MODELING OF WIDEBAND RADIO CHANNEL CHARACTERISTICS

- 6.1 Introduction
- 6.2 Wideband Time-Domain Statistical Modeling
 - 6.2.1 Wideband Models for Wide-Area Networks
 - 6.2.2 Wideband Models for Local Area Networks
 - 6.2.3 Direct Modeling of Path Arrivals and Amplitude
 - 6.2.4 UWB Models for Local Area Networks
 - 6.2.5 Simulation of AOA for MIMO Channels
- 6.3 Wideband Frequency-Domain Channel Modeling
 - 6.3.1 Autoregressive Modeling
 - 6.3.2 Statistical Modeling in the Frequency Domain
- 6.4 Comparison Between Statistical Models
- 6.5 Ray-Tracing Algorithms
 - 6.5.1 Reflection and Transmission Mechanisms
 - 6.5.2 Diffraction Mechanism
 - 6.5.3 Diffused Wall Scattering
 - 6.5.4 Two-Dimensional Ray Tracing in Small Indoor Areas
 - 6.5.5 Three-Dimensional Simulation in Urban Microcellular Environments
- 6.6 Direct Solution of Radio Propagation Equations
 - 6.6.1 Finite-Difference Time-Domain Model
- 6.7 Comparison of Deterministic and Statistical Modeling
- 6.8 Site-Specific Statistical Model
 - 6.8.1 Power of a Path with a Given Length
 - 6.8.2 General Formula for the Delay Power Spectrum
- Appendix 6A: GSM-Recommended Multipath Propagation Models
- Appendix 6B: Wideband Multipath Propagation Models
- Questions
- Problems
- Projects
 - Project 1: Simulation of Wideband JTC Model
 - Project 2: Channel Simulation Using Poisson Arrivals
 - Project 3: Channel Simulation Using $\Delta-K$ Arrival Model
 - Project 4: Channel Simulation Using AR Model

6.1 INTRODUCTION

In Chapters 3 to 5 we defined the parameters that characterize multipath fading, described the systems used for measuring these parameters, and presented the results of measurements made on various radio channels. The measurement results were divided into two categories, narrowband and wideband. The results of narrowband measurements addressed signal coverage through the power–distance relationship, and they related the Doppler spread to the movement of objects in the coverage area and the movement of portable or mobile terminals. We then discussed the statistics of the amplitude fluctuations of a received narrowband signal and its Doppler spectrum in indoor and outdoor areas and described methods for computer simulation of narrowband amplitude fluctuations. For the case of wideband signaling, we introduced various methods for measuring the multipath characteristics of the radio channel for portable and mobile users and presented results of some time- and frequency-domain measurements. We also discussed results of measurements of the root-mean-square (rms) multipath spread and the 3-dB width of the frequency correlation function in different buildings, as well as the effects of partitioning and short-time variations. Finally, we discussed advances in channel measurements for MIMO channels, UWB communications, and wireless positioning that are vital to development of the next-generation location-aware broadband wireless ad hoc networks.

In this chapter we describe methods for computer simulation of wideband radio propagation which includes the multipath characteristics of the channel. In the past, performance evaluations of communication systems were typically based on simple statistical models of channel and closed-form solutions, providing bit error rates (BERs) for different modulation techniques. With the rapid increase in computational power of computers and drastic reduction in their cost, software and hardware computer simulation is becoming an increasingly popular approach to performance evaluation. Many versatile software products are also becoming available for use in developing communication system simulations, including block-oriented simulation packages, which offer many conveniences to the user.

Ideally, a simulation should provide “snapshots” of wideband channel response, in the time, frequency, and space domains, at a rate twice the Doppler spread of the channel. A complete simulator of this form provides both static and dynamic behavior of the channel. Computer simulation of the channel is used for performance evaluation of modems, analysis of multiple access methods, placement of base stations in a cellular system, and analysis of interference in various networks. For some of these applications the channel response as a function of location is of primary importance, and a description of the static behavior of the channel is adequate. Channel models providing static snapshots of the channel impulse response at different locations, to be used for evaluations under various performance criteria, such as probability of outage or average probability of error of a specific modulation and coding technique over a prescribed area, also need to include the dynamic behavior of the channel. The dynamic behavior of the channel is also needed for detailed analysis of the behavior of the adaptive functions of modems, such as automatic gain control (AGC), equalization, and timing recovery.

There are two basic approaches to simulating wideband radio propagation characteristics: (1) measurement-based statistical modeling and (2) direct analytical solution of the radio propagation equations. Measurement-based statistical models are based on a

mathematical description using several parameters. The parameter values are evaluated for each measurement of the wideband channel characteristics, and the statistics for the parameters over a large database are used to complete the model for a given coverage area. Statistics gathered from measurements in typical areas are extended to develop a more generalized model for all coverage areas. Statistical models generally do not incorporate details of the siting of buildings in an outdoor coverage area or the layout of rooms within a building. Instead, they classify all areas into a limited number of broadly designated environments and all buildings into a few classes of buildings.

In modem performance evaluations, the system designer is usually concerned with overall performance over typical areas or typical buildings, and statistical models usually serve the purpose reasonably well. In some other application, such as microcellular or indoor installations, where proper siting of antennas is an important issue, building-specific radio propagation models offer a more precise tool for determining optimum antenna locations. Building-specific radio propagation models are based on direct solution of the radio propagation equations, with boundaries defined by a map of a coverage area or the layout plan of a building. The technique known as *ray tracing* provides a simple approximation for analysis of radio wave propagation. Another approach is the numerical solution of Maxwell equations using the *finite-difference time-domain (FDTD) technique*. Ray-tracing algorithms are also very useful for analysis of the angle of arrival of the paths for MIMO applications and the TOA of the DLOS path needed for the popular TOA-based geolocation systems.

To compare the results of various computer simulation techniques, several approaches might be taken. The most obvious is to compare the measured and simulated channel responses in typical locations. This method is not well suited to evaluation of statistical models because statistical models do not relate the channel response to a specific location. However, for assessing building-specific radio propagation models, this method is very useful. Another approach to evaluating the results of a simulation method is to compare empirical data with the cumulative distribution functions (CDFs) of the rms delay spread and multipath power produced by the simulation. Yet another approach to comparing radio propagation models is to evaluate the performance of a particular modem over the measured and modeled channels. Standard modulation techniques such as BPSK and wideband techniques such as direct-sequence spread-spectrum or nonspread signaling with adaptive equalization can be used as benchmarks in these evaluation approaches.

In the following sections we describe various methods for the simulation of wideband radio channel characteristics and compare simulation results with the results of wideband measurements. Channel models for wideband characteristics of wireless channels are divided into statistical and building-specific deterministic models. The time-domain statistical models, described in Section 6.2, model channel impulse response based on empirical data and provide a tool to reproduce channel profiles using computer simulations. The frequency-domain statistical models described in Section 6.3 provide statistical models developed from empirical data to be used for generating the frequency response of the channel. Building-specific deterministic models are intended to solve Maxwell's propagation equations for boundaries defined by the layout of a building or an area. The ray-tracing algorithm, described in Section 6.5, provides an approximate ray-optics solution to the Maxwell equations. Simple ray-tracing algorithms using direct transmission and reflections were given as examples in Chapter 3. In Section 6.5 we provide more detailed equations and include transmission

through objects, diffraction, and scattering. Section 6.6 provides a brief discussion of the finite-difference time-domain (FDTD) approach for direct numerical calculation of Maxwell's equations. Sections 6.4 and 6.7 are devoted to comparisons among these techniques. Section 6.8 introduces a statistical approach for calculation of the wideband characteristics of a channel using the ray-tracing concept.

6.2 WIDEBAND TIME-DOMAIN STATISTICAL MODELING

Time-domain techniques using wideband statistical models are the most popular methods for computer simulation of indoor and outdoor radio systems for both wide-area cellular networks and the local area WLAN and WPAN applications. Standards-setting bodies usually recommend a generalized and simple time-domain wideband statistical model for simulation of the radio channel characteristics that are pertinent to performance evaluation of modems and provide a good estimate of the coverage of the network.

Figure 6.1 shows the behavior of the wideband characteristics of a typical multipath fading channel as terminals move away from a base station or an access point. In close proximity to the base station or access point, the DLOS path between the transmitter and receiver dominates all other paths and we have a small multipath spread. As the distance increases, the other paths become gradually stronger, and at a certain point when an obstacle appears between the transmitter and receiver, the DLOS path is no longer the strongest path, and after a certain distance, DLOS may no longer be detected. The total power that is the phasor sum of all paths fluctuates according to the multipath fading phenomenon described in Chapter 4. Wireless channels considered in this book are slowly time varying; therefore, channel multipath parameters change

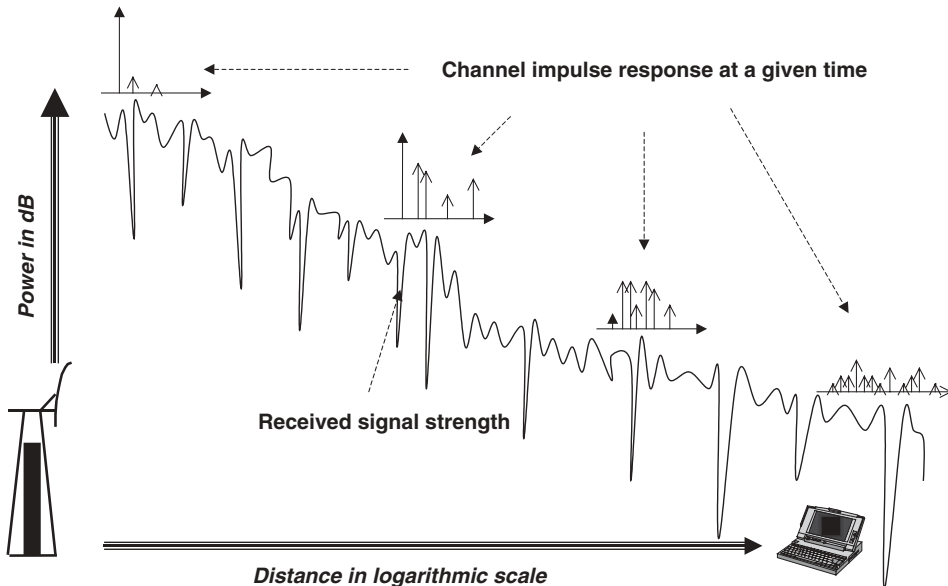


FIGURE 6.1 Relation between wideband static and narrowband dynamic behaviors of the wireless channels.

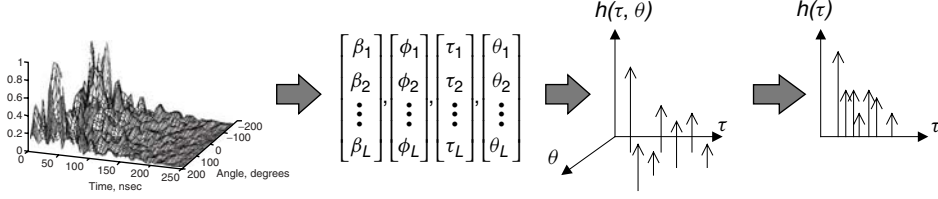


FIGURE 6.2 Extraction of channel multipath parameters from two-dimensional measurements of radio propagations.

slowly. As a result, we can separate the static and dynamic behavior of the channel and model them separately. In other words, we can use more complex wideband measurements for modeling the behavior of normalized channel impulse response and use extensive narrowband measurements of the received signal strength for analysis of the dynamic behavior. Referring back to Fig. 6.1, we can collect adequate channel impulse responses in different locations to develop models for the statistical behavior of multipath parameters, and we collect measurements of the received signal strength to develop path-loss models and models for the Doppler spectrum of the channel. Since the latter part of measurement and modeling was done in Chapter 4, in the rest of this section we focus on the modeling of channel multipath parameters.

With the most advanced measurement systems we can measure the two-dimensional wideband characteristics of the channel as shown in Fig. 6.2. From each measurement we can determine the two-dimensional peaks of all the paths and their associated magnitude, phase, delay, and angle of arrival. As we discussed in Chapter 5, using estimates of the two-dimensional channel parameters associated with the paths, we can represent the overall channel impulse response with a simple mathematical representation:

$$h(\tau, \theta) = \sum_{i=1}^L \beta_i e^{j\phi_i} \delta(\tau - \tau_i) \delta(\theta - \theta_i) \quad (6.2.1)$$

Since measurement of the angle of arrival is very challenging, most measurements of the channel characteristics are taken only for the delay variable, where the impulse response is represented by

$$h(\tau) = \sum_{i=1}^L \beta_i e^{j\phi_i} \delta(\tau - \tau_i) \quad (6.2.2)$$

The mathematical formulation in Eq. (6.2.2) was first suggested for statistical modeling of the urban radio channel by Turin [Tur72] and later used for statistical modeling of the indoor radio channel [Sal87b, Gan91a, Gan91b, Rap91b, Yeg91, Has93a]. The simple and generalized models recommended by the GSM standards body for mobile radio channel modeling and by the Joint Technical Committee (JTC) for PCS channels are based on the same mathematical formulation [GSM91, JTC94]. More recently, these models are extended to modeling the angle of arrival for MIMO systems [Spe00, Tin00], modeling the TOA for geolocation systems [Pah98], and UWB channel modeling [Cas02].

In terms of classical channel modeling, described in Section 3.5, separation of static and dynamic behavior in a slowly time-varying channel means that we can decompose the scattering function into two functions: the delay power $Q(\tau)$ and the Doppler spectrum $D(\lambda)$. Then the scattering function is given by

$$S(\tau, \lambda) = R_{hH}(\tau, \lambda) = Q(\tau)D(\lambda) \quad (6.2.3)$$

Modeling and simulation of Doppler spectrums $D(\lambda)$ was discussed in Chapter 4. In this chapter we focus on models for simulation of the delay power spectrum $Q(\tau)$. For urban and indoor radio channels represented by a discrete channel impulse response model, given by Eq. (6.2.1), the delay power spectrum is the average of the channel impulse responses:

$$Q(\tau) = \overline{|h(\tau)|^2} = \sum_{i=1}^L \overline{|\beta_i|^2} \delta(\tau - \tau_i) \quad (6.2.4)$$

If we complement specifications for Eq. (6.2.3) with a path-loss model, we have a complete set of analytical tools to simulate a channel for both coverage and performance evaluation purposes.

In terms of hardware and software simulations, separation of static and dynamic behavior means that we can implement the behavior in the delay variable with a tapped delay line and use filtered Gaussian noise or the Clarke–Jakes model to simulate the variations of each tap. This concept is illustrated in Fig. 6.3. In Fig. 6.3a we have paths with different arrival delays implemented in parallel branches, and Fig. 6.3b shows implementation of the amplitude and phase of each path using a filtered Gaussian noise method. The filter is designed to shape the spectrum to the prescribed Doppler spectrum $D(\lambda)$. The simulated complex channel fluctuations in Fig. 6.3b are scaled with the strength of the path so that the overall channel response in Fig. 6.3a provides for the delay power spectrum defined in Eq. (6.2.4). In general, the delay τ_i is a random variable, but for simplicity of implementation, traditional standardization organizations assume fixed values for the delay.

The main objective in the development of a model for the wideband characteristics of a channel is to develop a foundation for design and comparative performance evaluation of wireless modems. Traditionally, performance analysis was carried out using analytical equations, and the analytical equations were calculated using digital computers. As the speed of computers and digital hardware in general increased, models were also used for real-time hardware and computer software simulations of channel behavior. One of the major challenges for wireless standardization organizations is to compare and select the best modem design for physical layer implementation among the variety of systems proposed. To achieve a fair comparison among these proposed alternatives, a commonly accepted channel model is needed. After completion of the standard, these models are used by manufacturers for the design and performance evaluation of their products. Since the bandwidth and environments in which these channel models are used are different, most standardization groups develop their own standards. Wide-area cellular networks operating in licensed bands have lower bandwidths, GSM has 200 kHz, and IS-95 has 1.25 MHz; the third-generation systems are all under 10 MHz of bandwidth. Wireless LANs use unlicensed ISM and U-NII bands, and the minimum bandwidths of the systems are around 20 MHz. Wireless PANs, operating in UWB frequency bands, use a bandwidth on the order of gigahertz. As a result, we

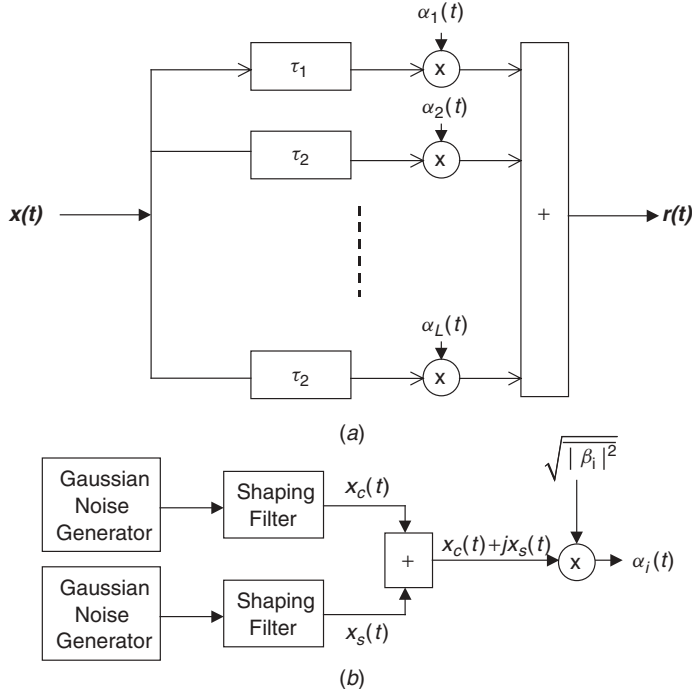


FIGURE 6.3 Block diagram of a wideband channel simulator: (a) the tapped delay line representing the static behavior; (b) narrowband channel fluctuation simulator representing dynamic behavior.

have different channel models for wide- and local-area networks, and we treat them in separate subsections below.

6.2.1 Wideband Models for Wide-Area Networks

A standards committee usually recommends a set of propagation conditions or channel characteristics to be used for hardware or software simulation of the channel. Such a recommendation provides a common basis for comparative evaluation of alternative modulation schemes, adaptive equalization techniques, link-layer protocols, and access methods under consideration for adoption in a particular standard. These recommendations generally comprise two parts: a path-loss model and a wideband propagation model. Path-loss models, discussed in Chapter 4, provide equations to relate the average received power to the distance between transmitters and receivers in different environments.

To define a channel model, we need to specify the scattering function of Eq. (6.2.4). When we want to adopt a time-delay profile for a particular standard, we have to define the number of paths in the delay power spectrum, $Q(\tau)$. Fewer paths would be easier for simulation of the channel on a digital computer. The distance between the adjacent paths in Eq. (6.2.4) should be at most as high as the inverse of the bandwidth of the measurement system so that in the measurements, multipath components are isolated. The number of paths should be large enough that the rms multipath spread of the defined

channel profile fits the measured multipath spread of the application environment. The Doppler spectrum, $D(\lambda)$, should be assigned so that it reflects the speed of movement of the terminal and the multipath characteristics of the application environment.

Wideband propagation models provide a procedure for modeling wideband characteristics in different areas. Wideband models generally assume that the channel is subject to wide-sense stationary uncorrelated scattering (WSSUS), defined by scattering equation (6.2.4), in which $Q(\tau)$ is the discrete delay power spectrum and $D(f)$ is the continuous Doppler power spectrum of the channel. The discrete delay power spectrum is defined by a set of taps with specified arrival delays and average relative powers. The smaller number of taps in the model is preferred for computer simulations. The distance between the adjacent taps should be at most as high as the inverse of the bandwidth of the system so that in the simulated multipath, components are isolated. The Doppler power spectrum is defined by a continuous frequency function that specifies the distribution function and the spectrum of local shadow fading. The application environments are separated into different classes, and for each class a numerical table specifies the characteristics of individual taps. Each tap is implemented using the techniques described in Chapter 4 for simulation of narrowband signal characteristics. As we discuss later, for both indoor and urban radio channels, the path arrivals are random and correlated, which contradicts the WSSUS fixed-tap model. However, for all practical purposes, these simplified models are adequate to represent the channel for evaluation of the various techniques incorporated into wireless standards.

GSM-Recommended Model. The GSM group defined a set of channel profiles with discrete delay power spectrums of various lengths for rural areas, urban areas, and hilly terrains [GSM91]. The Doppler spectrum choices for each path or tap of the model are either Rician or the classical Rayleigh. In a manner similar to the simulation of narrowband signals, the Doppler power spectrum for the classical Rayleigh model is

$$D(f) = \frac{1}{2\pi f_m} \left[1 - \left(\frac{f}{f_m} \right)^2 \right]^{-1/2}, \quad -f_m < f < f_m \quad (6.2.5)$$

where $f_m = v_m/\lambda$ is the Doppler spread, v_m is the mobile vehicle velocity, and λ is the wavelength at the carrier frequency. The Rician spectrum is the sum of the classical Doppler spectrum and one direct path, weighted so that the total multipath power is equal to that of a direct path alone:

$$D(f) = \frac{0.41}{2\pi f_m} \left[1 - \left(\frac{f}{f_m} \right)^2 \right]^{-1/2} + 0.91\delta(f - 0.7f_m), \quad -f_m < f < f_m \quad (6.2.6)$$

To simulate the channel, the absolute power at each location is determined from the Okumura–Hata path-loss model, and similar to Fig. 6.3, each tap is implemented using the methods described in Section 4.5 for the simulation of narrowband signals. The following example provides more insight into the typical delay power spectrum or delay profile recommended by this standardization organization.

Example 6.1: GSM Delay Power Spectrum for Rural Areas Table 6.1 gives model parameters for “typical rural areas” as recommended in the GSM standard [GSM91].

TABLE 6.1 Typical Values of the Arrival Delay and Average Power for Rural Areas as Recommended by the GSM

Tap Number	Relative Time (μs)		Average Relative Power (dB)		Doppler Spectrum
	(1)	(2)	(1)	(2)	
1	0.0	0.0	0.0	0.0	Rice
2	0.1	0.2	-40	-2.0	Class
3	0.2	0.4	-8.0	-10.0	Class
4	0.3	0.6	-12.0	-20.0	Class
5	0.4		-16.0		Class
6	0.5		-20.0		Class

Source: [GSM91].

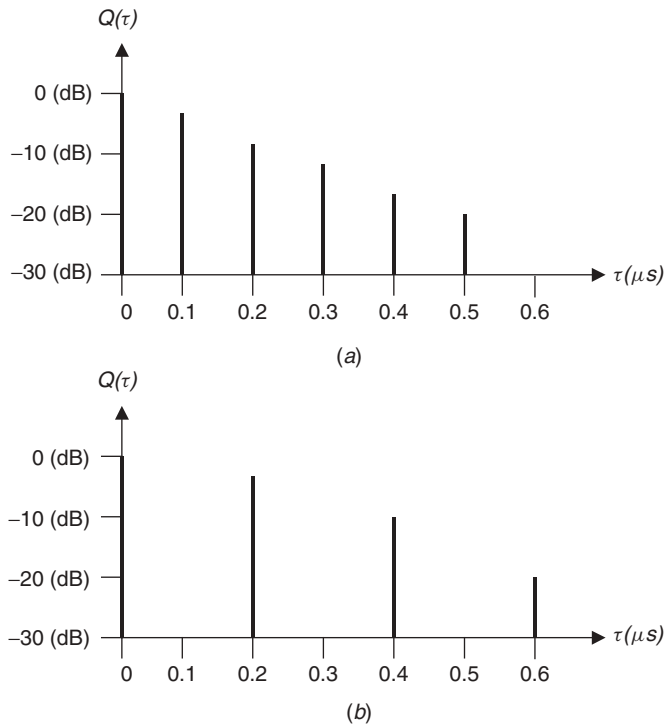


FIGURE 6.4 Two options for delay power spectra recommended by the GSM committee: (a) six taps; (b) four taps.

This model defines the discrete delay power spectrum with six taps, each with two alternative tap settings. Values for the delay and average relative power for the two choices are shown in the columns labeled (1) and (2). Figure 6.4 shows the two delay power spectra recommended for the rural areas by the GSM committee. Both provide the same rms delay spread, but the six-tap model provides a more refined model at

the expense of additional hardware for implementation. Since we want to evaluate the effects of multipath on the design of different modems, the two models should provide similar results. The bandwidth of the GSM channel is 200 kHz, resulting in pulse lengths of approximately 5 μ s. The multipath spread of the channel is around 10% of this value, which does not cause significant intersymbol interference. The first path in both profiles is assumed to have Rician distribution because they are assumed to be direct LOS. The remaining paths are assumed to have classical Rayleigh distributions.

Appendix 6A provides GSM-recommended tables for typical hilly terrain, and urban areas, together with tap settings to be used for testing receivers employing equalization. Equalizers are needed when the delay spread gets close to the pulse duration. Using these tables, one can implement hardware or software simulation of the mobile radio channel as recommended by the GSM standard. A slightly different version of this model is recommended by the COST207 committee for the simulation of GSM channels [COS86].

JTC Recommendation for PCS Bands. A more elaborate and comprehensive model is recommended by the PCS Joint Technical Committee (JTC) for simulation of radio propagation in different areas for 1900-MHz PCS bands [JTC94a]. This recommendation includes parameters for both indoor and outdoor channels. The path-loss model for this recommendation was discussed in Chapter 4. Here we discuss the multipath profile structures defined in [JTC94b]. This document provides a more straightforward presentation of multipath profiles, as it is compared with the earlier version provided in [JTC94a]. The general structure of the JTC model is the same as that of the GSM model, but the JTC model is more comprehensive. The JTC model divides the environments into one indoor and two outdoor classes. Indoor areas, in turn, are divided into residential, office, and commercial areas. The outdoor areas include urban high-rise, urban/suburban low rise, and outdoor residential areas. Each class of outdoor areas is divided into other classes, specified by the transmitter antenna height with respect to the tops of buildings. Each tap is simulated in the same way as we described for narrow-band signals. The model defines two types of Doppler spectra—classical Clarke–Jakes and flat—for each tap of the discrete-time model. The classical Clarke–Jakes spectra are similar to the spectra used in the GSM model. The flat spectrum is used for simulation of the Doppler spectrum in indoor areas and is defined by

$$D(f) = \frac{1}{2\pi f_m}, \quad -f_m < f < f_m \quad (6.2.7)$$

Because in the same area the multipath characteristics can be quite different from one radio link to another, this model suggests three different types of channel profiles for each environment, providing a wide variety of rms multipath delay spreads for each class of area. Table 6.2 shows the expected rms delay spreads of individual classes of profiles for all nine environments.

Example 6.2: JTC Model for Indoor Residential Areas Table 6.3 shows the relative delay, average power, and Doppler spectrum of the JTC-recommended taps for indoor residential areas. The tap gains are selected to generate the recommended rms delay spreads given in Table 6.2. Figure 6.5 shows the discrete delay power profiles related to residential areas. The rms delay spread for the three classes of profiles are 20, 70, and

TABLE 6.2 Delay Spread Parameters and the Probability of Selection of a Class of Impulse Delay Profile in All Areas as Recommended by the JTC

Environment	S_A (ns)	S_B (ns)	S_C (ns)	Table
Indoor residential	20	70	150	6.4
Indoor office	35	100	460	6B.1
Indoor commercial	55	150	500	6B.2
Outdoor urban high-rise; low antenna	100	750	1,800	6B.3
Outdoor urban/suburban low-rise; low antenna	100	750	1,500	6B.4
Outdoor residential; low antenna	70	460	850	6B.5
Outdoor urban high-rise; high antenna	500	3,250	8,000	6B.6
Outdoor urban/suburban low-rise; high antenna	400	4,000	12,000	6B.7
Outdoor residential high antenna	350	2,260	6,450	6B.8

Source: [JTC94b].

TABLE 6.3 Typical Arrival Delay and Average Power for the Taps in the Three Channel Models Suggested for the Residential Indoor Areas by the JTC

Tap	Channel A		Channel B		Channel C		Doppler Spectrum
	Relative Delay (ns)	Average Power (dB)	Relative Delay (ns)	Average Power (dB)	Relative Delay (ns)	Average Power (dB)	
1	0	0	0	0	0	0	Flat
2	100	-13.8	100	-6.0	100	-0.2	Flat
3			200	-11.9	200	-5.4	Flat
4			300	-17.9	400	-6.9	Flat
5					500	-24.5	Flat
6					600	-29.7	Flat

Source: [JTC94b].

150 ns, respectively. The tap gains are selected with a minimum of 100 ns spacing, and the amplitudes are selected to fit the rms delay spreads in Table 6.2 with 3% accuracy. To develop this model, a measurement system of 10-MHz bandwidth can be used to provide for a resolution of around 100 ns.

The JTC recommendation document [JTC94b] also provides specifications for the infinite impulse response (IIR) digital filters used for shaping the Doppler spectrum. Readers interested in implementation of such filters can refer to Project 2 in Chapter 4. Appendix 6B provides the details of the tap gains and delay power profiles for all of the other indoor and outdoor areas.

6.2.2 Wideband Models for Local Area Networks

Statistical modeling of the multipath profiles for the indoor radio channel was introduced by Saleh and Valenzuela at Bell Laboratories [Sal87b]. Based on limited measurements, they came up with a statistical time-domain model for simulation of the

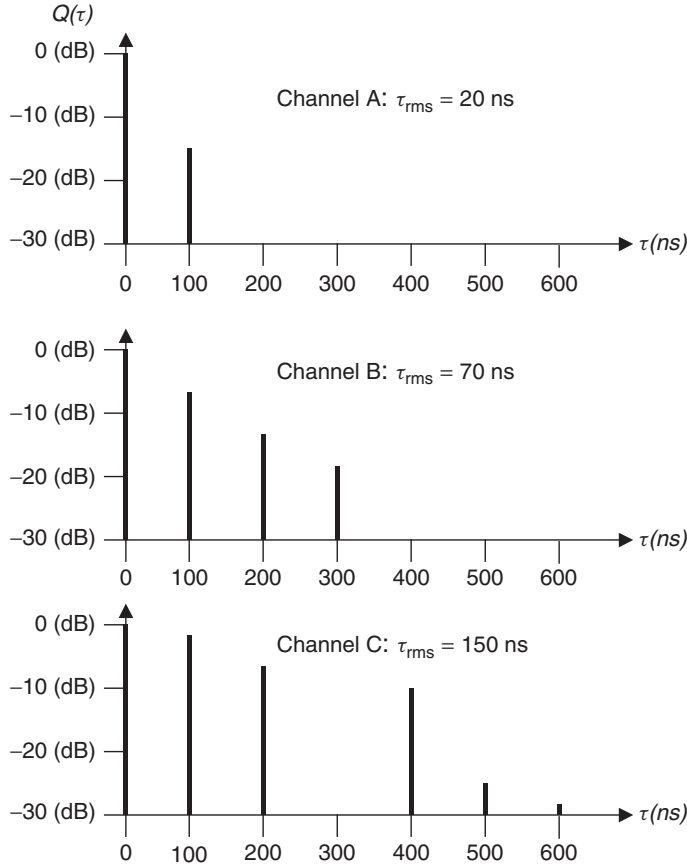


FIGURE 6.5 Three classes of channel profiles for indoor residential areas, specified by the JTC committee [JTC94b].

indoor radio channel. This work and its follow-up academic work [Gan91a, Gan91b, Rap91b, Yeg91] were used in the late 1990s for measurement and modeling of the angle of arrival [Spe00, Tin00]. More recently, with the popularity of the WLAN and WPAN applications, and introduction of more advanced MIMO and UWB technologies for them, these models were used by the IEEE 802.11 and IEEE 802.15 community to develop practical models for comparative performance evaluation of different implementation alternatives. In this section we start by introducing the simple IEEE 802.11b and basic Saleh–Valenzuela model. We show the shortcomings of these models for UWB and MIMO applications before we introduce the Intel model for IEEE 802.15 UWB applications and Spencer’s model used for the IEEE 802.11n model.

IEEE 802.11b Model. Similar to the GSM and JTC models, the IEEE 802.11 model documented in [IEE00] assumes discrete channel impulse responses. Similar to the Saleh–Valenzuela model, the delay power spectrum for the IEEE 802.11b model holds that an exponentially decaying function and path amplitude form a Rayleigh distribution. But unlike any of those models, the distance between the amplitude samples is constant in the IEEE 802.11 model. Starting with Eq. (6.2.2), the discrete channel impulse

response of the IEEE 802.11 model is given by

$$h(\tau) = \sum_{k=1}^{K_{\max}} \beta_k \delta(\tau - kT_s) e^{j\phi_k} \quad (6.2.8a)$$

where

$$k_{\max} = \left\lceil \frac{10\tau_{\text{rms}}}{T_s} \right\rceil, \quad \overline{|\beta_k|^2} = \overline{|\beta_0|^2} e^{-kT_s/\tau_{\text{rms}}}, \quad \overline{|\beta_0|^2} = \frac{1 - e^{-T_s/\tau_{\text{rms}}}}{1 - e^{-(k_{\max}+1)T_s/\tau_{\text{rms}}}} \quad (6.2.8b)$$

in which $\lceil \cdot \rceil$ means the closest integer of “ $\cdot$,” τ_{rms} is the rms delay spread in the given indoor area, and T_s is the sampling interval considered for realization of the channel impulse response of Eq. (6.2.8a). Similar to Eq. (6.2.4), the delay power spectrum is given by

$$Q(\tau) = \sum_{k=1}^{K_{\max}} |\beta_k|^2 \delta(\tau - kT_s) \quad (6.2.8c)$$

Figure 6.6 shows the general concept for the definition of this exponential profile for the delay power spectrum. The filter taps are independent complex Gaussian variables with an average power profile that decays exponentially. For the simulation of the

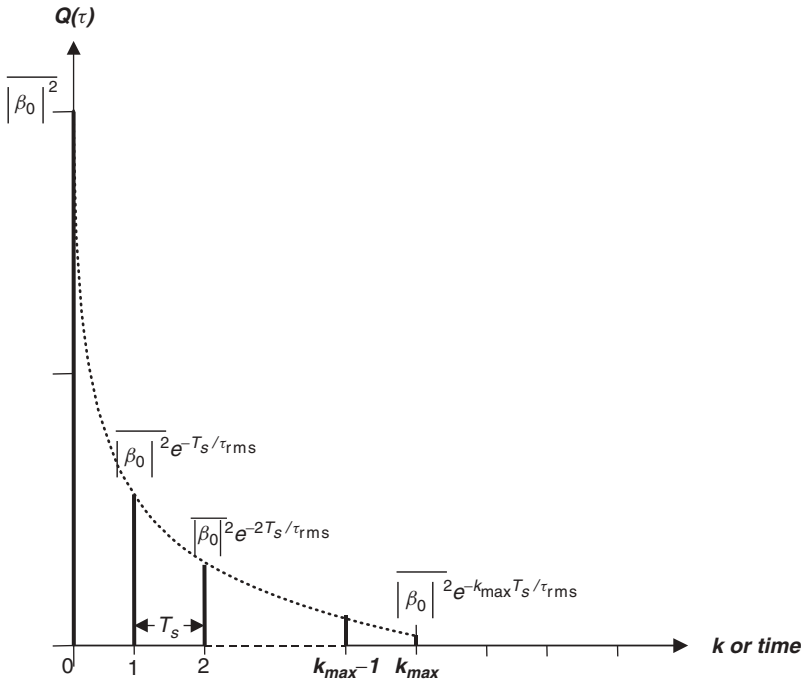


FIGURE 6.6 Normalized delay power spectrum recommended by the IEEE 802.11b community.

channel, each sample of the channel impulse response is generated from a filtered complex Gaussian random variable,

$$h(k) = h(\tau)|_{\tau=kT_s} = N(0, \frac{1}{2}\overline{|\beta_k|^2}) + jN(0, \frac{1}{2}\overline{|\beta_k|^2}) \quad \text{for } k = 0, 1, \dots, k_{\max} \quad (6.2.9)$$

which creates Rayleigh amplitude fading with variance $\overline{|\beta_k|^2}$ and a uniform distribution between $\{0, 2\pi\}$ for the phase ϕ_k . The value of $\overline{|\beta_0|^2}$ in Eq. (6.2.8b) is selected so that the addition of all samples of the delay power spectrum in Eq. (6.2.8c) is normalized to 1.

Saleh–Valenzuela Model. The basic model introduced by Saleh and Valenzuela [Sal87b] characterizes the channel impulse response with Eq. (6.2.2). As discussed earlier in this chapter, this model is based on a limited database collected in one of the Bell Lab buildings, and in its simplest form it assumes that the amplitude of the path follows a Rayleigh distribution, the arrival of the paths forms a Poisson random variable, and the envelope of the delay power spectrum forms an exponential function. In a more complete description, they also assume that the paths arrive in clusters, the strength of clusters forms another exponential function with slower decay, and the arrival of clusters follows another Poisson distribution with slower rate. This model originally motivated further measurements and modeling in a variety of buildings [Gan91a, Gan91b], as described in Section 6.2.3. More recently, this model motivated modeling of the AOA [Spe00] that is described at the end of this part of our discussion.

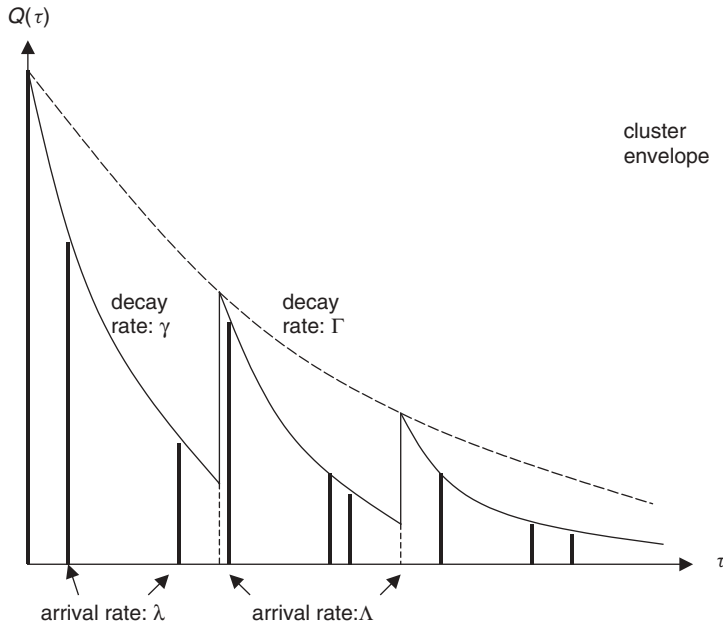


FIGURE 6.7 Basic concept for the Saleh–Valenzuela's model.

Figure 6.7 illustrates the basic concept of clusters and ray arrivals in the clusters. The overall impulse response of Eq. (6.2.2) is now represented by

$$h(\tau) = \sum_{l=0}^{\infty} \sum_{k=1}^{\infty} \beta_{kl} \delta(\tau - T_l - \tau_{kl}) e^{j\phi_{kl}} \quad (6.2.10a)$$

where the sum over l represents the clusters, the sum over k represents the ray arrivals within each cluster, and T_l is the delay of the l th cluster. The strength of the paths within the clusters is determined from

$$\overline{|\beta_{kl}|^2} = \overline{|\beta_{00}|^2} e^{-T_l/\Gamma} e^{-\tau_{kl}/\gamma} \quad (6.2.10b)$$

in which $\overline{|\beta_{00}|^2}$ is the average power of the first path in the first cluster and Γ and γ are the decay rates associated with the clusters and the rays within the clusters, respectively. Since the arrivals of the clusters and the rays within the clusters form Poisson processes, the interarrival rate of the clusters and rays form exponential distributions given by

$$\begin{aligned} p\left(\frac{T_l}{T_{l-1}}\right) &= \Lambda e^{-\Lambda(T_l - T_{l-1})} \\ p\left(\frac{\tau_{kl}}{\tau_{(k-1)l}}\right) &= \lambda e^{-\lambda(\tau_{kl} - \tau_{(k-1)l})} \end{aligned} \quad (6.2.10c)$$

in which Λ and λ represent the cluster and ray arrival rates. The parameters estimated from the data collected by Saleh and Valenzuela [Sal87] are $\Gamma = 60$ ns, $\gamma = 20$ ns, $1/\Lambda = 300$ ns, and $1/\lambda = 5$ ns. To determine the parameters of the model, Saleh and Valenzuela adjust them to fit the cumulative distribution function of the measured data collected at Bell Lab. Figure 6.8 shows the results of simulation versus the measurement data [Sal87b].

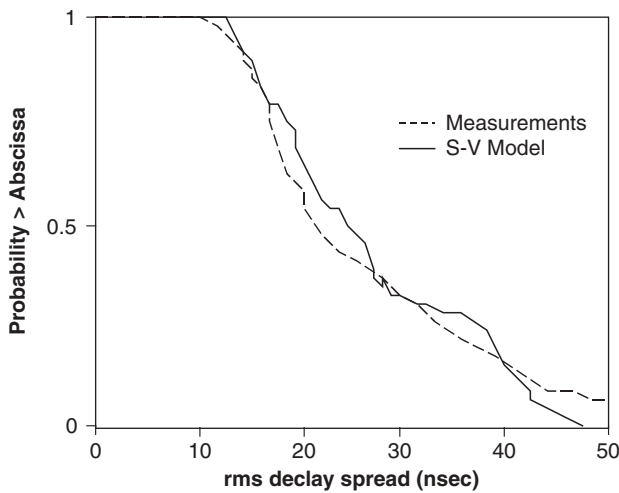


FIGURE 6.8 Cumulative distribution function of the measurements at Bell Labs and the results of the Saleh-Valenzuela model [Sal87b].

6.2.3 Direct Modeling of Path Arrivals and Amplitude

As we discussed in Section 6.2.2, Saleh and Valenzuela's model is based on the assumption that paths arrive in clusters with Poisson distribution and that the amplitude of each path is Rayleigh distributed. The Rayleigh model is a good fit for narrowband signals because arriving paths with random phase and amplitude adds up to create multipath fading. When the bandwidth of the system increases to UWB signals, similar to Example 3.6, the received signal power averages over a wide frequency spectrum, eliminating the effects of frequency-selective multipath fading and resulting in stable received signal strength. With the elimination of Rayleigh multipath fading, the only cause of changes in the received power of individual paths is the lognormal shadow fading. In addition, if we assume that paths arrive in clusters, it is counterintuitive to assume that they are uncorrelated. The Poisson model, on the other hand, assumes uncorrelated arrivals and so cannot be the best model for the arrival of the paths. Direct modeling of the arrival and amplitude statistics based on wideband indoor radio measurements was first reported in [Gan91a, Gan91b], and we provide a summary of these results in this section. These results are important because they laid a foundation for modification of the Saleh–Valenzuela model to fit UWB channel characteristics that we discuss in Chapter 12.

Path Arrival Times. A simple statistical model for path arrival is a *Poisson process*, the model typically used for characterizing random arrivals in queuing theory analysis. On indoor and outdoor radio links, if the objects causing the multipath are located randomly throughout the space surrounding the link, the Poisson distribution should provide a good model for path arrivals. However, the results of several studies of urban [Tur72, Suz77] and indoor [Gan89, Yeg91, Has93b] radio environments have shown that the Poisson distribution does not closely match the results of empirical measurements. This observation suggests that on indoor and urban radio channels, the spatial distribution of the objects causing multipath cannot be described accurately as being totally random. In this section, closely following the experimental results of Ganesh and Pahlavan [Gan89], we provide an explanation of this phenomenon.

To evaluate the accuracy of the Poisson model for path arrivals, we examine the results, described in Chapter 5, of wideband indoor radio propagation measurements made in manufacturing and office areas. The path arrival distribution given by the Poisson model is compared with the empirical data to determine the degree of closeness. The time axis of each measured time-domain channel profile is divided into bins of width $\Delta = 5$ ns, which is the pulse width used by the measurement system. The existence of a path in a bin is determined by comparing the peak value of the signal in each bin with a certain threshold set according to the level of the background noise. If the peak value is higher than the threshold, we declare that a path exists in the bin. The number L of paths in the first N bins of each measured profile is determined. Then the probability of having L paths over all the measured profiles is calculated (the first path, which always exists and serves as the reference for the delay times, is not included in the calculation). To determine the empirical path index distribution, the probability of receiving l paths in the first N bins $P_N(l)$ is plotted against l . This procedure is repeated for $N = 5, 10, 15$, and 20 bins.

The Poisson process is a one-parameter model of “totally random” events occurring at a fixed average rate λ . The probability $P_N(l)$ for a theoretical Poisson path index

distribution is given by

$$P_N(l) = \frac{\lambda^l}{l!} e^{-\lambda} \quad (6.2.11a)$$

where l is the path index and λ is the mean path arrival rate, given by

$$\lambda = \sum_{i=1}^N r_i \quad (6.2.11b)$$

In this equation, r_i is the path occurrence probability for bin i , defined as the ratio of the number of times we have detected a path in bin i to the total number of profiles used for statistical modeling.

Figures 6.9 and 6.10 provide a comparison between the empirical path index distributions and the theoretical Poisson path index distributions [Gan89] for $N = 5, 10, 15$, and 20 bins. The figures correspond to the manufacturing floor areas and office areas, respectively. For clarity, the results are plotted as continuous curves, although they have values only for integer path numbers. We observe considerable discrepancy between the empirical and Poisson distributions for all values of N , irrespective of the environment. This discrepancy reflects a tendency for the paths to arrive in groups rather than in a random manner.

To explain these discrepancies, a modified Poisson model, also referred to as the $\Delta-K$ model,¹ was proposed by Suzuki [Suz77] for characterizing urban radio channels.

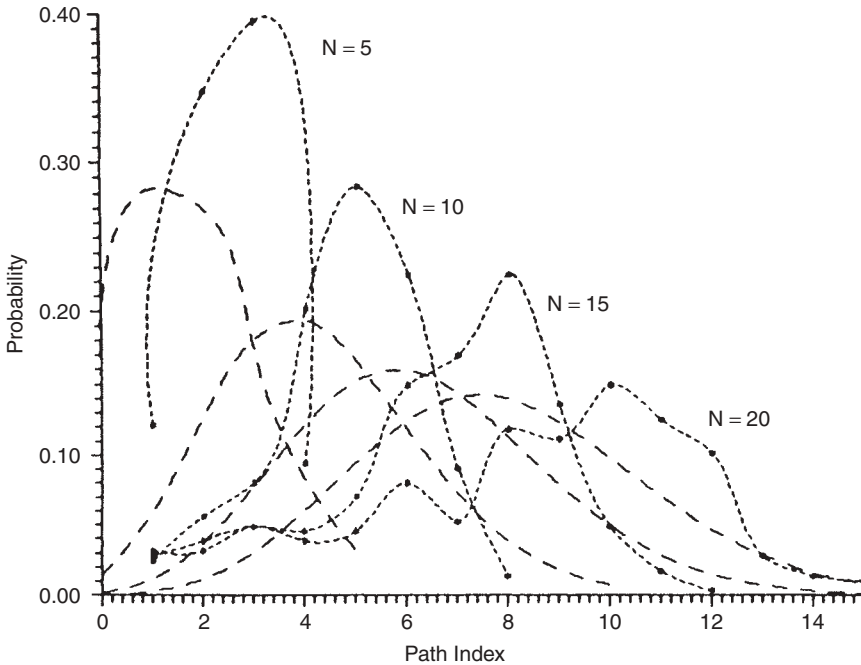


FIGURE 6.9 Empirical (· · ·) and theoretical (—) Poisson path index for manufacturing floors.

¹Since Δ is the width of the bins and K is a representative of correlation among neighboring paths, the model is called $\Delta-K$.

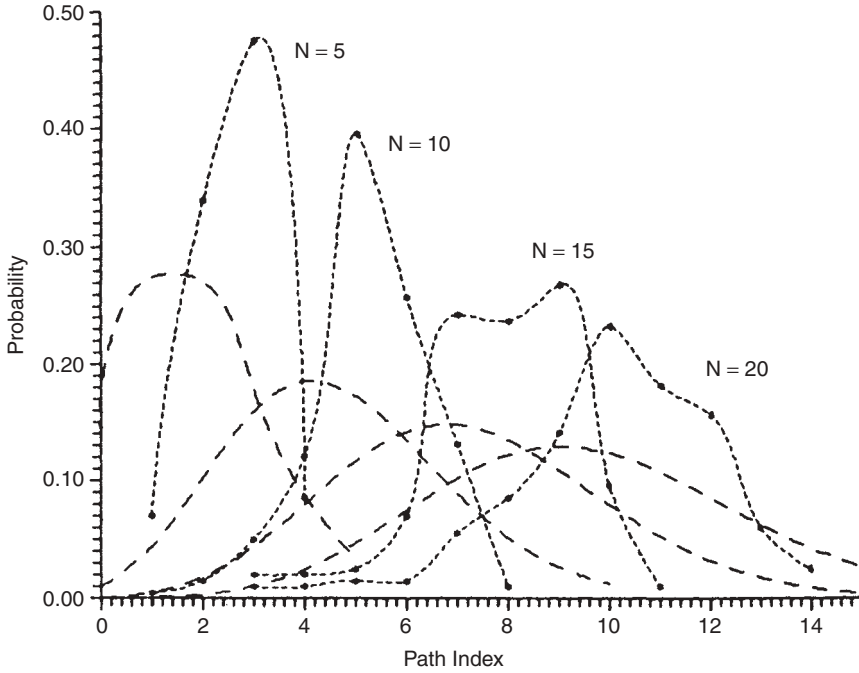


FIGURE 6.10 Empirical (···) and theoretical (—) Poisson path index for offices.

This model was subsequently extended to indoor radio propagation [Gan89]. Figure 6.11 summarizes the modified Poisson process. For the modified Poisson process, the probability of having a path in bin i is given by λ_i if there was no path in the $(i - 1)$ st bin, or by $K_N \lambda_i$ if there was a path in the $(i - 1)$ st bin. The “underlying” probabilities of path occurrences λ_i are related to the empirical path occurrence probabilities r_i by

$$\lambda_i = \frac{r_i}{(K_N - 1)r_{i-1} + 1}, \quad i \neq 1 \quad (6.2.12a)$$

where $\lambda_1 = r_1$.

The modified Poisson path index distribution is related to the $\{\lambda_i\}$ by the following recursive equations [Suz77]:

$$\begin{aligned} P_i(l) &= P_{1,i}(l) + P_{2,i}(l) \\ P_{2,i+1}(l) &= P_{2,i}(l-1)K_N\lambda_{i+1} + P_{1,i}(l-1)\lambda_{i+1} \\ P_{1,i+1}(l) &= P_{2,i}(l)(1 - K_N\lambda_{i+1}) + P_{1,i}(l)(1 - \lambda_{i+1}) \end{aligned} \quad (6.2.12b)$$

where $P_i(l)$ is the probability of having l paths in the first i bins, $P_{1,i}(l)$ is the probability of having l paths in the first i bins conditioned on having no path in the i th bin, and $P_{2,i}(l)$ is the probability of having l paths in the first i bins conditioned on having one path in the i th bin. The process begins in bin 1, where $P_{1,1}(0) = 1 - \lambda_1$, $P_{2,1}(1) = \lambda_1$, $P_{1,1}(1) = 0$ for $l \geq 1$, and $P_{2,1}(l) = 0$ for $l \geq 2$ or $l \leq 0$. Starting with a small value of K_N and minimizing the mean-squared error between the empirical distribution and

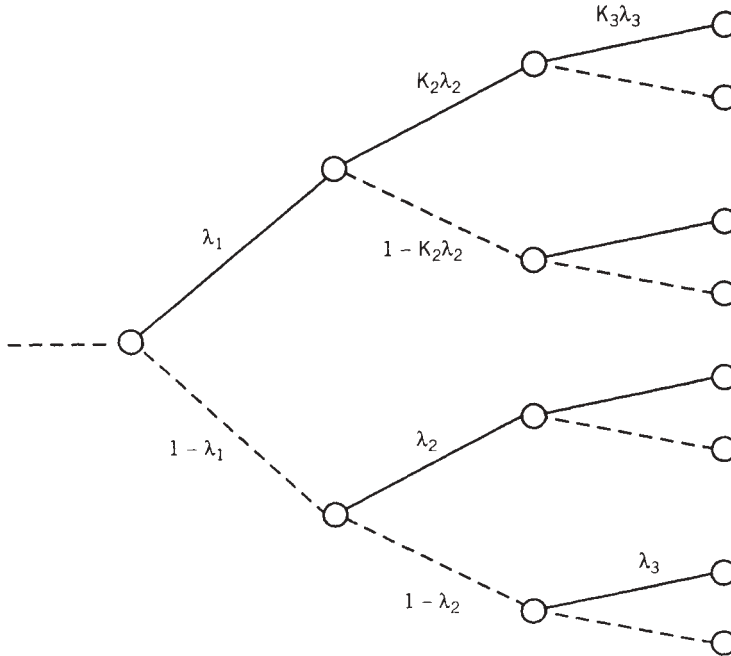


FIGURE 6.11 Tree structure representing the modified Poisson process.

theoretical modified Poisson path index distribution found using the equations above, optimum values of K_N are found from the data for $N = 5, 10, 15$, and 20 bins. To aid in simulation, optimum values of K_N , which are functions of the number of bins N , are replaced by new parameters $\{K_i\} (i = 1, 2, \dots, 20)$, which are functions of the bin numbers $\{i\}$ [Has79]. The $\{K_i\}$ are determined by linear interpolation. For final calculation of the modified Poisson path index distribution, the equations above are used again with interpolated $\{K_i\} (i = 1, 2, \dots, 20)$ replacing $K_N (N = 5, 10, 15, 20)$. Table 6.4 shows optimum values of K_i and λ_i calculated for the measurements made in the manufacturing floors and college office areas.

Figures 6.12 and 6.13 provide a comparison of the empirical path index distributions with the modified Poisson distributions for the manufacturing floors and college office areas, respectively. The curve fittings show considerable improvement over those shown in Figs. 6.9 and 6.10 for the Poisson model. This suggests that the paths do not arrive randomly but in groups, and the presence of a path at a given delay is greatly influenced by the presence or absence of a path in earlier bins. In mathematical terms, the modified Poisson model utilizes the empirical probability of occurrence for each bin, whereas the Poisson model simply uses the sum of the probabilities of occurrence for all bins.

Two other approaches are used to modify the Poisson arrival model. The first approach, suggested in [Sal87b] for indoor radio propagation, assumes that the paths arrive in clusters. The path arrivals in each cluster, and the arrivals of the clusters, are both assumed to be Poisson processes. The problem encountered with this approach is that there is no reliable way of directly identifying the clusters from the results of measurements. This prevents us from developing a logical and systematic means of

TABLE 6.4 Parameters K and λ for the Modified Poisson Process in Manufacturing Floors and Office Areas

Bin	Manufacturing Floors		College Office Areas	
	K_i	λ_i	K_i	λ_i
1	0.964	0.4513889	0.697	0.3636364
2	0.932	0.5418182	0.666	0.6408464
3	0.900	0.5217417	0.634	0.6288999
4	0.868	0.4942650	0.602	0.7164743
5	0.836	0.5146308	0.570	0.7313251
6	0.804	0.5442811	0.539	0.7230882
7	0.772	0.5484897	0.507	0.7350161
8	0.740	0.5127087	0.475	0.8171411
9	0.708	0.5711924	0.443	0.7717720
10	0.676	0.3649115	0.411	0.6645216
11	1.340	0.3444129	0.589	0.6733878
12	1.440	0.2928466	0.595	0.6594670
13	1.540	0.3592564	0.602	0.6470692
14	1.636	0.2782416	0.608	0.6743552
15	1.732	0.1192904	0.614	0.4608082
16	5.796	0.1300228	1.056	0.4402552
17	6.388	0.1044048	1.117	0.4432566
18	6.980	0.1019357	1.177	0.3898038
19	7.572	0.0901229	1.238	0.3576177
20	8.164	0.0810876	1.299	0.3297661

determining the model parameters from measurement data. In only a small fraction of measurements can one recognize a pattern of having more than one cluster. We should also caution that the model developed in [Sal87b] is based on a rather limited set of measurement data.

The second approach to modifying the Poisson process is to analyze interarrival delays. The interarrival delay for the Poisson distribution is exponentially distributed, but one may consider other distributions for the interarrival times. It is shown in [Yeg91], based on extensive measurements on manufacturing floors, that the Weibull distribution best fits the interarrival delays of the time-domain model. The Weibull distribution has multiple parameters and therefore can provide a better fit to any empirical distribution than can the one-parameter Poisson model. However, unlike the Poisson model and its variations, the Weibull-based model lacks an obvious physical interpretation.

Path Amplitudes. The simplest method of modeling path amplitudes is to assume that each measured path is the phasor sum of several paths arriving so close to one another that they are not distinguishable by the measurement system. With this assumption, the amplitude fluctuation of each path follows a statistical pattern similar to that of the amplitude of a narrowband signal. The small-scale variations form Rayleigh and Rician distributions for the obstructed-line-of-sight (OLOS) and LOS

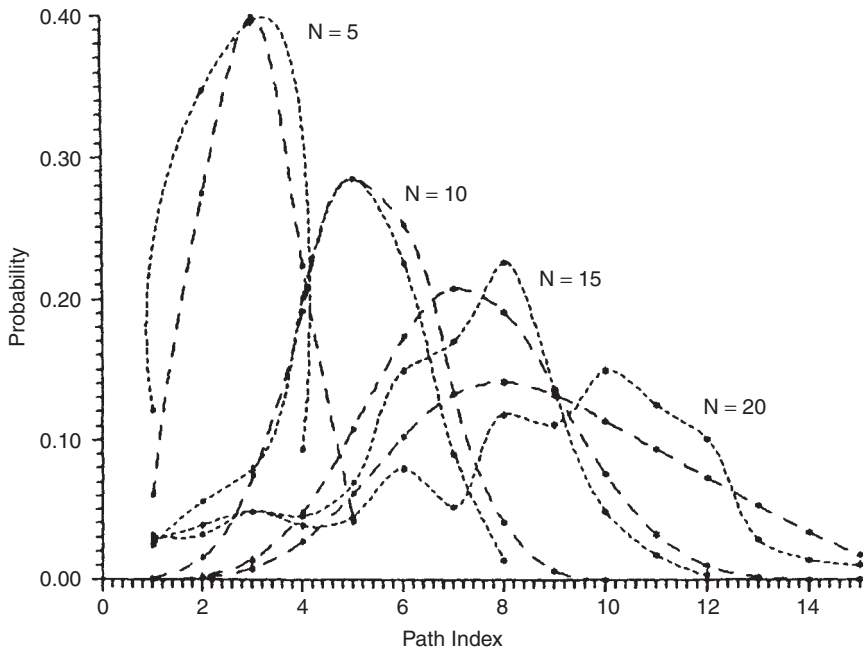


FIGURE 6.12 Empirical (···) and theoretical modified (—) Poisson path index for manufacturing floors. An asterisk denotes that a clustering property is exhibited.

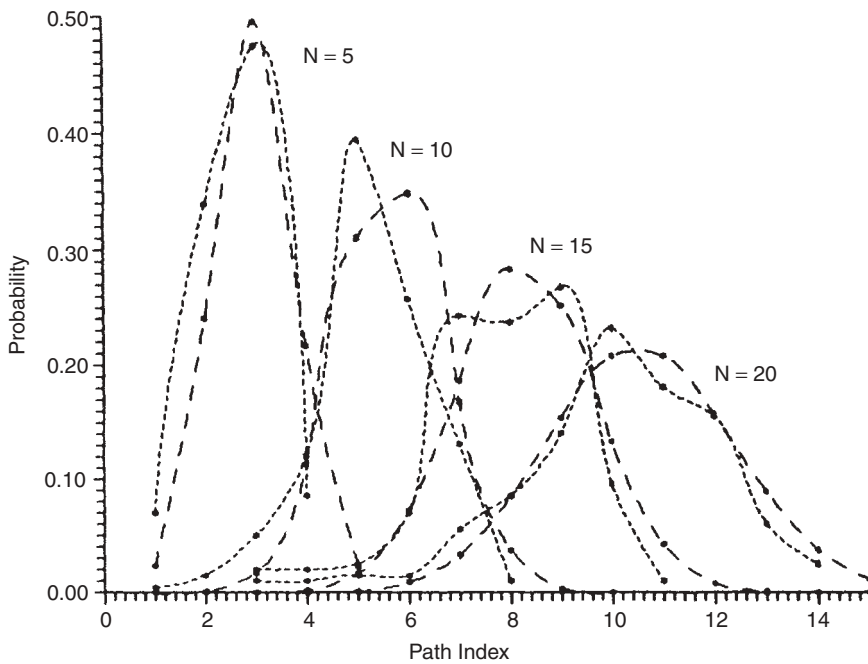


FIGURE 6.13 Empirical (···) and theoretical modified (—) Poisson path index for offices.

cases, respectively. Large-scale variation of the mean of the amplitude fluctuations is modeled by a lognormal distribution. The Doppler spectrum of each path would follow the Jakes spectrum for mobile radio applications and would follow uniform spectra for indoor wireless applications.

To justify the validity of these assumptions, we examine the distribution of path amplitudes measured in the two wideband indoor radio propagation experiments discussed in Section 6.2.2. The discussion closely follows the experimental results reported by Ganesh and Pahlavan [Gan91b]. We divide the time scale into 5-ns bins and record the path amplitudes in the bins found to contain paths. The statistics of the amplitude fluctuations are then analyzed for each bin. The curve-fitting approach introduced in Chapter 4 is used to find the distribution of path amplitudes in each bin. Figures 6.14 and 6.15 give comparisons between the theoretical and experimental distributions for bin 1 in manufacturing floor areas and bin 5 in college office areas, respectively. The horizontal axis is normalized to the measured median signal value in decibels. Evaluation of the distribution over individual bins reveals that the lognormal and Suzuki distribution functions provide the closest fit to the measured data. The lognormal assumption is consistent with the previous models, and computer simulation with this approach is simpler than with the other models. Therefore, we judge this method to be the preferred approach to modeling of received amplitudes.

The inhomogeneities of the radio channel result in variations in the mean and variance of the amplitudes from one delay to another. To simulate these changing parameters, we need to know the distribution of their variations. Scatter plots of the mean and standard deviation of the lognormal distribution versus delay were fitted to decaying exponentials of the form $Ae^{-\tau/T} + B$, where T is the decay rate, τ is the delay, and A and B are constants [Gan91a]. Note that the decay rate is defined with respect to the bin number, which models the arrival delay relative to the delay of the first arriving path. Figures 6.16 and 6.17 show the mean and standard deviation

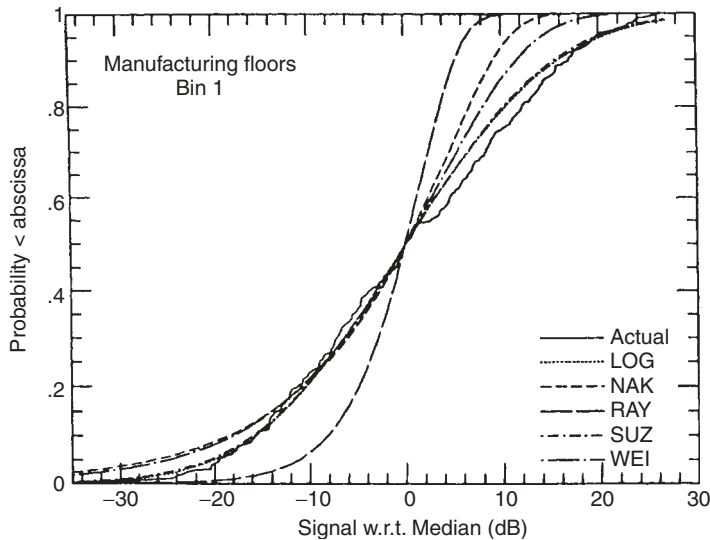


FIGURE 6.14 Theoretical amplitude CDFs and the empirical CDF for the measured data in bin 1.

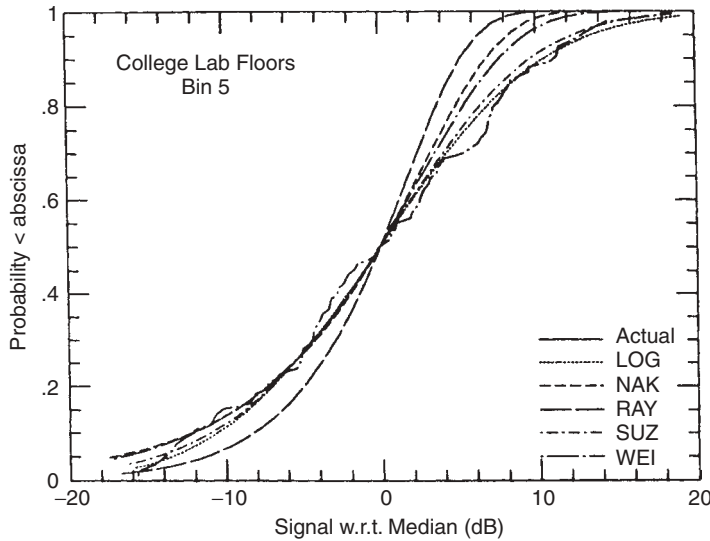


FIGURE 6.15 Theoretical amplitude CDFs and the empirical CDF for the measured data in bin 5.

and their exponential fits for the manufacturing floor and college office areas, respectively. For the manufacturing areas, the decay rate T was 5.0 for the mean and 4.0 for the standard deviation. In these areas most of the received power is concentrated in the earliest-arriving paths, resulting in a faster decay of the received power with delay. On the other hand, the college office areas exhibited a wider spread of power in delay and thus a slower decay for the received power with delay. In the college offices, the decay rate T was 35.5 for the mean and 28.3 for the standard deviation. Note that the scatter plots give the values of the mean and standard deviation for the path amplitude given the existence of a path at that delay. The modified Poisson process determines whether or not a path exists at a given delay. Figure 6.18 shows the average received power versus the delay of the path arrival and the best-fit exponential function found for measurements made in the manufacturing and college office areas. This function is the equivalent spatial delay power spectrum in which the power at different delays is determined by averaging over the measurements in an area rather than by classical averaging over time. As we will see in the following section, the equivalent delay power spectrum is useful in the calculation of modem BERs.

Simulation of the Channel Impulse Response. Here we examine the sensitivity of computer simulation results to the measurement-based statistical models used to represent the path delays and amplitudes in the simulation. The database used here is the set of measurements made in the manufacturing and office areas discussed in Chapter 5. To observe the influence of a statistical model on computer simulation results, we consider two statistical models and compare the CDFs of the rms delay spreads derived from simulations with the results from empirical data. The first statistical model assumes Poisson and Rayleigh distributions for the arrival times and amplitudes of the paths, respectively. This model is similar to the model suggested

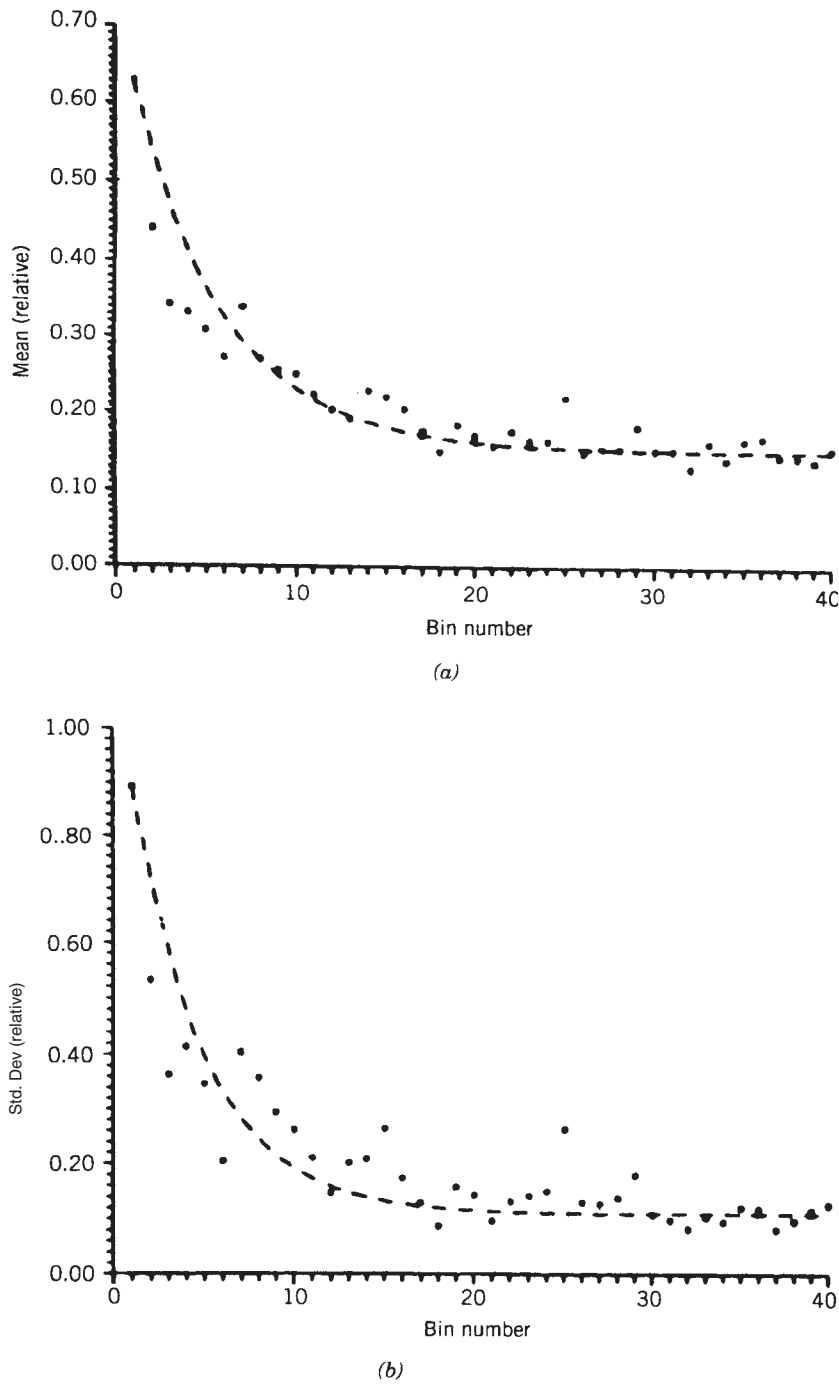


FIGURE 6.16 Mean (a) and standard deviation (b) of the arriving path amplitudes as a function of the path arrival delay, and best exponential fit, for manufacturing floors.

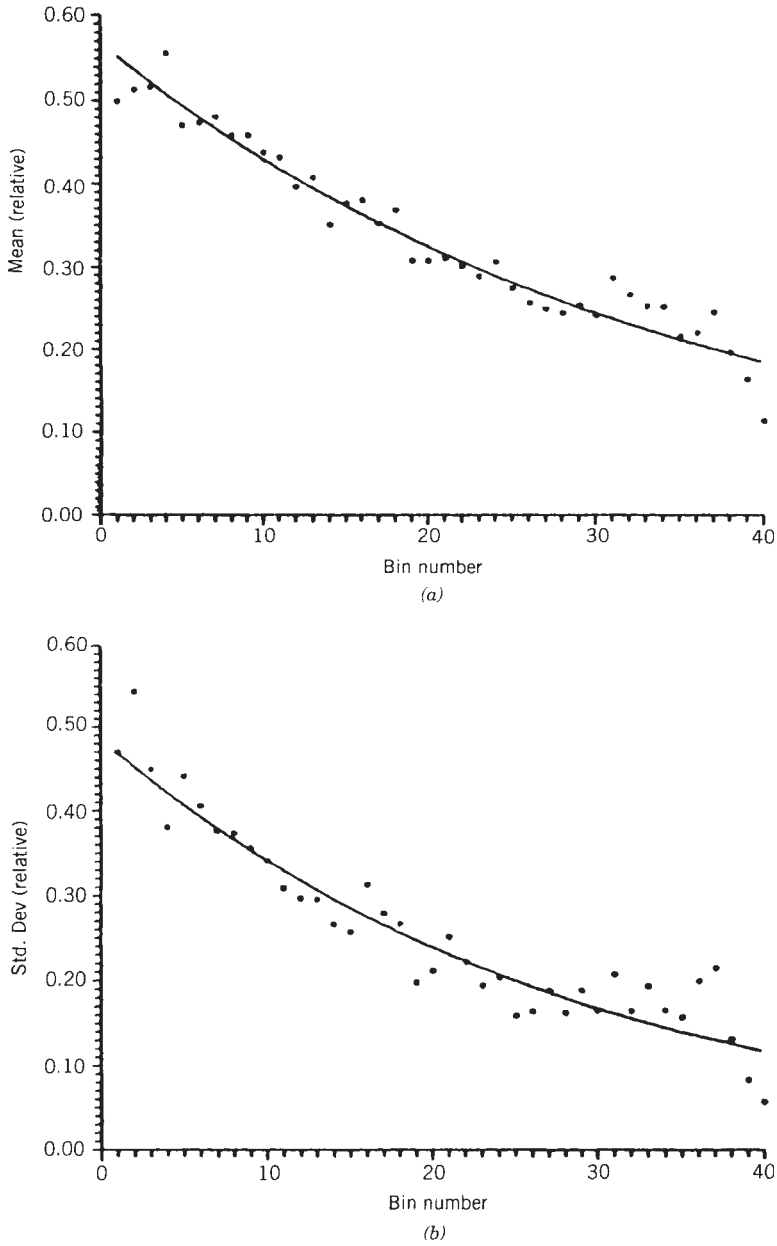


FIGURE 6.17 Mean (a) and standard deviation (b) of the arriving path amplitudes as a function of the path arrival delay, and best exponential fit, for office areas.

in [Sal87b] except that the statistics of the magnitudes and arrival times of the paths are extracted from the empirical data. Because we were unable to identify clusters directly from the individual measurements, we have assumed only one cluster. The second model, suggested in [Gan91b], assumes a modified Poisson and a lognormal distribution for the arrival times and amplitudes of the paths, respectively.

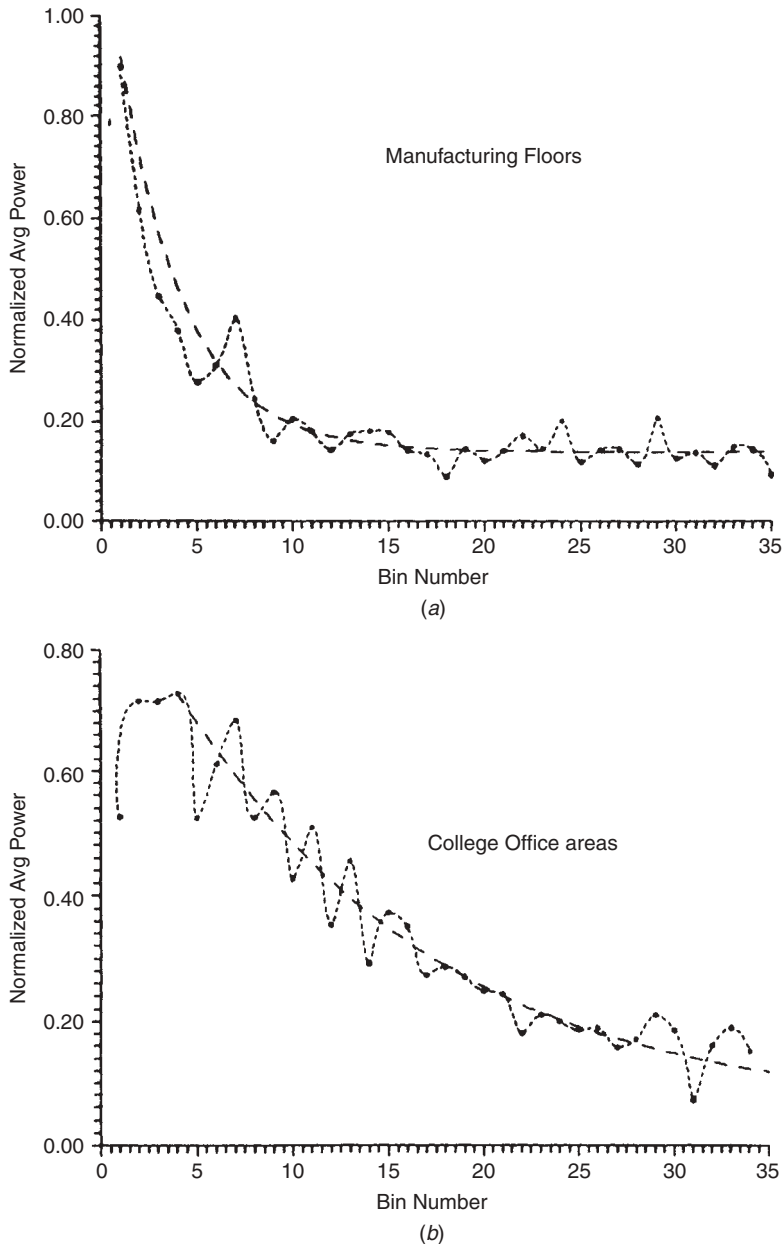


FIGURE 6.18 Average received power versus delay path arrival, and best exponential fit: (a) manufacturing floors; (b) college office areas.

For the Poisson-Rayleigh model, the mean path arrival rates obtained from the measurement data gathered in the manufacturing and office areas were used to determine the presence or absence of a path in any bin. The measured signal powers in each bin were then used to determine the Rayleigh amplitude of an existing path. For the modified Poisson-lognormal model, Fig. 6.11 and Table 6.4 were used to determine

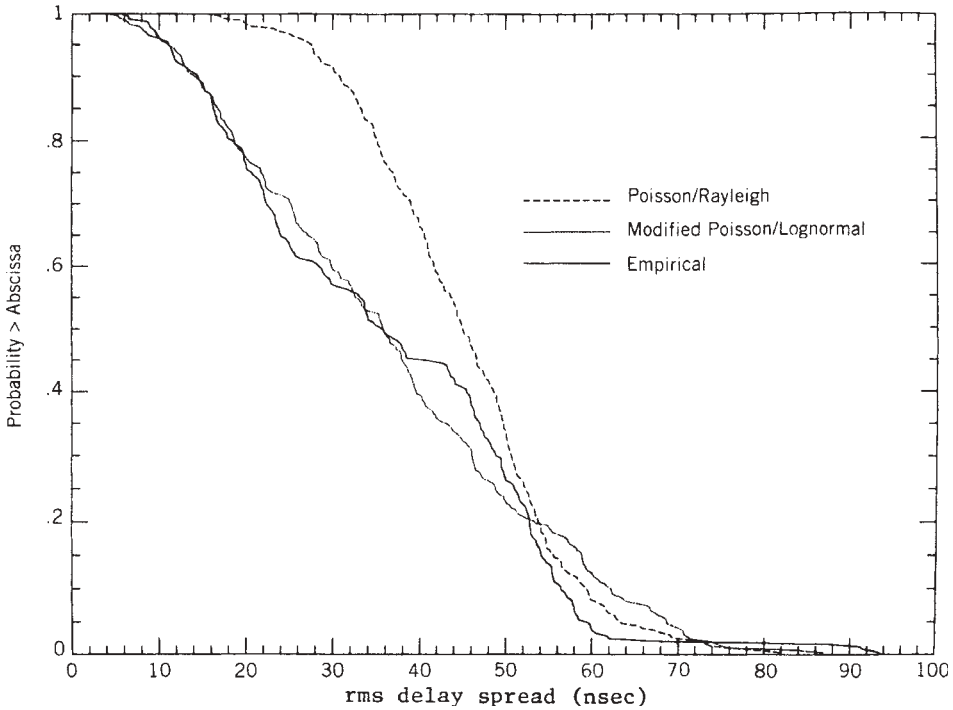


FIGURE 6.19 CDFs of the rms multipath delay spread for the Poisson-Rayleigh and modified Poisson-lognormal models and the results of empirical measurements in manufacturing areas.

the presence of a path in a bin, and the amplitudes of the paths were determined by exponential fits for the mean and variance of the lognormal distribution discussed in Section 6.2.2.

The channel profiles were simulated for the Poisson-Rayleigh and modified Poisson-lognormal distribution models. The rms delay spread of the results of the two simulations and the results from the empirical data for manufacturing and office areas are shown in Figs. 6.19 and 6.20, respectively. The dashed lines in these figures are the cumulative distributions of the rms delay spreads computed from the two simulations; the solid lines represent the results of actual measurements. The match between the empirical and simulated distributions for the Poisson-Rayleigh distribution is seen to be very good. The match between the modified Poisson-lognormal simulation and the empirical results is even better than that provided by the Poisson-Rayleigh simulation.

6.2.4 UWB Models for Local Area Networks

Observations discussed in Section 6.2.3 turned out to be useful for the recent modeling of UWB channel characteristics in the IEEE 802.15 committee. Based on measurement of the UWB channel characteristics, a Δ - K model for arrival times and a lognormal distribution for the amplitudes of the paths were studied by the committee. We provide more details of this model in Chapter 12.

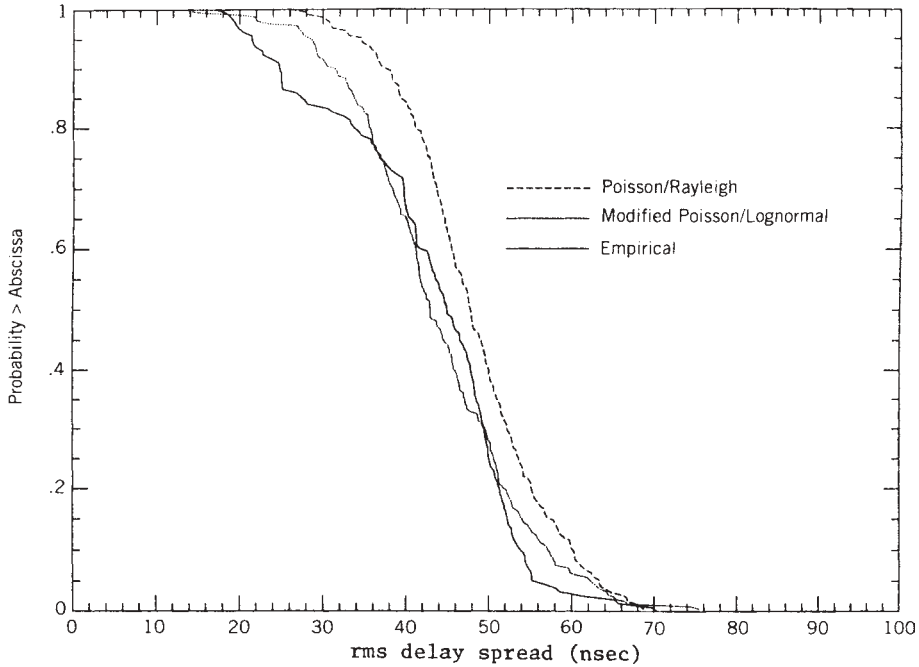


FIGURE 6.20 CDFs of the rms multipath delay spread for the Poisson–Rayleigh and modified Poisson–lognormal models and the results of empirical measurements in office areas.

6.2.5 Simulation of AOA for MIMO Channels

In this section we discuss new radio channel models emerging for multiple-input multiple-output (MIMO) antenna systems. Research and development in the WLAN industry use MIMO and OFDM technology to increase the data rate and coverage of WLANs. Researchers in cellular systems expect that fourth-generation cellular systems using MIMO technology with smart antennas and adaptive antenna arrays will increase the capacity of existing systems by an order of magnitude. Analysis of MIMO systems requires channel models to reflect the *angle of arrival* (AOA) of multipath components in order to steer antenna beams in the right directions. Until recently, modeling the AOA of the paths had not attracted serious attention because (1) there was no real demand for a popular application such as MIMO to drive the effort, and (2) measurement of the angle of arrival of the paths is far more complex than measurement of the time of arrival of the paths. In Section 5.4.2 we described several techniques for measurement of the AOA; in the remainder of this section we discuss modeling of the AOA with particular attention to the IEEE 802.11n channel model developed for MIMO systems. Early literature on the modeling of AOA is reported in [Hed97, Zwi98]. A more comprehensive model that became popular within the IEEE 802.11n standardization activities is introduced by Spencer et al. in [Spe00], which we describe below. This model is an extension of the Saleh–Valenzuela model described in Section 6.2.1, which includes the angle of arrival in the overall multipath profile.

Spencer’s Model for AOA. The Saleh–Valenzuela model assumes that paths arrive in clusters and that the rays within the clusters and the clusters themselves form two

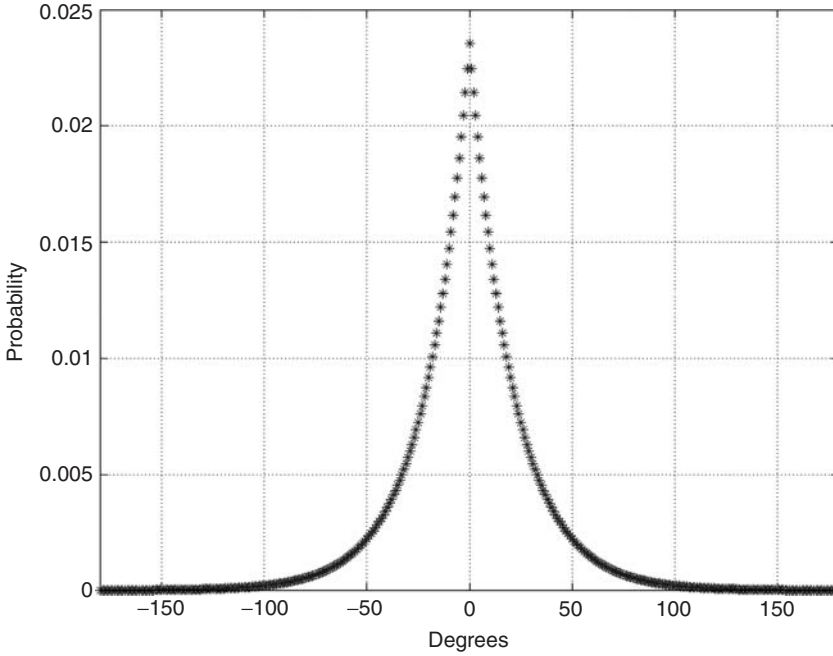


FIGURE 6.21 Laplacian distribution function with a standard deviation of 30° .

independent Poisson processes. Spencer's model associates an angle of arrival with each path of the Saleh–Valenzuela model given in Eq. (6.2.10). The angles are also assumed to form clusters. Starting from Saleh–Valenzuela's model in Eq. (6.2.10), the overall impulse response of Spencer's model is given by

$$h(\theta, \tau) = \sum_{l=0}^{\infty} \sum_{k=1}^{\infty} \beta_{kl} \delta(\tau - T_l - \tau_{kl}) \delta(\theta - \Theta_l - \omega_{kl}) e^{j\phi_{kl}} \quad (6.2.13a)$$

in which Θ_l and ω_{kl} are the angle of the cluster and the ray within the cluster, respectively. The distribution of the arrival angle of a cluster Θ_l is assumed to be uniform between 0 and 2π , and the distribution of the angle of arrival of the rays within the clusters is assumed to have a Laplacian distribution given by

$$p(\omega_{kl}) = \frac{1}{\sqrt{2\pi}} e^{-|\sqrt{2}\omega_{kl}/\sigma|} \quad (6.2.13b)$$

where σ is the variance of the arrivals. A typical value for the variance of the arrival angles within a cluster is around 22° [Spe00]. Figure 6.21 shows a typical Laplacian distribution function with its two-sided exponential appearance. To simulate the AOA of the rays within a cluster, the AOA of the first path is selected from the uniformly distributed random variable, and the following paths deviate from this value according to the Laplacian distribution function.

Example 6.3: IEEE 802.11 MIMO Model A modified version of Spencer's model is recommended by IEEE 802.11n for MIMO channels. As shown in Fig. 6.22, this

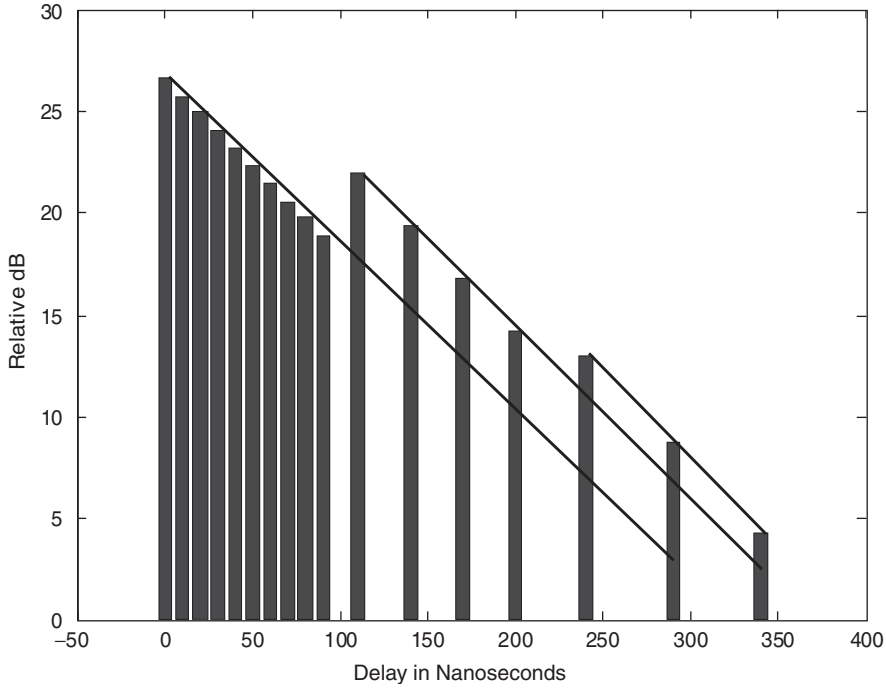


FIGURE 6.22 Implementation of Spencer's model in the IEEE 802.11 MIMO model [IEE04].

model assumes a fixed number of clusters and fixed delays between the rays. The AOA of the model, however, follows the Laplacian distribution. The model specifies five environments, with rms delay spreads of 15, 30, 50, 100, and 150 ns, respectively. The first two are identified with two clusters and the next three specify three, four, and six clusters. The Doppler spectrum recommended by the IEEE 802.11n group is the bell-shaped spectrum defined in Section 4.5.3.

6.3 WIDEBAND FREQUENCY-DOMAIN CHANNEL MODELING

In frequency-domain modeling, channel frequency response measurements are used to develop a statistical model for computer simulation of the channel. In this section we describe a particular approach to frequency-domain modeling based on autoregressive modeling. With this approach, a statistical autoregressive model is developed from measurement data and is used for computer simulation of the channel frequency response. Here we describe this method in the context of modeling an indoor radio channel, and the discussion follows closely the experimental work of Howard and Pahlavan [Pah90b, How92].

Autoregressive modeling of time-domain signals is a standard technique in the field of digital signal processing [Mar87b]. Here we apply the technique to samples of the frequency response of a channel. In mathematical terms, an autoregressive (AR) model maps a large set of data points representing a sample of a stochastic process onto a limited number of filter poles representing an *AR process*. To develop the model, we first

determine the locations of the poles of the AR process for each measured frequency response. Then we determine the statistics of the poles over the measurement set. With this approach, the entire database obtained from a measurement experiment is mapped onto a few parameters representing statistics of the locations of the poles of the frequency-domain AR process. To produce a sample profile in a simulation, the poles are regenerated from the statistics, and the AR model defined by these poles is driven with white noise to produce a sample frequency response. As we show in the following discussion, the pole locations are related to the amplitude and arrival time of a cluster of paths in the time domain.

6.3.1 Autoregressive Modeling

With the AR model, the frequency response at each location is a realization of an AR process of order p given by the equation

$$H(f_n; t) - \sum_{i=1}^p a_i H(f_n; t-i) = V(f_n) \quad (6.3.1)$$

where $H(f_n; t)$ is the n th sample of the complex frequency-domain measurement at a given location and $V(f_n)$ is a complex white noise process representing the error between the actual frequency response value at frequency f_n and its estimate based on the last p samples of the frequency response. The parameters of the AR model are the complex constants $\{a_i\}$. Taking the z -transform of Eq. (6.3.1), we can view the AR process $H(f_n; t)$ as the output of a linear filter with transfer function

$$G(z) = \frac{1}{1 - \sum_{i=1}^p a_i z^{-i}} = \prod_{i=1}^p \frac{1}{1 - p_i z^{-1}} \quad (6.3.2)$$

driven by a zero-mean white noise process $V(f_n)$. Given the mathematical form of $G(z)$, the AR model is often referred to as an *all-pole model*. Using the AR or all-pole model, the channel frequency response represented by the N measurement samples is described by the p parameters of the AR model or the locations of the p poles of $G(z)$, where typically $N \gg p$.

The AR parameters $\{a_i\}$ are solutions of the Yule–Walker equations [Mar87b]:

$$R(-l) - \sum_{i=1}^p a_i R(i-l) = 0, \quad 1 < l < p \quad (6.3.3)$$

in which $R(k) = R_{Hh}(k, 0)$ is the frequency correlation function, defined as

$$R(k) = \frac{1}{N} \sum_{i=1}^{N-k} H^*(f_i, t) H(f_{i-k}, t), \quad k \geq 0 \quad (6.3.4)$$

The variance of the zero-mean white noise process $V(f_n)$ is the same as the minimum mean-squared error of the predictor output, which is given by

$$\sigma_v^2 = R(0) - \sum_{i=1}^p a_i R(i) \quad (6.3.5)$$

Mapping a Time-Domain Profile to an AR Model. In time-domain modeling, the channel is characterized by the power-delay profile, such as those shown in Fig. 6.5. To find the appropriate poles related to a power-delay profile, we need to extract the correlation in frequency $R(k)$ from the time-delay profile $|h(\tau_i; t)|^2$. Since the frequency correlation function is the Fourier transform of the magnitude squared of the power-delay profile, these two functions are related by

$$R(k) = \sum_{i=1}^L |h(\tau_i, t)|^2 e^{-j \frac{2\pi}{N} k \tau_i}$$

where N is the total number of points used in the frequency-domain measurement. $R(k)$ values obtained in this manner can be used in Eq. (6.3.3) to determine the AR parameters. Then using Eq. (6.3.2), the locations of the poles are identified. In [Ali94] this method is used to map the JTC model to an equivalent frequency-domain model identified with the locations of the poles of an AR process.

Order of the AR Process. To examine the accuracy of the model, results of frequency-domain measurements introduced in Chapter 5 are examined. In general, the order of the AR process will be different for measurements made at different locations. Therefore, the first step in setting up a model is to determine the minimum order of the process that accommodates all the measurements. Using standard AR *order-selection criteria* such as the Akaike information criteria [Mar87b], the fifth-order process has been shown to provide an upper bound on the order of the frequency-domain measurements of an indoor radio channel [How91]. This bound is very conservative, and to determine a more realistic order, we examine the results of measurements directly. Figure 6.23 shows the five pole locations of the fifth-order AR process for the four frequency-domain measurement experiments described in Chapter 5. For a fifth-order model, typically, the magnitude of the largest pole is 0.97 or greater, the magnitude of the second pole is between 0.57 and 0.98, and the magnitudes of the remaining three poles are around 0.5, making them insignificant.

In conventional parametric spectral estimation, a pole close to the unit circle represents significant power at the frequency related to the angle of the pole. The frequency is calculated as $f = |p_i|/2\pi T_s$ and the spectrum is periodic with period $1/T_s$, where T_s is the sampling time. In our application, the model is applied in the frequency domain. The time-domain response obtained from the inverse Fourier transform is a periodic function, and a pole close to the unit circle represents significant power at the path delay related to the angle of the pole. The time-domain response is a periodic function with period $1/f_s$, and the corresponding delay is calculated as $\tau = -|p_i|/2\pi f_s$, where f_s is the sampling rate in the frequency domain: that is, the distance between two consecutive frequencies at which the frequency response was measured. For the measurements and modeling discussed in this chapter, $f_s = 0.25$ MHz, which covers a span of $1/(0.25 \text{ MHz}) = 4000$ ns for the time-domain measurements. In all measurements the significant paths were found to lie in the interval of delays from 150 to 550 ns, which corresponds to the angular range $-3\pi/40$ to $-11\pi/40$ in the z -domain. As shown in the scatter plots in Fig. 6.23, at most two significant poles exist in this range. Based on these observations we conclude that a two-pole model is adequate to represent the indoor radio channel for these experiments.

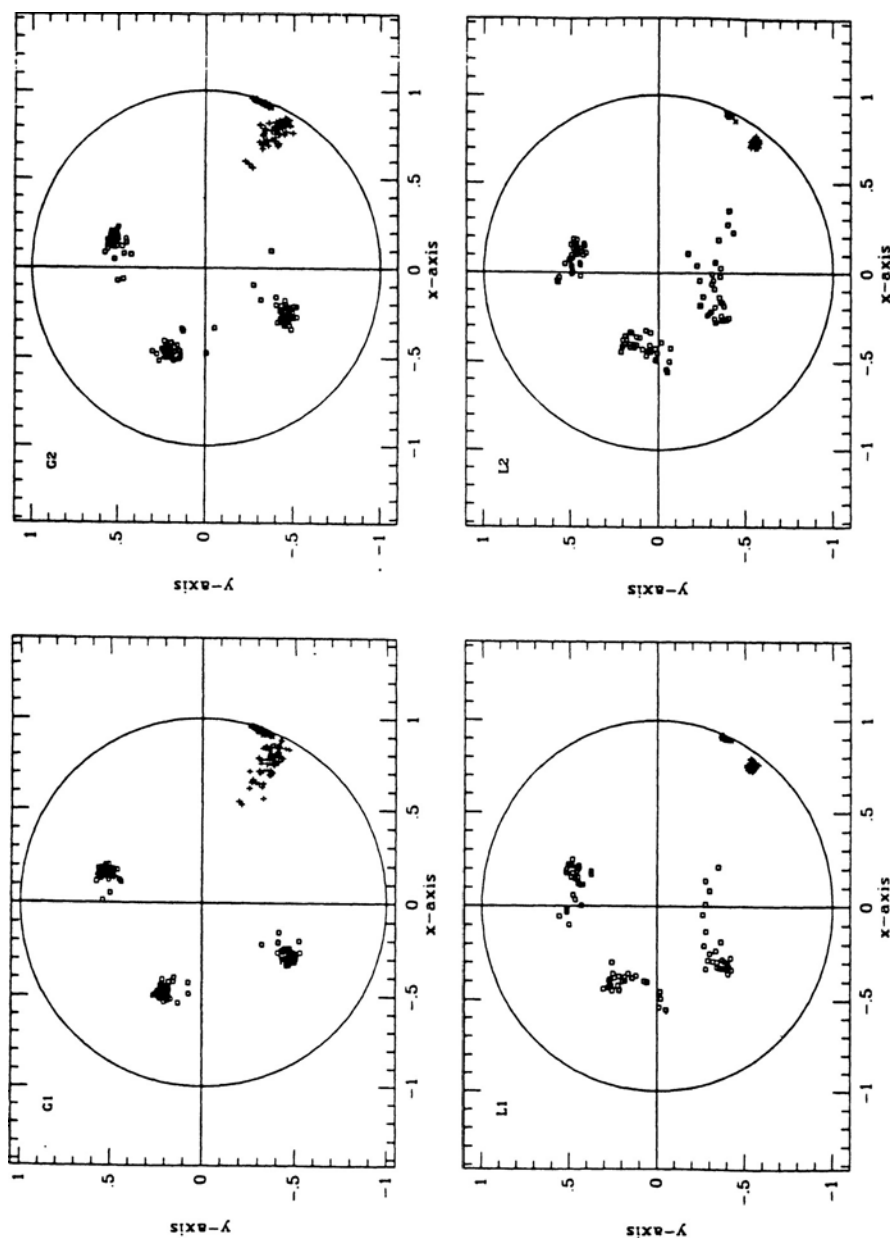


FIGURE 6.23 Scatter diagrams of the locations of the poles for the AR process representing the frequency-domain characteristics of several indoor radio channels.

Accuracy of the Second-Order Process. To demonstrate that a two-pole model is adequate, the first two poles of the fifth-order model are used to regenerate the frequency responses of the channel. Figures 6.24 and 6.25 represent a sample frequency response and a sample time response, respectively, obtained from a second-order AR model. The time response also represents the inverse transform of the two-pole transfer function of the AR process. The presence of two “humps” corresponds to two major poles, with each hump representing a cluster of arriving paths. Thus, two major poles can be interpreted as two significant clusters of multipath arrivals. The interpretation of a pole as defining a cluster of paths, and the distance from the unit circle as defining the power in the cluster, provides a useful physical interpretation for this AR model. The observation of more than one cluster in time-domain measurements of indoor radio channels was first reported in [Sal87b]. However, as we discussed earlier, it is difficult to distinguish adjacent clusters in direct examination of the data, because the clusters may overlap or the trailing cluster may be too small to detect [Sal87a]. Careful examination of the time-domain measurements in [Sal87] had shown that the existence of two clusters can be observed in only a fraction of the measured time-domain channel responses. However, by using frequency-domain AR modeling, two clusters are identified for all the measurements.

Figure 6.26 shows the CDF of the 3-dB width of the frequency correlation function for both the original measurements (solid line) and the measurements regenerated from the two-pole model (short-dashed line) of an experiment done in a small office. Figure 6.27 shows the CDF of the rms delay spread for both the original measurements (solid line) and the regenerated measurements (short-dashed line). Because the parameters of the model are determined from the frequency-domain autocorrelation function, the frequency-domain statistics agree more closely with the original measurements

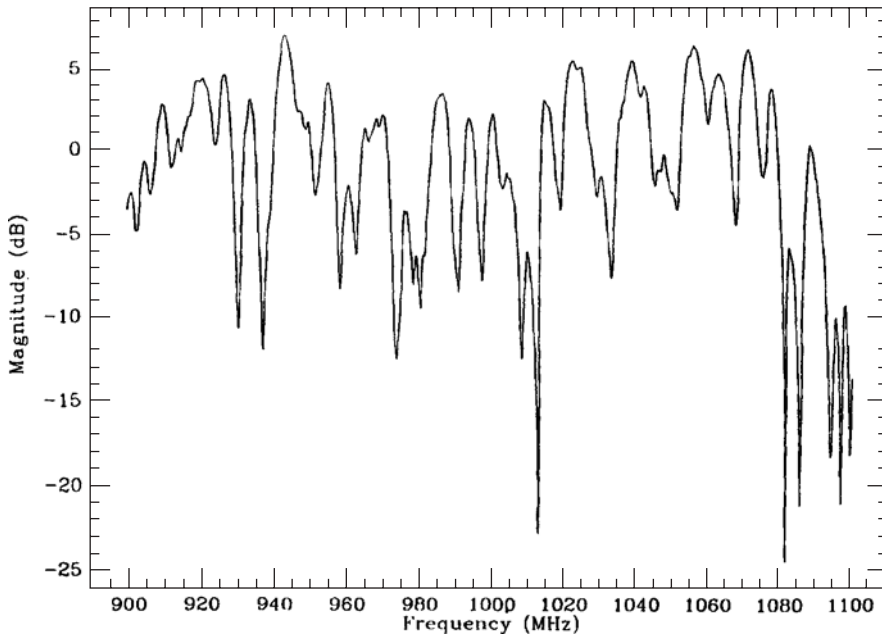


FIGURE 6.24 Sample frequency response regenerated from AR modeling.

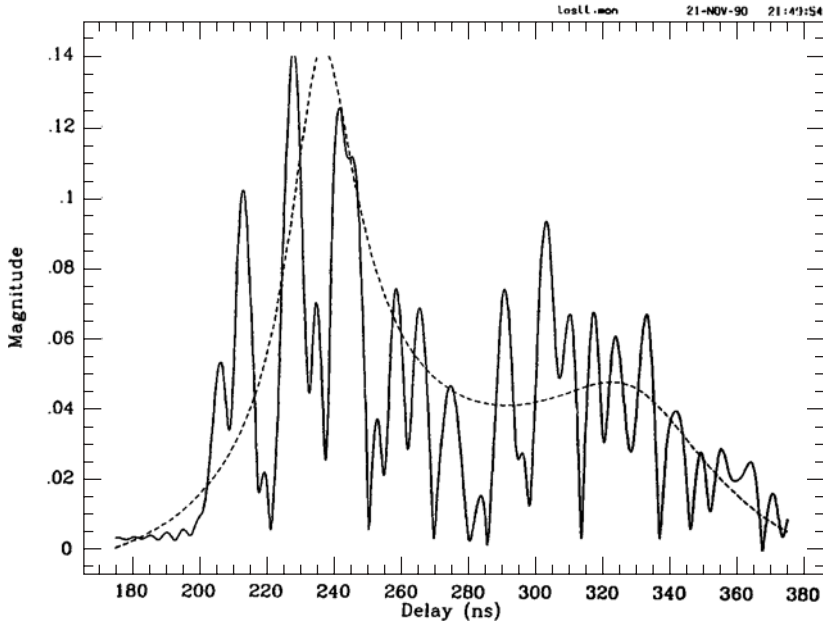


FIGURE 6.25 Sample time response obtained by taking the inverse Fourier transform of the frequency response shown in Fig. 6.16. The dashed line represents the inverse transform of the transfer function of the AR process. The two peaks correspond to the two poles.

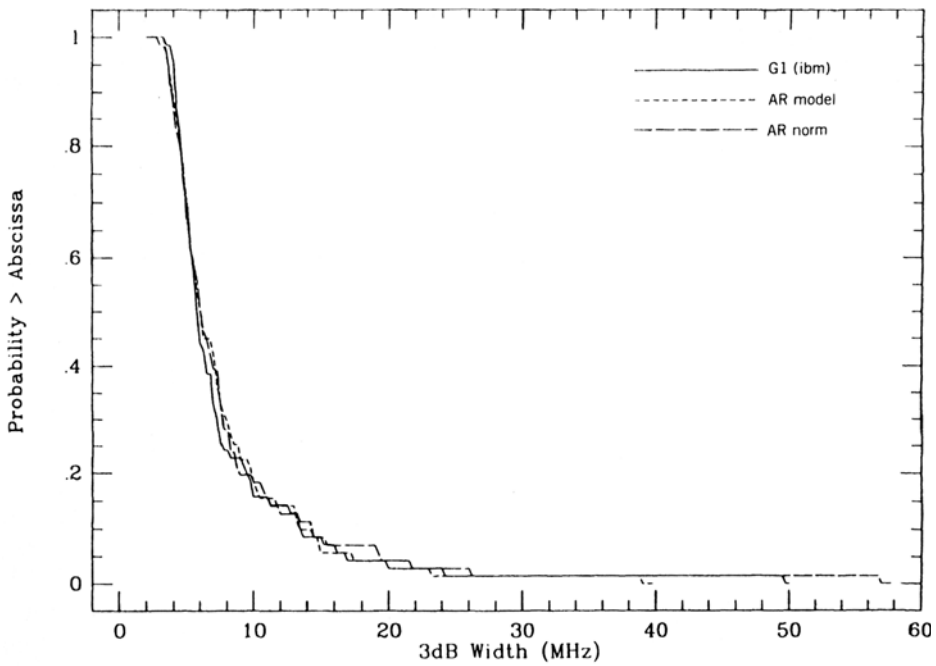


FIGURE 6.26 CDFs of the 3-dB width of the frequency correlation function for the results of AR modeling, normalized AR modeling, and the empirical data collected in area G1.

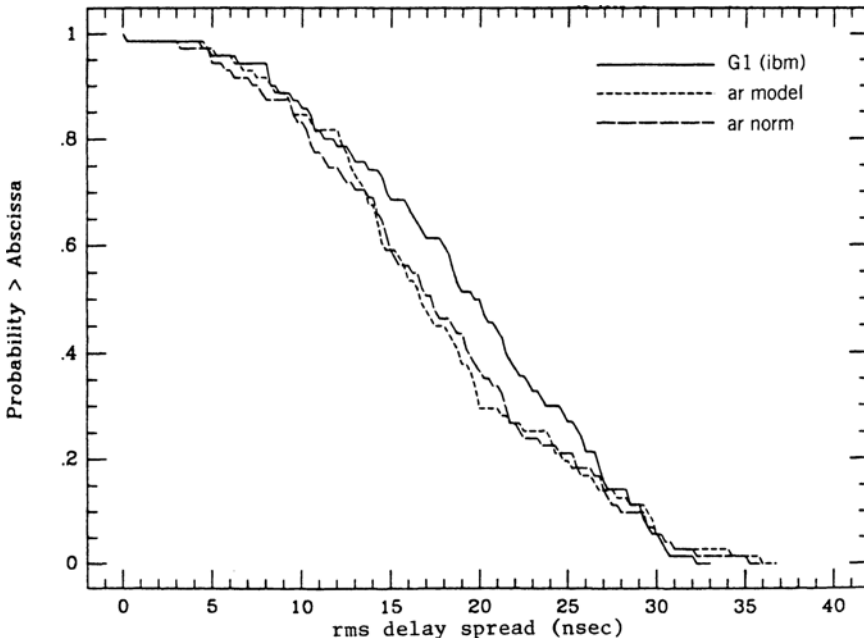


FIGURE 6.27 CDFs of the rms multipath delay spread for the results of AR modeling, normalized AR modeling, and the empirical data collected in area G1.

than do the time-domain statistics. This analysis is extended to the remainder of the measured data, and the results support the foregoing conclusions.

The third CDF shown in Figs. 6.26 and 6.27 (long-dashed lines) corresponds to the normalized AR model. In the normalized AR model, the angle of the first pole for all the measurements is fixed at -18° , which is equivalent to normalizing the arrival of the first cluster at 200 ns. The angle of the second pole is adjusted to maintain the original angular separation in the AR model while the magnitudes of the poles are left unchanged. This normalization does not affect the rms delay spread or the 3-dB width of the frequency correlation function, and it reduces the complexity of the model. After normalization, the channel that was represented by two complex parameters (magnitude and angle for both poles 1 and 2) will be represented with one real and one complex parameter (magnitudes of poles 1 and 2 and the angular difference).

Table 6.5 summarizes the statistics of the locations of the first two poles for each of the five experiments described in Chapter 5. Experiments G1 and G2 were spatial measurements made in two different offices, L1 and L2 were two local experiments in which the distance between the transmitter and receiver was relatively fixed, and M1 was a mixed experiment in which nine measurements were taken around each location and the locations were chosen in several rooms of an office area. As shown in Table 6.5, for all the experiments, the magnitude of pole 1 has a larger mean value and smaller variations than the magnitude of pole 2, indicating that the first cluster is always present, whereas the second cluster is not always significant. The poles for local experiments have smaller variations than the poles for spatial experiments, indicating smaller changes in the channel multipath characteristics in the local environment. To relate the rms delay spread to the pole locations for our measurements, we calculate

TABLE 6.5 Statistics of the Location of the First Two Poles for Each of Five Frequency-Domain Measurement Experiments

Experiment	Parameter	Mean	Standard Deviation	Minimum	Maximum
G1	$ p_1 $	0.9871	0.006832	0.9648	0.9982
	$\text{ANG}(p_1)$	-18.82	1.989	-22.75	-14.73
	$ p_2 $	0.8380	0.08995	0.5701	0.9780
	$\text{ANG}(p_2) - \text{AN}$	-6.355	1.355	-9.376	-2.459
G2	$ p_1 $	0.9878	0.005093	0.9750	0.9986
	$\text{ANG}(p_1)$	-19.04	1.839	-22.44	-14.94
	$ p_2 $	0.8624	0.07776	0.6224	0.9599
	$\text{ANG}(p_2) - \text{AN}$	-8.239	1.477	-9.213	-3.549
L1	$ p_1 $	0.9850	0.004760	0.9758	0.9942
	$\text{ANG}(p_1)$	-22.80	0.9938	-25.36	-21.42
	$ p_2 $	0.9332	0.01783	0.8979	0.9599
	$\text{ANG}(p_2) - \text{AN}$	-12.95	0.7876	-14.70	-11.48
L2	$ p_1 $	0.9797	0.004941	0.9632	0.9871
	$\text{ANG}(p_1)$	-24.59	0.7449	-27.34	-23.57
	$ p_2 $	0.9268	0.01551	0.8880	0.9533
	$\text{ANG}(p_2) - \text{AN}$	-12.52	0.7932	-13.84	-10.96
M1	$ p_1 $	0.9860	0.005735	0.9660	0.9964
	$\text{ANG}(p_1)$	-18.50	1.41	-22.9	-15.17
	$ p_2 $	0.8614	0.07405	0.5997	0.9739
	$\text{ANG}(p_2) - \text{AN}$	-8.483	1.667	-12.98	-0.176

the correlation between these parameters. The correlation of the magnitude of pole 1 with the rms delay spread reveals high negative values; thus, as pole 1 approaches the unit circle, the rms delay spread decreases. In contrast, the magnitude or angle of pole 2 shows weak correlation with the rms delay spread. Given two clusters of arrivals, the minimum rms delay spread occurs when the paths of cluster 1 are much stronger than the paths of cluster 2, while the maximum rms delay spread occurs when the amplitudes of the paths are more nearly equal. Because the second cluster represents energy arriving at a later time (from a greater distance), the magnitude of pole 1 is almost always larger than the magnitude of pole 2. Therefore, in the two-pole model, if the magnitude of pole 1 decreases toward the magnitude of pole 2, this increases the rms delay spread. Whereas the magnitude of pole 2 may in some cases increase to create the same effect, there are also cases where pole 2 is insignificant relative to pole 1, and its movement is uncorrelated with the rms delay spread value. These observations will be exploited in the statistical modeling that we describe in the next section. Given the inverse relationship between rms delay spread and the 3-dB width of the frequency correlation function, similar results regarding the 3-dB widths are observable.

6.3.2 Statistical Modeling in the Frequency Domain

Assuming that the frequency response of the channel forms a second-order AR process, modeling of the channel can be reduced to statistical characterization of the two poles of the model and the variance of the white noise process driving the two-pole filter.

Five real values are required to represent the two complex poles and the variance of the driving noise. In this section we introduce four models, and in each model it is assumed that the four values used to represent the two complex poles are statistically independent Gaussian random variables defined in a prescribed range. The mean, variance, and range of each value are determined from the empirical data for a given experiment. The angle of the first pole, representing the arrival time of the first cluster, is normalized to the arbitrary value of -18° , which corresponds to the first cluster being centered at 200 ns. Note that in time-domain modeling, the modeled arrival delays are the delays relative to arrival of the first path. In the four models described here, we reference the delays to the arrival of the first cluster.

The four second-order models are defined as follows:

1. Model 1 assumes that the magnitude of the first pole and the angle and magnitude of the second pole are all Gaussian random variables with means and variances determined from the values measured in an experiment.
2. Model 2 assumes that the angle of pole 2 is fixed at its average as determined from an experiment, and the magnitudes of the two poles are treated as random variables.
3. Model 3 assumes that the second pole is fixed at the average angle and magnitude of the second pole determined from an experiment, and only the magnitude of the first pole is considered as a random variable.
4. Model 4 assumes that both poles are fixed at the average magnitudes and angles of the poles determined from an experiment.

These models reflect the relative importance of the locations of poles 1 and 2 in their influence on the rms delay spread and 3-dB width of the frequency correlation function. Model 4 with two fixed poles is the simplest of the four models, whereas model 1, with random magnitude of the two poles and angle of the second pole, is the most complex. Table 6.5, discussed earlier, shows the mean, standard deviation, and range of all parameters obtained from the five experiments.

For computer simulation of the channel, any one of the four models can be used, with the means and standard deviations provided in Table 6.5, to generate the two poles (the poles lying outside the measurement range are discarded). The poles are then used to form the AR process. The variance of the noise driving the process is determined from

$$\sigma_v^2 = \frac{R(0)}{R_g(0)} \quad (6.3.6)$$

where the received power, $R(0)$, is determined from a lognormally distributed random variable with mean and standard deviation given in Table 6.5. The $R_g(n)$ is the inverse z -transform of $G(z)G^*(z^{-1})$ with the $G(z)$ defined in Eq. (6.3.2) and is the autocorrelation of $g(n) = Z^{-1}[G(a)]$. $R_g(0)$ is related to pole 1(p_1) and pole 2(p_2) by

$$R_g(0) = \frac{1 - |P_1|^2|p_2|^2}{(1 - |p_1|^2)(1 - |p_2|^2)|1 - p_1p_2^*|^2} \quad (6.3.7)$$

The frequency-response measurement is generated by using Eq. (6.3.1) with $p = 2$, $a_1 = p_1 + p_2$, and $a_2 = -p_1p_2$.

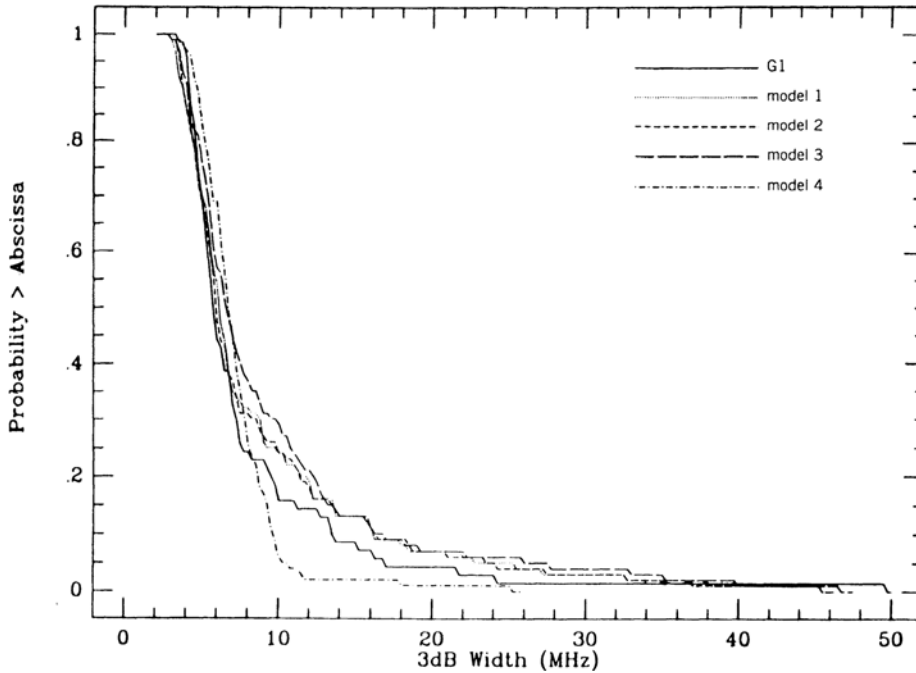


FIGURE 6.28 CDFs of the 3-dB width of the frequency correlation function for the four statistical AR models and the empirical data collected in area G1.

Figure 6.28 shows a graph of the CDF of the 3-dB width of the frequency correlation function obtained from 100 simulated measurements for each of the four models of experiment G1 and the actual CDF determined from the 70 measurements for G1. Figure 6.29 shows corresponding results for the rms delay spreads. In the two figures it is seen that the first three models provide very similar results in either the time or frequency domain. In these models, the magnitude of the first pole changes for each measurement regenerated. It can be seen that the simpler model 4, with its fixed poles, cannot accommodate the entire range of variations of 3-dB width or rms delay spread. In [How92] this analysis is extended to the other four models described in Table 6.5, and it is concluded that for global and mixed measurements at least, the magnitude of the first pole should be considered random, whereas local measurements can be regenerated using a model having fixed poles.

6.4 COMPARISON BETWEEN STATISTICAL MODELS

Here we draw comparisons between the time- and frequency-domain approaches to statistical channel modeling. The time-domain approach is by far the most popular statistical approach used for modeling indoor and urban radio channels. In time-domain modeling there is a straightforward relationship between multipath propagation and the results of measurements, and therefore a physical interpretation of the model is quite evident. The amplitudes and arrival delays of the paths are determined directly from the results of measurements. However, systematic identification of the path clusters from

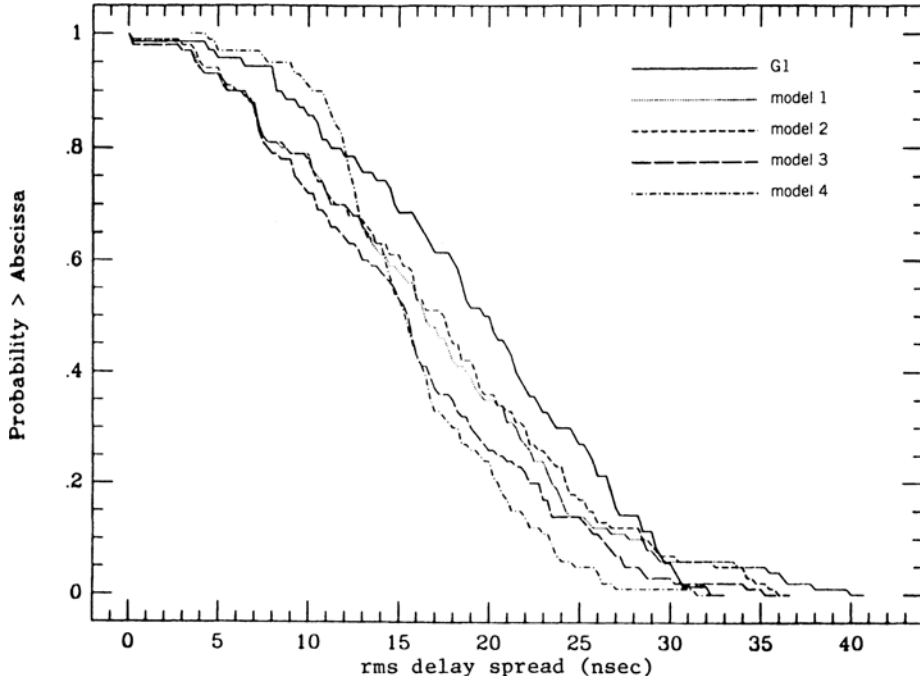


FIGURE 6.29 CDFs of the rms multipath delay spread for the four statistical AR models and the empirical data collected in area G1.

measurement data is typically not a simple task. With frequency-domain modeling, physical interpretation of a model is not so straightforward, because the path arrivals are not directly related to the results of frequency-domain measurements. The second-order frequency-domain model associates the arriving paths with two clusters, and the arrival delays of the clusters are evaluated indirectly by determining the angular positions of the poles of the AR process.

Computation of the parameters of the frequency-domain AR model is straightforward and generally accurate. The frequency-domain AR model is defined by the statistics of the pole locations. First, the pole locations are calculated by solving a set of linear equations involving the complex-frequency autocorrelation function. Then the location of each pole is modeled as a random variable. These calculations are systematic and routine, and efficient algorithms are available for doing them. In contrast, computation of parameters for the time-domain model is rather tedious and prone to inaccuracy. In time-domain modeling, path amplitudes and arrival times should be modeled statistically. Each time response has to be processed with a peak detection algorithm to determine the path locations and amplitudes. Peak detection involves calculation of a threshold to distinguish real peaks from those created spuriously by noise, and the resulting statistical model is sensitive to these details. The arrivals of the detected peaks are used for calculation of the parameters of the modified Poisson process. This calculation is iterative and requires careful attention to ensure accurate results. The amplitudes of the individual paths are then analyzed for best fit to one of the alternative statistical distributions. The mean and variance of the statistical model selected are then determined by best fit to another exponential curve.

As far as the results of simulations are concerned, in time-domain modeling the CDF of the rms delay spread of the modified Poisson–lognormal model have been found to fit measurement data more closely than the Poisson–Rayleigh model. In frequency-domain AR modeling, simulation results show close agreement with the CDF of the 3-dB width of the frequency correlation function of the measurement data, but the agreement is not as close for the CDF of the rms delay spread of the channel.

The time-domain channel impulse responses obtained from frequency-domain measurements typically represent higher numbers of arriving paths than those of responses obtained from direct time-domain measurements. Therefore, the results of time-domain modeling using frequency-domain measurement systems have to be evaluated carefully.

In a block-oriented computer simulation environment, a simulation of an AR model uses a filter block and a fast Fourier transform (FFT) block to generate the frequency- and time-domain responses. These two blocks are included in all block-oriented simulation packages. The time-domain simulation involves generation of a modified Poisson process and amplitude and phase distributions that are not typically provided in block-oriented simulation packages. Therefore, the implementation of the time-domain method in a block-oriented simulation environment is more complicated.

6.5 RAY-TRACING ALGORITHMS

In Chapter 3 we introduced the two-dimensional ray-tracing technique through a simple example of modeling radio propagation within a room using free-space transmission and multiple reflections from walls. We also showed how, given the paths connecting the transmitter and receiver, one can determine the response of the channel to narrow transmitted pulses. In addition to free-space transmission and simple reflection, more complex ray-tracing algorithms include mechanisms of diffraction, diffuse wall scattering, and transmission through various materials. In this section we provide a more detailed treatment of signal reflection and introduce three other mechanisms, while describing more complex examples of propagation in indoor and outdoor areas. Appropriate combinations of these mechanisms can be used for simulation in a particular coverage area. For example, in small indoor areas with soft-surfaced walls, reflection and transmission are the dominant mechanisms for radio propagation at frequencies around 1 GHz; whereas for macrocellular high-rise urban canyons with antennas installed above roof level, diffraction is the main mechanism for signal propagation.

6.5.1 Reflection and Transmission Mechanisms

Figure 6.30 shows reflected and transmitted paths in a simple propagation example. As a ray meets a wall, two paths emerge, one reflected and the other transmitted through the wall. To determine the reflected and transmitted paths in a ray-tracing algorithm, we assume walls, ceilings, floors, and other objects to be dark mirrors, and we use the theory of geometric optics to identify the paths. There are two general approaches that can be employed to calculate these paths. The first uses optical images of the transmitter and receiver [McK91, Rus91, Law92]. In this approach, reflections of the transmitted signal by various reflecting objects in the floor plan are described by images of the transmitter, and these images are used with images of the receiver to find all the paths to the receiver.

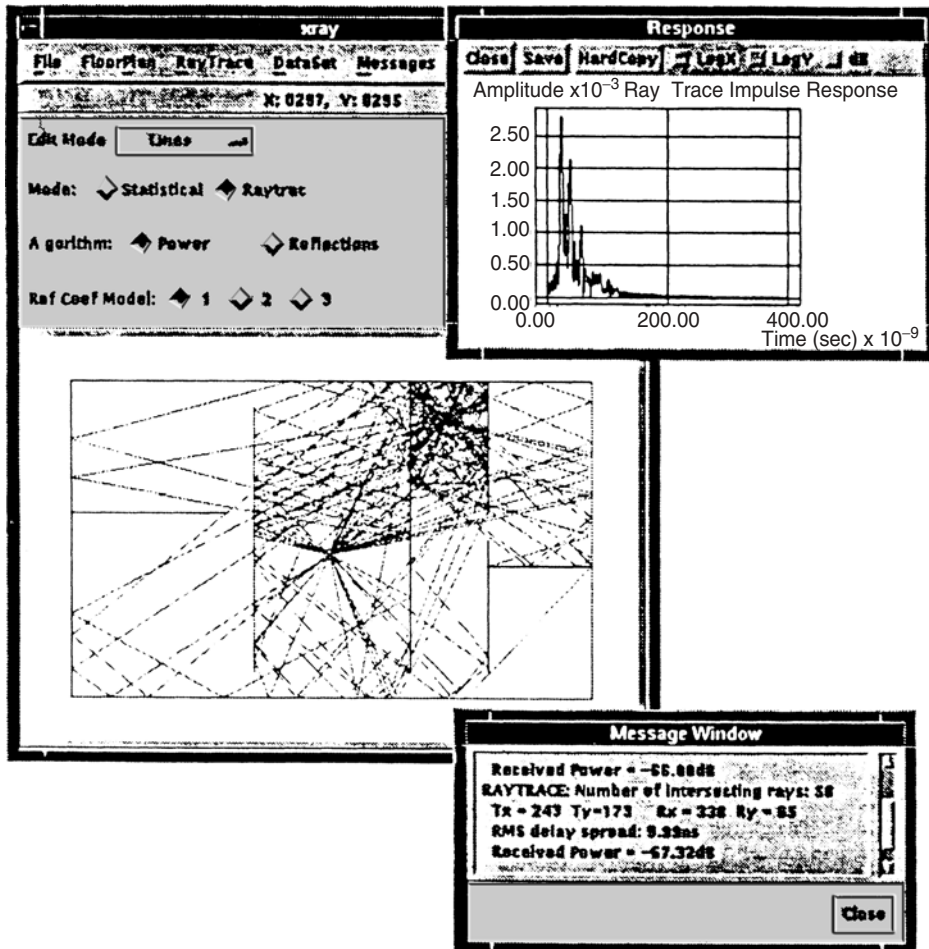


FIGURE 6.30 Sample graphical interface for the two-dimensional ray-tracing algorithm. The windows identify the trace of the paths, channel impulse response, channel parameters, and simulation parameters. (From [Hol92c] © IEEE.)

The path distance and direction are then used to determine the magnitude, phase, and delay of each path arriving at the receiver. The two drawbacks in this method are as follows: (1) It is difficult to determine all image points if we must consider many reflections of the transmitted signal, and (2) it is difficult to automate the imaging procedure with a computer program. Thus, to implement this approach efficiently, special procedures are employed to cancel the unnecessary images. This approach is best suited to environments, such as large hilly urban areas, where only a few dominant reflectors need be considered.

The second method for determining the reflected and transmitted paths is through the application of *ray-shooting techniques* [Des72, Ike91]. The ray-shooting algorithm is an intuitively simple approach to the problem of multipath propagation. A pincushion of rays is sent out from the transmitter, and the progress of each ray is traced through the environment until the ray has either intersected the receiver or has lost

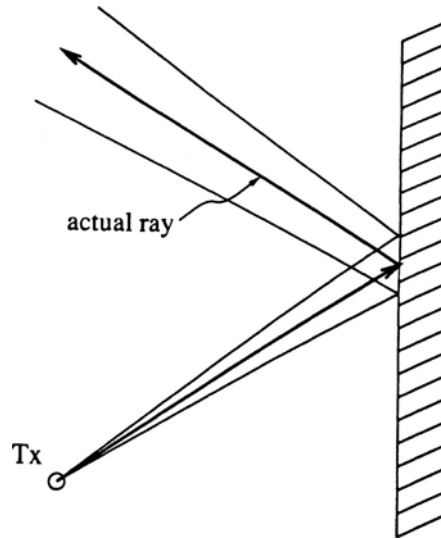


FIGURE 6.31 Cone structure used for the ray-shooting method.

enough power that its contribution to the received signal is negligible. The time of arrival, intensity, phase, and direction of arrival are recorded for each ray that intersects the receiver. Once every ray has been traced to completion, the channel impulse response is formed. In the implementation of this algorithm, as shown in Fig. 6.31, the rays are actually treated as cones. This implementation has a few potential problems: (1) When a ray strikes near an edge of a wall it is difficult to trace the reflected and transmitted rays; and (2) this method of ray tracing does not ensure that every signal path between the transmitter and receiver is considered. By taking special care in designing the algorithm, the harmful effects of the first problem can be avoided. The second problem is found to be minor in most applications. The ray-shooting approach is found to be suitable for areas having many irregular reflectors, such as small indoor areas.

Calculation of Reflection Coefficients. When a traced ray meets a wall or other surface, some of the energy is reflected, some is transmitted through the surface, and some is absorbed. There are three rays that can be defined at the wall: the incident ray, the reflected ray, and the transmitted ray. Figures 6.32 and 6.33 represent all the electrical field components of the incident ray, reflected ray, and the transmitted or refracted rays. Figure 6.32 represents the horizontal polarization components, and Fig. 6.33 represents the vertical polarization field components. The angle of the transmitted ray is assumed to be the same as that of the incident ray, and the angle of the reflected ray is determined by the orientation of the reflecting surface. The amounts of energy reflected and transmitted are based on the structural material as well as the angle of incidence. The phase of the ray is also affected by a reflection or a transmission. We can assume a 180° phase reversal of the reflected ray and no phase change in the refracted ray. This assumption is valid as long as we consider high-frequency electromagnetic radiation. Over a smooth surface the reflection coefficients for the horizontal and vertical

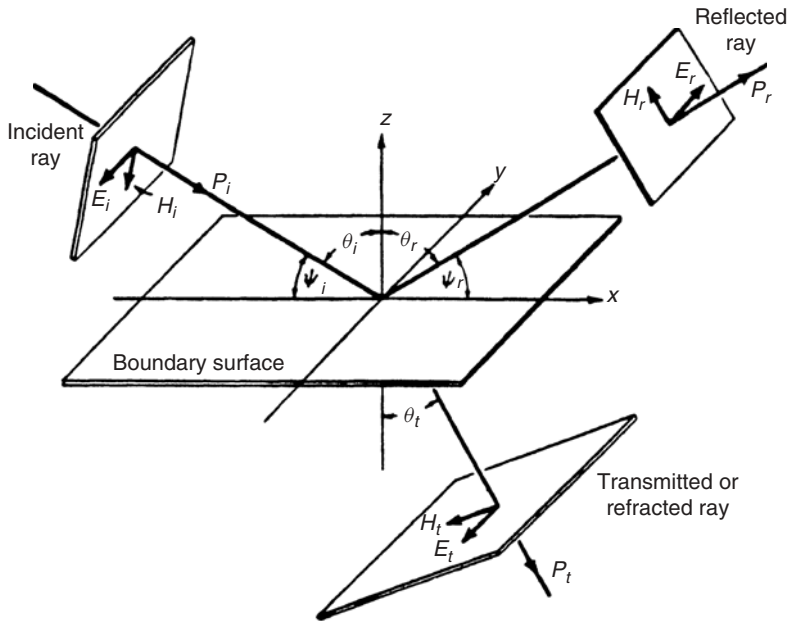


FIGURE 6.32 Various field elements for the transmission and reflection mechanism with horizontal polarization.

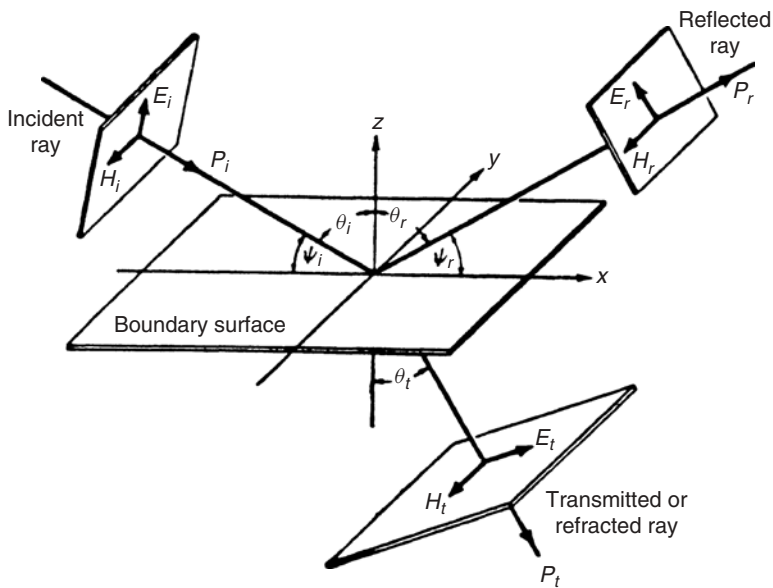


FIGURE 6.33 Various field elements for the transmission and reflection mechanism with vertical polarization.

components are given by [And93]

$$R_{s,h} = \frac{E_{r,h}}{E_{i,h}} = \frac{\sin \psi - \sqrt{\varepsilon - \cos^2 \psi}}{\sin \psi + \sqrt{\varepsilon - \cos^2 \psi}} \quad (6.5.1)$$

and

$$R_{s,c} = \frac{E_{r,c}}{E_{i,v}} = \frac{\varepsilon \sin \psi - \sqrt{\varepsilon - \cos^2 \psi}}{\varepsilon \sin \psi + \sqrt{\varepsilon - \cos^2 \psi}} \quad (6.5.2)$$

where ψ is the angle of incidence and ε is the complex permittivity given by

$$\varepsilon = \varepsilon_r - j60\sigma\lambda \quad (6.5.3)$$

where ε_r is the normalized relative dielectric constant of the reflecting surface, σ is the conductivity of the reflecting surface, and λ is the wavelength of the incident ray. Examples of the magnitude and phase of the reflection coefficients for different angles of incident are given in Figs. 6.34 and 6.35.

If the surface is rough, an additional surface roughness factor ρ is included in the equations and the reflection factors are then given by

$$R_h = \rho R_{s,h} \quad (6.5.4)$$

and

$$R_v = \rho R_{s,v} \quad (6.5.5)$$

where the roughness factor is given by [Cha82] as

$$\rho = e^{-\delta}$$

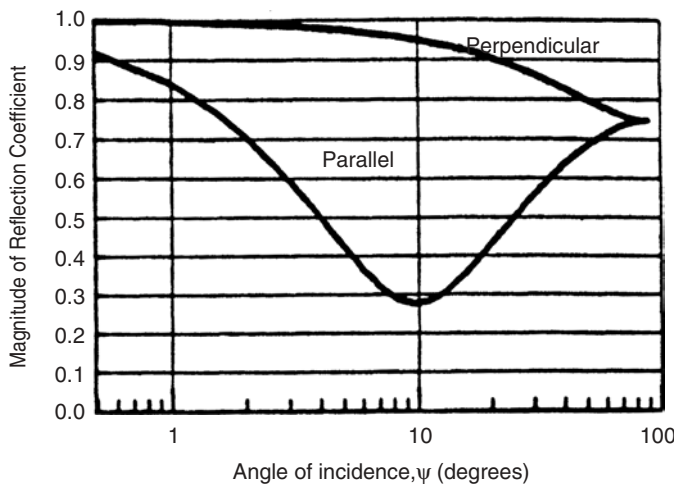


FIGURE 6.34 Magnitude of the reflection coefficient for $\sigma = 1.0$, $\varepsilon_r = 15$, and $f = 600$ MHz. (From [And93] © IEEE.)

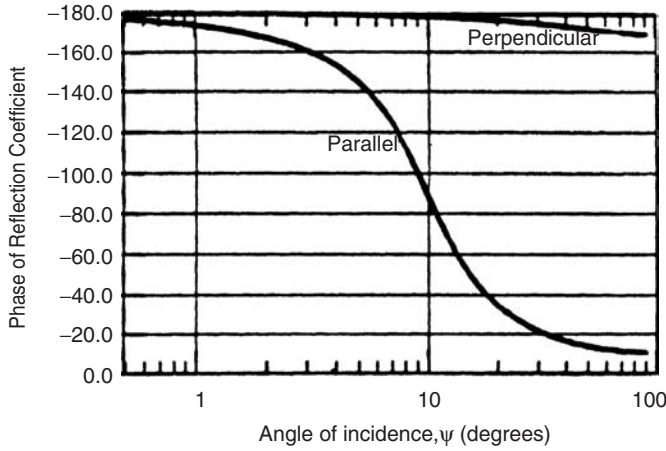


FIGURE 6.35 Phase of the reflection coefficient for $\sigma = 1.0$, $\epsilon_r = 15$, and $f = 600$ MHz. (From [And93] © IEEE.)

and

$$\delta = \frac{4\pi \Delta h}{\lambda} \sin \psi$$

where Δh is the standard deviation of the normal distribution assumed for the surface.

Calculation of the Wall-Transmission Coefficient. The *wall-transmission coefficient* is calculated from the reflection coefficient and the attenuation through the wall. A thick concrete wall has a very high transmission loss, almost totally blocking the transmission, while a glass wall may have the loss of only a fraction of a decibel. For outdoor applications where both transmitter and receiver are outside, but in the vicinity of one or more buildings, transmissions through building walls are not included in the ray-tracing models. Instead, it is assumed that most of the energy of the rays transmitted through the outside wall of a building is absorbed by other walls within that building and is thus rendered negligible. For outdoor-to-indoor communications, as we observed in the discussion of narrowband signal power loss, a path loss of 10 to 20 dB is usually included in the calculations. Propagation after the initial loss follows the same rules as those described for indoor radio propagation. For indoor communications, the transmission through walls constitutes an important component of the received signal, especially for obstructed-line-of-sight (OLOS) environments. In the remainder of this section we describe three simple models for transmission through walls on indoor radio channels, and then we introduce the slab model for transmission.

The three transmission models we describe here are simple and straightforward, although each is an approximation rather than an exact representation of an actual wall. In all three models, Eqs. (6.5.1) and (6.5.2) are used to determine the reflection coefficients of the structural material. The differences among the three models lie in the way that the transmission coefficients are determined.

For the first model it is assumed that there is no energy lost in the wall itself; that is, all of the energy is either transmitted or reflected. The transmission coefficient for

both horizontal and vertical components of this model is given by

$$T_{(h,v)} = \sqrt{1 + R_{(h,v)}^2} \quad (6.5.6)$$

This equation is derived from the fact that total power must be conserved in the system. Given that the reflection coefficient is determined by assuming a single planar boundary, this equation is not accurate. The second model assumes that there will always be some loss in the material and includes loss for the first and second layers of the wall. This is a simple *slab model* which includes the boundaries encountered at both sides of the wall. The horizontal and vertical transmission coefficients in this case are given by

$$T_{(h,v)} = \sqrt{(1 - R_{1,(h,v)}^2)(1 - R_{2,(h,v)}^2)} \quad (6.5.7)$$

The values of R_1 and R_2 are the reflection coefficients at the boundaries of the wall. R_1 is the transmission coefficient at the “front” boundary between the air and the wall, whereas R_2 is the transmission coefficient at the “back” boundary between the wall and the air. The two terms are multiplied together to obtain the total transmission through the wall. This model does not take into account the thickness of the material or the fact that some of the energy is reflected off the second boundary and back out of the material. A more elaborate model for the slab is discussed later. The third and last model assumes a 3-dB loss through the wall, as defined by

$$T_{(h,v)} = \sqrt{0.5(1 - R_{(h,v)}^2)} \quad (6.5.8)$$

Figure 6.36 shows plots of the reflection coefficient and transmission coefficients for all three models, considering only the horizontal component. The plot shows the reflection coefficient from Eq. (6.5.1) as well as the transmission coefficient calculated using each of the three models. The permittivity of the wall is assumed to be 20 F/m in this case.

Slab Model. Figure 6.37 shows a model for reflection and transmission in a slab. The wave transmitted through the first layer of the slab is reflected several times, and each reflection from either side of the slab causes another transmission to the air. Therefore, each incident ray gives rise to multiple reflected and transmitted rays, and the strengths of these rays are functions of the material and the thickness of the slab as well as the signal frequency and the angle of the incident ray. If we assume that all the reflected and transmitted rays are added to form a single ray, it is possible to model the slab with a single reflection coefficient and a single transmission coefficient. These overall transmission and reflection coefficients are given by [Bur83, Yan94b]

$$R_{(h,v)} = \frac{R_{s,(h,v)}(1 - P_d^2 P_a)}{1 - R_{s,(h,v)}^2 P_d^2 P_a} \quad (6.5.9)$$

and

$$T_{(h,v)} = \frac{(1 - R_{s,(h,v)})P_d P_t}{1 - R_{s,(h,v)}^2 P_d^2 P_a} \quad (6.5.10)$$

where $R_{s,(h,v)}$ denotes the reflection coefficients for horizontal and vertical polarization of smooth surfaces as given by Eqs. (6.5.1) and (6.5.2). The other three parameters are

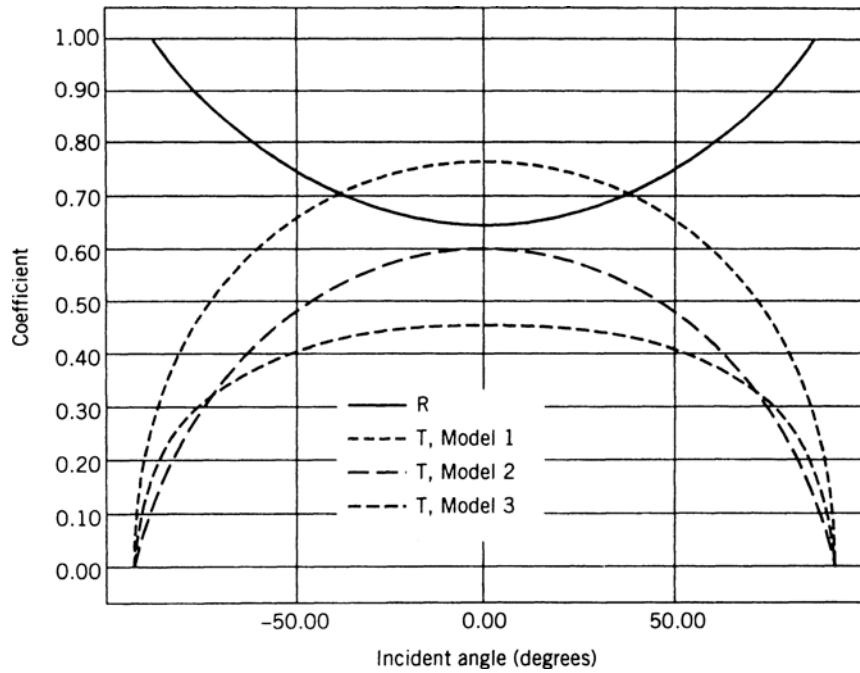


FIGURE 6.36 Reflection and transmission coefficients for three simple transmission models.

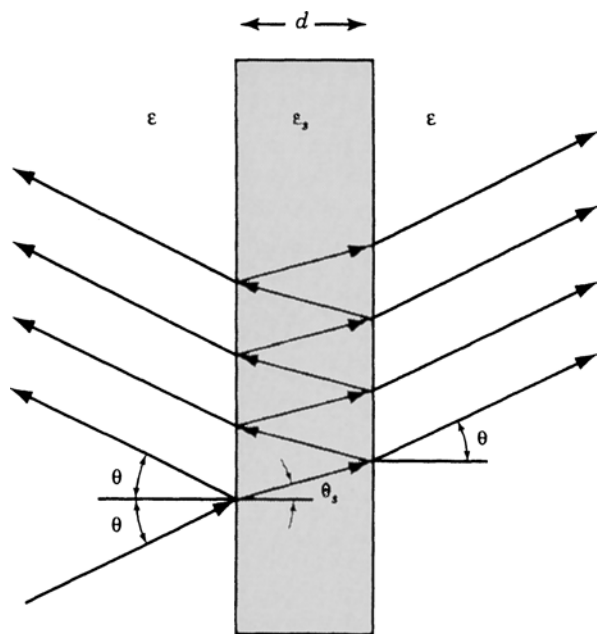


FIGURE 6.37 Reflection and transmission in a slab.

given by

$$P_d = e^{-jk'l}, \quad P_a = e^{j2kl \sin^2 \theta}, \quad P_t = e^{-jkl}$$

and $l = d / \cos \theta$, where d is the thickness of the slab and where $k = 2\pi/\lambda$ and $k' = k\sqrt{\epsilon}$ are the propagation constants in free space and the slab, respectively.

6.5.2 Diffraction Mechanism

Figure 6.38 shows a typical diffraction scenario. The ray striking the corner at angle ϕ is diffracted in all directions, and one of the diffracted rays arrives at the receiver from angle ϕ' . The diffracted ray is attenuated by the *diffraction coefficient*. In outdoor radio applications, diffraction occurs at roof edges and the corners of walls. In indoor radio propagation it usually occurs at the corners of rooms, at bent surfaces of walls, in corridors, and at the edges of windows. The energy contributions from diffracted paths can be strong if a first- or second-order diffracted path exists, while other paths arriving with multiple reflections or transmissions are heavily attenuated. For example, if the transmitter and receiver are located at opposite sides of a building, it usually takes many reflections to establish a path connecting the transmitter and receiver, and transmission through the building involves many intervening walls. However, a second-order diffraction at two sides of the roof can provide a diffracted path of only second order.

The complete solution to the diffraction problem for perfectly conducting and finite-conducting wedges using the uniform theory of diffraction (UTD) is treated in [Lue84a, Lue84b]. The diffraction coefficients given in [Lue84b] are as follows:

$$\begin{aligned} D_{h,v} = & \frac{-e^{-j(\pi/4)}}{2n\sqrt{2\pi k}} \left[\cot \frac{\pi + (\phi - \phi')}{2n} F(kLa^+(\phi - \phi')) \right. \\ & + \cot \frac{\pi - (\phi + \phi')}{2n} F(kLa^-(\phi - \phi')) \\ & + R_{o(h,v)} \cot \frac{\pi - (\phi + \phi')}{2n} F(kLa^-(\phi + \phi')) \\ & \left. + R_{n(h,v)} \cot \frac{\pi + (\phi - \phi')}{2n} F(kLa^+(\phi - \phi')) \right] \end{aligned} \quad (6.5.11)$$

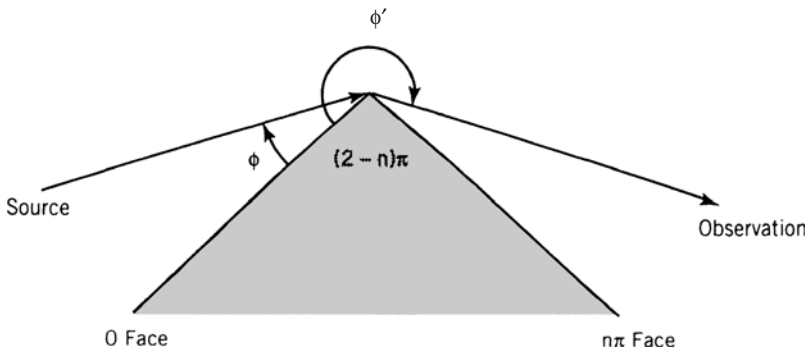


FIGURE 6.38 Two-dimensional wedge diffraction geometry.

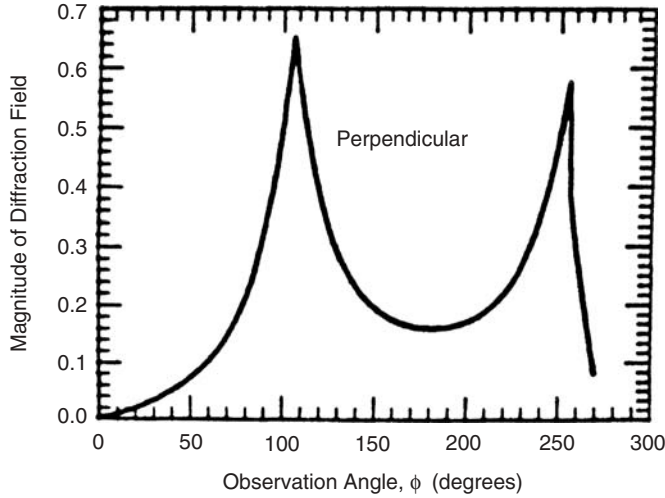


FIGURE 6.39 Magnitude of diffracted field for $\phi' = 75^\circ$, $\sigma_r = 1.0$, and $\varepsilon_r = 15$. (From [And93] © IEEE.)

where $R_{o(h,c)}$ and $R_{n(h,v)}$ are the reflection coefficients of the incident wedge face and the opposite wedge face as determined from Eqs. (6.5.1) and (6.5.2). The function $F(x)$ is the Fresnel integral

$$F(x) = 2j\sqrt{x} e^{jx} \int_{\sqrt{x}}^{\infty} e^{-j\tau^2} d\tau$$

and the parameter L is defined by

$$L = \frac{lr}{l+r}$$

where l and r are the distances of the transmitter and the receiver, respectively, from the diffracting edge. The remaining parameters are given by

$$a^\pm(\phi) = 2 \cos^2 \frac{2n\pi N^\pm - \phi}{2}$$

and $N^\pm$ are the closest integers satisfying the equation

$$N^\pm = \frac{\phi \pm \pi}{2n}$$

Figure 6.39 shows the magnitude of the diffracted field for a 90° wedge angle, $\phi' = 75^\circ$, $\sigma = 1$, and $\varepsilon = 15$.

6.5.3 Diffused Wall Scattering

It is well known that when an electromagnetic wave meets a rough or nonuniform surface, it scatters. The scattering of radio signals has been studied extensively for many

years, particularly in the field of radar [Eav87]. As the frequency of the impinging wave increases, the wavelength decreases and the roughness of the surface causes increased scattering. At optical frequencies, scattering is the dominant source of electromagnetic wave propagation, as discussed in greater detail in Chapter 10. To model the scattering process, we may treat the intersection of the ray and the surface as a radiating point. The signal radiated from this point will have a special radiation pattern, and within the pattern the signal propagates just as it would from a radiating point in free space. The power is inversely proportional to the square of the distance from the radiating point. Inclusion of the effects of scattering in a transmission analysis can require substantial computational power. Optical signals are confined within a room; and as we will see later, inclusion of a detailed scattering model with Lambertian laws is numerically feasible. For radio signals, the structures are more complicated, and direct inclusion of scattering is generally not computationally feasible. However, wavelengths are longer at radio frequencies than at optical frequencies, resulting in smaller amounts of scattering. In these cases it is possible to account for scattering by including an additional loss factor in the transmission analysis [And93].

6.5.4 Two-Dimensional Ray Tracing in Small Indoor Areas

In this section we compare the results obtained from a two-dimensional ray-tracing algorithm with the results of measurements made in a small indoor office area. Results of frequency-domain measurements described in Chapter 5 are used as a reference in evaluating the accuracy of the ray-tracing algorithm. The model uses the two-dimensional reflection and transmission mechanism to trace the rays using the ray-shooting technique. This model provides a low-cost means of doing propagation analysis for small indoor areas used for wireless local area network (WLAN) applications. As we discussed earlier, diffraction does not play a major role in most indoor radio propagation scenarios. The diffraction effect may influence propagation significantly in some locations in corridors when the LOS is blocked and the received signal involves multiple reflections and transmissions. However, this is not a likely situation for indoor WLAN applications, where terminals are typically used in reasonably open work areas. One of the major costs in ray tracing is the development of an accurate electronic map of the area. To minimize the cost of preparing the map, one may use the two-dimensional layout plans that are generally available for office, commercial, or residential buildings. Ideally, one would use a scanner to read the plan into the computer.

To examine the accuracy of the simple two-dimensional ray-tracing algorithm, we look at a typical indoor area, shown in Fig. 6.40, which was used for measurement and modeling in Chapters 4 and 5. The software and the interface graphics for this model were developed as part of the work reported in [Hol92a, Hol92b]. Figure 6.30 shows a sample measured ray tracing, the associated impulse response, and the message window seen on the computer screen. A total of 621 measurements were taken near the center of the building in an area comprising several labs and offices. The receiver was located at the center of Room 1 (Fig. 6.40), and the transmitter was moved to different locations in several of the surrounding rooms as well as in Room 1. A simplified model of the actual floor plan was entered into the ray-tracing program to simulate the environment. The model contains all of the walls, doors, and windows in the area being simulated. The simulation was performed for the same locations at which the measurements were taken.

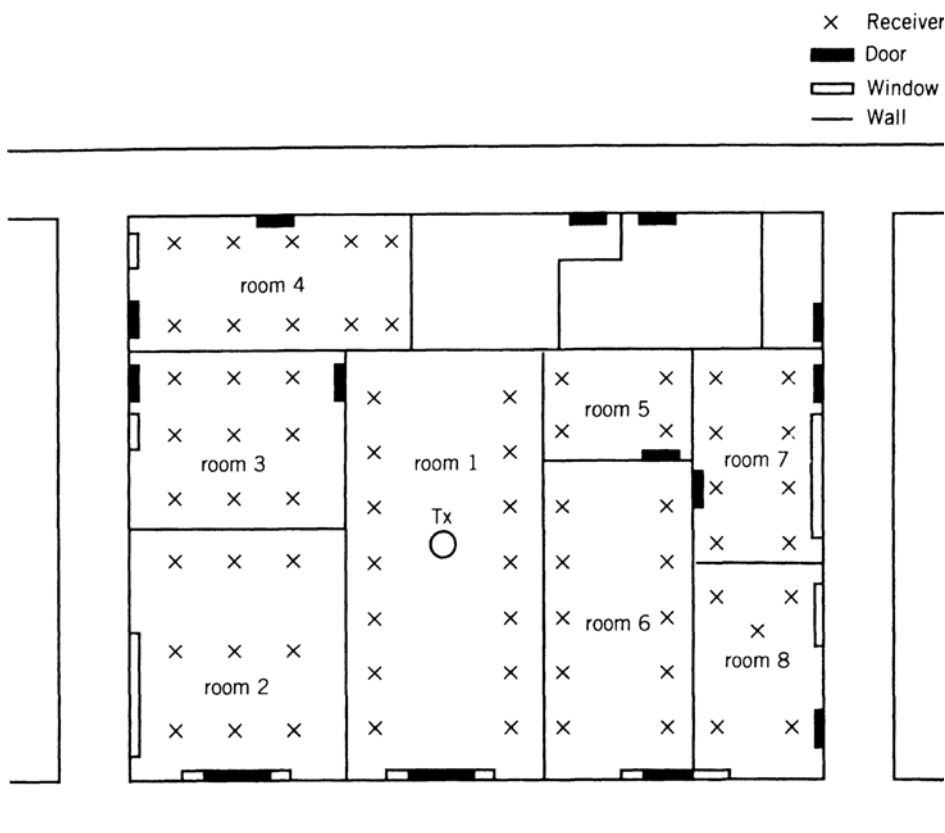


FIGURE 6.40 Floor plan of one section of the second floor of the Atwater Kent Laboratories. The dimension of the internal area is 12×16 m.

The 621 measurements consisted of nine evenly spaced measurements in each of 69 locations distributed throughout the area shown in Fig. 6.40. The 69 locations are marked with an “x” in the figure, and nine measurements were taken in a $2\text{-} \times 2\text{-ft}$ area surrounding each of the marked locations. The average normalized received power and the average rms delay spread for each local area were calculated. The simulation consisted of calculating the average received power and rms delay spread in each of the 69 locations. Approximately 200 receiver locations, evenly distributed over the same $2\text{-} \times 2\text{-ft}$ area, were simulated and the results from the 200 locations were averaged together. Figures 6.41 and 6.42 show the average received power and rms delay spread at each of the 69 locations for both the measurements and the simulation. These graphs show that the received power and rms delay spread follow essentially the same trends from one location to the other.

Based on these propagation models, received powers predicted by the models are examined by comparing them with measured data. Figure 6.43 shows the point-to-point comparison between predicted and measured power in all 621 locations for two- and three-dimensional ray-tracing models. It was determined that if two- and three-dimensional ray-tracing models are used, differences between the two predicted powers are less than 1 dB, and consequently, the two-dimensional model is judged to be valid

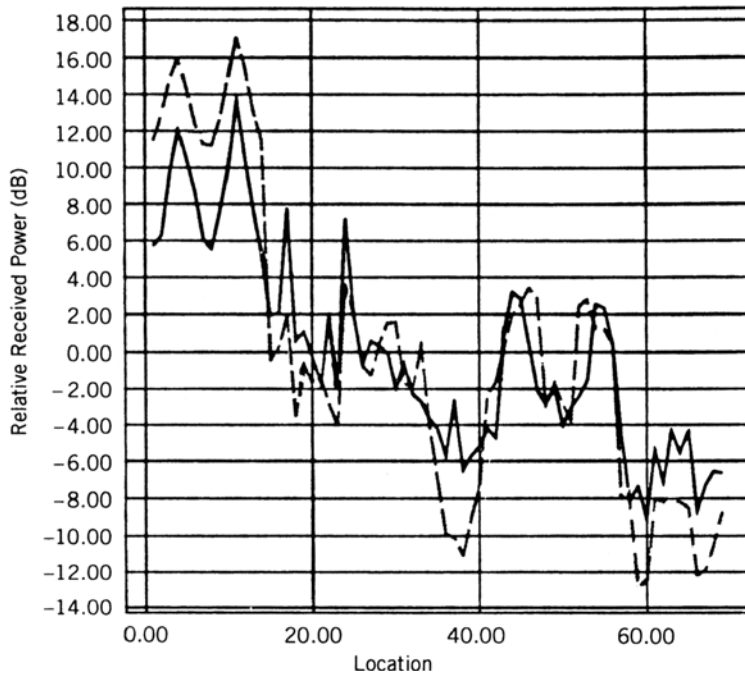


FIGURE 6.41 Average received power versus location for ray tracing, and the results of measurements. Solid line, measurements; dashed line, simulation.

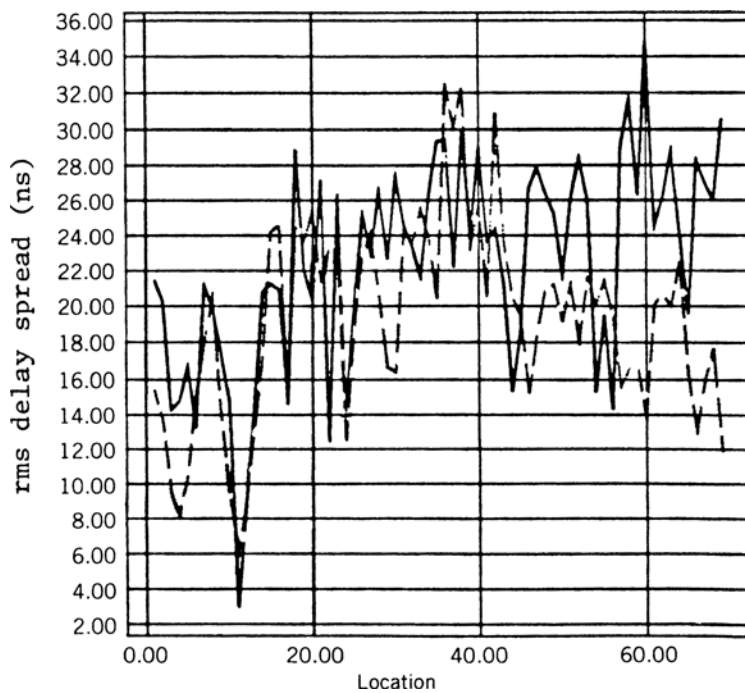


FIGURE 6.42 Average rms delay spread versus location for the ray tracing, and the results of measurements. Solid line, measurements; dashed line, simulation.

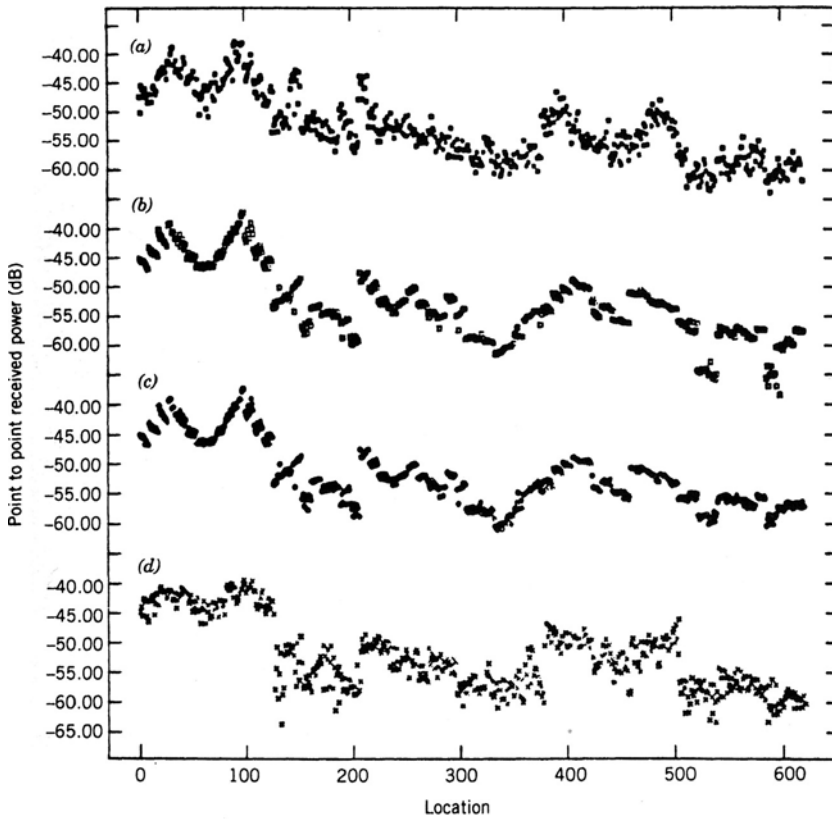


FIGURE 6.43 Received power for (a) measurements, (b) two-dimensional ray tracing, (c) three-dimensional ray tracing, and (d) two-dimensional FDTD modeling at 621 points in the area shown by Fig. 6.36. FDTD modeling is discussed in Section 6.6.

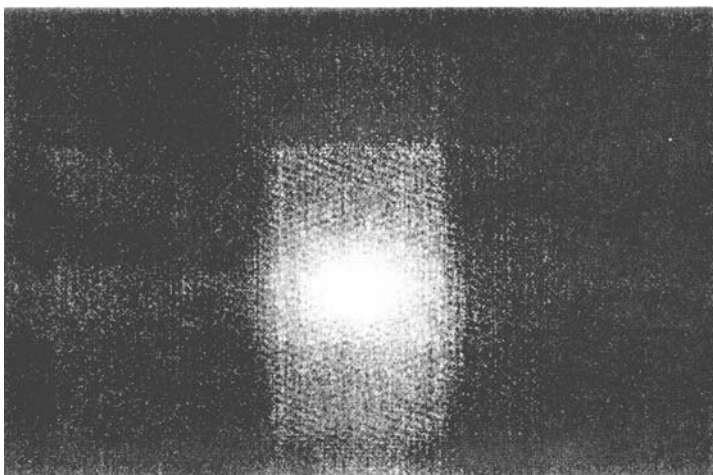


FIGURE 6.44 Power coverage prediction using two-dimensional ray tracing in the area shown by Fig. 6.36.

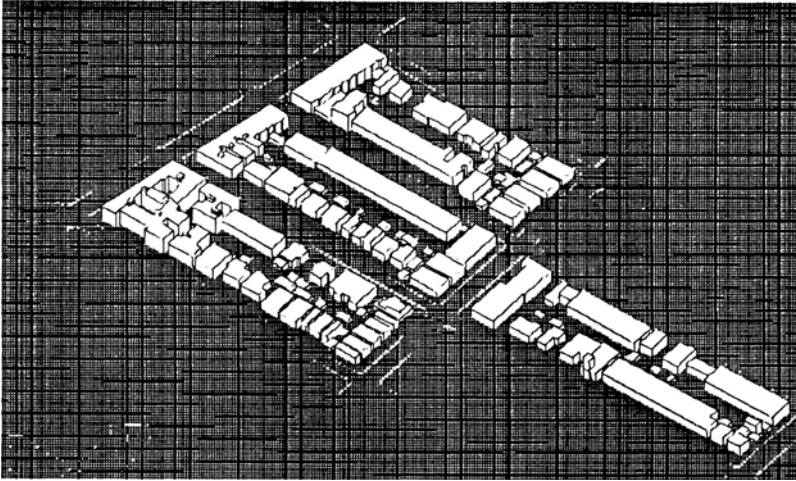


FIGURE 6.45 Four-block area in Brooklyn, New York, displayed by AutoCAD. The area is approximately 400×250 m.

and suitable for power prediction for this area. Judging from this figure, we see that the ray-tracing model is in good agreement with the measurements. The standard deviation of prediction error with the ray-tracing model is 2.3 dB. Figure 6.44 depicts the power coverage prediction for the area shown in Fig. 6.40.

6.5.5 Three-Dimensional Simulation in Urban Microcellular Environments

Here we consider the simulation of a typical residential area, located on 71st and 72nd Streets, between Fort Hamilton Parkway and 10th Avenue, in Kings County, New York. This area consists mostly of two- and three-story private homes, although there are two four-story apartment buildings on the southern corner of 72nd Street and Fort Hamilton Parkway. Extensive measurements have been made in this area, and a three-dimensional description of these city blocks has been transferred to AutoCAD format [Yan94b]. Figure 6.45 depicts the overlays of the four blocks as displayed by AutoCAD. This simplified overlay provides a good model of realistic residential buildings, although it neglects complex roof structures, trees, and fences.

The simulation is carried out with locations of mobile base stations (MBSs) and personal subscriber units (PSUs), antenna patterns, and antenna heights kept the same as those used during the actual measurements. Figure 6.46 shows that the MBS receiver is located in the backyard of a residence next to a telephone pole while the PSU is positioned at more than 40 different locations. The MBS receiving antenna is 18 ft high, and the PSU transmit antenna height is maintained at either 6 or 15 ft above the ground, providing two measurements for each location.

Figure 6.47 shows a comparison of path losses as determined by measurement and simulation in 80 receiver–transmitter configurations. The results show that power prediction errors are no more than 10 dB at 70% of the locations, while the remainder are in the range 10 to 20 dB. The accuracy of these simulation results is comparable to the most accurate models utilized in the literature.

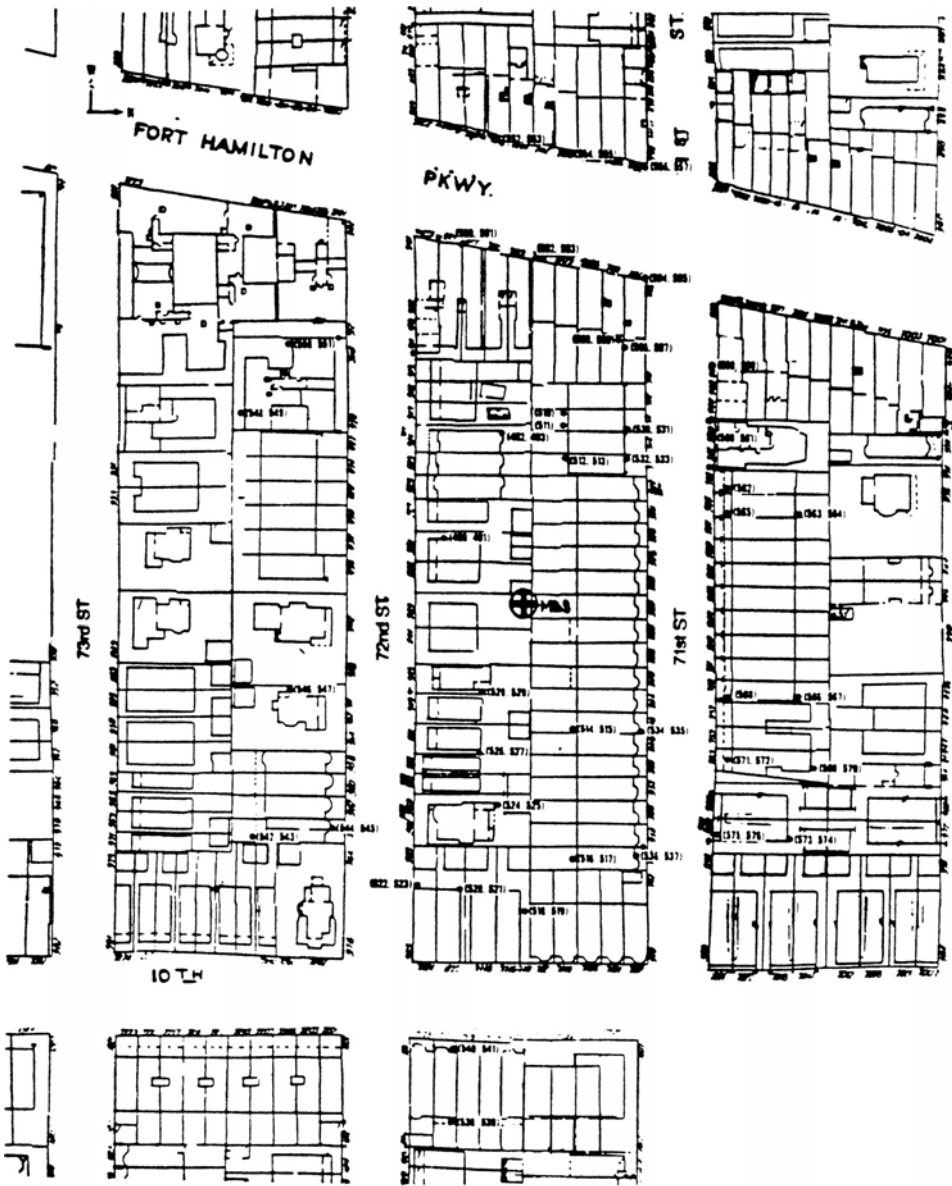


FIGURE 6.46 Location codes for measurements made with a mobile base station located in the backyard next to the telephone pole at 939 72nd Street.

Realistic estimation of power coverage area is important in planning layouts of wireless communication systems. Figure 6.48 shows an example of power coverage predicted by simulations based on the simplified model described above. For more detailed treatments of ray tracing, the reader may refer to [Leb89, Law91, Hol92a, Law92, Leb92, Rap92, Ros92, Ros93, Yan93a, Yan93b].

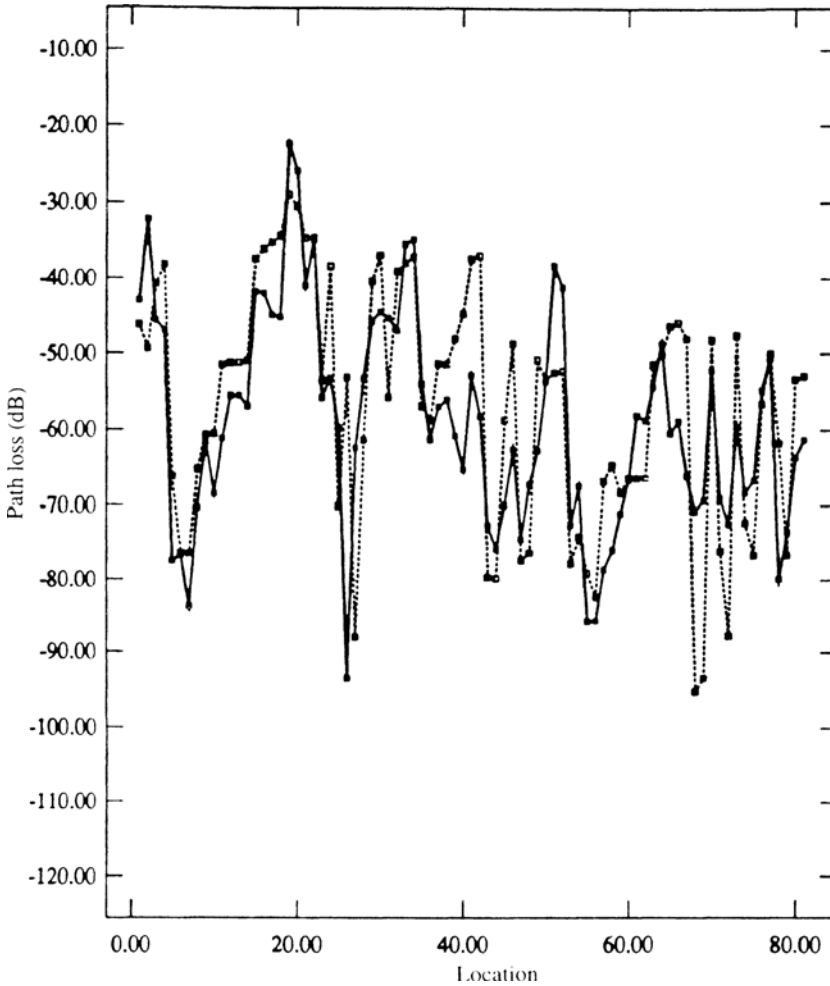


FIGURE 6.47 Path-loss measurements and results of three-dimensional ray-tracing simulations: ●—, measurement; ■—, simulation.

6.6 DIRECT SOLUTION OF RADIO PROPAGATION EQUATIONS

Although ray-tracing models can efficiently predict radio propagation characteristics for indoor and outdoor applications, these techniques are only approximations to the direct solution of electromagnetic wave propagation equations. The ideal method of simulating radio propagation is to solve Maxwell's equations numerically [Yee66, Taf75, Hol83, Fus90, Har92, Lee93]. The numerical solution of these differential equations over a designated area requires selection of a number of points at which the solution is to be determined iteratively. A systematic method of selecting these points is to draw a grid over the area, using certain specified rules, define meshes within the grid, and solve the equations at the nodes of the meshes. The dimension of the meshes within the grid is on the order of a wavelength at the carrier frequency. An increase in the carrier



FIGURE 6.48 Power coverage for the area shown in Fig. 6.41.

frequency or in the area for which the solution is to be obtained will increase the number of nodes exponentially. Therefore, numerical solution typically requires (1) large amounts of memory to keep the track of the solution at all locations and (2) extensive calculations to update the solution at successive instants of time. Given the steady advancements in high-performance computing, the memory size and computational speed of computers are increasing rapidly, and this increases the feasibility of direct solution of Maxwell's equations for coverage areas of substantial size.

With today's computational capabilities, direct solution of the three-dimensional Maxwell's equations for areas with dimensions on the order of several meters and carrier frequencies around 1 GHz is generally not feasible. However, much work has been done with *finite-difference time-domain (FDTD) numerical methods* for performing direct solutions with reasonable computational complexity. The FDTD method is described in the next section.

6.6.1 Finite-Difference Time-Domain Model

The FDTD method is probably the most straightforward and most widely used method for numerical solution of Maxwell's equations. With this method, Maxwell's equations are approximated by a set of finite-difference equations. By placing the electric and magnetic fields on a staggered grid and defining appropriate initial conditions, the FDTD algorithm employs the central differences to approximate both spatial and temporal derivatives, and it solves Maxwell's equations directly. The distribution of electric

and magnetic fields over the entire grid is calculated incrementally in time; and when the simulation is finished, the propagation characteristics are known at every location in the area under study.

In its original form, the FDTD algorithm used a rectangular grid [Yee66, Taf75], and curves and slanted lines were approximated by “staircases,” which can introduce large errors unless the grid size is made very small. To circumvent this difficulty, a nonorthogonal quadrilateral grid was introduced by Holland [Hol83] in 1983. Later, Fusco [Fus90] applied this approach successfully to solving two-dimensional scattering problems. Recently, Lee and others [Har92, Lee93] have modified the approach to a more computationally efficient form that shows significant improvement in accuracy over the rectangular FDTD algorithm. In this section we describe results obtained with a simple two-dimensional FDTD model of reasonable computational complexity, which was suggested in [Yan93a] and is suitable for modeling areas with dimensions on the order of 10 m. A nonorthogonal mesh rather than a rectangular mesh is used in the FDTD model to reduce the mesh density required to achieve a certain precision and a further reduction of the computational load. The computational time needed with this method was comparable to that of a three-dimensional ray-tracing algorithm.

With this method, the indoor radio propagation problem is formulated as an E -field polarization scattering problem in two dimensions. Both the transmitting and receiving antennas are assumed to be ideally omnidirectional. The source electric field is specified as a function of time at the location of the transmitting antenna. By solving the Maxwell's equations numerically using the nonorthogonal FDTD algorithm, the electric field distribution in the area is calculated incrementally in time. Finally, from the spatial and temporal field distributions, the propagation characteristics are obtained. The area used for examination of this algorithm, shown in Fig. 6.40, is that portion of the Atwater Kent Laboratories at Worcester Polytechnic Institute that we had used previously for frequency-domain measurements, statistical modeling, and ray tracing.

Figure 6.43 shows the point-to-point comparison between predicted and measured power in all 621 measurement locations for two- and three-dimensional ray tracing and for the FDTD models. From this figure we see that both ray-tracing models and the FDTD model are in good agreement with the measurements. The standard deviations of prediction error with the ray-tracing and FDTD models are 2.3 and 2.7 dB, respectively, indicating that ray-tracing provides a more accurate estimate of the power. Figure 6.49 shows the cumulative distribution functions of rms delay spreads of measurements and simulations for all the locations. It is shown that the FDTD model can provide a more accurate prediction of rms delay spreads than the ray-tracing model.

6.7 COMPARISON OF DETERMINISTIC AND STATISTICAL MODELING

Development of a statistical model can involve a substantial effort to collect accurate databases for different environments of interest. However, implementation of statistical models on a computer requires minimal computational resources. In principle, development of deterministic models does not require that any measurements be made. In practice, some measurements are needed to check the accuracy of the model and to determine the values of model parameters such as conductivity and permittivity of structural materials. The implementation of both ray-tracing and FDTD models, however, always requires extensive computational resources.

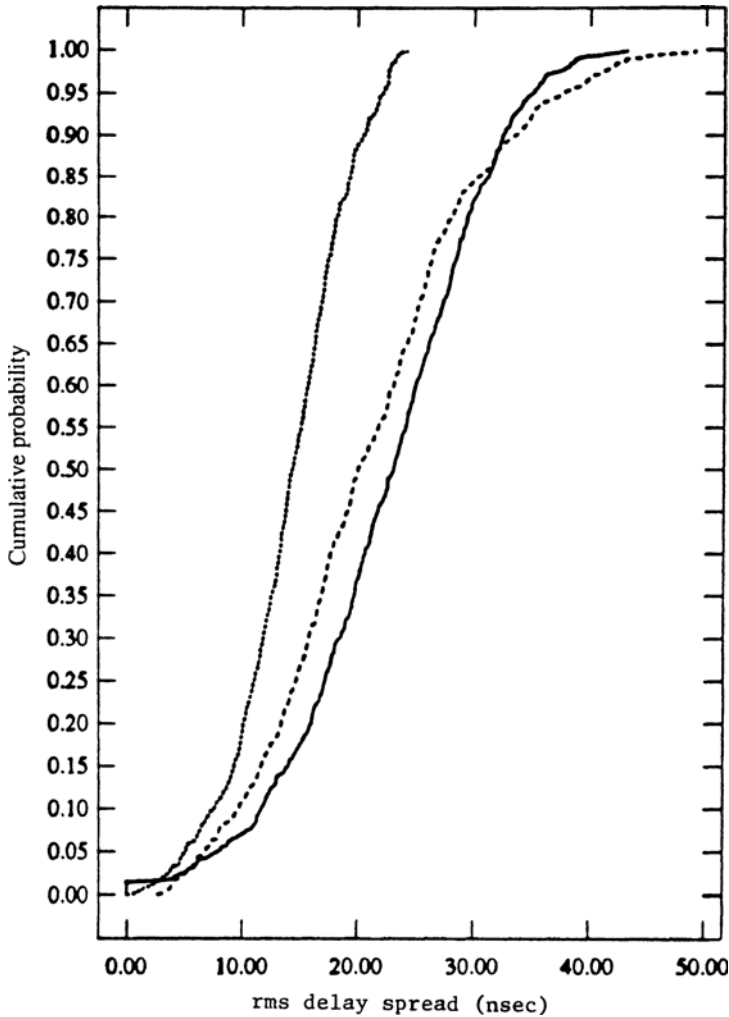


FIGURE 6.49 Comparison of the CDFs of the rms delay spreads for the measurements (—), three-dimensional ray-tracing (···), and two-dimensional FDTD models (---).

Computation time required with the ray-tracing method grows exponentially with the complexity of the details included in the layout of the area, but the size of the area does not necessarily increase the computation time. As a result, the computation time required for modeling a small area with many walls might be as great as for a large area containing only a few buildings. As we noted earlier, application of the ray-tracing technique will in general involve different analytical approaches for different application areas. In indoor areas the transmission and reflections are the dominant mechanisms, whereas in outdoor applications the diffraction and reflections are the most important mechanisms. Computation time for the FDTD technique is proportional to the size of the area, and the addition of structural details does not affect the computation significantly. However, the number of nodes used for computation is related exponentially to the size of the area and the frequency of operation. In general, the ray-tracing technique

is suitable for radio propagation studies in large areas, whereas the FDTD approach is appropriate only for smaller areas.

Statistical models have the advantage of being computationally efficient and easy to generalize. Furthermore, the development of a statistical model is independent of the layout and structural details of the coverage area, which removes the requirement for investing time and resources in surveying outdoor areas or building layouts for each application. The statistical time-domain models in particular are the popular choice of standards bodies, and their recommended models generally include selectable sets of parameters derived from earlier measurements in representative categories of coverage areas. However, a disadvantage of statistical models is that they cannot provide the relationship between the layout of a building or an outdoor area and the detailed channel response in a specific location. Building-specific radio propagation modeling is feasible only with deterministic approaches such as ray tracing and FDTD. Applications such as the optimum siting of antennas, or the analysis of systems using sectorized antennas, are best served by use of deterministic ray-tracing models, which can provide realistic estimates of the azimuthal distribution of rays received in a multipath environment.

6.8 SITE-SPECIFIC STATISTICAL MODEL

The ray-tracing algorithm has been used for predicting radio propagation characteristics, such as the average received power and the multipath behavior. Despite its accuracy, ray tracing is computationally intensive, due to the massive number of ray-tracing operations required to determine the multipath behavior of the channel. In this section, a model originally proposed in [Ali98, Ali02] is presented that is based on the statistical derivation of the ray-tracing operation. This model assumes that each path consists of a number of rays having the same traveling length arriving at the receiver at the same delay. The pattern and length of the rays in these paths are related to statistical parameters of the site-specific features of indoor environment. Using this relationship the delay power spectrum of the channel is calculated based on the density and transmission and reflection coefficients of the walls in the building.

6.8.1 Power of a Path with a Given Length

Figure 6.50 shows the basic computation behind ray-tracing algorithms and three example rays in a sample indoor area where transmission and reflections are the only dominant propagation mechanisms. The first ray is the DLOS ray of length L with two transmissions, the second ray has one reflection and three transmissions, and the third ray has three reflections and four transmissions. For a given ray traveling l meters from the transmitter (Tx) and undergoing n intersections (m reflections and $n - m$ transmissions), the received path power is given by

$$P(l, n, m) = P_0 l^{-2} R^{2m} T^{2(n-m)} \quad (6.8.1)$$

where R and T are the mean reflection and transmission coefficients, respectively, $P_0 = G_t G_r (\lambda/4\pi)^2$ is the free-space normalized received power at distance 1 m [see Eq. (3.2.1)]. The analysis in [Ali02] assumes that for a transmitter and receiver located L meters apart, we have infinite rays with the same length l . With this assumption, the

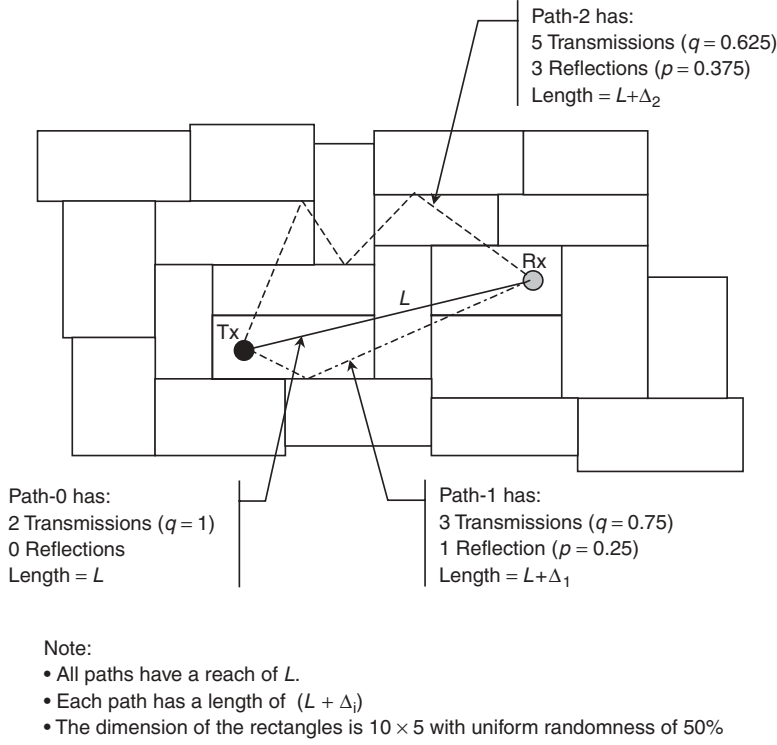


FIGURE 6.50 Basic concept for analysis of the statistical behavior of paths in an indoor area [Ali02].

mean path power can be expressed as

$$P(l) = P_0 l^{-2} \sum_{n=0}^{\infty} \sum_{m=0}^n f(n, m|l) R^{2m} T^{2(n-m)} \quad (6.8.2)$$

where $f(n, m|l)$ is the probability density function of a path that intersects n objects after traveling distance l with m reflections and $n - m$ transmissions.

To find the $f(n, m|l)$, [Ali02] assumes that the number of reflection and transmissions are independent and exclusive. Hence, $f(n, m|l)$ can be decomposed as a multiplication of two functions:

$$f(n, m|l) = f_1(n|l) f_2(m|n, l) \quad (6.8.3)$$

where $f_1(n|l)$ is the PDF for a path that has undergone n intersections after traveling distance l . This function is further assumed to have a Poisson distribution:

$$f_1(n|l) = \frac{(\lambda l)^n}{n!} e^{-\lambda l} \quad (6.8.4)$$

where $1/\lambda$ is the *mean free distance*² between two intersections, which depends on the building layout. The second function $f_2(m|n, l)$, on the other hand, gives the

²The *mean free distance* is defined as the mean distance that a ray can travel before it intersects with an object.

probability of having exactly m reflections and $n - m$ transmissions in path length l , and it is assumed to have a binomial distribution,

$$f_2(m|n, l) = \binom{n}{m} p^m(l) q^{n-m}(l) \quad (6.8.5)$$

where $p(l)$ and $q(l)$ are the probabilities of reflection and transmission, respectively. Figure 6.50 provides the value of these probabilities for the three example rays discussed earlier. Substituting Eqs. (6.8.4) and (6.8.5) into Eq. (6.8.2), after a few algebraic manipulations we have

$$P(l) = P_0 l^{-2} e^{-\lambda l} e^{\lambda l(qT^2 + pR^2)} \quad (6.8.6)$$

This equation gives an explicit relationship between the average power of a path with site-specific details and the building layout via λ , and the floor materials via R and T . By estimating the values of these parameters based on the location of both the transmitter and the receiver, Eq. (6.8.6) can be applied to predict the power of a path versus distance.

To use Eq. (6.8.6) for predicting the power of a multipath arrival knowing the location of the transmitter and the receiver, it is important to know how p and q change with path distance l . To do that, let L denote the transmitter–receiver distance; therefore, $l = L + \Delta$, where Δ is the difference between the total path length and the transmitter–receiver distance. As shown in Figure 6.50, for the DLOS path $p(l = L) = p(\Delta = 0) = 0$, because LOS cannot undergo any reflections. As the length of the path increases, the number of reflections increases, which results in an increase in p and a reduction in q . In Fig. 6.50, for example, p has increased from 0 to 0.25 and 0.375 as the length of the path increases, and consequently, Δ increases. When Δ becomes very large, one expects that p approaches 0.5, because the probability of reflection or transmission should be the same. Using this argument [Ali02] suggests that we model p and q by

$$p(\Delta) = \frac{1 - e^{-\lambda \Delta}}{2} \Rightarrow q(\Delta) = \frac{1 + e^{-\lambda \Delta}}{2} \quad (6.8.7)$$

This models fits for $\Delta = 0$ and very large values of Δ , and the results of Monte Carlo simulation in [Ali02] confirms the appropriateness of this model.

6.8.2 General Formula for the Delay Power Spectrum

Substituting Eq. (6.8.7) into Eq. (6.8.6), and knowing the parameters, a power profile can be predicated for particular L . Then

$$P(l) = P(L + \Delta) = P_0 l^{-2} e^{-\lambda l} e^{\lambda l(T^2 + R^2)/2} e^{\lambda l(T^2 - R^2)e^{-\Delta \lambda}/2} \quad (6.8.8)$$

Since Δ associates with the distance beyond the DLOS, the delay associated to this distance is the delay between the arrival time of the DLOS path and the paths with

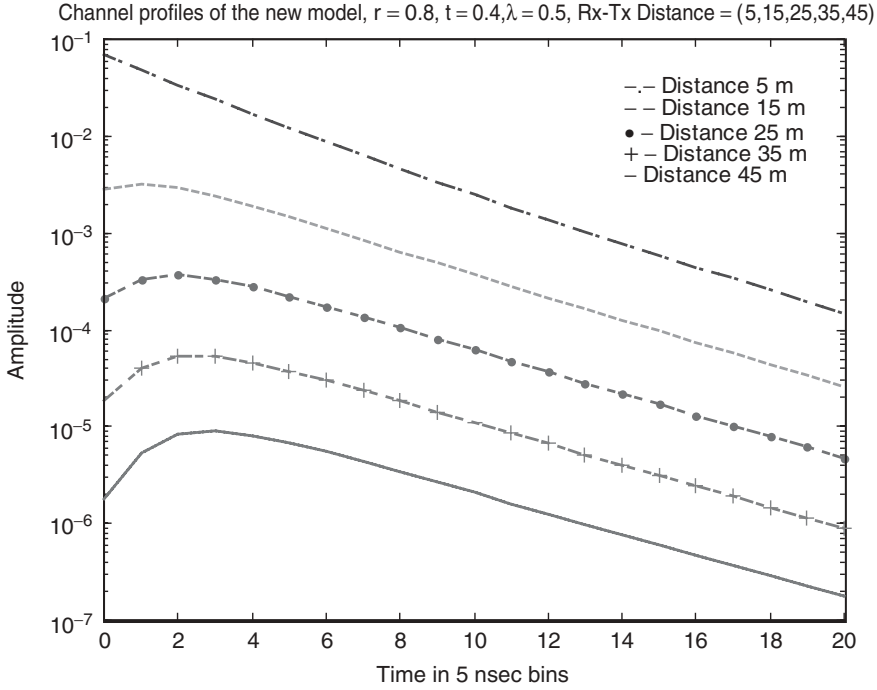


FIGURE 6.51 Calculated delay power spectrum at various distances [Ali02].

length l corresponding with the excess arrival delay of $\tau = \Delta/c$, where c is the speed of light. Hence, Eq. (6.8.8) can be written as

$$Q(\tau) = P(L, \tau) = P_0(L + c\tau)^{-2} e^{-\lambda(L+c\tau)} e^{\lambda(L+c\tau)(T^2+R^2)/2} e^{\lambda(L+c\tau)(T^2-R^2)} e^{-\lambda c\tau/2} \quad (6.8.9)$$

This equation represents the delay power spectrum or the power-delay profile for indoor radio channel. Figure 6.51 represents five delay power profiles calculated from Eq. (6.8.9) for distances of 5, 15, 25, 35, and 45 m, respectively. In short distances the strongest received power is around the DLOS path; as the distance increases, the peak received power shifts away from the DLOS path. In this figure the reflection and transmission coefficients are $R = 0.8$ and $T = 0.4$ and $\lambda = 0.5$. The most influential parameter is the Tx–Rx distance; λ ranks second. T and R have a roughly similar effect. More details of calculations of these parameters for a given building layout are available in [Ali98, Ali02].

The area under Eq. (6.8.9) is the total received power,

$$P_r = \int_{L/c}^{\infty} Q(\tau) d\tau \quad (6.8.10)$$

Figure 6.52 shows a comparison between the results of ray tracing and calculation of the received signal strength for different locations on the second floor of the Atwater Kent Laboratories at Worcester Polytechnic Institute (Fig. 6.40). We see the pattern of two slopes and a breakpoint around 10 m. Figure 6.53 shows the results of calculations

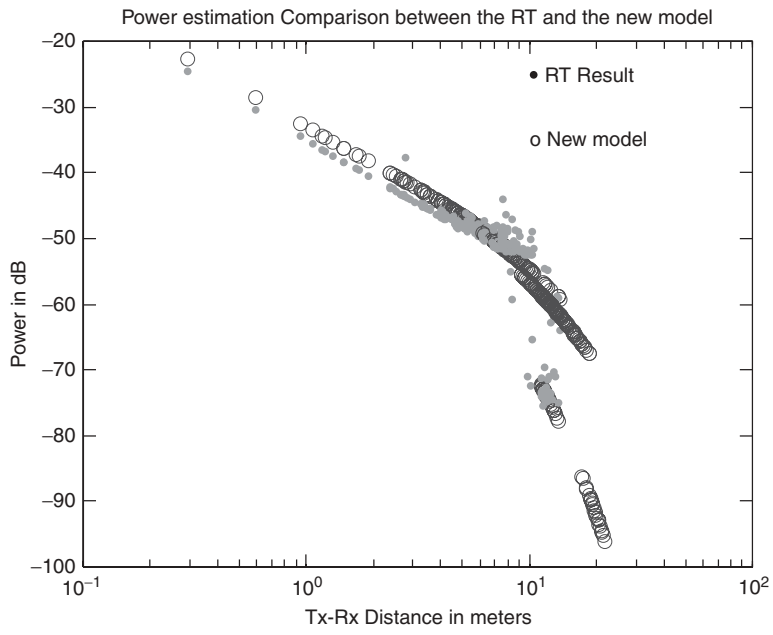


FIGURE 6.52 Measured power for the model presented in [Ali02] versus the results of ray tracing.

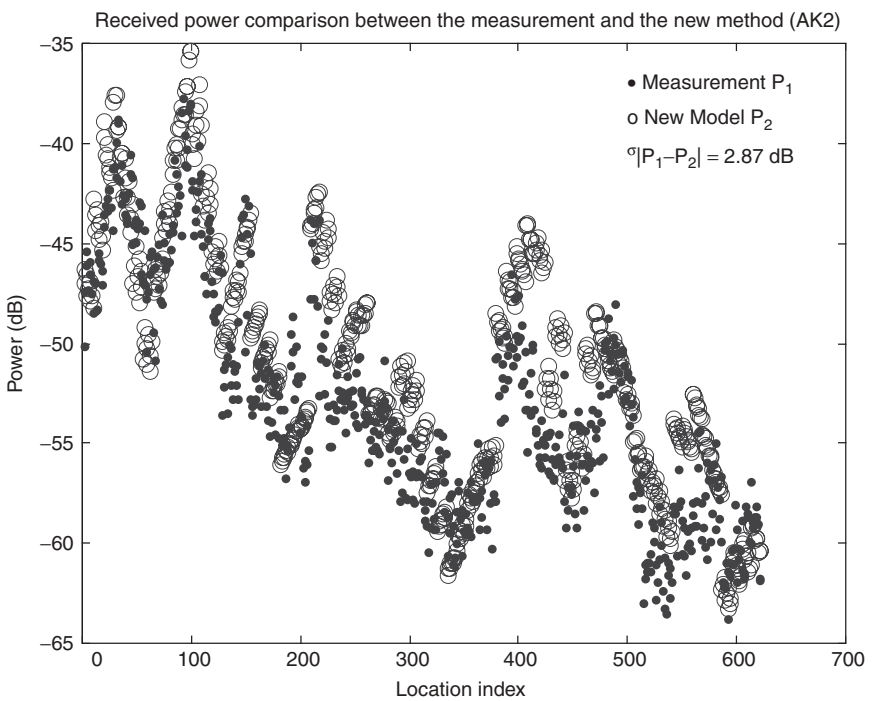


FIGURE 6.53 Received power for the model presented in [Ali02] and the measurements in 621 locations shown in Fig. 6.36.

using Eq. (6.8.9), and measurements in 621 locations on the second floor of the Atwater Kent Laboratories show a very close fit to the results of measurements. Figures 6.43 and 6.53 can be used to compare the results with the results of the ray-tracing and FDTD approaches. This technique provides a much better fit for the received signal strength than do the gradient methods introduced in Chapter 4.

APPENDIX 6A: GSM-RECOMMENDED MULTIPATH PROPAGATION MODELS

Parameters recommended by the GSM standards organization for modeling and simulation of mobile radio channels in various environments are shown in Table 6A.1 to 6A.3. The sets of model parameters for outdoor areas are identified as RAx for rural areas, HTx for hilly terrain, and TUX for urban areas. A fourth set of parameters, EQx, is provided for standardized testing of adaptive equalizers. For each of the four cases, a profile of six tap settings is provided, and for the HTx and TUX cases, a 12-tap profile is also provided. The six-tap models were defined by GSM to provide for situations in which a given simulator does not have the capability to implement a 12-tap model. For each model, two equivalent alternative profiles of tap settings

TABLE 6A.1 GSM Model Parameters for Hilly Terrain

Number of Taps	Relative Time (μ s)		Average Relative Power (dB)		Doppler Spectrum
	(1)	(2)	(1)	(2)	
Typical case for hilly terrain (HTx) (12-tap setting)					
1	0.0	0.0	-10.0	-10.0	Classical
2	0.1	0.2	-8.0	-8.0	Classical
3	0.3	0.4	-6.0	-6.0	Classical
4	0.5	0.6	-4.0	-4.0	Classical
5	0.7	0.8	0.0	0.0	Classical
6	1.0	2.0	0.0	0.0	Classical
7	1.3	2.4	-4.0	-4.0	Classical
8	15.0	15.0	-8.0	-8.0	Classical
9	15.2	15.2	-9.0	-9.0	Classical
10	15.7	15.8	-10.0	-10.0	Classical
11	17.2	17.2	-12.0	-12.0	Classical
12	20.0	20.0	-14.0	-14.0	Classical
Reduced setting (six taps)					
1	0.0	0.0	0.0	0.0	Classical
2	0.1	0.2	-1.5	-2.0	Classical
3	0.3	0.4	-4.5	-4.0	Classical
4	0.5	0.6	-7.5	-7.0	Classical
5	15.0	15.0	-8.0	-6.0	Classical
6	17.2	17.2	-17.7	-12.0	Classical

Source: Annex 3 of GSM Recommendation 05.05.

TABLE 6A.2 GSM Model Parameters for Urban Area

Number of Taps	Relative Time (μ s)		Average Relative Power (dB)		Doppler Spectrum
	(1)	(2)	(1)	(2)	
<i>Typical case for urban area (TUx) (12-tap setting)</i>					
1	0.0	0.0	−4.0	−4.0	Classical
2	0.1	0.2	−3.0	−3.0	Classical
3	0.3	0.4	0.0	0.0	Classical
4	0.5	0.6	−2.6	−2.0	Classical
5	0.8	0.8	−3.0	−3.0	Classical
6	1.1	1.2	−5.0	−5.0	Classical
7	1.3	1.4	−7.0	−7.0	Classical
8	1.7	1.8	−5.0	−5.0	Classical
9	2.3	2.4	−6.5	−6.0	Classical
10	3.1	3.0	−8.6	−9.0	Classical
11	3.2	3.2	−11.0	−11.0	Classical
12	5.0	5.0	−10.0	−10.0	Classical
<i>Reduced TUx setting (six taps)</i>					
1	0.0	0.0	−3.0	−3.0	Classical
2	0.2	0.2	0.0	0.0	Classical
3	0.5	0.6	−2.0	−2.0	Classical
4	1.6	1.6	−6.0	−6.0	Classical
5	2.3	2.4	−8.0	−8.0	Classical
6	5.0	5.0	−10.0	−10.0	Classical

Source: Annex 3 of GSM Recommendation 05.05.

TABLE 6A.3 GSM Parameters for Testing Adaptive Equalizers, Profile for Equalization Test (EQx), Six-Tap Setting

Number of Taps	Relative Time (μ s)	Average Relative Power (dB)	Doppler Spectrum
1	0.0	0.0	Classical
2	3.2	0.0	Classical
3	6.4	0.0	Classical
4	9.6	0.0	Classical
5	12.8	0.0	Classical
6	16.0	0.0	Classical

Source: Annex 3 of GSM Recommendation 05.05.

are given, indicated, respectively, by (1) and (2) in appropriate columns. The Doppler spectrum to be applied at each tap, identified in the rightmost column of each table, is either a Rician (RICE) or classical Doppler (CLASS) spectrum, as discussed in Section 6.2.

APPENDIX 6B: WIDEBAND MULTIPATH PROPAGATION MODELS

Parameters recommended by the Joint Technical Committee (JTC) for PCS Air Interface Standards for time-domain simulation of PCS channels in various environments are shown in Tables 6B.1 to 6B.8. The various categories of communication environments are identified in Table 6.3 and discussed briefly in Section 6.2.

TABLE 6B.1 JTC Channel Model Parameters for Indoor Office Areas

Number of Taps	Channel A		Channel B		Channel C		Doppler Spectrum
	Relative Delay (ns)	Average Power (dB)	Relative Delay (ns)	Average Power (dB)	Relative Delay (ns)	Average Power (dB)	
1	0	0	0	0	0	0	Flat
2	100	−8.5	100	−3.6	200	−1.4	Flat
3			200	−7.2	500	−2.4	Flat
4			300	−10.8	700	−4.8	Flat
5			500	−18.0	1100	−1.0	Flat
6			700	−25.2	2400	−16.3	Flat

TABLE 6B.2 JTC Channel Model Parameters for Indoor Commercial Areas

Number of Taps	Channel A		Channel B		Channel C		Doppler Spectrum
	Relative Delay (ns)	Average Power (dB)	Relative Delay (ns)	Average Power (dB)	Relative Delay (ns)	Average Power (dB)	
1	0	0	0	0	0	0	Flat
2	100	−5.9	100	−0.2	200	−4.9	Flat
3	200	−14.6	200	−5.4	500	−3.8	Flat
4			400	−6.9	700	−1.8	Flat
5			500	−24.5	2100	−21.7	Flat
6			700	−29.7	2700	−11.5	Flat

TABLE 6B.3 JTC Channel Model Parameters for Outdoor Urban High-Rise Areas: Low Antenna

Number of Taps	Channel A		Channel B		Channel C		Doppler Spectrum
	Relative Delay (ns)	Average Power (dB)	Relative Delay (ns)	Average Power (dB)	Relative Delay (ns)	Average Power (dB)	
1	0	0	0	0	0	0	Classical
2	100	−3.6	200	−0.9	500	−2.1	Classical
3	200	−7.2	800	−4.9	800	−12.4	Classical
4	300	−10.8	1,200	−8.4	2,200	−4.1	Classical
5	500	−18.0	2,300	−7.8	7,000	−11.1	Classical
6	700	−25.2	3,700	−23.9	10,000	−19.7	Classical

**TABLE 6B.4 JTC Channel Model Parameters for Outdoor Urban Low-Rise Areas:
Low Antenna**

Number of Taps	Channel A		Channel B		Channel C		Doppler Spectrum
	Relative Delay (ns)	Average Power (dB)	Relative Delay (ns)	Average Power (dB)	Relative Delay (ns)	Average Power (dB)	
1	0	0	0	0	0	0	Classical
2	100	−3.6	200	−0.9	500	−0.2	Classical
3	200	−7.2	800	−4.9	800	−1.5	Classical
4	300	−10.8	1,200	−8.4	2,200	−2.6	Classical
5	500	−18.0	2,300	−7.8	7,000	−11.8	Classical
6	700	−25.2	3,700	−23.9	10,000	−6.9	Classical

**TABLE 6B.5 JTC Channel Model Parameters for Outdoor Residential Areas:
Low Antenna**

Number of Taps	Channel A		Channel B		Channel C		Doppler Spectrum
	Relative Delay (ns)	Average Power (dB)	Relative Delay (ns)	Average Power (dB)	Relative Delay (ns)	Average Power (dB)	
1	0	0	0	0	0	0	Classical
2	100	−6.0	200	−1.4	200	−1.3	Classical
3	200	−11.9	500	−2.4	400	−2.7	Classical
4	300	−17.9	700	−4.8	2000	−13.4	Classical
5			1100	−1.0	2900	−6.1	Classical
6			2400	−16.3	4000	−16.8	Classical

**TABLE 6B.6 JTC Channel Model Parameters for Outdoor Urban High-Rise Areas:
Low Antenna**

Number of Taps	Channel A		Channel B		Channel C		Doppler Spectrum
	Relative Delay (ns)	Average Power (dB)	Relative Delay (ns)	Average Power (dB)	Relative Delay (ns)	Average Power (dB)	
1	0	0	0	0.0	0	−1.2	Classical
2	200	−4.9	200	−0.3	700	0	Classical
3	500	−3.8	1,300	−1.6	4,000	−6.7	Classical
4	700	−1.8	6,500	−6.1	10,000	−3.6	Classical
5	2,100	−21.7	13,500	−9.3	22,000	−4.3	Classical
6	2,700	−11.5	19,000	−23.6	50,000	−19.5	Classical

TABLE 6B.7 JTC Channel Model Parameters for Outdoor Urban/Suburban Low-Rise Areas: High Antenna

Number of Taps	Channel A		Channel B		Channel C		Doppler Spectrum
	Relative Delay (ns)	Average Power (dB)	Relative Delay (ns)	Average Power (dB)	Relative Delay (ns)	Average Power (dB)	
1	0	−1.6	0	−2.5	0	−4.8	Classical
2	100	−5.1	300	0	300	−0.3	Classical
3	200	0	8,900	−12.8	800	−7.4	Classical
4	500	−7.6	12,900	−10.0	8,000	0	Classical
5	1,200	−6.9	17,100	−25.2	27,000	−6.5	Classical
6	1,600	−27.6	20,000	−16.0	55,000	−9.8	Classical

TABLE 6B.8 JTC Channel Model Parameters for Outdoor Residential Areas: High Antenna

Number of Taps	Channel A		Channel B		Channel C		Doppler Spectrum
	Relative Delay (ns)	Average Power (dB)	Relative Delay (ns)	Average Power (dB)	Relative Delay (ns)	Average Power (dB)	
1	0	−3.8	0	0	0	−1.2	Classical
2	100	0	200	−0.3	2,500	−1.6	Classical
3	500	−6.6	1,300	−2.3	13,000	0	Classical
4	800	−1.2	5,200	−0.2	22,000	−12.5	Classical
5	1,300	−18.4	12,000	−20.8	37,000	−21.4	Classical
7	1,700	−23.7					Classical

QUESTIONS

- Why do we need wideband channel modeling and simulations?
- Although the arrival times of signal paths are known to be random, the models recommended by most traditional standards bodies assume fixed arrival times for the paths. How are modem performance evaluations affected by the assumption of fixed arrival times in these models? Are these models useful in assessing the distance measurement errors in TOA-based geolocation systems?
- What is the classical channel characteristic function used for modeling of radio channel propagation for wireless applications discussed in this book? How is it decomposed into simpler functions for slowly time-varying channels?
- Name three advantages of time-domain statistical modeling.
- Name three advantages of frequency-domain statistical modeling.
- Which form of statistical modeling is commonly used by standards bodies to represent radio propagation media?

- (g) What are the three major classes of environments in the JTC channel model?
- (h) What are the three major categories within each environment?
- (i) How many classes of channel profiles are considered in the GSM channel model?
- (j) Why does the JTC model provide three classes of profiles for each specific area in an environment?
- (k) What is the difference between IEEE 802.11b and the Saleh–Valenzuela model in terms of the models for the arrivals and amplitudes of the paths?
- (l) What is the difference between the Spencer and Saleh–Valenzuela models in describing the path behavior and the use of the models in performance evaluation of emerging technologies?
- (m) Why are models used for WLAN applications not directly applicable to the UWB system?
- (n) How does ray tracing relate to the direct solution of Maxwell's equations?
- (o) How can we use the results of ray tracing in the analysis of MIMO and indoor geolocation systems?
- (p) Discuss advantages and disadvantages of ray tracing relative to statistical modeling.
- (q) Name all the mechanisms for radio propagation. What are the major mechanisms used in ray tracing of the indoor radio channel? What are the major mechanisms used in ray tracing for microcellular outdoor areas?

PROBLEMS

1. If an exponential distribution with variance 2 is used for simulation of the inter-arrival times of a Poisson random process, what is the average arrival rate of the process? Using MatLAB or an alternative computational tool, generate 10 samples of Poisson arrivals and plot them versus time.
2. In the Poisson arrival process, if the time axis is divided into small intervals referred to as *bins*, the existence of arrivals in the bins forms a binomial distribution. This principle is commonly used for the simulation of Poisson path arrivals in computer simulation of wideband radio channel characteristics.
If we observe a Poisson process for T seconds and the average of path arrivals per second is μ , the probability of arrival of l paths in T seconds is given by

$$P(L = l, T) = \frac{(\mu T)^l}{l!} e^{-\mu T} \quad (\text{P6.1})$$

Assume that T_m is the measurable delay spread of the channel (the span of time in which we are able to detect a path in background noise) and that Δt is the resolution of the measurement system (base of the pulses used for measurement). If we use the resolution as the bin interval and we have N bins, then $T_m = N \Delta t$.

- (a) Comparing Eq. (P6.1) with Eq. (6.2.2), observe that $\lambda = \mu T_m$.

- (b) From Eq. (P6.1), show that the probability of path occurrence in a bin is given by

$$r_i = 1 - P(L = 0, \Delta t) = \mu \Delta t e^{-\mu \Delta t}$$

- (c) Using the exponential series expansion, show that for narrow bin widths where $\mu \Delta t$ is a very small number, $r_i = \mu \Delta t$.
- (d) Show that Eq. (6.2.3) holds if the approximation in part (c) holds.
- (e) Using the discussion above, describe a procedure for simulation of the Poisson path arrivals that uses only uniformly distributed random variables to determine the existence of paths in the bins.
3. For frequency-selective fading channels it is sometimes necessary to simulate a notch in the passband of the channel. A convenient model has been developed by Rummler for application in the microwave LOS channels. In this model the frequency response of the channel is represented by

$$H(j\omega) = a[1 - be^{-j(\omega - \omega_0)\tau}]$$

- (a) Assume that $a = 1$ and determine b so that the channel produces a 30-dB notch at the frequency f_0 .
- (b) For $\tau = 6.3$ ns and $f_0 = 20$ MHz, sketch the amplitude and phase response of the transfer function of the channel for $0 < f < 100$ MHz.
4. Another approach to simulating a deep notch is to use the power series in the frequency domain. In the two-term power series model for microwave LOS, the channel frequency response is represented by

$$H(j\omega) = A_0 + (A_1 + jB_1)j\omega$$

- (a) Assume that $A_0 = 1$, and determine the values of A_1 and B_1 that produce an infinite decibel notch at $f_0 = 20$ MHz [$|H(j\omega_0)|^2 = 0$].
- (b) Sketch the magnitude and phase of the frequency response for the parameters determined in part (a) in the range $0 < f < 100$ MHz.
- (c) Describe a method to solve part (a) for any depth of the notch, where the notch depth is defined with respect to $H(0)$.
5. Calculate the rms delay spread of the GSM rural area channel profiles with six and four taps shown in Fig. 6.4 and Table 6.1. Is the value of the rms delay spreads different? Explain.
6. Consider the IEEE 802.11b model described in Section 6.2.2.

- (a) Using general parameters of the delay power spectrum given by Eq. (6.2.9b), show that area under the delay power spectrum function, given by Eq. (6.2.9c), is normalized to 1:

$$\int_{-\infty}^{\infty} Q(\tau) d\tau = 1$$

- (b) Sketch $Q(\tau)$ with $T_s = 100$ ns in an environment with an rms delay spread of 80 ns.

7. Consider the model in [Ali02] described in Section 6.2.8. In an open indoor area with metallic walls, all of the energy of the incident wave is reflected and nothing is transmitted through the medium.
- (a) In this environment, what are the values of reflection coefficient R , transmission coefficient T , probability of reflection p , and probability of transmission q ?
 - (b) Find the average received power in a path calculated from Eq. (6.8.6), and discuss the intuitive correctness of your results.
 - (c) State the delay power spectrum calculated from Eq. (6.8.9), and sketch the results in linear scales for a mean free distance of $\lambda = 0.5$ and a DLOS distance $L = 10$ m.
 - (d) Repeat part (c) for $\lambda = 0.1$ and 2 m and discuss the effects of λ on the shape of the delay power spectrum.

PROJECTS

Project 1: Simulation of Wideband JTC Model

This project is a continuation of Project 3 in chapter 4, in which we simulated the fluctuation of a Rayleigh fading channel using the Clarke–Jakes model and the filtering of Gaussian noise. The traditional method for simulation of the wideband characteristics of a signal, shown in Fig. 6.3, is to use several taps with different delay and power values, each simulating a narrowband channel independently. In this project we use the Gaussian filter channel simulator of Project 3 in Chapter 4 with a tap delay line to implement one of JTC recommendations channel models for indoor office areas.

- (a) Use MatLAB or an alternative computational tool to simulate channel B in the indoor office areas with the JTC model described in Table 6B.1.
- (b) Generate the CDF of the rms delay spread using 100 samples of simulated channel profiles. Compare your CDF with the results of empirical measurements reported in Fig. 6.11. Explain how the model fits the result of measurements.
- (c) Calculate the rms delay spread of the channel from the definition of the channel profile in the table, and compare the result with the average of the rms delay spreads of all simulated samples of the channel profile used in the CDF curve.

Project 2: Channel Simulation Using Poisson Arrivals

In this project we simulate a one-cluster channel with the Poisson arrival time described in Section 6.3.3.

- (a) Use Table 6.4 to determine the arrival rate of the experiment in the manufacturing areas if the arrivals are assumed to be Poisson.
- (b) Use MatLAB or an alternative computational tool to simulate the Poisson–Rayleigh profiles used in Fig. 6.19.
- (c) Generate the CDF of the rms delay spread using 100 samples, and compare the results with those presented in Fig. 6.19.

Project 3: Channel Simulation Using Δ - K Arrival Model

In this project we simulate a one-cluster channel model with the modified Poisson Δ - K arrival times described in Section 6.2.3.

- (a) Use MatLAB or an alternative computational tool and parameters given in Table 6.4 to regenerate the path occurrence probabilities shown in Fig. 6.12.
- (b) Simulate the modified Poisson/lognormal profiles used in Fig. 6.19.
- (c) Generate the CDF of the rms delay spread using 100 samples, and compare the results with those presented in Fig. 6.19.

Project 4: Channel Simulation Using AR Model

In this project we simulate the frequency-domain channel model introduced in Section 6.3.1. For implementation of this project we use the results of Project 1 in Chapter 5 for generating channel impulse responses from the simulated frequency response of the channel.

- (a) Use MatLAB or an alternative computational tool and the statistics provided in Table 6.5 to simulate the frequency-domain AR model 2 for experiment G1 described in Section 6.3.1. Generate a sample frequency response and a sample channel impulse response using the chirp Z -transform described in Project 1 in Chapter 5.
- (b) Use the peak detection algorithm of Project 1 in Chapter 5 to determine the location and amplitude of the sample simulated channel impulse response, and calculate the rms delay spread of the channel impulse response.
- (c) Generate the CDF of the rms delay spread using 100 samples of the simulated channel impulse responses. Compare the results of your simulation with those shown in Fig. 6.29. Also, determine the average of all rms delay spreads calculated for individual time profiles.
- (d) In Fig. 6.26, the dashed line is the inverse transform of the transfer function for the AR process defined in Eq. (6.3.2), and it is given by

$$\tilde{h}(\tau) = G(z)|_{z=e^{j2\pi\tau}} = \prod_{i=1}^p \frac{1}{1 - p_i e^{-j2\pi\tau}}$$

If we consider the inverse Fourier transform of a transfer function with poles located at the average location of the poles, $\overline{p_i}$, as the delay power spectrum of the channel, we can simply derive a multicluster delay power spectrum for the channel, given by

$$\mathcal{Q}(\tau) = |\tilde{h}(\tau)|^2 = \left| \prod_{i=1}^p \frac{1}{1 - \overline{p_i} e^{-j2\pi\tau}} \right|^2$$

Sketch the double clustered delay power spectrum of the channel using the average locations of the poles in the area G1. Calculate the rms delay spread of the channel from this newly defined delay power spectrum, and compare the results with the average of the rms delay spreads found from 100 simulations in part (c).

PART III

MODEM DESIGN

The diversity and complexity of transmission techniques in wireless systems are far greater than are in wired networks. This part of the book provides comprehensive coverage of major transmission techniques employed in voice-oriented cellular and PCS systems as well as in data-oriented mobile data, WLAN, and WPAN systems. These modulation techniques are logically divided into narrowband, wideband, and spread-spectrum techniques.

Chapter 7: Narrowband Modem Technology

In this chapter we introduce the basic modulation and coding techniques used in a variety of modems. We begin with traditional FSK, PSK, and QAM modulation techniques, which have been favored for high-speed data communications in WLANs, WPANs, and mobile data networks. Then we introduce GMSK and $\pi/4$ -QPSK modulation techniques, used in voice-oriented wide-area digital cellular networks. The chapter concludes with a discussion of practical aspects of modem design, including synchronization and pulse shaping.

Chapter 8: Fading, Diversity, and Coding

We begin this chapter by considering the limitations encountered in applying traditional transmission techniques to fading multipath radio channels. We discuss the harmful effects of fast multipath fading on the error-rate performance of radio modems, and we introduce diversity and coding techniques as approaches to counteracting the effects of multipath fading. We also introduce recently developed space-time coding and multiple-input multiple-output (MIMO) techniques.

Chapter 9: Broadband Modem Technologies

In this chapter we describe the application of signal processing algorithms in the receiver for improving modem performance and achieving reliable broadband communications in fading multipath conditions. After demonstrating how frequency-selective multipath fading degrades the performance of basic modems designed for narrow-band communications, we introduce a statistical discrete-time model for the behavior of frequency-selective multipath fading channels. The remainder of the chapter is devoted to broadband wireless communication techniques operating over frequency-selective fading channels. We describe discrete matched filters and maximum likelihood sequence estimation, different types of equalizers, sectored antennas, multicarrier modulation, orthogonal frequency-division modulation (OFDM), STC (space-time coding)-MIMO, and MIMO-OFDM techniques for application in broadband modems.

Chapter 10: Spread-Spectrum and CDMA Technology

In this chapter we introduce the basic principles of direct-sequence and frequency-hopping spread-spectrum techniques and show how they are used in code-division multiple-access (CDMA) systems. Next, we analyze the effects of interference and multipath fading on performance of spread-spectrum systems. Then we introduce a number of important topics, such as M -ary orthogonal coding, multicarrier CDMA, multiuser detection, interference cancellation, and MIMO-CDMA techniques.

NARROWBAND MODEM TECHNOLOGY

- 7.1 Introduction
- 7.2 Basic Modulation Techniques
 - 7.2.1 Framework for Analysis
 - 7.2.2 On–Off Keying
 - 7.2.3 Frequency Shift Keying
 - 7.2.4 Phase Shift Keying
 - 7.2.5 Pulse Amplitude Modulation
 - 7.2.6 Quadrature Amplitude Modulation
 - 7.2.7 Multiphase Modulation
 - 7.2.8 Partial-Response Signaling
 - 7.2.9 Trellis-Coded Modulation
 - 7.2.10 Comparison of Modulation Methods
- 7.3 Theoretical Limits and Practical Impairments
 - 7.3.1 Theoretical Limits of Communication Performance
 - 7.3.2 Transmission Channel Impairments
- 7.4 Traditional Modems for Wide-Area Wireless Networks
 - 7.4.1 Requirements for Radio Modems
 - 7.4.2 Digital Frequency Modulation
 - 7.4.3 OQPSK, MSK, and GMSK
 - 7.4.4 $\pi/4$ -Shift QPSK
- 7.5 Other Aspects of Modem Implementation
 - 7.5.1 Power Control
 - 7.5.2 Carrier and Timing Recovery
 - 7.5.3 Pulse Shaping
- Questions
- Problems
- Projects
 - Project 1: Error Rate and Phase Jitter in QPSK Modulation
 - Project 2: Error Rate and Phase Jitter in 16-QAM Modulation
 - Project 3: Design of Raised Cosine Matched Filters

7.1 INTRODUCTION

In this chapter we describe various modulator and demodulator (modem) technologies. We begin with a discussion of basic modem techniques: that is, techniques widely employed in standard voiceband modems for use in the wired public telephone network. Then we discuss the application of basic modem techniques to radio channels, where multipath and fading limit performance and give rise to certain design issues and enhancement techniques that are specific to the radio environment.

The application of modem technology to radio systems has evolved along a path similar to that followed in PSTN data transmission [Pah88c, Bla91] but has lagged several years behind wireline developments. This has, of course, been due to the special characteristics of radio propagation, which create a much harsher environment for data transmission than we encounter in the telephone network. Location-dependent power variations, fading, and multipath act to limit the data rates and performance achievable over a radio channel. Consequently, the earliest of modem applications in radio systems used only the simplest wireline modem techniques, although much more advanced techniques were already in use in the PSTN. For example, while voiceband wireline modems incorporating adaptive equalization were coming into general use in the early 1970s, adaptive equalization was not applied successfully to radio systems until the late 1970s. Similarly, whereas the technique of trellis-coded modulation (TCM) has been used in standard commercial modems for several years now, the application of TCM to mobile radio systems is just beginning. Thus, the transfer of evolving wireline modem technology from the wired network to wireless systems is a continual process, with radio applications employing certain design features and enhancements that are specific to the special characteristics of that environment.

A number of the wireline modem techniques that we describe here are being incorporated into standards for several new mobile wireless systems, including digital cellular systems, digital land-mobile radio systems, and emerging systems for personal communications services (PCSs). In the wireless local area network (WLAN) market, lack of an unlicensed frequency band without restriction on the choice of modulation technique held back application of the more sophisticated modem design technologies in those systems. As the market for portable data terminals, notebook computers, and portable facsimile devices grows, the demand for a band specifically assigned to personal wireless data communications increases. Given the 1993 and 1997 FCC designation of certain 2-GHz PCS and 5-GHz U-NII bands for data services, combined with the ever-growing demand for higher speeds and wider and more robust coverage for wireless data communications, there will undoubtedly be a migration of the more sophisticated wireline modem technologies into wireless data communication products.

Before we move on to the remainder of this chapter, it is worthwhile to review a very basic point: why a radio system must include modulation for data transmission. Although this discussion will be viewed by some readers as unnecessary, it will serve to highlight a few issues that are important in examining differences among different wireless systems designed for the same type of service or application.

In the case of data transmission over a telephone line, it is clear that we need a modulator to convert a stream of digital data to a line signal that is compatible with the passband characteristics of a voice circuit. A typical telephone voice circuit passes signals only in the frequency interval from about 300 Hz to about 3300 Hz, and thus a modem creates a line signal positioned appropriately in that band. Most existing

modem standards utilize a carrier frequency of 1800 Hz, which is about at the center of the passband of this channel. In the case of a radio system, however, one might well ask why we do not simply apply the data stream to a power amplifier feeding a transmitting antenna. There are at least three reasons why we use modulation as an intermediate step.

First, we must consider the fact that for effective signal radiation, the length of the transmitting antenna must be comparable to the wavelength of the signal to be transmitted. Therefore, for example, a binary stream at 3 kb/s has a bandwidth of about 3 kHz, which implies an antenna of length around 100 km, a totally impractical design. But if we modulate the data stream onto a 3-GHz carrier, the appropriate antenna length will be around 10 cm. Furthermore, because the effective bandwidth of an antenna, normalized as a percentage of its center frequency, stay roughly constant from one wavelength to another, the absolute value of the available bandwidth increases proportionally with the center frequency. Thus, we want to modulate the data stream onto an appropriate radio frequency (or onto some intermediate frequency, which is then up-converted to the radio frequency), where antenna size is practical and where antenna bandwidth is ample, or at least adequate, relative to the data rate we wish to transmit.

Second, by modulating the data stream onto a carrier, we can ensure the orderly coexistence of multiple signals in a given spectral band by arranging the carriers in a frequency-division multiplex (FDM) format. In fact, FDM is the oldest and simplest method of providing multiple-user access to a shared segment of frequency spectrum.

Third, particular forms of modulation may provide an especially effective way of reducing interference among intended users or interference from unwanted sources. We refer here to spread-spectrum modulation, which was developed to defeat intentional jamming of military communication systems [Sch82, Pri83]. Spread-spectrum modulation is being used for both wireless voice and data communications. In wireless data communications, the availability of frequency bands (ISM bands) for spread-spectrum modulation has motivated the implementation of a number of spread-spectrum WLAN products. In the wireless portable and mobile radio industries, spread-spectrum modulation is used with a multiuser access technique known as *code-division multiple access* (CDMA) to increase the number of users supported in an allocated frequency band.

In Section 7.2 we describe basic modem techniques and show that a number of the techniques are closely related. We show the transmitting and receiving functions that are performed in all standard modems. We show how modem signal constellations of increasing complexity can provide corresponding increases in bandwidth efficiency: that is, increased data rate per unit of available bandwidth. Increasing bandwidth efficiency has been an important objective in the wireline modem industry, because the bandwidth of telephone channels is fixed by the design of the public network, whereas the demand for ever-higher data rates has increased unabated. The most sophisticated voiceband modems have been developed for use on leased lines, where the modems are part of a data network. Leased lines are more expensive than ordinary dial-up lines, and an increase in the achievable data rate will reduce the required number of leased lines, thus lowering the cost of operating the network. As a result, major users of these networks, such as banks and airline companies, have always been willing to make large investments to steadily update the data rate capabilities of their leased circuits. In the case of radio modems, an increase in the bandwidth efficiency will reduce the

bandwidth required for a given data rate. For the wireless portable and mobile radio industry, where service is provided to a very large number of customers, increasing the bandwidth efficiency allows more simultaneous users in the allocated bandwidth, increasing the revenues of the operating company.

In Section 7.3 we review the theoretical limits on achievable data rate and communication efficiency in an ideal Gaussian noise channel, as defined by Shannon's capacity formula. We then describe briefly the real-world channel impairments that constrain the data rates achievable in practice.

In Section 7.4 we begin by addressing the major issues that arise in the selection of a modulation technique for use on a radio channel. Then we describe the standard radio modems that are being used in existing and emerging new wireless networks. Other modem technologies, utilizing spread-spectrum and infrared techniques, are discussed in Chapters 9 and 14, respectively.

7.2 BASIC MODULATION TECHNIQUES

In this section we outline basic modulation techniques used primarily in standard voiceband modems. It is not our intention here to provide an exhaustive treatment of modulation techniques but rather, to explain the basic forms of modulation, to make comparisons among the various techniques, and to describe the principal functions that must be implemented in any modem. First we provide the framework for analysis of modem techniques, and then we describe various approaches to implementing a modem. More detailed treatments of modulation techniques and modem design may be found in a number of references, including [Feh87, Bin88, Pro01].

7.2.1 Framework for Analysis

In this subsection we discuss certain topics that underlie our subsequent discussions of modem techniques and achievable performance. Our discussion covers matched filtering, different measures and interpretations of SNR, and the criteria used for error-rate analysis for different modems.

Matched Filtering. In establishing a digital communication link, the information digits arriving in sequence are mapped into symbols, which are in turn represented by pulses for transmission over the link. At the receiving end, each pulse waveform received is filtered and then sampled. The sampled signal is used for making a decision on the transmitted symbol, which is then mapped back to information digits. As a result, the basic elements of digital communication are built around pulse transmission and the design of the receiver filter. It is well known that on an additive noise channel, the optimum receiver filter, one that maximizes the SNR after sampling at the receiver, is a *matched filter*. Figure 7.1a shows a block diagram of the optimum receiver for pulse transmission. The transmitted pulse representing the i th symbol is $s_i(t)$, and it is disturbed by additive white Gaussian noise (AWGN) $\eta(t)$ with two-sided spectral density N_0 . The received signal is passed through the filter and sampled at time $t = T$. The sampled signal $z(T)$ has a deterministic component $\sqrt{E_{si}}$, where

$$E_{si} = \int_{-\infty}^{\infty} |s_i(t)|^2 dt \quad (7.2.1)$$

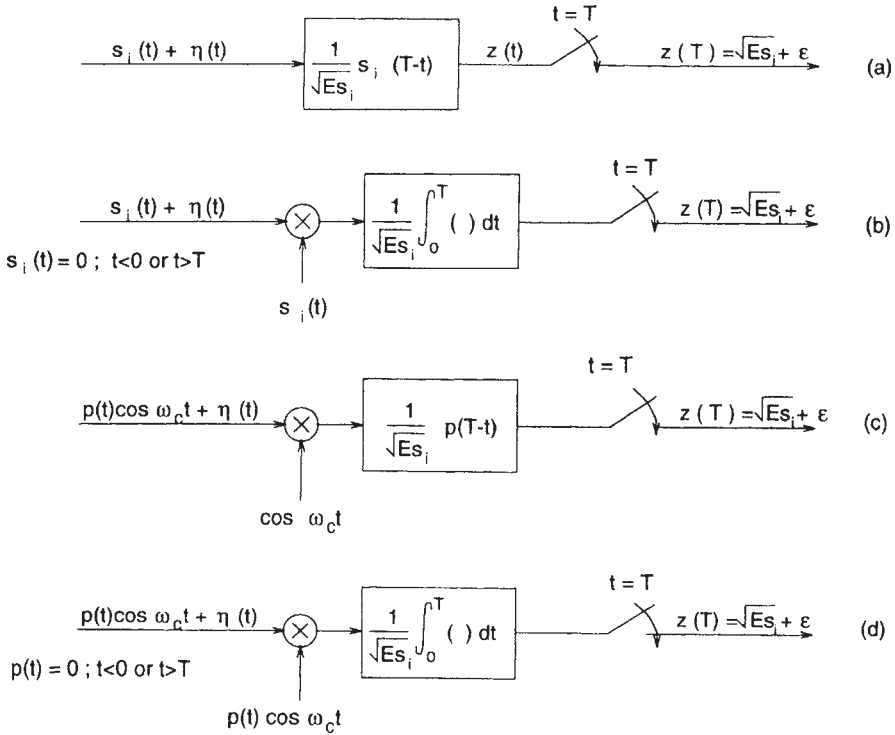


FIGURE 7.1 Matched filter. (a) Block diagram of the optimum receiver for pulse transmission. (b) Matched filter implemented with a correlator when the transmitted pulse is time limited. (c) Matched filter implemented with a mixer followed by a filter matched to $p(t)$. (d) Matched filter implemented with a correlator when $p(t)$ is time limited.

which is referred to as the *energy per symbol*; it also has a random component ε with variance

$$|\varepsilon|^2 = N_0$$

The impulse response of the matched filter is

$$f(t) = \frac{1}{\sqrt{E_{si}}} s_i(T - t) \quad (7.2.2)$$

which is the transmitted pulse waveform reversed and shifted to time $t = T$. The SNR per symbol, γ_{si} , at the output of the matched filter is then given by

$$\gamma_{si} = \frac{E_{si}}{N_0} \quad (7.2.3)$$

which is the ratio of the square of the deterministic component of the sampled signal to the variance of the random component. The energy per symbol E_{si} is an indicator of the transmitted power, and N_0 , the power spectral density of the AWGN, is an indicator of the noise power. Different modulation techniques can be based on different-sized modulation symbol alphabets and employ different pulse waveforms. Thus, to provide

a more meaningful measure, the SNR and energy per symbol are usually averaged over all symbols and are represented by γ_s and E_s , respectively. The error rate of any modulation technique is a function of γ_s . For different modulation techniques, the mapping of the digits into symbols is different, resulting in different equations for calculation of the error rate.

If the transmitted pulse is time limited, the matched filter can, equivalently, be implemented with a correlator, as shown in Fig. 7.1b. In this figure, $S_i(t) = 0$ if $t > T$ or $t < 0$. If the transmitted symbol is a radio-frequency (RF) pulse of the form $S_i(t) = p(t) \cos \omega_c t$, the matched filter will be implemented with a mixer followed by a filter matched to $p(t)$, as shown in Fig. 7.1c. As with the reception of baseband pulses, if $p(t)$ is time limited, the matched filter can be implemented with a correlator, as shown in Fig. 7.1d. The matched filter implementation is commonly used in voiceband modems, whereas the correlator implementation is used in direct-sequence, spread-spectrum modulation systems. The pulse $p(t)$ is a baseband pulse with energy $E_p = 2E_{si}$, so that the SNR at the baseband is given by $E_p/N_0 = 2E_{si}/N_0$.

Alternative Interpretations of the SNR. To have a basis for evaluating the communication performance achievable with various modulation and demodulation methods, we must carefully define our terminology for signals and noise. It is important to distinguish between two measures of signal energy. The measure of signal energy used to define γ_s in Section 7.1 is the average signal energy per channel symbol, commonly denoted by E_s . In most cases of M -ary modulation, M is a power of 2, say $M = 2^m$, and thus each M -ary channel symbol carries m bits of information. Therefore, we can define signal energy per bit as $E_b = E_s/m$ and the SNR per bit as $\gamma_b = \gamma_s/m$, which in effect normalizes the symbol energy in the channel to the individual bits in the reconstituted data stream appearing at the output of the demodulator. Both measures of signal energy are useful in making comparisons among the error rates of alternative communication techniques and systems.

Another approach is to define the SNR as the ratio of the received power to noise power in the communication channel. This method relates the SNR directly to the requirements on transmitted power. Here we are comparing modulation schemes based on the ratio of signal power S to noise power N in the transmission channel. We can relate channel S/N to the modulation parameters by assuming that the bandwidth is W and the symbol duration is T_s , which yields

$$\frac{S}{N} = \frac{E_s/T_s}{N_0 W} = \frac{R_s}{W} \gamma_s \quad (7.2.4)$$

where $R_s = 1/T_s$ is the symbol transmission rate. This equation relates the received signal-to-noise power ratio S/N to the SNR per symbol γ_s , usually employed for the calculation of error rate. Equation (7.2.4) is derived for the ideal case in which the receiver filter is matched to the transmit pulse-shaping filter. If the filters are not matched [e.g., if the transmitter uses a filter with raised-cosine frequency rolloff (discussed in Section 7.5), but the receiver uses a brick-wall filter], the required S/N for the same E_s/N_0 will change by 1 to 2 dB.

The symbol transmission rate is $R_s = R_b/m$, where R_b is the bit transmission rate of the system. Thus, defining S/N in terms of the bit rate, we have

$$\frac{S}{N} = \frac{E_b R_b}{N_0 W} = \frac{R_b}{W} \gamma_b \quad (7.2.5)$$

which relates the signal-to-noise power ratio to the SNR per bit, the spectral density of the additive noise, the data bit rate, and the occupied bandwidth. The ratio of the transmission data rate to the occupied bandwidth, $\eta = R_b/W$, is referred to as the *bandwidth efficiency* of the modulation technique. The ratio $\gamma_b = E_b/N_0$ is the SNR per bit, the quantity ordinarily used for calculation of the error rate of the system. (The quantity E_b/N_0 is sometimes referred to as the *energy contrast ratio*, because E_b and N_0 both have units of energy.) The ratio S/N is the received SNR parameter, which provides a measure of the transmitted power. The parameters usually examined in evaluations of alternative modulation techniques are (1) the required minimum transmitted power for acceptable performance, and (2) the bandwidth efficiency of the modulation technique. For a modulation technique with a bandwidth efficiency of $\eta = 1$ bit/s per hertz, we have $S/N = E_b/N_0$; that is, the signal-to-noise power ratio is the same as the ratio of the energy per bit to the spectral density of the additive noise. In the communications literature, different combinations of the measures of SNR described above are used, and the reader must have an accurate understanding of the various parameters in order to make proper comparisons among systems. For example, the power ratio S/N is generally used in the satellite communications literature and is sometimes referred to as the *carrier-to-noise ratio* (CNR or C/N). The energy ratio E_s/N_0 is the parameter ordinarily used in the literature on voiceband data communications.

Error Rate as a Performance Criterion. The standard performance criterion in digital communications is the probability of the bit error or bit error rate (BER) of a modem. Some voiceband modem applications, such as the transfer of financial data, permit error rates no greater than 10^{-5} , whereas other applications, such as digitized voice in cellular or mobile radio systems, will tolerate error rates as high as 10^{-2} to 10^{-3} . Meanwhile, high-fidelity digital audio systems (e.g., compact disk players) demand error rates on the order of 10^{-8} . From the design standpoint, for a given modulation and coding scheme there is a one-to-one correspondence between the BER and the received signal-to-noise power ratio S/N . From a user standpoint, S/N is not the favorite criterion for the performance evaluation of digital communication links, because users measure the quality of a system by the number of errors in the received bits and prefer to avoid the technical details of modulation or coding. However, using received S/N rather than BER will allow us to relate our performance criteria to the required transmitted power, which is very important for battery-operated wireless operations. For analog communications the received S/N is the usual measure of performance quality. An 18-dB S/N is typically required for analog mobile radio systems, and 30-dB S/N is expected in FM broadcasting systems. In comparing analog and digital systems, we need to translate these performance criteria into a common basis that makes S/N the convenient criterion for comparing these systems. In comparing digital systems with one another, the BER is used most of the time.

The error rate for a digital modulation technique is almost always expressed in the form of an *exponential function* or *complementary error function* (erfc), where the erfc function is defined as

$$\text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt$$

For the exponential function we have the BER or probability of bit error P_b , given by the general expression

$$P_b = Ae^{-B\gamma_b} \quad (7.2.6)$$

where A and B are given values appropriate to each specific modulation technique. The erfc function is bounded by an exponential function; as a result, for a large class of modulations we have

$$P_b = A \text{erfc}\sqrt{B\gamma_b} < Ae^{-B\gamma_b} \quad (7.2.7)$$

Figure 7.2 shows the two functions given in Eqs. (7.2.6) and (7.2.7) for $A = \frac{1}{2}$ and $B = 1$. With these parameter values, the two equations give the probability of error for DPSK and BPSK modulations, respectively, to be discussed later. The exponential approximation is an asymptotic bound providing a close approximation (less than a 1-dB error in γ) for the low error rates required in most practical applications. It is convenient if we assume that the error rate is an exponential function with parameters A and $B\gamma_b$. The error rate is related linearly to the parameter A and exponentially to $B\gamma_b$. Furthermore, we observe that the value of A varies over a limited range from one modulation scheme to another, and only order-of-magnitude changes in the error rate are considered significant. As a result, we may ignore A and compare various modulation techniques on the basis of the value of $B\gamma_b$ needed to provide an acceptable error rate. This approach will allow us to use SNR as the basis for comparing modulation methods rather than the precisely calculated error rate. For example, from Fig. 7.2 we see that either BPSK or DPSK modulation for steady signals in AWGN requires a

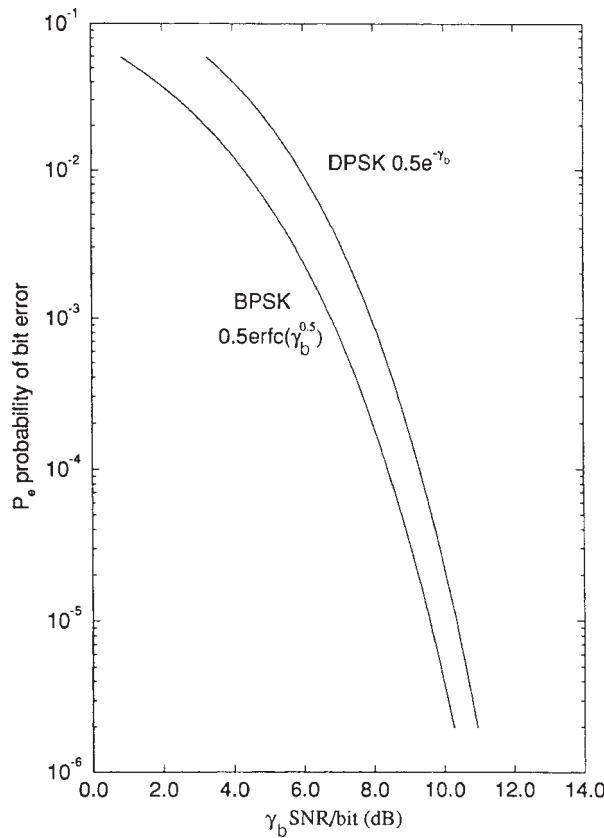


FIGURE 7.2 Comparison of the erfc function and its exponential bound.

γ_b value of around 10 dB to provide an error rate of 10^{-5} . Using SNR rather than the error rate has two advantages. First, SNR is the criterion used for assessing both digital and analog modulation techniques. Therefore, using SNR we may compare, for example, the analog Advanced Mobile Phone Service (AMPS) cellular system with the IS-54 TDMA digital cellular system. Second, SNR is related directly to the transmitted power, which is an important design parameter. As a rule of thumb, in the middle of the erfc curve each 3-dB change in γ_b will change the error rate by approximately two orders of magnitude.

In a modulation scheme based on a multiple-symbol alphabet, it is conventional to represent the symbols in a diagram known as a *signal constellation*, where the dimensions in the diagram are expressed in terms of the square root of the energy of the transmitted symbols. If we assume coherent symbol detection, the error rate in these cases is approximated by $0.5 \operatorname{erfc}(d/2\sqrt{N_0})$, where d is the *minimum distance* between the points in the constellation. To determine the error rate of a modulation technique as represented by its signal constellation, the minimum distance is expressed as a function of average energy in the constellation. Then, by substituting d into $0.5 \operatorname{erfc}(d/2\sqrt{N_0})$, one finds the error rate expressed in terms of average energy in the constellation.

In transmission on radio channels the signal is subject to fading over a wide dynamic range. In these cases the *average SNR* may be used as the performance parameter of interest. We shall see later in the chapter that the relationship between the average SNR and the average error rate over a fading channel does not follow the relatively steep exponential or erfc curves of the steady-signal AWGN channel. Therefore, the average SNR is not a good indicator of performance in fading. Instead, the average error rate provides a more meaningful performance parameter. On channels such as troposcatter or HF, where the system is subject to fading over time while the terminals are held fixed, the average error rate is defined as the average over time. In portable and mobile applications, where the error rate changes from one location to another, the average error rate is defined as the average over a range of locations. Thus, we see that the average SNR may imply either temporal or spatial averaging, depending on the system under consideration and the physical mechanisms producing signal fading.

Another important performance criterion for systems operating on fading channels is the *probability of outage*. The probability of outage is the percentage of time or locations at which the modem performance is unacceptable. The acceptable level of performance is defined by a required BER or SNR level, which we term the *performance threshold* or simply the *threshold*. In mobile applications a 1% outage probability is usually considered acceptable. This will subject the terminal to unacceptable performance in 1% of the locations in a service area.

In the remainder of this chapter we review various modulation techniques designed for operation on steady-signal additive noise channels, examine the effects of fading and multipath, and finally, describe standard modem technologies used in the portable and mobile radio industries.

7.2.2 On–Off Keying

The simplest form of carrier modulation is *on–off keying* (OOK), in which the modulator simply turns a fixed-amplitude carrier signal on or off in accordance with the value of each information bit to be transmitted. Let us say that the carrier is turned on for a

1 and off for a 0, as shown in Fig. 7.3a. Demodulation of the OOK modulated signal can be done coherently with a carrier reference, as shown in Fig. 7.3b. To represent a 1, the symbol $p(t) \cos \omega_c t$ is transmitted, where $p(t)$ is the general pulse shape and ω_c is the radian carrier frequency. For this modulation, as shown in Fig. 7.3b, the signal sampled at the output of the matched filter is given by

$$z(T) = a_i \sqrt{E_{s1}} + \varepsilon$$

where a_i can take the value 0 or 1 and

$$E_{s1} = 2E_s = \int_0^T |p(t) \cos \omega_c t|^2 dt$$

In Fig. 7.3a, $p(t)$ is a rectangular pulse spanning the symbol transmission time. Figure 7.3d shows the signal constellation for the OOK signal. In this constellation the average energy per bit over all symbols is $E_b = E_s = E_{s1}/2$. The minimum distance is expressed in terms of average energy in the constellation as $d = \sqrt{E_{s1}} = \sqrt{2E_s}$. The probability of error for the coherent implementation is $0.5 \operatorname{erfc}(d/2\sqrt{N_0})$, which yields

$$P_b = \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{E_b}{2N_0}} \right) = \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{\gamma_b}{2}} \right) < \frac{1}{2} e^{-\gamma_b/2} \quad (7.2.8)$$

The signal received can also be detected noncoherently without a carrier reference by using a simple envelope detector, as shown in Fig. 7.3c. Noncoherent reception provides a simpler implementation, but the output SNR of the noncoherent receiver is 3 dB lower than that of the coherent receiver. The bandwidth efficiency of OOK depends on the chosen pulse shape $p(t)$. For a rectangular pulse shape, if we define the transmission bandwidth as the bandwidth between the first zero crossings of the spectrum, the bandwidth efficiency is $\eta = R_b/W = 0.5$. If ideal $(\sin x)/x$ pulses (having a rectangular spectrum) are used, the bandwidth efficiency increases to $\eta = 1$.

Although this method of modulation is indeed very simple, use of the scheme poses some nontrivial problems in the design of the demodulator. For efficient reception of the OOK signal in additive noise, the demodulator must set a detection threshold

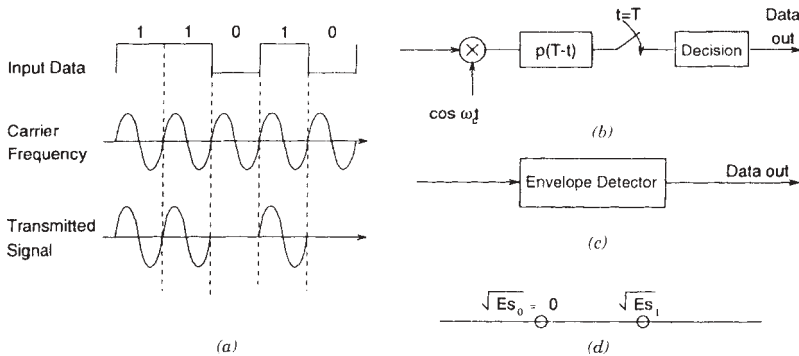


FIGURE 7.3 On-off keying: (a) modulation; (b) coherent matched filter; (c) envelope detector; (d) signal constellation.

at a level that depends on the received signal strength. Thus, in a communications environment where received signal strength can vary with time or location or both, the detection threshold must be varied accordingly. Furthermore, long strings of zeros (carrier off in Fig. 7.3a) cannot be distinguished from a no-transmission state. Finally, the BER performance of OOK modulation is poorer than is achievable with other modulation techniques that are almost as simple.

The OOK modulation method is used in certain wireless information networks (some optical WLANs in particular), where light-emitting diodes (LEDs) and photo detectors offer practical and inexpensive transmitter and noncoherent receiver implementations. The transmitted light can be thought of as a carrier that is modulated simply by turning the LED on and off. The photo detector can be thought of as a noncoherent envelope detector demodulating the transmitted signal by eliminating the optical carrier signal and detecting only the signal amplitude.

7.2.3 Frequency Shift Keying

The second simplest form of modulation is *frequency shift keying* (FSK), which uses two signal tones. In each bit interval, the modulator sends a pulse of one tone or the other in accordance with whether the information bit is 1 or 0. An FSK modulator implementation simply requires two oscillators; and it switches between oscillators in accordance with the information bit to be transmitted, as shown in Fig. 7.4a. FSK signals can be demodulated coherently by correlating the received signal over each pulse interval with the two tones, sampling the result, and selecting the larger of the two outputs. Figure 7.4b shows a coherent receiver for binary FSK, where the two symbols are represented by $s_1(t) = p(t) \cos \omega_1 t$ and $s_0(t) = p(t) \cos \omega_0 t$. The receiver consists of two branches matched to the two transmitted symbols. The sampled output of the two branches of the receiver in Fig. 7.4b are given by

$$z_0(T) = \sqrt{E_s} + \varepsilon_0, \quad z_1(T) = \sqrt{E_s} + \varepsilon_1$$

where

$$E_s = \int_0^T |p(t) \cos \omega_1 t|^2 dt = \int_0^T |p(t) \cos \omega_0 t|^2 dt$$

The occupied bandwidth and consequently, the bandwidth efficiency of the FSK scheme depend on the separation between the center frequencies of the two tones. For proper operation of the system, the two symbols must be orthogonal so that the signal intended for one detector branch does not cause interference (crosstalk) on the other branch. The orthogonality requirement is expressed mathematically as

$$\int_0^T s_1(t)s_0(t) dt = 0$$

where T is the pulse duration.

The signal constellation for FSK modulation is shown in Fig. 7.4d, where the orthogonality of the two signals is represented by placing the signals on orthogonal axes. The average energy over two symbols is given by $E_b = E_s = E_{si}$ and $d = \sqrt{2E_s}$. The relationship between average energy in the constellation and the minimum distance, and

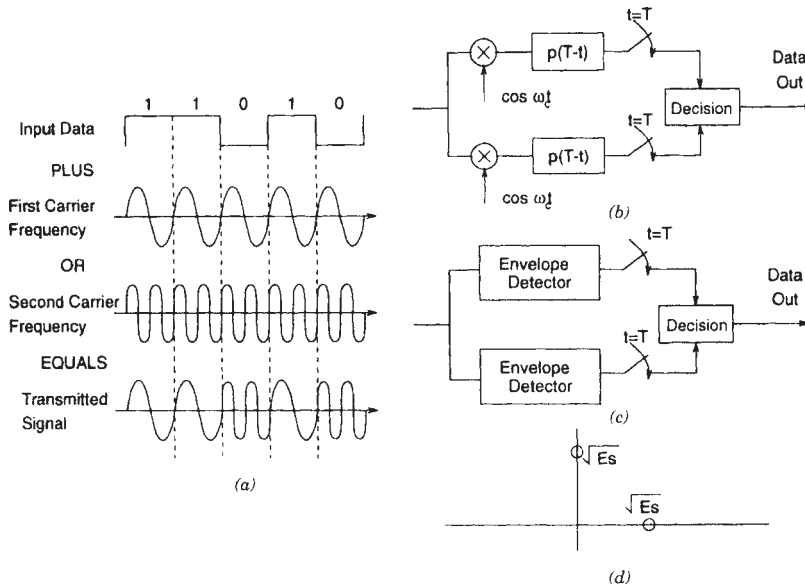


FIGURE 7.4 Binary frequency shift keying: (a) modulation; (b) coherent detection; (c) envelope detection; (d) signal constellation.

consequently, the error rate, remains the same as for OOK. The symbol decision is made by comparing the outputs of the two matched filters. In FSK reception, one of the branches contains the signal plus the additive noise, whereas the other branch contains only the additive noise. Therefore, the variance of the noise involved in the symbol decision is twice the variance of the noise in each branch.

FSK can be implemented in either a coherent or noncoherent form; the noncoherent form is shown in Fig. 7.4c. The choice between the two affects the minimum frequency spacing between the tones required to achieve orthogonality. With coherent FSK the pulses are generated and demodulated with known phases. In this case it can be shown that orthogonality is achieved if the two tones are separated by any integer multiple of $1/(2T)$ hertz, where T is the duration in seconds of each FSK pulse. However, most applications use noncoherent FSK, in which the detector operates without knowledge of the received signal phase. Thus, it is necessary that the tones be spaced by an integer multiple of $1/T$ hertz to achieve orthogonality with arbitrary signal phases. As a practical matter, the tones may be spaced at any integer multiple of the minimum orthogonal spacing, but the most efficient use of bandwidth is achieved with the minimum tone spacing applicable to either coherent or noncoherent operation. With either form of binary FSK, implemented with minimum orthogonal spacing, we can say that the signal bandwidth is approximately equal to the channel signaling rate, $1/T$, which places the bandwidth efficiency at $\eta = 1$. The noncoherent implementation of FSK suffers an SNR disadvantage of about 3 dB relative to coherent FSK.

Coherent FSK with frequency spacing $1/(2T)$ hertz is referred to as *minimum shift keying* (MSK). The MSK scheme is the most bandwidth-efficient form of FSK, and a special version of this modulation, called *Gaussian filtered MSK* (GMSK), is widely used in the portable and mobile radio industries. This topic is treated in greater detail later in the chapter.

For M -ary FSK modulation we have $M = 2^m$ orthogonal signals, where m is the number of bits per symbol. The receiver for this case consists of M parallel matched filters. If all the signals have the same energy, the average signal energy remains the same as for binary FSK, but the noise involved in making each symbol decision is M times the noise in each branch, resulting in an increase in energy per bit by the factor $M/2m$ relative to binary FSK, to maintain the same error rate. The required bandwidth is M times that of OOK, while the bandwidth efficiency is $m/(M - 1)$ times that of OOK.

The 4-ary FSK modulation format is used in many wireless applications, including WLANs at 18 to 19 GHz and digital land-mobile radios operating in VHF and UHF bands. A practical advantage of FSK is the availability of low-cost FM radios for analog voice applications such as AMPS and land-mobile radio. To modify the system to accommodate data transmission, one need only organize the data into a stream of four-level pulses and use them as an input to the FM modulator. At the receiving end, the four-level symbol stream is extracted at the output of a simple frequency-discriminator detector. This approach provides for easy integration of voice and data services in a unit having low production cost.

7.2.4 Phase Shift Keying

In binary *phase shift keying* (PSK), there is only one signal oscillator with a constant known phase, and information is conveyed in each bit interval T either by leaving the signal phase unchanged or by shifting the phase 180° relative to the oscillator phase, in accordance with the bit value to be transmitted. Binary PSK modulation, which is sometimes called *antipodal signaling*, is shown in Fig. 7.5a, where the two transmitted symbols are $\pm p(t) \cos \omega_c t$. For this modulation, as shown in Fig. 7.5b, the sampled signal at the output of the matched filter is given by

$$z(T) = a_i \sqrt{E_s} + \varepsilon$$

where $a_i = \pm 1$ and $E_s = \int_0^T |p(t) \cos \omega_c t|^2 dt$.

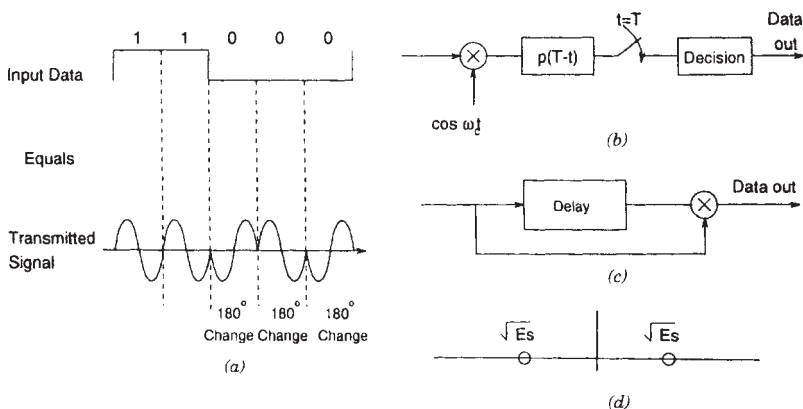


FIGURE 7.5 Phase shift keying: (a) modulation; (b) coherent matched filter detection; (c) differential detection; (d) signal constellation.

Optimum coherent detection of PSK signals is done using a matched filter followed by a sampler, as shown in Fig. 7.5*b*; the sampled output of the matched filter is compared with a zero threshold to determine the polarity of the transmitted signal. The phase reference needed is extracted from the received waveform using a phase-locked loop. The signal constellation for PSK is shown in Fig. 7.5*d*. Similar to FSK, both symbols have the same energy, and the average energy per bit is given by $E_b = E_s = E_{si}$. However, for the same average energy in the constellation, the minimum distance is $d = 2\sqrt{E_s}$ and the squared minimum distance is twice that of coherent FSK or OOK. This increase in normalized distance gives a 3-dB SNR advantage to PSK over FSK or OOK. In fact, it can be shown that the binary PSK format is the optimum binary signal set for communication in AWGN [Woz65]. Noncoherent implementation of PSK involves differential modulation at the transmitter and differential demodulation at the receiver, and it is referred to as *differential PSK* (DPSK). The differential demodulator is shown in Fig. 7.5*c*. DPSK modulation suffers a 1- to 2-dB performance disadvantage relative to PSK at the levels of error rate required for most system applications. As with OOK and binary FSK, the bandwidth of a PSK signal is roughly $1/T$, where T is the PSK symbol duration, which results in a bandwidth efficiency of $\eta = 1$.

BPSK modulation is the building block for multi-amplitude/phase modulation and coding techniques that are used in the most sophisticated voiceband modems. DPSK modulation is the building block for many of the radio modems designed to operate in harsh multipath fading environments.

7.2.5 Pulse Amplitude Modulation

Next we examine a form of nonbinary digital modulation, called *pulse amplitude modulation* (PAM), which can be viewed as an extension of binary PSK. This modulation technique was developed for use in telephone carrier systems. PAM uses one signal oscillator with known fixed phase but allows the transmitted signal amplitude to have any of a set of discrete values (levels) $\{a_i\}$, where $i = 1, 2, 3, \dots, M$ and $M = 2^m$, with m being the number of bits encoded into a symbol. In a PAM transmitter, shown in the upper half of Fig. 7.6*a*, information bits are buffered and encoded into a stream of pulse amplitudes $\{a_i\}$. If, for example, $M = 4$, two information bits at a time are

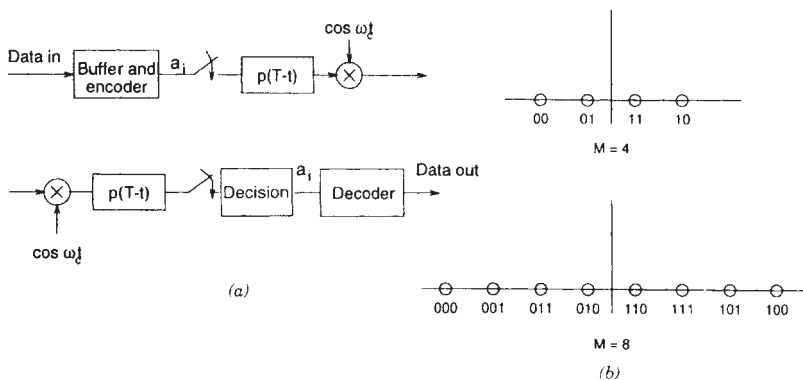


FIGURE 7.6 Pulse amplitude modulation: (a) transmitter and receiver; (b) signal constellations for $M = 4$ and $M = 8$.

buffered and encoded into one of four amplitude levels. If $M = 8$, three bits at a time are buffered and encoded into one of eight levels, and so on. The encoding process is most easily done using a simple look-up table. The encoded amplitude is then applied to a fixed pulse shape, which we denote as $p(t)$. The amplitude-modulated pulse is next multiplied by the carrier signal $\cos \omega_c t$ and transmitted on the channel. The transmitted symbol in each symbol interval is given by $a_i p(t) \cos \omega_c t$, where one amplitude-modulated pulse is transmitted every T seconds. The choice of pulse shape is very important in the design of a PAM system, and we will say more about this when we discuss the demodulator. Note that because m information bits are encoded onto each transmitted pulse, the symbol rate in the channel is lower than the source information bit rate by the factor m , and thus the bandwidth efficiency is $\eta = m$. Because the bandwidth of the transmitted signal is determined by the pulse-shaping filter, regardless of the number of pulse amplitudes, PAM provides a very effective way of increasing the transmitted data rate within a fixed bandwidth.

In PAM, the pulse amplitudes are chosen with uniform spacing and are arranged symmetrically about zero. Figure 7.6b illustrates the signal constellation for the cases $M = 4$ and $M = 8$, with allowed pulse amplitudes denoted by circles on the horizontal axis. It should be clear that if $M = 2$, this is equivalent to binary PSK modulation. The relationship between the minimum distance and the average energy per symbol in this constellation is given by

$$d^2 = \frac{12}{4^m - 1} E_s^m \quad (7.2.9)$$

where m is the number of bits per symbol and E_s^m is the average energy per symbol for a constellation with 2^m symbols. The average energy in the constellation can be written in the following recursive form:

$$E_s^{m+1} = 4E_s^m + \frac{1}{3}d^2$$

In other words, the transmitted power must be increased by a factor of 4 (6 dB) to compensate for the performance degradation caused by sending an additional bit per symbol.

Example 7.1: Performance of a 4-PAM Modem Figure 7.7 shows the signal constellation for a 4-PAM modem with $m = 2$ and $M = 4$. In this modem the received sampled signal at the output of the matched filter is given by

$$z(T) = a_i \sqrt{E} + \varepsilon$$

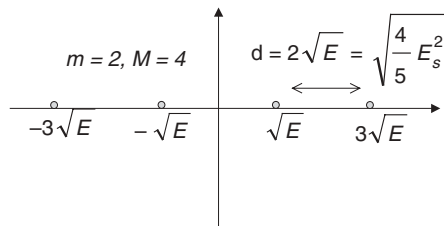


FIGURE 7.7 4-PAM constellation.

where $a_i = \pm 1, \pm 3$ and E is an auxiliary parameter facilitating representation of the samples received. From Eq. (7.2.9) the relationship between the minimum distance and the average energy in the constellation, E_s^2 , is given by

$$d^2 = \frac{12}{4^2 - 1} E_s^2 = \frac{4}{5} E_s^2$$

The probability of symbol error can be approximated by the relationship between the minimum distance in the constellation and the variance of the additive noise:

$$P_s \approx 0.5 \operatorname{erfc} \left(\frac{d}{2\sqrt{N_0}} \right) = 0.5 \operatorname{erfc} \left(\sqrt{\frac{E_s^2}{5N_0}} \right)$$

At the receiver, shown in the lower part of Fig. 7.6a, the received signal is multiplied by the carrier signal $\cos \omega_c t$ and passed through a filter matched to the transmitted pulse shape. The multiplication by $\cos \omega_c t$ produces two realizations of the received signal: One is centered about zero frequency, which is the baseband signal, and the other is centered about $2f_c$. We want the baseband signal, which is recovered by using the transmitter pulse shape $p(t)$ as the matched filter. Given that we sample at the optimum time, the output is the transmitted amplitude a_i . The amplitude detected is then decoded to the appropriate set of information bits, which can be done with the decoding version of the encoding look-up table.

7.2.6 Quadrature Amplitude Modulation

In the preceding discussion we saw that the use of M -level PAM provides bandwidth efficiency proportional to m bits per channel symbol. We now describe a modulation scheme that doubles the bandwidth efficiency of PAM simply by applying the same amplitude levels on both the sine and cosine of the carrier, producing a transmitted signal of the form

$$s(t) = a_i p(t) \cos \omega_c t + b_i p(t) \sin \omega_c t \quad (7.2.10)$$

Because the transmitted signal consists of two PAM pulse streams in phase quadrature, this modulation scheme is called *quadrature amplitude modulation* (QAM). This form of modulation was first used in a 9600-bit/s commercial modem introduced to the market in the early 1970s [Pah88c].

Because the quadrature channels are orthogonal, the modulator can be designed with two signal branches, each configured exactly as in a PAM modem, one channel modulating the cosine of the carrier, the other the sine. At the receiver, the two channels are prevented from interfering with one another by their orthogonality. The data rate for the QAM modem is simply the sum of the data rates on the two channels, but the signal bandwidth, which is determined by the pulse shape $p(t)$, is unchanged from the single-channel PAM signal. Thus, the bandwidth efficiency of the QAM design is twice that of the PAM design and is given by $\eta = m = 2n$, where $m = 2n$ is the number of bits for each point in the constellation and n is the number of bits in each dimension.

Note that if we use antipodal amplitude values on each branch, we have two binary PSK signals in phase quadrature. This modulation, called *quadrature phase shift keying* (QPSK), is commonly used, with a number of variations, in digital radio systems. QAM

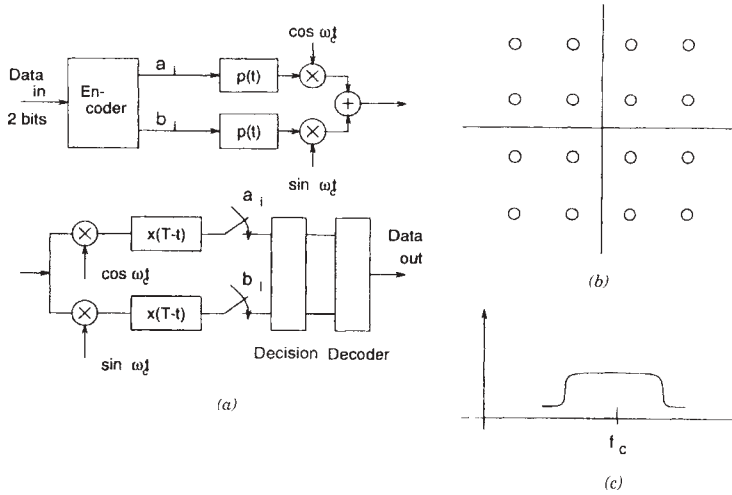


FIGURE 7.8 Quadrature amplitude modulation: (a) transmitter and receiver; (b) signal constellation for $M = 16$; (c) power spectrum.

modulation is the predominant modulation technique in use in voiceband modems, and it is expected eventually to find adoption in the radio communications industry. There are a number of other modulation schemes that are implemented with versions of the two-branch modem structure. Important examples are minimum shift keying (MSK), Gaussian low-pass filtered MSK (GMSK), and $\pi/4$ -shift QPSK, each of which is characterized by a particular form of pulse shaping used on the quadrature branches.

Figure 7.8a shows a block diagram of a QAM modulator and demodulator. Figure 7.8b shows an example of 16-ary QAM signal constellations that transmits two bits on each branch in each symbol interval, for a total of four bits per symbol interval. The voiceband modem mentioned in the preceding paragraph uses the 16-ary QAM signal constellation and a symbol transmission rate of 2400 baud, yielding a data rate of 9600 bits/s. The relationship between the minimum distance and the average energy over all QAM symbols is given by

$$d^2 = \frac{6}{2^m - 1} E_s^m \quad (7.2.11)$$

This equation leads to the recursive equation

$$E_s^{m+1} = 2E_s^m + \frac{d^2}{6}$$

indicating a requirement of twofold (3 dB) additional power for transmitting one additional bit in the constellation. This is 3 dB better than the 6 dB per added bit required with PAM.

Example 7.2: Performance of a 16-QAM Modem Figure 7.9 shows the signal constellation of a 16-QAM modem with $m = 4$ and $M = 16$. In this modem the sampled signal received at the output of the matched filter is given by

$$z(T) = c_i \sqrt{E} + \varepsilon$$

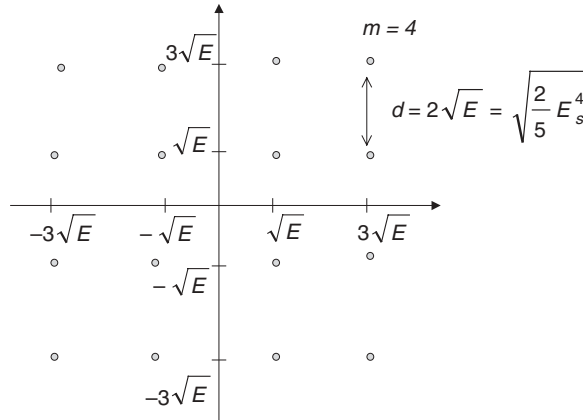


FIGURE 7.9 16-QAM constellation.

where $c_i = a_i + jb_i$ is the two-dimensional information transmitted, expressed in complex notation; $a_i, b_i = \pm 1, \pm 3$ are in-phase and quadrature-transmitted information; and E is an auxiliary parameter that facilitates representation of the samples received. From Eq. (7.2.11), the relationship between the minimum distance and the average energy in the constellation, E_s^2 , is given by

$$d^2 = \frac{6}{2^4 - 1} E_s^4 = \frac{2}{5} E_s^4$$

The probability of symbol error can be approximated by the relationship between the minimum distance in the constellation and the variance of the additive noise:

$$P_s \approx 0.5 \operatorname{erfc} \left(\frac{d}{2\sqrt{N_0}} \right) = 0.5 \operatorname{erfc} \left(\sqrt{\frac{E_s^4}{10N_0}} \right).$$

In general, if we use Eq. (7.2.11), we have

$$P_s \approx 0.5 \operatorname{erfc} \left(\frac{d}{2\sqrt{N_0}} \right) = 0.5 \operatorname{erfc} \left(\sqrt{\frac{3E_s^m}{2(2^m - 1)N_0}} \right)$$

As we will see later, a more accurate estimate replaces 0.5 with M . However, as we mentioned earlier, we are not as concerned about the effect of the linear multiplier on the overall error rate as we are about the orders-of-magnitude differences caused by changes in the signal-to-noise ratio.

In the implementation of digital modems, the real and imaginary parts of the symbols transmitted are modulated onto sine and cosine functions so as to achieve orthogonality. When the orthogonal channels are transmitted simultaneously, the peak in transmit power occurs at the peak of the pulse-shaping filter. To maximize power efficiency, it is desirable in many radio communication channels to transmit at full power, but at the same time, power amplifiers exhibit increasing nonlinearity as they are driven

near their peak power limits. To deal with this problem, the drive level to the power amplifier is adjusted to keep the peak signal power at a specified margin relative to the peak power limit of the amplifier. Consequently, the average power is kept at an even lower level, which reduces the overall power efficiency. Therefore, in many cases the modulation is staggered so that the transmitted pulses for the real and imaginary parts of the signals have a relative time delay of $T/2$ seconds. Staggering reduces the peak-to-average power ratio, allowing the average power to be set closer to the nonlinear range of the amplifier, achieving better overall power efficiency.

7.2.7 Multiphase Modulation

M -phase PSK modulation can also be implemented with a two-branch structure, when $M = 2^m$, with m the number of bits per symbol. All the M-PSK symbols transmitted have the same energy $E_s = mE_b$, and the average energy per symbol is the same as the energy of any individual symbol. As a result, the signals in the constellation are located on a circle with radius $\sqrt{E_s}$. As shown in Fig. 7.10, the minimum distance for M-PSK modulation is given by

$$d = 2\sqrt{E_s} \sin \frac{\pi}{M} \quad (7.2.12)$$

and

$$P_s \approx 0.5 \operatorname{erfc} \left(\sqrt{\frac{E_s}{N_0} \sin \frac{\pi}{M}} \right)$$

For $M = 2$ and $M = 4$ we have BPSK and QPSK (4-QAM), respectively. An 8-ary PSK modulator can be structured as two quadrature branches, with three amplitude levels: $0, \pm\sqrt{E_s/2}$, and $\pm\sqrt{E_s}$.

Because all symbols have the same amplitude, PSK modulation is less sensitive to nonlinearities in the channel. As a result, PSK is widely used on power-limited radio channels such as satellite channels, where the amplifiers are driven close to their nonlinear regions of operation in order to maximize power efficiency.

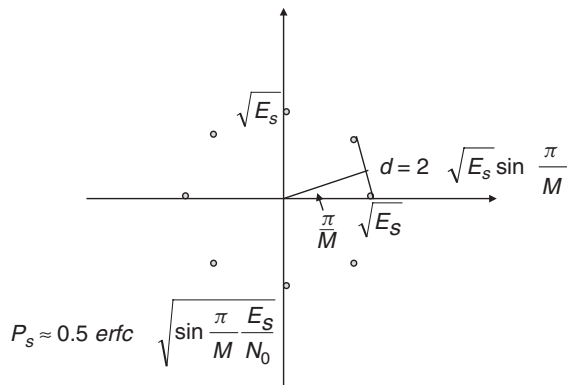


FIGURE 7.10 Minimum distance and symbol error rate in M-PSK modulation.

7.2.8 Partial-Response Signaling

The transmission of two symbols per second per hertz with QAM requires ideal pulse-shaping filters: that is, filters that eliminate intersymbol interference completely at the sampling instants. If we can remove the constraint of having no intersymbol interference at sampling instants, the system can be designed with physically realizable filters, which are generally easier to implement than are ideal pulse-shaping filters. Signaling techniques that allow symbol transmission at a rate equal to two symbols per hertz of bandwidth, with controlled amounts of intersymbol interference, are called *partial-response signaling techniques*. Partial response signaling was introduced in the 1960s for use in wireline modems and has sometimes been applied in radio modems. With partial-response signaling, the bandwidth efficiency of ideal QAM is achieved with a realizable filter. As shown in Section 7.5.3, the pulse shape for partial response signaling is designed so that the information content of one transmitted symbol is distributed over two sample intervals. The partial-response form of the filter allows additional noise into the system, which reduces the SNR per bit by a factor of $(\pi/4)^2 = 2.1$ dB. Otherwise, the error rates for one- or two-dimensional partial-response signals are given by the same equations as for PAM and QAM signaling. However, the transmitted waveforms are different, and as we will see later, the performance analysis for frequency-selective fading channels is different from that of PAM or QAM. For more detailed discussions of partial response signaling, see [Luc68, Feh87, Pro01]. As shown in [Bel84, Pah85a], the performance of quadrature partial response (QPR) and staggered QPR (SQPR) signaling over frequency-selective fading channels is inferior to those of QPSK and staggered QPSK (SQPSK).

7.2.9 Trellis-Coded Modulation

In classical communication systems, error control is provided by coding the input data bits and then modulating a carrier with the coded signal. To keep the data rate unchanged, one should compensate for the error-correction parity bits by increasing the transmission rate. In bandlimited channels such as voiceband channels, an increase in transmission rate requires an increase in the number of points in the constellation, resulting in a higher symbol error rate. For many years it was believed that if the data rate remained the same, practical error-control codes could not compensate for the performance loss caused by increasing the number of points in the signal constellation. As a result, coding techniques were not employed in voiceband modems.

About two decades ago, renewed attention was given to the concept of coding for bandlimited channels, spurred by the development of a combined modulation and coding technique now referred to as *trellis-coded modulation* (TCM) [Ung82, Ung87]. The principal advantage of TCM over modulation schemes used with traditional error-correction coding is its ability to achieve improved power efficiency without the customary bandwidth expansion introduced by the use of coding. Various versions of TCM can improve the performance of a modem by 3 to 6 dB on steady-signal channels. The eight-state trellis code with a nominal gain of 4 dB is perhaps the most attractive, because more complex trellis codes offer little additional improvement but with extensive additional implementation complexity. A version of TCM that can resolve 90° phase ambiguity [Wei84a, Wei84b] has been adopted by CCITT as a standard for QAM voiceband modems [CCI84a]. A comprehensive treatment of coded and uncoded signal constellations for bandlimited channels is available in [For84].

A historical overview of the development of the TCM for wireline modems is available in [Pah88c]. Standard coded and uncoded QAM signal constellations can be modified to improve the performance of a wireline modem in the presence of “nonuniform” noises such as those arising from phase jitter or nonlinear quantization [Pah91].

The TCM technique is an extension of QAM in which the number of points in the constellation is increased to create redundancy. These extra symbols enable the transmitter to create dependency between successive transmitted symbols, and in this way, only certain sequences of symbols are valid. The sequence of symbols received is compared with all valid sequences, and the sequence with maximum likelihood is chosen. The efficient search method under the maximum likelihood criterion is the Viterbi algorithm [Vit67]. Implementation of the Viterbi algorithm for TCM is computationally complex. The CCITT-recommended TCM technique almost doubles the processing power required for the implementation of a modem.

In addition to its applications in wireline modems, TCM has also been studied for application to fading channels, particularly mobile satellite channels [Div87, Div88a, Div88b, McL88, Moh89, Sch89, Big91]. An important point to be noted in the application of TCM to fading channels is that the criteria for designing optimum trellis codes for fading channels are different from the design criteria for steady-signal AWGN channels. This point is discussed in detail in [Big91].

7.2.10 Comparison of Modulation Methods

The probabilities of bit error for the most common binary modulation techniques are given by

$$\begin{aligned}
 \text{FSK or OOK-CD :} \quad P_b &= \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{\gamma_b}{2}} \right) \\
 \text{BPSK-CD :} \quad P_b &= \frac{1}{2} \operatorname{erfc} (\sqrt{\gamma_b}) \\
 \text{DPSK-NCD :} \quad P_b &= \frac{1}{2} e^{-\gamma_b} \\
 \text{FSK-NCD :} \quad P_b &= \frac{1}{2} e^{-\gamma_b/2}
 \end{aligned} \tag{7.2.13}$$

where CD denotes coherent demodulation, NCD denotes noncoherent demodulation, and in each case steady-signal reception in AWGN is assumed. The four formulas in Eq. (7.2.13) are plotted in Fig. 7.11. As shown in the figure, the best BER performance is achieved with coherent BPSK. The BER performance achieved by coherent FSK (or OOK) is exactly 3 dB poorer than coherent PSK, simply reflecting the doubled noise level associated with the detection of two orthogonal signals in FSK, as contrasted with the detection of antipodal signals in PSK.

It can be seen in Fig. 7.11 that DPSK provides a somewhat poorer performance than does coherent PSK, but at high SNR values the curves are very close together. The relationship between the PSK and DPSK curves is in fact given by the analytical bound shown in Eq. (7.2.7). A simple heuristic explanation of the nearly identical performance of PSK and DPSK at high SNR values is as follows: Whereas a PSK pulse is demodulated with an ideal noiseless phase reference, each DPSK pulse is in effect demodulated using the (noisy) previous pulse as its phase reference. As the SNR

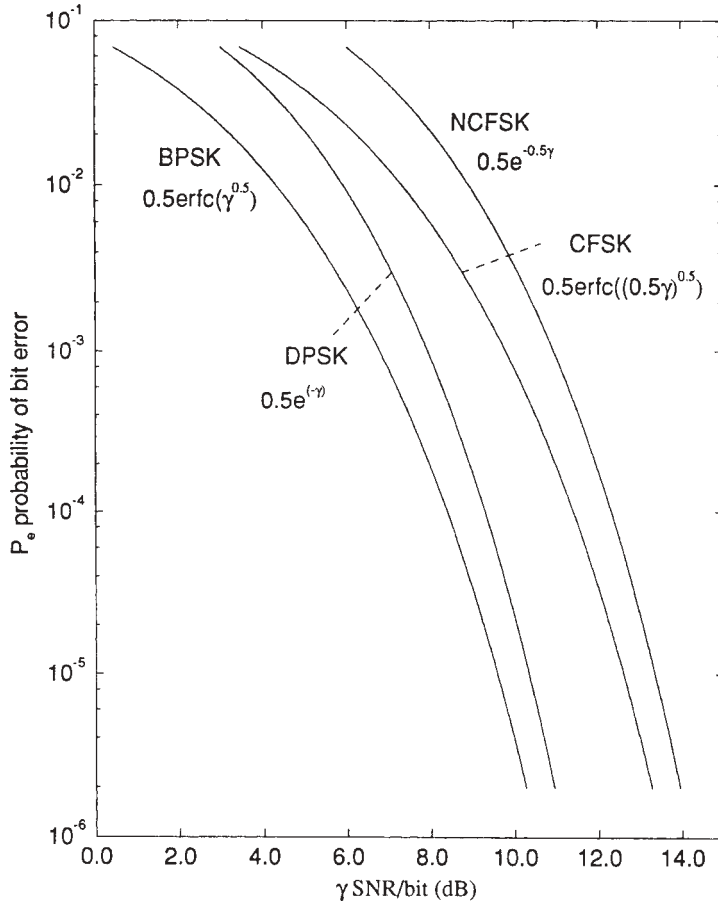


FIGURE 7.11 Comparison of error rates for four binary modulation techniques.

increases, the previous pulse becomes steadily less noisy and thus becomes more like the ideal noiseless phase reference. Noncoherent FSK is the highest of the four BER curves in Fig. 7.11, and it is in fact exactly 3 dB poorer than DPSK. (It is left as an exercise for the reader to explain the exact 3-dB difference between FSK and DPSK BER performance.)

For nonbinary signal constellations, we assume that $M = 2^m$ symbols, with m the number of bits conveyed in each symbol. The approximate equations for the probability of symbol error with coherent detection in AWGN are

$$\begin{aligned}
 \text{M-PSK :} \quad P_s &\simeq \text{erfc} \left(\sqrt{\sin^2 \frac{\pi}{M} m \gamma_b} \right) \\
 \text{M-FSK :} \quad P_s &\simeq M \text{erfc} \left(\sqrt{\frac{m \gamma_b}{M}} \right) \\
 \text{M-QAM :} \quad P_s &\simeq 2 \text{erfc} \left(\sqrt{\frac{3}{2(2^m - 1)} m \gamma_b} \right)
 \end{aligned} \tag{7.2.14}$$

The bandwidth efficiency of M-PSK and QAM for pulse shaping with ideal $(\sin x)/x$ pulses is $\eta = m$, whereas the bandwidth efficiency of M-FSK with tones spaced $1/T$ hertz apart is $\eta = m/M - 1$.

Writing Eqs. (7.2.14) in terms of the signal-to-noise power ratio S/N , we have

$$\begin{aligned} \text{M-PSK :} \quad P_s &\simeq \text{erfc} \left(\sqrt{\sin^2 \frac{\pi}{M} \frac{S}{N}} \right) \\ \text{M-FSK :} \quad P_s &\simeq M \text{erfc} \left(\sqrt{\frac{M-1}{M} \frac{S}{N}} \right) \\ \text{M-QAM :} \quad P_s &\simeq 2 \text{erfc} \left(\sqrt{\frac{3}{2(M-1)} \frac{S}{N}} \right) \end{aligned} \quad (7.2.15)$$

For M-PSK and M-QAM modulations, the bit error probability calculation depends on the encoding scheme for the symbols in the constellation: that is, the mapping of information bits onto modulation symbols. If the symbols are Gray-coded, each symbol error that is a transition to an adjacent symbol in the constellation causes only one bit error. Thus, given a reasonably high SNR, we may assume that the bit error probability is m times smaller than the symbol error probability. Generally, m is a small number and the symbol error rate provides a reasonable approximation to the bit error rate. For M-FSK modulation, the symbol error probability can be converted to a bit error probability in the corresponding m -bit groups by assuming that when an M -ary symbol is in error, each of the $2^m - 1$ incorrect symbols is equally likely. Then each bit in the erroneous symbol has $2^m - 1$ chances out of the $M - 1$ possibilities to be in error. This leads to the relationship

$$P_b = \frac{2^m - 1}{M - 1} P_s$$

Curves of P_s versus γ_b for M -ary PSK with coherent demodulation are shown in Fig. 7.12. In this presentation it can be seen that at very low values of symbol error probability P_s , binary and 4-ary PSK operate at essentially the same levels of E_b/N_0 . In the approximation given by Eq. (7.2.14), the probability of symbol error versus γ_b is the same for binary PSK (BPSK) and 4-ary or quadrature PSK (QPSK). If we were to plot the exact bit error probability P_b instead of P_s , the BPSK and QPSK curves would be identical, because coherent QPSK is equivalent to two orthogonal BPSK channels. This means that QPSK allows us to double the data rate of BPSK with no increase in bandwidth and no penalty in communication efficiency. However, as the M -ary phase constellation changes from four to eight phases, there is a loss of communication efficiency of nearly 4 dB. For each further doubling of the PSK signal constellation, there is a steady growth in the corresponding loss in communication efficiency.

Figure 7.13 shows symbol error probability versus γ_b for a selection of M -ary PSK and M -ary QAM modulation systems. Note that 4-ary PSK and 4-ary QAM give exactly the same symbol error probability performance; this is to be expected, because they are really the same scheme, as we pointed out earlier. Note, however, that as M values are increased, the QAM constellations have better communication efficiency than that of the PSK constellations for the same number M of signal points. The reason for this is the relatively more efficient “packing” of signal points in a rectangular QAM constellation.

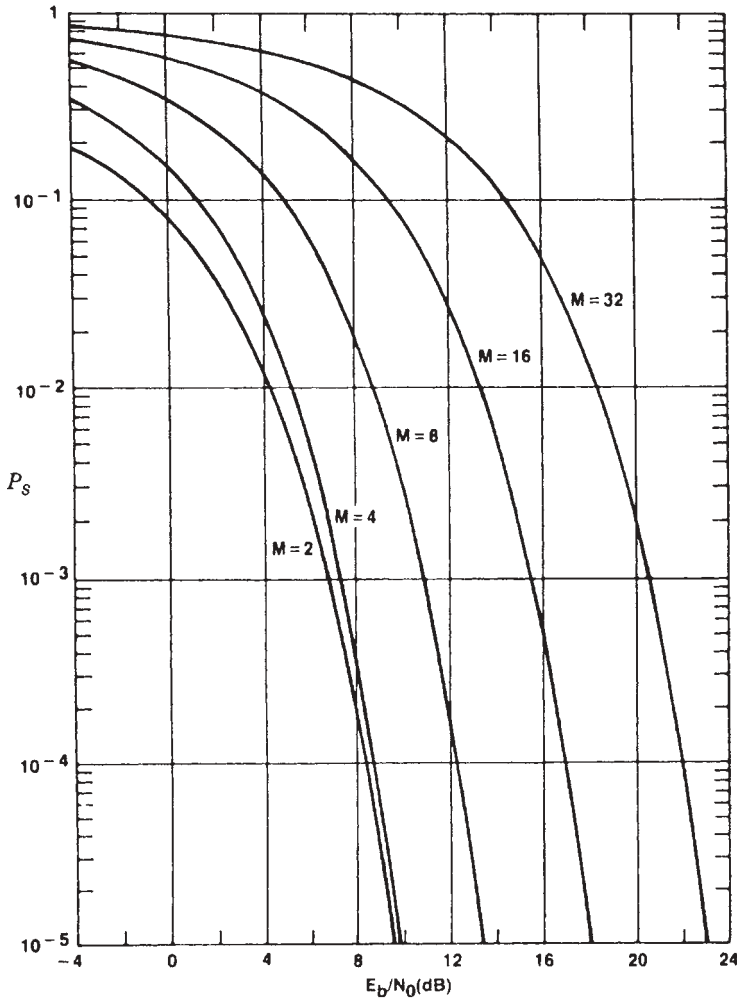


FIGURE 7.12 Symbol error probability versus E_b/N_0 for coherent demodulation of PSK and M -ary PSK. $E_b/N_0 = (S/N)(W/R_b)$.

For a thorough treatment of the effects of various packing strategies on the performance of the multidimensional modulation techniques, the reader is referred to [For84].

Next, let us compare performance curves on the basis of the signal-to-noise power ratio, S/N , where we recall from Section 7.2.1 that S/N is related to γ_b , the SNR per bit, by $S/N = (R_b/W)\gamma_b$, with W the signal bandwidth and R_b the data rate in bits per second. As we noted in Section 7.2.1, the signal-to-noise power ratio required is a measure of the received power before processing and is directly related to the power requirements for the transmitter.

Figure 7.14 shows a set of performance curves calculated for a selection of modulation schemes operating on a steady-signal channel with AWGN. The curves show $\log_{10}$ of the probability of symbol error P_s as a function of received signal-to-noise power ratio, S/N , for BPSK, QPSK, M -ary PSK for M up to 64, QAM for M up

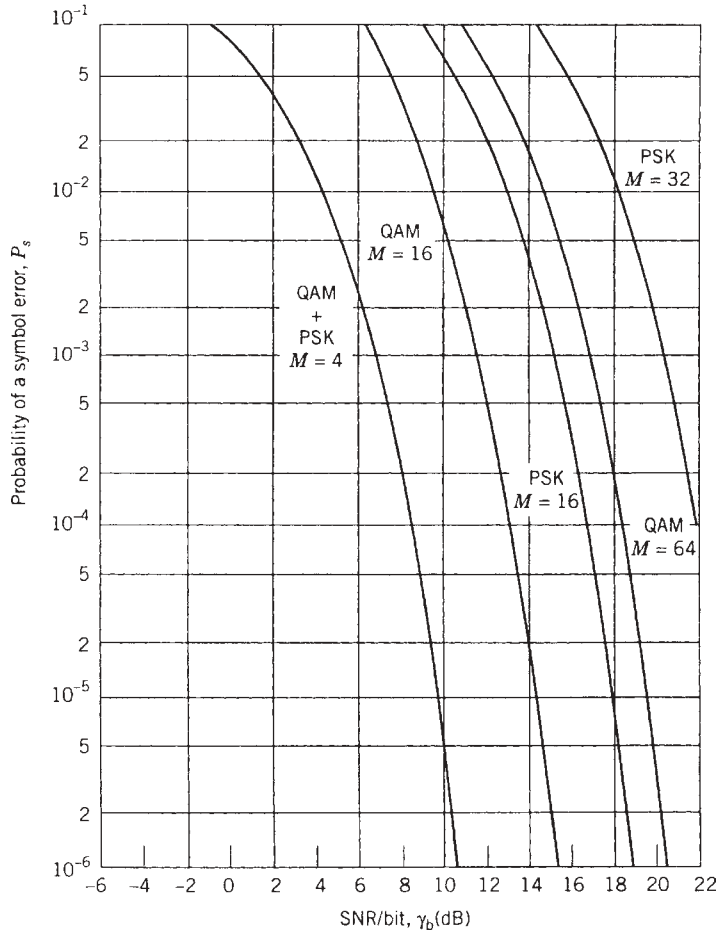


FIGURE 7.13 Comparison of symbol error probabilities versus SNR per bit for M -ary PSK and M -ary QAM signal constellations.

to 1024, and a selection of N -ary quadrature partial response (QPRS) systems for N up to $63 \times 63 = 3969$. To compare the performances, let us consider a BER value of 10^{-5} for each modulation scheme. For nonbinary modulations, although Fig. 7.14 gives probability of symbol error, we can assume the use of Gray coding, as discussed above, and use the approximation for the bit error probability $P_b = (1/m)P_s$, where $m = \log_2 M$ and P_s is the M -ary symbol error probability.

Let us compare BPSK with QPSK. For BPSK the S/N required at 10^{-5} is somewhat less than 10 dB, whereas QPSK requires just less than 13 dB. A comparison of the SNR required per symbol is really a comparison of the signal power required, if we assume that the background noise is the same in each case. Therefore, we see that QPSK requires 3 dB, twice the signal power required by BPSK, to achieve the same error rate in the same noise background. This is exactly what we should expect when we recall the earlier description of QPSK as two BPSK channels operating on orthogonal phases of the same carrier. For the higher-order M -ary PSK and M -ary QAM curves, it can be seen that increasing signal power is required as the number of points in the

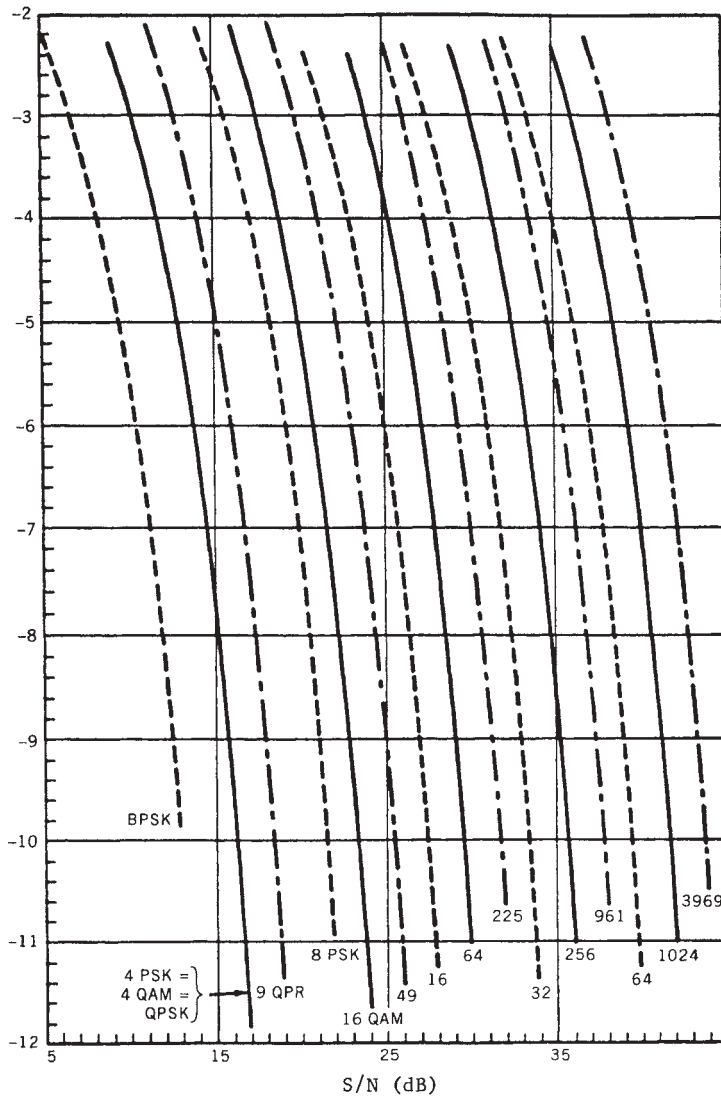


FIGURE 7.14 Symbol error probability versus S/N for coherent demodulation of M -ary PSK (---), M -ary QAM (—), and N -ary QPRS (—). (From [Kuc85] © IEEE.)

signal constellation grows. In particular, observe that as we discussed in Section 7.2.6, we pay a 3-dB penalty for each additional bit contained in a QAM signal.

Example 7.3: Modulation in IEEE 802.11g Assume that we have an IEEE 802.11g WLAN device with a receiver noise level of -80 dBm, operating in an open area with a distance power gradient of 2.0. The 802.11g standard provides several modulation options, including 64-QAM, 16-QAM, and QPSK. From Fig 7.14 the SNR value required to support a symbol error rate of 10^{-5} for a 64-QAM modem system is around 26 dB; thus, the received signal strength required is -54 dBm. If we reduce the points in the constellation to 16-QAM, the signal strength requirement drops to

about -48 dBm. In an open area where the distance power gradient is 2.0, as we saw in Chapter 3, the path loss is 6 dB per octave of distance. Therefore, reducing the modulation from 64- to 16-QAM, the coverage of the modem doubles. If we continue by lowering the number of points in the constellation to 4, with QPSK modulation, we can achieve double the coverage of 16-QAM and quadruple the coverage of 64-QAM. By reducing the points in the constellation from 64 (6 bits per symbol) to 16 (4 bits per symbol) and QPSK (2 bits per symbol), the data rate decreases to two-thirds and one-third, respectively, of the data rate achieved with 64-QAM. In designing systems for radio channels, we must always compromise on the data rate to increase the coverage, and vice versa. Another approach to this compromise is to reduce the error-rate requirement by adding parity-check coding to the transmitted symbols. For example, if we go from 10^{-5} to 10^{-3} as the required error rate, as shown in Fig. 7.14, the SNR requirement drops almost another 3 dB, which can help to increase the coverage. Adding coding in its turn reduces the effective data rate of the system. The effects of coding on the performance of radio modems are discussed in Section 8.4.

Figure 7.15 shows curves of P_s versus SNR for M -ary FSK with coherent demodulation. Here we see that the performance becomes poorer as the number of tones increases. Increasing the number of FSK tones does not affect the minimum distance between signals, because the tones are always chosen to be orthogonal over the symbol interval T . But increasing M does increase the number of “competitors” for the correct signal in the demodulation process, and this is exhibited in the multiplicative factor M seen in the M -FSK formula in Eq. (7.2.15).

7.3 THEORETICAL LIMITS AND PRACTICAL IMPAIRMENTS

In this section we review the theoretical limits on achievable data rate and communication efficiency as defined by the Shannon capacity for steady-signal Gaussian noise channels. We then review briefly the real-world channel impairments encountered in practical situations.

7.3.1 Theoretical Limits of Communication Performance

In our discussions thus far we have shown how the design of the modem signal constellation relates to the achievable data rate and the energy efficiency of the modem design. Therefore, it is useful to consider the ultimate limits on data rate and efficiency that are theoretically achievable. This is best done by examining Shannon’s well-known formula for the capacity of a bandlimited continuous AWGN channel:

$$C = W \log_2(1 + S/N) \quad \text{bits/s}$$

where C is the maximum achievable information transfer rate of the channel, W is the channel bandwidth in hertz, and S/N is the signal-to-noise power ratio in the bandwidth [Sha48]. Stated succinctly, the essence of Shannon’s work on channel capacity is as follows: “If we take increasingly long sequences of source information bits and map them into correspondingly long transmission waveforms, the error rate in the delivered data can be brought arbitrarily close to zero, as long as we do not attempt to transmit data at a rate higher than C . Therefore, at any nonzero level of channel

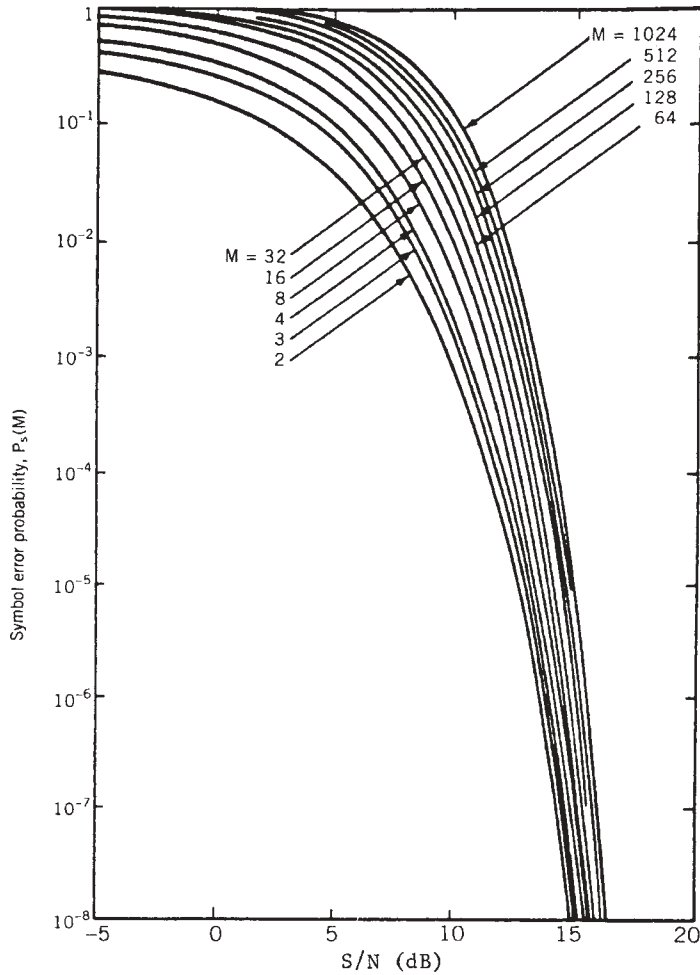


FIGURE 7.15 Symbol error probability versus. S/N for coherent M -ary FSK signaling.

signal-to-noise ratio S/N , there is some nonzero information transfer rate below which arbitrarily accurate communication can in principle be achieved.” The significance of Shannon’s result, which is called the *channel coding theorem*, is that channel noise does not inherently limit the accuracy with which communication can be achieved, only the rate at which information can reliably be transmitted [Sha49, Sha59].

We can readily apply the capacity formula to the case of a voiceband telephone channel, which has a bandwidth of 3 to 4 kHz and a typical S/N value of 28 dB for conditioned leased lines. This yields a theoretical channel capacity of about 30 kb/s. It will therefore be useful to determine how closely this limit can be approached with various signal constellations. In comparing various modulation schemes with the Shannon limit, it is instructive to rewrite the capacity formula in the form

$$\frac{C}{W} = \log_2 \left(1 + \frac{S}{N_0 C} \frac{C}{W} \right) \quad \text{bits/s per hertz}$$

where N_0 is the one-sided power spectral density of the white Gaussian noise. We can further rewrite the formula as

$$\frac{S}{N_0 C} = \frac{W}{C} (2^{C/W} - 1) \quad (7.3.1)$$

For transmission at capacity, the signal power is $S = C E_{b_{\min}}$ and the left side of Eq. (7.3.1) becomes

$$\frac{S}{N_0 C} = \frac{E_{b_{\min}}}{N_0}$$

where $E_{b_{\min}}$ is the minimum transmitted energy per source information bit required for reliable communication. Finally, we rewrite Eq. (7.3.1) as

$$\frac{E_{b_{\min}}}{N_0} = \frac{W}{C} (2^{C/W} - 1) \quad (7.3.2)$$

This equation describes channel capacity in terms of two convenient normalized parameters, $E_{b_{\min}}/N_0$ and C/W . The first parameter is the minimum value of SNR per source information bit required for reliable transmission of data at capacity over an AWGN channel of bandwidth W . The second parameter, C/W , simply normalizes the channel capacity to an arbitrary bandwidth and represents the maximum achievable value of bandwidth efficiency; its reciprocal, W/C , is the bandwidth expansion factor for operation at capacity. Therefore, Eq. (7.3.2) expresses channel capacity in terms of two parameters defining the achievable limits of communications efficiency as measured by SNR per bit and bandwidth utilization. This now provides us with a convenient framework for assessing the communications efficiency of any modulation scheme chosen.

In Fig. 7.16 we show the capacity formula as a plot of R/W versus E_b/N_0 , where R is the information rate in the channel, with $R = C$ at channel capacity. Note that the lower portion of the scale is expanded for convenience in drawing the figure. This figure essentially represents a bandwidth versus efficiency plane, and the capacity curve divides the plane into two regions. The shaded area to the left of the curve defines the region in which reliable communication cannot be achieved; that is, no modulation or coding scheme can be devised to operate in that region with low BER in the data delivered. In the right-hand area of the figure, which defines the region of achievable signal designs, design points are shown for several modulation methods, which we discussed earlier. For all the cases shown, the BER delivered is 10^{-5} . The displacement of each design point from the capacity boundary indicates how close the communication efficiency of the corresponding modulation scheme comes to the capacity limit. The horizontal displacement measures the shortfall in terms of SNR per bit, while the vertical displacement measures the shortfall in terms of bandwidth utilization. Note that if we were to plot the modem design points for a lower level of BER delivered, the points would all move to the right (i.e., farther away from the capacity boundary), whereas if we used a higher BER value, they would move closer to the capacity boundary.

It is conventional to call the region $R/W > 1$ the *bandwidth-limited region* of operation and to call the region $R/W < 1$ the *power-limited region* of operation. The bandwidth-limited region includes all the modulation schemes we have described for use on voiceband telephone circuits, where rigid channel bandwidth limitations are

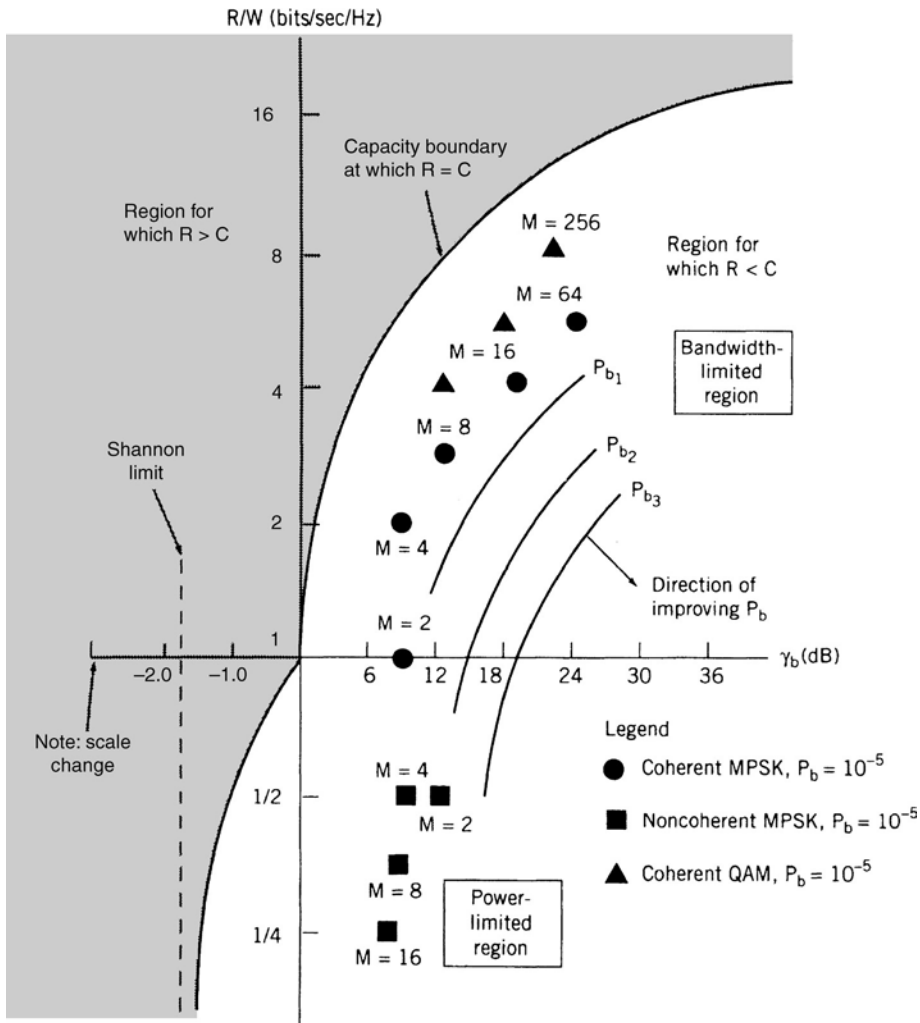


FIGURE 7.16 Channel capacity and a comparison of several modulation methods at bit error probability equal to 10^{-5} . (From [Sk101] © Prentice Hall PTR.)

imposed by the existing design of the public network. There we see that the M -ary modem signal constellations provide steadily increasing bandwidth utilization as M is increased. It can be seen from the figure that the QAM schemes are closest to the capacity boundary.

As can be seen from the figure, the Shannon capacity formula shows that the greatest energy efficiency is achieved in the power-limited region, where the bandwidth must be made very large relative to the information rate. In the limiting case of very large bandwidth and C/W approaching zero, E_b/N_0 approaches $\ln 2$ or -1.6 dB, which is called the *Shannon limit*. In the power-limited region, we show design points for noncoherent binary and M -ary FSK modulations. Because these modulations are bandwidth-expansive, they are not used in modern voiceband modems. However, binary FSK was used in the earliest voiceband modems, as was binary PSK. As modem

technology developed, the industry evolved modems with roughly the succession of modulation methods seen as we progress to the right and upward in the bandwidth-limited region of Fig. 7.16. The current state of the art in high-rate modem technology for the PSTN is high-order QAM with TCM, and data rates up to 33.6 kb/s are now being achieved.

Many radio systems must also be treated as bandwidth-limited media, due in most cases to ever-increasing demands for access to the available spectrum in various frequency bands. Consequently, increasingly sophisticated modem techniques are migrating from the voiceband applications for which they were developed, to various radio applications. As examples, M -ary PSK modulation with adaptive equalization is now a proven technique for radio applications, and there have been applications using 16-ary QAM with trellis coding in some frequency bands.

7.3.2 Transmission Channel Impairments

In reality, most data communication channels are not described accurately by the steady-signal AWGN model. To describe real-world channel impairments, it is again helpful to use signal constellation diagrams. This is done in Fig. 7.17, where several common forms of channel impairment are illustrated by their effects on a 4-ary PSK signal constellation. The first diagram shows an undisturbed constellation, representing a distortion-free channel with no measurable noise. The second diagram shows the constellation with each of the four points smeared into a circular “cloud” caused by additive Gaussian noise. The third diagram shows the effects of *phase jitter*, a prevalent effect on most channels. The phase jitter effect is a continuous periodic smearing of the signal phase, with little or no effect on the signal amplitude. The last diagram shows the effects of harmonic distortion, again common on channels with nonlinearities, which results in nonperiodic smearing of the signal amplitude, with somewhat less effect on signal phase. This appears as an elliptical cloud around each signal point in the constellation. Another impairment observed on some channels, termed *gain jitter*, is a random-appearing amplitude modulation similar in its effect to harmonic distortion.

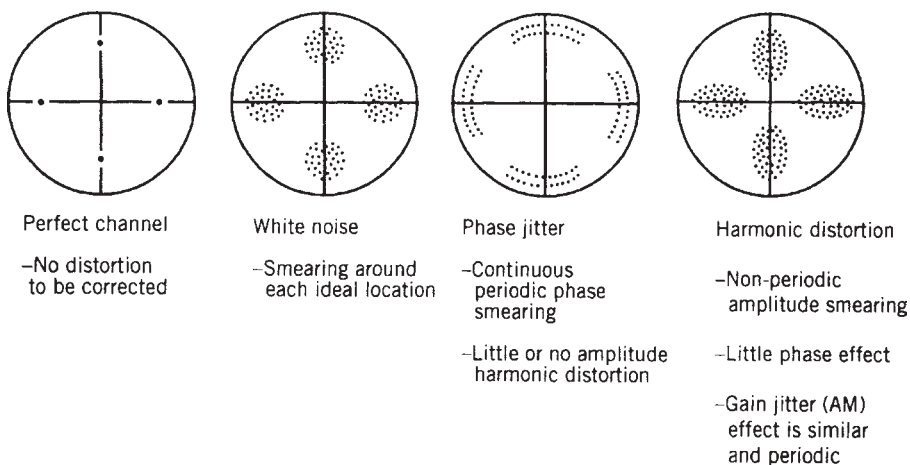


FIGURE 7.17 Transmission channel impairments and their effects on a four-phase modem signal constellation.

Although the impairments depicted in Fig. 7.17 are shown individually, they will in general appear in various combinations on different channels. To visualize the effects of these impairments on modem performance, one can pass decision boundaries through the origin at 45° to the right and left of the vertical axis. Received signal points crossing one or more decision boundaries result in demodulated symbol errors.

All of the channel impairments shown in Fig. 7.17 are found on both wired and wireless channels. Wireless channels also suffer from large amplitude fluctuations caused by signal fading. We treat this topic in detail in Chapter 8. Another category of transmission impairment is one that can be described equally well for wireline and radio systems as distortion due to nonideal channel frequency-response characteristics. That is, when the channel has nonflat amplitude and delay response over the bandwidth occupied by the transmitted signal, the channel acts as a nonideal filter, causing intersymbol interference in the received symbol stream. Intersymbol interference imposes the principal limitation on achievable data rates on bandlimited channels. The difference between the amplitude and phase distortion in wireline and wireless channels is that distortions in wireline channels are on the edges of the band, whereas a radio channel may be subjected to frequency-selective fading even in the midregion of the band.

7.4 TRADITIONAL MODEMS FOR WIDE-AREA WIRELESS NETWORKS

In our earlier discussion of modem technology, we described the progression of modem techniques, ranging from simple OOK and FSK through PSK, PAM, QAM, partial response, and TCM. In this section we describe modulation techniques that have been adopted in most of the developing standards for second-generation wireless information networks. In principle, the modulation techniques discussed earlier are applicable to all wireline and wireless modems. That is, there are basic design issues that are common to both wireline and wireless systems. In general, we would like to transmit data with the highest achievable data rate and with the least expenditure of signal power. In other words, we usually want to maximize both bandwidth efficiency and power efficiency. However, the emphasis on these two objectives varies from one application to another, and there are certain details that are specific to particular applications.

In voiceband telephone channels, high bandwidth efficiency has a direct economic advantage to the user, because it can reduce connect time or avoid the necessity of leasing additional circuits to support the application at hand. The typical telephone channel is less hostile than a typical radio channel, providing a fertile environment for examining complex modulation techniques and signal-processing algorithms. Specific impairments seen on telephone channels are amplitude and delay distortion, phase jitter, frequency offset, and effects of nonlinearities. Many of the practical design elements of wireline modems have been developed to deal efficiently with these categories of impairments.

In radio systems, bandwidth efficiency is also an important consideration, because the radio spectrum is limited and many operational bands are becoming increasingly crowded. Radio channels are characterized by multipath fading and Doppler spread, and a key impediment in the radio environment is the relatively high levels of average signal power needed to overcome fading. However, there are other considerations that affect the selection of a modem technique for a wireless application. In the next subsection we discuss requirements for radio modems in greater detail before going on to a description of the modem techniques that are in most widespread use in evolving wireless networks.

7.4.1 Requirements for Radio Modems

There are a number of considerations that enter into the choice of a modulation technique for use in a wireless application, and here we review briefly the key requirements. These requirements can vary somewhat from one system to another, depending on type of system, the requirements for delivered service, and users' equipment constraints.

Bandwidth Efficiency. Most wireless networks that support mobile users have a need for bandwidth-efficient modulation, and this requirement grows steadily in importance each year. For example, land-mobile radio (LMR) communications in licensed VHF and UHF bands, which has long provided service for public safety organizations and commercial fleet dispatch users, is a rapidly expanding market. Although these networks were designed originally for FM voice service, they are seeing growing use for digital voice and data services. The growth in demand for mobile data services, combined with the need to provide ever more channels, makes bandwidth efficiency a crucial issue. Standard channel spacing in most LMR bands is 25 kHz, and most mobile data applications are supported by data devices operating at data rates of 2400 bits/s or lower. However, a joint public safety and U.S. government LMR standard is currently under evaluation which will specify a digital transmission format for use with 12.5-kHz channel spacing and data rates as high as 9600 bits/s [FS93].

Other well-known developments are the early initiatives in North America, Europe, and the Far East to define 2G standards for digital cellular telephone services to replace the then-existing analog cellular systems. The driving force for those developments was the rapid growth in the market for cellular services, which had resulted in loading analog systems to full capacity during busy hours in some large metropolitan areas. A cellular carrier company is assigned a specified amount of licensed bandwidth in which to operate its system, and therefore an increase in system capacity leads directly to increased revenues. This defines another clear need for modulation techniques that provide efficient utilization of available bandwidth.

An area of wireless communications network development where the modulation bandwidth efficiency is not yet as critical is that of WLANs. Unlike cellular systems, which often have been used to support circuit-mode services (services, like standard telephone service, in which a user has access to the full user-channel capacity for the duration of the call connection), WLANs are typically used for burst-mode traffic. Because of the bursty nature of the user data, the aggregate data traffic on a WLAN rarely approaches system capacity. Furthermore, many WLAN products operate in the unlicensed ISM bands, where the same frequencies are reused again and again even in relatively close geographic areas. It is for these reasons that the WLAN industry has placed relatively little emphasis on bandwidth-efficient modulation techniques.

Power Efficiency. Power efficiency is another parameter that may not be of major importance in some of wireless applications, such as WLANs used to interconnect stationary workstations, because these types of equipment are typically powered from the ac power sources already available in the office or factory environment. However, in most other applications, such as digital cellular, cordless phones, and mobile data services, power translates into battery size and recharging intervals, and even more important to the mobile user, into the size and weight of portable terminals. Thus, power efficiency is important in most personal and mobile communication systems,

and this will become increasingly important as consumers become accustomed to the convenience of small handheld communication devices.

There are two facets of the power requirement: One is the power needed to operate the electronics in the terminal, and the other is the amount of power needed at the input to the power amplifier in order to radiate a given amount of signal power from the antenna. The radiated signal power, of course, translates directly into signal coverage and is a function of the data rate and the complexity of the receiver. Higher data rates require higher operating levels of SNR. More complex systems using TCM or adaptive equalization require less transmission power. However, more complex receiver design increases the power consumed by the electronics and consequently, reduces battery life. In some applications a compromise has to be made between the complexity of the receiver and the electronic power consumption. For example, for handheld local communication devices, some manufacturers avoid the use of complex speech coding techniques in order to hold down battery consumption. Also, in the design of high-speed data communication networks for laptop or pen-pad computers, some designers find it difficult to justify the additional electronic power consumption required for inclusion of adaptive algorithms.

In spread-spectrum CDMA systems, power efficiency and overall system bandwidth efficiency are closely related. The use of a more power-efficient modulation method allows a system to operate at lower SNR. The performance of a CDMA system is limited by the interference from other users on the system, and an improvement in power efficiency in turn increases the bandwidth efficiency of the system. The discussions that follow address only the issue of efficiency through the power amplifier; they do not deal with the issue of power consumption in the electronics of a wireless terminal.

Out-of-Band Radiation. An important issue in the selection of a modulation technique for a radio modem is the amount of transmitted signal energy lying outside the main lobe of the signal spectrum. This issue requires more careful definition than we undertake here, but the key point is that in many multiuser radio systems, the performance is limited by adjacent-channel interference rather than additive noise. For example, in the design of VHF/UHF land-mobile radio systems, the major design parameter is adjacent-channel interference (ACI), which is the interference that a transmitting radio presents to the user channels immediately above and below the transmitting user's channel. ACI will determine the geographic area over which mobile users can be served by a single base station. This is because a low level of ACI will permit a distant mobile transmitter to reach the base station with a weak signal while another mobile much closer to the base station is transmitting in an adjacent channel. Thus, ACI specifications indirectly influence system capacity and cost. Evaluation of the ACI involves the characteristics of the transmitting and receiving channel filters, nonlinearities in the transmitter, and of course the height and rolloff characteristics of the skirts of the transmitted signal spectrum. Radio manufacturers strive to design radios that keep ACI below a specified level, typically -60 dB, and the out-of-band spectral power of the modulation scheme is the principal ingredient in achieving that goal. In contrast, the out-of-band signal power in voiceband modems is not as critical, and a voiceband modem manufacturer would be satisfied with an out-of-band power around -40 dB.

Resistance to Multipath. Another important issue in the design of a radio modem is sensitivity to multipath. Various modulation techniques have different degrees of

resistance to multipath. This was a major issue in the development of digital cellular standards, where it was necessary that each standard be written to accommodate the worst-case multipath conditions likely to be encountered by users over the entire geographic region of usage for that standard. Considerable attention is also being given to multipath specifications as part of the standardization of the air interface for 2-GHz PCS systems in the United States [JTC94a,b].

Constant Envelope Modulation. Most mobile radio products are designed with class C power amplifiers, which provide the highest power efficiency among the common types of power amplifiers. However, class C amplifiers are highly nonlinear, so it is necessary that the signal to be amplified is constant-envelope or as nearly so as is practical. The reason for this is that any amplitude fluctuations in the input signal will give rise to spectral widening of the output signal, in turn causing increased ACI. It is because of these considerations that frequency modulation has remained in widespread use in the mobile radio industry. Although analog FM mobile radio systems were originally designed for analog voice, they have been extended to data service simply by feeding baseband digital data streams to the frequency modulator. This in effect is a method of FSK, where input amplitude levels correspond to transmitted tones. An FM signal is by its very nature constant-envelope; however, it is not spectrally efficient, due to its high sidelobes. Thus, as the needs for greater bandwidth efficiency have grown, efforts have been made to design modulation schemes that are less wasteful of bandwidth while preserving (or nearly so) the constant-envelope nature of FM. To conform to spectrum constraints, it is necessary in some systems to apply filtering to the modulated waveform before power amplification, and the filtering produces amplitude variations. In order that undesirable out-of-band spectral components not be generated, it is then necessary that the amplification be linear. Consequently, such non-constant-envelope filtered signals are commonly referred to as *linear modulation systems*.

In the subsections that follow, we describe three modulation techniques that are specified in prominent standards for radio modems: four-level-FM, GMSK, and $\pi/4$ -shift QPSK. We shall see that the first two are constant-envelope modulations, while the third, which is specified in a filtered version in the IS-54 TDMA digital cellular standard, is a form of linear modulation. Although today's implementations of linear modulation schemes require linear amplifiers, and thus suffer a loss in efficiency relative to class C amplification, much research work is in progress to develop new methods of power amplification that combine near-linearity with power efficiency approaching class C characteristics.

7.4.2 Digital Frequency Modulation

As we noted earlier in this section, FM is the predominant form of modulation used in the mobile radio industry. Although FM has long been used for carrying analog voice over radio systems, newer digital systems have also been based on FM, specifically multilevel digital FM. A widely used format is four-level FM, which is equivalent to 4-ary FSK. A typical modulator implementation will use direct modulation, in which the four-level baseband digital signal is applied directly to the voltage-controlled oscillator (VCO). This provides a relatively simple design that is compatible with class C power amplification and permits demodulation with a simple frequency discriminator followed by a sampler. A disadvantage of digital FM is that the spectral skirts are relatively high,

and therefore it becomes difficult to move carrier frequencies closer together while complying with limits on ACI. However, it is possible to reduce the spectral skirts by filtering the baseband signal before it is applied to the VCO. Because this filtering affects the frequency excursions, not the amplitude of the signal, the constant-envelope nature of the transmitted waveform is preserved. Nevertheless, in the development of some of the new wireless networks, multilevel FM has not been judged to provide sufficient bandwidth efficiency, so other modulation techniques have been adopted. We describe these alternative techniques in the following subsections.

7.4.3 OQPSK, MSK, and GMSK

Here we describe briefly a set of closely related modulation techniques: offset quadrature phase shift keying (OQPSK), minimum shift keying (MSK), and Gaussian MSK (GMSK). Two of these modulations, MSK and GMSK, can be explained by direct reference to FSK modulation, but it is useful to begin our discussion with quadrature phase shift keying (QPSK), which we mentioned briefly in Section 7.2.6. The overall block diagram of QAM systems was given in Fig. 7.8 and the general description of the transmitted signal was given by Eq. (7.2.10). For QPSK (4-QAM) modulation, the signal constellation has four points on a circle, each point associated with two data bits, and each of the information streams transmitted, $\{a_i\}$ and $\{b_i\}$, takes values $\pm$ each stream carrying one of the two encoded data bits per symbol. Therefore, if $\{d_k\}$ is the sequence of the incoming data bit stream, $\{a_i\}$ represents the even bits of the stream and $\{b_i\}$ the odd bits. Figure 7.18 is a redrawing of the transmitter part of Fig. 7.8a, representing the incoming data $\{d_k\}$.

As shown in Fig. 7.18, the QPSK signal transmitted can be written as

$$s(t) = d_k p(t) \cos \omega_c t + d_{k+1} p(t) \sin \omega_c t, \quad k = 0, 2, 4, \dots$$

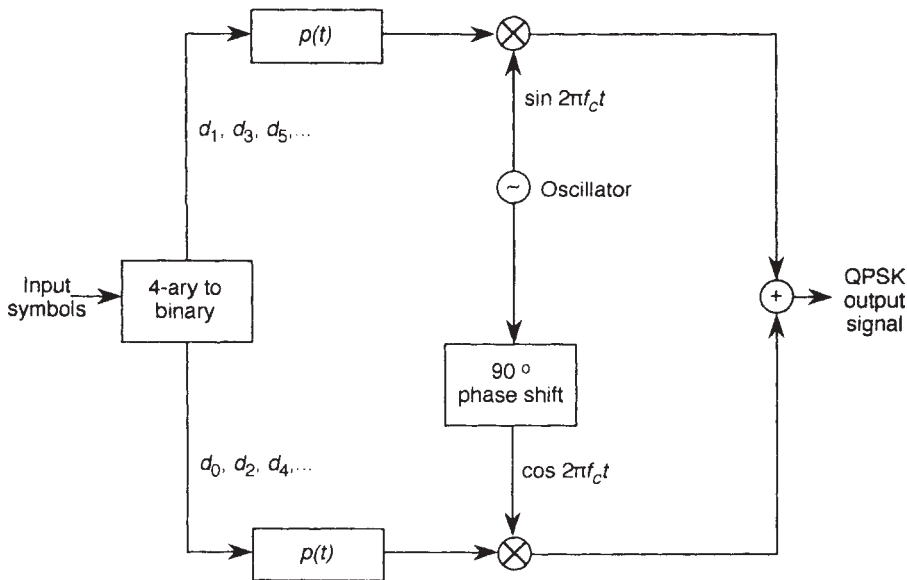


FIGURE 7.18 Quaternary phase shift keying.

where $\{d_k\}$ is the stream of data bits, which we have divided into even-numbered bits $d_0, d_2, d_4, \dots$ and odd-numbered bits $d_1, d_3, d_5, \dots$. On the right-hand side of the equation above, the first term is called the *in-phase signal* and the second term is called the *quadrature signal*. The spectrum of $s(t)$ is simply the sum of the spectra of the two terms; and because each term represents binary PSK signaling, the overall spectrum has the same shape as that of binary PSK. Therefore, the bandwidth efficiency of QPSK is twice that of BPSK. Modems using QPSK modulation and variations of this technique have been used in many different radio systems. One common variation of basic QPSK is differential QPSK, in which the information to be transmitted is sent as successive changes in the signal phase. Another important variation of this modulation technique is offset QPSK, which we discuss next.

OQPSK. Figure 7.19 shows the basic structure of offset QPSK (OQPSK), which is also referred to as *staggered QPSK*. In this figure we denote the streams of even- and odd-numbered data bits as $d_I(t)$ and $d_Q(t)$, respectively. In an OQPSK modulator, instead of applying the source data bits to the two branches simultaneously every T seconds (Fig. 7.18), the two branches are offset by $T/2$ seconds. The QPSK and OQPSK transmitted waveforms are compared in Fig. 7.20. The benefit of this scheme can be seen by referring to the QPSK modulator shown in Fig. 7.18. In ordinary QPSK modulation, the data bit polarity can change sign simultaneously on the two branches, which results in a 180° phase shift in the transmitted waveform. (In a QPSK stream carrying random data, one-fourth of the symbol-to-symbol phase transitions will be 180° phase shifts.) When the QPSK-modulated signal is filtered, as must typically be done to constrain its radiated spectrum, the resulting signal exhibits significant amplitude variations. The presence of these amplitude variations in turn rules out the use of highly nonlinear but power-efficient amplifiers. Nonlinear amplification would simply regenerate the out-of-band spectral components, which the filtering is meant to suppress.

However, referring to Fig. 7.19, we see, that with OQPSK, the data bits can change polarity on only one branch at a time, every $T/2$ seconds, and thus the phase of the modulated signal can change by at most 90° from one symbol to the next. The end

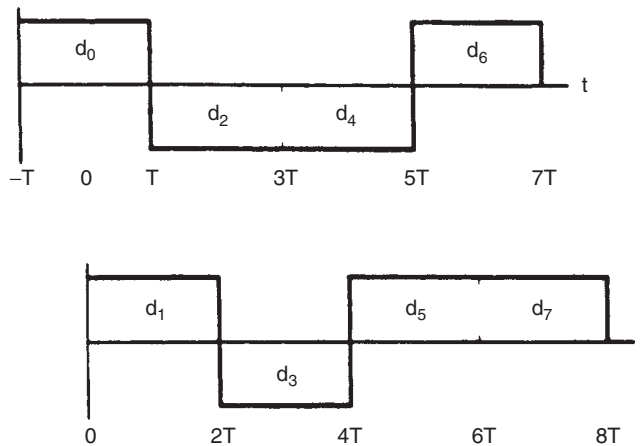


FIGURE 7.19 Offset QPSK data streams.

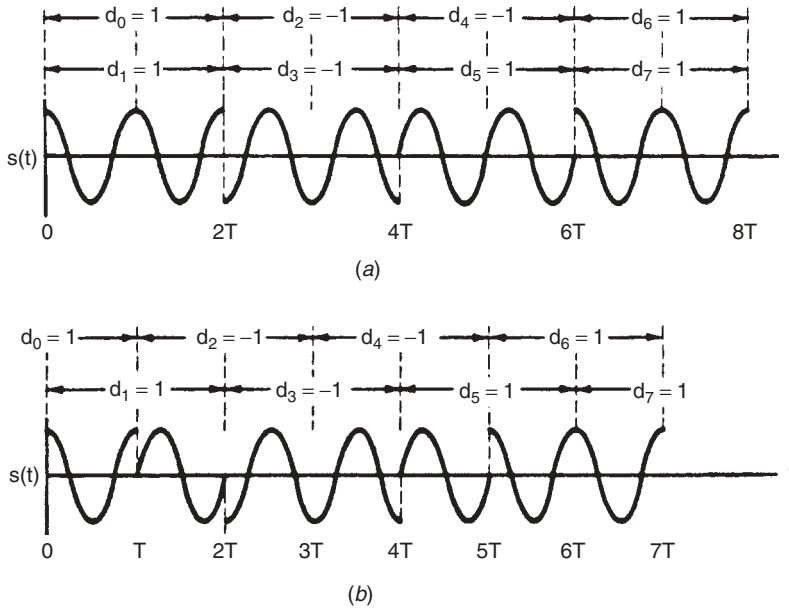


FIGURE 7.20 Comparison of QPSK (a) and OQPSK (b) waveforms. (From [Pas79] © IEEE.)

result of this modification is much lower variations in the envelope of the modulated signal after passing through a bandlimiting filter. Thus, the filtered signal can then be put through a nonlinear amplifier with very little growth in the out-of-band spectral components. Note that because the offset alignment does not change the spectra of the branch signals, each is the spectrum of BPSK modulation with keying interval T , and therefore the spectrum of OQPSK is identical to the spectrum of basic QPSK. However, the spectra of the two signals are very different after bandlimiting and nonlinear amplification.

Because the offset time alignment does not affect the orthogonality of the two branches of the modulator, the theoretical BER performance of OQPSK is identical to that of QPSK for the same received signal and noise power. But because OQPSK can be used with class C power amplification, it can be implemented with considerably less prime power in the transmitter. OQPSK was first used in satellite systems because of its compatibility with the highly nonlinear amplifiers used in satellite repeaters and because of the critical importance of minimizing power consumption onboard the satellite. Power consumption is also important in mobile communications systems, where low weight and long battery life are important to users of handheld radios. As shown in Chapter 9, the equalizers used for OQPSK modulation and their performance on multipath channels are slightly different from those of standard equalizers used for QPSK.

Thus, we see that modulation techniques that constrain the instantaneous phase transitions in the transmitted waveform can yield important benefits in the design of power-efficient systems. A further improvement of this nature is provided by the modulation technique we examine next, which is closely related to OQPSK.

MSK. Minimum shift keying (MSK) can be described either as a special case of frequency modulation or as a variation of OQPSK. Perhaps the simplest definition

of MSK is that it is phase-continuous coherent binary FSK with modulation index $m = 0.5$. The modulation index is defined by $m = T\Delta f$, where Δf is the frequency spacing between the FSK tones, and therefore the tone spacing is $1/2T$, the minimum spacing for which orthogonality over the symbol interval T can be achieved. This is what gives MSK its name. Analytically, we can write the transmitted MSK signal as

$$s(t) = \cos \left[2\pi \left(f_c + \frac{d_k}{4T} \right) t + x_k \right], \quad kT < t < (k+1)T \quad (7.4.1)$$

where f_c is the carrier frequency, $\{d_k\}$ ($k = 0, 1, 2, 3, \dots$) is the sequence of data bits, and x_k is a value of phase, which is constant over the k th successive T -second symbol interval. During each interval, x_k is either 0 or π , to meet the requirement that the phase be continuous from the end of one symbol interval to the start of the next. This requirement is satisfied if x_k is given by the following recursion:

$$x_k = \left[x_{k-1} + \frac{\pi^k}{2}(d_{k-1} - d_k) \right] \bmod 2\pi \quad (7.4.2)$$

Examining Eq. (7.4.1), we see that the frequency in each symbol interval is either $f_c + 1/(4T)$ or $f_c - 1/(4T)$, in accordance with the data bit value to be transmitted in that interval. Therefore, the tone spacing is $1/2T$, the minimum tone spacing for signal orthogonality over the interval T .

One way of implementing MSK is to use phase-continuous FSK with a modulation index of 0.5, as represented by Eq. (7.4.1) and as shown in Fig. 7.21. At the receiving end, for best BER performance, the signal is demodulated by coherent FSK demodulation. Alternatively, the MSK signal can be detected with a frequency discriminator followed by a slicer, a simpler implementation that yields somewhat poorer BER performance than does coherent FSK demodulation. In practice, in radio modems, MSK has typically not been implemented using the FM approach depicted in Fig. 7.21, because it requires the modulation index to be implemented very precisely in order for phase coherence to be preserved. (This, of course, is not important if frequency discriminator detection is to be used.) Consequently, a quadrature modulator structure is ordinarily used, as described next. Our presentation follows those of Pasupathy [Pas79] and Sklar [Skl88].

Using standard trigonometric identities, we can rewrite Eq. (7.4.1) in the form

$$s(t) = a_k \cos \frac{\pi t}{2T} \cos 2\pi f_c t + b_k \sin \frac{\pi t}{2T} \sin 2\pi f_c t, \quad iT < t < (i+1)T \quad (7.4.3)$$

where

$$a_k = \cos x_k = \pm 1, \quad b_k = d_k \cos x_k = \pm 1 \quad (7.4.4)$$

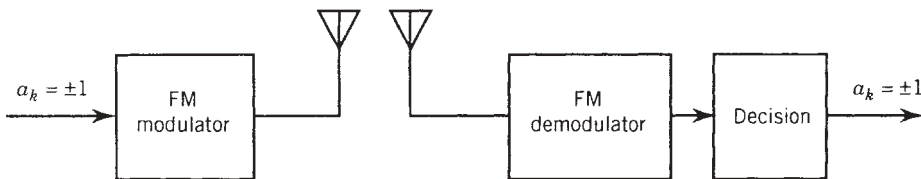


FIGURE 7.21 MSK implemented as phase-continuous FSK.

In this quadrature form, the in-phase signal component is represented by $a_k \cos(\pi t/2T) \cos 2\pi f_c t$, where $\cos(\pi t/2T)$ is described as sinusoidal symbol weighting and a_k depends on the data through Eq. (7.4.4). Similarly, the quadrature signal component is given by $b_k \sin(\pi t/2T) \sin 2\pi f_c t$, where $\sin(\pi t/2T)$ is a sinusoidal symbol weighting and b_k also depends on the data through Eq. (7.4.4). At first examination of Eq. (7.4.4), it might appear that a_k and b_k can change every T seconds, because the data bit value d_k can certainly change every T seconds. However, because of the continuous-phase constraint, as ensured by Eq. (7.4.2), a_k can change sign only at the zero crossings of $\sin(\pi t/2T)$, and b_k can change sign only at the zero crossings of $\cos(\pi t/2T)$. Thus, the symbol weighting in either the in-phase or quadrature channel is a half-cycle sinusoidal pulse of duration $2T$ seconds with alternating sign. As with OQPSK, the two channels of this modulator are offset by T seconds. Therefore, we can consider MSK modulation to be a form of OQPSK having sinusoidal symbol weighting on the quadrature channels and a special relationship between the source data stream d_k and the binary values applied to the quadrature channels. For steady-signal reception in AWGN, the BER performance for MSK is exactly the same as for OQPSK and QPSK. The composition of the MSK waveform is shown in Fig. 7.22.

As with OQPSK, MSK is an offset constant-envelope modulation, but MSK has the additional advantage that it is phase continuous, and therefore even the $\pm 90^\circ$ shifts of OQPSK are eliminated. In fact, it is readily seen from Eqs. (7.4.1) and (7.4.2) that in each T -second interval, the phase of the signal moves ahead of or behind the carrier phase at a constant rate for a total excursion of exactly $\pi/2$ radians over the interval. The phase continuity of MSK results in a signal spectrum with tails somewhat lower than those of OQPSK. The power spectral density $G(f)$ for MSK modulation is given [Pro01] by

$$G(f) = \frac{16A^2T}{\pi^2} \left(\frac{\cos 2\pi fT}{1 - 16f^2T^2} \right)^2$$

where A is the amplitude of the MSK signal.

Figure 7.23 shows the normalized power density spectra for BPSK, QPSK, OQPSK, and MSK. As can be seen in the figure, MSK has lower sidelobes than QPSK or OQPSK, but the main lobe of the MSK spectrum is broader. At the 3-dB power points, the MSK main lobe is about 30 per cent wider than that of QPSK.

GMSK. Murota and Hirade [Mur81] observed that the MSK signal spectrum could be made even more compact, and the spectral skirts lowered further, by implementing the modulation in its direct FM form with a low-pass filter applied to the data stream before modulation. This technique provides a smoothing of the phase transitions at symbol boundaries; and because the filtering is done before modulation, the constant-envelope property of the signal is preserved. The specific filter characteristic proposed by Murota and Hirade is the Gaussian low-pass filter. This modulation technique, shown in Fig. 7.24, is called *Gaussian filtered MSK* or simply *Gaussian MSK* (GMSK). The rapid roll-off achievable with Gaussian filters provides spectra that are more compact than that of strict-sense MSK. Because the transform of a Gaussian function is also Gaussian, the impulse response of the premodulation filter is Gaussian. The tails of the Gaussian time-domain function remain above zero, and this results in some intersymbol interference. The choice of roll-off characteristic for the Gaussian filter involves a trade-off between spectral confinement and performance

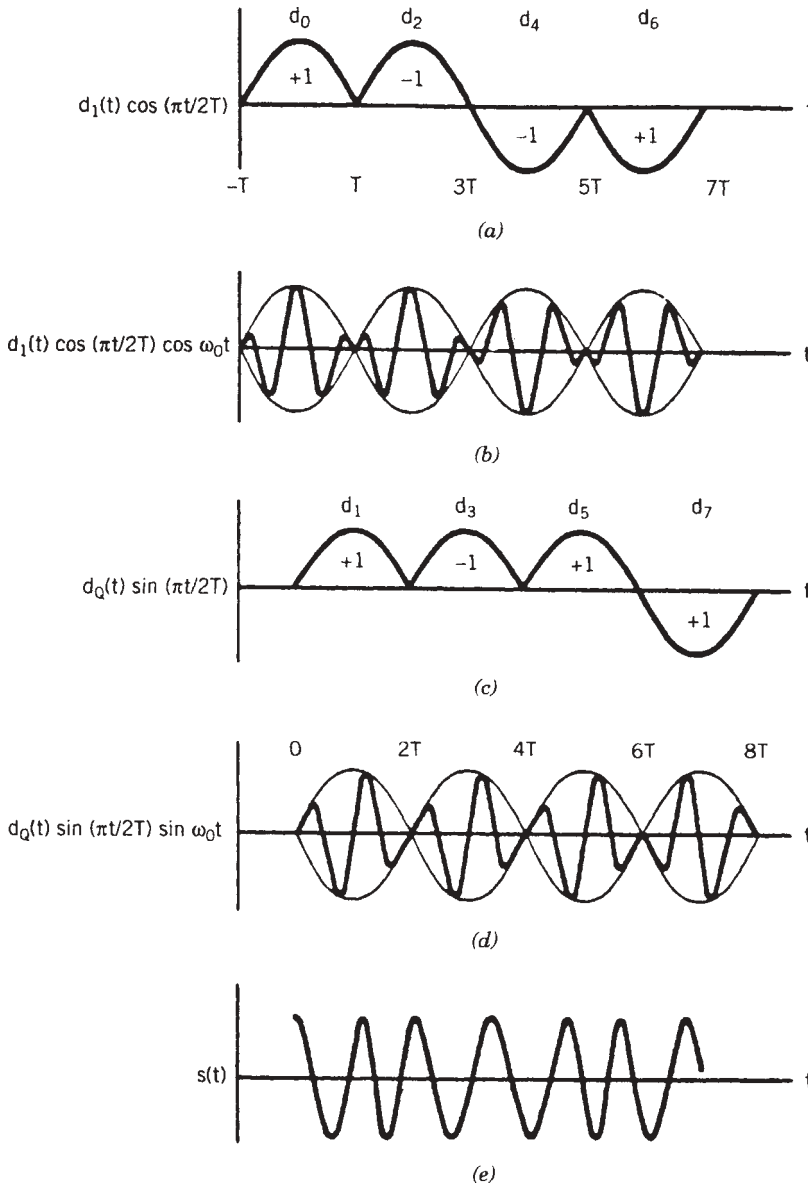


FIGURE 7.22 MSK waveform composition: (a) modified I bit stream; (b) I bit stream times carrier; (c) modified Q bit stream; (d) Q bit stream times carrier; (e) MSK waveforms. (From [Pas79] © IEEE.)

loss. Figure 7.25 shows power spectra for GMSK versus the normalized frequency difference from the carrier center frequency, $(f - f_c)T$, with a normalized 3-dB filter bandwidth $B_b T$ as a parameter. The parameter selection $B_b T = \infty$ effectively removes the filter and thus corresponds to pure MSK. Decreasing values of $B_b T$ produce corresponding narrowing of the power spectrum while producing increasing degradation in BER performance relative to unfiltered MSK. The spectral plots presented here

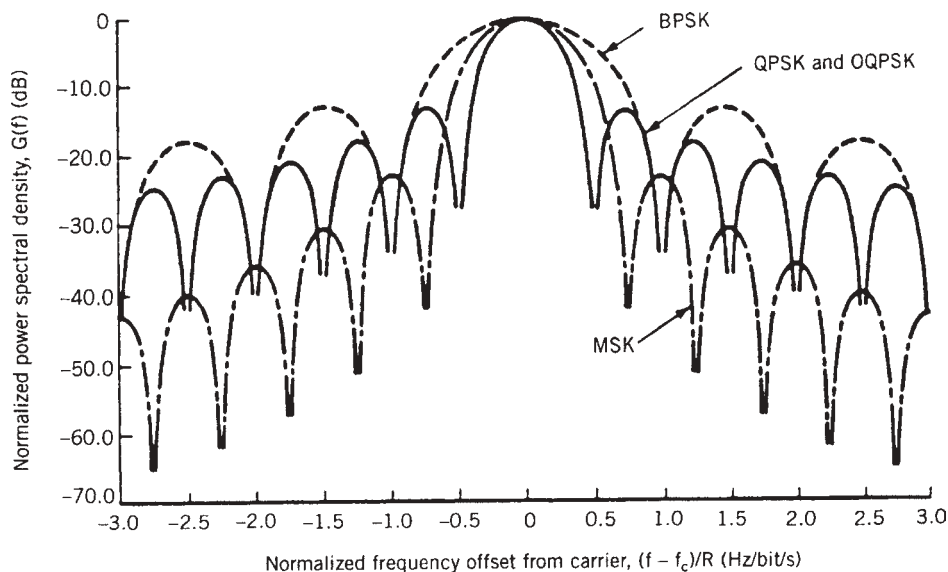


FIGURE 7.23 Normalized power density spectra for BPSK, QPSK, OQPSK, and MSK. (From [Amo80] © 1980 IEEE.)

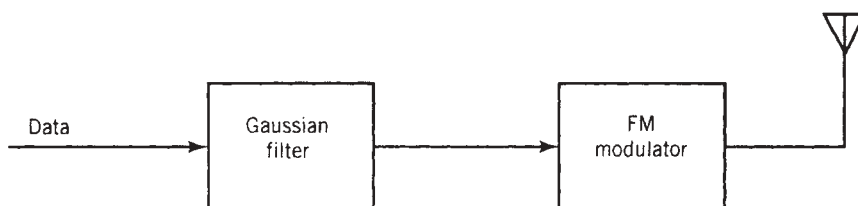


FIGURE 7.24 Gaussian MSK modulation.

represent a practical range of bandwidth selections. In their paper, Murota and Hirade suggest $B_b T = 0.25$ as an optimum choice and indicate that this produces a performance degradation of no more than 0.7 dB relative to MSK. The GMSK modulation, with bandwidth parameter $B_b T = 0.3$, has been adopted for GSM, the Pan-European Digital Cellular Standard [Rah93, Hau94].

It can be seen from the spectral plots that as the Gaussian filter bandwidth is reduced, both the sidelobes and the width of the main lobe are reduced. The narrowing of the main lobe relative to a fixed data rate yields an improvement in bandwidth efficiency, measured by bits per second per hertz. Although GMSK provides some improvement in bandwidth efficiency relative to MSK, about 1.6 rather than 1.4, the bandwidth efficiency of QPSK or OQPSK is still better, at about 1.8 bits/s per hertz. However, GMSK does provide an important advantage in the low level of its spectral sidelobes, which in turn translates into superior ACI performance.

Table 7.1 gives an occupied bandwidth for several percentage values of total signal power, with $B_b T$ also a variable parameter. The table provides a convenient means of assessing how the selection of premodulation filter bandwidth affects the distribution of transmitted signal power between the main lobe and the sidelobes of the spectrum.

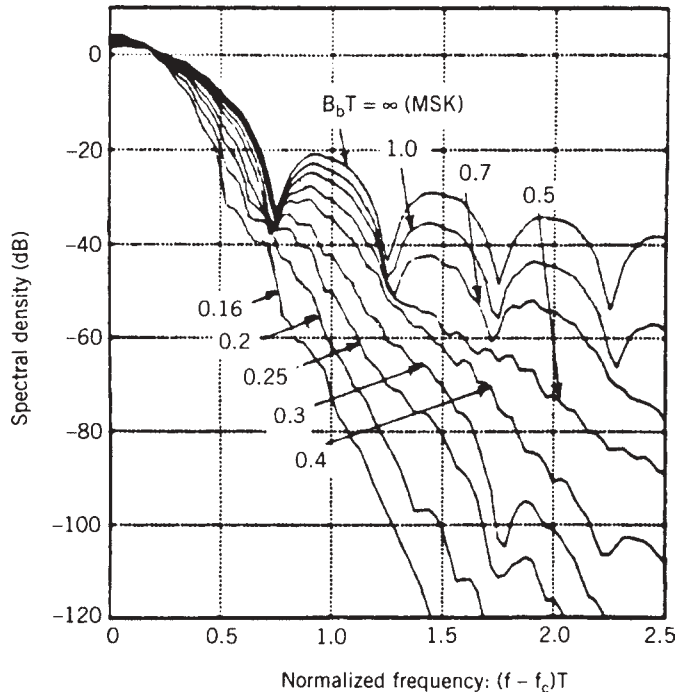


FIGURE 7.25 Power spectral density of GMSK. B_bT is the normalized 3-dB bandwidth of the premodulation Gaussian LPF; T is the unit bit duration. (From [Mur81] © IEEE.)

TABLE 7.1 Occupied Bandwidth Normalized by Bit Rate for Specified Percentage Power

B_bT	Power (%)			
	90	99	99.9	99.99
0.2	0.52	0.79	0.99	1.22
0.25	0.57	0.86	1.09	1.37
0.5	0.69	1.04	1.33	2.08
MSK	0.78	1.20	2.76	6.00
TFM	0.52	0.79	1.02	1.37
Occupied bandwidth				

Source: [Mur81] © IEEE.

Example 7.4: GMSK Modulation in GSM The GSM system for digital cellular communications uses GMSK modulation with a normalized bandwidth parameter $B_bT = 0.3$. A plot of the power spectral density for this case is included among the plots in Fig. 7.25. For steady-signal reception in AWGN, the bit error probability for GMSK was shown by [Mur81] to be

$$P_b = \frac{1}{2} \operatorname{erfc}(\sqrt{\alpha \gamma_b})$$

where the parameter α is a constant related to the normalized bandwidth $B_b T$. Values of α were determined to be

$$\alpha = \begin{cases} 0.68 & \text{for GMSK with } B_b T = 0.25 \\ 0.85 & \text{for simple MSK } (B_b T \rightarrow \infty) \end{cases}$$

From these values of α it can be seen that the performance of GMSK with $\alpha = 0.25$ degrades from that of MSK by about 1 dB, and thus with the $\alpha = 0.3$ value of the GSM design, the degradation is even smaller. Experimental BER test results given in [Mur81] indicate that with the use of orthogonal coherent detection of the GMSK signal, with $\alpha = 0.25$, BER performance degrades by about 1.6 dB relative to ideal coherent BPSK.

As with MSK, GMSK can be viewed as a form of digital FM and therefore can be demodulated with a simple limiter–discriminator detector [Hir84, Eln86]. However, because of its close relationship to ordinary MSK, best performance is obtained by demodulating the received GMSK signal with a two-branch coherent demodulator very much the same as that used with MSK. In GMSK reception, extra care must be taken in the design of the demodulator to ensure reliable carrier and timing recovery given the use of premodulation filtering. These issues are discussed in some detail in [Mur81].

In a signal-fading environment, coherent detection of GMSK has been shown to exhibit high error floors or irreducible error rates [Hir79]. Thus, there has been much work on developing various forms of differential detection for GMSK [Ogo81, Sim84, Yon86a, Yon86b, Yon88].

7.4.4 $\pi/4$ -Shift QPSK

In our previous discussions we noted that multilevel FM, which has been used widely in mobile radio systems, in large part due to its amenability to nonlinear amplification and limiter-discriminator detection, does not provide the bandwidth efficiency required in modern digital radio systems. We saw that QPSK provides much better bandwidth efficiency, and the staggered version, OQPSK, also provides greater compatibility with nonlinear amplification, due to its lower amplitude fluctuations after filtering. However, a drawback of OQPSK modulation is that it may require coherent demodulation and may suffer performance degradations on radio channels with large Doppler shifts [Feh91]. These limitations relate directly to the $T/2$ -staggered structure of the OQPSK signal. The search for nonstaggered modulation schemes having low postfiltering amplitude variations led to work by Akaiwa and Nagata [Aka87] and others, including [Liu89, Goo90, Liu90], on $\pi/4$ -shift QPSK modulation.

Simply described, $\pi/4$ -shift QPSK is a form of QPSK modulation in which the QPSK signal constellation is shifted by 45° each symbol interval T . This means that phase transitions from one symbol to the next are restricted to $\pm\pi/4$ and $\pm3\pi/4$. By eliminating the $\pm\pi$ transitions of QPSK, the amplitude variations after filtering are reduced significantly. Also, the presence of a phase transition between every pair of successive symbols makes it easier to synchronize the demodulator. Figure 7.26a shows the eight possible phase states of the $\pi/4$ -shift QPSK modulated carrier at the sampling instants. The eight phases represent the four-phase QPSK constellation in its two shifted positions. The four dashed lines radiating from each point on the circle

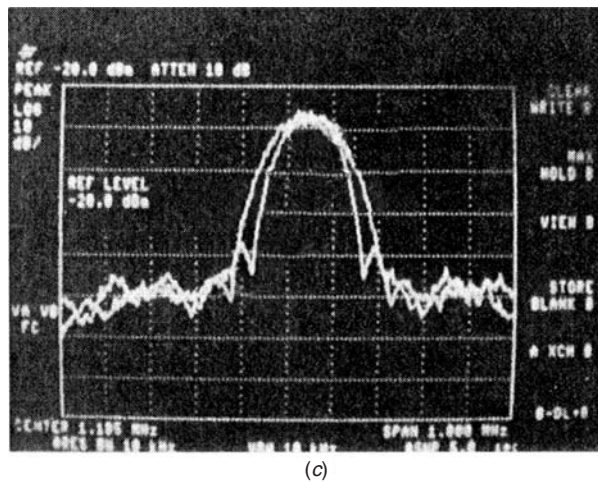
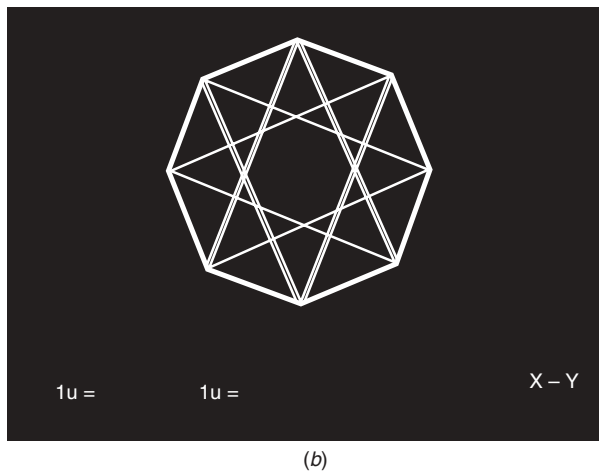
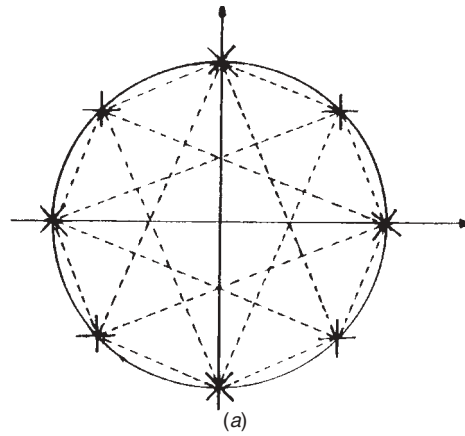


FIGURE 7.26 $\pi/4$ -QPSK modulation: (a) possible phase states of the $\pi/4$ -QPSK modulated carrier at sampling instants; (b) signal constellation with sinusoidal shaping; (c) spectrum of the $\pi/4$ -QPSK signal (upper trace) compared with that of an SQAM signal (lower trace). (From [Feh 91] © IEEE.)

indicate the allowed phase transitions. Figure 7.26*b* shows the signal constellation and the transitions as displayed on an oscilloscope. In the implementation depicted, the modulation is implemented with sine-wave pulse shaping [Feh91]. Figure 7.26*c* shows spectra measured with two versions of $\pi/4$ -shift QPSK, nonlinearly amplified. The upper trace is strict-sense $\pi/4$ -shift QPSK, while the lower trace is the spectrum for sine-wave pulse-shaped $\pi/4$ -shift QPSK (SQAM).

Thus, $\pi/4$ -shift QPSK modulation provides the bandwidth efficiency of QPSK together with a diminished range of amplitude fluctuations. Furthermore, the $\pi/4$ -shift QPSK modulation has the advantage that it can be implemented with coherent, differentially coherent, or discriminator detection [Liu90]. These multiple advantages of $\pi/4$ -shift QPSK led to its adoption as the U.S. digital cellular (USDC) TDMA standard [TIA92] as well as the Japanese digital cellular standard [Nak90] and the standard for trans-European trunked radio (TETRA).

Although it is not essential, $\pi/4$ -shift QPSK modulation is frequently implemented with differential encoding, because this permits the use of differential detection in the receiver, although coherent detection may also be used to achieve optimum performance. This scheme is termed *differential $\pi/4$ -shift QPSK*, denoted simply as $\pi/4$ -DQPSK. A good approximation to the probability of bit error for $\pi/4$ -DQPSK, derived in [Mil98], is given by

$$P_b \simeq \frac{1}{2} \operatorname{erfc} \left(\sqrt{0.5858 \frac{E_b}{N_0}} \right)$$

which is in the general format of Eq. (7.2.7). Compared with the basic binary modulation techniques in additive noise channels given in Eq. (7.2.13), the performance of $\pi/4$ -DQPSK is slightly better than FSK but not as good as BPSK-CD. The main advantage of this modulation over traditional QPSK remains that of having better continuity in the phase transitions, resulting in an envelope that is closer to being constant. This performance is also very close to the performance of the GMSK discussed in Example 7.4.

The use of differential detection avoids the complexity required to extract a coherent carrier reference reliably under multipath fading conditions. Figure 7.27 shows a block diagram of a baseband differential detector for $\pi/4$ -DQPSK. Figure 7.28 shows the BER performance of $\pi/4$ -DQPSK in a flat, slowly fading channel corrupted by AWGN and co-channel interference. As we explain in Chapter 8, in a fading channel we need a much higher signal-to-noise ratio to achieve the same average error rates as those observed in additive Gaussian noise channels.

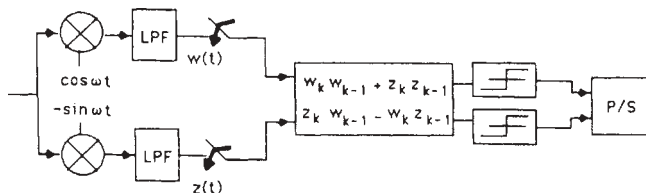


FIGURE 7.27 Baseband differential detector for $\pi/4$ -DQPSK. The low-pass filters are assumed to be square-root raised cosine. (From [Feh91] © IEEE.)

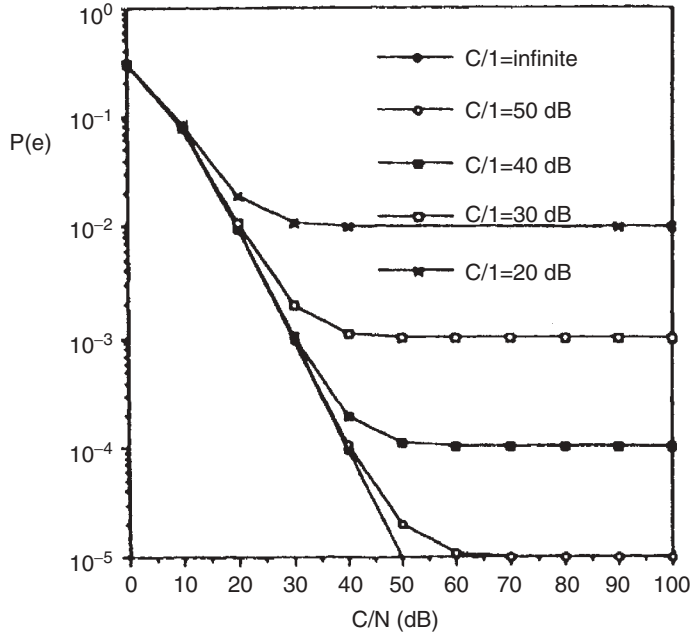


FIGURE 7.28 $P(e)$ versus C/N for $\pi/4$ -DQPSK in a flat, slowly fading channel, with Gaussian noise and co-channel interference. x, $C/I = 20$ dB; $\square$, $C/I = 30$ dB; $\blacksquare$, $C/I = 40$ dB; $\circ$, $C/I = 50$ dB; $\bullet$, $C/I = \infty$. (From [Feh91] © IEEE.)

Example 7.5: Modulation in the USDC Standard The IS-136 standard for the USDC TDMA digital cellular system specifies the use of $\pi/4$ -DQPSK modulation, implemented with raised-cosine amplitude shaping and a roll-off factor β of 0.35. Using common terminology, amplitude shaping in the USDC waveform is often referred to as having *excess bandwidth* of 35%. The standard also specifies that Gray coding be used in mapping of data bit pairs to the adjacent phases in the $\pi/4$ -QPSK signal constellation. In accordance with the USDC standard, the binary data stream $\{b_m\}$ entering the modulator is converted by a serial-to-parallel converter into two separate binary streams $\{X_k\}$ and $\{Y_k\}$, constituting the odd- and even-numbered bit streams, respectively. The data bits are then differentially encoded; that is, symbols are transmitted as changes in phase rather than as absolute phases. The binary streams $\{X_k\}$ and $\{Y_k\}$ are differentially encoded into in-phase and quadrature signal values $\{I_k\}$ and $\{Q_k\}$, respectively, by the following:

$$\begin{aligned} I_k &= I_{k-1} \cos[\Delta\Phi(X_k, Y_k)] - Q_{k-1} \sin[\Delta\Phi(X_k, Y_k)] \\ Q_k &= I_{k-1} \sin[\Delta\Phi(X_k, Y_k)] - Q_{k-1} \cos[\Delta\Phi(X_k, Y_k)] \end{aligned}$$

where $\{I_{k-1}, Q_{k-1}\}$ are the in-phase and quadrature amplitudes in the preceding pulse interval. The phase-change delta value is determined as shown in Table 7.2. The differentially encoded signal values $\{I_{k-1}, Q_{k-1}\}$ can each take on one of five values, 0, ± 1 , $\pm 1/\sqrt{2}$, thus producing the signal constellation shown in Fig. 7.26. Further details on the USDC modulation specification may be found in [TIA92].

TABLE 7.2 Phase Changes in the USDC Modulation

X_k	Y_k	$\Delta\Phi$
1	1	$-\frac{3\pi}{4}$
0	1	$\frac{3\pi}{4}$
0	0	$\frac{\pi}{4}$
1	0	$-\frac{\pi}{4}$

7.5 OTHER ASPECTS OF MODEM IMPLEMENTATION

We have devoted most of this chapter to describing the principal signal modulation techniques employed in narrowband modems, including techniques used in traditional wireline communications and in wireless communications as well. In this concluding section, we address certain additional aspects that must be considered in the design of a narrowband modem.

Figure 7.29 shows a block diagram of a transmitter and a receiver for a typical modem. The transmitter consists of a coder, a pulse-shaping filter, a modulator, and a power control element. The receiver consists of a decoder, the second part of the pulse-shaping filter, an automatic gain (power) control circuit, a timing recovery circuit, a phase recovery circuit, a demodulator, and a box called a *digital processor*. The encoder will perform both source and channel coding plus data scrambling, if utilized.

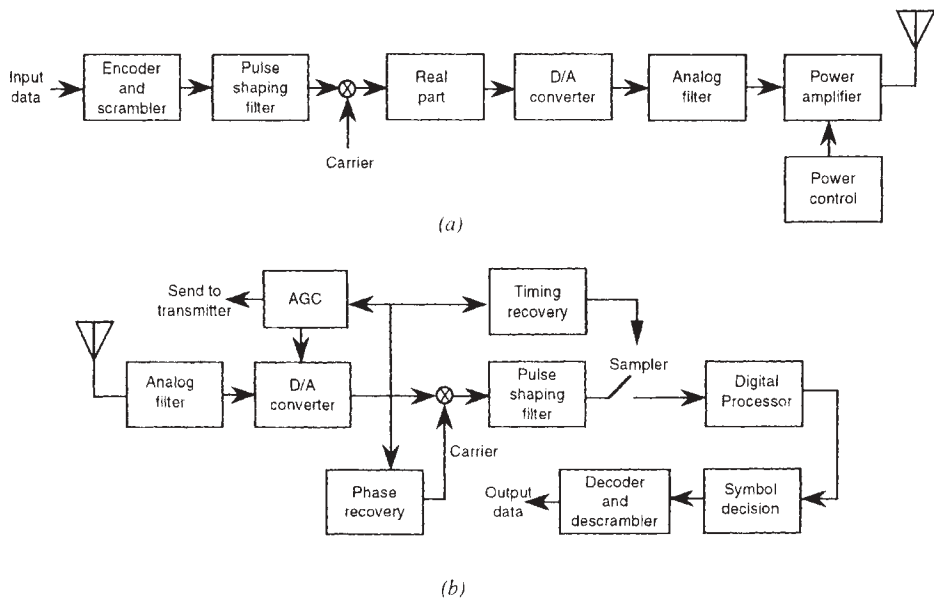


FIGURE 7.29 Block diagrams for a typical modem: (a) transmitter; (b) receiver.

Numerous applications of signal processing are identified in the figure. These applications can be divided into three categories: (1) signal processing needed for proper operation of the modem, (2) signal processing used to mitigate the effects of frequency selective fading or cancel interference in the channel, and (3) source-coding algorithms used for information compression.

Proper operation of a receiver requires power control normally done with automatic gain control (AGC) circuits, recovery of a timing reference for the transmitted pulses, recovery of the transmitted carrier phase for coherent modulation, and pulse shaping to optimize the use of channel bandwidth and to minimize the *intersymbol interference* (ISI). In many modern wireless information networks, transmitter power control is implemented as well.

The purpose of source coding is to minimize the volume of information transferred, to save transmission time and to accommodate as many users as possible in a given bandwidth. Earlier in this chapter we discussed various modulation techniques used to achieve high bandwidth efficiency, measured by the information transmission rate achievable in a given bandwidth. Source coding is another approach to increasing the bandwidth efficiency of the system, one that is independent of the transmission channel or modulation technique. The ongoing evolution of wireless information networks is toward multimedia systems integrating data, voice, and image/video services in a unified network. Data compression techniques are now used extensively in voiceband data communications, and in some of the new wireless data service standards as well.

Speech coding is an important function in digital cellular communications, where the support of a large number of users requires transmitting digitized speech at the lowest possible rates. Cellular systems use sophisticated model-based speech-coding algorithms to compress speech signals to rates as low as 7 kb/s. In wireless systems such as cordless telephone, where bandwidth utilization efficiency is less critical but where power consumption and battery life are important, simpler speech-coding algorithms, such as adaptive differential PCM (ADPCM) with speech-coding rates around 32 kb/s, have been used. With respect to image coding, both fixed- and moving-image coding techniques are being used in the implementation of wireless services. Section 15.6 provides an overview of the leading speech-coding algorithms used in wireless information networks. In addition to the signal modulation and demodulation functions, the most basic signal-processing functions performed in a modem are power control, carrier and timing recovery, and pulse shaping. We next provide an overview of these operations.

7.5.1 Power Control

The traditional method of controlling received power is the technique known as *automatic gain control* (AGC). Most modern radio modems for wireless network applications include transmitter power control. With any modem using multiple signal amplitudes (e.g., versions of QAM), some form of gain control is required as part of demodulation. This function is required because, at a minimum, received levels cannot in general be expected to be the same as transmitted levels. Therefore, some form of gain adjustment is required to establish correct decision threshold levels relative to the received signal levels. This problem is compounded in modems operating on radio channels, where the channel can exhibit wide variations in received signal power, including fading to depths of 30 dB. In practice, this is accomplished with an AGC circuit.

In a typical digitally implemented modem, the received signal is passed through an antialiasing filter to eliminate out-of-band noise and is then sampled with an A/D converter. The samples are squared and then low-pass-filtered to provide a smoothed estimate of the power. The filter time constant is chosen to be long enough to ensure that the gain control does not respond to amplitude variations occurring on a symbol-by-symbol basis, but instead, maintains an essentially fixed average output power level. The output power is compared with a reference, and the difference is used to adjust the step size of the A/D converter. In this way, the range of received signal is adjusted in accordance with average received power, and the digital representation of the received signal remains essentially independent of power fluctuations in the channel. A detailed treatment of AGC in modem design may be found in [Bin88].

Transmitter power control is the simplest method for counteracting the effects of fading. As the signal goes into a fade, the transmitted power is increased to compensate for the performance degradation due to the fading. In an interference-limited environment such as a cellular telephone or PCS network, reduction in the average transmitted power will reduce the interference, which will in turn increase the number of simultaneous users that can be supported in a service area. For example, in a cellular telephone network, the transmitter power level of each mobile terminal is being controlled constantly by its serving base station. This ensures that the mobile terminal transmits only as much power as is needed to maintain good quality of service, which results in significant extension of battery life and controls co-channel interference on mobile-to-base channels. Power control is especially important in code-division multiple access (CDMA) spread-spectrum communication systems, where user signals in each cell coverage area are overlaid onto each other in the same frequency channel. Since every user signal is a direct interferer with all other users on that channel, power control critically affects the number of user signals that can occupy the channel simultaneously.

The common methods of transmit power control are the open- and closed-loop approaches. In the open-loop scheme, the sum of the transmitted and received power in a two-way communication is held constant. Assuming a reciprocal channel and two-way communication, a weak received signal is an indicator of deep fading. An increase in the transmitted power compensates for the fading. With closed-loop power control the receiver informs the transmitter about the level of the received signal, which allows the transmitter to adjust its power accordingly. The open-loop method converges more rapidly, but as a practical matter, not all channels are reciprocal. We discuss power control further in Chapters 10 and 11.

7.5.2 Carrier and Timing Recovery

The most important ingredient in accomplishing accurate sampling and data demodulation is the provision of an accurate timing reference at the receiver. This timing requirement is present whether the form of demodulation is coherent, differentially coherent, or noncoherent. In the case of coherent demodulation, carrier recovery is also required, since the constellation of transmitted signals is defined relative to a fixed phase reference.

The requirements for carrier and timing recovery are closely related. The carrier recovery function provides an estimate of the received carrier phase, and the timing recovery function provides an estimate of the correct symbol sampling times. Inaccuracy in recovering and tracking the signal phase will cause “smearing” of the received

signals around their assigned points in the signal constellation, thus increasing the probability of incorrect symbol decisions in additive noise. Similarly, timing errors will cause the demodulator to sample the received symbol stream at time instants misaligned from the maximum “eye openings,” again making the symbol decisions more vulnerable to additive noise. Carrier and timing recovery can be examined as separate functions, and in typical modem implementations, they act separately while sharing some common circuits. Also, some techniques have been investigated for joint carrier and timing recovery. Here we outline briefly the most commonly used carrier and timing recovery techniques.

A simple way to provide a reference carrier phase is to transmit a special tone, termed a *pilot tone*, and extract it at the receiver with a narrowband filter. The pilot tone extracted can then be used to synchronize the receiver’s local oscillator to the frequency and phase of the received signal. However, this method wastes a portion of the transmitted signal power and requires stringent narrowband filtering at the receiver. (It should be noted that certain digital signals having asymmetric spectra, such as single sideband, require the use of a pilot tone in demodulation [Luc68]. However, such modulation schemes are not employed in the wireless systems we consider in this book.) To avoid these problems, the carrier frequency and phase can be obtained from the received signal itself using a phase-locked loop (PLL), and this is the approach typically used in modem designs.

Designs for PLLs fall into three broad categories: squaring loops, Costas loops, and decision-directed feedback loops. The *squaring loops* approach is the simplest and can be applied to a BPSK signal by filtering the received signal to an intermediate frequency (IF) and passing the filtered output to a square-law device. The squaring operation removes the binary modulation and generates a frequency component at $2f_c$, twice the received carrier frequency. The frequency component at $2f_c$ is then filtered and used to drive a voltage-controlled oscillator (VCO), keeping it in phase with the arriving signal. The VCO output signal is then frequency-divided by 2 to produce a coherent phase reference at the carrier frequency. It must be noted that the frequency-divider output has a 180° phase ambiguity relative to the received carrier, which can be taken care of by differential precoding of the data bits before transmission.

A *Costas loop* eliminates the squaring operation (and some attendant design issues) by providing both in-phase and quadrature carrier-frequency signal components at the output of the VCO. The quadrature components are multiplied by the received signal, and those quadrature products are low-pass filtered and multiplied to form an error signal driving the VCO. As with the squaring loop, the Costas loop VCO has a 180° phase ambiguity relative to the received carrier, which can be taken care of by differential precoding of the data bits before transmission. For a squaring loop or Costas loop implemented with the same filter parameters, it has been shown that the performance of the two loops will be identical. Both squaring loops and Costas loops offer the advantage of relatively rapid phase acquisition and tracking as compared with decision-directed techniques, which we discuss next.

In *decision-directed carrier recovery*, sometimes referred to as *data-aided carrier recovery*, the basic Costas loop configuration is modified to utilize symbol decisions in forming the error signal driving the VCO. As long as the error rate in the symbol decisions is reasonably low, say $<10^{-4}$, the decision-directed method will provide performance superior to the squaring loop or strict-sense Costas loop. In some applications, at startup, use of decision-directed carrier recovery may require the use of an

initial training sequence. Thus this technique might not be preferred in situations where rapid carrier recovery is critical. However, in moderate-to-high SNR conditions, where the fed-back symbol decisions have a low error rate, the decision-directed loop is the preferred carrier recovery technique. Due to superior performance in the presence of noise and phase jitter, decision-directed carrier recovery methods have been adopted for virtually all voiceband modems designed for wireline applications. Detailed treatments of PLL theory and design can be found in a number of references, including [Lin73, Bin88, Pah88b, Pro01].

As stated earlier, proper recovery and tracking of symbol timing are critically important for reliable modem operation. Since carrier recovery and timing recovery are closely related functions, typical modem implementations use some form of PLL in performing both functions. As with carrier recovery, we can distinguish two categories of timing recovery methods, non-decision-directed and decision-directed. The simpler non-decision-directed methods typically apply a nonlinearity to the received signal, followed by a VCO or a voltage-controlled clock (VCC), which controls the sampling times for input to the control loop. Another non-decision-directed technique is the *early-late gate synchronizer*, which exploits the symmetry properties of the signal at the output of the matched filter [Pro01]. As in decision-directed carrier recovery methods, decision-directed timing recovery methods utilize previously detected data symbols in the estimation of optimum sampling times and achieve performance superior to that of non-decision-directed methods. Detailed treatments of timing recovery methods can be found in [Lin73, Bin88, Pah88b, Git92, Pro01], and in references cited therein. Considerable research has also been done on timing recovery techniques for use in modems utilizing adaptive equalization. For details the reader is referred to [Kob71a, Kob71b, Fra74, Lyo75, Maz75, Bin88, Pah88b].

An alternative to coherent modulation and demodulation is to use differentially coherent modulation and demodulation. In differentially coherent forms of modulation, it is not necessary that the carrier phase be known. This is because the data bits to be transmitted are mapped into successive differences in the transmitted channel symbols; this is called differential PSK (DPSK). A direct extension of DPSK is differential PAM. In demodulating the DPSK signal or the differential PAM signal, it is not necessary to recover the carrier phase, and thus the PLL is not needed. Rather, each symbol received is used as a reference for the succeeding symbol. This simplifies the receiver design in systems (e.g., some radio systems) where it is difficult to reliably recover the received carrier phase. However, simplification of the receiver design comes with a cost in performance. Differential detection will typically result in a performance loss of about 1 to 2 dB relative to coherent demodulation of the same signal. Recall, however, that symbol timing recovery is still required for demodulation of the differentially modulated signals. DPSK and other forms of differential modulation have been used for a number of years in both wireline and radio modems. Most modems currently in use in radio systems use differentially coherent forms of modulation. For example, the USDC TDMA system uses a differential form of QPSK, $\pi/4$ -shift DQPSK. However, there have been important advances in recent years in the use of coherent modulation techniques on radio channels.

7.5.3 Pulse Shaping

Next, we consider the pulse-shaping function shown in Fig. 7.29, split between the transmitter and receiver sections of a modem. In particular, we examine the choice

of pulse-shaping filter. If we were to choose an arbitrary shape for $f(t)$, a stream of PAM pulses might overlap one another such that a sample of an individual received pulse will be interfered with by many of the neighboring pulses. This effect is called *intersymbol interference* (ISI). Ideally, we would like $f(t)$ to be rectangular over the pulse interval $(0, T)$. However, the sidelobes of the spectrum of that pulse shape, which has the functional form $\sin(\pi_f T)/\pi_f T$, decrease slowly with frequency, and this can lead to unacceptable signal interference in adjacent user channels. Instead, we prefer a pulse shape $f(t)$ that has its peak at time $t = 0$ and value zero at sampling times $t = T, 2T, 3T, \dots$. Pulses having this characteristic are referred to as *Nyquist pulses*, and filters having impulse responses with this characteristic are called *Nyquist filters*. With transmission of Nyquist pulses at uniform intervals of T seconds, each received pulse can be sampled free of intersymbol interference from other transmitted pulses. A widely studied class of Nyquist filters is the class of *raised-cosine filters*. The frequency response of a raised-cosine filter has a flat amplitude portion and sinusoidal roll-off to zero. The raised-cosine roll-off characteristic is one that can be realized without difficulty in a practical design. Raised-cosine filters have been used extensively in modems designed for both wireline and radio systems.

The spectral characteristic of a raised-cosine filter is given by

$$F\{f(t)\} = \begin{cases} T, & 0 \leq |f| \leq \frac{1-\beta}{2T} \\ \frac{T}{2} \left[1 - \sin \frac{\pi T}{\beta} \left(|f| - \frac{1}{2T} \right) \right], & \frac{1-\beta}{2T} \leq |f| \leq \frac{1+\beta}{2T} \end{cases} \quad (7.5.1)$$

where β , the *roll-off factor*, can range between 0 and 1. The corresponding time-domain Nyquist pulse is

$$f(t) = \frac{\sin \pi t/T}{\pi t/T} \times \frac{\cos \beta \pi t/T}{1 - 4\beta^2 t^2/T^2} \quad (7.5.2)$$

A few examples of raised-cosine frequency responses and their corresponding impulse responses are shown in Fig. 7.30 for selected values of the roll-off parameter β . As can be seen in the figure, small values of β yield the sharpest spectral roll-off characteristics, with $\beta = 0$ corresponding to a rectangular spectrum. The roll-off value $\beta = 1$ eliminates the flat portion of the spectrum and yields a pure raised-cosine shape. Although it is the frequency-domain characteristic that has the raised-cosine shape, the corresponding time-domain impulse responses are often called *raised-cosine pulses*.

In practice, a pair of matched filters is used to implement the raised-cosine spectrum, one each at the transmitter and receiver. The two filters have the same the same spectrum shape, the square root of the raised-cosine spectrum. Thus, Eq. (7.5.2) represents the overall impulse response used for analysis of the system. For design purposes the impulse response of the square root of the raised-cosine frequency function is needed, which is given by

$$f_2(-t) = f_1(t) = \frac{\sin[\pi(1-\beta)t] + 4\beta t \cos[\pi(1+\beta)t]}{\pi[1 - (4\beta t)^2]} \quad (7.5.3)$$

The simplest approach to digital implementation of this filter is to use the samples of this function as the discrete impulse response of a finite impulse response (FIR), discrete, pulse-shaping filter. To make the samples finite and to control the sidelobes

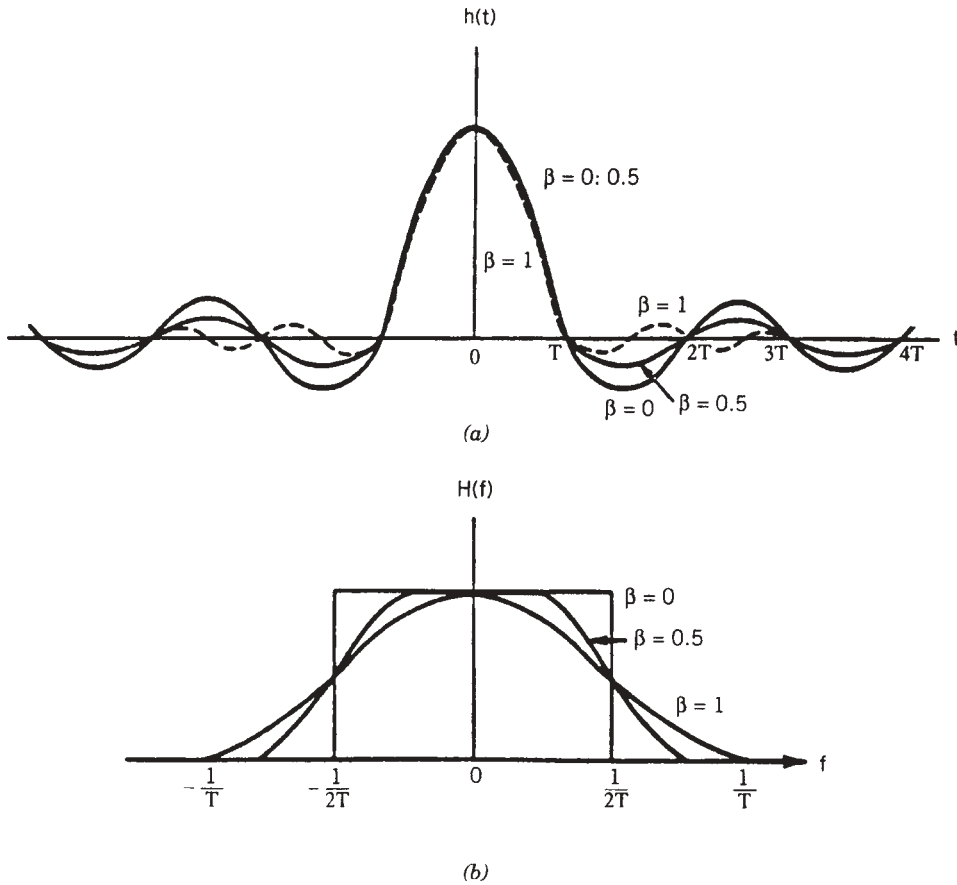


FIGURE 7.30 (a) Time- and (b) frequency-domain representations of raised-cosine pulses.

of the spectrum, a window is applied to the time-domain representation. In voiceband modems a roll-off factor of around 0.2 with 20 to 30 taps is typical, and the sidelobes are designed to be more than 40 dB below the main lobe. For radio modems, analog filters with higher roll-off factors and lower sidelobes are used.

The maximum bandwidth efficiency of 1.0 symbol/s per hertz for QAM modulation corresponds to the case of $\beta = 0$, which is not practically feasible. To preserve the same bandwidth efficiency with a realizable filter, one may use *partial response signaling*. The channel impulse response of a partial response system has the value 1.0 for two consecutive samples spaced by T seconds, and 0 at all other samples. As a result, in the absence of channel distortions the received samples are $\{a_k + a_{k-1}\}$ rather than $\{a_k\}$. This can be viewed as known “forced” ISI, which creates no problem for detection as long as the decisions are based on the last two received sampled pulses. The frequency-domain representation of the half-cosine pulse-shaping filters used for pulse shaping in partial-response systems is given by

$$F\{f(t)\} = \begin{cases} 2T \cos \pi T f, & |f| \leq 1/2T \\ 0, & |f| \geq 1/2T \end{cases} \quad (7.5.4)$$

The time-domain representation for this waveform is

$$f(t) = \frac{\pi}{4} \frac{\cos \pi t/T}{1 - 4t^2/T^2} \quad (7.5.5)$$

More extensive discussion of partial response signaling, with pertinent references, may be found in [Feh87, Pro01].

Several approaches have been used for the optimal design of digital pulse-shaping filters. Digital linear-phase FIR filters are discussed in [Mul73] with special attention to zero ISI and minimum stopband attenuation. An iterative technique using the steepest-descent algorithm to design a pair of zero-ISI matched filters with maximum spectral power in the passband is available in [Che82]. Other methods, using linear programming [Sal82, Lia85] and modified Remez exchange algorithm [Rab78], are also available in the literature.

QUESTIONS

- (a) What is a matched filter, and how does it help in digital communication over an additive white Gaussian noise channel? What is the signal-to-noise ratio of the sampled output of a matched filter, and how is it related to the energy of the transmitted pulse and the variance of the background noise?
- (b) What are the two most popular functions used in calculation of the error rate for a given signal-to-noise ratio? How are they related to one another mathematically?
- (c) Sketch the signal constellations for on–off keying and binary PSK, letting the signal amplitude be A volts in each case. Using these diagrams, explain the difference in the P_b formulas for these modulation methods, as shown in Eq. (7.2.13).
- (d) To maintain approximately the same error rate with the addition of one bit per symbol to a PAM system, how much increase in transmission power is needed? What is the additional power for transmission of one additional bit per symbol in a QAM system?
- (e) What is trellis-coded modulation, and how does it change the signal constellation and performance of a modem?
- (f) Give an equation for calculation of the error rate of the $\pi/4$ -QPSK modulation scheme in AWGN.
- (g) Why are constant-envelope modulation techniques preferred for use on radio channels?
- (h) Why is bandwidth efficiency important for WIN systems?
- (i) Why is power efficiency important for WIN systems?
- (j) Explain why binary and 4-ary FSK modulations exhibit the same bandwidth efficiency.
- (k) Why was GMSK modulation adopted for several leading WIN systems?
- (l) What is the major advantage of OQPSK modulation relative to QPSK?

- (m) Explain why the signal constellation for $\pi/4$ -QPSK modulation, shown in Fig. 7.26, has eight points rather than the four points of standard QPSK. How many points do we expect in the signal constellation of OQPSK?
- (n) Discuss the relative advantages and disadvantages of $\pi/4$ -QPSK and GMSK modulation.
- (o) Why are multi-amplitude and multiphase modulation techniques not very popular in wide-area cellular wireless networks but used in WLAN and WPAN systems?

PROBLEMS

1. The standard pulse shape used in most short-distance cable communication applications such as RS232 is a rectangular pulse.
 - (a) The matched filter receiver for the rectangular pulse transmission is usually implemented with an integrate-and-dump circuit. Prepare a block diagram for this receiver, and explain how it works and why it is a matched filter.
 - (b) If the voltage used for the amplitude of the pulses is $\pm A$ volts, the data rate is R , and the variance of the received noise is N_0 , determine the signal-to-noise ratio after sampling at the receiver.
 - (c) An integrate-and-dump circuit can be implemented with a simple RC low-pass filter and a switch. Sketch the circuit diagram for this matched filter receiver.
2. Draw the signal constellations for binary PSK and QPSK modulation. Using the figure and the formula for binary PSK bit error probability P_b in AWGN, derive the P_b formula for QPSK.
3. Show how two binary PSK transmissions can operate simultaneously in the same radio channel by using two carriers at the same frequency in phase quadrature. Draw a block diagram for transmitters and receivers. Give the overall data rate for this quadrature carrier system as a function of channel bandwidth available.
4. Consider the quadrature carrier system described in Problem 3. Let the data rates on the quadrature carriers be 2400 and 9600 bits/s, and assume that the system is operating in AWGN. Determine the relative amplitudes of the two quadrature carriers that will provide the same value of E_b/N_0 on each quadrature channel.
5. Assume that we have a BPSK modem operating on a wireless voiceband channel with a symbol transmission rate of 2400 symbols/s and a bandwidth efficiency of 1. We want to increase the data rate to 19.2 kb/s.
 - (a) If we increase the number of points in the constellation until the data rate becomes 19.2 kb/s while the baud rate remains at 2400 symbols/s, what is the number of points in the constellation? What is the bandwidth efficiency of the modulation technique? What is the additional power requirement for the transmitter to keep the quality the same as before? The approximations used in Section 7.2 can be applied.
 - (b) Repeat part (a) if we use a trellis-coded modulation that doubles the number of points in the constellation and has an overall coding gain of 3 dB.

- (c) If we increase the symbol transmission rate to 19,200 and use the same BPSK modulation, what is the additional power requirement for the transmitter to maintain the same quality as a 2400-bit/s modem?
6. IEEE 802.11a and g use multiple modulation techniques in the same transmission bandwidth to provide different data rates. When the mobile terminal is close to the access point, a 64-QAM modulation is used, and as the modem goes to the coverage boundary of the access point, a BPSK modulation is used that requires substantially lower received signal strength to operate.
- (a) If the data rate for the BPSK system is 12 Mb/s, what is the data rate of the 64-QAM modem?
- (b) What is the difference between the received signal strength requirement of the 64-QAM and BPSK modulation techniques? Approximations used in Section 7.2 can be applied.
- (c) If the coverage with 64-QAM is D meters, what is the coverage with a BPSK modem when we operate in a large indoor open area with a distance–power gradient of $\alpha = 2$?
- (d) Repeat part (b) for an indoor office area with a distance–power gradient of $\alpha = 3$.
7. Consider the 16-PSK constellation and the rectangular 16-QAM constellation shown in Fig. P7.1. We have two transmitters, each using one of these constellations. In order to have approximately the same performance (error rates) at the receiver, what should be the difference (in decibels) in the transmitted power for the two transmitters? What are the advantages and disadvantages of 16-PSK versus 16-QAM modems?

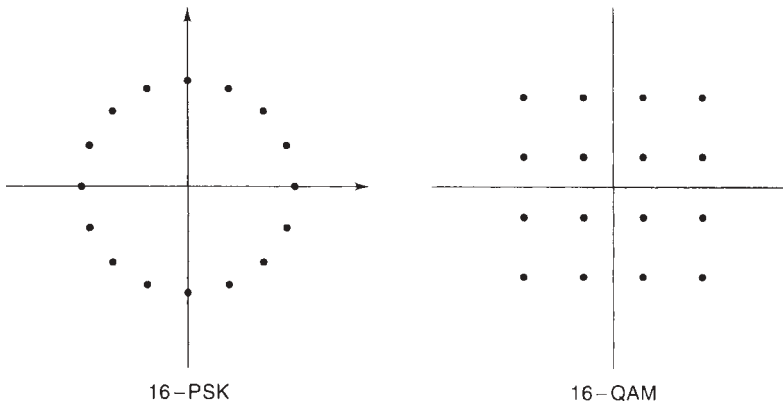


FIGURE P7.1 Signal constellations of 16-PSK and 16-QAM.

8. Suppose that you are asked to design a 16.8-kb/s constellation for a wireless voice-band modem. The symbol rate is 2400 symbols/s and the SNR received is 28 dB.
- (a) For an uncoded constellation, find the alphabet size and select a constellation. Give your reasons for the selection. Using the asymptotic bound, give the symbol error rate for the modem.
- (b) Repeat part (a) for a trellis-coded constellation with a 3-dB coding gain.

9. (a) Use MATLAB to plot the three formulas in Eq. (7.2.14) for $M = 16$. The format of the plots should be similar to Fig 7.2. Find the γ_b values for which each plot provides an error rate of 10^{-5} .
- (b) Plot Eq. (7.2.15) for $M = 16$, and determine the signal-to-noise ratio for which each plot provides an error rate of 10^{-5} .
- (c) Compare your results with those given in Figs. 7.13 and 7.14. If there are discrepancies, explain them.
10. Consider the MSK version of coherent FSK modulation, and assume that transmission begins with an initial phase of π radians. Draw a phase-state diagram for MSK and determine the terminal phase for each of the following pairs of input data bits: (a) 00; (b) 01; (c) 10; (d) 11.
11. (a) Using Poisson's sum formula

$$\sum_{k=-\infty}^{+\infty} p(t - kT) = \frac{1}{T} \sum_{m=-\infty}^{\infty} P\left(\frac{2\pi}{T}m\right) e^{j(2\pi mt/T)}$$

where $P(\omega)$ is the Fourier transform of $p(t)$, show that

$$\sum_{m=-\infty}^{+\infty} |p(t - mT)|^2 = \frac{1}{T} \sum_{n=-\infty}^{+\infty} Z_n e^{j(2\pi nt/T)}$$

where

$$Z_n = \frac{1}{2\pi} \int_{-\infty}^{+\infty} P(\omega) P^*\left(\omega - n\frac{2\pi}{T}\right) d\omega$$

- (b) Show that $Z_{-n} = Z_n^*$ for all n .
- (c) For $p(t)$, a raised cosine pulse with $0 < \alpha < 1$, show that

$$\sum_{m=-\infty}^{+\infty} |p(t - mT)|^2 = Z_0 + 2\text{Re}(Z_1 e^{j2\pi t/T})$$

PROJECTS

Project 1: Error Rate and Phase Jitter in QPSK Modulation

- (a) Sketch a typical QPSK signal constellation and assign the 2-bit binary codes to each point in the constellation. Define decision lines for the received signal constellation so that the receiver can distinguish received noisy symbols from each other.
- (b) In Section 7.2 we stated that the probability of symbol error for a multi-amplitude, multiphase modem with coherent detection can be approximated by $P_s = 0.5 \text{erfc}(d/2\sqrt{N_0})$, where d is the minimum distance between the points in the constellation and N_0 is the variance of the additive Gaussian noise. Use this equation to calculate the probability of error for the QPSK modem. Observe that if we consider the signal constellation and the decision lines of part (a), this

equation can be modified to $P_s = 0.5 \operatorname{erfc}(\delta/8\sqrt{N_0})$, where δ is the minimum distance of a point in the constellation from a decision line.

- (c) Use MATLAB or an alternative computation tool to plot the probability of symbol error versus signal to noise ratio in dB. What are the signal to noise ratios (in dB) for the probabilities of symbol error of 10^{-2} and 10^{-3} ? Let us refer to these two SNRs as SNR-2 and SNR-3.
- (d) Simulate transmission of the QPSK signal corrupted by additive Gaussian noise for 10,000 transmitted bits. Generate random binary bits and use each two bits to select a symbol in the constellation of part (a), add complex additive white Gaussian noise to the symbol so that the signal to noise ratio in dB is SNR-2, and use the decision lines to detect the symbols. Find the number of erroneous symbol decisions and divide it by the total number of symbols to calculate the symbol error rate. Compare the error rate with the expected error rate of 10^{-2} .
- (e) Repeat part (d) for SNR-3 and error rate of 10^{-3} .
- (f) Assume that a channel produces a fixed phase error θ . Give an equation for calculation of the probability of error for a QPSK modem operating over this channel. Use the minimum distance from a decision line, δ , and Eq. $0.5 \operatorname{erfc}(\delta/8\sqrt{N_0})$ for the calculation.
- (g) Assume that the received SNR is 10 dB, and sketch the probability of error versus the phase error $0 < \theta < \pi/4$.
- (h) Repeat parts (c), (d), and (e) for a channel with a phase error of $\theta = \pi/8$.

Project 2: Error Rate and Phase Jitter in 16-QAM Modulation

- (a) For a 16-QAM modem, use MATLAB to plot the probability of symbol error, P_s , versus E_s/N_0 , where N_0 is the variance of the noise and E_s is the average energy per transmitted symbol.
- (b) Repeat part (a) for a 10° phase error at the receiver, and compare the results with those of part (a). What are the SNRs (in dB) for the probabilities of symbol error of 10^{-2} and 10^{-3} ?
- (c) Repeat parts (a) and (b) for 64-QAM.

Project 3: Design of Raised Cosine Matched Filters

The impulse response of a pair of matched filters that results in a raised cosine spectrum is given in Eq. (7.5.2). A simple way to design these filters is to window the sampled version of the filter impulse response and design an FIR filter with the windowed sampled impulse response. In this project we examine the time and frequency response of this approach for a particular design specification. Assume that the rolloff factor of the raised cosine pulse is 0.1, the length of each FIR filter is 23 taps, and the sampling rate is $T/4$, with $1/T$ the transmission rate of the pulses.

- (a) If the design uses a rectangular window, sketch the overall back-to-back impulse response of the transmitter and receiver filters. What is the variance of the ISI caused by the filter if the center tap is normalized to 1?

- (b) Using MATLAB or an alternative computation tool, calculate and plot the frequency response of the filter. What is the minimum attenuation in the sidelobes and the percentage of power in the main lobe?
- (c) Repeat parts (a) and (b) for a triangular window.
- (d) If the design criterion is to minimize the out-of-band component of the spectrum, which window is your choice? If the design criteria is to minimize the ISI, which window is your choice?

8

FADING, DIVERSITY, AND CODING

- 8.1 Introduction
- 8.2 Radio Communication on Flat Rayleigh Fading Channels
- 8.3 Diversity Combining
 - 8.3.1 Performance Evaluation
 - 8.3.2 Special Cases
- 8.4 Error-Control Coding for Wireless Channels
 - 8.4.1 Error-Detection and FEC Block Coding
 - 8.4.2 Convolutional FEC Coding
 - 8.4.3 ARQ Schemes
 - 8.4.4 Effects of Fading on Coding Performance
- 8.5 Space-Time Coding
- 8.6 MIMO and STC
 - 8.6.1 Design of Codes for MIMO Systems
 - 8.6.2 Capacity Limits for MIMO Systems
 - 8.6.3 Practical Considerations for MIMO Systems
- Questions
- Problems
- Projects
 - Project 1: Error Rate on a Fading Channel for QPSK Modulation
 - Project 2: MRC for BPSK Modulation

8.1 INTRODUCTION

Having discussed the principal modem techniques in Chapter 7, in this chapter we start by considering the limitations encountered in applying these techniques to fading-multipath radio channels. The fundamental problem to be dealt with here is multipath, which causes power fluctuations, frequency-selective fading, and multipath delay spread in the received signal. The signal fluctuations cause an increase in the signal power required, relative to steady-signal operation, to achieve the same overall BER performance. If it occurs in the midregion of the band, frequency-selective fading, can

disable proper operation of the modem. Time dispersion of the signal due to multipath puts a limit on the speed at which modulated symbols can be transmitted in the channel. In this chapter we discuss the effects of power fluctuations, and then we explain diversity and coding techniques as approaches to counteracting the effects of power fluctuations. More specifically, we examine the effects imposed on the performance of standard modem techniques by the characteristics of flat Rayleigh fading radio channels. We then discuss the effects of diversity combining methods and coding techniques on the performance of radio modems operating on these channels. In Chapter 9 we discuss the effects of frequency-selective fading and the methods used to increase the data rate and improve performance in the presence of frequency-selective fading.

Our analysis here is based on flat Rayleigh fading, the statistical model most commonly used to describe the behavior of fading on radio channels, a model for which some closed-form solution and simple approximations have been derived. These simple approximations are helpful in gaining an intuitive understanding of the effects of fading on the performance of a modem and how diversity and coding help to improve the performance in fading. In the general analysis of the performance of modems in additive white Gaussian noise, described in Chapter 7, the assumed performance criterion was the symbol or bit error rate. There we related the error rate for different modulation techniques to the received signal-to-noise ratio (SNR), which was assumed to be a fixed value. In fading channels the received SNR is a random variable resulting in a bit error rate that is also a random variable. As a result, the performance criterion commonly used for fading channels is either the average error rate over all possible SNR values, or the probability of the error rate exceeding a specified threshold value, and we refer to this as the *probability of outage*.

In this chapter our analysis begins with Section 8.2, which presents a mathematical framework for performance evaluation of communication over a flat Rayleigh fading channel. Since the relation between the error rate and any given value of the received SNR remains the same as in the steady-signal case (Chapter 7), we show how these results can be used for calculation of the average error rate and the probability of outage. Using these results, we demonstrate how multipath fading causes significant performance degradation. In Section 8.3 we discuss diversity techniques and their effectiveness in counteracting the performance degradations caused by multipath fading. The simplest form of diversity is to increase the number of antennas at the receiver. Since these antennas receive signals affected by different multipath conditions, the fading patterns of the received signals are different, providing for diversity in the received signal. In this section we show how a receiver can take advantage of the diversified received signal to improve overall performance. In Section 8.4 we describe how different coding techniques are used to improve performance in fading. On a fading channel, most errors occur when the signal goes into a deep fade. If we scramble the signal and add coding, we can recover from many of these errors. In this section we provide an overview of selected coding techniques and how they improve performance over fading channels. In Section 8.5 we describe *space-time coding* (STC), a relatively recent coding technique being used for performance enhancements to mobile radio modems. The STC technique, as its name suggests, combines coding with antenna design. In Section 8.6 we discuss the combined use of STC and MIMO techniques. Multiple-input multiple-output antenna systems are being given much attention in the development of emerging high-capacity wireless networks. The use of STC and MIMO,

used with diversity combining techniques, can yield a considerable improvement in communication performance over that of fading channels.

8.2 RADIO COMMUNICATION ON FLAT RAYLEIGH FADING CHANNELS

Here we consider the case of frequency-nonselective or flat-fading channels. As the name implies, *flat fading* is a form of fading in which all the frequency components of the transmitted signal rise and fall in exact unison. Let us recall the discussion in Chapter 3 of the classical uncorrelated scattering model of a multipath fading channel, where we defined the root mean square (rms) delay spread of the channel, τ_{rms} , and referred to its reciprocal as the coherence bandwidth of the channel. We noted there that the symbol transmission rate should be much smaller than the coherence bandwidth of the channel if multipath distortion of transmitted symbols is to be made negligible. The assumption of flat fading, therefore, is simply the assumption that the transmission bandwidth is significantly smaller than the coherence bandwidth of the channel. If the contrary were true, and the transmission bandwidth were comparable to or wider than the coherence bandwidth, we would describe the fading model as frequency selective or nonflat fading.

In pulse transmission under the assumption of flat fading, we can characterize the received sampled signal at the output of the matched filter, shown in Fig. 7.1, as

$$z(T) = a_i \alpha \sqrt{E_s} + \varepsilon \quad (8.2.1)$$

where α is the channel gain factor imposed on the signal by flat fading and $\{a_i\}$ is the set of transmitted information symbols, for example, for BPSK modulation, $a_i = \pm 1$. Note that in this expression the additive noise ε is assumed to be unaffected by channel fading. In the flat-fading model, the channel gain factor α is a random variable that is described completely by a probability density function $f_A(\alpha)$. As noted in earlier chapters, many fading radio channels are accurately characterized by the Rayleigh model of fading. This is analytically convenient, because closed-form solutions for calculation of average error rates are readily available for a number of common modulation techniques. With Rayleigh fading, the probability density function of the magnitude of α is given by the Rayleigh distribution

$$f_A(\alpha) = \frac{\alpha}{\Gamma} e^{-\alpha^2/2\Gamma} \quad (8.2.2)$$

where Γ is the mean-squared amplitude of the channel gain factor. In this situation, the SNR per bit for BPSK modulation is given by

$$\gamma_b = \frac{|\alpha|^2 E_b}{N_0}$$

In contrast with data communication over wireline circuits, here the γ_b is a random variable that changes with time or spatial movements of the transmitter or receiver. Because E_b and N_0 have fixed values, the probability density function of γ_b will follow the probability density function of $|\alpha|^2$. Given that the channel gain factor $|\alpha|$ has a

Rayleigh distribution, $|\alpha|^2$ has an exponential distribution, the chi-square distribution with two degrees of freedom [Pro01]. The probability density function of γ_b is then given by

$$f_{\Gamma}(\gamma_b) = \frac{1}{\gamma_b} e^{-\gamma_b/\bar{\gamma}_b} \quad (8.2.3)$$

As we discussed earlier, there are two performance criteria for digital communication over fading channels: probability of outage and average probability of error. *Probability of outage* is the probability that the modem performs more poorly than a specified threshold. The threshold for most digital communications applications is usually defined by a certain error rate, which we will call $P_{e,th}$. For a given modulation technique, we may use the error probability formula or the error rate curve for a nonfading channel to determine the corresponding value of γ_b , which we refer to as γ_{th} .

Figure 8.1 shows the basic parameters related to performance evaluation of modems over flat-fading channels. Due to the fading effect, the signal-to-noise ratio, γ_b , fluctuates in time randomly with a Rayleigh distribution. When γ_b crosses the specified threshold, γ_{th} , the error rate drops below the acceptable error rate of $P_{e,th}$. The outage probability is the fraction of time during which the error rate is unacceptable. The average SNR is related to the transmitted power, and it can be adjusted by changing the transmitted power. The error rate when the signal is above the threshold is always very small (close to zero), and when it is below the threshold it is very high (close to 0.5). Therefore, most errors occur during deep fades, when the signal level crosses the threshold. This is a very important observation, because as we show later in the chapter, to remedy the effects of fading, we need to find methods that can recover bits corrupted by intervals of deep fading.

The *average error rate* and the outage rate may look different, but they represent the same phenomenon and in many cases they look similar. If we define the average error rate below and above the outage threshold by $\overline{P_{b,above}}$ and $\overline{P_{b,below}}$, respectively, the average error rate and the probability of outage are related by

$$\overline{P_b} = (1 - P_{out}) \overline{P_{b,above}} + P_{out} \overline{P_{b,below}}$$

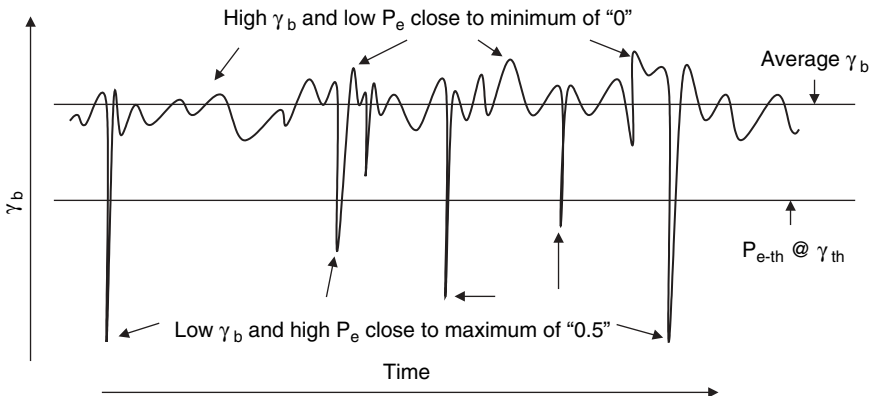


FIGURE 8.1 Relation among error rate, outage rate, and fading characteristics.

Under the condition that outage rate and average error rates above the threshold are very small and the average error rate below the threshold is close to $\frac{1}{2}$, we have

$$\overline{P_b} \approx P_{\text{out}} \overline{P_{b \text{ below}}} \approx 0.5 P_{\text{out}}$$

which demonstrates the direct relationship between the two performance measures. However, this relationship does not hold at all times, and thus we use the more general formulation that follows.

If the error probability for the modulation technique used on a nonfading channel is described by the general exponential function

$$P_e = A e^{-B\gamma_b} \quad (8.2.4)$$

where $\gamma_b = E_b/N_0$ is the received SNR per bit if the channel is nonfading, we have

$$\gamma_{\text{th}} = -\frac{1}{B} \ln \frac{P_{e,\text{th}}}{A} \quad (8.2.5)$$

The probability of outage P_{out} is the probability that γ_b , having the probability density function given by Eq (8.2.3), is less than γ_{out} :

$$P_{\text{out}} = \int_0^{\gamma_{\text{th}}} f_{\Gamma}(\gamma_b) d\gamma_b = 1 - e^{-\gamma_{\text{th}}/\overline{\gamma}_b} = 1 - \left(\frac{P_{e,\text{th}}}{A} \right)^{1/B\overline{\gamma}_b} \quad (8.2.6)$$

If the error probability for the chosen modulation method on a nonfading channel is given by

$$P_e = A \operatorname{erfc}(\sqrt{B\gamma_b}) \quad (8.2.7)$$

the inverse mapping to P_{out} , similar to Eq. (8.2.5), is not analytically feasible. In this case either the asymptotic exponential bound for erfc can be used in conjunction with Eq. (8.2.7) or plotted curves of the error probability can be used for numerical inverse mapping. The value γ_{th} determined from a plotted curve is then substituted into the integral over $f_{\Gamma}(\gamma_b)$ in Eq. (8.2.6). Figure 8.2 shows the probability of outage versus threshold for several modulation techniques.

Now let us consider the average probability of error. Given our assumption of the flat Rayleigh fading model and fixed noise level, the received SNR per bit is a random variable with the exponential probability density function of Eq. (8.2.3). Therefore, we can find the average probability of error in fading by averaging P_e given by Eq. (8.2.4) or Eq. (8.2.7) over the probability density function of γ_b , given by Eq. (8.2.3). Both integrals have closed-form solutions. For the bit error probability of the form of Eq. (8.2.7), this averaging is given by the integration

$$\overline{P_b} = P(\overline{\gamma}_b) = \frac{A}{\overline{\gamma}_b} \int_0^{\infty} \operatorname{erfc} \sqrt{B\gamma_b} e^{-\gamma_b/\overline{\gamma}_b} d\gamma_b = A \left[1 - \sqrt{\frac{B\overline{\gamma}_b}{1+B\overline{\gamma}_b}} \right] \simeq \frac{A}{2B\overline{\gamma}_b} \quad (8.2.8)$$

For the bit error probability of the form of Eq. (8.2.4), the averaging is given by the integration

$$\overline{P_b} = P(\overline{\gamma}_b) = \frac{A}{\overline{\gamma}_b} \int_0^{\infty} e^{-B\gamma_b} e^{-\gamma_b/\overline{\gamma}_b} d\gamma_b = \frac{A}{1+B\overline{\gamma}_b} \simeq \frac{A}{B\overline{\gamma}_b} \quad (8.2.9)$$

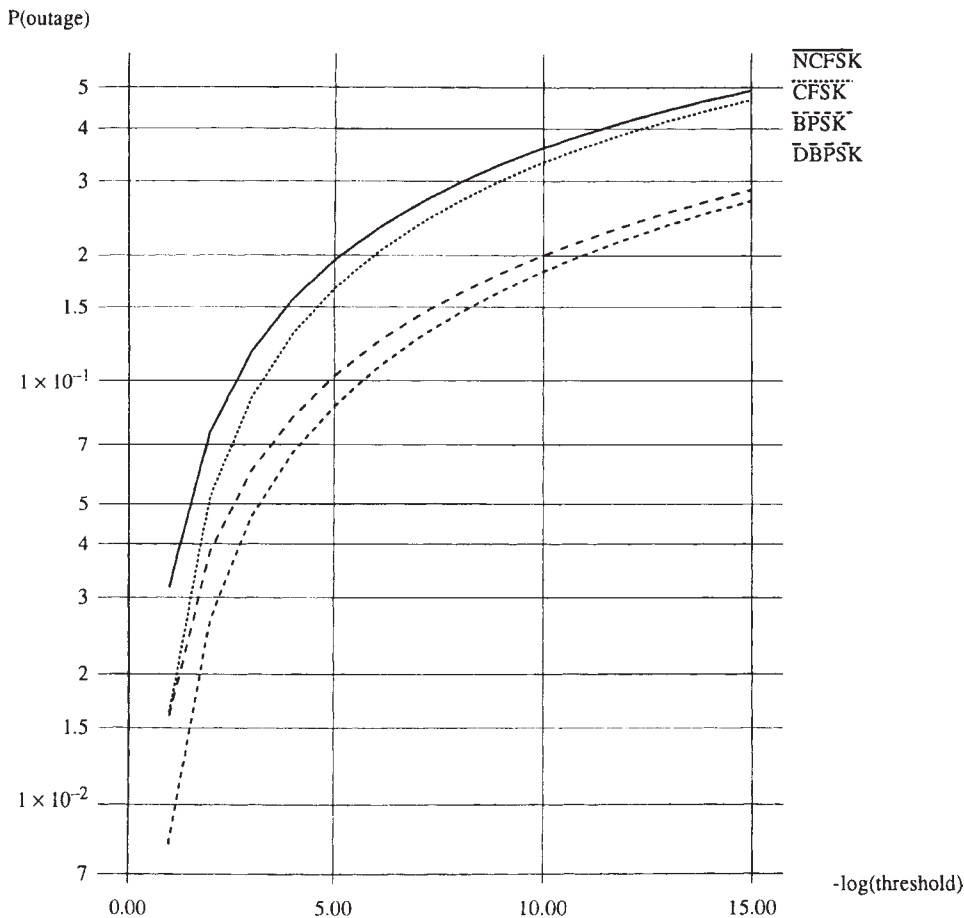


FIGURE 8.2 Probability of outage versus threshold for basic modulation methods.

which gives a 3-dB poorer performance than that of Eq. (8.2.8). In both cases the average error probability is reduced by only one decade per 10 dB of increase in $\bar{\gamma}_b$. This is to be contrasted with the exponential reduction of error probability with increasing $\bar{\gamma}_b$ on nonfading channels. This clearly indicates the need for substantial additional power to provide the same average error probability on fading versus nonfading channels. This increase in required signal power is referred to as the *fade margin*.

Figure 8.3 shows the probability of bit error versus SNR per bit for coherent BPSK modulation for both the nonfading and flat Rayleigh fading cases. As can be seen from the figure, the presence of signal fading causes a large increase in the SNR required to achieve reasonable levels of BER. For example, to achieve BER equal to 10^{-5} , we require about 10 dB on a nonfading channel but require nearly 45 dB SNR in fading, a penalty of 35 dB. The reason for this large SNR penalty can be understood by examining the probability density function for the received signal power in fading, given by Eq. (8.2.3), where the exponential form of the distribution places some of the signal power at very low levels, where the instantaneous BER is near 0.5. For higher levels of signal power, the error rates are negligible. Therefore, the average BER is

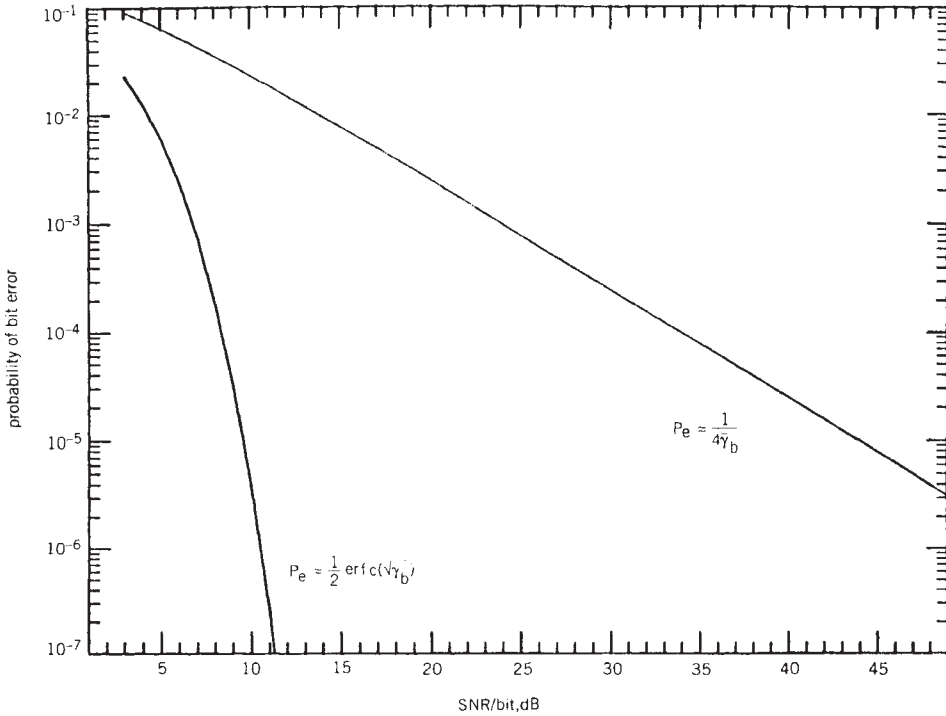


FIGURE 8.3 Probability of bit error for BPSK modulation on nonfading (left-hand curve) and fading (right-hand curve) channels.

dominated by the intervals of time in which the SNR is low and the BER is high. As a result, a large increase in average signal power, which greatly widens the distribution $f_{\Gamma}(\gamma)$, is needed to reduce the probability that the instantaneous signal power lies in the region of high BER.

Because averaging over Rayleigh fading involves integration over an exponential power distribution function, Laplace transform tables can be used for evaluation of the integral. Laplace transform tables are bountiful, and a wide variety of closed forms can be found for application to many different modulation techniques. As a result, most closed-form solutions available in the communications literature have been derived for Rayleigh fading channels. In the next section we provide a selection of these derivations that are widely used in different applications to obtain formulas for performance of various modulation techniques over fading channels.

8.3 DIVERSITY COMBINING

As we observed in earlier chapters, multipath fading is manifested as signal amplitude fluctuations over a wide dynamic range. In particular, during short periods of time, the channel goes into deep fades, causing a significant number of errors that virtually dominate the overall average error rate of the system. To compensate for the effects of fading when operating with a fixed-power transmitter, the power must typically be increased by several orders of magnitude relative to nonfading operation. This increase

in power protects the system during the short intervals of time when the channel is deeply faded. A more effective method of counteracting the effects of fading is to use diversity techniques in transmission and reception of the signal. The concept here is to provide multiple received signals whose fading patterns are different. With the use of diversity, the probability that all the received signals are in a fade at the same time reduces significantly, which in turn can yield a large reduction in the average error rate of the system.

Figure 8.4 shows fluctuations in two branches of a diversity channel and how they help the reduction of overall average error rate and outage probability. When one of the branches is in deep fade, causing a large number of errors, the correct data can be retrieved from the other branch. In a diversity channel, large numbers of errors can occur when all branches are in deep fade at the same time. Since the probability of a deep fade occurring in all branches is much lower than that in only one branch, the error rate on a diversity channel is much less than on a single-branch fading channel. The occurrence of deep fading on all branches is a function of correlation among different branches and of the number of diversity channels. As the correlation among the diversity branches decreases and the number of branches increases, the error rate decreases.

Diversity can be provided spatially by using multiple antennas, in frequency by providing signal replicas at different carrier frequencies, or in time by providing signal replicas with different arrival times. It is conventional to refer to the diversity components as *diversity branches*. We assume that the same symbol is received from different branches, with each branch exposed to a separate random fluctuation. This has the effect of reducing the probability that the received signal will be faded simultaneously on all the branches; this in turn reduces the overall outage probability as well as the average BER.



FIGURE 8.4 Fading in two branches of a diversity channel.

A variety of techniques are available for reception of the diversity signals. With *selection diversity*, one signal is chosen from the set of diversity branches, usually on the basis of received signal strength. With *linear combining*, as the name suggests, the diversity branches are simply summed together before demodulation. In the optimum method of combining, called *maximal-ratio combining*, the diversity branches are weighted prior to summing them, each weight being proportional to the received branch signal amplitude. The maximal-ratio combiner for the diversity channel can be considered equivalent to a discrete matched filter receiver, in the sense that it provides the optimum postdemodulation SNR for the received signal, which in this case is made up of diversity components.

8.3.1 Performance Evaluation

In this section we provide an analytical framework for calculation of the error probability achieved by the use of diversity reception. The channel is assumed to be a Rayleigh fading channel, and most of the derivations are based on the use of maximal-ratio combining (MRC). In later chapters we describe various innovative techniques for providing diversity with different modulation methods. However, the equations used for performance calculations will be those introduced in this section. As a basis for discussing the performance improvements obtained through diversity, let us assume that we have a Rayleigh fading channel and are using diversity of order D ; that is, the transmitted signal is arriving from D independent diversity branches, each equipped with a matched filter. The set of sampled signals received at one instant of time from the diversity branches, after sampling at the output of the matched filters, is given by the following vector equation:

$$z_j(T) = a_i \alpha_j \sqrt{E_s} + \varepsilon_j, \quad 0 < j \leq D \quad (8.3.1)$$

where E_s is the energy of the transmitted symbol, α_j is the amplitude fluctuation in the j th branch caused by flat fading in that branch, and ε_j is the AWGN associated with the j th branch. With the use of MRC, the received signal in each branch is scaled by the amplitude of that branch, α_j , and the scaled signals are added to form $z(T)$, the sample used for decision making:

$$z(T) = a_i \sum_{j=1}^D |\alpha_j|^2 \sqrt{E_s} + \sum_{j=1}^D \alpha_j^* \varepsilon_j \quad (8.3.1a)$$

Figure 8.5 is the block diagram of a typical MRC receiver for a single transmitter antenna, and two receiver antennas operating in a flat-fading channel. The transmitted signal is exposed to two different flat-fading gains. Before combining the signals from the two branches, the received signal after matched filtering is scaled with the strength of the channel to form the decision variable. To detect the transmitted symbol using a maximum likelihood receiver, the distance of the received symbol from all the points in the constellation is determined and the symbol with minimum distance is declared to be the detected symbol. In cellular applications it is much more convenient to add multiple antennas at the base station than in the mobile terminal, due to obvious space limitations, particularly in handsets.

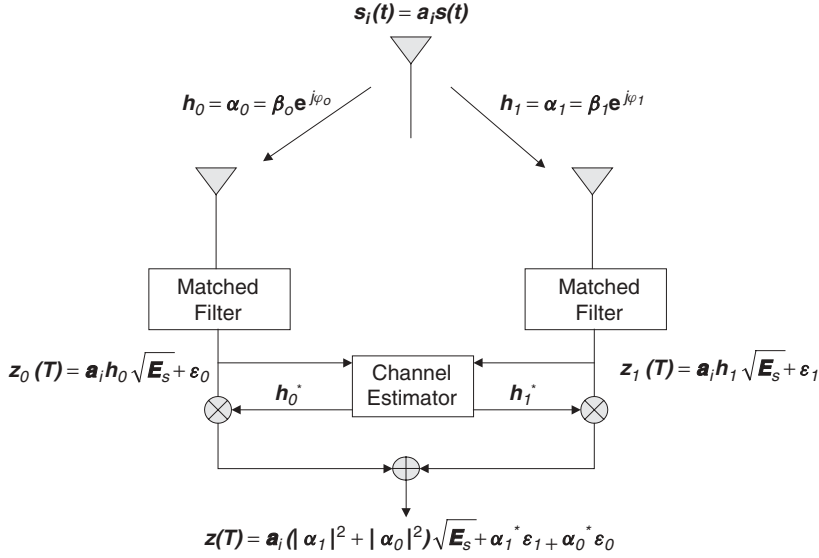


FIGURE 8.5 Block diagram of an MRC receiver with one transmitter and two receiver antennas.

The SNR per bit is then given by

$$\gamma_b = \sum_{j=1}^D |\alpha_j|^2 \frac{E_b}{N_0} \quad (8.3.2)$$

which is a random variable following the distribution function of $\sum_{j=1}^D |\alpha_j|^2$ in which $|\alpha_j|^2$ is distributed exponentially. If we assume that the average of $\sum_{j=1}^D |\alpha_j|^2$ is normalized to 1, the average SNR per bit is then given by

$$\overline{\gamma_b} = \frac{E_b}{N_0}$$

It can be shown [Pah79, Pah80] that the most general form for the probability distribution function of γ_b is given by

$$f_{\Gamma}(\gamma_b) = \sum_{n=1}^D \frac{A_n}{2\lambda_n} e^{-\gamma_b/2\lambda_n} \quad (8.3.3)$$

where $\{\lambda_n\}$ are eigenvalues of the $D \times D$ branch amplitude covariance matrix whose elements are defined by $r_{ij} = E\{\alpha_j \alpha_i^*\}$ and the $\{A_n\}$ are defined by

$$A_n = \prod_{\substack{k=1 \\ k \neq n}}^D \frac{1}{1 - \lambda_k / \lambda_n}$$

Given the probability distribution function of the SNR per bit, as given in Eq. (8.3.3), the probability of outage with diversity reception of order D is given by

$$P_{\text{out}} = \int_0^{\gamma_{\text{out}}} f_{\Gamma}(\gamma_b) d\gamma_b = 1 - \sum_{n=1}^D A_n e^{-\gamma_{\text{out}}/2\lambda_n} \quad (8.3.4)$$

where γ_{out} , the threshold SNR for a given modulation technique, is determined by the same method as that used for flat fading and described in Section 8.2. The average probability of error for coherent PSK, expressed in terms of $\overline{\gamma}_b$ and determined from Eq. (8.2.8) using the density function in Eq. (8.3.3), is given by [Pah79, Pah90c]

$$\overline{P}_b = \sum_{n=1}^D \frac{A_n}{2} \left[1 - \left(\frac{2\overline{\gamma}_b\lambda_n}{1 + 2\overline{\gamma}_b\lambda_n} \right)^{1/2} \right] \quad (8.3.5)$$

This equation incorporates the correlation among all diversity branches and assumes that the values of all eigenvalues are different.

8.3.2 Special Cases

Let us now assume that the amplitudes of the signals received on different branches are all uncorrelated Rayleigh-distributed random variables. For our first case, we also assume that the same average signal power is received on each diversity branch and that the average SNR on each branch is denoted by $\overline{\gamma}_b$. The probability distribution function of the post-combining SNR is then given by

$$f_{\Gamma}(\gamma_b) = \frac{1}{(D-1)! \overline{\gamma}_b^D} \gamma_b^{D-1} e^{-\gamma_b/\overline{\gamma}_b} \quad (8.3.6)$$

The probability of outage at the post-combining SNR level γ_{out} is then given by

$$P_{\text{out}} = \int_0^{\gamma_{\text{out}}} f_{\Gamma}(\gamma_b) d\gamma_b = 1 - \frac{1}{e^{\gamma_{\text{out}}/\overline{\gamma}_b}} \sum_{j=1}^D \frac{1}{(j-1)!} \frac{\gamma_{\text{out}}}{\overline{\gamma}_b} \quad (8.3.7)$$

The average probability of error for coherent PSK demodulation at the maximal-ratio combiner output is given by

$$\overline{P}_b = [P(\overline{\gamma}_b)]^D \sum_{l=1}^D \binom{D-2+l}{l-1} [1 + P(\overline{\gamma}_b)]^{l-1} \approx \left(\frac{1}{4\overline{\gamma}_b} \right)^D \binom{2D-1}{D} \quad (8.3.8)$$

where $P(\overline{\gamma}_b)$ is found from Eq. (8.2.8) or (8.2.9) and is the equation giving the average error probability for a specific modulation technique used in the system. We use the standard notation for a binomial coefficient:

$$\binom{N}{k} = \frac{N!}{(N-k)! k!}$$

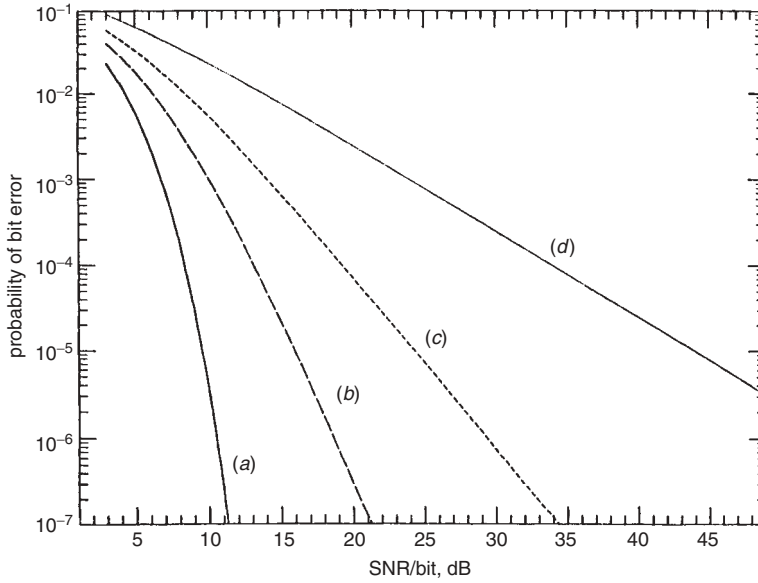


FIGURE 8.6 Probability of bit error for BPSK modulation for no fading and for Rayleigh fading with various orders of diversity: (a) no fading; (b) $D = 4$ fading; (c) $D = 2$ fading; (d) $D = 1$ fading.

The expression in Eq. (8.3.8) shows that average BER performance at the maximal-ratio combiner output improves exponentially with increasing D , the order of diversity. Figure 8.6 shows the average BER $\overline{P}_b$ versus average SNR per bit $\overline{\gamma}_b$, for different orders of diversity, with an assumption of independent equal-power Rayleigh fading on diversity branches. Included in the figure is the BER curve for steady-signal reception. As we saw earlier, with a single antenna, we lose 30 to 35 dB in performance relative to steady-signal reception at reasonable levels of BER. With two independent diversity branches, the performance loss is reduced to about 25 dB, and with four orders of diversity, the SNR penalty is reduced to around 10 dB. With additional orders of diversity the penalty relative to nonfading can be reduced further. There will, of course, be a practical limit to the order of diversity implemented because, for example, one cannot put an arbitrarily large number of antennas into a communications terminal.

If the average received power value from different branches is unequal but the amplitudes are uncorrelated Rayleigh-distributed random variables, the average error probability for a maximum ratio combining with D paths is given by [Pro01]

$$\overline{P}_b = \frac{1}{2} \sum_{i=1}^D \pi_i \left(1 - \sqrt{\frac{\overline{\gamma}_{bi}}{1 + \overline{\gamma}_{bi}}} \right) \quad (8.3.9)$$

where

$$\pi_i = \prod_{\substack{j=1 \\ j \neq i}}^D \frac{\overline{\gamma}_{bi}}{\overline{\gamma}_{bi} - \overline{\gamma}_{bj}}$$

and $\overline{\gamma}_{bi}$ is the average signal power received on path i .

For selection diversity, where the strongest path is selected for making the symbol decision, the average probability of PSK error for equal-power uncorrelated branches is given by the following closed-form equation [Sch66]:

$$\overline{P}_b = D \sum_{k=1}^D \binom{D-2}{k-1} \frac{(-1)^{k-1}}{k} P\left(\frac{\overline{\gamma}_b}{k}\right) \quad (8.3.10)$$

Calculation of the average probability of error for equal-gain combining is relatively difficult. However, the results given above for BER with maximal-ratio combining and selective combining can be used as lower and upper bounds, respectively, on the BER achieved with equal-gain combining. The actual BER performance for equal-gain combining in most cases of interest is closer to the performance of maximal-ratio combining.

The treatment of diversity combining techniques in this chapter is somewhat abstract. In subsequent chapters we discuss practical methods for providing diversity in real systems.

8.4 ERROR-CONTROL CODING FOR WIRELESS CHANNELS

A fundamental problem in digital communication over wireless networks is that the wireless environment typically produces levels of BER in delivered data which are several orders of magnitude poorer than that usually found on wireline channels. Added to this is the fact that in the wireless environment, channel errors often occur in bursts, essentially coinciding with the incidence of deep fades or even signal blockage on the link. The result of these effects is that average untreated error rates on wireless channels will often be at a relatively high level (perhaps around 10^{-2} on outdoor cellular and mobile radio channels) and will occur as mixtures of random-appearing and bursty patterns, where within the bursts the local BER will approach 50%. Such error characteristics may be marginally acceptable for some user applications, such as digital voice service, but may be completely unacceptable for other applications, such as message or data services or for critical system functions relying on over-the-air exchange of control signaling. Because of this, most digital wireless systems utilize one or more forms of *error-control coding* as one means of mitigating the effects of multipath fading on received bit quality.

From a broad perspective, error-control coding involves the addition of systematic redundancy to the transmitted data and the use of that redundancy at the receiver to improve the quality of the data delivered relative to uncoded transmission. As described in these terms, coding has some similarity to diversity transmission and reception. In fact, diversity transmission of information symbols, say on multiple frequencies or in multiple time slots, is essentially the same as repetition coding, a simple form of error-control coding. However, an extensive array of coding techniques are available which make much more efficient use of redundancy than can be obtained with simple repetition. In this subsection we describe briefly the principal coding techniques used in the principal wireless systems and discuss the issues arising in the application of coding to fading channels.

Error-control coding techniques used in wireless systems can be grouped broadly into three categories:

1. Error-detection coding using block codes

2. Forward-error correction (FEC) coding using both block and convolutional codes
3. Automatic repeat request (ARQ) schemes

8.4.1 Error-Detection and FEC Block Coding

In block coding for error detection or error correction, k information digits are used with a prescribed encoding rule to calculate a set of $n - k$ *parity-check digits*, which are transmitted along with the information digits on the channel as an n -digit *code block* or *codeword*. It is customary to refer to a parity-check block code as an (n, k) *code*, where n is the *block length* and k is the number of information digits encoded in each code block. The *code rate* is defined as the ratio $R = k/n$. The encoding rule, defined by a *parity-check matrix* or an equivalent *generator polynomial*, determines the mathematical structure of the code and hence its capabilities for detecting and correcting errors. Note that in this description we have used the term *digits* to denote the fact that block codes can be constructed on various nonbinary alphabets as well as on the binary alphabet (0,1). In all cases, encoding and decoding are performed with *finite-field arithmetic*, with *modulo-2 arithmetic* as the special case. The mathematical details of coding theory, along with discussions of widely used coding techniques, may be found in a number of texts, including [Pet72, Mac77, Mic85, Bla03, Lin04]. The application of coding techniques to fading channels is discussed in [Cla81, Big91, Pro01] and references cited therein.

The received version of the transmitted codeword is called simply the *received word*, to allow for the possibility that it may not be the intended word or even a valid codeword (as defined by the encoding rule), due to errors occurring during transmission. The *decoder* follows the demodulator and operates on the received word to determine whether the received information and check digits satisfy the encoding rule, and it uses any observed discrepancy to detect and possibly correct transmission errors.

A given block code may be used for error detection only or, given a sufficient amount of redundancy in the design, for error detection and correction, commonly referred to by the acronym *EDAC*. In error-detection decoding, the parity check bits are recomputed from the k information bits received and compared with the check bits received from the channel. If the check bits received are in agreement with those recomputed, the k information bits contained in that block are accepted as correct. Cyclic redundancy check (CRC) codes are widely used error-detection codes that can be applied to information-bit strings of arbitrary length [Cas93, L'Ga00]. Each CRC code is defined by its *generator polynomial*, which establishes its encoding rule. A number of CRC codes have been standardized, and microchips are available for implementing some of the more popular codes.

It is important to note that no error-detection code can be designed to detect all received error patterns successfully. In particular, an error pattern that changes the transmitted codeword into a different valid codeword is always undetectable. Thus, an apparently successful outcome of the error-detection decoding procedure does not guarantee that the code block was received correctly, because some error patterns will produce undetectable false decodings. For many error-detection block codes, the probability P_{FD} of false decoding is upper-bounded by a function of the number of parity bits $n - k$ [Wol82]:

$$P_{FD} \leq 2^{-(n-k)}$$

It should be noted here that the bound given above does not necessarily hold for an arbitrarily chosen error-detection code. For some codes, in certain ranges of channel BER, the probability of false decoding rises above the $2^{-(n-k)}$ “bound.” This issue is dealt with at some length in a number of papers, including [Wol82, Cas93].

Error-detection codes are incorporated into the transmitted signal structure for some of the wireless standards. For example, the USDC and GSM digital cellular standards include error-detection coding on selected bits in each frame of digitized speech. At the receiving end, where speech frames are being synthesized into analog speech, techniques such as frame interpolation can be employed to mask the effects of an occasional segment of detected error bits. Error-detection codes are also used in retransmission protocols, as we discuss in a later paragraph. In addition to the protection of compressed speech bits, CRC coding is used in various traffic and control channels in the major digital cellular standards and in the IEEE 802.11 standards for WLANs. The mathematical treatment of the structure and performance capabilities of block codes is based on the concepts of Hamming weight and Hamming distance. The *Hamming weight* (or simply, *weight*) of a codeword is the number of nonzero elements in the codeword. Similarly, the weight of an error pattern is the number of errors in the pattern. The error-correction capability of a block code is determined by its *minimum Hamming distance* (or simply, *minimum distance*), where the Hamming distance between two codewords is the number of symbol positions in which two codewords differ. A block code with minimum distance $d_{\min}$ can correct any error pattern having weight up to $t_{\max}$, where

$$t_{\max} = \left\lfloor \frac{d_{\min} - 1}{2} \right\rfloor$$

where $\lfloor x \rfloor$ denotes the largest integer not greater than x .

FEC coding techniques can be implemented to perform either *hard-* or *soft-decision decoding*, depending on the amount of information conveyed to the decoder with each demodulated symbol. In the simplest implementations, the demodulator makes a definite decision on each received symbol and passes the bits or symbols to a hard-decision decoder. Hard-decision decoding algorithms are essentially efficient algebraic equation-solving routines, although simple look-up tables are sometimes used for decoding very short block codes. Single-error-correcting codes are sometimes implemented with simple shift-register encoders and decoders. Here the codewords are represented as polynomials, and encoding and decoding are done using polynomial multiplication and division operations [Pet72, Lin04].

Soft-decision FEC decoding begins with soft-decision demodulation, in which the demodulator output is quantized to Q levels, where Q is greater than the size of the transmission alphabet. Quantization incurs loss of information, and thus soft-decision demodulation preserves information that can profitably be utilized with appropriate decoding algorithms. Soft-decision decoding algorithms more nearly resemble signal correlation or matched-filtering operations than equation-solving routines. A number of efficient soft-decision techniques have been devised for decoding block codes, and several are described in [Mic85, Skl01, Lin04]. It has been shown that the soft-decision Viterbi decoding algorithm, used widely for decoding convolutional codes, can also be used to perform optimum soft-decision decoding for some block codes [Wol78].

Soft-decision decoding provides better performance than hard-decision decoding at a cost of increased demodulator and decoder complexity. The range of performance improvement achievable will depend to a great extent on the characteristics of

the transmission channel. On steady-signal AWGN channels, the theoretical limit on SNR improvement achievable with soft-decision decoding is 3 dB [Cha72]. However, practical experience shows that actual improvements of 1 to 2 dB are feasible with algorithms of reasonable complexity [Mic85]. Much greater SNR improvements are achievable on fading channels, as we show later.

When a coding technique is specified in a wireless standard, no particular method of decoding is mandated, because manufacturers can choose to implement various receiver techniques without affecting interoperability with other manufacturers' products. Whether the use of soft-decision decoding is justified in any given application is a matter of trading off between improved performance, which may translate into widened signal coverage, and increased complexity, which may be reflected in size, weight, and cost of a communication terminal. A great many choices are available to a manufacturer. For example, a mobile terminal manufacturer may choose to implement hard-decision decoding in a mobile terminal, to minimize the complexity and cost of the subscriber equipment. At the same time, a base-station equipment manufacturer may choose to implement soft-decision decoding to maximize the geographic range over which mobile subscribers can operate.

8.4.2 Convolutional FEC Coding

In encoding a convolutional code, the encoder accepts information bits as a continuous stream and generates a continuous stream of encoded bits at a higher bit rate. The information stream is fed to the encoder b bits at a time, where b is usually in the range 1 to 6. The encoder operates on the current b -bit input and some number of immediately preceding b -bit inputs to produce V output bits with $V > b$. Thus, the code rate is $R = b/V$. The number of successive b -bit segments of information bits over which each encoding step operates is called the *length* of the code, which we also denote by k . The encoder for a convolutional code might be thought of as a form of digital filter with memory extending $k - 1$ symbols into the past. A typical binary convolutional code is one having $b = 1$, $V = 2$ or 3 , and k in the range 4 to 7.

As with block codes, convolutional codes may be decoded with either hard- or soft-decision decoding, and the performance advantage for soft-decision decoding will again vary with channel characteristics. Although a number of different algorithms are available for decoding convolutional codes, the most frequently used is the Viterbi algorithm [Vit67], which is in fact a *maximum likelihood decoding algorithm* for the steady-signal AWGN channel [For73]. Microchip Viterbi decoders are available from a number of manufacturers, and some of the chips provide both hard- and soft-decision decoding options.

We shall not delve any further into the details of code design here, because these details are amply treated in a number of texts, including those cited earlier in this section. Our brief discussion has served our primary purpose, which is to introduce the parameters: code rate, block length, and constraint length. These are key parameters in the selection of a coding technique, because the reciprocal of code rate provides a measure of required bandwidth expansion, while the block length or constraint length gives a measure of complexity of the required encoding and (more important) decoding operations.

TABLE 8.1 Examples of Coding Techniques Used in Wireless Systems

Type of Code	Comments
Hamming codes	Length = $2^m - 1$, $m = 2, 3, 4, \dots$ Minimum distance $d_{\min} = 3$
BCH codes	Length = $2^m - 1$, $m = 2, 3, 4, \dots$ $d_{\min} \geq 2t - 1$, t any integer Number of parity checks: $n - k \leq mt$
Golay (23, 12) code	$n = 23$, $k = 12$, $d_{\min} = 7$, $t = 3$
Reed–Solomon (RS) codes (q -ary)	$q = p^m$, p prime, m integer $N = q - 1$, $K = 1, 2, 3, \dots, N - 1$ $d_{\min} = N - K + 1$
Walsh–Hadamard codes	$n = 2^m$, $d_{\min} = n/2$
Binary convolutional codes	Typically used code rates: $\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{8}$ Typical constraint lengths: $k = 5, 6, 7$
Turbo codes	Two short encoders, with interleaving

Table 8.1 lists some of the coding schemes used in wireless systems. The first row in the table shows the Hamming codes, an infinite class of single-error-correcting ($d_{\min} = 3$) codes. The “natural” length of a Hamming code is $n = 2^m - 1$, where m can be any integer greater than 1, and we use the term *natural length* to mean the block length prescribed by the formal mathematical definition of the code. Any Hamming code can be shortened by replacing some of the information bits with assumed zeros. Hamming codes and shortened versions thereof are used in the link-layer coding specifications for a number of wireless standards, such as IS-136 and IS-95.

The second row in the table shows Bose–Chaudhuri–Hocquenghem (BCH) codes, an infinite class of binary multiple-error-correcting codes. For any positive integers m and $t < n/2$, there exists a binary BCH code with natural block length $n = 2^m - 1$ and minimum distance $d \geq 2t - 1$ having no more than mt parity-check bits. Each such code can correct up to t random errors per codeword and thus is a t -error-correcting code. BCH codes with $t = 1$ are identical to Hamming codes. BCH codes have been studied and utilized extensively, and much work has been done on developing efficient decoding algorithms for the codes. As with the Hamming codes, the BCH codes can be shortened from their natural lengths as needed, and thus the BCH class provides a rich assortment of codes with various block lengths and degrees of error-correction capability. Various BCH codes, including shortened versions, have been incorporated into wireless system specifications.

The next row in the table shows the three-error-correcting (23, 12) Golay code. This code has the special property that with hard-decision decoding, all received error patterns will be decoded; that is, no received word will be declared as having a “detectable but uncorrectable” error pattern. This property defines the Golay code as a *perfect code*. A frequently used variation of the Golay code is the (24, 12) distance-8 code, often called the *extended Golay code*, which is obtained by appending an overall parity check to the distance-7 (23, 12) code. The extended code is found to be attractive in some applications because its code rate k/n is exactly equal to 0.5. Both the basic and

extended Golay codes have been used widely for a number of years, and decoders are available as commercial chips. Both versions of the Golay code can be shortened as well, and several versions of both codes have been incorporated into wireless system specifications, such as the APCO/TIA standard for digital land-mobile radio [TIA93a].

The next row in the table shows Reed–Solomon (RS) codes, which are an important class of nonbinary block codes. The symbols in RS codewords are drawn from an alphabet of size $q = p^m$, where p can be any prime and m is an integer. In most applications $p = 2$, so that $q = 2^m$, and m bits are mapped into each q -ary symbol. An (N, K) RS code has block length $N = q - 1$, and the number of symbols in a codeword can have any value $K = 1, 2, 3, \dots, N - 1$. As with the Hamming codes and binary BCH codes, the RS codes can be shortened to lengths smaller than the natural length. Thus, a wide range of code designs is available within this class of codes. An important property of the RS codes is that any (N, K) RS code has the largest minimum distance achievable with the given values of N and K : specifically, $d_{\min} = N - K + 1$. For this reason, RS codes are described as *maximum-distance separable codes*. RS codes have proved to be very effective for use on channels where errors occur in bursts, or as mixtures of random errors and bursts. The RS codes can be defined as a special subclass of nonbinary BCH codes, and several of the efficient decoding algorithms developed for BCH codes can be applied to RS codes as well [Mic85, Wic94, Lin04]. RS codes are used in the link-layer coding structure specified for the CDPD cellular data standard [CDP93] and have also been specified for use in the APCO/TIA standard for digital land-mobile radio [TIA93a].

Walsh codes, also called *Hadamard codes*, are constructed by selecting as codewords rows of special square matrixes called *Hadamard matrixes*. A Hadamard matrix of order n is an $n \times n$ matrix of $+1$'s and -1 's, where all pairs of rows are orthogonal. Hadamard matrixes exist only for orders 1 and 2 and multiples of 4. (That they exist for all multiples of 4 has been conjectured but not yet proved.) If the matrix elements $\{\pm 1\}$ are replaced by 1's and 0's, respectively, a Hadamard matrix has one row of all 0's, and the remaining rows each have an equal number of 1's and 0's. Walsh or Hadamard codes can be constructed for block lengths $n = 2^m$, where m is an integer, all nonzero codewords having Hamming weight $n/2$ and all pairs of codewords being separated by Hamming distance $n/2$. The cdmaOne cellular system uses Walsh codes of order 64, where each Walsh codeword provides one of 64 channels on an RF carrier.

The next row in Table 8.1 shows binary convolutional codes, which are included in the major digital cellular standards. Although convolutional codes are designed to encode continuous streams of data, they are readily adapted to a block-structured transmission format simply by truncating the information bit stream and inserting known *tail bits* into the encoder, which facilitates decoding the information bits at the end of the block. The tail bits represent overhead in the transmission format and thus affect overall bandwidth efficiency. However, in most practical cases the impact is small.

The last row in Table 8.1 shows Turbo codes, a powerful convolutional coding technique introduced in 1993 [Ber93]. Turbo coding uses parallel concatenation of two short, recursive convolutional codes, together with interleaving and iterative *maximum a posteriori probability* (MAP) decoding, to provide low post-decoding bit error rates at low values of E_b/N_0 [Sk197, Hee99, Han02]. Turbo codes are utilized in the cdma2000 high-data-rate (HDR) standard [Ste01, Han02]. Also, a Turbo code has been

adopted by the Third-Generation Partnership Project (3GP) in the UMTS 3G cellular standard [ET500].

8.4.3 ARQ Schemes

A long-standing and widely used method of error control combines error-detection block coding with retransmission on request in a technique called *automatic repeat request* (ARQ). If forward-error correction coding is used in conjunction with an ARQ protocol, the technique is referred to as *hybrid ARQ*. Over the years, many variations on the ARQ technique have been studied, and detailed treatments can be found in a number of references, including [Lin04, Com84, Lin84, Tan88, Dos92]. ARQ techniques are particularly well suited to any channel where errors tend to occur in bursts, and to fading radio channels in particular. In essence, ARQ is a method of adapting the effective information transmission rate to the conditions of the channel. That is, when the channel transmission quality is high, most of the code blocks are received correctly on the first try, and information is carried over the channel at a rate at or near the maximum rate allowed by the transmission format. When channel quality degrades due to fading, signal blockage, or other temporary signal disruption, code blocks are received with detected errors, and transmission of new data is slowed down or even halted (“flow-controlled”), while erroneous blocks are retransmitted, perhaps multiple times, until the channel returns to a state of good transmission quality. Effective applications of ARQ are not limited to fading channels, of course. Forms of ARQ are incorporated into all of the common contention-based multiuser access protocols, such as ALOHA, CSMA and others, where the principal source of errors is collision between different users’ packets (see Chapter 11). ARQ protocols are also part of the design of standard data network protocols, such as BISYNC, X.25, and TCP/IP.

A key figure of merit for an ARQ system is its *throughput efficiency* or simply *throughput*, which is defined as the ratio of the average number of information bits accepted at the receiver to the maximum data transmission rate on the channel. The achievable throughput is determined to a large extent by the retransmission strategy chosen from the several strategies are available. The relative advantages and disadvantages of one retransmission strategy relative to another are influenced somewhat by the detailed error-clustering characteristics of the channel at hand. Here we briefly define the principal retransmission strategies and comment on their effectiveness on fading channels.

The three basic types of ARQ strategies are stop-and-wait, go-back- N , and selective repeat. The simplest strategy is *stop-and-wait*, in which the transmitter stops after transmitting each data block and waits until an acknowledgment (ACK) or retransmission request (NAK) is sent back from the receiver, or a timer expires. In full-duplex transmission, ACKs and NAKs are sent along with data blocks on the return channel. Typically, only error detection (rather than FEC) is implemented at the receiving end, although hybrid forms of stop-and-wait have been proposed for some applications [Com84, Lin04]. An obvious potential problem with stop-and-wait ARQ is that if the transmitter must be idle while waiting for acknowledgments, throughput will suffer, and if round-trip delays are long, throughput can suffer appreciably.

The problem of idling with stop-and-wait ARQ is alleviated with the use of a slightly more complex strategy called *continuous ARQ* or *go-back- N* . This is the retransmission protocol in predominant use in packet-switching networks. Here the transmitter does

not wait for ACKs or NAKs, but instead, transmits code blocks continuously until receipt of a NAK or expiration of a timer. Then the transmitter stops, backs up to the code block that was not received successfully, and restarts the transmission with that block. All N blocks that were transmitted in the time interval between the original transmission and the receipt of the NAK receipt or timer expiration are sent again in sequence. The throughput enhancement achieved with the pipelining nature of go-back- N can be pronounced. However, many of the blocks that are retransmitted may already have been received successfully, as many as all $N - 1$ following the one received with detected errors. Thus, additional throughput gains can be realized if only those blocks found to be in error are retransmitted. This is the essence of the strategy we describe next.

The throughput inefficiency caused by retransmission of error-free blocks can be overcome by *continuous ARQ with selective repeats*, commonly termed *selective repeat* (SR) or sometimes, *selective reject*. Here only a NAK'd frame, or a frame for which the timer has expired without receiving an ACK, need be retransmitted. In most applications the SR scheme provides the best throughput performance of the three basic ARQ strategies; however, its implementation is appreciably more complex than that of the other two. In particular, buffer management for the SR scheme is rather involved, because a reordering of blocks is required at the receiving end before releasing data to the user interface. Because of the relative complexity and cost of this protocol, it has not received wide commercial adoption. However, with modern developments in VLSI technology, the SR scheme is viewed increasingly as a cost-effective ARQ strategy. Recent studies of ARQ protocols for application to mobile and cellular radio networks have shown the SR strategy to be superior to go-back- N . The margin of performance is small for slow-fading conditions but becomes significantly greater in fast fading [Cha91, TIA93b]. The performance improvement achievable with one protocol over another can be very important in networks where maximum signal coverage and information throughput are needed. A version of SR (based on Tannenbaum's Protocol 6 [Tan88]) is specified in the asynchronous data service in the USDC digital cellular standard [TIA94].

8.4.4 Effects of Fading on Coding Performance

As we have noted in earlier discussions, the principal issue arising in the application of error-control coding techniques on wireless channels is that of error clustering, that is, the occurrence of errors in bursts or clusters of varying density as a direct result of the nonuniformity of signal propagation in the wireless environment. The key point here is that error-control codes generally perform better against statistically independent errors than against clustered errors. One consequence of this is that the performance of the code will vary with the temporal characteristics of the fading. Let us say, for example, that we transmit codewords with their symbols appearing in direct consecutive order in the channel. If the fading rate is very slow relative to the duration of a code block, so that a deep fade in effect "submerges" the entire block, clearly the code will be rendered ineffective, and only retransmission (if it is being used) can provide successful delivery of the contained information. However, if the fading rate in the channel is speeded up, with fade durations made comparable to the symbol interval (in this discussion we are ignoring associated phase distortion effects), errors in adjacent symbols may appear nearly independent, and the code may yield

satisfactory performance. This variation in performance with fading rate can be seen in the results of a number of investigations coding applied to mobile and cellular radio channels, where coding performance improves as vehicle velocity increases [Iye93]. Figures 8.7 and 8.8 show coding performance measured in computer simulations of a rate- $\frac{1}{2}$ convolutional code operating on a mobile radio channel with various mobile vehicle speeds. The code in each case is the rate- $\frac{1}{2}$ code specified in the USDC digital cellular standard. Figure 8.7 gives output BER versus channel SNR for hard-decision decoding, whereas Fig. 8.8 gives the corresponding results for soft-decision decoding of the convolutional code. It can be seen from the figures that BER performance improves as vehicle speed increases and channel errors become less correlated, and that greater improvements are achieved with soft-decision decoding.

A technique that can be used to reduce the statistical dependence of errors in a code block is *interleaving*. With interleaving, the symbols contained in one code block are not transmitted in consecutive order but instead are interspersed among other transmitted symbols so that a signal fade is less likely to impose a dense burst of errors on an individual code block. If the system design allows interleaving of code blocks over a sufficiently long time span, errors affecting individual blocks can be made to appear essentially independent, enhancing code performance. However, practical system considerations often rule out long interleaving spans.

For example, if the system is providing digital voice service, interleaving over multiple voice frames may impose an unacceptable time delay in a conversation. The USDC standard, for example, provides for interleaving over only two 20-ms voice frames, in the interest of minimizing time delay. Even in a digital data service, the use of large interleaving spans in turn requires large data buffers, which have a complexity and cost impact on the user terminal. Thus, the benefits of interleaving must be balanced against the need to satisfy user requirements properly and to achieve a cost-effective product design.

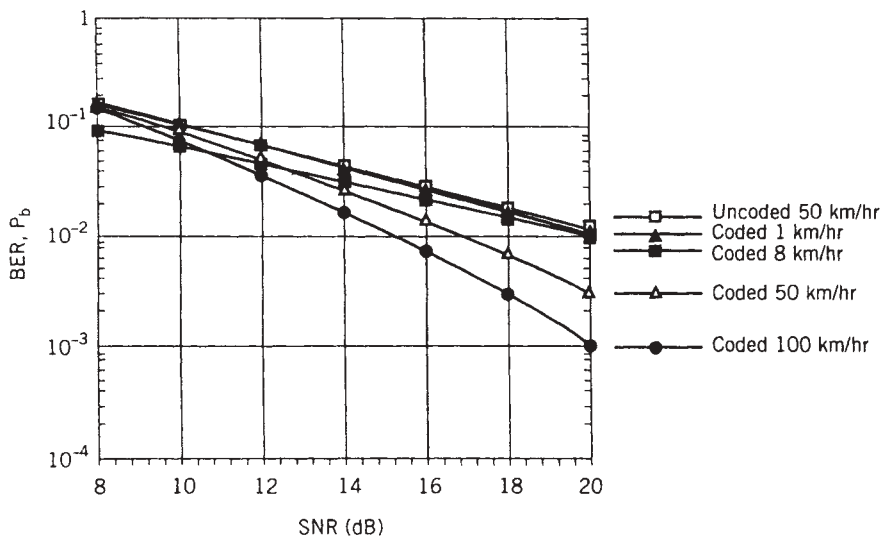


FIGURE 8.7 BER performance for the IS-54 rate- $\frac{1}{2}$ convolutional code on a simulated mobile radio channel (hard-decision decoding) [Iye93].

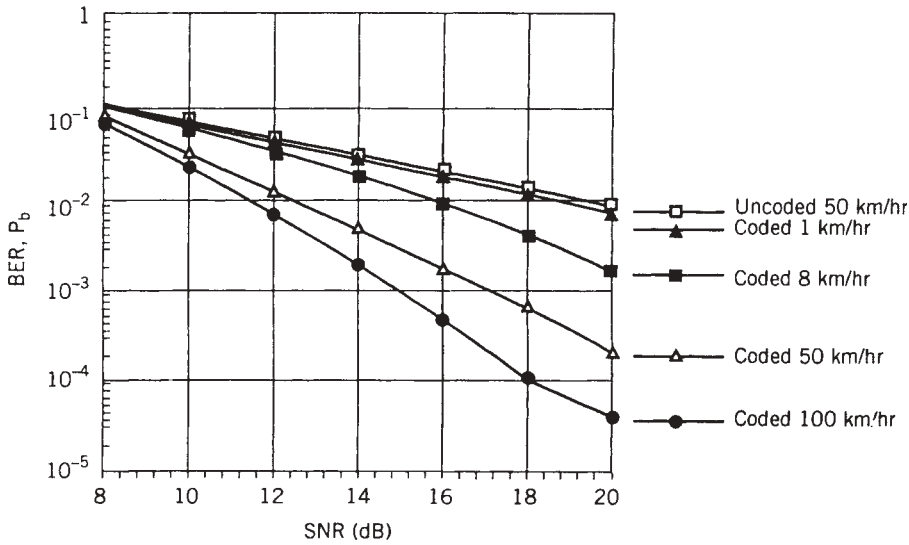


FIGURE 8.8 BER performance for the IS-54 rate- $\frac{1}{2}$ convolutional code on a simulated mobile radio channel (soft-decision decoding) [Iye93].

There is one type of system in which independence between consecutive errors can be achieved automatically; this is a system using frequency hopping. As we noted in an earlier chapter, fading can be made independent from one frequency to another, provided that the frequencies are sufficiently separated. In the ideal case, if the carrier is hopped to a new frequency in each consecutive symbol interval, the error-control code will be dealing with independent errors and will exhibit a corresponding improvement in performance. The GSM system is designed to operate with frequency hopping, and in service networks where it has been implemented, it has proved to enhance performance. The IEEE 802.11 standards also include frequency hopping.

The use of coding redundancy with multifrequency operation is very suggestive of frequency diversity operation, and on this point we return to the topic of the connections between diversity and coding as it is employed on fading channels. We do this in the context of examining soft-decision decoding, which we have already noted offers significant performance gains relative to hard-decision decoding on fading channels. The reason that soft-decision decoding is more beneficial on fading channels than on steady-signal channels is that in the fading case the soft-decision demodulator output conveys information about the instantaneous level of signal fading imposed on that particular symbol, and this in turn represents a “quality metric” for that symbol. By using the symbol quality metrics appropriately in decoding, one can achieve more reliable decoding than is achieved with use of only hard-decision demodulator outputs. Furthermore, one can show that given the use of coding on a Rayleigh fading channel, there is a direct connection between soft-decision decoding and diversity combining. Specifically, one can show the performance achieved with optimum soft-decision decoding of a block code having minimum distance $d_{\min}$ is equivalent to optimum diversity combining with order of diversity equal to $d_{\min}$ [Pro01]. The significance of this result can be appreciated by considering the example of using the (24, 12) extended Golay code on a Rayleigh fading channel. This is a rate- $\frac{1}{2}$ code with minimum distance 8. Therefore,

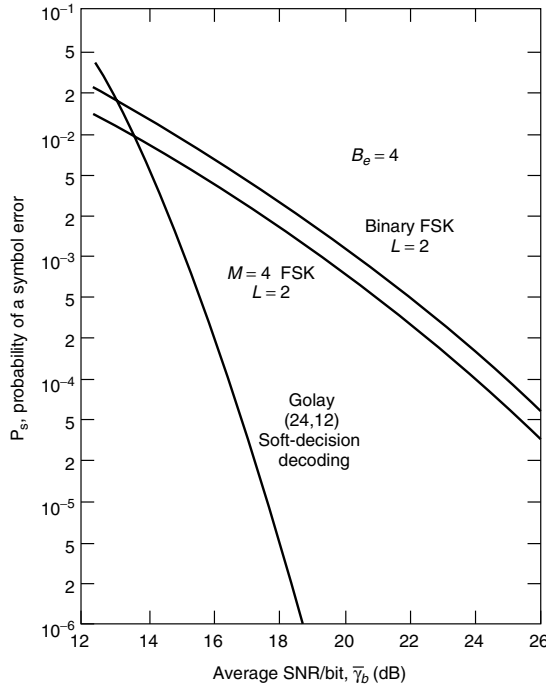


FIGURE 8.9 BER performance obtained with conventional dual-diversity as compared with rate- $\frac{1}{2}$ coding for bandwidth expansion factor $B_c = 4$. (From [Pro01] © McGraw-Hill with permission.)

the bandwidth expansion factor needed for use of this code is 2, as with dual diversity, but optimum soft-decision decoding, with independent bit-to-bit fading, yields performance equivalent to eighth-order diversity. The results obtained for this example are shown in Fig. 8.9. The figure shows average probability of bit error versus average SNR per bit $\bar{\gamma}_b$ for binary and 4-ary FSK, both with dual diversity, and Golay coding used with binary FSK and optimum soft-decision decoding. At $\bar{P}_b = 10^{-5}$ the Golay code outperforms dual-diversity 4-ary FSK by about 10 dB. Therefore, rate- $\frac{1}{2}$ coding can be used to make much more effective use of available bandwidth than does dual diversity, at the cost of the greater complexity required for a decoding implementation and a factor-of-2 reduction in bandwidth efficiency of the system.

8.5 SPACE-TIME CODING

Space-time coding (STC) techniques are used for wireless communication systems with multiple transmitting antennas and single or multiple receiving antennas. STC is realized by introducing temporal and spatial correlation into the signals transmitted from different antennas. Using STC does not require increasing the total transmitted power or transmission bandwidth. The overall diversity gain of the STC technique results from combining the time diversity obtained from coding with the space diversity obtained from using multiple antennas. In cellular networks a number of antennas can

be deployed in the base station while the mobile terminal receiver usually has one main antenna with some support from another antenna that may be implemented around the circuit board or the terminal cover cage. In traditional multiple-base-station antenna systems, all transmit antennas carry the same signal, and the signal received at each receiver antenna is the summation of all received signals from different transmitting antennas. As a result, the mobile station has no choice but to implement the optimum MRC for the received signal from different transmitter antennas. The basic principle of STC is to encode the transmitted symbols from different antennas at the base station and modify the receiver to take advantage of the space and time diversity of the arriving signal to implement an MRC of the multiple transmitter antennas. Using STC at the base station we can improve the performance of the downlink (base to mobile) channel significantly to support asymmetric applications such as Internet access, where the downlink data stream operates at a much higher rate than that of the uplink data stream.

Using STC, significant increases in throughput over a single-antenna system are possible with only two antennas at the base station and one or two antennas at the mobile terminal. It can be implemented for block [Ala98] or convolutional codes [Nag98, Tar98] with simple receiver structures. To show the basic concept of STC, we describe a simple two transmitting and one receiving antenna block coding system known as the *Alamouti coded STC system* [Ala98]. Figure 8.10 illustrates the operation of the two-transmitter, one-receiver antenna system and should be compared with the one-transmitter, two-receiver MRC antenna system shown in Fig. 8.5. There are two parts to be specified for the system: (1) how to encode the transmitted symbols for different transmitter antennas, and (2) how to combine the received signal form the decision variables. The transmitted block code operates on a sequence of two symbols $\{s_0, s_1\}$,

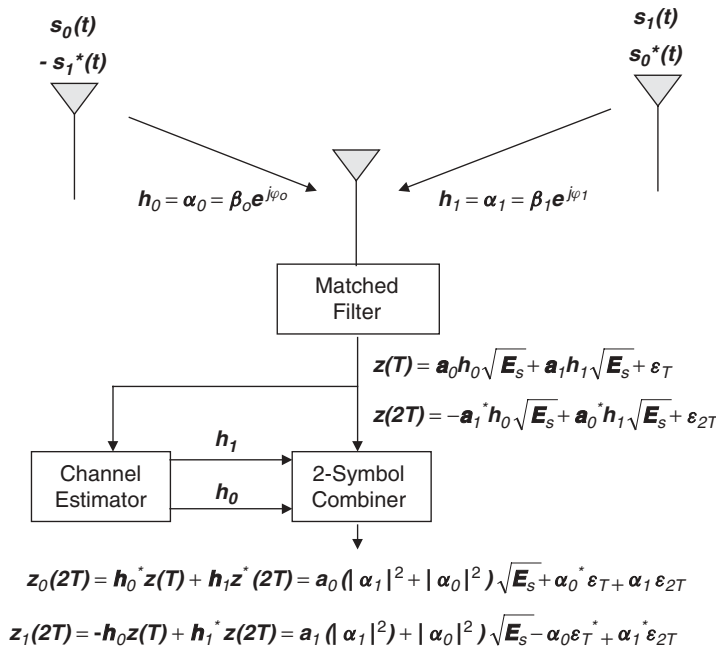


FIGURE 8.10 Simple Alamouti code for a two-transmitter, one-receiver antenna system.

this sequence being time-coded into two sequences $\{s_0, -s_1^*\}$ and $\{s_1, s_0^*\}$ to be then space-coded by the first and second antennas, respectively. Each of these two sequences has two symbols, which are transmitted in two consecutive time slots of one of the antennas in parallel with transmission of the other sequence in the other antenna. Assuming a flat-fading channel, we have two channel gain factors, h_0 and h_1 , which are samples of independent complex Gaussian processes each of whose magnitude obeys a Rayleigh distribution. Since all mobile channels are slow-fading channels, the value of the channel gains during transmission of two symbols remains the same. Therefore, the received signals after matched filtering for the first and second symbols are given by

$$\begin{aligned} z(T) &= a_0 h_0 \sqrt{E_s} + a_1 h_1 \sqrt{E_s} + \varepsilon_T \\ z(2T) &= -a_1^* h_0 \sqrt{E_s} + a_0^* h_1 \sqrt{E_s} + \varepsilon_{2T} \end{aligned}$$

To form the decision variables for the two transmitted symbols, we need to combine these two samples with an estimate of the channel impulse response. The decisions are made after two symbol transmissions from both antennas are completed, and they are given by

$$\begin{aligned} z_0(2T) &= h_0^* z(T) + h_1 z^*(2T) = a_0(|\alpha_1|^2 + |\alpha_0|^2) \sqrt{E_s} + \alpha_0^* \varepsilon_T + \alpha_1 \varepsilon_{2T} \\ z_1(2T) &= -h_0 z(T) + h_1^* z(2T) = a_1(|\alpha_1|^2 + |\alpha_0|^2) \sqrt{E_s} - \alpha_0 \varepsilon_T^* + \alpha_1^* \varepsilon_{2T} \end{aligned}$$

The statistical behavior of these decision variables is identical, and the same as the statistical behavior of the decision variable of the traditional MRC receiver derived in Section 8.3 and denoted by $z(T)$ in Eq. (8.3.1a). Consequently and not surprisingly, results of simulations for two transmitting and one receiving antennas provided in [Ala98] are identical to the results of the analysis of MRC in Section 8.3, shown in Fig. 8.6c, for one transmitting and two receiving antennas. A simple MIMO scheme for two transmitting and two receiving antennas was also introduced in [Ala98], and the simulation results show that the performance is identical to that shown in Fig. 8.6c. Therefore, Alamouti has shown that with his simple block-coded STC for two transmitting and one and two receiving antennas, one can obtain the same diversity performance as achieved with optimum MRC with two and four receiving antennas. More elegant approaches that combine transmitting diversity with channel coding, similar to trellis-coded modulation (TCM), are also available in [Nag98, Tar98]. A good overview of STC and its applications is provided in [Dha02].

Initial STC research efforts were focused on narrowband communication over flat-fading channels. The next step of research in this field is examination of STC techniques in multipath fading channels for broadband wireless communication. In a multiuser environment this involves complicated signal-processing algorithms for channel estimation, joint equalization and coding, and multiuser interference cancellation, and this has motivated a substantial amount of research work in recent years.

8.6 MIMO AND STC

Multiple-input multiple output (MIMO) antenna systems have recently emerged as one of the most promising technologies for the next generation of wireless networks.

As we described in Section 6.2.4, the IEEE 802.11 community, anticipating future developments of MIMO systems, is already working to develop channel models for performance evaluation of such systems. In general, MIMO systems combine the transmitting scheme and the detection process so that the overall performance of the system is improved. In Section 8.3 we demonstrated the large reduction in the SNR requirement resulting from the diversity provided by a single-input multiple-output (SIMO) antenna or any other diversity system. In Section 8.5 we introduced STC for implementation of MIMO systems, and we gave an example of a multiple-input single-output (MISO) STC block code. In this section we resume the discussion begun in Section 8.5, but with greater emphasis on the general understanding and limits of the MIMO systems when they are applied to narrowband modems. We leave the description of MIMO systems for broadband modems to Chapter 9.

In general, a MIMO system uses multiple transmitting and receiving antennas. If we denote the number of transmitting antennas by M and the number of receiving antennas by N , the incoming blocks of information bits are divided among M branches of the transmitting antenna system. The vector of M blocks of data is then coded and modulated to form a vector of M modulated symbols which are then transmitted from different antennas. The signals received from M transmitters at each of the N branches of the receiver antennas are fed to a signal processing unit that ultimately recovers the transmitted information bits from the signals received from all the antennas. This operation is well described with an example for $M = N = 3$, shown in Fig. 8.11 [Ges03]. In this figure the modulation technique is QPSK, which uses blocks of data carrying 2 bits each. The modulation and mapping unit at the transmitter generates the M-STC-coded signals for transmission. The received N-signal from the N receiving antennas is processed by the signal processing unit at the receiver to form estimates of the M transmitted symbols. Decisions made on the estimated transmitted symbols are then combined into a serial format to reproduce the transmitted bit stream recovered. Using this system, even though the signal constellation received at the antenna is very noisy, receiver signal processing produces a clear signal constellation.

8.6.1 Design of Codes for MIMO Systems

The main challenge in the design of MIMO systems is to find a modulation and mapping technique that allows a reasonable signal processing unit at the receiver to detect the transmitted symbols distributed among the transmitter antennas. In Section 8.5 we introduced the simple and practical $M = 2$, $N = 1$ Alamouti MISO-STC with orthogonal symbols and we noted that another simple 2×2 orthogonal MIMO-STC is introduced in [Ala98]. Design of orthogonal block codes for a larger number of antennas is rather challenging; however, other researchers have introduced more practical nonorthogonal solutions. The first coding technique of this type was introduced by Foschini in [Fos99] and today is known as the *diagonal Bell Labs layered space-time* (D-BLAST) *algorithm*. In this system the incoming data stream is demultiplexed into M data streams. Each stream is encoded independently from other streams. Rather than sending each stream through a separate antenna, the transmitted bit streams are periodically cycled across all transmitter antennas. This cyclic operation proceeds for N symbols. To further clarify this coding technique, consider the 3×3 MIMO system of Fig. 8.11. In its simplest form, assuming no coding on each specific branch stream, the encoder at the transmitter forms a sequence of blocks of length

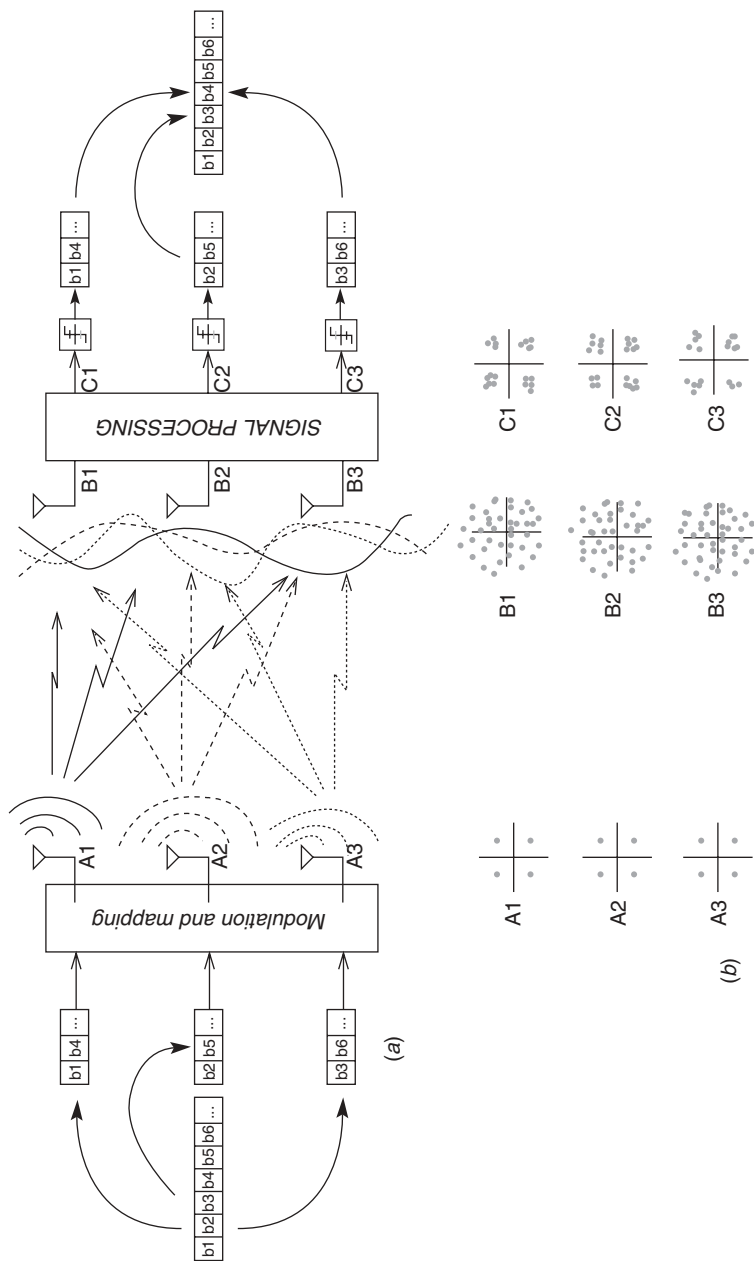


FIGURE 8.11 MIMO example: (a) modulation and demodulation; (b) signal constellations [Ges03] (© IEEE).

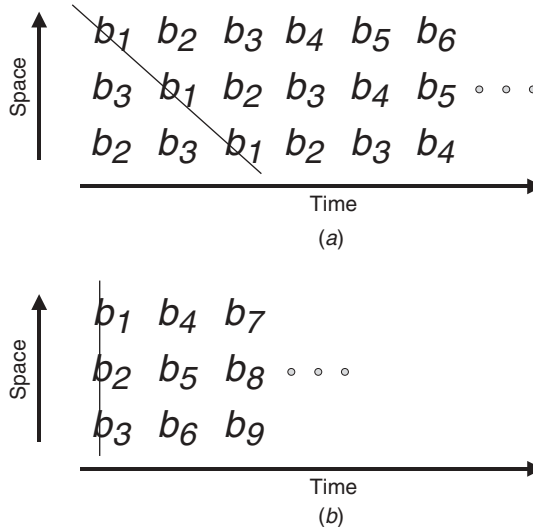


FIGURE 8.12 Encoding scheme in (a) D-BLAST and (b) V-BLAST.

three, $\{b_1, b_2, b_3\}$, and transmits that with antenna A_1 . At the same time, it transmits $\{b_2, b_3, b_1\}$ with antenna A_2 and $\{b_3, b_1, b_2\}$ with antenna A_3 . Figure 8.12.a illustrates this operation and why it is referred to as D-BLAST. In this transmission system all the transmitted symbols are exposed to all $M \times N$ fading channels between the M transmitting and N receiving antennas. The challenge here is the design of a receiver algorithm that can best take advantage of this diversity and extract all the transmitted symbols with the lowest possible error rate. The decoding process for this system is complex and is explained in [Fos99].

A modified version of D-BLAST, known as *vertical* or *V-BLAST*, was introduced by Wolnianski et al. [Wol98]. The main difference between the D- and V-BLAST is in the encoding process. In V-BLAST, the cyclic format of the transmitted streams in different antennas provides an additional redundancy at the expense of a more complex architecture in D-BLAST. This additional coding, however, leads to higher bandwidth efficiency for D-BLAST. As shown in Fig. 8.12b, in V-BLAST the incoming data blocks are simply transmitted without any interstream coding. The V-BLAST decoding techniques, described in [Wol98], take advantage of conventional adaptive antenna array techniques. These techniques operate very similarly to adaptive equalization and have been used for the past several decades in military applications, where they are referred to as *interference-nulling techniques*. Nulling techniques treat the signal received from one of the transmitted streams as the desired signal and the remainder of the signals as interference. Weights are applied to the signals received from different antennas, similar to the tap weights used as an equalizer, so that the performance is optimized. The algorithm used to update the antenna weights, again similar to adaptive equalization, can be either mean-squared error (MSE) or zero forcing (ZF). The architectures used for implementation of nulling techniques are—also similar to equalization—either linear, or nonlinear decision feedback. Equalization techniques are discussed in more detail in Chapter 9, and that analysis is directly applicable to the V-BLAST decoding techniques described in [Wol98]. In the V-BLAST system, since

the interference is caused by other information bits carried on known channels, they can be estimated and subtracted from the incoming signal. This interference cancellation process further improves the performance of a V-BLAST system. The laboratory prototype of V-BLAST, reported in [Wol98], operating at a 30-kHz channel spacing used in AMPS and USDC systems, has demonstrated the capability of operating at data rates up to 621 kb/s at 1.9 GHz.

8.6.2 Capacity Limits for MIMO Systems

In Section 8.2 we used the probability of error versus SNR performance curves for traditional modulation techniques to show how Rayleigh fading degrades the performance of a modem and how a diversity system recovers from that degradation. A more abstract approach to determining performance bounds, commonly used by information theorists, is to find the channel capacity limits using Shannon's bound, which was introduced in Section 7.3.1. As we showed in that section, the normalized channel capacity in bits per second per hertz is given by

$$\frac{C}{W} = \log_2 \left(1 + \frac{S}{N_0 W} \right)$$

This equation is for the single-input single-output (SISO) additive white Gaussian noise channel. In a fading channel this equation becomes

$$\frac{C}{W} = \log_2 \left(1 + \frac{S}{N_0 W} |\alpha|^2 \right)$$

where $|\alpha|^2$ is a random variable representing the amplitude fluctuations due to fading. In a SIMO channel, we have

$$\frac{C}{W} = \log_2 \left(1 + \frac{S}{N_0 W} \sum_{i=1}^N |\alpha_i|^2 \right)$$

where α_i is the complex gain of the i th received diversity branch. Continuing in this manner, the capacity of a MISO channel is given by

$$\frac{C}{W} = \log_2 \left(1 + \frac{S}{N_0 W} \frac{1}{M} \sum_{i=1}^N |\alpha_i|^2 \right)$$

where the factor M comes into the picture because transmitted power has to be divided among all antennas to keep the overall power constant. Following the original derivations for the capacity of MIMO provided in [Tel95, Fos99], the capacity of the MIMO channel is given by

$$\frac{C}{W} = \sum_{i=1}^N \log_2 \left(1 + \frac{S}{N_0 W} \frac{1}{M} \lambda_i \right)$$

where λ_i are the eigenvalues of the $M \times N$ cross-correlation matrix of the channel gains between the elements of the transmitter and receiver antennas. Several bounds

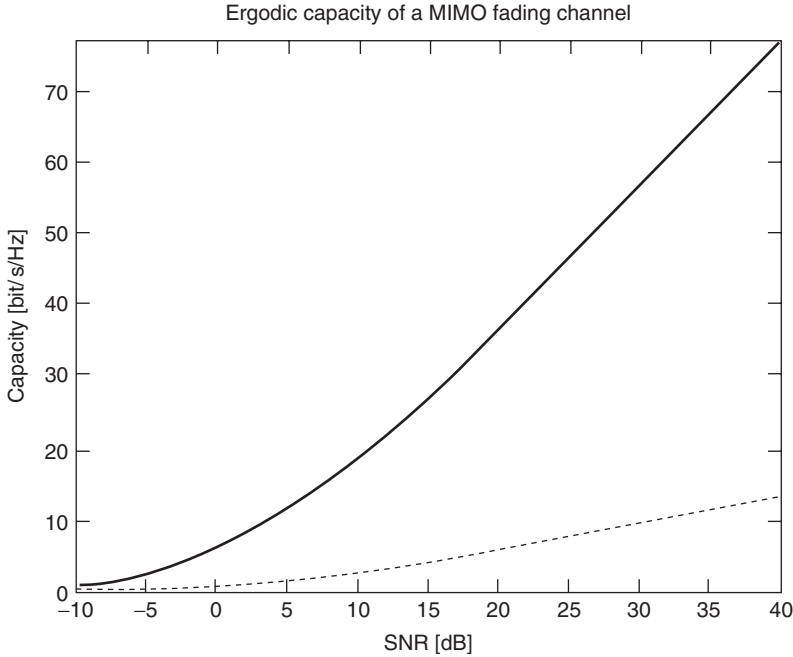


FIGURE 8.13 Comparison of the capacity of SISO and 6×6 MIMO under ideal conditions [Hol01].

on the capacity of MIMO have been developed. Maximum capacity is achieved when each of the N transmitted signals is received with each of the N receiving antennas without interference. This implies ideal nulling of the received signal from all transmitters and ideal interference cancellation at the receiver. A bound calculated this way is derived in [Hol01], and the result of that calculation is shown in Fig. 8.13. Another bound on the performance of practical implementations of MIMO using D-BLAST and V-BLAST is provided in [Fos99]. Figure 8.14 shows a comparison of the two systems for the limiting case of an infinite number of receiving antennas and a capacity normalized to the number of transmitter antennas. Variations of these bounds are commonly used in the literature to compare the fundamental performance bounds of different MIMO systems.

8.6.3 Practical Considerations for MIMO Systems

The receiver structures and performance bounds for MIMO systems provided earlier in this section are obtained assuming a flat-fading channel and ideal practical conditions. The performance of the STC and MIMO techniques on frequency-selective fading channels degrades significantly and we need to modify the receiver structures to exploit fully the space-time diversity provided by these techniques. In Chapters 9 and 10 we give some examples for performance evaluation and modified receiver structures for STC and MIMO systems. In this section we provide a brief description of practical considerations for the implementation of STC and MIMO receivers. In practical situations, performance enhancements of MIMO come at the expense of increased receiver

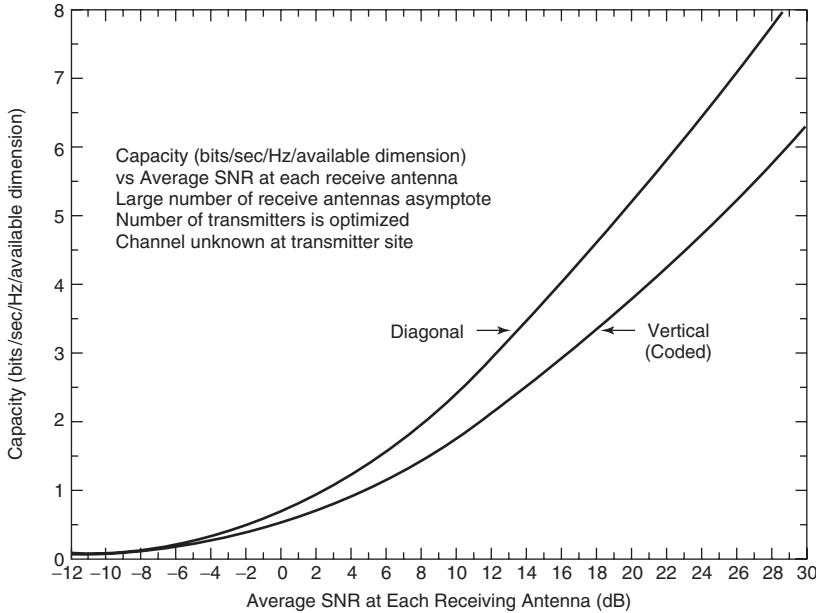


FIGURE 8.14 Capacity bounds on the performance of D-BLAST and V-BLAST systems [Fos99] (© IEEE).

complexity both in the base station and the mobile terminals. In addition, various factors, such as incorrect channel estimation, the presence of correlation among antenna elements, and higher Doppler frequencies tend to degrade the ideal system performance. In the remainder of this section, following discussions in [Ges03], we address additional complexities for the design of the antenna and receiver for MIMO systems.

The number of antenna elements and interelement spacing are two practical issues related to implementation of MIMO systems. Base stations with large numbers of antennas and wide spacing between the elements are difficult to install and maintain. In practice, MIMO antennas in the base stations use about four elements installed within a few meters of one another. In theory, as was shown in Section 3.2.2, if we change the spacing by one wavelength, λ , the phase changes by 2π radians. Since for 2 GHz the wavelength is around 15 cm, one can assume that antenna separations of that order provides reasonable independence among the received signals in the two elements. In practice, MIMO systems use an overall size of 10λ (1.5 m) for a four-element antenna. The wide spacing is preferred because base stations are usually mounted at elevated positions where the presence of local scatterers needed to decorrelate the fading cannot always be guaranteed. For the mobile terminal, the achievable spacing is sufficient to ensure a fair amount of uncorrelated fading because the terminal is typically situated among local scatterers, and quite often there is no direct propagation path [Ges03]. For the mobile terminal, the number of installed antennas depends on the size of the terminal. For a laptop computer, one can easily use four patch antennas in a square with $\lambda/2$ (7.5 cm) sides embedded in the casing. For a small mobile telephone, however, one would have difficulty installing more than one stick-out antenna. Some manufacturers add a circular antenna around the circuit board to increase the diversified spatial reception. Some manufacturers embed the antenna inside the casing to improve

product appearance and customer appeal, and this makes spacing requirements even more challenging.

Another important practical issue for implementation of the MIMO systems is that a MIMO receiver requires a number of accurate channel estimators. The channel estimation function adds to the complexity of the receiver because it must track a full matrix of channel conditions. In frequency-selective fading each element of the channel estimator matrix needs to track several paths or several tones across the frequency band of the system, and that further increases the computational load. Another important concern comes from the need for extra RF hardware and sophisticated receiver separation algorithms. These complexities not only create challenges in achieving low cost and compact design of mobile terminals but also affect battery life, a very important issue for the mobile user.

QUESTIONS

- (a) What are the physical meanings of the average probability of error and the probability of outage in a fading channel?
- (b) Using Fig. 8.1, explain why the average error rate in fading channels is much higher than the error rate in non-fading channels. Then, use Fig. 8.2 to explain why diversity is so effective in mitigating the effects of fading.
- (c) Name the three major techniques for diversity combining and explain the advantages and disadvantages of these techniques.
- (d) Explain why coding is so effective in improving the performance in fading channels and discuss why scramblers are used with coding to improve protection in fading channels.
- (e) Why do we usually use error-correcting codes for packet data communications and error-detecting codes for digital voice telephony applications?
- (f) Why are ARQ codes popular in data applications but not in voice applications?
- (g) Explain the principal differences between turbo coding and traditional convolutional coding.
- (h) If a data block is protected using the CRC-16 code, compute an approximate upper bound on the probability of accepting an erroneous block of data.
- (i) What is new in STC, and how does it relate to MIMO?
- (j) What is the difference between D-BLAST and V-BLAST systems, and why are they popular in wireless communications?

PROBLEMS

1. Sketch the probability of outage versus the error rate threshold for BPSK modulation with one and two orders of diversity and average SNRs of 10, 20, 30, and 40 dB. Assume that the channel exhibits flat Rayleigh fading and that the two

- diversity branches are independent, with the same average power. Use logarithmic scales for both axes.
2. (a) Develop an equation for approximate calculation of the probability of outage of M -ary coherent PSK modulation over a flat Rayleigh fading channel. The equation should be written in terms of the average received SNR per bit γ_b and threshold SNRs γ_{th} .
 - (b) For $M = 16$, plot the probability of outage versus threshold SNR γ_{th} for average SNRs $\gamma_b = 10, 20$, and 30 dB. For an acceptable error rate threshold of 10^{-5} , compare the outage rates for the three plots.
 - (c) For $M = 16$, plot the probability of outage versus average SNR γ_b for threshold SNRs $\gamma_{th} = 10, 20$, and 30 dB. For an average error rate of 10^{-5} , compare the outage rates for the three plots.
 3. (a) Provide an equation for the approximate calculation of the probability of outage of M -ary coherent QAM modulation over a flat Rayleigh fading channel. The equation should be written in terms of the average received SNR per bit $\overline{\gamma}_b$ and the threshold SNR γ_{th} .
 - (b) For $M = 16$, plot the probability of outage versus threshold SNR γ_{th} for average SNRs $\overline{\gamma}_b = 10, 20$, and 30 dB. For an acceptable error rate threshold of 10^{-5} , compare the outage rates for the three plots.
 - (c) For $M = 16$, plot the probability of outage versus average SNR γ_b for threshold SNRs $\gamma_{th} = 10, 20$, and 30 dB. For an average error rate of 10^{-5} , compare the outage rates for the three plots.
 4. (a) What average received SNR is required for a QPSK modem operating over a flat-fading radio channel to have an outage rate of 10^{-2} relative to the threshold error rate of 10^{-4} ?
 - (b) If we change the acceptable threshold level in part (a) to 10^{-2} , what improvement in the outage rate would we see?
 - (c) If we increase the transmitted power in part (a) fourfold (6 dB), how much improvement would be seen in the outage rate?
 - (d) If we operate the modem over a fixed wireline channel, what SNR would be required to maintain an error rate of 10^{-2} ? How much improvement in the error rate would occur if we increase the power by 6 dB?
 5. A coherent BPSK receiver operates in a Rayleigh fading channel with two antennas employing maximal ratio combining. The signals received from the two antennas are correlated, and the eigenvalues of the covariance matrix are 0.3 and 0.7.
 - (a) Plot the average probability of error on a logarithmic scale versus the average SNR in decibels.
 - (b) Plot the probability of outage on a logarithmic scale versus the average SNR in decibels.
 - (c) Repeat parts (a) and (b) when the two diversity branches are uncorrelated and the power received in both branches is the same.
 - (d) Assuming that the ratios of two eigenvalues are given by $r = \lambda_1/\lambda_2$, where $\lambda_2 = 1 - \lambda_1$, plot the average probability of error versus the ratio of the eigenvalues for $0 < \lambda_1 < 1$ and an average SNR per bit $\overline{\gamma}_b = 30$ dB. Use

the plots to explain the impact of correlation between the diversity branches on the performance of a receiver with two-branch diversity.

(*Hint:* Note that the calculations involve subtraction of two exponentials with values very close to each other. For an accurate computation, either very high precision computation or series expansions may be needed.)

6. Use MatLAB or an alternative computational tool to plot the probability of error versus the average SNR of a binary DPSK-NCD modem over a Rician fading channel with $k = 6$ dB.
7. Use MathCAD or MatLAB to plot the probability of error versus the average signal-to-noise ratio of a binary DPSK-NCD modem over a zero-mean lognormal fading channel.

PROJECTS

Project 1: Error Rate on a Fading Channel for QPSK Modulation

This project combines the results of Project 2 in Chapter 4 and Project 1 in Chapter 7 to analyze the performance of a QPSK modem over a fading channel.

- (a) Give the value of the parameters A and B in Eq. (8.2.8) for calculation of the probability of error for the QPSK modulation. Plot the average bit error rate versus average SNR for this modulation technique.
- (b) Use MatLAB or an alternative computation tool to plot the average probability of symbol error versus SNR (in dB). What are the SNRs (in dB) for the probability of symbol error of 10^{-2} and 10^{-3} ? Let us refer to these two SNRs as SNR-2 and SNR-3.
- (c) Simulate transmission of the QPSK signal over a Rayleigh fading channel corrupted by additive Gaussian noise for 10,000 transmitted bits. For simulation of the channel use the results of Project 2 in Chapter 4 and for the implementation of the transmitted symbols and the detection process use the results of Project 1 in Chapter 7. Run the simulation for the average SNRs of SNR-2 and SNR-3 that were found in part (b). Compare the error rates observed in the simulation with the expected error rates of 10^{-2} and 10^{-3} .

Project 2: MRC for BPSK Modulation

This project expands Project 1 for the implementation of an MRC receiver described in Fig. 8.5.

- (a) Use Eq. (8.2.17) to give the probability of error of an MRC receiver for the BPSK modulation assuming dual diversity and independent fading channels. Plot the average bit error rate versus average SNR for this MRC receiver.
- (b) Use MatLAB or an alternative computation tool to plot the average probability of symbol error versus SNR (in dB). What are the SNRs (in dB) for the probability of symbol error of 10^{-2} and 10^{-3} ? Let us refer to these two SNRs as SNR-2 and SNR-3.

- (c) Simulate transmission of the BPSK signal over two independent branches of Rayleigh fading channel corrupted by additive Gaussian noise for 10,000 transmitted bits. For the design of the MRC receiver use Fig. 8.5. Run the simulation for the average SNRs of SNR-2 and SNR-3 that were found in part (c). Compare the error rates observed in the simulation with the expected error rates of 10^{-2} and 10^{-3} .
- (d) Repeat part (c) for STC using Alamouti's codes described in Fig. 8.10. Discuss the effectiveness of the codes in improving the average error rate of the modem.

9

BROADBAND MODEM TECHNOLOGIES

- 9.1 Introduction
- 9.2 Effects of Frequency-Selective Multipath Fading
 - 9.2.1 Effects of Frequency Selectivity
 - 9.2.2 Multipath Effects and Data-Rate Limitations
- 9.3 Discrete Multipath Fading Channel Model
 - 9.3.1 Adaptive Channel Measurement
- 9.4 Adaptive Discrete Matched Filter
 - 9.4.1 Performance Prediction for the DMF
 - 9.4.2 Adaptive MLSE
- 9.5 Adaptive Equalization
 - 9.5.1 Equalizer Architectures
 - 9.5.2 Adaptive Algorithms for Equalizers
 - 9.5.3 Performance of a DFE in Fading Multipath Channels
- 9.6 Sectorized Antennas
 - 9.6.1 Performance Prediction in an LOS Environment
 - 9.6.2 Performance Prediction in an OLOS Environment
- 9.7 Multicarrier, OFDM, and Frequency Diversity
 - 9.7.1 Multirate Transmission
 - 9.7.2 Multiampitude and Multiphase Modulation and Coding
- 9.8 Comparison of Traditional Broadband Modems
- 9.9 MIMO in Frequency-Selective Fading
 - 9.9.1 STC-MIMO and Equalization
 - 9.9.2 MIMO OFDM
- Appendix 9A: Analysis of the Equalizers
 - Questions
 - Problems
 - Projects
 - Project 1. Performance Analysis of an Equalizer
 - Project 2. Simulation of a Simplified IEEE 802.11a/g OFDM

9.1 INTRODUCTION

In Chapter 7 we presented the basic narrowband modulation techniques and requirements for radio modem design. In Chapter 8 we discussed the effects of fading on the performance of narrowband modems and showed how diversity techniques counteract the harmful effects of random amplitude fluctuation caused by multipath fading. In using narrowband modems the frequency response of the channel over the bandwidth of the transmitted symbol is flat. Since the frequency response of the channel is flat, the transmitted symbols preserve their basic waveform at the receiver and get disturbed only by the additive noise and envelope fading of the channel. This condition holds when the bandwidth of the transmitted signal is much less than the coherent bandwidth of the channel. In broadband communications the bandwidth of the transmitted signal grows beyond a small fraction of the coherent bandwidth of the channel, and the frequency response of the channel is no longer flat over the transmission bandwidth of the communication symbols. This situation allows frequency-selective fading to affect only certain portions of the transmission bandwidth, causing significant changes in the shape of the transmitted symbol, stretching it into previous and following transmitted symbols and causing intersymbol interference. Therefore, frequency-selective fading channels deliver a distorted form of the transmitted symbol stream to the receiver. As the data rate is increased, the channel appears more frequency selective and modem performance degrades further. To recover the transmitted symbols and improve the quality of the information detected at higher data rates in frequency-selective fading channels, we need adaptive signal-processing algorithms that can counteract the harmful effects of channel distortions and allow broadband communications at required levels of quality.

Since the mid-1950s, many different types of adaptive receiving techniques have been studied for a variety of frequency-selective fading multipath channels. These techniques increase achievable modem data rates by overcoming to varying degrees the frequency-selective multipath fading characteristics of the channel. These techniques are in effect intended to take advantage of the in-band diversity provided by the frequency-selective multipath nature of the received signal. Historically, the most important adaptive receiver for fading multipath channels is the RAKE system invented by Robert Price and Paul Green of the MIT Lincoln Laboratory [Pri58]. The RAKE system was originally designed for teletype communication with a baud rate of 90 chips/s operating over an ionospheric channel. The envelope of the transmitted binary FSK signal was a 10-kHz direct-sequence spread-spectrum signal. The received signal was passed through two tapped delay lines—the matched filters—for mark and space frequencies, and the outputs were compared for decision making. The tap gains of the delay lines were adaptively adjusted by cross-correlating the received signal with both mark and space reference signals at the receiver. Later, other versions of RAKE receivers were examined for urban radio [Kam81], HF radio [Bel88], and indoor radio [Cha93] channels. The most popular implementation of the RAKE receiver was in the Qualcomm CDMA system that was adopted as the IS-95 standard. Further details on design and performance evaluation of the modern RAKE receivers used in second- and third-generation systems are provided in Chapter 10.

During the 1970s and 1980s, other techniques emerged in the development of a new generation of radios for frequency-selective fading multipath channels. Time gating of the transmitted pulse to avoid ISI, with adaptive discrete matched filtering (DMF) of each pulse received, was the approach taken for a family of military troposcatter

radios [Con78, Pah80]. The adaptive decision feedback equalizer (DFE) was another approach investigated for application to troposcatter [Grz75, Mon84], microwave line of sight [Bel84, Pah85a], HF [Fal85], and indoor radio [Sex89a] channels. Finally, an adaptive version of the MLSE technique [For72] was investigated for troposcatter in [Bel69] and for HF in [Cha75]. Similar to a RAKE receiver, DMF, DFE, and MLSE modems intend to take advantage of the time dispersion to provide for time diversity in the received signal from different paths.

Another approach to dealing with the problem of frequency-selective multipath distortion is one that has been used in HF radio modems for more than 40 years, known by several names, including *multicarrier modulation* (MCM) or *multitone modulation* [Bin90]. Computationally, efficient implementation of MCM using fast Fourier transform (FFT) algorithms, referred to as *orthogonal frequency-division multiplexing* (OFDM) *modulation*, was introduced unsuccessfully for 19.2-kb/s voiceband modem standards in the early 1980s [Pah88c]. In the early 1990s, OFDM was examined for radio communications [Har93], and in the late 1990s it was adopted for the IEEE 802.11a and subsequently, IEEE 802.11g standards for WLANs. The concept here is very simple: Instead of modulating a single carrier at rate R symbols/s, we use N carriers spaced by about R/N hertz and modulate each of the carriers at the rate R/N symbols/s. We can select N so that the bandwidth of individual carriers is less than fractions of the coherent bandwidth, whereas the bandwidth of the entire system is comparable with the coherent bandwidth. In this way, individual carriers are supporting narrowband modems operating over flat-fading channels, whereas the complete MCM system is a broadband modem. In MCM the subchannels provide a form of frequency diversity, which can be exploited by applying error-control coding across symbols in different subchannels or implementing an adaptive system adjusting the power of individual channels to the channel conditions for that particular subcarrier.

In addition to time and frequency diversity, broadband wireless modems also use spatial antenna diversity and time-duration coding diversity by employing sectored antennas and STC and MIMO techniques. In the early 1990s, Motorola's Altair project used sectored antenna systems for implementation of a 10-Mb/s broadband modem for a WLAN system operating at 18 GHz. Three-sectored antennas are very popular in second- and third-generation system installations. In the early 2000s, using STC and MIMO for the high-data-rate forward channel of third-generation systems attracted attention, and the WLAN industry started considering combinations of MIMO and OFDM to increase the data rate of these systems beyond 100 Mb/s.

The emphasis in this chapter is on applications of signal-processing algorithms in receivers to improve modem performance in fading multipath wireless media and to achieve reliable broadband communications. We start our discussions in Section 9.2 by demonstrating how frequency-selective multipath fading degrades the performance of basic modems designed for narrowband communications. Since all signal processing in wireless modems is performed digitally, in Section 9.3 we develop a statistical discrete-time model for the behavior of frequency-selective multipath fading channels. In this section we also introduce basic techniques that can be used to estimate the channel characteristics modeled by a discrete tapped delay line. The rest of the chapter is devoted to broadband wireless communication techniques operating over frequency-selective fading channels. In Section 9.4 we describe DMF and MLSE, in Section 9.5 we introduce different types of equalizers, in Section 9.6 we describe sectored antennas, in Section 9.7 we describe MCM and OFDM, and in Section 9.7 we compare these

traditional techniques. Section 9.9 is devoted to STC MIMO and MIMO OFDM as the latest advancements in broadband modem design technologies.

9.2 EFFECTS OF FREQUENCY-SELECTIVE MULTIPATH FADING

As the data rate of a modem increases beyond fractions of the rms multipath spread of the channel, the channel becomes frequency selective. In a frequency-selective fading multipath channel, a null may occur in the passband of the channel. Equivalently, we can say that the data rate is high enough with respect to the multipath spread that the multipath causes performance degradation due to ISI [Bel63b]. The performance degradation caused by ISI forces the performance curves into flat areas where any increase in the SNR does not improve the error-rate performance of the modem. The error rate obtained at these values is sometimes referred to as the *irreducible error rate* of the system.

To represent the effects of frequency-selective multipath fading quantitatively, we provide two examples in the following two sections. The first example shows the effects of frequency-selective fading on the performance of typical radio modems by evaluating the performance of the modem with a deep null in various locations in the passband of the channel. In the following section we consider a hypothetical two-path model and examine the effects of the distance and relative amplitude of the paths on the error rate of the modem.

9.2.1 Effects of Frequency Selectivity

Figure 9.1 shows the results of simulations performed to model a microwave LOS link exhibiting a spectral null in its amplitude frequency response [Bel84, Pah85a]. In these simulations, the spectral null was placed at various positions in the frequency band, ranging from the center to the edge of the band. Results are shown for four different modulation techniques. It can be seen from the figure that the performance degradation is greatest when the spectral null is at the center of the band, resulting in a BER for PSK modulation approaching 0.5, a very poor level of quality. With the null at the band edge, the BER is around 10^{-5} , an acceptable level of quality. The four modulations simulated are quadrature partial response (QPR), staggered QPR, QPSK, and staggered QPSK. It can be seen from the figure that QPSK is more resistant to the effects of the spectral null than either QPR or staggered QPR, but the degradation of QPSK performance is nevertheless considerable when the spectral null is at the center of the band.

9.2.2 Multipath Effects and Data-Rate Limitations

While signal fading, which arises from channel multipath, penalizes performance by greatly increasing the SNR required, the multipath also has a direct effect on performance in the form of intersymbol interference (ISI), as shown in Fig. 9.2. The top part of this figure shows the arrival of three consecutive pulses in a steady channel with no multipath. At the sampling time 0 only the sample associated with the symbol “1” has a nonzero value. At following sampling times of $T, 2T, 3T, \dots$, the samples from the symbol 1 have a zero value. As a result, the symbol 1 has no contribution in decisions

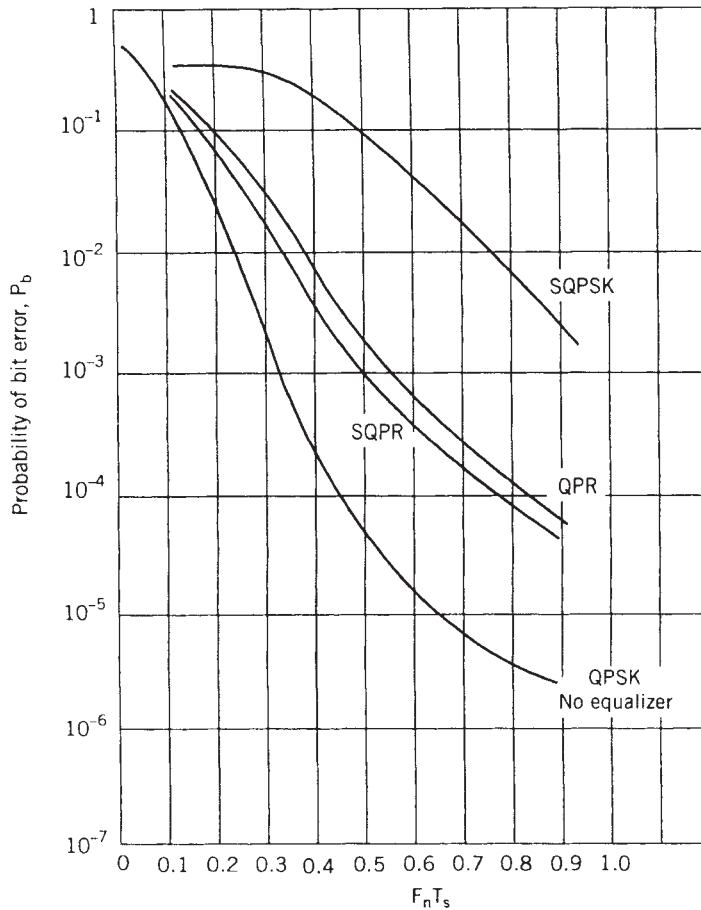


FIGURE 9.1 BER versus location of the null in a frequency-selective fading microwave channel.

made for other symbols. The lower part of Fig. 9.2 shows the received pulses in a two-path channel and the overall received pulse associated with one transmitted symbol, which is the summation of the two pulses received from the two paths. In this channel samples of the received overall pulse have nonzero values at $T, 2T, 3T, \dots$; this causes interference in decisions made on other symbols. Therefore, multipath causes intersymbol interference.

As we increase the symbol signaling rate in a multipath channel, the received symbols increasingly flow into one another, and this places an upper limit on rate at which symbols can be transmitted. Let us consider a theoretical example that will serve to illustrate this point. Figure 9.3 shows performance results obtained with a two-path channel model, where a_1 is the amplitude of one path, and the amplitude of the second path, a_2 , is allowed to take on various values relative to a_1 . The curves show the received bit-error probability for different values of separation between the two paths normalized by the symbol duration. The figure shows that for values of a_2 up to about 20% of a_1 , the ISI has very little effect on performance. However, for higher

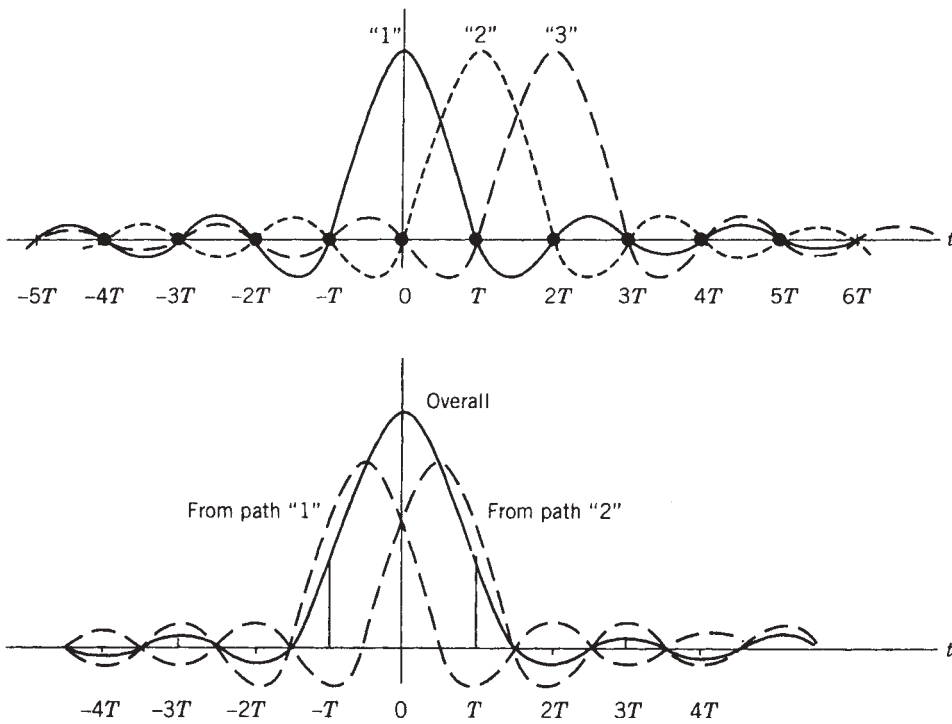


FIGURE 9.2 Intersymbol interference caused by multipath.

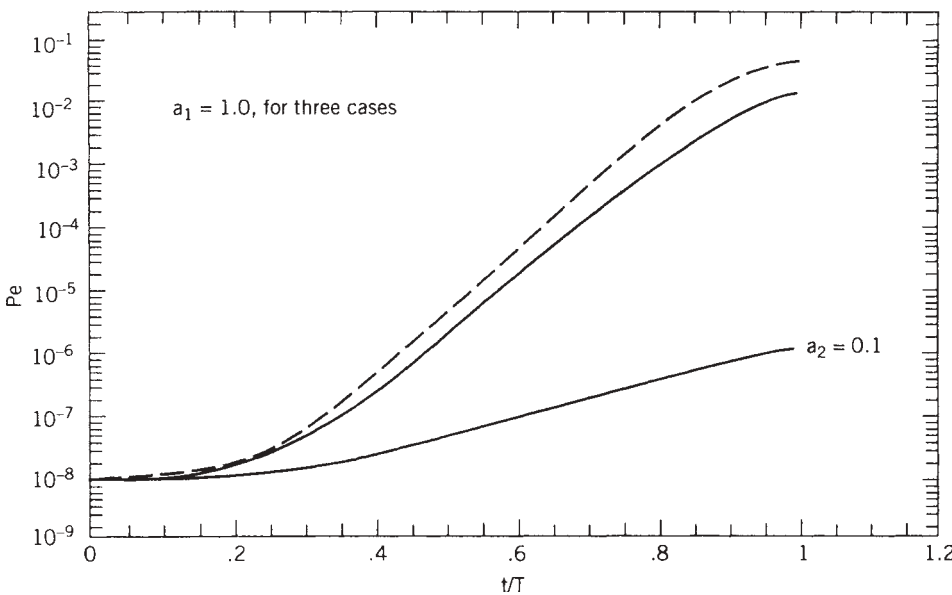


FIGURE 9.3 Probability of error versus normalized delay for a two-path channel model and three cases of relative path strengths. SNR = 15dB for all three cases when $t/T = 0$.

relative values of a_2 and when $t = T$, we find orders of magnitude of degradation in BER performance. These results give a general indication of how multipath affects performance as a function of data rate. For very low data rates the multipath has very little effect, but as the data rate increases, the performance can degrade markedly.

In communications environments where the received BER can vary widely with time or with location of the transmitter or receiver, the overall average received BER is not necessarily a useful measure of system performance. A performance measure which is often more useful is the *outage probability*, which is defined as the probability that the received BER lies above some preselected threshold of acceptable performance. In Table 9.1 we show outage probabilities that have been computed from signal measurements made in five factory locations (results presented in Chapter 5). The third, fourth, and fifth columns in the table give outage probabilities computed for BER thresholds of 10^{-2} , 10^{-4} , and 10^{-8} , respectively. These BER values represent performance thresholds that might be deemed acceptable for different data applications. The outage probabilities are calculated for BPSK modulation and are given for various data rates in the five manufacturing areas. The BER for each individual location

TABLE 9.1 Outage Probabilities Computed from Signal Measurements Made in Five Factory Locations

Area	Data Rate (MHz)	Outage for BER		
		10^{-2}	10^{-4}	10^{-8}
A	0.1	0.00093	0.00241	0.00444
	1.0	0.00130	0.00296	0.00630
	4.0	0.00741	0.01204	0.01778
	8.0	0.02667	0.03426	0.04333
	16.0	0.09778	0.10907	0.11926
B	0.1	0.00311	0.00867	0.02289
	1.0	0.00400	0.01089	0.02556
	4.0	0.02333	0.03333	0.05378
	8.0	0.07378	0.09044	0.11733
	16.0	0.23667	0.26733	0.30378
C	0.1	0.00067	0.00173	0.00440
	1.0	0.00680	0.00867	0.01387
	4.0	0.08347	0.09040	0.10227
	8.0	0.25293	0.26680	0.28453
	16.0	0.60373	0.61867	0.63600
D	0.1	0.00067	0.00200	0.00556
	1.0	0.01311	0.01800	0.02400
	4.0	0.13867	0.15178	0.16444
	8.0	0.39222	0.40889	0.42400
	16.0	0.67178	0.68289	0.69333
E	0.1	0.00015	0.00076	0.00182
	1.0	0.00273	0.00424	0.00606
	4.0	0.04045	0.04333	0.04788
	8.0	0.12121	0.12742	0.13439
	16.0	0.29136	0.29682	0.30515

Source: [How91].

was calculated using the measured channel impulse response in that location. All the points were calculated for 40-dB received SNR, in order to eliminate the effects of additive noise and include only the effects of multipath. All points were measured for transmitter-to-receiver distances below about 50 m. If we accept outage rates around 1% and set 10^{-4} as the outage threshold on BER, a data rate on the order of 1 Mb/s is supported in all five areas.

These results, which can be regarded as typical of many indoor wireless communications environments, indicate that a data rate of 1 Mb/s is often feasible in these situations. However, for applications such as WLANs we wish to provide higher data rates in the same environments. We can, of course, replace BPSK modulation with QPSK, which doubles the data rate, but we may want even greater increases for our intended application in order to make the data rate comparable to those of WLANs. In the remainder of this chapter we describe several ways in which this can be achieved. Analysis of this situation requires an understanding of the implicit in-band diversity in frequency-selective fading channels, which we describe first.

9.3 DISCRETE MULTIPATH FADING CHANNEL MODEL

Multipath has the harmful effect of causing ISI, but at the same time the signals arriving from different paths are exposed to different fading patterns. The multipath signals can be regarded as a form of diversity, and a smart receiver can use this diversity to improve its performance. In this section we assume that we have a wide-sense stationary uncorrelated scattering (WSSUS) frequency-selective fading multipath channel, and we show how we can take advantage of the multipath to provide diversity for the received signal. This diversity is not provided explicitly with multiple antennas or frequencies or with repeated transmissions, and thus it is referred to as *internal* or *implicit diversity*.

The baseband model that we consider, shown in Fig. 9.4, is a general model of a high-speed digital communication system operating over a fading multipath channel. Closely following Pahlavan and Matthews [Pah90c], in this discussion the model, extending from the information source in the transmitter to the input of the digital processor in the receiver, will be recast as a discrete tapped delay line.

The data sequence $\{a_k\}$ modulates an impulse train $\delta(t - kT)$, $k = 1, 2, 3, \dots$, where T is the symbol period. The impulse train acts as an input to a filter whose impulse response, $f(t)$, is the fundamental transmitted symbol. The output of the filter,

$$\sum_k a_k f(t - kT)$$

is the transmitted signal. It is assumed that $f(t)$ is bandlimited to $f_0 = W/2$ Hz.

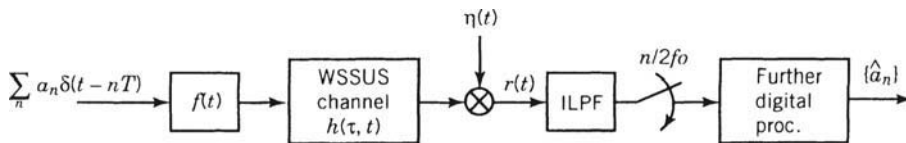


FIGURE 9.4 Overall baseband model of digital communication over a WSSUS fading multipath channel.

The signal passes through a fading multipath channel with time-variant impulse response $h(\tau, t)$. The function $h(\tau, t)$ represents the path gain associated with delay τ at time t . White complex Gaussian noise $\eta(t)$ is added to the information-bearing signal to form the received signal

$$r(t) = \sum_k a_k g(t, t - kT) + \eta(t) \quad (9.3.1)$$

where

$$g(u, t) = \int h(\tau, t) f(u - \tau) d\tau$$

The receiver would normally perform matched filtering by pulse-shaping filtering followed by a sampler at the output, operating at the symbol rate of $1/T$ samples/s. This is an optimal way to process $r(t)$ in the absence of frequency-selective fading, in which the received waveform is continually changing, and the matched filtering with the pulse-shaping filter does not provide optimum filtering. Because $f(t)$ and hence $g(t)$ are bandlimited, it is also possible to pass $r(t)$ through an ideal low-pass filter followed by a sampler operating at the Nyquist rate. In this manner the matched filtering can be deferred to a subsequent digital processor.

The sampled output is

$$r(t)|_{t=n/2f_0} = \sum_k a_k g(t - kT, t)|_{t=n/2f_0} + \eta(t)|_{t=n/2f_0}$$

or

$$r_n = \sum_k a_k g(t - kT, t)|_{t=n/2f_0} + \eta_n \quad (9.3.2)$$

Equation (9.3.2) is not in a convenient form because it represents sampling at one rate and transmission at another rate. Therefore, we define $M = 2f_0T$ and increase f_0 somewhat, if necessary, to ensure that M is an integer. We now define the discrete sequence

$$x_n = \begin{cases} a_{n/M}, & n/M \text{ an integer} \\ 0, & \text{otherwise} \end{cases} \quad (9.3.3)$$

which consists of the data interleaved with $M - 1$ zeros. Then the sampled output may be written compactly as

$$r_n = \sum_k x_k g_{n-k}(n) + \eta_n = \sum_k x_{n-k} g_k(n) + \eta_n \quad (9.3.4a)$$

where

$$g_k(n) = g(u, t)|_{u=k/2f_0, t=n/2f_0}$$

Because $g(u, t)$ is bandlimited, it cannot also be time limited. However, in all practical situations, $g(u, t)$ will have a finite duration by some suitable engineering definition. Let L denote the number of samples in that duration.¹ The model corresponding

¹Because $E[|g_k(n)|^2]$ is the average power associated with a tap weight, one method of choosing L is to retain all taps whose power is within “ x dB” of the largest tap power. The criterion of 10 dB was found to be a good choice for the numerical results presented later.

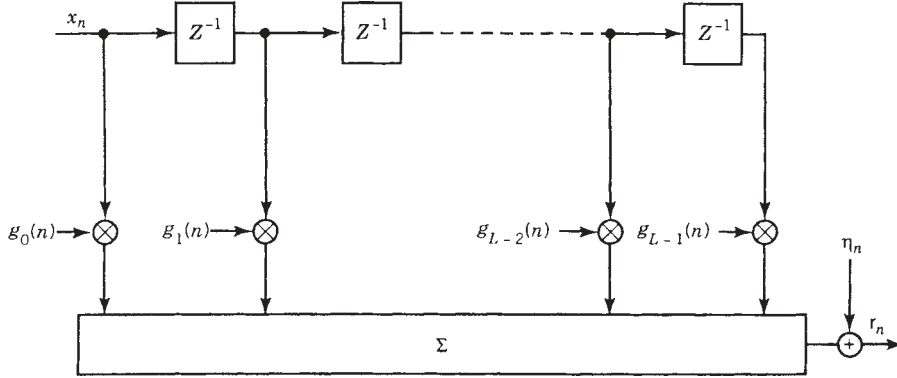


FIGURE 9.5 Discrete channel model representing transmitter filtering, the WSSUS channel, additive noise, and prefiltering at the receiver.

to Eq. (9.3.4a) is shown in Fig. 9.5, where sufficient delay is added to make the model causal.

This is a multirate discrete system because the rate at which information enters the system is not the same as the sampling rate before digital processing. The transmitted, symbol, x_n , appears at the input of the equivalent discrete-time channel model every M samples, and we have L samples of r_n for each transmitted information symbol x_n . If the transmission system is designed so that $M \geq L$, there is no ISI and there is a vector of sampled received sequences with length L corresponding to each single symbol a_n . Therefore, Eq. (9.3.4a) reduces to

$$r_{n-k} = a_n g_k(n) + \eta_{n-k} \quad 0 \leq k < L \quad (9.3.4b)$$

at the baud rate.² In vector form this equation is written as $\mathbf{R} = a_n \mathbf{G} + \boldsymbol{\eta}$. For each transmitted symbol a vector of length L is received, resulting in L -order implicit diversity. To normalize Eqs. (9.3.4a) and (9.3.4b), throughout the analysis it is assumed that $a_n = \pm\sqrt{E_b}$, where E_b is the transmitted energy per bit,

$$\sum_{k=0}^{L-1} E\{|g_k(n)|^2\} = 1$$

and $E\{|\eta_n|^2\} = N_0$. Then the average SNR per bit is $\bar{\gamma}_b = E_b/N_0$. In the absence of ISI the optimum receiver is a discrete matched filter (DMF), which is a maximal-ratio combiner for the implicit diversity of the channel. In the presence of ISI, we have $M < L$ and additional signal processing is needed to eliminate the ISI. Maximum likelihood sequence estimation (MLSE) and decision feedback equalization (DFE) are the two methods considered in this case.

9.3.1 Adaptive Channel Measurement

The channel measurement techniques that we studied earlier were used for the development of statistical models for radio propagation. In this section we introduce another

²Note that during the transmission time of several symbols the channel is assumed to be fixed; that is, $g_k(n) = g_k(n+k)$ for small values of k .

concept in channel measurement. Here we measure the taps of the discrete time-domain channel model developed in Section 9.3. The results of these measurements are used to implement adaptive receivers that take advantage of the in-band diversity. In the next section we introduce two specific receivers, the *adaptive discrete matched filter* (ADMF) and the adaptive MLSE, using the adaptive channel estimator introduced in this section as a part of the receiver.

Estimation of the taps of the overall channel impulse response requires a reference signal. The reference signal can be a transmitted probe sequence known to the receiver, or it can be some sequence of detected symbols. The probe signal can be time-multiplexed with the transmitted information symbols, or it can be overlaid onto the traffic channel as a weak spread-spectrum signal. The effects of the transmitted reference signal on the overall performance of an ADMF are analyzed in [Pah90c].

MMSE Solution for a Channel. The input to the channel estimator is given by

$$r_n = \sum_{k=0}^{L-1} g_k(n)x_{n-k} + \eta_n \quad (9.3.5)$$

where $g_k(n)$ is the equivalent discrete channel impulse response, x_n is the reference signal, and η_n is the additive noise. The channel estimator forms estimates of the channel impulse response $\hat{g}_k(n)$ and the received signal $\hat{r}_n$, which are related as follows:

$$\hat{r}_n = \sum_{k=0}^{L-1} \hat{g}_k(n)x_{n-k}$$

It is well known [Pah79, Pro89] that the minimum mean-squared error (MMSE) estimate of the channel impulse response that minimizes $E\{|r_n - \hat{r}_n|^2\}$ satisfies the following linear set of simultaneous equations:

$$\mathbf{A}\hat{\mathbf{G}} = \mathbf{b} \quad (9.3.6a)$$

where $\hat{\mathbf{G}}$ is the vector of MMSE estimates of the tap gains, and the elements of matrix $\mathbf{A}$ and the vector $\mathbf{b}$ are defined by

$$\begin{aligned} A_{ij} &= E\{x_{n-j}x_{n-i}\}, & 0 \leq i, j < L \\ b_i &= E\{r_n x_{n-i}\}, & 0 \leq i < L \end{aligned} \quad (9.3.6b)$$

For a reference signal with statistically independent symbols, such as a pseudonoise (PN) sequence, $\mathbf{A}$ is a diagonal matrix with E_b the value of all diagonal elements. The solution to Eq. (9.3.6a) is then

$$\hat{g}_k(n) = \frac{1}{E_b} E\{x_{n-k}r_n\} = g_k(n), \quad 0 \leq k < L \quad (9.3.7)$$

Implementation Using Cross-Correlation, and the LMS Algorithm. There are two ways to implement the channel estimator discussed above. One is to use cross-correlation to calculate Eq. (9.3.7) directly. The second is to use an equalizer-like channel estimator described in [Mag73]. This estimator is implemented by applying the

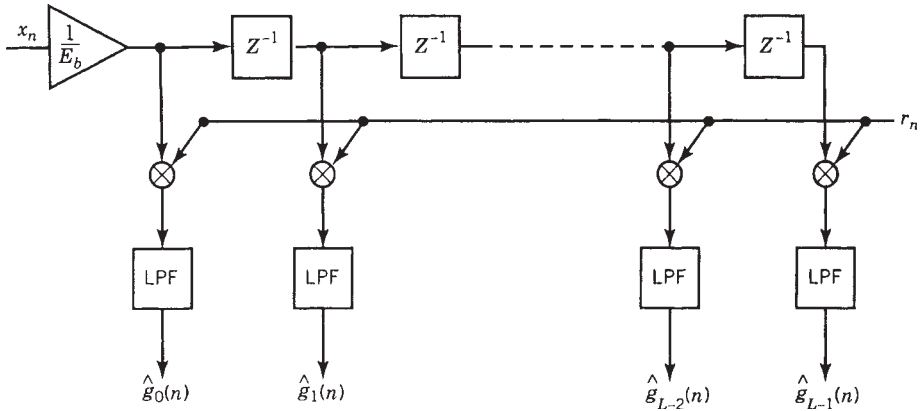


FIGURE 9.6 Discrete channel estimation using cross-correlation [Pah90c].

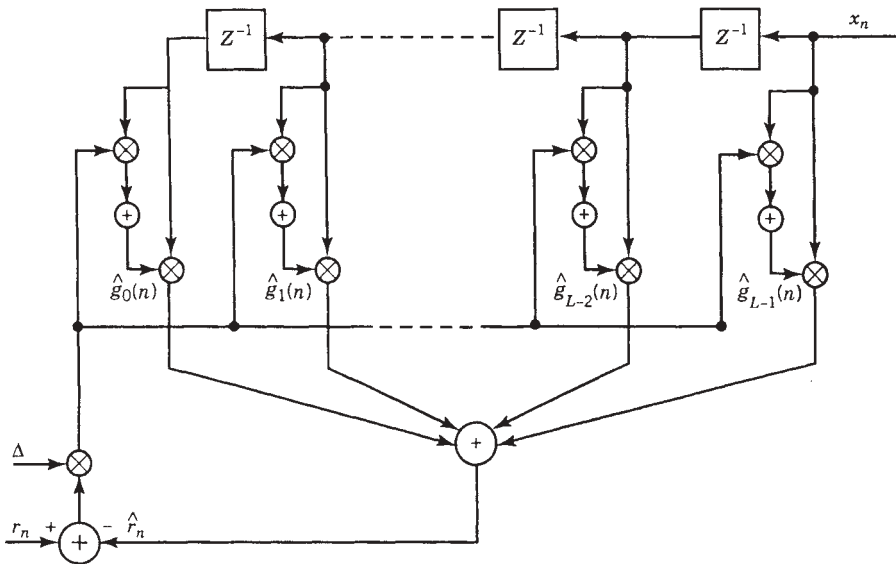


FIGURE 9.7 Adaptive channel estimator using the LMS algorithm.

steepest-descent algorithm to minimization of mean-squared error. The steepest-descent algorithm was first introduced and named the least-mean-squared (LMS) algorithm by Widrow [Wid66, Pro89].

Figure 9.6 shows the implementation of discrete channel estimation using cross-correlation. The reference signal is passed through a tapped-delay line (TDL), and the tapped signals are multiplied with the received signal and then passed through a bank of low-pass filters. To normalize the output of the filters to actual samples of the channel impulse response, one of the signals is scaled to $1/E_b$, as indicated in Eq. (9.3.7).

Figure 9.7 shows equalizer-like channel estimation using the LMS algorithm. Here the reference signal is passed through a TDL, and the channel tap gains are adjusted

according to

$$\hat{g}_j(n+1) = \hat{g}_j(n) + \Delta e_n x_{n-j} \quad (9.3.8)$$

where Δ is the adjustment rate parameter and

$$e_n = r_n - \hat{r}_n = r_n - \sum_{k=0}^{L-1} \hat{g}_k(n) x_{n-k} \quad (9.3.9)$$

The adaptive channel measurement techniques used with MLSE are described in [Mag73, Ung74], with application to slowly time-varying channels. Analysis of adaptation techniques used for fading multipath radio channels is available in [Pah79].

9.4 ADAPTIVE DISCRETE MATCHED FILTER

The simplest way to avoid ISI is to reduce the symbol rate relative to the transmission bandwidth, which also reduces the bandwidth efficiency of the system. There are two approaches to implementing this technique. One approach is to provide a time guard between the transmitted symbols so that signals arriving along different paths and associated with different symbols do not interfere with one another. The time guard should be equivalent to the maximum multipath delay spread of the channel so as to avoid the ISI completely, as shown in Fig. 9.8. In this case the bandwidth efficiency of the system is reduced $T/(T + \tau_m)$, where T is the duration of the pulse and τ_m is the maximum multipath spread of the channel. As explained in Chapter 10, a second approach to avoiding ISI is to use direct-sequence spread-spectrum transmission. The adaptive matched filter in this case is referred to as a *RAKE receiver*. The advantage of spread spectrum for multiuser applications is that the loss in the bandwidth efficiency can be overcome by the use of code-division multiple access. Time gating has been used in less bandwidth-sensitive applications such as troposcatter communication, whereas the spread-spectrum technique is used in a number of recently developed wireless networks.

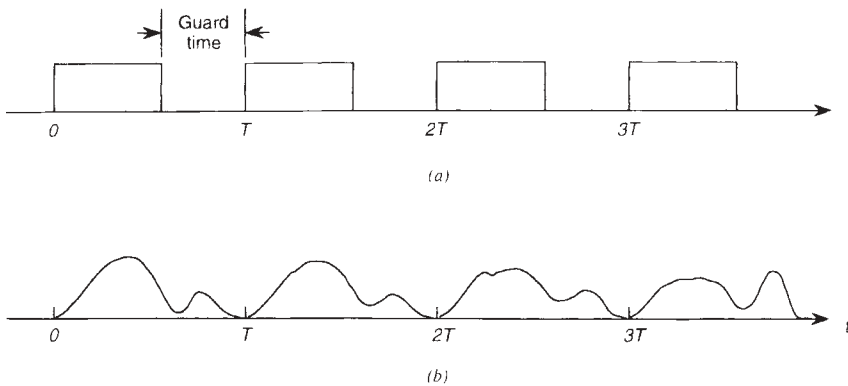


FIGURE 9.8 Time gating, a technique for avoiding intersymbol interference: (a) transmitted symbols; (b) received symbols.

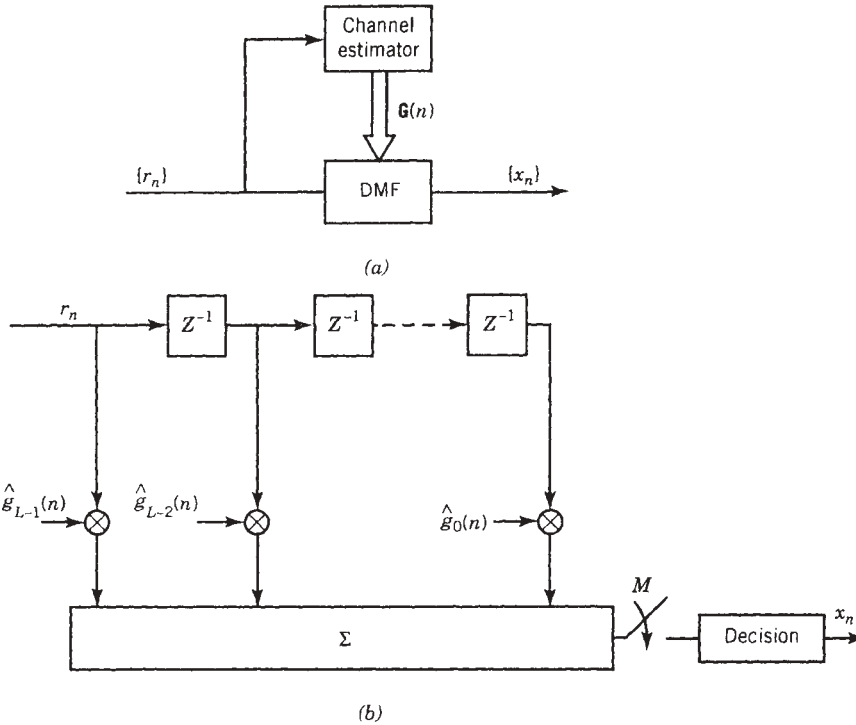


FIGURE 9.9 Adaptive matched filtering: (a) general block diagram for the adaptive discrete matched filter (ADMF); (b) implementation of the DMF using a finite impulse response digital filter.

In the absence of ISI, the optimum receiver is an adaptive discrete matched filter consisting of two parts: an adaptive channel estimator and a DMF that is a maximal ratio combiner of the implicit (in-band) diversity. Figure 9.9a shows a general block diagram for the adaptive DMF (ADMF). The adaptive channel estimator, described in Section 9.3, evaluates the instantaneous values of the taps in the equivalent discrete channel model. The DMF uses the estimated values of the channel impulse response to perform maximal ratio combining the diversity signal represented by the tapped channel model. Figure 9.9b shows the implementation of the DMF using a finite impulse response digital filter. The taps of the DMF are the complex conjugates of the taps of the discrete channel impulse response in reverse time order. If we consider the overall discrete model, this is the discrete match filter for the overall discrete channel impulse response, and thus is the optimum receiver in the absence of ISI. Another way of looking at the DMF is to consider each tap of the overall channel impulse response as a branch of the inband diversity. Then the DMF multiplies each branch with the complex conjugate of the amplitude of that branch, which constitutes maximal-ratio combining under the assumption of equal noise powers on the diversity branches.

9.4.1 Performance Prediction for the DMF

If we assume that the in-band diversity branches are all Rayleigh fading signals, Eq. (8.3.5), presented in Chapter 8 for calculation of maximal-ratio combining

performance, is applicable to performance analysis of the DMF with order of in-band diversity L replacing the order of external diversity D . With this equation the average probability of error for coherent PSK as a function of average signal-to-noise ratio $\bar{\gamma}_b$ was given by

$$\bar{P}_e = \sum_{n=0}^{L-1} \frac{A_n}{2} \left[1 - \left(\frac{2\bar{\gamma}_b \lambda_n}{1 + 2\bar{\gamma}_b \lambda_n} \right)^{1/2} \right] \quad (9.4.1)$$

where $\{\lambda_n\}$ are the eigenvalues of the covariance matrix of tap gains and

$$A_n = \prod_{\substack{k=0 \\ k \neq n}}^{L-1} \frac{1}{1 - \lambda_k / \lambda_n}$$

In terms of the parameters of the model, the elements of the covariance matrix are given by

$$r_{ij} = E\{g_i g_j^*\} = \int_{-\infty}^{\infty} f\left(\frac{i}{2f_0} - u\right) Q(u) f^*\left(\frac{j}{2f_0} - u\right) du$$

where $Q(\tau)$ is the delay power spectrum of the channel and $f(t)$ is the impulse response of the pulse-shaping filter.

Example 9.1: Error Rate for Exponential Delay Power Spectrum For a numerical example, let us assume that the delay power spectrum is given by

$$Q(\tau) = \begin{cases} b^2 r e^{-h\tau}, & \tau \geq 0 \\ 0, & \tau < 0 \end{cases}$$

where b is related to the channel rms multipath spread τ_{rms} through the expression

$$b = \frac{2\sqrt{2}}{\tau_{\text{rms}}}$$

This delay power spectrum is an approximation found by curve fitting to typical measurements made on troposcatter links [She75] and is consistent with theoretical predictions given in [Bel69].

The system considered here for performance prediction uses a 100% raised-cosine pulse with a bandwidth of 2 MHz. The samples at the receiver are taken at the Nyquist rate, 4 Mb/s. With the ratio of the rms multipath spread to symbol interval set at $\tau_{\text{rms}}/T = 0.11$, the equivalent tapped delay-line model for this example contains four significant tap gains. The average received power for each tap gain and the eigenvalues of the covariance matrix of the tap gains are given in [Pah79]. Using these values in Eq. (9.4.1), a plot of average PSK error probability versus $\bar{\gamma}_b$ is generated, as shown in Fig. 9.10. The figure also includes the performance calculated for a one-path case. The performance gain achieved through optimum exploitation of the in-band diversity is shown to be about 10 dB at $\bar{P}_e = 10^{-5}$.

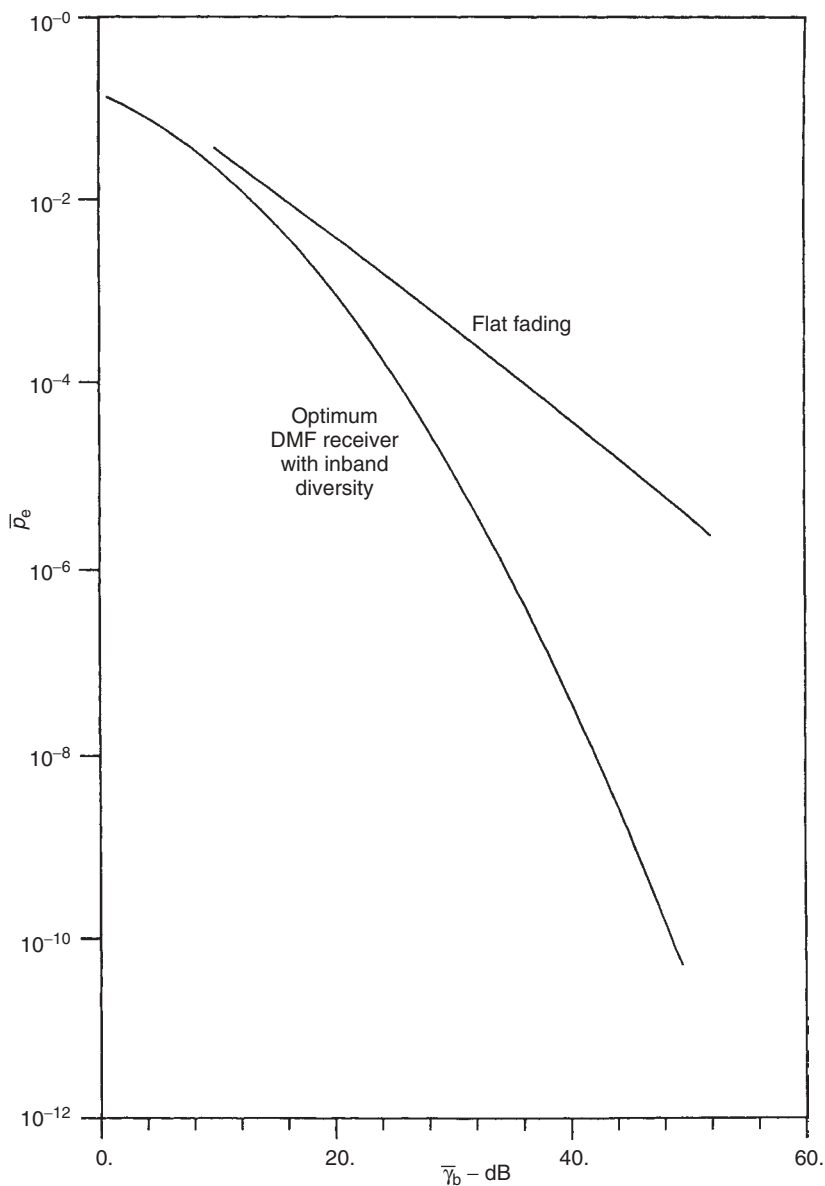


FIGURE 9.10 Performance of a discrete matched filter receiver operating on a troposcatter channel. The single-path flat-fading case is included for comparison.

In the DMF example above we assumed that the channel estimator makes perfect estimates of the tap weights in the channel model. In practice, the adaptive channel estimator introduces measurement noise, which somewhat degrades the performance of the system from the theoretical optimum. The effect of the noise depends on the approach used in providing the reference signal, on the Doppler spread of the channel, and on the parameters of the adaptive filters. The influence of these factors on ADMF performance is discussed in [Pah90c].

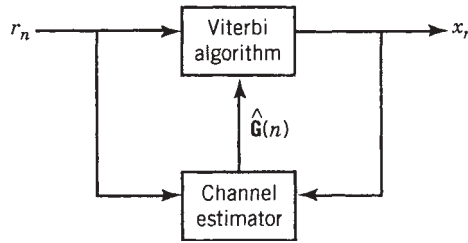


FIGURE 9.11 Adaptive maximum likelihood sequence estimation.

9.4.2 Adaptive MLSE

The MLSE is the optimum receiver in the presence of ISI. Given estimates of the channel impulse response, an MLSE receiver uses a trellis diagram with the Viterbi algorithm to obtain maximum-likelihood estimates of the transmitted symbols. The adaptive MLSE receiver shown in Fig 9.11 is similar to the ADMF. It consists of two parts, the adaptive channel estimator and the MLSE algorithm. The sampled channel impulse response is measured with the adaptive channel estimator described in Section 9.3. Given the samples of channel impulse response, the sequence of sampled received signals is compared with all possible received sequences. The maximum likelihood procedure [For72] is to determine the squared distance between the sequence of the sampled received signal and all possible received sequences and to determine the transmitted symbol sequence with minimal distance. The Viterbi algorithm [Vit67] is used as the computationally optimal and efficient solution to this search. This application of the Viterbi algorithm is similar to other well-known applications, including decoding of convolutional error-control codes and demodulation of TCM signals. If the ISI is negligible, the performance of the MLSE is the same as that of the DMF. With ISI, the performance of the ADMF provides a tight upper bound to the performance of the adaptive MLSE and can be used as an approximation for purposes of performance prediction.

The MLSE is the optimal method of canceling the ISI; however, the complexity of this receiver grows exponentially with the length of the channel impulse response, whereas the complexity of the equalizer grows only linearly with the length of the impulse response. For this reason, MLSE is an attractive option for channels with short impulse responses, but for longer impulse responses, equalizers are more practical. In the radio communication literature, MLSE is usually compared with decision feedback equalization (DFE), which is the equalizer of choice for frequency-selective fading multipath channels and is discussed in more detail below. Comparisons of MLSE versus DFE performance are given for telephone line modems in [Fal76b, Fal76c], for HF radio in [Fal85], and for troposcatter radio links in [Cha75, Mon77].

9.5 ADAPTIVE EQUALIZATION

A third technique for increasing the data rate in multipath fading environments is adaptive equalization. Frequency-selective fading channels produce ISI in the received signal. In these channels an increase in the power does not improve performance, because additional power amplifies the ISI in step with the desired signal. The

traditional method of compensating for ISI is to equalize the channel impairment by applying a filter at the receiver. In general, channel characteristics are subject to variations in time, which leads to a need for adaptive equalizers. Tapped delay line (TDL) equalizers are the most commonly used equalizers, and their detailed analysis and theory of operation are given in this section. For more detailed treatments of equalization techniques, the reader may refer to [Qur85, Pro01].

The principles of operation for TDL equalizers are the same as those for the MMSE channel estimator discussed earlier, and all of these techniques are considered special cases of adaptive filtering. Equalizers have wider applications in telecommunications, and they were introduced and applied earlier than the other techniques. For these reasons, more detailed attention in this chapter is devoted to equalizers. Here, issues arising in the application of TDL equalizers to radio modems, such as sensitivity to frequency-selective fading, sensitivity to timing recovery, variations related to choice of modulation technique, and bounds for performance evaluations, are discussed. We start with a review of various equalizer architectures and then examine the effects of equalization on the performance of radio modems over frequency-selective fading channels. This discussion is followed by analysis of performance over measured indoor radio channels. Here we provide a bound for calculation of the probability of outage and average error rate on radio channels.

9.5.1 Equalizer Architectures

Equalizer architectures considered for modems are linear transversal equalizers (LTEs), linear fractionally spaced equalizers (FSEs), decision feedback equalizers (DFEs), pass-band equalizers, and blind equalizers. We next discuss these architectures, with detailed emphasis on LTE, FSE, and DFE equalizers.

Linear Transversal Equalizer. The LTE, shown in Fig. 9.12, is the earliest of the TDL equalizers. The received signal is passed through a tapped delay line with tap spacing $\Delta = T$. The tapped signals are weighted and added to form the equalized output signal. The optimum tap gains are determined using either the zero-forcing or MMSE criterion. In a zero-forcing algorithm, the tap gains are adjusted to force the overall sampled

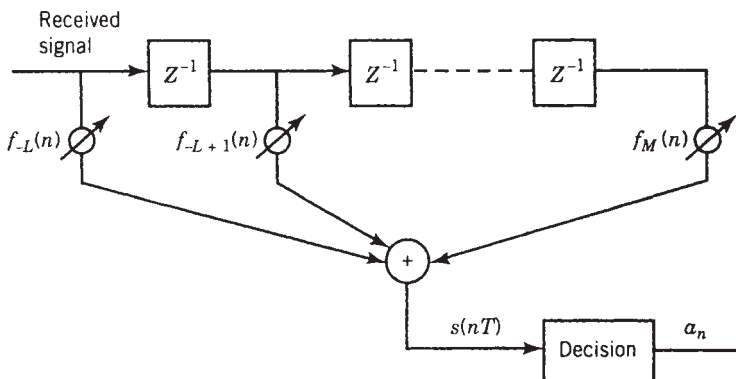


FIGURE 9.12 Tapped-delay-line structure for a linear transversal equalizer.

impulse response after equalization to have minimum ISI [Luc65, Luc66, Pro01]. For MMSE equalization, tap gains are adjusted to minimize the mean square of the error signal between equalized signal and the actual transmitted symbols [Pro01].

The equalizer is a discrete-time filter intended to compensate for the amplitude and phase distortions of the channel. One can see intuitively that for an infinite-tap equalizer, the sampled frequency response of the equalizer should be the inverse of the frequency response of the channel, which is a correct characterization of zero-forcing equalization. In MMSE equalization, the tap gains are set so as to minimize both channel distortions and additive noise, and therefore the frequency-response function of the equalizer depends in part on the variance of the noise.

Today, MMSE is the dominant criterion used in the design of equalization algorithms, and various versions of the equalizer architectures are developed with this criterion. As a result, we focus primarily on MMSE equalization.

Fractionally Spaced Equalizer. The structure of the FSE is the same as that of the LTE, shown in Fig. 9.12, except that in FSE the tap spacing is $\Delta = kT/n$, where k and n integers and $k < n$. This minor difference in spacing of the taps results in three basic advantages for a modem using FSE over a modem using LTE. (1) A modem with a FSE is relatively insensitive to timing error, (2) it does not require accurate pulse shaping, and (3) it is more effective in treating the channel phase distortions at the edges of the band. For these reasons, FSEs are widely used for both voiceband and radio modems.

It has been found that whereas an FSE will perform well in computer simulation with floating-point arithmetic, in real-time implementation with fixed-point arithmetic and limited word length, the tap gains tends to diverge. An investigation of tap gain divergence and its relationship to characteristics of the channel is given in [Git82]. In general, the simple solution to the tap gain divergence problem is the *tap-leakage algorithm* [Ung76, Git82]. This ad hoc solution introduces small corrections in the tap gains to prevent them from growing excessively.

Another issue concerning the comparison between LTE and FSE equalizers is that with the same amount of hardware complexity, an LTE spans a longer time interval, which intuitively suggests superior performance for an LTE. However, results of computer simulations reveal that with the same number of taps, an FSE performs at least as well as an LTE, and the FSE provides significantly better performance on channels with severe phase distortions at the edges of the band [Qur77].

Decision Feedback Equalizer. The DFE [Sal73, Bel79b] shown in Fig. 9.13 consists of two tapped-delay-line filters, referred to as *forward* and *backward equalizers*. The input to the forward equalizer is the received signal, and it operates similarly to the linear equalizers discussed earlier. The input to the backward equalizer section is the stream of detected symbols. The tap gains of this section are estimates of the channel sampled impulse response, including the forward equalizer, and this section cancels the ISI due to past samples. For finite-tap equalizers the backward equalizer of length M eliminates the ISI due to M past symbols.

A linear equalizer is notable, properly, to equalize a channel having a deep null in the passband. For such a channel, the linear equalizer applies high gain in its frequency response to compensate for the null, which in turn causes noise enhancement, as we noted earlier. However, the backward filter of a DFE does not suffer

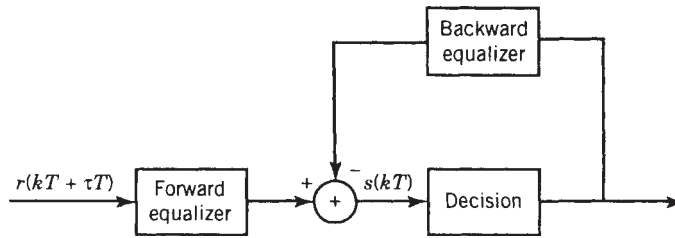


FIGURE 9.13 Decision feedback equalizer. The forward and backward equalizers are both tapped-delay-line equalizers.

from the noise enhancement problem, because it estimates the channel rather than its inverse. As a result, for channels with deep nulls in the passband, DFEs are superior to linear equalizers.

In frequency-selective fading radio channels, channels occasionally experience deep nulls in the passband, resulting in unsatisfactory performance for linear equalizers. For these channels, the DFE is the preferred design, and DFEs have been examined for troposcatter [Mon77], HF [Fal85], and microwave line-of-sight [Bel84, Pah85a] channels. On telephone channels, the most significant amplitude and delay distortion is found at the edges of the passband. As a result, a DFE, which is more effective against nulls in the middle of the passband, has little to offer over a linear equalizer with a large number of taps. Thus, linear equalization is the predominant design choice for voiceband data modems operating over telephone channels.

Blind Equalization. Standard equalizers are trained with a known sequence at the start of data transmission. After training, with high probability, the decisions at the output of the equalizer are the same as the transmitted symbols and are used as the reference for the tap gain adjustment. In many cases, both radio and telephone line modems need to be retrained without interruption of normal data transmission. The retraining algorithm should be independent of any particular information about the sequence of transmitted symbols. This equalization is usually referred to as *blind equalization*. A simple algorithm for blind equalization, which uses the square of the error signal for adjustment of equalizer tap weights, is given in [Fal76a], and today this algorithm is used widely to retrain voiceband modems after an interruption of data. Because in these algorithms reference to transmitted symbols is unavailable at the receiver, the convergence is about one order of magnitude slower than with standard LMS equalization algorithms. Additional analysis of the transient behavior of blind equalizers, applied to radio communications, is available in [God80, Fos85].

Bandpass Equalization. Bandpass equalization [Git73, Fal76a] is performed before demodulation—in contrast with standard baseband equalization, which is done after demodulation. When the phase reference of the oscillator in the demodulator is obtained from the data stream detected, the delay between demodulation and phase recovery is smaller in passband equalizers, resulting in faster tracking of phase variations and, consequently, a more robust modem. However, in modern digital designs it is possible to demodulate without a phase reference and to adjust the phase by multiplying the demodulated symbol with a numerical phasor that shifts the point in the constellation into its proper place. For these implementations, there is no difference between

baseband or bandpass equalization. In radio modems operating at high carrier frequencies, surface acoustic wave (SAW) devices sometimes provide a cost-effective solution for implementation of the equalizer. The SAW devices can operate at the high frequencies required; and given their cost advantage, passband equalization at either IF or RF is preferred to baseband equalization.

Example 9.2: Equalization of a Null in the Passband Here we provide some numerical examples to compare the effectiveness of different equalizers on frequency-selective fading channels. As an example applicable to radio communications, consider a frequency-selective fading channel whose frequency response is given by

$$H(j\omega) = A + B(j\omega)$$

where A and B are complex numbers. This model is used to represent frequency-selective fading in microwave line-of-sight (LOS) channels.

The transfer function of this channel has one zero at $s = -A/B$, which results in a null at the frequency $\omega = \text{Im}[-A/B]$. The depth of the notch is determined by $\text{Re}[-A/B]$. Figure 9.14 shows plots of the inverse of the MMSE versus normalized timing error for $A = 1$, $B = 0.4$, and various numbers of taps for a DFE working with an SQPR modem. The inverse of the MMSE is a measure of SNR after equalization, which can be viewed as a performance criterion for an equalizer. The SNR before equalization is 23 dB. The forward tap gains are associated with $\Delta = T/2$, and plots include various numbers of forward (N) and backward (M) tap gains. As the number of forward tap gains increases, the flat region of the curve widens, indicating a larger span of time in which the sampling can take place without significant performance degradation. The insensitivity to sampling time will significantly reduce the complexity of the timing recovery circuit. An increase in the number of backward taps will increase the SNR. For this particular example, three forward and three feedback taps are adequate to provide a wide region of allowed sampling times and a maximum SNR.

Figure 9.15 shows the probability of error versus location of a deep null in the passband of the channel for a QPSK modem. The channel model is the same as above, but the parameters are adjusted to force a null on frequency axis. The probability of error is calculated by determining the overall sampled impulse response after equalization and calculating the average error rate for all possible combinations of the ISI [Pah85a]. Plots include the modem without an equalizer, an FSE with five $T/2$ -spaced taps, and a DFE modem with three forward and three feedback taps. For the deep null near the center ($FT \simeq 0$), only the DFE shows an acceptable level of performance. As the null moves toward the edges of the passband, the performance of the other two modems improves rapidly. As a result, one can conclude that for robust performance over a range of channel conditions, a DFE is required.

9.5.2 Adaptive Algorithms for Equalizers

In Section 9.5.1 we analyzed the effectiveness of various equalizers under the MMSE criterion and found that the optimal tap gains in each case are determined from a set of linear equations. Direct solution of these equations requires measurement of the instantaneous overall channel impulse response, measurement of the variance of

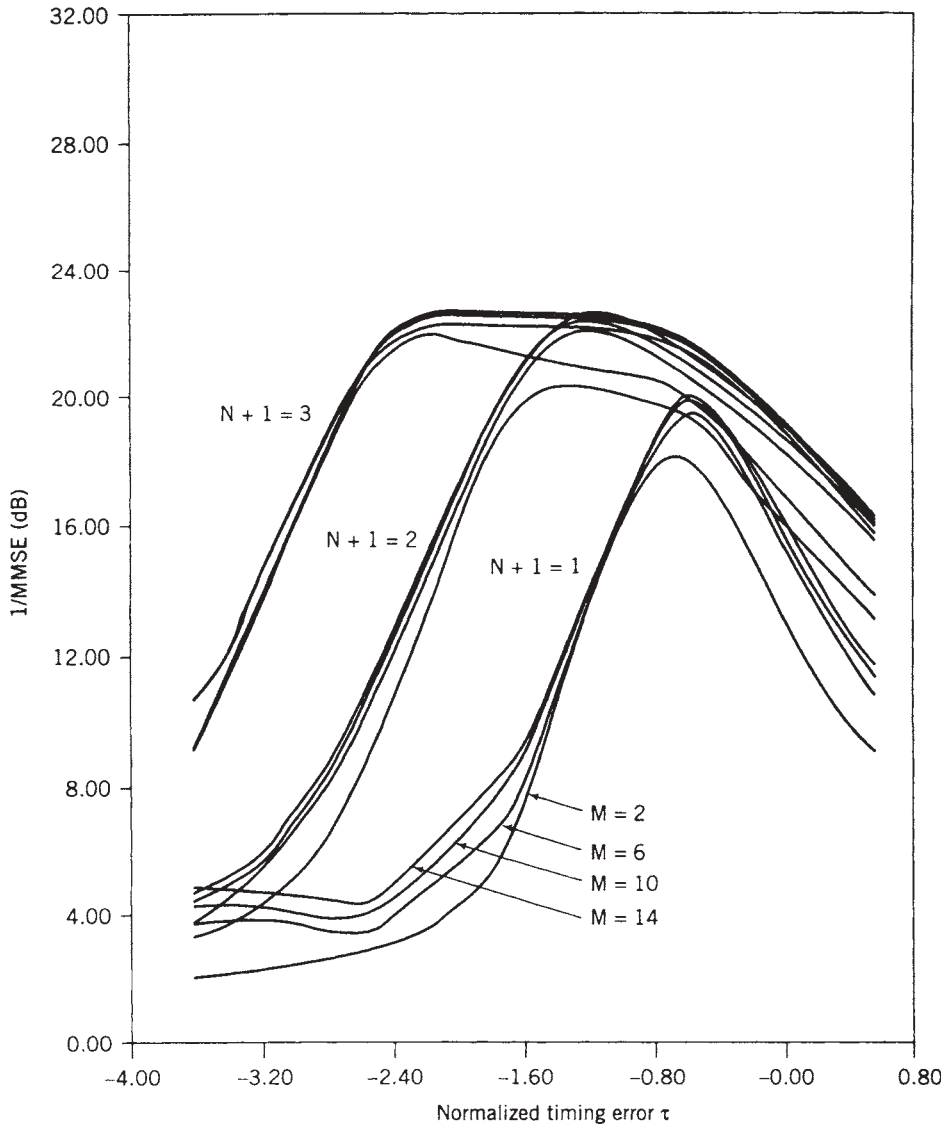


FIGURE 9.14 Inverse MMSE versus normalized timing error for a staggered quadrature partial response DFE operating over a frequency-selective microwave LOS channel. The inverse of the MMSE is a measure of SNR after equalization. $N + 1$ number of forward taps; M = number of feedback taps; decision variable SNR = 23 dB with no distortion [Bel84].

the noise, formation of the covariance matrix, and solution of a large set of linear equations. This method requires extensive numerical calculations, which is by no means economically attractive for real-time applications. For real-time computations we need a numerically efficient method of solution. Fortunately, both the peak distortion function used in the zero-forcing algorithm and the MSE function used in MMSE equalizers are convex functions of the tap gains, which allows application of the LMS algorithm. The LMS algorithm is computationally attractive and can easily adapt to channel variations.

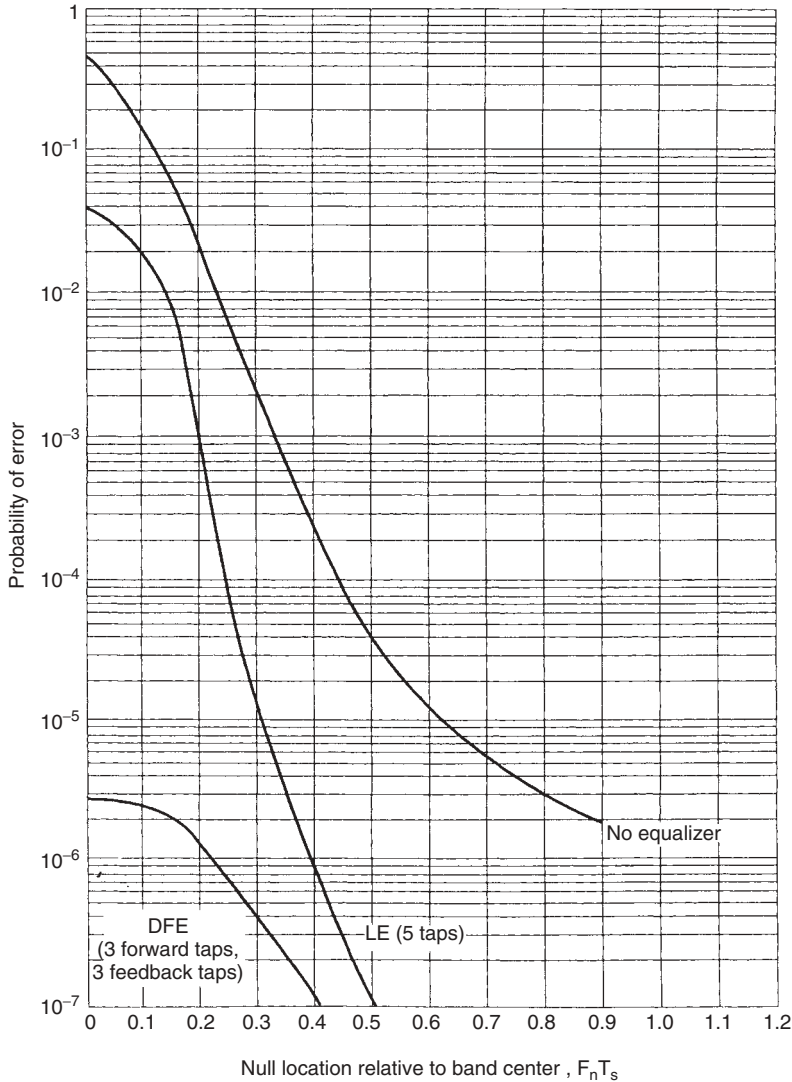


FIGURE 9.15 Probability of error versus location of a null in the passband of a microwave LOS channel for a DFE with three forward and three backward taps, an LTE with five taps, and the QPSK modem without equalization [Pah85a].

Application of the LMS algorithm to adaptive equalization was first done for the peak distortion criterion [Luc65, Luc66] and was subsequently investigated for adaptive MMSE equalization [Pro01].

The LMS algorithm implementation for the forward tap gains $\{f_l\}$ of the equalizer is given by

$$f_l(k+1) = f_l(k) + \mu e(kT) r(kT + \tau T - l\Delta)$$

where μ is a small constant known as the *step size*, $e(kT)$ is the error between the equalizer output and the transmitted symbol, and $r(t)$ is the received signal at the

input of the equalizer. For the feedback taps of the equalizer, we have

$$b_l(k+1) = b_l(k) + \mu e(kT)d_{k-l}$$

where d_k is the transmitted symbol. The LMS algorithms was found to be slow for some applications, and rapidly converging algorithms for fast startup equalization and fast tracking of rapidly changing channels were studied extensively. For an overview of these algorithms, see [Pah88b]. For discussion of the methods applied in voiceband modems, see [Qur85], and for detailed derivation of specific algorithms, see [Pro01].

9.5.3 Performance of a DFE in Fading Multipath Channels

Data rates on the order of 10 Mb/s are desirable for WLANs in order to make them compatible with existing wired or cabled LANs. Severe multipath fading, which is characteristic of indoor radio channels, limits the achievable data rates of BPSK modems to less than 1 Mb/s.

In this section we closely follow Pahlavan et al. [Pah93], who used measured channel profiles from several indoor sites to determine the probability of outage and the average probability of error of BPSK/DFE modems. It is shown that a BPSK/DFE modem with three forward and three feedback taps provides a data rate an order of magnitude higher than a BPSK modem operating without equalization. The results of simulation based on measured channel profiles are compared with performance predictions obtained from computer simulations and analytical bounds.

Analytical Performance Prediction. As we saw in the analysis of the DMF and the specific example in that section, if the channel is assumed WSSUS and the delay power spectrum is known, it is possible to develop an analytical framework for performance prediction. This analysis can be extended to performance evaluation of the DFE with an infinite number of feedback taps [Mon77]. If we wish to apply these analytical results to portable and mobile radio channels, we need to determine the delay power spectrum based on measured data. For fading multipath channels such as troposcatter, the multipath is caused by the reflection from small scattering elements in the atmosphere, which are changing continuously. In these channels, for a fixed location of the transmitter and receiver, the channel impulse response is a continuous-delay random process $h(\tau, t)$. As we explained in Chapter 3, if we assume that the channel is WSSUS, we have

$$E\{h(\tau, t)h^*(\tau_1, t_1)\} = Q(\tau, t - t_1)\delta(\tau - \tau_1)$$

where $Q(\tau, t)$ represents the received power as a function of multipath delay τ , measured at time t , and $\delta(\tau)$ is the delta function. For a slowly fading channel, where the Doppler spread of the channel is much lower than the data rate, $Q(\tau, t)$ is approximated by

$$Q(\tau, t) \simeq Q(\tau, 0) = Q(\tau)$$

in which $Q(\tau)$, referred to as the *delay power spectrum* of the channel [Bel69], represents the time-averaged received power as a function of the delay between arrivals of different paths. It can be viewed as the time average of the squared magnitude of the channel impulse response.

For the portable and mobile radio channels, multipath is caused by reflections from walls, ceiling, floor, furniture, and objects moving in the vicinity of the transmitter and receiver. As a result, the channel impulse response is a function of the location of the transmitter and receiver, and the delay associated with the arriving paths, which we denote by writing the impulse response as $h(\tau, x)$. Because the portable and mobile radio channels are represented by a set of channel impulse responses $h(\tau, x)$ measured in different locations in an area, we may define an equivalent delay power spectrum as the average of the channel impulse responses over all locations in an area. Then, similar to the troposcatter channel, if we assume that the channel is WSSUS over all locations, we have

$$E_x\{h(\tau, x)h^*(\tau_1, x)\} = C(\tau)\delta(\tau - \tau_1)$$

where $E_x\{\cdot\}$ denotes the average over the ensemble of all channel profiles observed in an area and $C(\tau)$ is the equivalent delay power spectrum of the channel.

For a set of measured profiles $h(\tau, x)$ from different locations in an environment, it is possible to determine the equivalent delay power spectrum, $C(\tau)$, experimentally. Starting from the arrival of the first path, the time axis is divided into bins of width equal to the transmitted pulse duration. The average received power over all profiles is calculated for each bin and plotted against the arrival time of the paths. Figure 9.16 shows the equivalent delay power spectrum measured from the indoor measurement database used in this section, and it also shows an exponential curve obtained as the best fit to the empirical data. The decay rate of the exponential curve is 15 ns. The exponential fit to the equivalent delay power spectrum is used for performance predictions.

Calculations of the BER based on the equivalent delay power spectrum are done for an infinite number of feedback taps, which provides a lower bound to the calculations based on real data or on the computer-generated channel impulse responses. The theoretical results presented for the BPSK/DFE modem are based on the analysis provided in [Mon77]. This work assumes an infinite number of feedback taps, which results in complete removal of the ISI due to past decisions. In this work, the probability density function of the SNR per bit after equalization, γ_b , is given by

$$f_\Gamma(\gamma_b) = \sum_{i=1}^D \sum_{k=0}^k \frac{A_{ik} \gamma_b^{k-1}}{(\Gamma_j)^k (k-1)!} \exp\left(\frac{-\gamma_b}{\Gamma_j}\right) \quad (9.5.1)$$

where $\Gamma_j = \lambda_j \overline{\gamma_b}$ ($\overline{\gamma_b}$ is the average SNR per bit) and λ_j represents the eigenvalues of the matrix $\mathbf{G}^{-1}\mathbf{C}(t_0)$. The elements of the matrix $\mathbf{C}(t_0)$ are given by

$$C_{kl}(t_0) = \int_{-\infty}^{\infty} g(t_0 - k\tau_s - u)g(t_0 - l\tau_s - u)Q(u) du$$

where τ_s is the equalizer tap spacing, $g(\tau)$ is the overall impulse response of the pulse-shaping filters (raised cosine, 50% roll-off), and $Q(u)$ is the delay power spectrum of the channel. The elements of the matrix $\mathbf{G}$ are given by

$$G_{kl} = g(k\tau_s - l\tau_s) + \overline{\gamma_b} S^2 \sum_{i=1}^I C_{kl}(t_0 - iT)$$

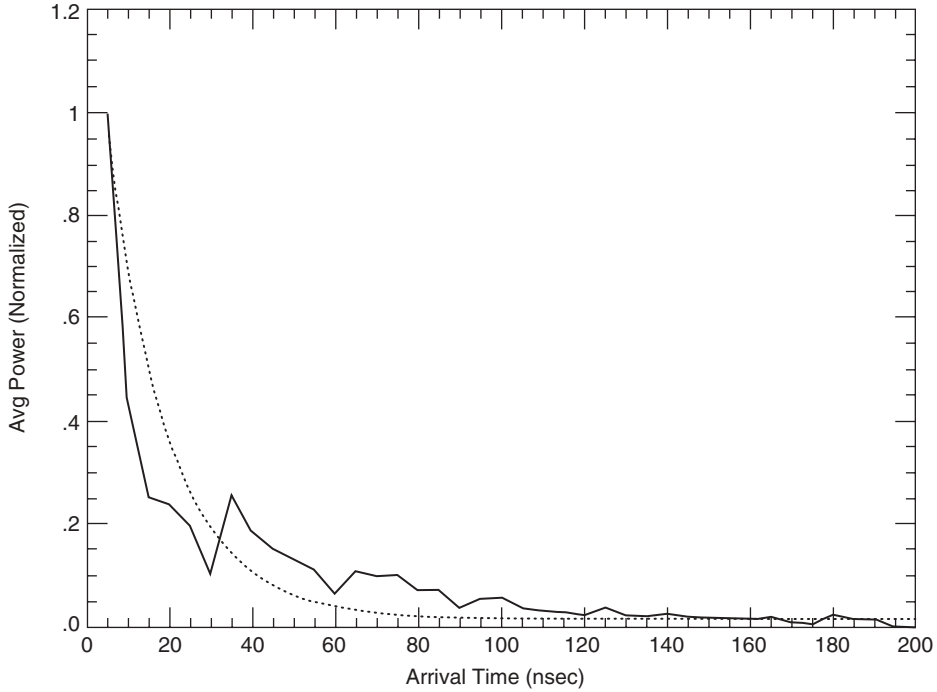


FIGURE 9.16 Equivalent delay power spectrum obtained from the database used in this chapter, along with an exponential fit to the empirical data.

where S^2 is the variance of the ISI, I is the number of future symbols with significant ISI, t_0 is the sampling instant, and T is the symbol period.

In our calculations these equations are used with the equivalent delay power spectrum $C(\tau)$ instead of the delay power spectrum $Q(\tau)$. The shape of the equivalent delay power spectrum is exponential with the decay rate of 15 ns, as shown in Fig. 9.16, and the equalizer uses three $T/2$ -spaced forward taps. Then the probability of outage is determined from [Sex89b] to be

$$\begin{aligned}
 P_{\text{out}} &= \int_0^{\gamma_{\text{out}}} f_{\Gamma}(\gamma) d\gamma \\
 &= 1 - \sum_{i=1}^D \sum_{k=0}^N \frac{A_{ik}}{\Gamma_j^k} \exp\left(-\frac{\gamma_{\text{out}}}{\Gamma_j} \sum_{i=0}^{k-1} \frac{\gamma_{\text{out}}^i}{i!} \Gamma_j^{k-i}\right)
 \end{aligned} \quad (9.5.2)$$

where γ_{out} is the SNR that would produce the threshold error rate if the channel were nonfading. Figure 9.17 shows the probability of outage versus the data rate using an equivalent delay power spectrum with one and two orders of diversity. It can be seen from the figure that the system with two orders of diversity performs more than three orders of magnitude better than the system with single diversity.

Figure 9.18 shows the probability of outage versus data rate at SNR = 40 dB and BER threshold 10^{-4} for the theoretical bound using the equivalent delay power spectrum, together with results from computer-generated channel profiles and the results

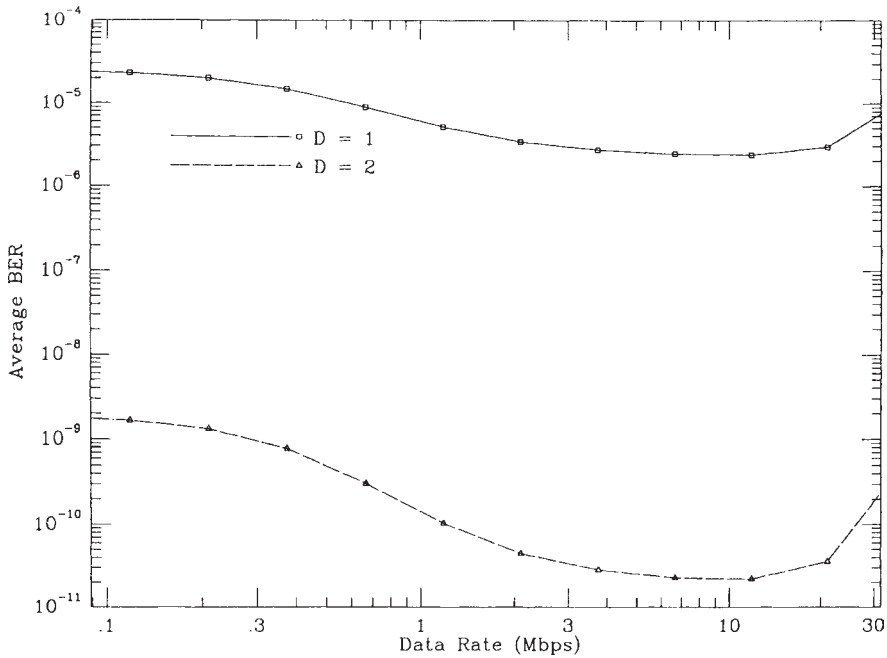


FIGURE 9.17 Probability of outage versus data rate for a BPSK/DFE modem with one and two orders of diversity for SNR equal to 40 dB. Theoretical results obtained using an equivalent delay power spectrum.

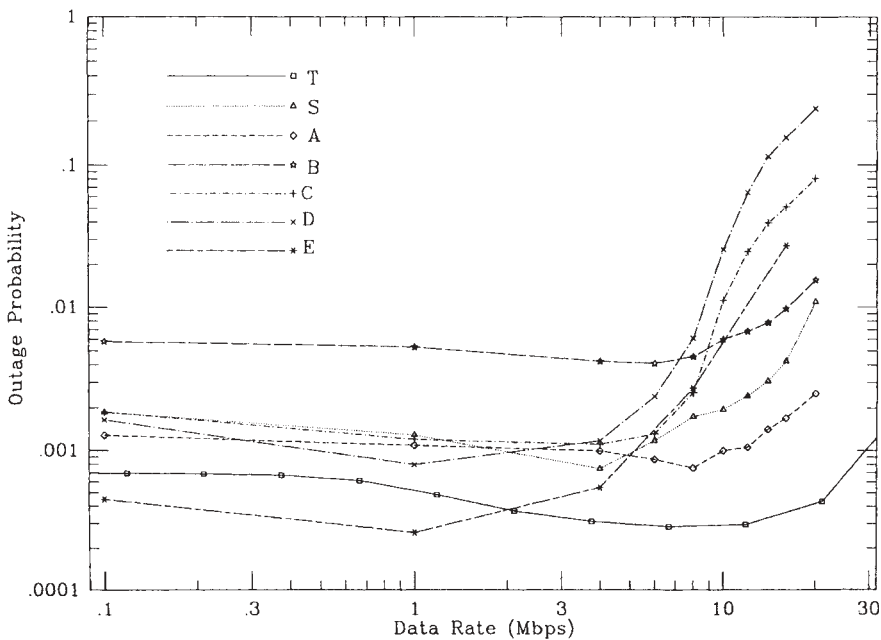


FIGURE 9.18 Probability of outage versus data rate for a BPSK/DFE modem in each of five areas for the computer-simulated channel model (S) and theoretical bound results (T). The SNR is 40 dB and the outage threshold is 10^{-4} . The theoretical bound uses the equivalent delay power spectrum.

for the measured channel profiles in the same five areas, A to E, described earlier. The simplified calculation using the equivalent delay power spectrum and an infinite number of feedback taps provides a close lower-bound approximation to the actual probability of outage with three forward and three feedback taps in four of the five areas. The results of performance analysis also show close agreement with the results obtained from the computer-simulated channel. At the lower data rates, the worst performance is observed in area B, containing a storage area partitioned with metallic screens. The best low-data-rate performance is observed for area E, which is a large open area used for final inspection of automobiles coming off the assembly line. At high data rates, area D, the “jungle,” yields the worst performance, whereas area A, with its considerable open space, shows the highest data rates.

The average probability of error for a BPSK/DFE modem [Mon77] is given by

$$\begin{aligned} \overline{P_e} &= \frac{1}{2} \int_0^\infty \operatorname{erfc}(\sqrt{\gamma_b}) f_\Gamma(\gamma_b) d\gamma_b \\ &= \sum_{i=1}^D \sum_{k=0}^N A_{ik} \left(\frac{1-u_j}{2} \right)^k \sum_{i=0}^{k-1} \binom{k-1+i}{i} \left(\frac{1+u_j}{2} \right)^k \end{aligned} \quad (9.5.3)$$

where

$$u_j = \left(\frac{\Gamma_j}{1 + \Gamma_j} \right)^{1/2}$$

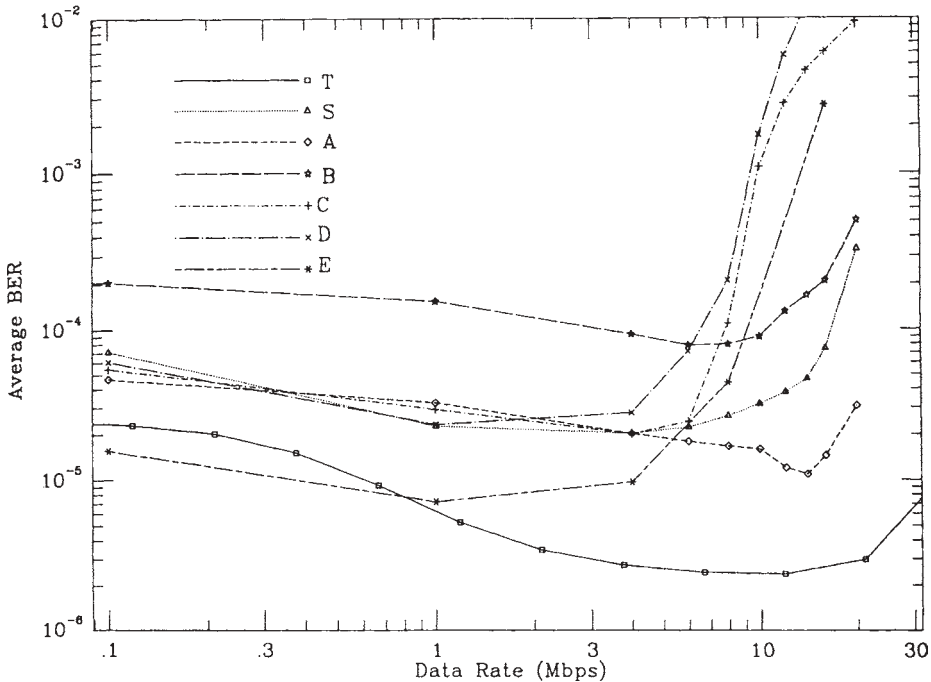


FIGURE 9.19 Average probability of error versus data rate for a BPSK/DFE modem in each of the five areas for the computer-simulated channel model (*S*) and for theoretical bound results (*T*). The SNR is 40 dB and the theoretical bound uses the equivalent delay power spectrum.

Figure 9.19 shows the average BER versus data rate at $\text{SNR} = 40 \text{ dB}$ for the theoretical bound using the equivalent delay power spectrum, together with results from computer-generated channel profiles and the results found for the measured channel profiles in the same five areas. For low data rates in area E, which is the most open area, the performance is better than the values predicted using the lower bound.

9.6 SECTORED ANTENNAS

One means of increasing the data rate is to use some form of antenna diversity, and three approaches to attain antenna diversity are depicted in Fig. 9.20. Spatial diversity can be implemented using multiple separate antennas of the same design, or polarization diversity can be used, which has been employed in some WLAN products. A third

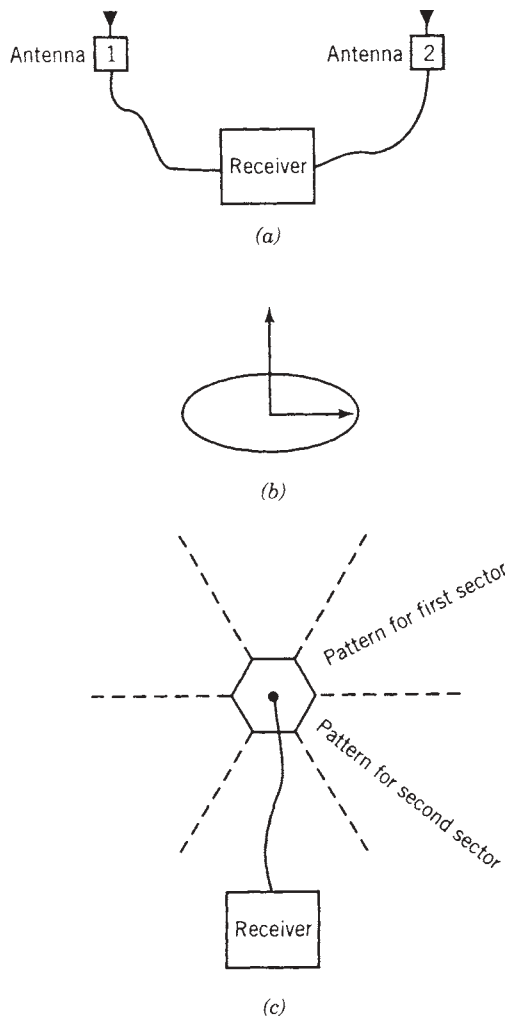


FIGURE 9.20 Three approaches to providing antenna diversity: (a) physical antenna diversity; (b) polarization diversity; (c) sectorized antenna diversity.

approach is to use a sectored antenna, as was done in the Altair WLAN product. The spatial diversity provided by multiple antennas, and the polarization diversity provided by orthogonally polarized antennas, will each reduce the power margin required to overcome the effects of power fluctuations caused by fading. The diversity analysis presented earlier in this chapter applies directly to the multiple-antenna and polarization diversity techniques. Therefore, in this section we give primary attention to the sectored-antenna technique.

A sectored antenna observes the signal arriving from different directions (paths) and selects the sector with maximum power. The use of sectored antennas has two advantages relative to the first two techniques: (1) A sectored antenna reduces the multipath in each sector, increasing the maximum data rate achievable on the channel, and (2) effective diversity can be provided without requiring wide physical separation between antennas, making compact product packaging feasible. As an example, the Altair WLAN product uses two six-sector antennas, one each at the transmitter and receiver, which provides a total of 36 effective sectors. With such a design, the array of all multiply reflected signal paths arriving at the receiver is divided into 36 subsets, and with appropriate signal processing (e.g., selection diversity), the receiver can extract the subset producing the least signal degradation, discarding all the other signal arrivals.

Performance analysis given in Section 9.5 for decision feedback equalization (DFE) with omnidirectional antennas over measured and modeled indoor radio channels showed that data rates on the order of 10 Mb/s are attainable with BPSK/DFE modems operating in indoor areas. It has been demonstrated that sectored antennas can provide similar enhancements for high-speed wireless data communications in indoor areas [Fre91a,b]. From one perspective, sectored antennas and DFE techniques represent two different approaches to exploiting diversity. A DFE modem uses the signals arriving from different paths to provide implicit or in-band diversity, which is, in effect, combined with an adaptively optimized FIR digital filter. In contrast, sectored antennas serve to create signal components of reduced multipath dispersion, from which the receiving system extracts the component of best quality.

In Section 9.5, measured channel profiles and statistical models based on measured channel profiles were used for performance prediction of the DFE modems. These measurements and modeling techniques cannot be applied to the analysis of the modems using sectored antennas, because all available wideband channel measurements and related statistical models have been developed from measurements made with omnidirectional antennas. These measurements identify all the arriving paths in the time domain but do not identify the directions from which the individual paths arrive. To evaluate the performance of a modem with a sectored antenna, the ray-tracing method, which provides the direction of the arriving paths, should be used to model the propagation characteristics. In the remainder of this section, following Yang et al. [Yan94c], we use the ray-tracing algorithm for the performance analysis of BPSK and BPSK/DFE modems operating with a sectored receiver antenna in a typical indoor area used for WLANs.

In analyzing the performance of sectored antennas, we assume that the receiver is equipped with a six-sector directional antenna whose polarizations are vertical. The i th idealized antenna pattern is defined by the function

$$g_i(\phi_k) = \begin{cases} f\left(\phi_k - \frac{\pi i}{3}\right), & \frac{\pi i}{3} - \frac{\theta}{2} \leq \phi_k < \frac{\pi i}{3} + \frac{\theta}{2} \\ 0, & \text{otherwise} \end{cases}$$

where $g_i(\phi_k)$ is the normalized power gain, ϕ_k is the antenna orientation angle, and θ is the 3-dB beamwidth of the antenna, which is around $\pi/3$ for a six-sector antenna. The channel impulse response used for each sector of the antenna includes all paths arriving in that direction with appropriate attenuation given by the antenna pattern. The overall impulse response for an omnidirectional antenna includes all paths. The technique used in Section 9.5 will be used for performance evaluation of systems with omnidirectional and sectored antennas, using the appropriate channel impulse response in each case. For the sectored antenna the impulse response associated with the sector receiving maximum power is used in the calculations. In other words, ideal selection combining is assumed to be used to process the diversity branches provided by the sectors of the antenna. Other diversity combining techniques or selection criteria may result in better performance in some cases, but selection diversity is considered the most attractive technique, due to its relative simplicity of implementation.

To compare the performance of omnidirectional and sectored antennas, the received unfaded SNR 1 m from the transmitter is assumed to be the same for both antennas. As an example of calculation of the unfaded SNR, assuming that the received bandwidth B is 10 MHz, the noise temperature T is 290 K (17°C), the transmitter power is 100 mW, and the received front-end noise figure is 9 dB. Then the thermal noise, kTB , is -104 dBm [Lee89] at the receiving antenna, and thus the net noise level is -95 dBm. If we assume that the attenuation 1 m from the transmitter is 35 dB, the unfaded SNR will be 80 dB for both omnidirectional and sectored antennas.

9.6.1 Performance Prediction in an LOS Environment

Office environments often include large, open inner areas. A simple square-room LOS environment will now be used to compare the performance of the four types of modems being examined. The receiver is assumed to be located at the center of the room, and the transmitter is located at many different locations throughout the room.

Figure 9.21 illustrates outage probabilities versus data rate for systems using different modems in a $30\text{-} \times 30\text{-m}$ room. The transmitter power is assumed to be 100 mW. This figure illustrates that if a system with an omnidirectional antenna is used in a $30\text{-} \times 30\text{-m}$ room, the data rate can reach only 3 Mb/s, with an outage rate of 0.01, for an acceptable error rate of 10^{-5} . If a six-sector antenna system is used, and the selection criterion is to choose the sector with the highest received power, a data rate of 20 Mb/s can be achieved, whereas a system with a DFE can only reach a data rate of less than 15 Mb/s. The results in the figure indicate that a DFE modem with three forward taps and three feedback taps can operate successfully at a data rate of around 10 Mb/s but shows very high outage probabilities at data rates beyond about 15 Mb/s. However, it can be seen that a DFE modem operating with a sectored antenna can operate successfully at data rates up to about 40 Mb/s.

Figure 9.22 shows outage probabilities versus unfaded SNR at 1 m from the transmitter for a BPSK modem operating at three different data rates with sectored antennas in a $30\text{-} \times 30\text{-m}$ room. It can be seen that the probabilities of outage show little sensitivity with respect to unfaded SNR when the transmitter power level is in the region of high unfaded SNR.

Figure 9.23 shows the probabilities of outage versus room size (square room with one wall length given) for different modems with a transmitter power of 100 mW at a data rate of 20 Mb/s. For a room smaller than 20×20 m, it is difficult to conclude from these results whether the sectored antenna or the DFE delivers better performance.

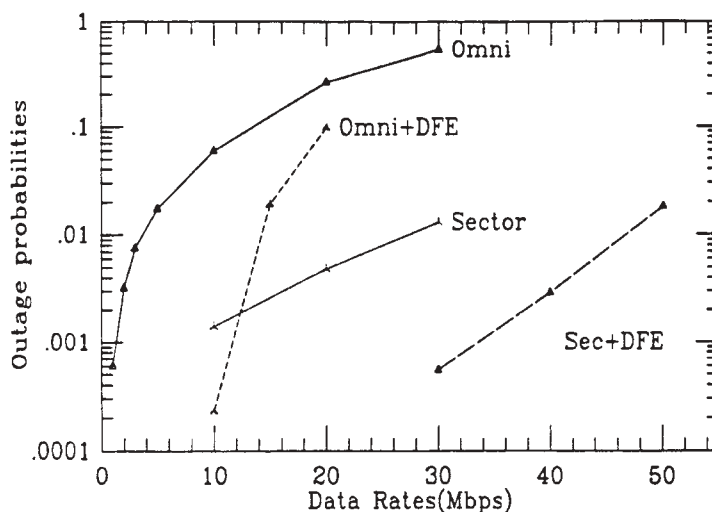


FIGURE 9.21 Outage probabilities versus data rate for systems with four different modems operating in a 30- × 30-m room.

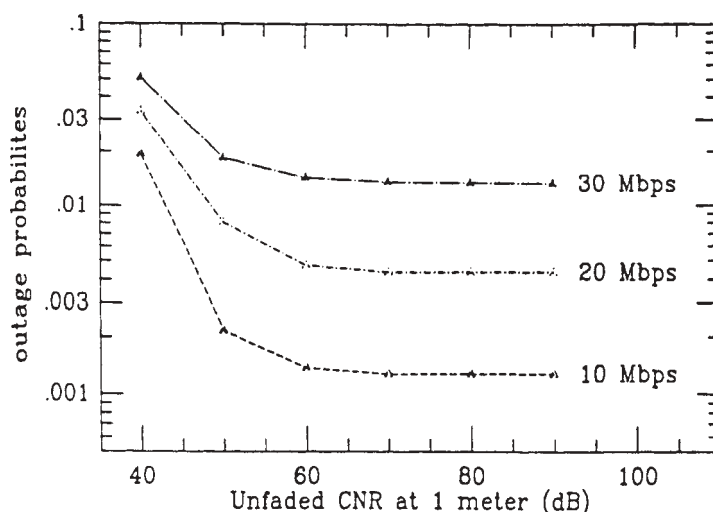


FIGURE 9.22 Outage probabilities versus unfaded CNR (or SNR) at 1 m from the transmitter given different data rates for a BPSK modem with sectored antennas in a 30- × 30-m room.

However, for room sizes larger than 30 × 30 m, significantly better performance will be achieved with a sectored antenna.

Figure 9.24 shows outage probabilities versus data rate for a BPSK modem operating at two different power levels with three different sectored antennas (see Fig. 9.25): (I) a nonoverlapped antenna span pattern, (II) an optimum antenna span pattern, and (III) the mathematically ideal pattern. The results in Fig. 9.24 show that the sectored antenna patterns provide significant benefits for modem performance and that a 10-dB change in the transmitted power causes little change in the outage probability. These

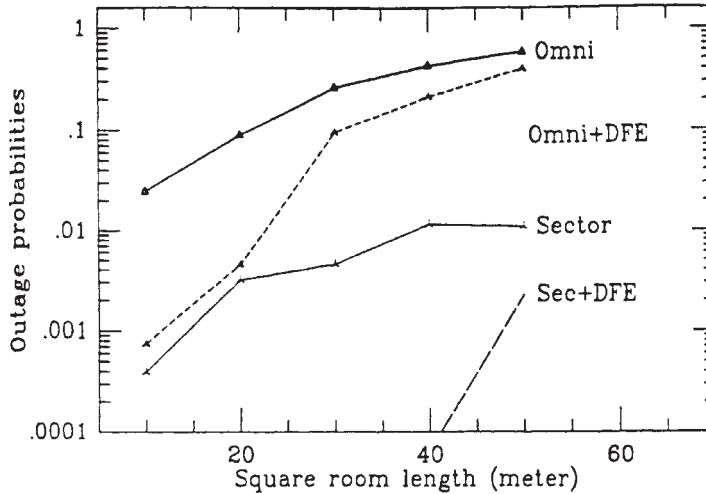


FIGURE 9.23 Outage probabilities versus room size for various modems with a transmitted power of 100 mW at a data rate of 20 Mb/s.

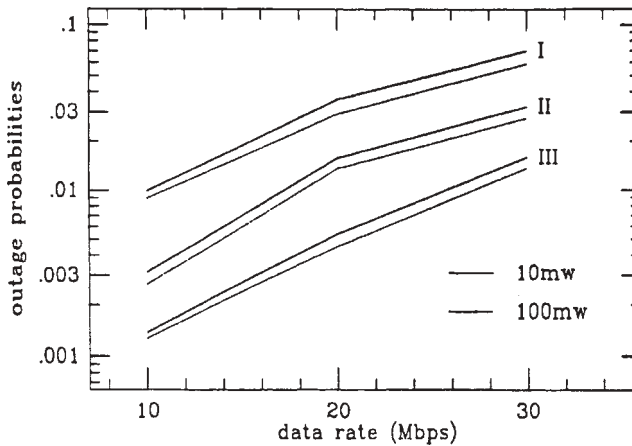


FIGURE 9.24 Outage probabilities versus data rate for a BPSK modem with different sectored antennas: (I) nonoverlapped span pattern; (II) optimum span pattern; (III) ideal pattern (see Fig. 9.30).

results also provide information on the performance gains achievable with different antenna patterns.

9.6.2 Performance Prediction in an OLOS Environment

The simulation results discussed above show that using a sectored antenna is an effective technique for counteracting the harmful effects of multipath. In this section we examine the sectored antenna and the DFE in an OLOS environment. The floor plan that was analyzed in Chapter 6 for two-dimensional ray tracing is used again here in Fig. 9.26, with the rooms numbered. The receiver is located at the center of room

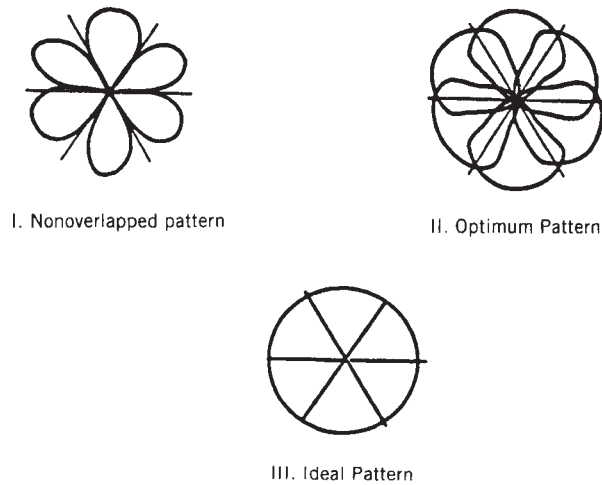


FIGURE 9.25 Three sectored antenna patterns: (I) nonoverlapped antenna-span pattern; (II) optimum antenna-span pattern; (III) ideal pattern.



FIGURE 9.26 Floor plan of one section of the second floor of the Atwater Kent Laboratories. This is the site of the indoor propagation experiments described earlier.

1 and the transmitter is moved to various locations in the floor plan. To obtain the outage probabilities, the number of simulations run for each room is about 5000. The maximum data rates are obtained for four different types of modems in each room, and an average maximum data rate is obtained for the overall floor plan.

TABLE 9.2 Maximum Data Rates for BPSK and BPSK/DFE Modems with and without Adaptive Antenna Arrays in Various Rooms

Room	Maximum Data Rate (Mb/s)			
	Omni	Omni + DFE	Sector	Sector + DFE
1	8	40	50	50
2	5	20	12	25
3	4	15	13	20
4	2	7	9	13
5	4	20	20	40
6	7	25	30	50
7	2	6	8	12
8	2	7	8	15
Overall	3	12	10	18

From Table 9.2 it can be seen that the average maximum data rate over the entire floor plan is 12 Mb/s for a BPSK/DFE modem with an omnidirectional antenna and 10 Mb/s for a BPSK modem with a sectored antenna. The DFEs with three forward taps and three feedback taps perform slightly better than the nonequalized modems with six-sector antennas. In a small LOS environment such as Room 1, high data rates can be achieved using either technique. If the receiver moves to one of the adjoining rooms, such as Room 2, 3, 5 or 6, the BPSK/DFE modem with omnidirectional antennas can still attain data rates above 15 Mb/s. However, the maximum achievable data rate for the BPSK modem with the sectored antennas in Room 3 drops to 12 Mb/s. In Rooms 4, 7, and 8, only a BPSK/DFE modem with sectored antennas can achieve a data rate above 10 Mb/s, and sectored antennas provide slightly better performance than DFE.

The results of Table 9.2 might seem to be overly optimistic but are, in fact, quite reasonable for the case examined here. The entire floor plan is confined to a relatively small area in which the maximum radius is about 10 m, the rms delay spread is less than 30 ns, and the power received is not exceptionally low in any location. The results shown for transmitter and receiver both in Room 1 show that any modem can achieve high data rates. In reality, of course the achievable data rates will be somewhat lower than are indicated here, due to the effects of phase and timing jitter.

The poorest performance is achieved in room 7. Figure 9.27 shows that if the data rate is below 12 Mb/s, the performance of the BPSK modem with sectored antennas is superior to that of a BPSK/DFE modem with an omnidirectional antenna. However, the sectored antenna seems to be less effective for data rates higher than 15 Mb/s in this worst case. The BPSK/DFE modem with sectored antennas offers relatively little advantage over the BPSK/DFE modem with omnidirectional antennas at very high data rates.

9.7 MULTICARRIER, OFDM, AND FREQUENCY DIVERSITY

Another approach to dealing with the problem of multipath distortion is one that has been used in radio modems for more than 30 years, known by several names, including *multicarrier modulation* (MCM) or *multitone modulation* [Bin90]. The concept here

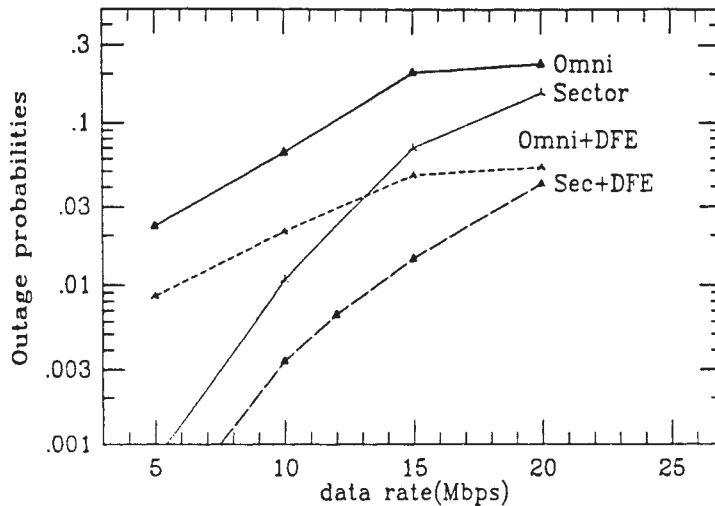


FIGURE 9.27 Outage probability versus data rate for four types of modems operating in Room 7.

is very simple: Instead of modulating a single carrier at rate R symbols/s, we use N carriers spaced by about R/N hertz and modulate each of the carriers at the rate R/N symbols/s. The advantage of this scheme is that on a multipath channel having a given value of multipath spread, say τ_m seconds, the multipath is in effect reduced relative to a symbol interval by the ratio $1/N$, and thus imposes less distortion in each demodulated symbol. If the symbols are made sufficiently long relative to the multipath spread, reliable demodulation performance can be achieved without the need for adaptive equalization. Therefore, on multipath (frequency-selective) fading channels, multicarrier modulation provides a simpler alternative to single-carrier modulation with adaptive equalization. An additional advantage of multicarrier modulation is that in frequency-selective fading, the subchannels provide a form of frequency diversity, which can be exploited by applying error-control coding across symbols in different subchannels.

Figure 9.28 represents a MCM system with the bandwidth of W that has N carriers and channel frequency response with frequency-selective fading. As shown in figure, carriers operating at different frequencies are exposed to different channel gains. Therefore, the received signals in individual carriers have different signal-to-noise ratios and bit-error-rate quality. If some redundancy in the carriers is imposed on the system, the errors caused by low SNRs in poor channels can be recovered. In addition, although the entire bandwidth used by the system is exposed to frequency-selective fading, individual channels are exposed only to flat fading, which does not cause ISI that needs computationally extensive receivers such as DFE or MLSE.

Example 9.3: Legacy Multicarrier in Military Applications Common examples of legacy multicarrier systems include several HF modems widely used by the U.S. military that operate with phase or frequency modulation on *parallel tones*, each tone or carrier typically modulated at 75 baud. As an example, let us postulate a 16-tone modem using QPSK modulation at 75 baud on each tone, for an overall data rate of 2400 bits/s. Consider a typical HF channel with multipath spread of $\tau_m = 2$ ms.

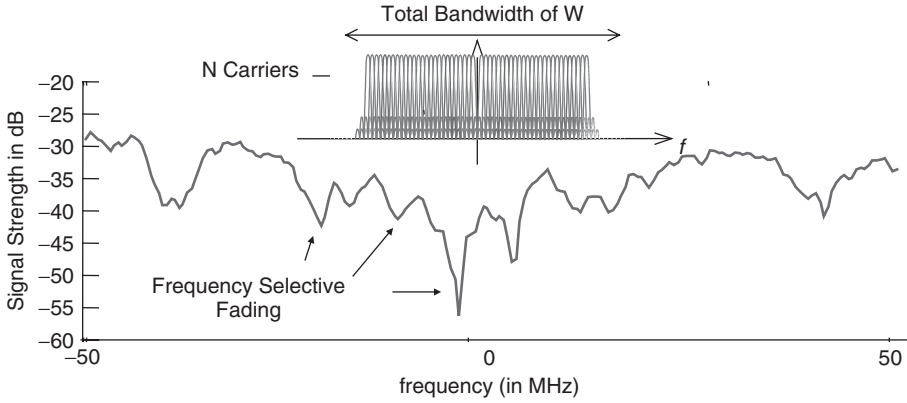


FIGURE 9.28 Frequency-selective fading and MCM.

If a single carrier were modulated at 1200 ($= 16 \times 75$) baud, the symbol duration would be 0.42 ms, for which the 2-ms multipath spread would cause enough intersymbol interference that adaptive equalization would be required for acceptable BER performance. However, with 16 channels modulated at 75 baud, the symbol interval in each channel is 13.3 ms and the multipath distortion can essentially be eliminated by providing “time guards” between successive symbols in each channel. (Frequency guard bands may also be required to avoid interference between adjacent subchannels.) Using exactly this approach, parallel-tone modems have been used very effectively on long-haul HF circuits for a number of years. Several papers describing the theoretical and measured performance of parallel-tone HF modems are included in [Bra75].

In the early 1990s, the multicarrier transmission approach was investigated for application to microwave radio [Cho90], land-mobile [Sas93], and indoor wireless [Har93, Yee93, Yan94a] communications. A brief discussion of some of the results presented in [Har93] will serve to illustrate the principles of designing a multicarrier system for a fading channel. In that paper the authors derive theoretical BER performance for a multicarrier system using DPSK modulation and differential detection on an indoor frequency-selective Rayleigh channel. Figure 9.29 shows block diagrams of the N -carrier transmitter and receiver that were analyzed. The modulator and demodulator are implemented with an inverse discrete Fourier transformer (IDFT) and discrete Fourier transformer (DFT), respectively, instead of with N oscillators. The DFT technique for multicarrier transmission allows minimum frequency separation between carriers and efficient demodulation with computationally efficient DFT algorithms [Wei71]. The transmitted signal $s(t)$ can be written as

$$s(t) = \sum_{i=-\infty}^{\infty} \sum_{k=0}^{N-1} \text{Re}[c_{ki} e^{j2\pi f_k(1-iT)}] f(t - iT) \quad (9.7.1)$$

In this equation, f_k is the frequency of the k th carrier,

$$f_k = f_0 + \frac{k}{t_s} \quad (9.7.2)$$

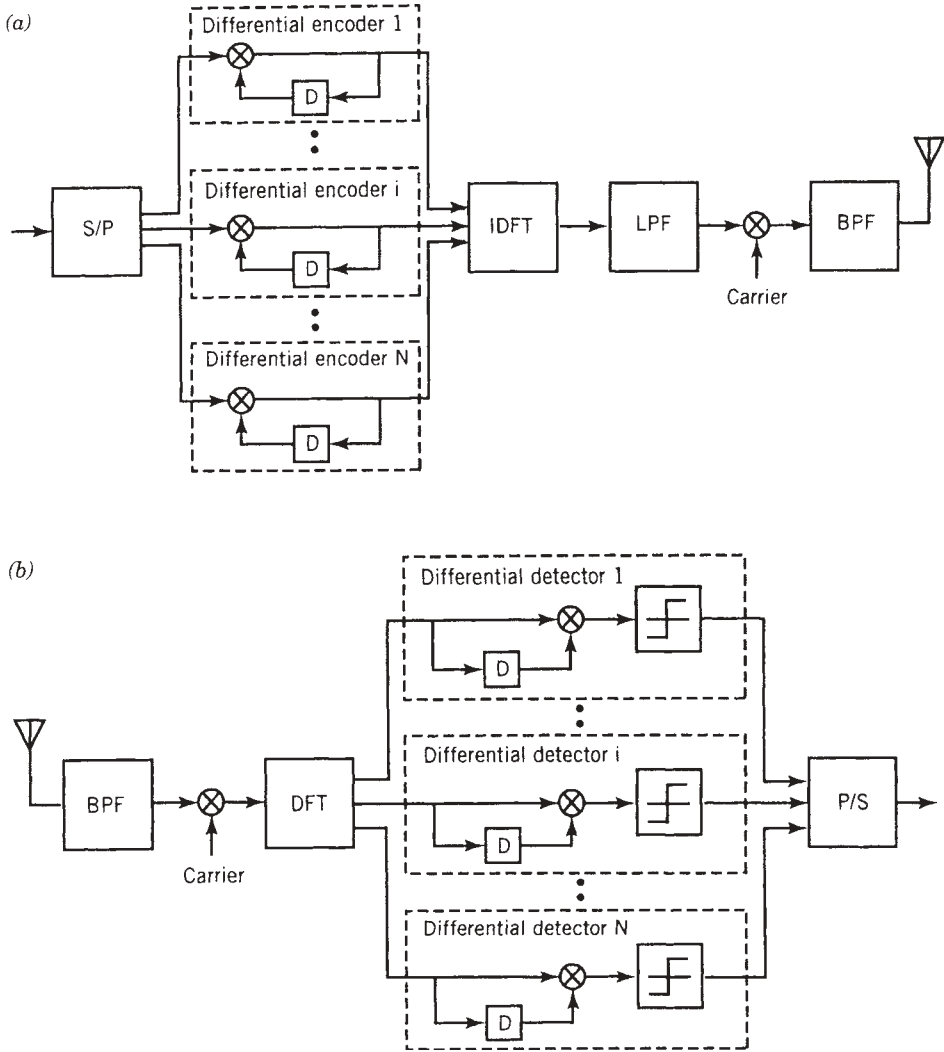


FIGURE 9.29 Multicarrier transmission system: (a) transmitter; (b) receiver. S/P, serial/parallel converter; D, unit delay; IDFT, inverse Fourier transform; LPF, low-pass filter; BPF, bandpass filter; DFT, Fourier transform; P/S, parallel/serial converter; $\square$, decision blocks. (From [Har93] © IEEE.)

where f_0 is the lowest carrier frequency. The function $f(t)$ is the pulse waveform for each transmitted symbol, defined as

$$f(t) = \begin{cases} 1, & -\Delta \leq t \leq t_S \\ 0, & t < -\Delta, t > t_S \end{cases} \quad (9.7.3)$$

where Δ is the time-guard interval, t_S is the observation interval, $T = \Delta + t_S$ is the total symbol duration, and c_{ki} is the output of the k th differential encoder in the time interval $(iT - \Delta, iT + t_S)$. In [Har93], M -ary DPSK modulation is used on each of the subcarriers. Therefore, the transmitted signal $x(t)$ is the sum of NM -ary DPSK

signals with symbol duration T and with carriers separated by $1/t_s$ hertz. The $1/t_s$ -hertz separation of carriers ensures orthogonality (when the receiver is correctly time synchronized) of symbols demodulated on different carriers.

The guard-time implementation described above is referred to as a *receiver-gated time guard*, in which length T symbols are transmitted in continuous succession but the receiver is time-gated “on” only during the observation interval t_s . An alternative implementation is a *transmitter-gated time guard*, in which the transmitter is actually turned off for the interval Δ in each symbol interval T in order to keep successive symbols separated in the multipath channel. These two time-guard approaches result in somewhat different synchronization characteristics, an issue that is treated in detail in [Bel65].

In the current literature the FFT implementation for MCM above is referred to as *orthogonal frequency-division multiplexing* (OFDM) because the symbols transmitted in different carriers are orthogonal to one another. If the carriers are coded to provide redundancy, the modulation technique is referred to as *coded OFDM* (COFDM). At the time of this writing, OFDM is the most popular transmission technique for broadband wireless and wireline communications and is considered for standard transmission of digital audio and video broadcasting, high-definition digital television, digital subscriber line, broadband cable, and broadband power-line modems. In the terrestrial wireless networks considered in this book, OFDM has been adopted by the IEEE 802.11a/HIPERLAN2 and IEEE 802.11g standardization committees and is being considered for next-generation digital cellular networks. For more details descriptions of OFDM and its applications, the reader may refer to [Bah99, Roh99, Pan02].

Example 9.4: 802.11 OFDM The OFDM system recommended by the IEEE 802.11a and g and HIPERLAN2 standards, shown in Fig. 9.30a, has 64 carriers. These carriers are divided into 48 data carriers, four pilot carriers, and 12 virtual carriers. Total

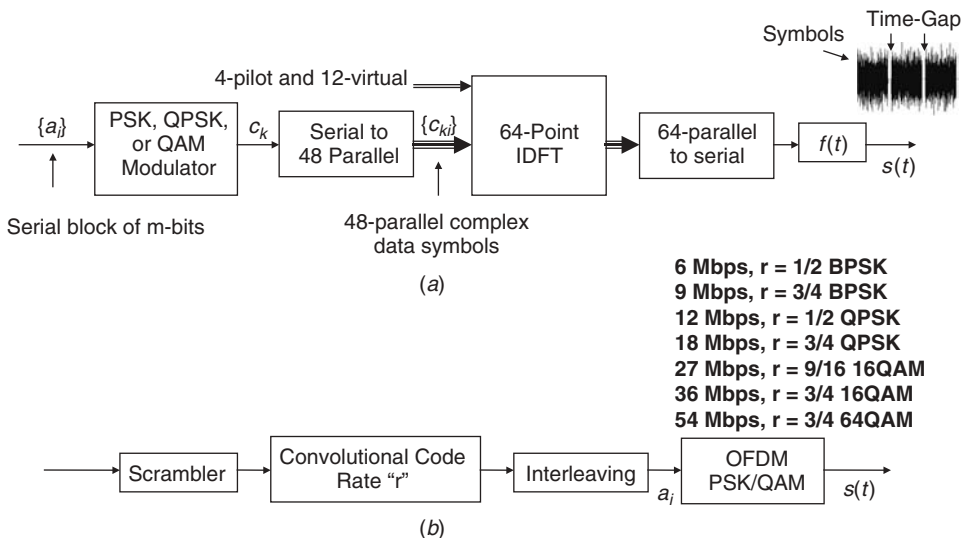


FIGURE 9.30 IEEE 802.11a and g and HIPERLAN2 baseband transmitter: (a) OFDM implementation; (b) overall modulation coding.

duration of the symbols is $T = 4000$ ns and the guard time between the symbols is $\Delta = 800$ ns, resulting in an observation interval of $t_s = 3200$ ns. Therefore, the carrier separation is $1/t_s = 31.25$ kHz and the total bandwidth used by the system is $W = N \times 1/t_s = 64 \times 31.25$ kHz = 20 MHz. Since every T seconds one symbol is transmitted, the effective channel coded data symbol transmission rate is $1/T = 250$ kilo symbol per second (kS/s) and the effective coded data transmission rate of the entire system is $48(\text{data} - \text{carriers}) \times 250$ kS/s = 12 MS/s. Readers should notice that the choice of the guard time is related to the excess multipath delay spread of the channel. WLANs are designed for indoor applications with a coverage of less than 100 m. As we observed from the results of wideband measurements in Chapter 5, the excess multipath delay spread in indoor areas is always less than several hundred nanoseconds, which is well accommodated by the 800 ns chosen by the standardization committee.

Continuing with our discussion of the receiver-gated multicarrier implementation, as treated in [Har93], we discuss briefly two system issues: (1) BER performance in multipath fading and (2) the overall bandwidth utilization efficiency achieved by the system design. System performance is directly related to the choice of time-guard interval Δ in relationship to the multipath characteristics of the channel. As a specific example, [Har93] considered a system using $N = 32$ carriers operating on an indoor wireless channel with the symbol duration on each carrier chosen as $T = 1/128 \times 10^{-3}$ s or 7.8125 μ s. With QDPSK modulation, this design yields a data rate of 256 kb/s per carrier, for an overall data rate of 8.192 Mb/s. The indoor channel was characterized by a nine-ray multipath model, with rays fluctuating independently with a Rayleigh distribution, and overall rms delay spreads of $\tau_{\text{rms}} = 50$ and 100 ns, typical values for such a channel. Figure 9.31 shows the theoretical BER performance versus normalized guard time (Δ/T) for the two assumed values of τ_{rms} and three values of $\bar{\gamma}_b$. These results were calculated for a case of frequency-selective,

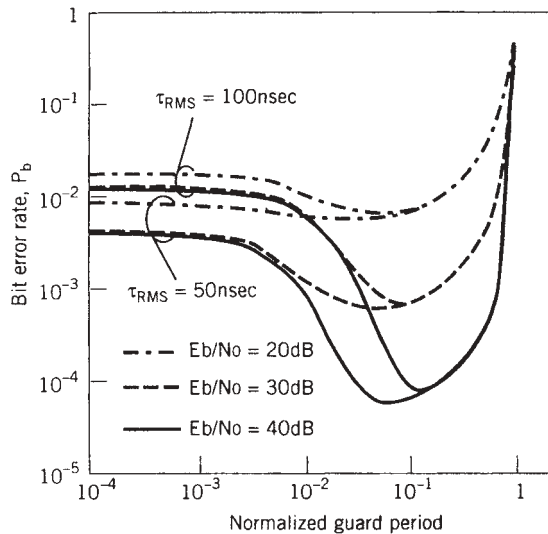


FIGURE 9.31 BER versus normalized guard time Δ/T for a multicarrier system operating on an indoor wireless channel. The number of carriers is 32, the modulation is QPSK, and the overall data rate is 9.92 Mb/s. (From [Har93] © IEEE.)

time-nonsselective fading, which means that the multipath rays are fixed in time. Note that for each combination of rms multipath spread and SNR there is an optimum value of normalized guard time for which BER is minimized. With shorter guard times, the BER increases due to the multipath distortion at the edges of the received symbols. With longer guard times, the BER increases due to the nonutilization of signal energy transmitted during the guard time. It can be seen from Fig. 9.31 that the optimum values of normalized guard time are in the range 0.03 to 0.1, or about 250 to 750 ns for the system considered here. As one would expect, smaller values of τ_{rms} result in shorter optimum guard times and lower levels of achievable BER.

Figure 9.32 shows BER performance versus normalized delay spread (τ_{rms}/T) for three different modulation techniques used with the same 32-carrier system and for a single-carrier QDPSK system operating at 8.192 Mb/s without equalization. The normalized guard time for each of the multicarrier cases is $\Delta/T = 3.03 \times 10^{-2}$, and the SNR is 40 dB for all cases. The results given in Figure 9.32 clearly show the greater robustness of the multicarrier systems relative to the nonequalized single-carrier system in the indoor multipath environment.

It is possible to combine the basic multicarrier transmission method with various modulation and coding techniques to achieve various system objectives. As one example, the classical paper [Yee93] describes a technique given the name *multicarrier code-division multiple access* (MC-CDMA) and analyzes its performance on indoor wireless channels. With this technique, symbols are transmitted in each time interval on multiple subcarriers, where each subcarrier is modulated with a $\pm\pi$ offset in accordance with a pseudonoise (PN) sequence.³ Thus, the signal structure is akin to conventional CDMA designs except that in MC-CDMA the transmitted signal has a

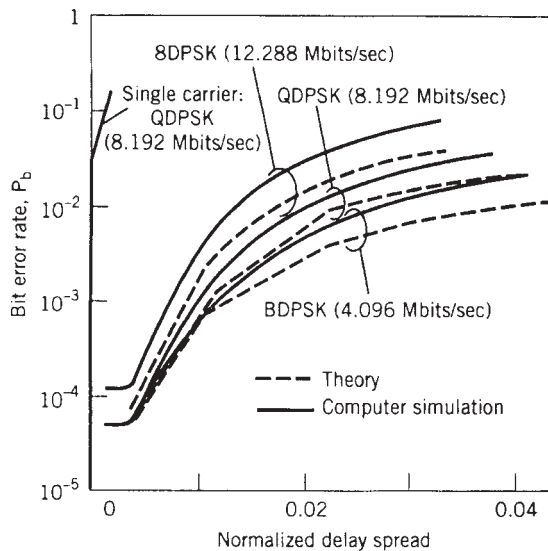


FIGURE 9.32 BER versus normalized delay spread τ_{rms}/T_s for multicarrier systems and a single-carrier system operating on a wireless indoor channel. The normalized guard time is $1/33$ for all cases. $E_b/N_0 = 40$ dB; $D/T_s = 1/33$; $T_s = 7.8/25$ ms. (From [Har93] © IEEE.)

³Details of CDMA techniques are described in Chapter 10.

PN structure in the frequency domain rather than in the time domain. Multiple access is provided for in this system by allowing different users to transmit in the same time interval with orthogonal PN codes. As with other multicarrier systems, the subchannel symbol duration is chosen to be long with respect to the multipath spread expected. The advantage for this technique over a more conventional CDMA technique using single-carrier transmission and PN coding in the time domain is that the PN codes can be detected with a relatively simple receiver structure that does not have to deal with intersymbol multipath interference. In [Yee93] the authors provide a detailed analysis of this technique, including an analysis of performance with the inclusion of diversity combining.

Yet another embellishment of the multicarrier technique is to incorporate the use of error-control coding to combat the effects of frequency-selective fading. The key concept here is to exploit the fact that in frequency-selective fading, there will be some degree of independence of the fading across the subcarriers, and by applying coding across the subcarriers, symbols on faded subcarriers can be corrected. In [Yan94a] this technique was studied for application to indoor wireless channels. Reed–Solomon (n, k) block codes were used to code across n carriers, where k of the carriers transmitted information symbols and the remaining $n - k$ carriers transmitted parity check symbols. The scheme was analyzed for use with BPSK and QPSK modulation on the subcarriers, and the results showed significant improvements in outage probabilities relative to uncoded multicarrier transmission.

9.7.1 Multirate Transmission

Yet another approach to increasing the data rate in the presence of multipath fading is to use a multirate modem, which provides one or more “fallback” modes of operation for increased reliability of communication under degraded channel conditions. For example, in an area where we have locations of good signal quality and others of poor quality, we can operate at higher rates in good locations and lower rates in poor locations. This is done in voiceband modems, for example, by using modes that operate with different numbers of points in their signal constellations. Table 9.3 shows the maximum data throughput achievable with single-rate and optimum double-rate modems operating with various orders of diversity [Win85, Zha90]. The results given in the table are based on an assumption of Rayleigh fading and a continuous multipath structure with a fixed rms multipath spread of 58 ns. We see from the table that we can

TABLE 9.3 Maximum Data Throughput (R_T) Achievable with Single- and Optimum Dual-Rate Modems Operating with Various Orders (M) of Diversity^a

M	Single-Rate R_T (Mb/s)	Dual-Rate R_T (Mb/s)	P_{out}	R_1/R_2
1	0.27	2.6	0.035	16.0
2	2.4	6.1	0.05	3.2
4	7.9	9.6	0.071	1.4
8	11.7	15.5	0.22	1.3

^aThe rms multipath spread is 58 ns for all cases.

use a single-rate modem with dual diversity and achieve 2.4 Mb/s, a 10-fold increase in data rate over nondiversity operation. However, we also see that by using a dual-rate modem with no diversity, we can achieve 2.6 Mb/s throughput. That is, we can achieve about the same data rate with either a single-rate modem and two antennas or a dual-rate modem and one antenna. Multirate modems can be incorporated into direct-sequence spread-spectrum systems where the processing gain can be adjusted in accordance with the environment. In this way we can operate with a fixed bandwidth but adjustable data rate. Another approach to adjusting the data rate with a fixed bandwidth is to use the multiampitude and multiphase modulation techniques. In this method the number of points in the constellation is increased as the channel condition improves. It is also possible to combine the two methods so as to increase the flexibility of the modem.

9.7.2 Multiampitude and Multiphase Modulation and Coding

Another technique for increasing data rate is to use multiampitude and multiphase modulation and coding. Table 9.4 shows selected parameters for modems, providing several steps of data-rate increase over QPSK modulation. Each example in the table utilizes trellis-coded modulation (TCM) as part of a combined modulation and coding design. In our preceding discussion on increasing data rates, we emphasized two objectives, one to compensate for the power loss caused by fading, the other to increase the data rate in the face of multipath constraints. One might well ask why we cannot use high-order multipoint signal constellations as is done in wireline modems. One problem is power fluctuations in the channel, which make it difficult to demodulate signal sets reliably with large numbers of points. Table 9.4 shows us that using 64-QAM modulation, which is currently a practical limit on modulation alphabet size for use in deep fading, we can achieve only a threefold increase in data rate over QPSK. However, we have seen that with the use of adaptive equalization or sectorized antennas, we can increase the data rate by a factor of 10 or more over the simplest systems. For high-speed WLANs, this threefold increase is not sufficiently attractive, especially when compared with other techniques. In the mobile communications industry, where even a factor of 2 or 3 translates into an increase in user channels by that amount, this approach is considered attractive and is being studied for application in new wireless networks.

The IEEE 802.11a and g and HIPERLAN2 standards use multirate transmission OFDM signals employing different channel coding and PSK and QAM modulation techniques to cover effective data transmission, rated from 6 to 54 Mb/s. Figure 9.30b

**TABLE 9.4 Selected Parameters for Modems
Providing Several Steps in Data-Rate Increase Over
QPSK Modulation**

Modulation	Data Rate	Coding
QPSK	R	8-TCM
8-PSK	$1.5R$	16-TCM
16-QAM	$2R$	32-TCM
64-QAM	$3R$	128-TCM

illustrates the modulation and coding schemes of the standard, which supports a variety of data rates. The incoming information bit stream is first scrambled to avoid problems caused by strings of bits with the same value. The scrambled bits are then encoded using a convolutional code with a rate that varies between $\frac{1}{2}$ and $\frac{3}{4}$. The encoded data are then passed through an interleaver to improve the performance in fading. The interleaved data bits are then grouped in 48 symbols of length 1 to 6 to cover modulation techniques ranging from BPSK (1 bit per symbol) up to 64-QAM (6 bits per symbol). With this variety of modulation and coding, one can support data rates between 6 and 54 Mb/s for the effective symbol transmission rate of 12 MS/s that was derived in Example 9.4. The following example will clarify how this scheme operates.

Example 9.5: Data Rate in IEEE 802.11 OFDM For the effective symbol transmission rate of 12 MS/s, if we use BPSK modulation with 1 bit per symbol and a convolutional code of rate $r = \frac{1}{2}$, the lowest effective data transmission rate is

$$R_{\text{BPSK}} = 12 \text{ MS/s} \times 1 \text{ bit/s} \times \frac{1}{2} = 6 \text{ Mb/s}$$

The highest effective rate is provided by 64-QAM modulation with 6 bits/s and an $r = \frac{3}{4}$ coding rate:

$$R_{\text{64-QAM}} = 12 \text{ MS/s} \times 6 \text{ bits/s} \times \frac{3}{4} = 54 \text{ Mb/s}$$

Since the symbol transmission rate and the guard time is fixed for all data rates, filters and the RF components remain the same and data rates are changed by a simple binary operation, changing the grouping and channel encoding of the data. The convolutional code has a memory of 7, which is kept fixed for all data rates. This controls the complexity of the receiver's Viterbi decoder used with the convolutional code. If rather than separate convolutional codes and multisymbol modulation, a TCM that combines the two were used, the performance would have been improved but the receiver complexity and overall design of the transmitter would have been more difficult.

The multirate system allows maintenance of a good error rate quality by reducing the data rate. When the mobile computer is close to the access point, the data rate is at its highest rate, 54 Mb/s. As the distance increases, the received signal strength decreases, and at a certain distance, the error rate of the highest data rate is no longer acceptable (let's say it goes below 10^{-5}). At this stage the data rate decreases to a second rate of 36 Mb/s, which uses 16-QAM. As shown in Section 7.2.6, the signal-strength requirement for QAM decreases 3 dB per bit; therefore going from 6-bit/s 64-QAM to 4-bit/s 16-QAM provides a 6-dB margin for signal strength. If we assume an open area with a distance–power gradient close to 2, as we showed in Section 3.2, a 6-dB edge doubles the coverage. Continuing this simplified case and reducing the modulation to BPSK and the coded to stronger, $r = \frac{1}{2}$ codes at 6 Mb/s operation, one expects to gain another 10 dB, for a total of approximately 16 dB of gain for adjustments of the modulation and coding gain. In an open area this gain may extend the coverage close to 1 decade of distance, which is considerable for WLAN operation. In lay terms, if the WLAN was designed to cover up to 100 m, with the highest data rate it need only cover up to 10 m.

9.8 COMPARISON OF TRADITIONAL BROADBAND MODEMS

In this section we provide a comparative performance evaluation of time-diversity techniques using DFE, space-diversity techniques using a sectored antenna system (SAS), and frequency-diversity techniques using OFDM technology. These results are based on [Fal96], which uses the two-dimensional ray-tracing software described in Section 6.5.4 on the floor plan shown in Fig. 6.40 to provide a comparative performance evaluation of broadband modem technologies applied to WLANs. The test area consists of seven rooms on the second floor of the Atwater Kent Laboratories at Worcester Polytechnic Institute. As described in Section 6.5.4, channel measurements in these rooms reported in [How90c] are used to calibrate a two-dimensional ray-tracing algorithm [Yan94c]. The algorithm is used to generate several hundreds of thousands of channel profiles in the test area. These profiles are used for performance evaluation of various modem design technologies operating in this environment. The QPSK modulation is used with different diversity techniques to provide a fair comparison of the effectiveness of the techniques. The performance criterion was the probability of an outage of 1% from a threshold level of 10^{-5} in the entire test area. The objective was to provide quantitative results relating bandwidth and power requirements to the maximum data rate of each of the major broadband modem design techniques described earlier in this chapter.

Figure 9.33 compares the required transmission power and bandwidth requirement of the OFDM, COFDM, DFE, SAS, and DFE/SAS techniques for a bandwidth-limited channel with a fixed bandwidth of 10 MHz. These results are useful for comparative performance evaluation of indoor broadband modem technologies using the 10-MHz unlicensed data-PCS bands at 1.9 GHz. The SAS technology was first used by Motorola in their pioneering Altair WLAN [Fre91a]. In our example we considered a single-input multiple-output (SIMO) system with an omnidirectional transmitter and six sectored receiver antennas. The DFE was adopted by the HIPERLAN1 standard [Wil95], and the one used in this performance evaluation has three forward and three feedback taps. The OFDM implementation of the MCM is the choice of 802.11a and g and HIPERLAN2. The performance here includes coded using Reed–Solomon codes of DM. The

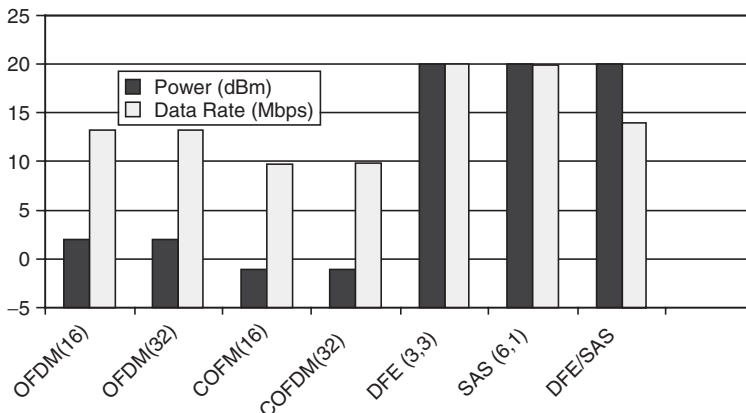


FIGURE 9.33 Performance of broadband modems in a bandlimited channel with a fixed bandwidth of 10 MHz [Fal96].

number of carriers in OFDM and COFDM are 16 and 32. To provide a guard band among the carriers, the transmitted signal is expanded 33% more than the required value. All these technologies can be combined for additional performance gains; the DFE/SAS technology is used as an example to demonstrate the combined effects of DFE and SAS. In Fig. 9.33, the light bar represents data rate in Mb/s and the dark bar shows the required power in dBm for various broadband modem technologies. The overall maximum-supportable data rate for all technologies in the figure is 20 Mb/s, which can be achieved by the DFE, SAS, or DFE/SAS systems. As shown in the figure, DFE and SAS require approximately 20 dBm (100 mW) of power to cover 99% of the seven-room test area with a maximum error rate of 10^{-5} . The DFE/SAS system, taking advantage of both time and space diversity, would need around 6 dB (four times) less transmit power at the expense of a more complex receiver. The data rate of the MCM or OFDM in the fixed bandwidth of 10 MHz is 13.3 Mb/s, while the effective data rate of COFDM, using a portion of the bits for parity check, is around 10 Mb/s. With the same performance, however, DFE and SAS need close to 20 dBm power to cover the test area. An OFDM system with either 16 or 32 taps requires close to 18 dB less power than does DFE or SAS. The DFE/SAS system has a lower power requirement, 6 dB, which is achieved at the expense of additional complexity for implementation. COFDM has approximately 3 dB of gain in the power requirement over the OFDM system, which is obtained at the expense of a one-fourth reduction in the data transmission rate. The two examples above lead us to the conclusion that for fixed-bandwidth channels, DFE and SAS provide the highest data rates, at the expense of considerably higher power consumption. The spread-spectrum systems provide better coverage at the expense of lowering the operating data rate.

Figure 9.34 shows the maximum data rate and bandwidth requirement for a power-limited channel with a fixed power of 20 dBm (100 mW). The broadband techniques, test area, and outage requirement in this case remain the same as those in Fig. 9.34. These results are useful for frequency bands such as U-NII bands, in which plenty of band is available, limitations on the power restrict the maximum achievable data rate. Figure 9.34 shows clearly that the maximum supportable data rates are achieved by

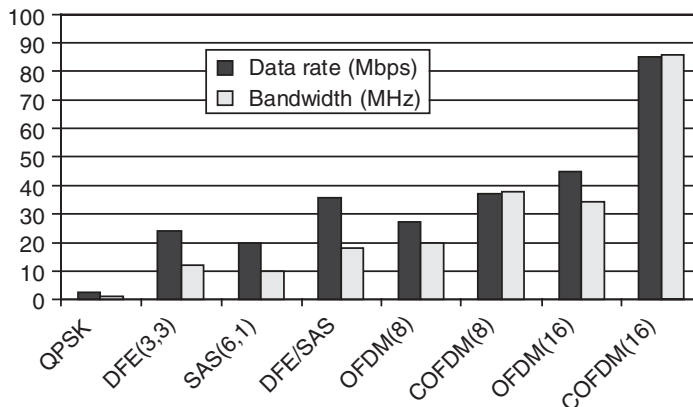


FIGURE 9.34 Performance of broadband modems in a power-limited channel with a fixed transmitting power of 20 dBm.

COFDM. An OFDM modem with 16 carriers can achieve data rates close to 42 Mb/s, and this data rate is doubled when coding is added to the system. The channel spacing and carrier bandwidths in Fig. 9.34 are kept the same as those used in Fig. 9.33. With eight carriers, the data rate drops to slightly more than half of the data rate for a 16-carrier system, but the coding is not as effective as before. The significant difference between the performance of eight carriers and 16 carriers reflects the fact that the bandwidth of the eight-carrier modem is not wide enough to take advantage of the frequency diversity in the signal received [Fal96]. DFE and SAS modems can achieve data rates on the order of 20 Mb/s, which is suitable for the HIPERLAN1 standardization objective of 23 Mb/s. These alternatives were debated in the standardization deliberations. The complexity of the SAS system for mobile applications was a drawback for adopting SAS, and DFE was selected as the modem technology for HIPERLAN1 [Wil95]. Higher data rates, on the order of 30 Mb/s, are obtained using DFE/SAS, at the expense of very complex implementation.

The results provided here are based on QPSK modulation; higher data rates are achievable with more-bandwidth-efficient QAM modulation techniques. To support video applications, the IEEE 802.11a and HIPERLAN2 standardization objectives were to support data rates above 50 Mb/s. As a result, as described in Example 8.4, they selected 64-OFDM with a QAM modulation technique that used up to 6 bits per symbol, described in Examples 8.4 and 8.5.

9.9 MIMO IN FREQUENCY-SELECTIVE FADING

The space-time coding techniques used with MIMO antennas, described in Chapter 8, were first designed for narrowband wireless communications. The performance predictions of the original STC codes were given over a flat-fading channel [Ala98, Tar98]. In frequency-selective multipath fading, the delay spread of the channel stretches transmitted symbols to the following symbols, causing ISI. The ISI caused by multipath arrivals will ruin the orthogonality of the arriving symbols in time, which is essential for optimal performance of STC codes. Therefore, the performance of STC codes in frequency-selective fading degrades significantly. To remedy this situation and maximize the effectiveness of the STC codes, one may employ a channel equalizer at the receiver to eliminate the ISI caused by frequency-selective fading, or the system should use OFDM modulation to eliminate the ISI in the individual carriers. In the next two sections we provide more specific details on how we can combine the STC, MIMO, channel equalization, and OFDM modulation to design a high-performance modem for frequency-selective fading channels.

9.9.1 STC-MIMO and Equalization

Since traditional equalization methods such as LTE, DFE, or MLSE, described in Section 9.6, are designed for SISO channels, we need to modify their implementation to adapt to MIMO channels. For example, for the STC system with two transmitter antennas and one receiver antenna described in Section 8.5, we need to design an equalizer that will equalize for each transmitter antenna given the nature of the received coded information symbols. For details of implementation of such receivers, one may refer to [Dha02]. Here we provide an example to compare the performances.

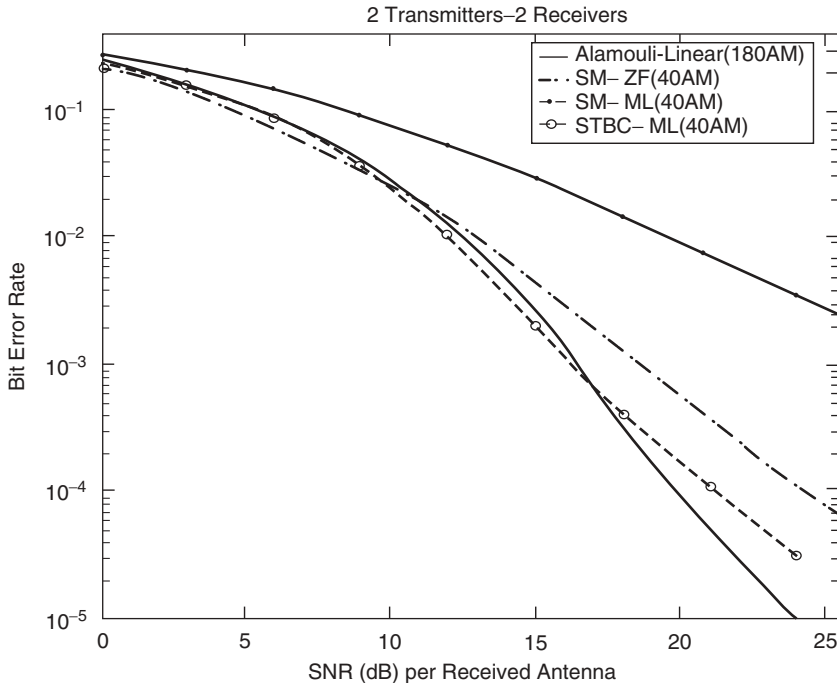


FIGURE 9.35 BER comparisons for various transmission techniques over 2×2 MIMO. At high SNR, from top to bottom: spatial multiplexing (SM)-ZF, SM-ML, STBC-ML, and Alamouti STBC [Ges03].

Example 9.6: STC, MIMO, and Equalization To compare the performance of pure space diversity with a nulling detector used in V-BLAST with simple STC, we use the results of simulation presented in [Ges03]. These results compare four different systems using 2×2 MIMO antennas with uncorrelated elements to support the same data transmission rate. Figure 9.35 illustrates the results of the simulation for this comparison. The four systems are 2×2 Alamouti code with 16-QAM, spatial multiplexing (SM) with ZF and ML detection using 4-QAM, and a more recent STBC code with ML decoding using 4-QAM. The Alamouti code is simulated using 16-QAM to have the same effective data rate as that of the other three systems. The performance of all of these systems is close, except that ZF algorithms do not perform as well as the others because the small number of receiver antennas does not allow effective nulling. Considering the practical aspects, Alamouti codes are very simple to implement, whereas SM systems require more complex receiver algorithms.

9.9.2 MIMO OFDM

In the late 1990s, STC codes and MIMO techniques were discovered and OFDM modulation became very popular in WLAN standardization. As a result, researchers began to develop approaches to combine these technologies for the next generation of WLAN systems. The objective of the LAN industry in general is to provide higher data links to keep up with the growth of speed and memory size of the evolving personal computer and to open new horizons for novel multimedia applications. Therefore,

WLAN researchers began to consider combining OFDM and STC to increase the data rate of the WLANs from the existing 54 Mb/s supported by OFDM modulation.

The MIMO techniques can be used either as STC [Ala98, Tar98] or as space-division multiplexing (SDM) [Fos99, Zel04]. STC, described in Section 8.6, increases the performance of the communication system by coding over the various transmitter antenna branches. The SDM, tailored to achieve higher data rates, uses independent data streams on various transmit branches simultaneously and at the same carrier frequency. As we discussed in Section 9.8, the IEEE 802.11a and g WLAN standards are based on OFDM. A high-data-rate extension of these standards could be based on SDM MIMO and OFDM [Zel04]. This approach leads to a combination of the data rate enhancement of SDM with the robustness of OFDM against frequency-selective fading. Most WLANs operate in a richly scattered multipath indoor environment (described in Chapter 6), which provides good conditions for a high MIMO capacity [Fos99]. The following example illustrates implementation and performance evaluation of a 3×3 MIMO OFDM system that can double the 54-Mb/s data rate of the IEEE 802.11a OFDM system.

Example 9.7: 3×3 MIMO OFDM for IEEE 802.11a Figure 9.36 is a block diagram of the MIMO OFDM system used in [Zel04] for WLAN application. The incoming bits are multiplexed among N_t branches of OFDM transmitters each having N_c subcarriers. Each branch performs encoding, interleaving, QAM mapping, and inverse DFT and adds a cyclic prefix before transmission. A preamble containing training sequences and pilot symbols on predefined subcarriers is inserted into every MIMO OFDM data symbol. The receiver uses the training sequence and the pilot signals to estimate and correct for frequency offset and symbol timing. Then the cyclic prefix is removed and DFT is performed in each receiver branch. The received subcarrier signals in each branch are routed to the MIMO detector to recover the processed transmitted symbols on that subcarrier. The symbols per transmitted stream are then demapped, deinterleaved, and finally, combined to recover the transmitted data stream.

Figure 9.37 shows the results of simulations to evaluate packet error rate (PER) performance for a 3×3 MIMO OFDM with data rates of 72 Mb/s in part (b), 108 in (c), 144 in (d), and 162 in (e) and its comparison with the 54-Mb/s IEEE 802.11a SISO system (Fig. 3.37a). The performance of 72- and 108-Mb/s 3×3 MIMO OFDM is better than the performance of the 54-Mb/s SISO OFDM used in IEEE 802.11a. Figure 3.37f is the same as Fig. 3.37b, the difference being that it is generated assuming perfect channel estimation at the receiver. Comparing part (b) with part (f), we observe that applying channel estimation results in a loss of more than 4 dB. Furthermore, we note that the curve falloff of MIMO systems is faster than that of the SISO system used in this example.

APPENDIX 9A: ANALYSIS OF THE EQUALIZERS

This appendix provides a general derivation for calculation of tap gains and minimum mean-squared error at the output of an adaptive equalizer. In these derivations, parameters are selected so that with a known transmitted pulse waveform and overall channel impulse response, one can calculate the minimum MSE at the output of an equalizer for different tap spacing, equalizer architecture, and sampling time error rates. The

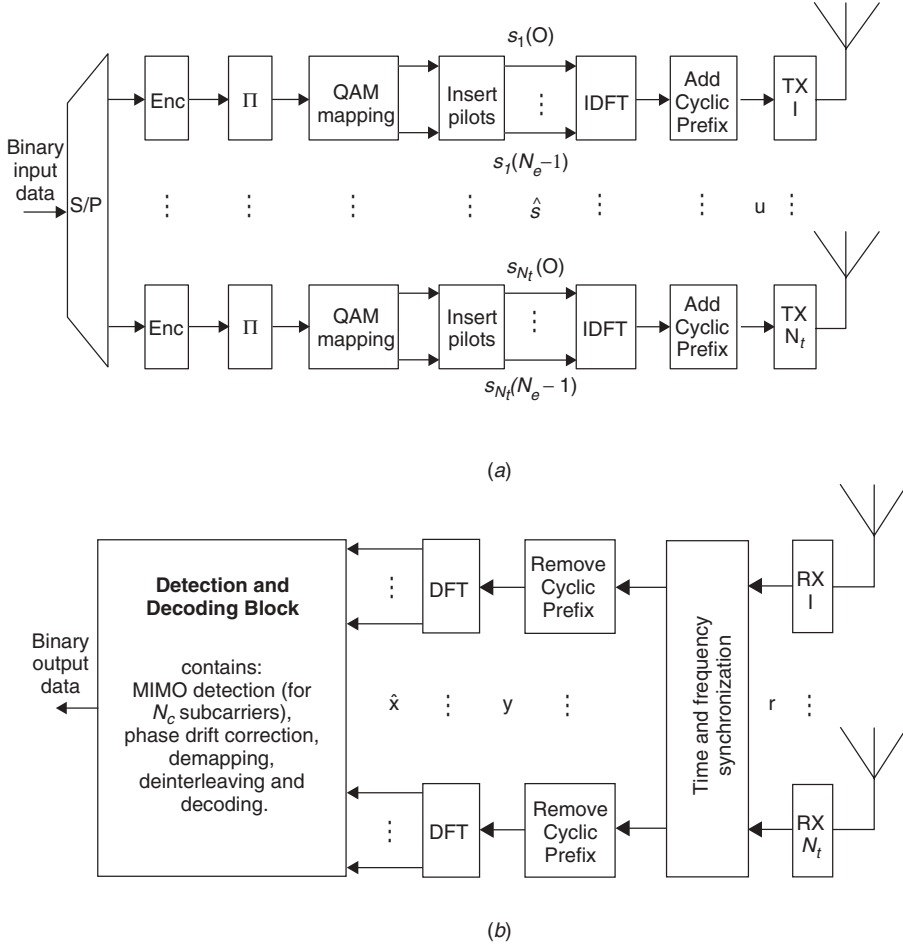


FIGURE 9.36 SDM MIMO-OFDM system block diagram: (a) transmitter; (b) receiver [Zel04].

minimum MSE at the output of the equalizer reflects the SNR after equalization that is used for comparative performance evaluation of different equalizers.

Calculation of the optimum tap gains for equalizers is treated thoroughly in the literature [Pro01]. Normally, the tap values that minimize the mean-squared error can be found from the solution of a set of linear equations. The following set of equations, derived in [Bel84, Pah88b], provides a unified solution for both the LTE and FSE as a function of tap spacing and timing error:

$$\sum_{j=-L}^N \left[f_j \overline{|a_k|^2} \sum_p g(pT + \tau T - j\Delta) g^*(pT + \tau T - l\Delta) + N_0 \delta_{jl} \right] = \overline{|a_k|^2} g^*(\tau T - l\Delta), \quad -L \leq l \leq N \quad (9A.1)$$

where the $\{f_i\}$ are the optimum tap gains of the equalizer; Δ is the tap spacing; $g(t) = f(t) * h(t)$ is the overall channel impulse response, including the modem filtering $f(t)$

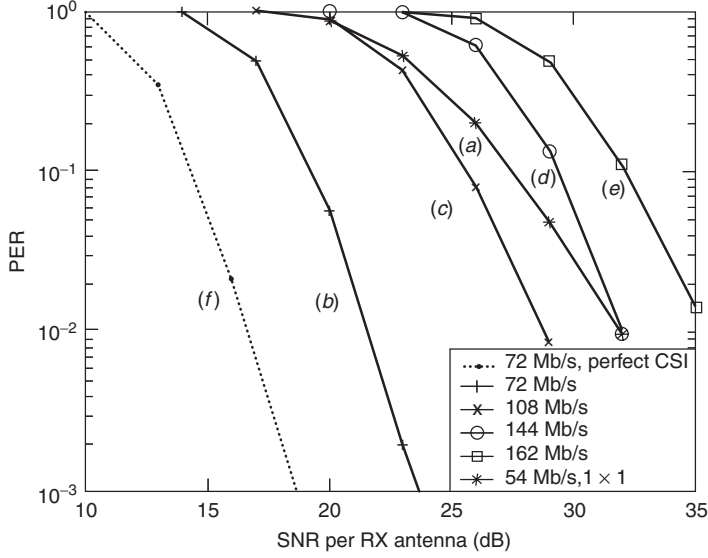


FIGURE 9.37 Performance of MIMO OFDM using IEEE 802.11a: (a) SISO with 64-QAM and $r = \frac{3}{4}$; (b) 3×3 MIMO with 16-QAM and $r = \frac{1}{2}$; (c) 3×3 MIMO with 16-QAM and $r = \frac{1}{2}$; (d) 3×3 MIMO with 64-QAM and $r = \frac{2}{3}$; (e) 3×3 MIMO with 64-QAM and $r = \frac{3}{4}$; (f) the same as part (b) with perfect channel estimation [Zel04].

and channel characteristic $h(t)$; τ is the normalized sampling-time error; and $\delta_{ji} = 1$ for $j = l$ and zero otherwise. The minimum mean-squared error calculated for this system is given by

$$\xi_{\min} = \overline{|a_k|^2} \left[1 - \sum_{i=-L}^N f_i g^*(\tau T - i\Delta) \right] \quad (9A.2)$$

where the $\{f_i\}$ are the optimal values of the tap gains found from Eq. (9A.1).

As shown in [Pah88b], for a DFE with $N + 1$ forward taps and M feedback taps, the optimum forward tap gains $\{f_i\}$ and feedback tap gains $\{b_i\}$ are the results of the solution of the following sets of linear equations:

$$\begin{aligned} \sum_{j=-L}^N \left[f_j \overline{|a_k|^2} \sum_p g(pT + \tau T - j\Delta) g^*(pT + \tau T - l\Delta) + N_0 \delta_{jl} \right] \\ - \overline{|a_k|^2} \sum_{j=1}^M b_j g^*(jT + \tau T - l\Delta) \\ = \overline{|a_k|^2} g^*(\tau T - l\Delta), \quad -N \leq l \leq 0 \end{aligned} \quad (9A.3)$$

and

$$b_l = \sum_{j=-N}^0 f_j g(lT - j\Delta + \tau T), \quad 1 \leq l \leq M \quad (9A.4)$$

The right-hand side of Eq. (9A.4) is a convolution of the sampled channel impulse response from the transmitter to the input of the forward equalizer taken at a time spacing of Δ , with the discrete-time impulse response of the forward equalizer. The results are sampled at time T , and only values on the right-hand side of the center sample are calculated; these samples are associated with ISI due to the past M transmitted symbols. Therefore, the optimum tap gains completely eliminate the ISI due to the past M samples. The value of the minimum MSE is

$$\xi_{\min} = \overline{|a_k^2|} \left[1 - \sum_{i=-L}^0 f_i g^*(\tau T - i\Delta) \right] \quad (9A.5)$$

At first glance, this equation suggests that the MMSE is independent of the backward tap gains, but this conclusion is incorrect. In Eq. (9A.3), when the optimum tap gains are determined, one minimizes the error jointly for both sets of tap gains. As a result, optimum values of forward tap gains are affected by the optimum values of backward tap gains, which influence the MMSE indirectly.

The equalizers introduced thus far are suitable for PAM, PSK, and QAM modulation techniques. For other modulation techniques the setup for optimum tap gains and MMSE can be different. Equations for the calculation of tap gains of the equalizer for SQPSK and SQPR are given in [Bel84]. A comparison of the performances of SQPSK, SQPR, QPS, and QPR over a frequency-selective microwave LOS channel is given in [Pah85a].

QUESTIONS

- (a) What was the first adaptive receiver for frequency-selective fading channels, what was the modulation technique for this system, and how it was implemented?
- (b) Name two methods for implementation of a DMF. Explain how these methods counteract the effects of frequency-selective multipath fading.
- (c) What is an adaptive MLSE, and how does it work?
- (d) What are the advantages and disadvantages of fractionally spaced equalizers?
- (e) Why is the DFE technique a popular choice for equalization in radio modems?
- (f) Name two applications in which fast-converging algorithms are useful in modem design.
- (g) Name two applications for blind equalization.
- (h) What typical data rates can a BPSK/DFE modem provide in LOS indoor areas?
- (i) Explain why a sectored antenna can be used in a WLAN to increase the maximum supportable symbol transmission rate.
- (j) Explain why multi-amplitude and multiphase modulation techniques are not very popular in wireless voice-oriented standards such as GSM or IS-95 but are used extensively in wireless data applications such as EDGE, HDR, or IEEE 802.11a and g.

- (k) Given that a bandwidth constraint does not exist, what is the data-rate limitation for a multicarrier system?
- (l) Why do mobile radio systems such as GSM use equalization rather than time gating and discrete matched filtering?
- (m) Explain the difference between implicit and explicit diversity. Which form of diversity is attainable in flat-fading channels?
- (n) What is the difference between MCM and OFDM modulations?
- (o) Why has OFDM emerged as the choice of most wideband modems used for wireless data communications?
- (p) Why is time gating used for implementation of an OFDM modem?
- (q) Explain how Doppler shift causes adjacent carrier interference in OFDM modems designed for wireless data communications.

PROBLEMS

1. (a) Give the block diagram of a discrete channel model for fading multipath channels and identify the tap gains of the model in terms of channel impulse response and front-end filtering at the transmitter and the receiver.
- (b) Give the block diagram of a channel estimator that estimates the taps of the discrete channel model.
- (c) Using the channel estimator in part (b), draw the DMF for the discrete channel model of part (a).
2. Assume that a coherent BPSK modem operating at the data rate $R = 1/T$ uses a raised-cosine pulse with 50% roll-off ($\alpha = 0.5$).
 - (a) Sketch the signal-to-intersymbol interference power (the variance of the sum of ISI terms) versus normalized timing error τ/T for $-T/2 < \tau < T/2$.
 - (b) Assume that the ISI noise forms a Gaussian-distributed interference and that the effects of the additive noise are negligible. Sketch the error rate versus normalized timing error τ/T for $-T/2 < \tau < T/2$.
 - (c) Repeat part (b) for a received SNR of 10 dB.
 - (d) Repeat part (a) by calculating the exact value of the probability of error. To calculate the exact value, every possible bit pattern causing ISI from neighboring symbols has to be considered separately. The error rate is the average of the bit error rates over all bit patterns.
3. Use the results of data rate versus power consumption for the central part of the Atwater Kent Laboratories building (see Figs. 9.33 and 9.34) to answer the following questions:
 - (a) For 10 MHz of bandwidth, what power requirement and maximum data rate are supported by OFDM(16) and DFE(3,3)?
 - (b) What are the maximum data rate and required bandwidth for OFDM(16) and COFDM(16) modulations to cover this area with 100 mW of power?
 - (c) Name two standards using DFE and OFDM for implementation of WLANs.

4. Consider the IEEE 802.11g standard, which uses BPSK and $r = \frac{3}{4}$ codes for 9-Mb/s information transmission and 16-QAM with the same coding for an actual payload data transmission rate of 36 Mb/s.
- What is the coded symbol transmission rate per subcarrier for each of the two modes? What is the bit transmission rate per subcarrier for each of the two modes?
 - If one switches from 36-Mb/s mode to 9-Mb/s mode, how much more (in decibels) path loss can it afford?
- If the system were covering up to 50 m with 36 Mb/s, what would be the coverage using the 9-Mb/s mode? Assume a distance power gradient of 2.5.
5. Using Fig. 9.4, determine the equivalent discrete channel model from the information source at the transmitter up to the digital signal processor at the receiver for a 10-Mb/s BPSK wireless modem using a raised-cosine pulse with a roll-off factor of 0.5. Assume that the channel is WSSUS and that its delay power spectrum is given by

$$Q(\tau) = T e^{-\tau/T}$$

where $T = 50$ ns.

- If we neglect the taps with power 10 dB below the tap with maximum power, how many taps are needed for the model?
 - What are the eigenvalues of the covariance matrix of the tap gains?
 - Repeat parts (a) and (b) for a 20-dB threshold.
6. The equivalent discrete channel model for a fading multipath channel has two taps. The modulation is DBPSK, the transmitted symbols are time-gated to avoid ISI, and the receiver has a two-tap DMF.
- What is the bandwidth efficiency of the system?
 - Sketch the probability of outage versus the threshold SNR γ_{out} if the eigenvalues of the covariance matrix for the equivalent discrete channel model are 0.4 and 0.6.
 - Sketch the average probability of error versus $\overline{\gamma_b}$.
 - Repeat parts (a) and (b) for eigenvalues of 0.1 and 0.9, and compare the results found with the two sets of eigenvalues.
7. A voiceband modem operates over a telephone channel. The equivalent low-pass received signal is given by

$$r(t) = \sum_p a_p h(t - pT) + \eta(t)$$

where $h(t)$ is the impulse response of the channel, including the echoes in the line; $\eta(t)$ is the additive white Gaussian noise with variance N_0 ; and a_p is the transmitted symbol value. For a BPSK system, $a_p = \pm 1$.

To eliminate the echoes for a BPSK modem, the structure shown in Fig. P9.1 is used.

- By minimizing the MSE, derive the normal equations and solve for the optimum taps of the canceler $\{c_i\}$ in terms of $h(t)$.

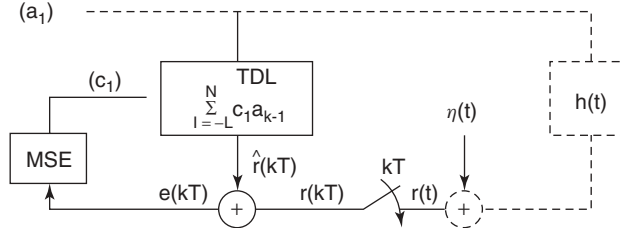


FIGURE P9.1

- (b) Determine the minimum value of the MSE, $\xi_{\min}$, in terms of $\{c_i\}$, $h(t)$, and N_0 . The equivalent baseband impulse response of a channel is given by

$$h(t, \tau) = A\delta(t) - B\delta(t - \tau)$$

A BPSK modem with coherent detection using raised-cosine pulses with $\alpha = 0$ and $E_S = 1$ is operating over this channel.

- (a) Assume that $A = B = 1$ and $E_S/N_0 = 10$ dB and plot the probability of error versus τ_{rms}/T for $0 < \tau_{\text{rms}}/T < 5$, where $R = 1/T$ is the bit rate of the channel. Use the Gaussian assumption for the calculation of error rate. With the Gaussian assumption, the ISI caused by channel multipath is treated as an additive Gaussian noise source. The total noise affecting the system is the additive thermal noise with variance N_0 plus the variance of the ISI term.
- (b) Repeat part (a) for $B = 0.5$, placing both plots in the same graph.
- (c) Assume that A and B are independent slowly time-varying Rayleigh random variables with variance 1 and $E_S/N_0 = 40$ dB and repeat part (a). Assume that the variance of B is 0.5 and repeat part (b).

PROJECTS

Project 1. Performance Analysis of an Equalizer

In this project we analyze the performance of an equalizer in a radio channel using MATLAB or an alternative computation tool. The analysis uses derivations provided in Appendix 9A to calculate the mean square error (MSE) of the received signal for variety of equalizers. The MSE is a measure of the signal to noise ratio of the equalized received signal. When it is compared with the MSE before equalization, it reveals the effectiveness of the equalizer to improve the performance of the receiver. The general equations provided in Appendix 9A allows calculation of performance of the linear, decision feedback, and fractionally spaced equalizers for different sampling time. One of the benefits of this project is to demonstrate the sensitivity of different equalizers to the sampling error. Less sensitivity to sampling time is desirable because it reduces the complexity of the time synchronization between the transmitter and the receiver.

Assume that we have a BPSK modem that uses ideal zero roll-off raised-cosine pulses for pulse shaping. The modem is operating over a microwave line-of-sight channel with a normalized ($T = 1$) frequency response of

$$H(j\omega) = 1 - 0.1e^{-j0.2\omega}$$

In this system the transmitted digits are $a_n = \pm 1$ and the received SNR ratio is $E_b/N_0 = 15$ dB, where E_b is the average received energy per bit and N_0 is the variance of the received noise after prefiltering at the receiver.

- Sketch and label the transmitted and the received pulse shapes used for communication over the channel. (*Hint*: You may use the delay property of the Fourier transform to determine the received overall impulse response).
- Sketch the MSE of the received signal versus τ/T , the normalized sampling time difference between the transmitter and the receiver, for $|\tau/T| \leq 1$. The MSE of the received signal is the variance of the difference between the desired detected symbols and the sampled signal used for detection.
- Repeat part (b) assuming a linear equalizer with three T -spaced taps is used at the receiver. Here you need to use the equation in Appendix 9A with the overall channel impulse response determined in part (a) to calculate the taps of the equalizer. These tap values are then used in the equation for calculation of the minimum MSE of the linear equalizer in the Appendix 9A.
- Repeat part (b) assuming a linear equalizer with three $T/2$ -spaced taps.
- Repeat part (b) assuming a DFE with two T -spaced forward taps and one feed-back tap.
- Repeat part (b) assuming a DFE with two T -spaced forward taps and one feed-back tap.

Project 2. Simulation of a Simplified IEEE 802.11a/g OFDM

The OFDM technique has been adopted in several wireless LAN standards, including IEEE 802.11a, IEEE 802.11g, HIPERLAN/2, as well as the local multipoint distribution service (LMDS) and digital audio broadcast (DAB) systems. In this project we implement IEEE 802.11a/g and HIPERLAN-2 OFDM modulation and demodulation techniques in the MATLAB software. The simulation model to be implemented in this project is a simplified version of the IEEE 802.11a standard. The complete standard document can be downloaded from *IEEE Xplore* (<http://ieeexplore.ieee.org>).

Figure P9.2 shows a simplified version of the IEEE 802.11 a/g OFDM system, which implements baseband OFDM modulation and demodulation techniques. A complete

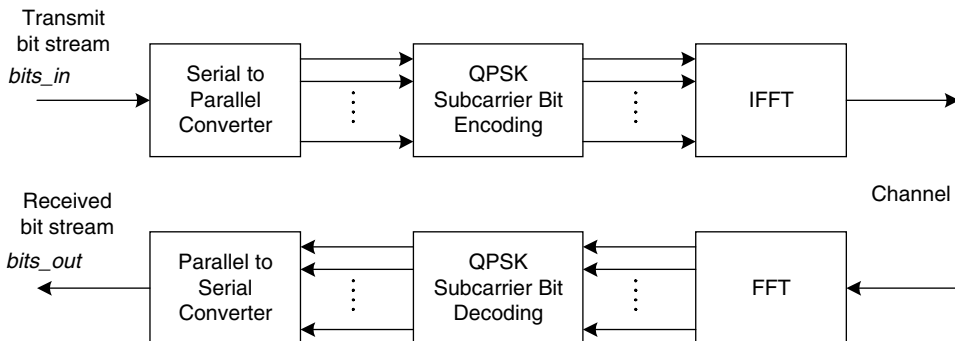


FIGURE P9.2 Simplified simulation model of baseband OFDM transceiver.

TABLE P9.1 Simulation Parameters for the OFDM Project

Parameter	Value
N _{st} : number of subcarriers	52
N _{FFT} : number of IFFT/FFT inputs	64
BW: bandwidth	20 MHz
delta_F: subcarrier frequency spacing	0.3125 MHz (= BW/N _{FFT})
T _{FFT} : IFFT/FFT period	3.2 us (= 1/delta_F)

simulation needs additional details to include error control coding, interleaving, cyclic prefix, pilot subcarriers, waveform shaping, carrier modulation, and the effects of radio propagation channels that are beyond the scope of this project.

The system model to be implemented in this project is shown in Fig. P9.2 and the simulation parameters are listed in Table P9.1. The serial transmit bit stream is first converted to parallel data by serial-to-parallel converter, which outputs two serial bits at each of its output lines (52 output lines in this simulation since N_{st} = 52). Then each group of two bits is encoded into a complex QPSK symbol with the constellation shown in Fig. P9.3. Following that, 52 complex QPSK symbols are mapped to the input of the IFFT block according to the mapping scheme shown in Fig. P9.4, where the *null* input are set to zero. The IFFT is performed to obtain a time-domain OFDM symbol (with 64 complex samples) of duration 3.2 us (T_{FFT} = 3.2us). At the receiver the transmitted bit streams can be recovered through the inverse process, as shown in the lower part of Fig. P9.2.

The followings are detailed requirements on the implementation and deliverables:

- (a) Implement all the blocks shown in Fig. P9.2, except IFFT and FFT, for which you may use the MATLAB function *ifft()* and *fft()*.
- (b) Generate a vector of 104 random binary bits, *bits_in*, and use this vector as the input to the system shown in Fig. P9.2 that you implemented. Receive the bit

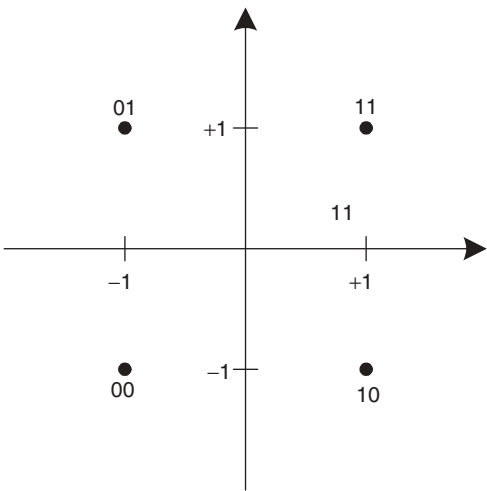


FIGURE P9.3 Constellation of QPSK for subcarrier bit encoding.

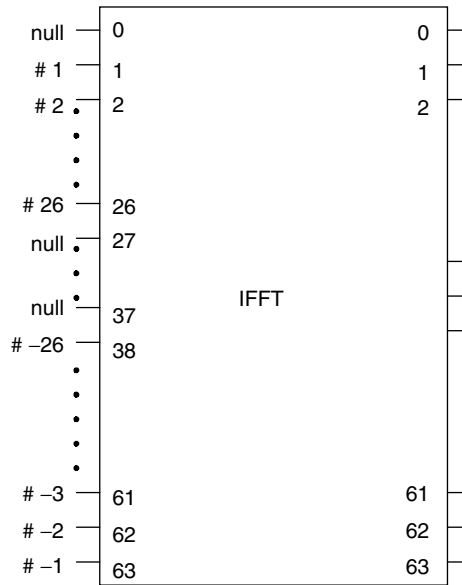


FIGURE P9.4 Inputs and outputs of IFFT.

vector at the output of the receiver, *bits_out*. Note that the bit vector *bits_in* and *bits_out* are both of the same length, 104. Verify in your MATLAB program that the received bit stream is exactly the same as the transmitted bits by observing the output of the following code segment, which should be included in your MATLAB program:

```
if sum(abs(bits_in - bits_out)) == 0
    disp('Transmitted bits are successfully received
        !');
else
    disp('Transmitted bits are received in ERROR !');
end;
```

- (c) Plot the real-part and the image-part of the time-domain OFDM symbol, i.e., the in-phase and quadrature signals, that you generated in your simulation in separate figures with a proper time axis. Since the duration for an OFDM symbol is $3.2 \mu\text{s}$ according to Table P9.1, the time axis should range from 0 to $3.2 \mu\text{s}$.
- (d) Plot the spectrum of the OFDM symbol that you generated in this simulation. Since the bandwidth is 20 MHz, the frequency axis of the plot should range from -10 MHz to 10 MHz .

The required submission for this project includes the MATLAB source code for the simulations and the generated plots.

10

SPREAD-SPECTRUM AND CDMA TECHNOLOGY

- 10.1 Introduction
 - 10.1.1 Spread-Spectrum Performance in Frequency-Selective Fading
 - 10.1.2 Spread Spectrum in Standards and Products
- 10.2 Principles of Frequency-Hopping Spread Spectrum
 - 10.2.1 FHSS in Frequency-Selective Fading
- 10.3 Principles of Direct-Sequence Spread Spectrum
 - 10.3.1 DSSS in Multipath Fading and the RAKE Receiver
 - 10.3.2 Performance of the RAKE Receiver
- 10.4 Interference in Spread-Spectrum Systems
 - 10.4.1 Interference in FHSS Systems
 - 10.4.2 Interference in DSSS Systems
 - 10.4.3 Effects of Coding
- 10.5 Performance of CDMA Systems
 - 10.5.1 Direct-Sequence CDMA
 - 10.5.2 M -Ary Orthogonal Codes
 - 10.5.3 Interference Cancellation and Multiuser Detection
 - 10.5.4 Multicarrier CDMA
 - 10.5.5 MIMO CDMA
 - 10.5.6 Frequency-Hopping CDMA
 - 10.5.7 Time-Hopping CDMA
- Questions
- Problems

10.1 INTRODUCTION

Spread-spectrum communications and code-division multiple access (CDMA) are now mature technologies with many highly developed subdisciplines, including modulation, coding, synchronization, and acquisition methods, and system implementation aspects. To treat all aspects of spread-spectrum communications thoroughly in one chapter is not feasible, nor is it our purpose here. Instead, we provide an overview of those aspects

of spread-spectrum communications that are specific to wireless information networks and principal issues that arise in the application of spread-spectrum techniques in a multiuser wireless environment.

The distinguishing characteristic of spread-spectrum communications is that the signals used for the transmission of information have a much wider bandwidth than that of the underlying information bit rate of the system. There are two basic traditional methods for implementing a spread-spectrum system: direct-sequence spread-spectrum (DSSS) and frequency-hopping spread-spectrum (FHSS). In this chapter we discuss the basic principles of DSSS and FHSS systems, the effects of interference and multipath fading on system performance, and CDMA and multicarrier CDMA (MC-CDMA), which are built on spread-spectrum transmission and reception. In Chapters 11 and 15 we discuss in greater detail the capacity of voice-oriented CDMA and the implementation of high-data-rate systems for Internet access using CDMA technology. In Chapter 12 we discuss using CDMA technology in UWB systems. Detailed treatments of other aspects of the design and analysis of spread-spectrum communication systems, as well as historical accounts of the origins of spread-spectrum technology, can be found in [Coo78, Dix84, Sim85, Gar00a, Pro01, Pah02a,b].

Since World War II, spread-spectrum technology has been used extensively for military communications, where it is attractive because of its resistance to interference and interception, as well as its amenability to high-resolution ranging [Sim85]. In the past several decades, commercial applications of spread-spectrum techniques have attracted considerable attention because of their amenability to CDMA operation, the possibility of spectral overlay, and the availability of unlicensed commercial bands allocated to this technology. The wide bandwidth of a spread-spectrum signal can be used to reduce the harmful effects of multipath, resulting in an increase in signal coverage and mobility. Spread spectrum for use in outdoor packet radio communications was introduced in late 1970s [Kah78], and for use with CDMA for outdoor packet data in the early 1980s [Ray81, Kav81, Mus82]. Spread-spectrum use for indoor applications was introduced in the early 1980s [Fer80]; use of spread spectrum and CDMA for office information networks was suggested [Pah84, Pah85b], analyzed, and implemented [Kav85, Kav87] for wireless PBX systems in the mid-1980s. After the May 1985 FCC release of unlicensed ISM bands for spread-spectrum technology trials [Mar85, Mar87a], various spread-spectrum commercial products, from low-speed fire safety devices to cordless telephone systems and high-speed wireless local area networks (WLANs), have appeared in the market. Since Qualcomm's introduction in the late 1980s of CDMA technology for licensed cellular bands, this technology has emerged as the choice for third-generation cellular networks. Today, both the voice- and data-oriented wireless information network industries have exploited spread-spectrum technology. However, the two industries have somewhat different motivations for adopting spread spectrum.

The voice-oriented digital cellular and personal communications services (PCS) industries use CDMA spread-spectrum as an alternative to TDMA/FDMA networks to increase system capacity, provide a more reliable service, provide soft handoff of cellular connections, and facilitate integration of services at variable data rates to support the quality of service requirements. In wireless data-oriented communications, spread-spectrum technology is used in the WLAN and wireless personal area networking (WPAN) industries. The first unlicensed bands available for high-speed data communication were the ISM bands, which originally were released to be used with

spread-spectrum transmission. That initiative resulted in adoption of spread-spectrum technology as the choice of the first IEEE 802.11 standard for 2-Mb/s WLANs and its follow-up IEEE 802.11b standard operating at 11 Mb/s. More recently, the IEEE 802.15 standard has considered CDMA technology as one of the options for implementation of low-power high-data-rate WPANs in UWB unlicensed operations.

One form of spread-spectrum communications receiving much attention in the wireless industry is CDMA. The CDMA technique is a well-proven technology that has been used in a number of military communication systems before it penetrated second-generation cellular networks in the late 1980s and become the dominant technology for third-generation cellular networks in late 1990s. In CDMA spread-spectrum transmission, user channels are created by providing different transmission codes for different users. The privacy of transmitted user information is readily provided by controlling the distribution of user-unique code sequences.

In summary, the principal advantages of spread-spectrum transmission are as follows:

1. Spread-spectrum signals can be overlaid onto bands where other systems are already operating, with minimal performance impact to or from the other systems.
2. The anti-multipath characteristics of spread-spectrum signaling and reception techniques are attractive in applications where multipath is likely to be prevalent. (Achieving good performance in frequency-selective fading may require the use of a RAKE receiver, which is in effect a matched filter for a multipath channel.)
3. The anti-interference characteristics of spread spectrum are important in some applications, such as networks operating on manufacturing floors, where the signal interference environment can be harsh.
4. Cellular systems designed with CDMA spread-spectrum technology offer greater operational flexibility and possibly a greater overall system capacity than those of systems built on TDMA access methods.
5. The convenience of unlicensed spread-spectrum operation in ISM bands is attractive to manufacturers and users alike.
6. High-resolution ranging achievable with spread-spectrum technology is attractive for emerging location-aware broadband wireless ad hoc networks.

10.1.1 Spread-Spectrum Performance in Frequency-Selective Fading

The performance of a spread-spectrum system in flat fading and in the absence of interference is similar to that of any other modulation. For any modulation method in flat fading, all spectral components of the transmitted signal fade in unison, and these power fluctuations make it necessary to provide additional transmitted power (fade margin) to bring the average BER performance to an acceptable level. As we discussed in Chapter 8, the effects of fading are reduced significantly, and the necessary fade margin reduced accordingly, if we use diversity-combining techniques at the receiver. The BER formulas given in Chapter 8 for conventional modulation techniques apply equally well to spread-spectrum modulation systems. In flat fading with high Doppler spread (i.e., in rapid flat fading), coding techniques can provide performance improvements similar to those provided by diversity combining. Just as diversity is most effective when the diversity branches fade independently, coding is most effective when the symbols in a codeword fade independently. If the Doppler spread is low, we can use

interleaving techniques to break up the clusters of consecutive errors produced by the slow fading, and thereby make the error-correcting codes operate more efficiently. However, in using spread-spectrum techniques on frequency-selective fading channels, new considerations arise, which we discuss next.

As we saw in Chapter 9 in discussing wideband modulation techniques, frequency-selective fading in the frequency domain is manifested as ISI in the time domain. As we showed, the effects of frequency-selective fading can be controlled by several means, including multiple and sectored antennas, adaptive equalization, and multirate modems. However, one of the major features of spread spectrum is its inherent resistance to frequency-selective fading. In addition, DSSS and FHSS systems both lend themselves to the design of receivers that actually take advantage of the multipath characteristics associated with frequency-selective fading. The approaches used for the design of these receivers are quite different for DSSS and FHSS systems, as we shall see. The effects of frequency-selective fading are, in a sense, similar to the effects of narrowband interference. With a narrowband interferer, a portion of the transmission bandwidth is hit by the interference, and the SNR in that part of the band is degraded relative to other parts of the band. In frequency-selective fading a portion of the band is affected by a deep null in the channel frequency response, and as with narrowband interference, the SNR in that part of the band is degraded. However, despite the close similarity of these situations, the methods used to deal with the two problems are not necessarily the same and we discuss them in separate sections.

10.1.2 Spread Spectrum in Standards and Products

In the voice-oriented cellular wireless networking industry spread-spectrum application started with QUALCOMM's adoption of the CDMA technology as an alternative replacing AMPS cellular network using FDMA technology. The first demonstration of technical feasibility by QUALCOMM was presented in 1989. In 1991 the company announced a common air interface (CAI) specification for its CDMA technology, and in 1993 TIA published this CAI as the interim standard IS-95. Today IS-95 is also referred to as cdmaOne. In the late 1980s, ITU formed a group study to specify wireless standards for high-speed data and multimedia services that is currently referred to as the 3G IMT-2000 standard. The most prominent proposals for IMT-2000 adopted CDMA technology. The two leading solutions are cdma2000, which evolved from the cdmaOne CAI, and the Pan European UMTS standard, which builds a wideband CDMA CAI on the successfully deployed GSM infrastructure. The key requirements for the IMT-2000 were improved voice quality, increased capacity, support of data rates of up to 384 kb/s everywhere and 2 Mb/s for local areas, and support of integrated multimedia applications. The CDMA technology was adopted because it provides a better quality of voice and is a more flexible air interface to customize multimedia applications.

In the WLAN industry the first experimental trial of the DSSS technology was reported by the HP laboratories in 1980 [Fer80]. After the release of ISM bands in 1985, several companies developed DSSS and FHSS WLAN products that formed the foundation for the first IEEE 802.11 standard, completed in 1997, for 1 and 2 Mb/s WLANs. In 1998, IEEE 802.11b adopted the complementary code keying (CCK) technology, which uses DSSS codes for M -ary orthogonal coding to achieve a data rate of 11 Mb/s. Today, IEEE 802.11a and g WLANs use the OFDM technology to achieve data rates up to 54 Mb/s that will divert this industry from spread-spectrum technologies.

In the WPAN industry, IEEE 802.15.1a adopted FHSS Bluetooth technology as its first standard for 1-Mb/s operation in 1998. The Bluetooth technology uses the FHSS CDMA technology, which allows multiple users adopting different codes in the same frequency bands. The FHSS used in the IEEE 802.11 standard uses three sets of nonoverlapping frequency pattern codes. Today, the IEEE 802.15.3a is considering UWB technology for WPAN applications, which is expected to support higher data rates, a larger number of users, and faster connection time than the Bluetooth technology. One of the two existing UWB technologies considered for IEEE 802.15.3a uses DSSS CDMA technology as part of its access mechanism. In addition to the cellular, WLAN, and WPAN industries, a number of consumer products, such as cordless telephones, monitoring systems, and controller units, have emerged in the market, utilizing spread-spectrum technologies in ISM bands.

10.2 PRINCIPLES OF FREQUENCY-HOPPING SPREAD SPECTRUM

The FHSS technique can also be described as a two-layer modulation technique. The first layer can be any standard digital modulation technique, and the second layer is M -ary FSK. The digitally modulated signal makes a *pseudorandom* or *pseudonoise* (PN) selection of one of the M frequencies as its carrier frequency. In other words, the carrier frequency of the digitally modulated data is hopped over a wide range of frequencies prescribed by a periodic PN code. The hopping of the carrier produces the desired spreading of signal spectrum transmitted. The changes in the carrier frequency do not affect the performance in additive noise, and the AWGN performance remains exactly the same as the performance of the digitally modulated system without frequency hopping.

Example 10.1: LFSR Codes and FHSS The maximal-length sequences or M -sequences that are created using maximal-length linear feedback shift registers (LFSRs) of length m are very popular in CDMA cellular networks. The name arises from the fact that M -sequences are the longest sequences that can be generated by an m -stage LFSR. The length of the PN sequence before it repeats itself is $2^m - 1$. These codes are represented by polynomials that denote the connections that are active in the LFSR. Figure 10.1a represents the shift register implementation of an LFSR with length $m = 3$. The polynomial $G(x) = 1 + x + x^3$ defines the code, and it represents feedback connections from the first (exponent 0) and second (exponent 1) shift register stages (but not the third). Figure 10.1b represents the states of the register in eight consecutive time intervals and associates a center frequency with each of these states. Figure 10.1c represents the time-frequency behavior of hops of a FHSS system using an LFSR of length 3. While information bits are transmitted, the center frequency of the carrier hops pseudorandomly.

Figure 10.2 shows a block diagram of a typical transmitter and receiver for a FHSS system. At the transmitter, the digital modulation and the modulation over the hop frequencies are implemented in two stages. The hop frequencies are selected randomly using a frequency synthesizer controlled by the PN code generator. A wideband filter is applied to the signal for spectral shaping before the signal is fed to the antenna. The receiver has a wideband front-end filter that accommodates the entire system

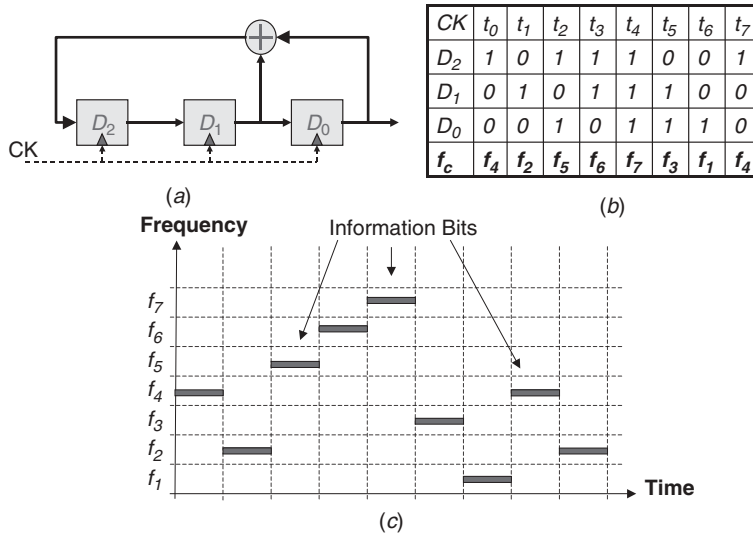


FIGURE 10.1 LFSR code of length 7 with generation function of $G(x) = x^3 + x + 1$: (a) three-memory-stage circuit for implementation; (b) states of the memory in time and frequency; (c) transmission of information bits in time and frequency in FHSS.

bandwidth. This filter is followed by a PN-code-controlled frequency synthesizer synchronized to the transmitter frequency synthesizer. After desynthesizing, the signal is passed through a noise-reduction bandpass filter with the same bandwidth as the transmitted information symbols. The final stage of the receiver is the data demodulator, which demodulates the first-stage digital modulation.

In an FHSS system the interval of time spent at each hop frequency is referred to as the *chip duration*. In contrast with a DSSS system, the chip duration in an FHSS system is not determined by the inverse of the bandwidth but instead, constitutes an independent design parameter. In commercial wireless packet data networks using FHSS, such as Bluetooth and IEEE 802.11, the dwell times are designed long enough to allow transmission of a full packet in each hop. In this way, retransmission of a corrupted packet occurs in a different center frequency allowing time-frequency diversity in reception of the packets.

Example 10.2: Packet Length in Bluetooth The hopping rate of Bluetooth devices, 1600 hops/s, provides for a dwell time of $1/1600 = 625 \mu\text{s}$. The data transmission rate of the Bluetooth devices is 1 Mb/s, which allows transmission of a packet of length less than 625 bits in each hop. Bluetooth also allows three- and five-slot packet data transmissions, for which the hop frequency remains unchanged.

Example 10.3: Packet Length in FHSS IEEE 802.11 The hopping rate of the FHSS IEEE 802.11 is 2.5 hops/s, which allows a dwell time of 400 ms for transmission of the data packets. Since the data rates specified by this standard are 1 and 2 Mb/s, the packet lengths can be as high as 800 kb or 100 kb. The actual maximum length of the data packets specified by the standard is about 4 kB.

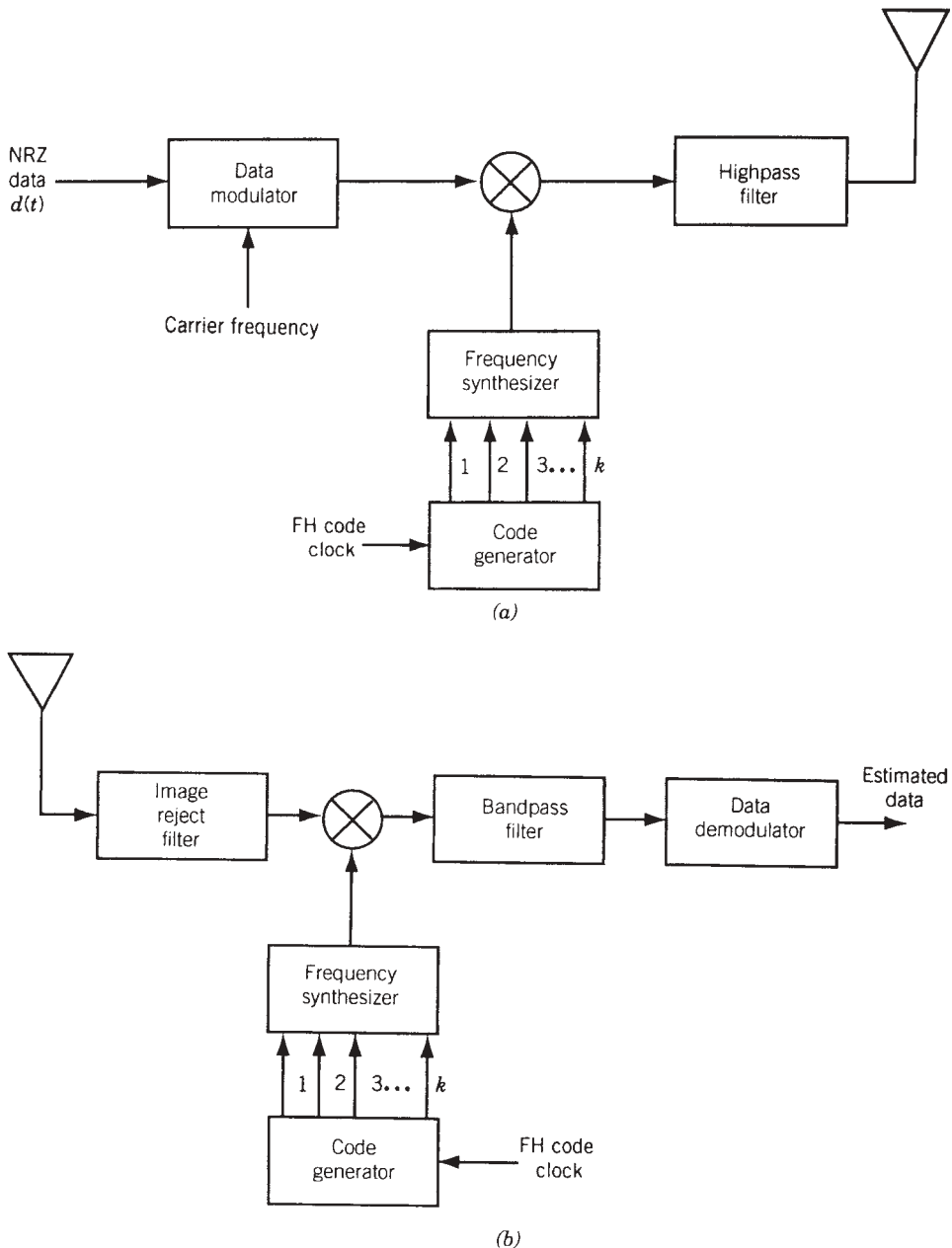


FIGURE 10.2 Block diagram of a frequency-hopping spread-spectrum modem: (a) transmitter; (b) receiver.

Just as with the DSSS technique, the FHSS technique can allow the coexistence of several systems with orthogonal codes in the same frequency band and can provide a degree of user-signal privacy by association of each user's signal with a randomly selected hopping pattern. One difference between the two spreading methods is that the DSSS technique uses the full system bandwidth throughout the entire transmission

time, whereas FHSS uses only a portion of the band at a time. In an FHSS/CDMA system, each user employs a different hopping pattern; in this system, interference occurs when two different users land on the same hop frequency. If the codes are random and independent of one another, the “hits” will occur with some calculable probability. If the codes are synchronized and the hopping patterns selected so that two users never hop to the same frequency at the same time, multiple-user interference is eliminated. The number of users in this case is limited by the number of frequency slots. This type of FHSS/CDMA scheme is equivalent to TDMA/FDMA with hopping of the center frequency, and it can be referred to as FH/TDMA.

10.2.1 FHSS in Frequency-Selective Fading

As is the case with DSSS transmission, FHSS transmission is also resistant to frequency-selective fading. However, performance analysis of FHSS systems in frequency-selective fading and the methods used to take advantage of in-band diversity are quite different from those for DSSS systems. Consider the wideband frequency response for a radio channel that exhibits frequency-selective fading: for example, as shown in Fig. 10.3. The multipath nature of the channel in the time domain accounts for the frequency-selective characteristic in the frequency domain. The depth of fading at certain frequencies can be as much as 30 to 40 dB below the average received signal power. As we saw in our discussion of statistical frequency-domain modeling, the rate of fluctuations in the frequency domain is proportional to the 3-dB width of the frequency correlation function, which is in turn inversely proportional to the rms multipath spread of the channel. In mathematical terms, fluctuations of the signal strength in the frequency domain follow the same patterns as fluctuations of the strength of a narrowband signal in the time domain.

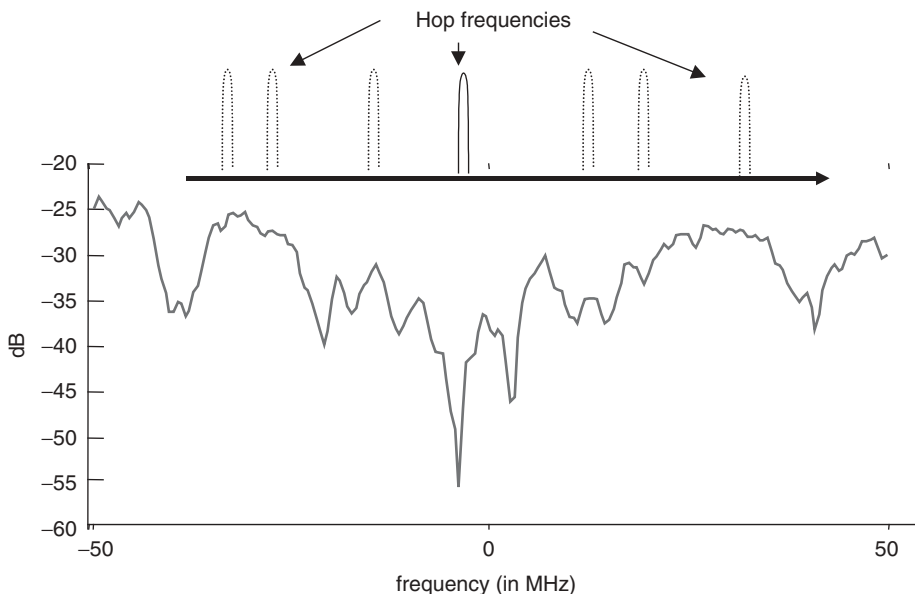


FIGURE 10.3 Frequency-selective fading and FHSS. Due to selective fading, the signal level in different hops change as much as 30 to 40 dB.

The relationship between the rate of fluctuations in the frequency domain and the rms multipath spread is the same as the relationship between the rate of variations in the time domain and the rms Doppler spread. In other words, there is a kind of *duality* between the time- and frequency-domain statistics of the channel response. Therefore, the Rice distribution for the number of fades crossing a certain threshold is also applicable to the number of frequency-selective fades crossing a certain threshold level, with rms multipath delay spread replacing the rms Doppler spread. Using the Rice fading distribution given by Eq. (4.3.2), with τ_{rms} replacing f_M , the average number of fades crossing a threshold A in a bandwidth W is given by

$$K = W\sqrt{2\pi}\tau_{\text{rms}}\rho e^{-\rho^2} \quad (10.2.1)$$

where $\rho = A/A_{\text{rms}}$ is the ratio of the threshold level to the rms value of the amplitudes. Similarly, we may use Eq. (4.3.3) to determine the average width of the fade in the frequency domain. For an FHSS system occupying a band of width W , if τ_{rms} is small enough so that $K < 1$, there is no deep notch in the operating band, and the channel is considered a flat-fading channel. For larger values of τ_{rms} , we have at least one deep fade in the operating band of the system, and the FHSS must operate in frequency-selective fading.

Let us now assume that there is only one deep null in the operating band of an FHSS system and that the width of the fade is such that it affects only one of the hop frequencies. The error rate of the modem is unacceptably high each time the signal hops to the faded frequency, but is extremely low for all the other hops. This situation is similar to that of the same system operating on a channel with a fixed narrowband interferer which disrupts only a single hop frequency. In both cases the SNR drops significantly on certain hops. Therefore, the performance analysis here would lead to the same equations as those used to analyze the FHSS system operating in narrowband interference, which is discussed later. Also, as in the case of narrowband interference, the FHSS system can be designed with appropriate coding and interleaving to correct the errors occurring on the hops into the frequency-selective fade. In an FHSS system, coding becomes more effective as the fading on neighboring hop frequencies becomes more independent, and this happens as the transmission bandwidth is increased. If the data rate of an FHSS system approaches the coherent bandwidth of the channel, a large number of hops will be affected by the frequency-selective fading; and even with very strong coding and long interleavers, error correction is inadequate and the output error rate becomes unacceptably high.

Example 10.4: FHSS in GSM In voice-oriented networks, there is often no retransmission mechanism. Corrupted packets are either discarded or are retained with a wrong value, in both cases causing distortions in the voice signal at the receiver. In TDMA systems, if the channel coincides with deep frequency-selective fading or when the co-channel interference (CCI) from another cell using the same frequency is excessive, distortion in the received voice signal will persist until the terminal moves adequately and the frequency-selective fading pattern is changed or the CCI is reduced. One method to reduce the duration of the frequency-selective fade or excessive CCI situations is to provide for a slow frequency-hopping pattern that forces a restriction on the duration of the frequency-selective fading or CCI effects. This option is exercised in the GSM system, which supports an optional frequency-hopping pattern of 217.6 hops/s on a frame-by-frame basis.

10.3 PRINCIPLES OF DIRECT-SEQUENCE SPREAD SPECTRUM

In DSSS transmission we may assume that the information signal is spread at base-band, and the spread signal is then modulated in a second stage. The received signal is first demodulated to recover the spread signal, and using a correlator the spread signal is despread to recover the original information signal. Figure 10.4 shows the four steps involved in this description. In this approach, the modulation technique is isolated from the spreading-and-despreading operation and we can discuss the base-band spreading and despreading separately. Figure 10.5 provides a simple example to describe spreading in a DSSS system. A square pulse with duration T_b represents a binary information digit in the time domain, and its Fourier transform is a sinc pulse with zero crossings spaced by $1/T_b$. The information signal is multiplied by a sequence of narrower pulses with time duration T_c and zero crossings spaced by $1/T_c$ to form the spread-spectrum signal. The narrow pulses are referred to as *chips*, and their amplitudes

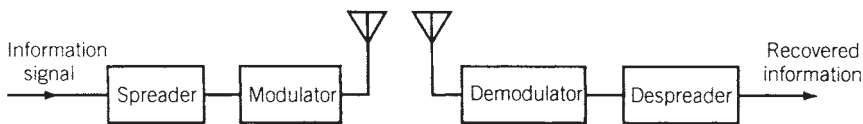


FIGURE 10.4 Simple block diagram of a direct-sequence spread-spectrum system.

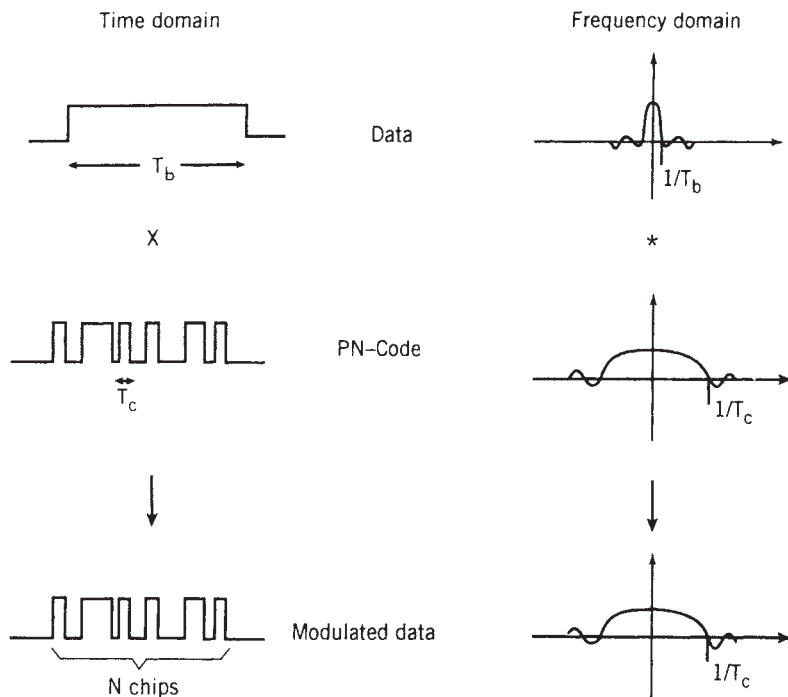


FIGURE 10.5 Simple examples to describe spreading and despreading in DSSS. If a TDMA system is used at the chip rate, we have N times more users. $R_b = 1/T_b$ (baud rate); $R_c = 1/T_c$ (chip rate).

are ± 1 .¹ The *bandwidth expansion factor* or *spreading factor* is $N = T_b/T_c$, the baud rate is $R_b = 1/T_b$, and the chip rate of the system is $R_c = 1/T_c$. Because the transmitted power is spread over a bandwidth N times wider than the information symbol rate, the spectral height of the signal is N times lower than it would be in nonspread transmission. The amplitudes of the chips are coded in a periodic random-appearing pattern referred to as the *spreading code*. Ideally, the spreading code is designed so that the chip amplitudes are statistically independent of one another. Because the chip sequence is coded to appear random, the sequences are referred to as *pseudorandom* or *pseudonoise* (PN) sequences or codes.

If we assume that the period of the spreading code is T_b , one period of the spreading signal is represented by

$$f(t) = \sum_{i=1}^N b_i p(t - iT_c) \quad (10.3.1)$$

where N is the number of chips per bit, $p(t)$ is the chip pulse shape, and $\{b_i\}$ are the values of the PN chips. The spreading signal is given by

$$s(t) = \sum_k f(t - kT_b) \quad (10.3.2)$$

The transmitted baseband signal is then given by

$$x(t) = \sum_n a_n f(t - nT_b) \quad (10.3.3)$$

where a_n is the information digit. The autocorrelation function of the transmitted signal and the duplicate of the periodic PN spreading signal is given by

$$R_{xs}(\tau) = \sum_n a_n R_{ff}(\tau - nT_b) \quad (10.3.4)$$

where

$$R_{ff}(\tau) = \int_{-\infty}^{\infty} f(t) f(t + \tau) dt \quad (10.3.5)$$

Figure 10.6 shows the autocorrelation function as given by Eq. (10.3.4). The correlation function is a periodic set of narrow pulses with width twice the chip duration, repeating at intervals of the bit duration T_b . If the transmitted bit is the symbol “1,” represented by a positive voltage value, the associated peak at the receiver is positive. If the transmitted symbol is the symbol “0,” represented by a negative voltage value, the peaks of the correlation function will be at a negative value. Therefore, the information bits represented by a constant voltage level with duration T_b at the input to the

¹Note that complex symbols with four phases are sometimes used for spreading. An example of such a spreading code used in the IEEE 802.11b is given in Example 10.25.

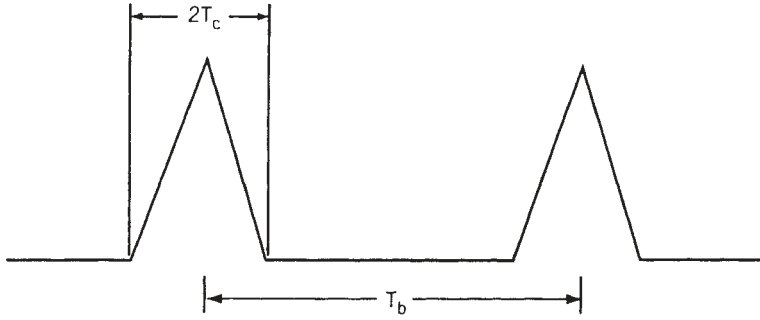


FIGURE 10.6 Autocorrelation function of the transmitted DSSS signal.

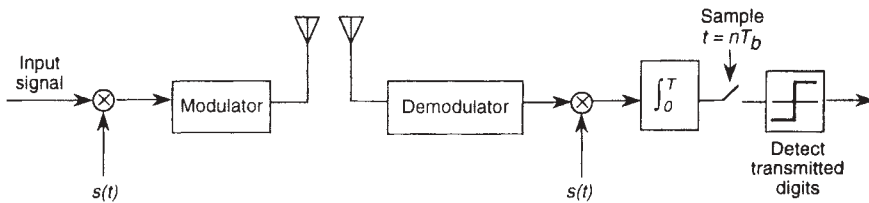


FIGURE 10.7 Block diagram of a DSSS transmitter and receiver.

transmitter are represented by narrow pulses with width twice the chip duration after cross-correlation at the receiver.

To implement a simple receiver, we need only generate the peak of the autocorrelation function. If it is a positive voltage, the transmitted information bit is a 1; if it is a negative voltage, the transmitted bit is a 0. The peak of the autocorrelation function occurs at $\tau = 0$; therefore, as shown in Fig. 10.7, we need only multiply the received signal by $s(t)$ and integrate it over one symbol interval. The sampled outputs of the integrator form the basis for the decision on the transmitted bit.

The structure of the spreading and despreading operations in Fig. 10.7 is similar to the structure of standard modulation and demodulation, but with a periodic PN spreading code replacing the usual carrier signal. The spread-spectrum operation can then be viewed as a two-layer modulation technique. The spread-spectrum modulation system multiplies the information signal by a random carrier (spreading signal), whereas the standard modulation multiplies the signal by a sinusoidal carrier. The standard demodulator correlates the received signal with a duplicate of the transmitted sinusoid, whereas the spread-spectrum demodulator (despreader) correlates the received signal with a duplicate of the random carrier. The difference between the two layers of modulation is that the generation and synchronization of the random carrier are more complicated processes than are required with the sinusoidal carrier. The modulated sinusoidal signal preserves the shape of the information signal spectrum and only shifts the spectrum to the carrier frequency. The spread-spectrum modulation spreads the spectrum of the information signal but does not shift its center frequency. The two-level modulation technique spreads the signal and shifts the center frequency.

Example 10.5: LFSR Codes in IS-95 The LFSR codes introduced in Example 10.1 are very popular spreading codes in DSSS systems. In the DSSS the output sequence

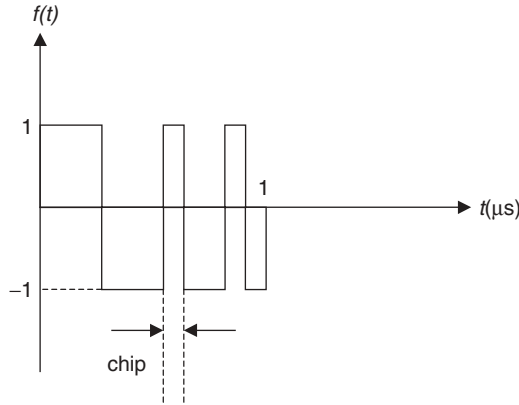


FIGURE 10.9 Eleven-chip Barker code used in the IEEE 802.11 DSSS standard.

entire T_b second. In this way we can claim that each chip actually carries a piece of the energy of a narrow pulse that is scrambled with a random code. The receiver knows the pattern (the spreading code) to put all these pieces together and construct the narrow pulse. This feature of the DSSS can be used to implement a pulse transmission system without actually transmitting a narrow pulse. The advantage of this approach over direct pulse transmission is that the DSSS approach avoids transmitting high-energy narrow pulses and instead, spreads the transmitted energy over a longer period of time. As we explained in Section 5.2.2, the DSSS pulse transmission techniques are very popular in measuring the multipath characteristics of the radio channels. The sliding correlator implementation of the DSSS correlator, described in Section 5.2.2, is also very popular in synchronization of DSSS modems, where we need to establish a timing reference between the transmitter and the receiver.

Another way of looking at DSSS modulation is to assume that it is a modulation technique whose pulse-shaping filter $f(t)$ is defined by Eq. (10.3.1). The transmitted baseband signal in this case is given by Eq. (10.3.3), which is the same as the transmitted signal for any baseband pulse transmission technique with a pulse-shaping filter given by $f(t)$. The difference between spread-spectrum and other modulation techniques is that the bandwidth of the pulse-shaping filter is approximately N times wider than with standard modulation with the same information symbol rate. As we saw in Chapter 7, the optimum receiver for the received pulses is a matched filter with impulse response $f(-\tau)$. The output of the matched filter is the cross-correlation given by Eq. (10.3.4) and shown in Fig. 10.6. The output of the correlator represents this function for all values of delay, whereas in other implementations discussed earlier we need generate only the peak of the correlation function.

Surface acoustic wave (SAW) devices are sometimes used for implementation of the impulse response $f(-\tau)$. The SAW device is a tapped delay line with taps set at the pseudonoise (PN) sequence values in the reverse order. In contrast with other implementations, this approach does not require code acquisition circuitry. Figure 10.10 shows a practical implementation of a SAW-based DSSS system using DBPSK modulation. Designs using SAW correlators are more expensive and less stable (due to temperature effects) than other designs, which makes them less attractive for application in commercial products. However, the simplicity of the SAW implementation has

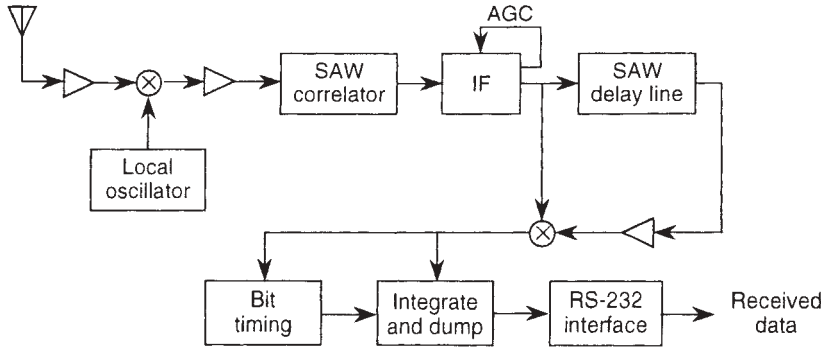


FIGURE 10.10 DSSS DBPSK receiver using SAW devices. (From [Fer80] © IEEE.)

made it an ideal candidate for use in experimental systems [Fer80, Kav87]. In commercial implementations the received signal is passed through a noise reduction filter followed by a sampler. The sampled digital signal is then processed for purposes of code tracking, timing recovery, phase recovery, and implementation of the correlator.

10.3.1 DSSS in Multipath Fading and the RAKE Receiver

Signal demodulation in a DSSS system is similar to adaptive MLSE or DFE in the sense that it takes advantage of the multipath to provide a form of in-band time diversity. We begin by considering the case of DSSS transmission on a single-path channel with no interference, for which the output of the receiver correlator was given by Eq (10.3.4). An example of the correlator output when the transmitted pulses are rectangular was shown in Fig. 10.6. The channel in this simple case is represented mathematically by an ideal impulse function. In a multipath environment the channel is represented by a linear summation of L multipath components, given by Eq. (5.2.3):

$$h(t) = \sum_{i=1}^L \beta_i \delta(t - \tau_i) e^{j\phi_i} \quad (10.3.6)$$

where $\beta_i(t)$, $\tau_i(t)$, and $\phi_i(t)$ represent the magnitude, arrival delay, and phase of the i th path in time. The output of the receiver correlator is then given by

$$R_{rs}(t) = \sum_n a_n \sum_{i=1}^L \beta_i R_{ff}(t - iT_b - \tau_i) e^{j\phi_i} \quad (10.3.7)$$

where a_n is the value of the n th information symbol, T_b the period of the spreading signal, and $R_{ff}(\tau)$ the autocorrelation function of the spreading signal, given in Eq. (10.3.5). Figure 10.11 illustrates the output of the correlator for the case of a three-path channel in a slowly time-varying channel in which multipath parameters does not change over a transmission of a few bits. By comparing this figure with Fig. 10.6, one can observe the effects of multipath on the receive correlator output. The system parameters in both figures are the same, but Fig. 10.6 represents a single-path channel,

whereas Fig. 10.11 represents a three-path multipath channel. Figure 10.12 shows several correlator output signals for the system shown in Fig. 10.10, in which the receiver correlator was implemented with a SAW device. In Fig. 10.12 one can see that there are up to five isolated paths in the channel, and the base of the correlation function is about 40 ns wide.

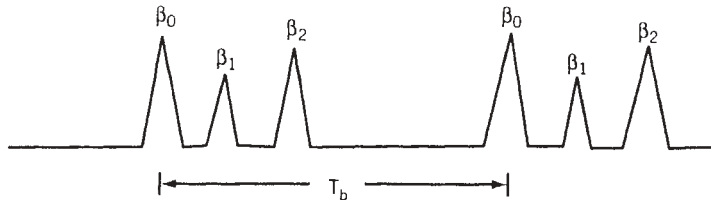


FIGURE 10.11 Output of the correlator with a three-path channel.

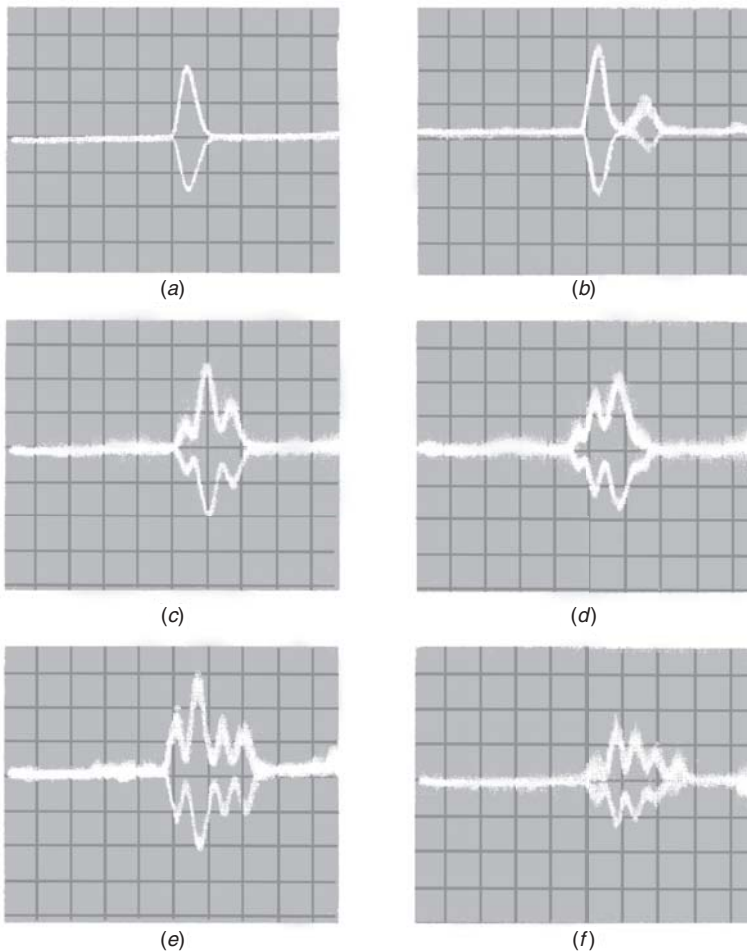


FIGURE 10.12 Sample correlated signal at the output of a system (shown in Fig. 10.10) that uses a SAW device correlator. (From [Fer80] © IEEE.)

For the example shown in Fig. 10.11, the interarrival delay among the multipath signals is greater than the width of the base of the autocorrelation function, and the delay spread is less than the information bit interval T_b . However, these conditions do not apply in every situation. For instance, if the delay spread of the channel is greater than T_b , this means that the data rate is greater than the coherence bandwidth of the channel, and we have ISI. To avoid interference between detected information symbols, the data rate should be kept well below the coherence bandwidth of the channel.² When consecutive signal paths arrive with delay differences greater than the chip duration T_c , the correlator output will exhibit separate peaks, as shown in Fig. 10.11. If the delay between two consecutive paths is significantly less than the chip duration, the two paths will merge and appear as one path equivalent to the phasor sum of the two actual paths. Thus, as the signal bandwidth is made narrower, the chip duration becomes correspondingly longer, and fewer isolated paths can be resolved at the correlator output. Of course, as paths merge, the fluctuations in their amplitudes and phases produce an overall fluctuation in the phasor sum, which we observe as fading.

Example 10.7: Multipath Reception in IS-95 DSSS The symbol transmission rate of the reverse channel of the IS-95 CDMA system is 4800 S/s, which provides for a 9600-bit/s offset QPSK transmission rate. The chip transmission rate of this standard is 1.25 megachips/s(Mc/s).

- (a) The base of the received pulses at the output of the correlator is $2T_c = 1/1.25$ Mc/s = 1600 ns, which is very large to resolve the indoor multipath conditions discussed in Chapter 5. The CDMA specifications considered for third-generation systems offer several times higher chip rates (wider bandwidths), which allows improved multipath resolution in indoor areas. In comparison, the chip rate of the IEEE 802.11 DSSS was 11 Mc/s, which resolves multipath spread on the order of 180 ns, which is much more suitable for indoor multipath conditions.
- (b) The distance between two peak of the correlator output at the IS-95 receivers is $T_b = 1/4800$ bits/s = 2.08 ms. As a result, channel multipath spreads on the order of 2 ms can not cause ISI in the system. Since the maximum multipath spread occurs in outdoor areas and never exceeds several tens of microseconds, this system does not suffer from an ISI problem.

If we operate with chip duration short enough to resolve individual paths, we can design a receiver to take advantage of multiple paths to provide diversity and enhance the reliability of the decision on each received information symbol. This sort of diversity is referred to as *implicit*, *internal*, *in-band*, or simply, *time diversity*. In a DSSS system, a receiver that combines the multipath components optimally as part of the decision process is referred to as a *RAKE receiver*. A typical RAKE receiver structure for a DSSS system is shown in Fig. 10.13. The received signal is passed through a tapped delay line, and the signal at each tap is passed through a correlator similar to the one used for standard DSSS receivers. The outputs of the correlators are then

²In a typical DSSS implementation, a scrambling code is used on top of the spreading code to randomize the ISI. The randomized ISI effects is then controlled by powerful coding.

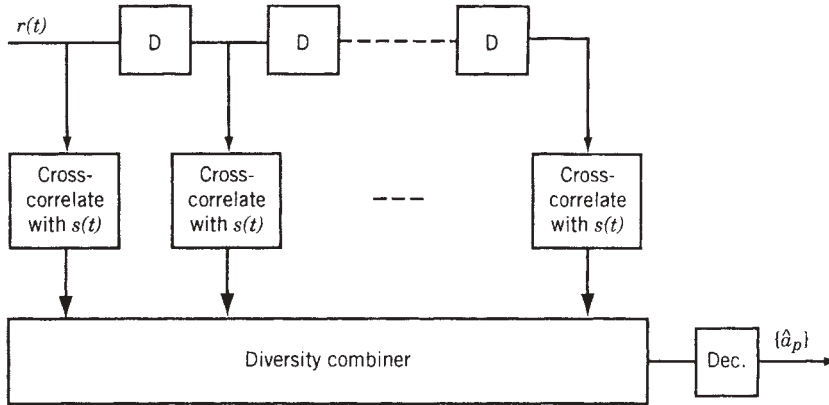


FIGURE 10.13 Typical RAKE receiver structure for a DSSS receiver.

brought together in a diversity combiner, whose output is the estimate of the transmitted information symbol.

In the original RAKE receiver [Pri58], the delay between the consecutive taps or “fingers” of the RAKE receiver was fixed at $W/2$ (where W is the chip bandwidth) to provide two samples of the overall correlation function for each chip period. Using this method for a rectangular chip pulse with triangular correlation function, we will have four samples of each triangle in the correlation function. Because the peaks are generally not aligned precisely at multiples of $T_c/2$, it is not possible to capture all the major peaks of the correlation function. But a good approximation of all major peaks will be provided by a RAKE receiver implemented with a sufficiently large number of taps. Modern digitally implemented receivers typically have only a few RAKE taps and the capability to adjust the tap locations. An algorithm is used to search for a few dominant peaks of the correlation function and then position the taps accordingly. If the receiver correlator is designed with a SAW device, there is no need for a tapped delay line. The output of the SAW device provides all values of the correlation function. As noted earlier, Fig. 10.12 shows a sample output of a SAW correlator receiver exhibiting several signal paths. Several preliminary developments of wireless DSSS systems have utilized SAW devices [Fer80]. However, SAW correlators have a stability problem, and it is difficult to integrate them into an ASIC-based modem design. For these reasons, as well as cost considerations, these devices are not widely used in available commercial wireless products.

A RAKE receiver can combine the arriving signal paths using any standard diversity combiner, such as a selective, equal-gain, square-law, or maximal ratio combiner (MRC). As we discussed in Chapter 8, the optimum diversity combiner is the MRC, which weighs the signal received from each branch by the signal amplitude at that branch. An MRC RAKE receiver is actually a discrete matched filter for a signal received on a multipath channel and thus provides optimum performance in the available bandwidth.³

³Note that this optimality is subject to a single-user environment with no interference and ideal autocorrelation codes with properties shown in Fig. 10.6.

Example 10.8: RAKE Reception in IS-95 The Qualcomm original receiver uses three moving fingers to implement the RAKE receiver. The original WaveLAN that was the model for implementation of the transmission system for the IEEE 802.11 standard did not use a RAKE receiver. Later, other manufacturers started implementation of RAKE receivers in wireless LANs to support more robust transmission.

In summary, in the operation of a DSSS system, multipath does not cause ISI unless the information symbol transmission rate approaches the coherence bandwidth of the channel. Also, it is possible to design a receiver that takes advantage of isolated arriving paths to improve or even optimize system performance. The wider the transmission bandwidth, the greater the number of isolated paths that can be resolved, and the greater the order of implicit diversity that can be utilized. The isolated paths will provide a source of implicit diversity to a DSSS receiver, which improves the performance of the system. On the other hand, if not utilized in some diversity-combining receiver, the signal arriving from each path is a wideband interference to signals arriving from other paths, which degrades the performance of the DSSS system.

10.3.2 Performance of the RAKE Receiver

Now we determine the received SNR for DSSS transmission on a frequency-selective multipath fading channel. The SNRs that we calculate apply to the peak of the correlation function and therefore represent the maximum achievable SNR at the output of a correlator. As shown by Eq. (10.3.7), the output of the receiver correlator includes a number of peaks, and the SNR is generally different for each peak. Each peak of the correlation function represents a branch of implicit or time diversity and is associated with signal arriving from one of the paths of the channel. Here we wish to determine the SNRs for all the diversity branches so that we can use them to analyze the performance of RAKE-type receivers. Because each diversity branch in effect represents a flat-fading channel, we begin by treating the SNR at each branch as a random variable.

We begin by determining the impact of multipath components as sources of interference to the signal received on any single path. Assume that we have L resolvable paths and that the amplitudes $\{\beta_i\}$ of the paths, shown in Eq. (10.3.6), are independent random variables. We denote the mean-squared value $E\{\beta_i^2\}$ of the magnitude of the i th path in the impulse response as $\overline{\beta_i^2}$. The received energy per bit for the i th branch is then given by $\overline{\beta_i^2} E_b$, where E_b is the transmitted signal energy per bit. Assuming that paths are uncorrelated, the average received power for the i th path is given by

$$P_{av,i} = \overline{\beta_i^2} E_b R_b$$

where R_b is the information bit rate. Assuming uncorrelated paths, the sum of interference power from the other $L - 1$ paths and the additive noise power before despreading at the receiver is

$$I_{av} = R_b E_b \sum_{i=1}^{L-1} \overline{\beta_i^2} + N_0 W$$

where N_0 is the one-sided background noise spectral density and W is the receiver bandwidth. After processing, this signal is multiplied by the despreading code and can

be approximated by Gaussian white noise with a spectral height of $I_0 = I_{av}/W$. Then the average received SNR per bit at i th branch is given by

$$\overline{\gamma}_i = \frac{\overline{\beta}_i^2 E_b}{I_0} = \frac{\overline{\beta}_i^2 E_b}{(R_b E_b / W) \sum_{i=1}^{L-1} \overline{\beta}_i^2 + N_0} = \frac{W}{R_b} \frac{P_{av, i}}{I_{av}} \quad (10.3.8)$$

If we assume the power to be the same on all branches, $\overline{\beta}_i^2 = \overline{\beta}^2$, we find the output average SNR per bit to be

$$\overline{\gamma}_i = \frac{\overline{\beta}^2 E_b}{[(L-1)/N] \overline{\beta}^2 E_b + N_0} \quad \text{all } i \quad (10.3.9)$$

where $N = W/R_b$ is the processing gain of the system. For an interference-dominated environment, the effect of the additive noise is negligible, so that $N_0 \simeq 0.1$, and we finally have

$$\overline{\gamma}_i \simeq \frac{N}{L-1} \quad (10.3.10)$$

Thus, we see that the received average SNR per bit is proportional to the processing gain and inversely proportional to the number of interfering signal paths. Note that for no multipath, $L = 1$, Eq. (10.3.10) gives the unrealistic result $\gamma_i = \infty$, due to the fact that we have neglected the effects of additive noise.

However, for the case $L = 2$, the equation yields $\gamma_i = N$, reflecting the fact that the impact of the equal-strength (0 dB) interferer is reduced by the system processing gain N . As we discussed earlier, a RAKE receiver can take advantage of the implicit diversity of the received multipath signal to improve performance significantly. Using a RAKE receiver, various methods for diversity combining can be applied to the implicit diversity just as they were applied to the explicit diversity provided by multiple antennas, as described in Section 8.3. If we assume that each path resolved by a RAKE receiver can be represented by a Rayleigh-distributed random variable, standard equations used for analyzing various diversity-combining methods can be applied directly to DSSS systems operating in frequency-selective fading. In carrying out the analysis, the implicit multipath diversity is incorporated into the order of diversity used in the calculation, and the SNR is adjusted to account for the multipath as an additional source of interference. Then, equations given in Section 7.4.2 can be used, with the equations given above for average SNR to predict the performance of a RAKE receiver.

If we assume that all paths have equal power and that the order of external diversity is 1 (single antenna), the average error rate of an MRC-RAKE receiver can be obtained by substituting Eq. (10.3.9) or (10.3.10) into Eq. (8.3.8). If the order of explicit diversity is D , the total order of diversity to be used in Eq. (8.3.8) is DL , which includes both implicit and explicit diversity. The SNRs, however, remain the same as given in Eq. (10.3.9), because we assume that only the paths arriving at the same antenna interfere with one another.

If the RAKE receiver uses selection combining, it searches among all arriving paths and selects the paths with the highest amplitudes. With D orders of explicit and L orders of implicit diversity, the average probability of error for a selective combiner

in Rayleigh fading and assuming equal average power in all branches is given by Eq. (8.3.10) as

$$\bar{P}_s = DL \sum_{k=1}^{DL} \binom{DL-2}{k-1} \frac{(-1)^{k-1}}{k} P\left(\frac{\bar{\gamma}_i}{k}\right) \quad (10.3.11)$$

where $P(\bar{\gamma}_i)$ is the expression providing the average error rate over the fading channel for the specific modulation technique used in the system, and $\bar{\gamma}_i$ is given by Eq. (10.3.9) or (10.3.10). In the simple design of a spread-spectrum wireless terminal using SAW correlators as the matched filter, shown in Fig. 10.10, the output of the correlator is integrated over all paths before the decision is made on the transmitted symbol. The integrator adds up all the path energies without any weighting factor. Therefore, this receiver is an equal-gain combiner of the implicit diversity components. Calculation of the average probability of error for equal-gain combiners is relatively difficult. Calculation of the average probability of error for the selective combiner and the MRC as discussed above can be used as lower and upper bounds, respectively, on the performance of the equal-gain combiner. The actual performance of the equal combiner is closer to the upper bound, that is, closer to the performance of the MRC.

Effects of Random Path Arrival. The methods discussed in this section for calculation of the average error rate are based on the assumption that the number of paths is fixed, the received power in all paths is the same, and the amplitude of each path is a Rayleigh-distributed random variable. As we saw in Chapter 4, the number of paths arriving at different locations is actually a random variable. Furthermore, the average received path powers are typically found to decay exponentially with increasing arrival delay, and received amplitudes are not necessarily Rayleigh. To address these issues properly, somewhat more complicated analysis is required.

If we assume that all paths have the same power but the number of paths is a random variable, the average probability of error is given by

$$\bar{P}_s = \sum_L \bar{P}_{b/L} P(L) \quad (10.3.12)$$

where $\bar{P}_{b/L}$ is the probability of error given that we have L equal power paths and $P(L)$ is the probability of having L paths in the profile. For a Poisson model of path arrivals, we have

$$P(L) = \frac{\lambda^L}{L!} e^{-\lambda} \quad (10.3.13)$$

where λ is the average number of arriving paths. If the values of the average received power from different paths are not equal, but the amplitudes are uncorrelated Rayleigh random variables, the average error rate for L paths is given by Eq. (8.3.9) with the average SNR per bit given by Eq. (10.3.8). If we assume that the path arrivals are random and that the received path powers are not equal, Eq. (10.3.12) cannot be used to average the error rate over a different number of paths as it was used when the average received power was assumed the same on all paths. For cases in which the path powers are unequal and the path arrivals are random, Eq. (10.3.12) is not applicable and other means must be found to determine the average error rate. A complete solution in such cases requires a Monte Carlo simulation to generate all the path arrival delays. From

a simulation of the arrival delays, the number of paths is determined and the power of each path $\overline{\beta_i^2}$ is calculated. The results are then used in Eqs. (10.3.8) and (8.3.9) to determine one value of the error rate. This simulation experiment is then repeated many times, and the resulting error rates are averaged over all the experiments to form an estimate of the overall average error rate.

Example 10.9: Performance of a RAKE Receiver To gain a better understanding of the equations above, we apply them to a typical example in which we have a set of profiles representing the channel. Saleh's model for indoor radio propagation [Sal87b] is useful for this purpose. As explained in Chapter 6, in this model the path arrivals are described by a Poisson process, the amplitudes are Rayleigh distributed, and the delay power spectrum is an exponential function. The average number of paths is $\overline{L} = \lambda T_m$, where λ is the arrival rate of the paths and T_m is the delay at which the exponential delay power spectrum drops below the SNR of the channel. The rms multipath delay spread of the channel is also determined easily from the shape and parameters of the delay power spectrum. The normalized average energy per path is $\beta^2 = 1/\overline{L}$, which is the normalized power divided by the average number of paths. Using these three parameters or using the results of simulations, one can use four different methods to determine the performance of a DSSS system.

Method 1 assumes that all paths have the same power and uses Eqs. (10.3.9) and (8.3.8) with $\overline{L}$ as the number of paths. Method 2 also assumes that the average power is the same for all paths and calculates the number of paths from $\hat{L} = |\overline{\tau}_{\text{rms}}/T_c| + 1$, where $|x|$ denotes the integer closest to x and $\overline{\tau}_{\text{rms}}$ is the average of τ_{rms} taken over all multipath profiles [Kav87, Pro89]. Method 2 also uses Eq. (8.3.8) for the error rate. Method 3 uses Eqs. (10.3.12) and (10.3.13) and assumes that the number of paths is a random variable but that the average power is the same for all paths. Equation (8.3.8) is used for the calculation of $P_{S/L}$, with Eq. (10.3.9) providing the value of $\overline{\gamma}_b$. Method 4 generates a set of channel profiles and calculates the error rate for each profile using Eq. (8.3.9), with Eq. (10.3.8) providing $\overline{\gamma}_b$. Figure 10.14 compares error rates calculated by these four methods.

All the calculations described in Example 10.9 are based on the assumption that the amplitude of each path is a Rayleigh-distributed random variable, and the arrivals of the paths are described by a Poisson process. As we saw in Chapters 4 and 6, neither of these assumptions is necessarily valid, and in most cases the empirical data indicate a lognormal amplitude distribution and modified Poisson arrival times. The most accurate way to predict performance is to take the results of measurements of the channel impulse response and calculate the error rate directly by generating a set of channel impulse responses. This method is investigated in [Cha93].

Effects of IPI and ISI. In the calculation of the SNR presented earlier in this section, we assume that the duration of the transmitted symbols are significantly larger than the multipath spread, so that there is no intersymbol interference (ISI) and the autocorrelation function of the received signal is ideal, resulting in zero sidelobes. These assumptions simplify calculation of the SNR significantly. We can use simple Eq. (10.3.9) or even simpler Eq. (10.3.10) as the SNR in average error-rate calculations for the discrete matched filter (DMF) for different diversity-combining techniques, introduced in Chapter 8, to calculate the performance of different RAKE receivers, as we did in Example 10.9. In practical situations, sidelobes of the autocorrelation function and

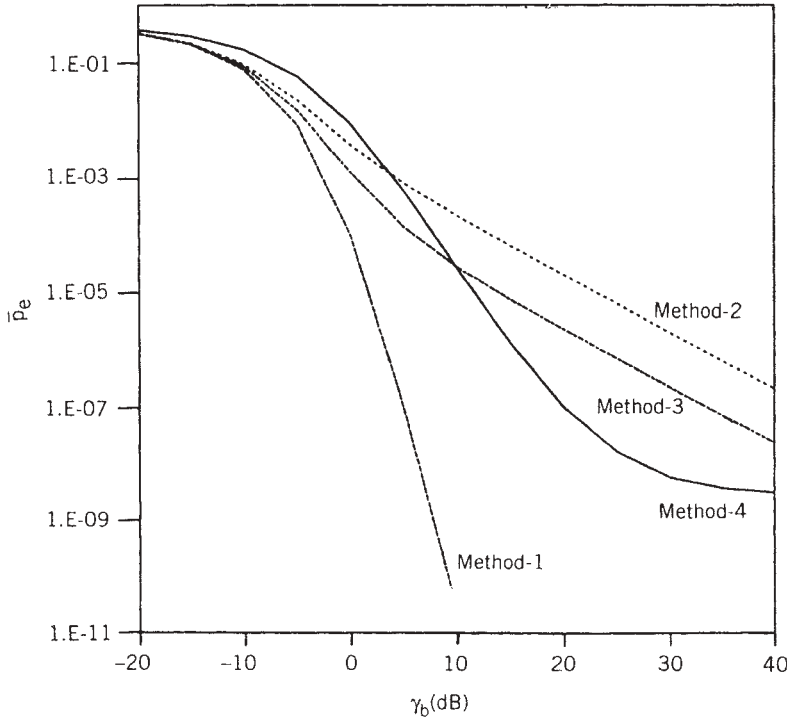


FIGURE 10.14 Comparison among the four methods for the calculation of the error rate: single user, threshold = -10 dB; bandwidth = 75 MHz; spreading code length = 255; order of external diversity = 1.

pulse-shaping filters have nonzero tails, causing ISI and additional interference among neighboring paths, which is referred to as *interpath interference* (IPI). Detailed calculation of the effects of ISI and IPI in the performance of RAKE receivers has attracted attention for the design of optimal RAKE receivers. A generalized description of IPI and ISI for discrete matched filter receivers is available in [Kaa99]. Explanation of the details of these calculations is beyond the scope of this book, and we proceed by describing their results in the following example.

Example 10.10: Performance in a Two-Path Channel In [Kaa99] the general expression for calculation of the IPI and ISI is used to calculate the SNR, which is then used for calculation of the maximum ratio combiner DMF described in Section 9.4. Similar to Fig 9.3, they assume a two-path channel and determine the performance as a function of the normalized distance between the two paths. To include the effects of fading, they assume that paths are exposed to Rayleigh fading. Since they address the general DMF problem, they can apply it to any arbitrary waveform. They have selected two waveforms, both with the same duration T_b . The first is a narrowband rectangular waveform and the second is a DSSS signal generated by a 15-chip m -sequence, similar to the one described in Example 10.5. To make the cases similar, one can assume that the first waveform is a DSSS with a spreading factor of 1. Figure 10.15 illustrates the average probability of error as a function of normalized distance between the two paths considering different noise components and an average received SNR of 20 dB.

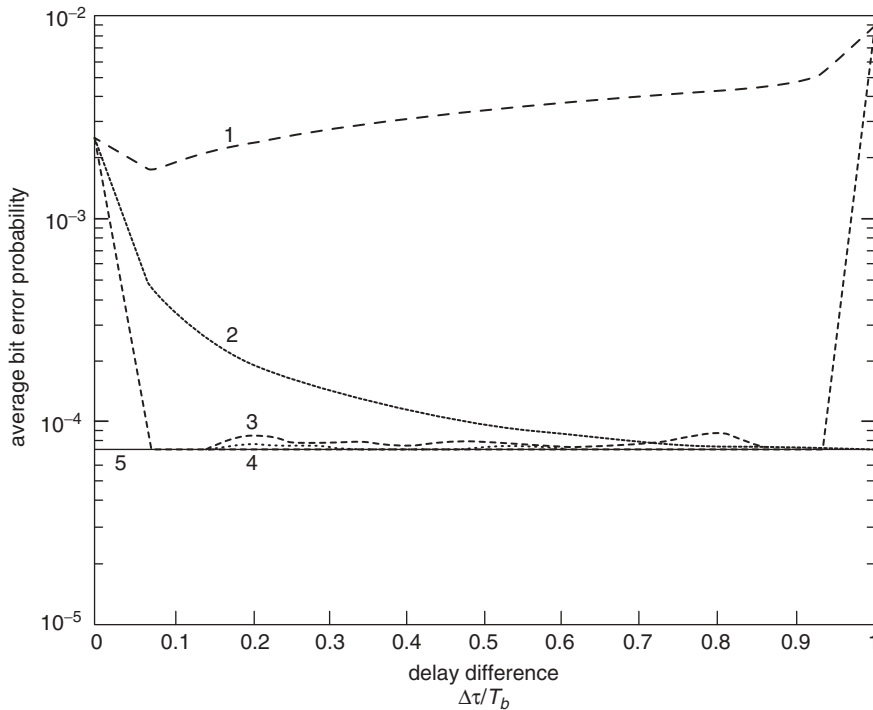


FIGURE 10.15 Effects of IPI and ISI in performance of a RAKE receiver.

Curves 1 and 2 correspond to the narrowband rectangular waveform, and curves 3 and 4 are for the 15-chip m -sequence. Curves 1 and 3 include both IPI and ISI as the source of noise. Curves 2 and 4 neglect the ISI and consider only the IPI. Curve 5 shows the ideal case in which there is no ISI or IPI. Obviously, the performance of the DSSS with RAKE receiver is much better. The interesting observation is that when the two paths are closer than the chip duration, performance is degraded significantly because the RAKE receiver cannot resolve the paths and take advantage of the in-band diversity. When the delay spread gets close to the bit duration (right-hand side of the figure), the second-arriving path of each symbol combines with the first-arriving path of the next symbol, causing significant degradation. In between the two ends of the curve, the overall performance of the RAKE receiver is close to the ideal performance neglecting the effects of ISI and IPI.

Example 10.11: Effects of the Spreading Factor Different applications in a multimedia network demand different data rates and error-rate requirements. A popular method to support different data rates in a DSSS system is to change the spreading factor while the chip rate remains the same, as the spreading factor decreases the data rate of the user increases. With this approach, the RF and signal-processing parts of the receiver remain the same, and the spreading factor is adjusted with simple digital operations. Figure 10.16 illustrates the effects of the spreading factor on the performance of a RAKE receiver over a two-path Rayleigh fading channel where the two

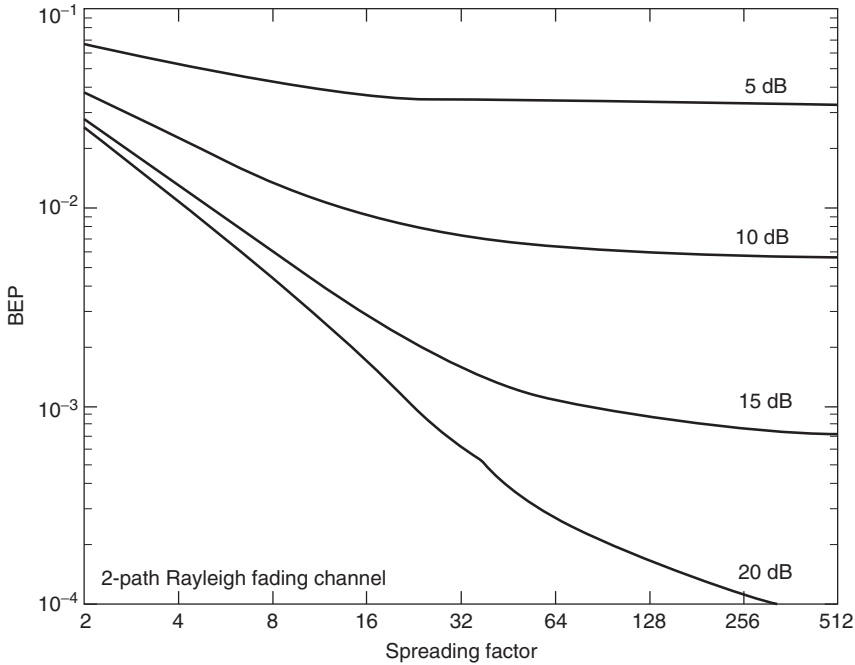


FIGURE 10.16 Performance of a RAKE receiver in a two-path Rayleigh fading channel for various spreading factors and SNRs [Lat98b].

paths are one chip apart. Since the two paths are close to one another, the effects of IPI are more significant than the ISI. As the spreading factor increases, IPI decreases and the performance is improved. Even for a 10-dB SNR, typically used for RAKE receivers, IPI causes considerable degradation.

In Example 10.11 we showed the performance in a two-path channel when both paths have equal power. In Example 10.12 we describe the effects of the number of paths as well as the effects of uneven path strengths.

Example 10.12: Effects of the Shape of the Delay Power Spectrum Figures 10.17 to 10.20 illustrates the effects of number of paths and the decay factor of the multipath profile. Figure 10.17 illustrates the average probability of error versus average SNR of a RAKE receiver operating in a channel with a uniform delay power spectrum for different number of paths. The performance in additive white Gaussian noise is also provided as a reference bound for the performance of the receiver. As the number of paths increases, the order of diversity increases and the performance improves. In ideal conditions, neglecting the effects of ISI and IPI, this performance should be the same as that of Fig. 8.6. Due to the ISI and IPI, the performance improvement for RAKE receiver is slower than an ideal MRC; however, with 10 paths the average probability of error of 10^{-6} is less than 4 dB worse than the performance in AWGN. Figure 10.18 represents the same type of plot for a delay power spectrum with 3-dB decay in each step. This time the performance with 10 taps is close to 8 dB away from the performance in AWGN. Figures 10.19 and 10.20 compare

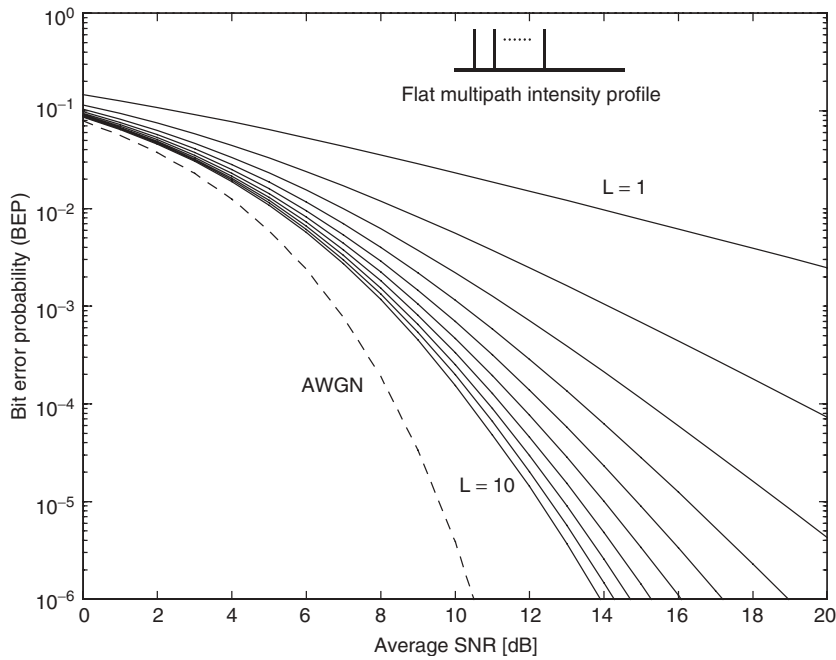


FIGURE 10.17 Average probability of error versus average SNR of a RAKE receiver operating in a channel with a uniform delay power spectrum for different numbers of paths. (Courtesy of Matti Latva-Aho of the University of Oulu, Finland.)

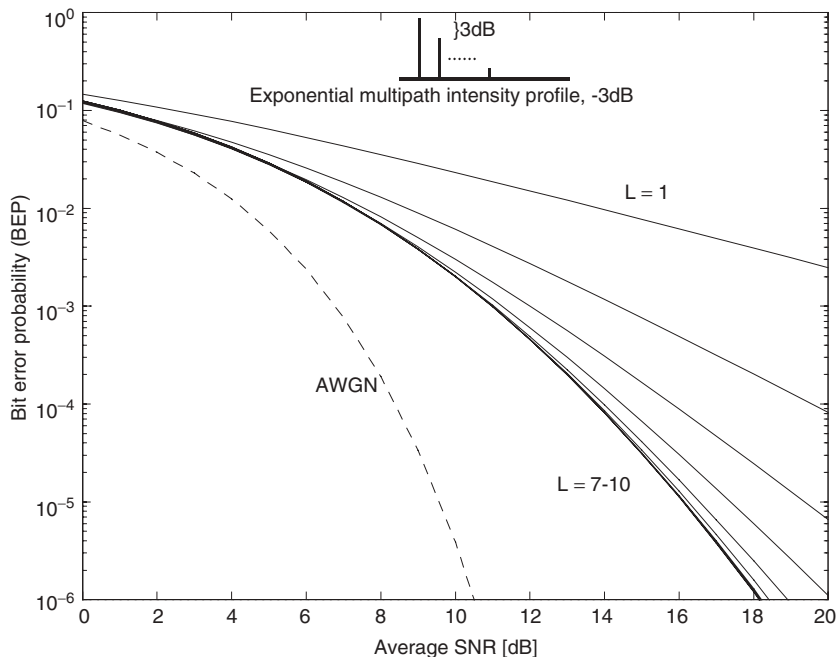


FIGURE 10.18 Average probability of error versus average SNR of a RAKE receiver operating in a channel with an exponential delay power spectrum for different numbers of paths. (Courtesy of Matti Latva-Aho of the University of Oulu, Finland.)

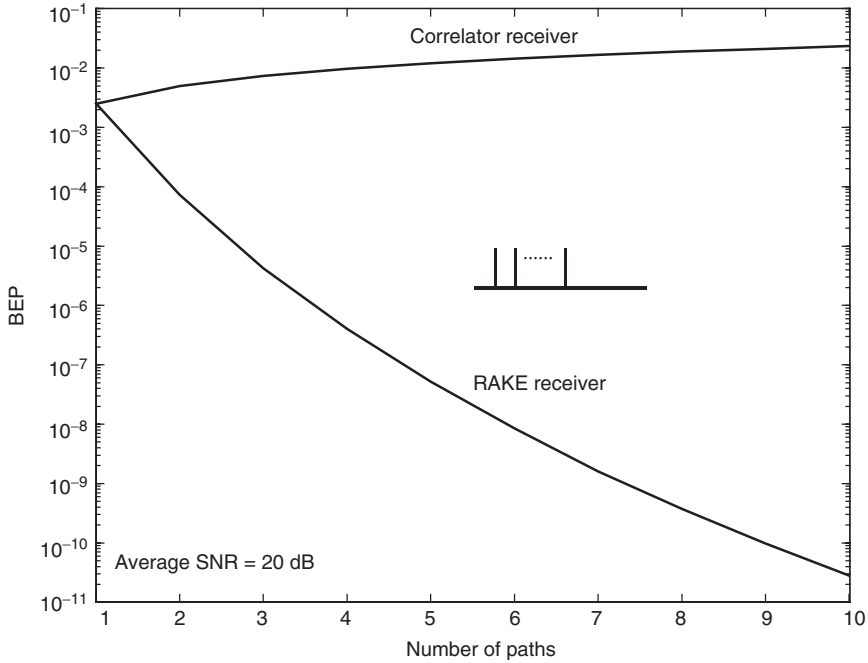


FIGURE 10.19 Performance of a correlator versus RAKE receiver in a uniform multipath spread environment with $\text{SNR} = 20$ dB. (Courtesy of Matti Latva-Aho of the University of Oulu, Finland.)

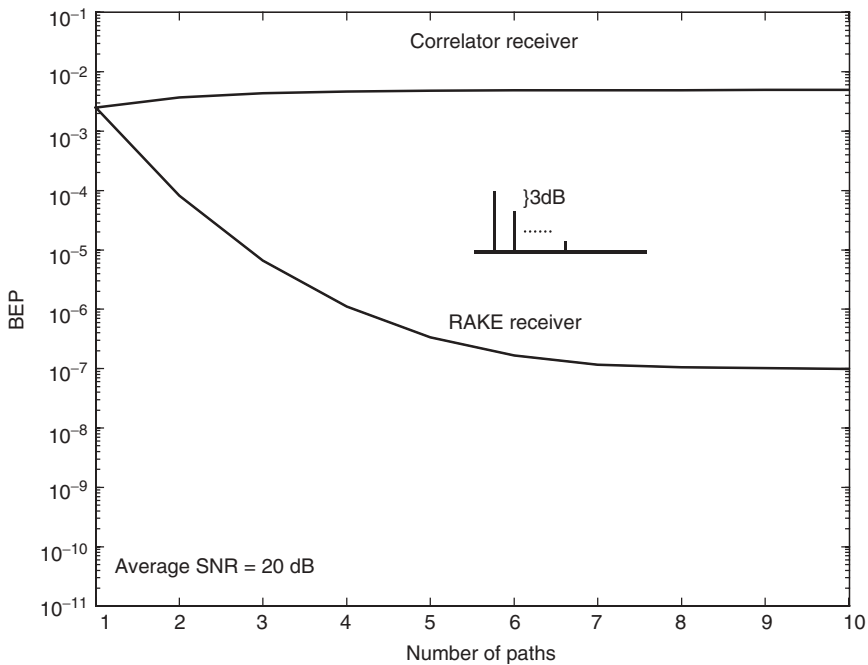


FIGURE 10.20 Performance of correlator versus RAKE receiver in an exponential multipath spread environment with $\text{SNR} = 20$ dB. (Courtesy of Matti Latva-Aho of the University of Oulu, Finland.)

the performance of a simple correlator with the RAKE receiver with $\text{SNR} = 20$ dB for the two delay power spectrums. As the number of paths increases, the performance of the correlator degrades, due to the interferences. The RAKE receiver takes advantage of the in-band diversity to improve performance when the number of paths is increased.

Example 10.13: Performance Evaluation Using Ray-Tracing Results Similar to Section 9.8, we can use channel impulse responses obtained by ray tracing for performance evaluation of spread-spectrum systems. Figure 10.21 compares the minimum power requirement and maximum data rates achievable in a 10-MHz band for four different spread-spectrum systems using BPSK modulation [Fal96]. The first pair of columns of this figure is a DSSS system with a four-tap RAKE receiver, using a 15-chip M -sequence, and the following pairs represent a 31-chip DSSS with a four-tap RAKE, and a 15- and a 31-hop FHSS. The maximum data rate is the bandwidth divided by the code length: that is, 1.33 Mb/s for code length 15 and 0.66 Mb/s for code length 31. For specific DSSS or FHSS systems, the power requirement is almost 6 dB (fourfold) less when we almost double (31/15) the processing gain or spreading factor. Therefore, as expected, either in additive noise or multipath channels, the power requirement is inversely proportional to the square of the processing gain. This is due to the fact the processing gain is a measure of the amplitude of the detected signal after processing, and the power is always related to the square of the amplitude. When we compare DSSS and FHSS with the same processing gain, due to the time-diversity combining ability of RAKE receivers, the DSSS systems perform close to 4 dB better than the FHSS systems.

If we were able to code the bits over different hops, so that the errors occurred in a deep fade would be corrected from the correct data obtained from other hops, the frequency diversity of the received signal, similar to coded OFDM, would enjoy the same improvements as DSSS systems using RAKE receivers. To take advantage of frequency diversity, as we discussed in Section 10.2, IEEE 802.11 or Bluetooth systems using FHSS send one packet per hop, so that the corrupted packets can be transmitted in a different hop frequency.

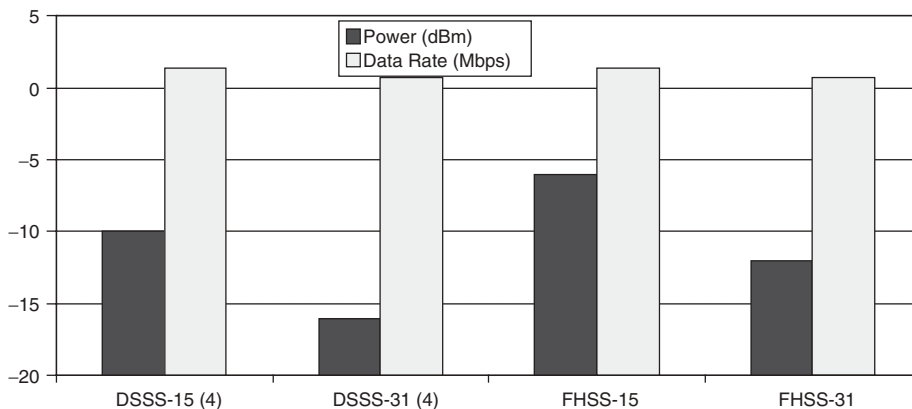


FIGURE 10.21 Performance of spread spectrum modems in a bandlimited channel with a fixed bandwidth of 10 MHz.

Effects of Channel Measurement Error. As shown in Section 9.4, implementation of the optimum maximum ratio combiner RAKE receiver requires a channel estimator. In Section 9.4 we described how a channel estimator is implemented, and Fig. 9.9 shows how the results of channel estimation are used for implementation of a DMF. The RAKE receiver is a special implementation of the DMF that takes advantage of the DSSS technology; therefore, maximum ratio combiner design needs a channel estimator similar to those described in Section 9.4. Channel estimators work based on known reference signals, and a practical issue in implementation of a channel estimator is how to provide this reference signal. Example 10.14 provides more insight in this issue.

Example 10.14: Training Sequence One of the earliest analyses of the impacts of various training strategies on the performance of a RAKE receiver is available in [Pah79, Pah80]. Three techniques to train an adaptive channel estimator are examined

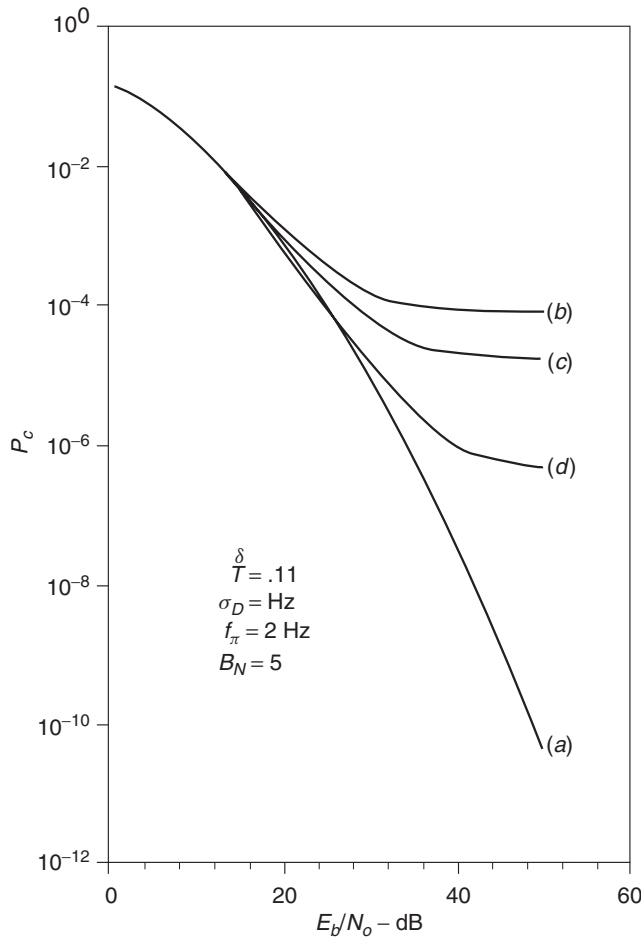


FIGURE 10.22 Performance of the RAKE receiver with various channel measurement techniques: (a) ideal; (b) serial probe; (c) parallel probe; (d) decision directed.

in this work: (1) a reference signal is added as a DSSS parallel probe signal to the transmitted signal, (2) a reference is time-multiplexed with the transmitted signal as a serial probe, and (3) the estimator is used in a decision-directed mode, where detected symbols are used as the reference signal for channel measurement. Figure 10.22 compares the performance of an ideal RAKE receiver with a receiver using a serial probe, a parallel probe, or a decision-directed channel estimator. The parameters of the channel are similar to those described in Example 9.1.

Any method used for channel estimation will produce a measurement error, which will restrict the performance of a RAKE receiver. Similar to the ISI, the measurement error is proportional to the received signal strength, and as we increase the transmitted power, these noise levels will also increase. When we increase the signal-to-background noise ratio, at a certain point the measurement noise becomes more than the background noise and the SNR; consequently, the probability of error remains flat. These flat error rates, caused by noises that are proportional to the signal level, are sometimes referred to as *irreducible* error rates.

10.4 INTERFERENCE IN SPREAD-SPECTRUM SYSTEMS

Analysis of the effects of interference in spread-spectrum systems is similar to analysis of the effects of additive noise. However, interference and additive noise have differences that affect the results of the analysis. The additive noise considered in communication systems is the thermal noise generated by the random motion of electrons in the resistive elements of the receiver circuit. The spectral height of this noise is constant and its bandwidth is unlimited. As we increase the transmission bandwidth of the system, the thermal noise power added to the received signal increases accordingly. However, the source of interference is another transmitter operating in the band of our system. For the sake of this discussion, let us consider the interfering transmitter as having fixed transmission power. In general, the bandwidth of the interfering signal is less than or at most equal to the bandwidth of the desired signal, and therefore the interference power is independent of the transmission bandwidth of our system. The spectrum of the interfering signal need not conform to any particular shape. The bandwidth of the interfering signal might be as narrow as the bandwidth of a steady sinusoid generated by an intentional or unintentional source up to a bandwidth as wide as the transmission bandwidth of a similar wideband system coexisting with the intended user in a CDMA environment.

The literature on military communication systems offers many detailed analyses of the performance of spread-spectrum systems in the presence of various intentional interferers or jammers [Sim85]. These jammers are designed to disrupt the operation of a spread-spectrum system, and they can employ relatively sophisticated techniques, such as multitone jamming and pulsed jamming. In civilian applications, the interference is neither intentional nor sophisticated. Most often, the interferer is simply another system designed to operate in a portion of, or in the entire band of, operation of our system, and users are generally willing to cooperate so as to minimize the mutual interference. To evaluate the effects of interference, we consider narrowband and wideband interference and compare the results with the performance in additive noise. The wideband interferer in commercial applications is another spread-spectrum user sharing the same operating band. Narrowband interference usually occurs in an

overlay situation where a wideband spread-spectrum system is to coexist with a group of narrowband systems operating in the same band.

10.4.1 Interference in FHSS Systems

The spectrum of additive thermal noise is flat and covers all the hop frequencies with the same spectral height. As a result, the received SNR for a FHSS modem on a channel corrupted only by additive noise is the same at every hop. This SNR is given by

$$\gamma_b = \frac{E_b}{N_0} \quad (10.4.1)$$

where N_0 is the height of the two-sided noise power spectral density and E_b is the energy per bit:

$$E_b = T_b P_{av} = \frac{P_{av}}{R_b} \quad (10.4.2)$$

with P_{av} the average received power, T_b the bit duration, and R_b the bit rate of the system. The error rate of the FHSS system in additive Gaussian noise remains the same as the error rate of the same system without frequency hopping and is computed from the BER equations for standard modems given in Chapter 7.

Although it is true that frequency hopping does not improve BER performance in additive noise, we shall show that it protects the system, in somewhat different ways, against narrowband and wideband interference. A narrowband interferer interferes with the signal at only one hop, and the signal at all other hops remains unaffected, regardless of the power level of the interferer. For the wideband interferer the power of the interference is spread over the entire band, and only a portion of this power interferes with the desired signal at each hop.

For a simple mathematical representation of narrowband and wideband interference, consider Fig. 10.23. Both narrowband and wideband interference are assumed to have an ideal rectangular power spectrum. The bandwidth of the wideband interference is assumed to be the same as the transmission bandwidth W , and its height is represented by I_0 . The bandwidth W_1 of the narrowband interferer is assumed to be narrower than the bandwidth at each hop; that is, $W_1 \leq W/N$, where N is the bandwidth expansion factor of the FHSS system. The power spectral height of the wideband interference is represented by I_n . The total received interference power is given by

$$I_{av} = I_n W_1 + I_0 W \quad (10.4.3)$$

and is assumed to be the same for both narrowband and wideband interference cases.

Effects of Wideband Interference. Within the transmission band, the power spectrum of wideband interference is the same as the power spectrum of additive noise having spectral height I_0 . Therefore, the performance of the system in wideband interference is the same as the performance on an additive noise channel with SNR per bit given by

$$\gamma_{b,WI} = \frac{E_b}{I_0} = \frac{W}{R_b} \frac{P_{av}}{I_{av}} \quad (10.4.4)$$

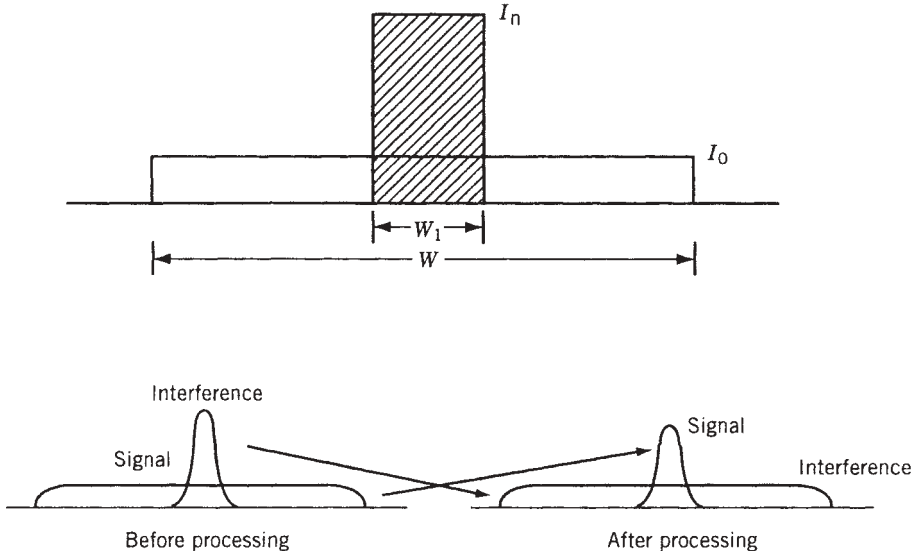


FIGURE 10.23 Simple mathematical representation of narrowband and wideband interference.

With a fixed ratio of signal to interference power, P_{av}/I_{av} , the SNR γ_b used for calculation of the error rate is $N = W/R_b$ times higher than the actual ratio of the signal power to interference power. If the occupied bandwidth at each hop is the same as the bit rate R_b , then N is the number of hops in the transmission band or the bandwidth expansion factor and is referred to as the *processing gain*. With a fixed interference power, as the transmission bandwidth increases, the spectral height of a uniformly distributed wideband interferer decreases. Reduction of the spectral height of the wideband interference, in turn, increases the received signal to interference ratio at each hop. Equation (10.4.4) shows that by adjusting the number of hops, we may design an FHSS system to overlay onto an existing wideband system. The existing system is viewed as a wideband interferer to the intended FHSS system, and the FHSS system will be a narrowband interferer to the existing system. The possibility of overlay is one of the attractive features of the spread-spectrum technique for commercial applications.

Example 10.15: IEEE 802.11b Interference with Bluetooth Devices Figure 10.24a shows the spectrum of overlapping IEEE 802.11b and Bluetooth channels, both operating in the 2.4-GHz ISM bands. IEEE 802.11b has three nonoverlapping bands, each $W = 26$ MHz wide. Bluetooth is an FHSS system that chooses one of the 78 nonoverlapping $R_b = 1$ Mb/s channels in the same band. Figure 10.24b shows the interference scenario in which at terminal BT-1 the desired received signal, P_{av} , from the BT-2 is interfered with the interference signal from the IEEE 802.11b terminal, I_{av} . Considering the transmitted power from the devices and the discussions in Section 4.2.1, we have $P_{av} = K P_{BT} r^{-\alpha}$ and $I_{av} = K P_b D^{-\alpha}$, and using Eq. (10.4.4) we have

$$\gamma_{b,WI} = \frac{W}{R_b} \frac{K P_{BT} r^{-\alpha}}{K P_b D^{-\alpha}} = \frac{W}{R_b} \frac{P_{BT}}{P_b} \left(\frac{D}{r} \right)^{\alpha} = 0.26 \left(\frac{D}{r} \right)^{\alpha}$$

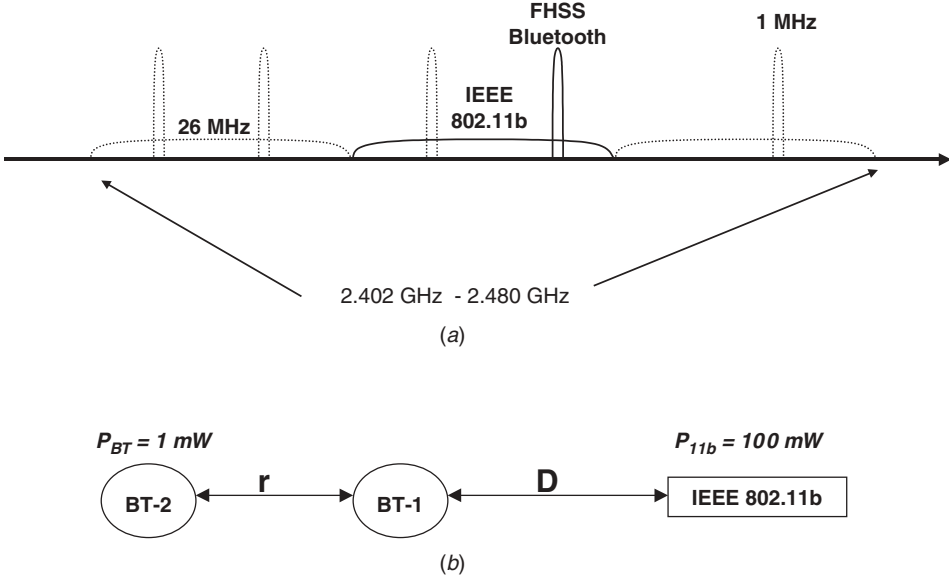


FIGURE 10.24 (a) Overlapping IEEE 802.11b and FHSS Bluetooth spectrum; (b) operation scenario of the terminals.

Using Fig. 7.11 for noncoherent FSK modulation, used in Bluetooth, we need a 10-dB signal-to-interference ratio for a bit error rate of around 1%. Therefore, in free space with a distance–power gradient of $\alpha = 2$, if the 802.11 device comes closer than $D/r = 6.2$ times the distance between the two Bluetooth devices, it will interfere with proper communication between those devices.

If we rewrite Eq. (10.4.4) in the form

$$\frac{I_{av}}{P_{av}} = \frac{W}{R_b} \frac{1}{E_b/I_0} \quad (10.4.5)$$

the value of I_{av}/P_{av} is referred to as the *interference margin*. The interference margin in decibels represents the maximum allowable power difference between the interference and the desired signal to maintain proper operation. Proper operation is identified with a minimum required E_b/I_0 value needed to deliver data with the maximum allowable error rate.

Effects of Narrowband Interference. For analysis of FHSS systems with narrowband interference, we assume that the interferer interferes with the signal at only one hop and that its spectrum is flat over the hop bandwidth. Therefore, the interference bandwidth for a one-bit-per-hertz modulation technique is given by $W_1 = W/N = R_b$. As shown in Fig. 10.23, the spectral height of the narrowband interference is $I_n = I_{av}/W_1$. Energy per bit, given by Eq. (10.4.2), is $E_b = P_{av}/R_b$. As a result, the SNR per bit for the hop that hits the narrowband interferer is given by

$$\gamma_{b,NI} = \frac{E_b}{I_n} = \frac{W_1}{R_b} \frac{P_{av}}{I_{av}} = \frac{P_{av}}{I_{av}} \quad (10.4.6)$$

which is the same as the SNR without frequency hopping. This means that we have no protection when the hop hits the interferer. Because there is assumed to be no interference for all other hops, the average signal-to-interference ratio over all hops is

$$\overline{\gamma}_{b,NI} = \frac{NE_b}{I_n} = \frac{NW_1 E_b}{I_{av}} = \frac{WE_b}{I_{av}} = \frac{E_b}{I_0} = \frac{W}{R} \frac{P_{av}}{I_{av}} \quad (10.4.7)$$

which is the same as the SNR in the case of wideband interference, as given by Eq. (10.4.4). Note that this does not mean that if the power of the narrowband and wideband interferences is the same, the average error rate is the same. We can substitute Eq. (10.4.4) into the formula for the probability of error of the modulation technique and determine the error rate of the FHSS system in wideband interference. However, we cannot substitute Eq. (10.4.7) into the same error rate formula to calculate the average error rate in the presence of narrowband interference. Instead, we should substitute Eq. (10.4.6) into the error-rate formula to obtain the error rate in the hop hit by the interferer. Since there is no error in other hops, the average error rate is N times less than this value. The following example will serve to clarify the situation.

Example 10.16: Error Rate in Narrowband Interference Assume that we have an FH/BPSK modem transmitting 10 hops, operating in an interference environment. Further assume that the effects of additive noise are negligible, the received interference power is at the same level as the received power of the desired signal, and the interference bandwidth is sufficiently narrow that it interferes with only one hop frequency. At this hop frequency the signal-to-interference ratio is $\gamma_b = 1$ and the BPSK error rate, given by $0.5 \operatorname{erfc}(\gamma_{b,NI})$, is approximately 10^{-1} . Other hops are disturbed only by additive noise, with negligible effect. The error rate at these hops is essentially zero. From the user's perspective, the average error rate over the 10 hops is then approximately 10^{-2} . Now consider a case in which the interferer is wideband with the same total power but with its power distributed uniformly over all hops. Here the interference power at each hop is 10 times smaller, but all hops are affected by the interference. Therefore, the SNR per bit at each hop is $\gamma_{b,WI} = 10$ dB, which produces an error rate of 10^{-5} in all hops, three orders of magnitude better than the average error rate over all hops in the presence of the narrowband interferer with the same power. In terms of signal power, the three-order-of-magnitude improvement in error rate of a BPSK-modulated signal accounts for approximately 5 dB in SNR. Thus, the impact on performance with narrowband interference in this example is 5 dB greater than with wideband interference.

In general, the average probability of error over all hops for a fixed narrowband interferer affecting only one hop is given by

$$P_{s,FHNI} = \frac{P_s(\gamma_{b,NI})}{N} \quad (10.4.8)$$

where $\gamma_{b,NI}$ is the signal-to-interference ratio in the hop that hits the narrowband interferer and is given by Eq. (10.4.6), and $P_s(\gamma_b)$ is the error-rate equation for the particular modulation technique used for the FHSS system. In Eq. (10.4.8), the probability of error is inversely proportional to the processing gain N . This equation should be compared

with the probability of error for wideband interference covering the entire band, which is given by

$$P_{s,\text{FWI}} = P_e(N\gamma_{b,\text{NI}}) = P_s(\gamma_{b,\text{WI}}) \quad (10.4.9)$$

In this equation the probability of error is an exponential function of the processing gain. Figure 10.25 shows the error rate as a function of N . The probability of error for narrowband interference is bounded by $P_{s,\text{FHNI}} = 1/(2N)$, which corresponds to the assumption that the information in the hop hitting the interferer is completely lost and $P_s = 0.5$ for that hop. If the interference is another FHSS modem signal with a random hopping pattern, the average fraction of hits remains at $1/N$. Although the FHSS interferer occupies the same band as the modem transmitting the desired signal, when the interference occurs it hits only one hop. Therefore, if the hopping patterns of the signal and interference are independent, the error rate for the desired signal is the same as the error rate in the case of narrowband interference, as given by Eq. (10.4.8). If the hopping pattern in the interfering system is in lock step with the pattern of the desired signal, the FHSS signal provides no protection and the processing gain is 0 dB. However, this is an unlikely scenario in commercial applications, where the interference is unintentional.

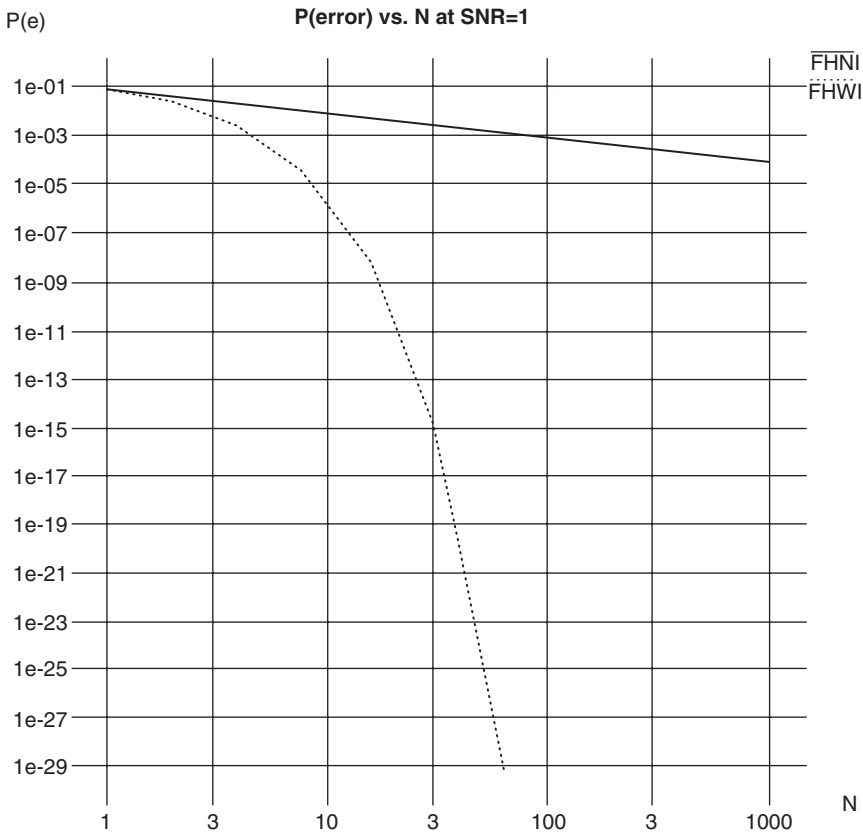


FIGURE 10.25 Error rate of FHSS-NI and FHSS-WI as a function of N .

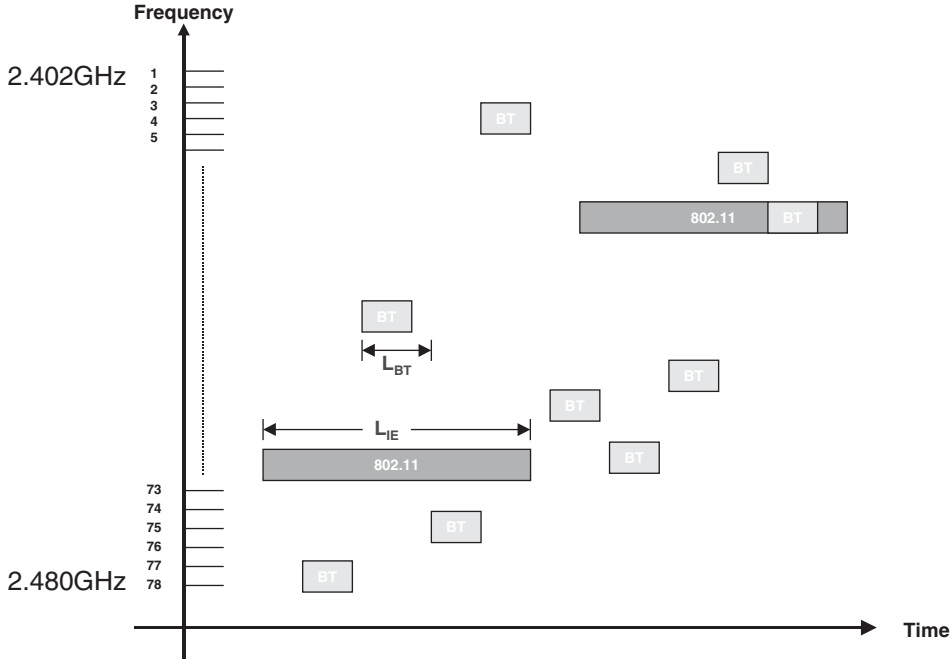


FIGURE 10.26 Time-frequency characteristics of the FHSS IEEE 802.11 and Bluetooth.

Example 10.17: Packet Collision Between 802.11 and Bluetooth Figure 10.26 illustrates packet transmission time-frequency behavior for the Bluetooth and IEEE 802.11 FHSS. Both devices use FHSS with a 1-MHz bandwidth in the 79 nonoverlapping channels in the ISM 2.4-GHz band. Bluetooth packets are shorter than 802.11 packets; during transmission of one 802.11 packet, the interfering Bluetooth device hops and sends one packet per hop several times. Assuming that L_{IE} is the length of the 802.11 packet and L_{BT} the length of a Bluetooth slot, the minimum number of Bluetooth hops occurring during transmission of one 802.11 packet is $n = \lceil L_{IE}/L_{BT} \rceil$, where $\lceil x \rceil$ represents the smallest integer greater than or equal to x . The maximum number of Bluetooth hops occurring in the duration of an 802.11 packet is $\lceil L_{IE}/L_{BT} \rceil + 1$. As shown in [Enn98], the probability of an 802.11 packet overlap with $n = \lceil L_{IE}/L_{BT} \rceil$ Bluetooth dwell periods of duration L_{BS} is

$$P_n = \frac{L_{IE}}{L_{BT}} - \left\lceil \frac{L_{IE}}{L_{BT}} \right\rceil$$

The probability that it overlaps with $n + 1 = \lceil L_{IE}/L_{BT} \rceil + 1$ dwell periods is

$$P_{n+1} = 1 - \frac{L_{IE}}{L_{BT}} + \left\lceil \frac{L_{IE}}{L_{BT}} \right\rceil$$

Considering these expressions, the probability of an 802.11 packet surviving Bluetooth interference, P_{survive} , is approximated by

$$P_{\text{survive}} = (1 - P_{\text{hit}})^n P_n + (1 - P_{\text{hit}})^{n+1} P_{n+1}$$

where P_{hit} is the probability of having the same frequency for both 802.11 and Bluetooth. The probability of collision is given by $P_{\text{collision}} = 1 - P_{\text{survive}}$. For a 1000-byte 802.11 packet at 2 Mb/s,

$$L_{\text{IE}} = \frac{1000(\text{bytes}) \times 8(\text{bits/byte})}{2(\text{Mb/s})} = 4 \text{ ms}$$

If Bluetooth is sending one-slot packets $L_{\text{BS}} = 625 \text{ } \mu\text{s}$. Therefore,

$$n = \left\lceil \frac{4 \text{ ms}}{625 \text{ } \mu\text{s}} \right\rceil = 6$$

and $P_n = 0.4$, which results in $P_{n+1} = 0.6$. Since the probability that a Bluetooth hop will occur at the operating frequency of the 802.11 FHSS system is $P_{\text{hit}} = 1/78 = 0.013$, we have

$$P_{\text{survive}} = (1 - 0.013)^6(0.4) + (1 - 0.013)^7(0.6) = 0.92$$

Therefore, the collision probability is 0.08, or 8%. Therefore, 802.11 FHSS and Bluetooth in close proximity to each other will communicate with some reduction in the throughput. However, for 802.11b and Bluetooth in close vicinity $P_{\text{hit}} = \frac{1}{3} = 0.33$, the probability of collision increases significantly (close to 50%), so that it virtually stops packet data communications.

10.4.2 Interference in DSSS Systems

Analysis of the effects of interference on the performance of a DSSS system can be done in a more systematic manner. Using the representation of the transmitted spread-spectrum signal given by Eq. (10.3.3), the received signal in a channel disturbed only by interference is given by

$$r(t) = \sum_n a_n f(t - nT_b) + i(t) \quad (10.4.10)$$

where a_n is the n th information symbol, $f(t)$ defined by Eq. (10.3.1) is the waveform representing one period of the spreading pattern of the signal, and $i(t)$ is the interference signal. After correlation at the receiver, we have

$$R_{rs}(t) = \sum_n a_n R_{ff}(t - nT_b) + z(t) \quad (10.4.11)$$

where $R_{ff}(t)$ is the autocorrelation function of $f(t)$, defined by Eq. (10.3.5).

The signal-to-interference ratio per bit at the output of the correlator is given by

$$\gamma_b = \frac{E_b}{I_0} \quad (10.4.12)$$

where the denominator in Eq. (10.4.12) is given by

$$I_0 = \frac{1}{E_c} \int_{-\infty}^{\infty} |P(f)|^2 \Phi_{zz}(f) df \quad (10.4.13)$$

in which $P(f)$ is the Fourier transform of $p(\tau)$, the impulse response of the pulse-shaping filter used to form the chips, $\Phi_{zz}(f)$ is the power spectral density of the interference signal $z(t)$ at the output of the correlator, and $E_c = E_b/N$ is the signal energy per chip. Finally, we determine the error rate of the DSSS system by substituting γ_b into the appropriate error-rate formula for the given modulation technique.

Effects of Wideband Interference. As in the analysis of FHSS transmission, we assume that the wideband interference has a flat spectrum covering the entire transmission band. If the power spectral density of the interfering signal is I , we have

$$\Phi_{zz}(f) = I \quad (10.4.14)$$

and we find I_0 , the interference component of γ_b , from Eq. (10.4.13) by substituting I for $\Phi_{zz}(f)$, which yields

$$I_0 = \frac{1}{E_c} \int_{-\infty}^{\infty} |P(f)|^2 I df = I \quad (10.4.15)$$

which is independent of the chip pulse shape. The signal-to-interference ratio (SIR) for the case of wideband interference is

$$\gamma_{b, \text{WI}} = \frac{E_b}{I_0} = \frac{W}{R_b} \frac{P_{\text{av}}}{I_{\text{av}}} \quad (10.4.16)$$

which is the same as the SIR for an FHSS system with wideband interference given by Eq. (10.4.4).

Note here that the processing gain W/R_b affects the SIR. For a fixed interference power, an increase in the operating bandwidth reduces I_0 , resulting in an increase in the SIR. Therefore, as with FHSS transmission in wideband interference, we can control the error rate of the received signal by adjusting the transmission bandwidth. On an additive Gaussian noise channel, $I_0 = N_0$ is constant and the SIR remains fixed as we increase the transmission bandwidth. Of course, if the interference consists of several other DSSS systems sharing the same band, the aggregate interference power is simply the sum of the interfering powers.

Effects of Narrowband Interference. To analyze the performance in narrowband interference, we assume that the interferer transmits only a carrier at the center frequency of our DSSS system. The power spectral density of the baseband equivalent of this signal is given by

$$\Phi_{zz}(f) = I_{\text{av}} \delta(f) \quad (10.4.17)$$

where I_{av} is the average power of the interfering signal. In this case we have

$$I_0 = \frac{1}{E_c} \int_{-\infty}^{\infty} |P(f)|^2 I_{\text{av}} \delta(f) df = \frac{I_{\text{av}}}{E_c} |P(0)|^2 \quad (10.4.18)$$

where

$$P(0) = \int_{-\infty}^{\infty} p(t) dt \quad (10.4.19)$$

with $p(t)$ the chip pulse waveform. The variance of the noise depends on the shape of the chip pulse waveform. The SIR is given by

$$\gamma_{b,NI} = \frac{E_c/|P(0)|^2}{R_b} \frac{P_{av}}{I_{av}} \quad (10.4.20)$$

For a rectangular pulse, $|P(0)|^2 = T_c E_c = E_c/W$ and

$$\gamma_{b,NI} = \frac{W}{R_b} \frac{P_{av}}{I_{av}} \quad (10.4.21)$$

which is the same as wideband interference in a DSSS system given by Eq. (10.4.16) or an FHSS system given by Eq. (10.4.4). Therefore, the performance of a DSSS system with rectangular pulses is the same for narrowband and wideband interference. This is in contrast with our observations in the analysis of FHSS systems. We saw there that the performance of an FHSS system in wideband interference is the same as the performance of a DSSS system, and an increase in the processing gain results in an exponential reduction in the error rate. However, we saw that the performance of an FHSS system in narrowband interference is quite different, and the error rate improves as a reciprocal function of the processing gain.

Example 10.18: Effects of Pulse Shape To observe the effect of choosing a chip pulse shape other than rectangular, let us assume a half sine wave as the pulse waveform. For this waveform, $|P(0)|^2 = 8T_c E_c/\pi^2$. Thus, the variance of the noise for half-sine-wave pulses is 0.9 dB smaller than in the case of rectangular pulses, which results in a 0.9-dB performance improvement in wideband interference.

Narrowband-Interference Suppression. The effects of narrowband interference on the performance of a DSSS system can be greatly reduced or nearly eliminated by the application of adaptive interference suppression techniques. One technique of this type is to apply a narrowband interference canceller to the received signal. This device is an adaptive filter, similar to an equalizer, that adjusts its parameters so as to place a deep notch at the location of the narrowband interference. The filter adjusts itself adaptively by using an optimization algorithm operating on the observed spectral or correlation properties of the received signal plus interference. Even if the spectral level of the narrowband interferer is very high relative to the desired DSSS signal, the filter can greatly reduce the interference, although the filtering will cause some change in the correlation properties of the underlying desired signal. If the bandwidth of the interferer is very narrow relative to the desired signal, the impact on demodulation of the desired signal is generally found to be very small.

A somewhat more general approach to narrowband interference suppression which has received considerable attention is one based on *transform-domain processing* (TDP). Briefly described, a TDP receiver for a DSSS system performs a Fourier transformation of the received signal plus interference and then suppresses interferers by adaptively

“excising” or “soft-limiting” appropriate interference spectral components [Gev89]. Narrowband interference suppression in FHSS systems is done in a manner closely related to the TDP technique. That is, techniques are employed to determine whether the signals on some hops are being subjected to interference, and those hop signals are then attenuated or completely excised. In the case of fast-FHSS systems, where there are multiple hops per information symbol, the situation is very much like that of a DSSS system with narrowband interference, and excision of hops hit by interference will greatly reduce the interference effects, with relatively little impact on the underlying FHSS signal. This is particularly true where coding can be applied in conjunction with narrowband interference excision. In slow-FHSS systems, where multiple information bits are contained in each hop, interference excision is less effective, because the excision of a single hop cancels several information bits together with the interference. However, even with slow frequency hopping, interference excision can provide performance benefits when applied in conjunction with coding designs using sufficiently long constraint lengths. Interference suppression techniques have been investigated and applied extensively in military communications systems. Applications to commercial wireless networks are under investigation, with particular attention to the issue of coexistence between new broadband CDMA systems and existing narrowband systems [Mil92].

10.4.3 Effects of Coding

As we saw earlier in this section, the equations used for calculation of the output SNR of various spread-spectrum systems in the presence of interference is not the same for all systems and all forms of interference. For an FHSS system in wideband interference and a DSSS system in narrowband or wideband interference, the signal-to-interference ratios given by Eqs. (10.4.4), (10.4.16), and (10.4.21) are identical. The same general formula, given in Eq. (10.4.7), is found for an *average* SIR for an FHSS system in narrowband interference, but the effect on FHSS system performance in narrowband interference is different from the other three cases. In particular, the probability of error for an FHSS system in narrowband interference, as given by Eq. (10.4.8), decreases as a reciprocal of the system processing gain N . However, for the FHSS system in wideband interference, the error probability decreases exponentially with processing gain, as shown by Eq. (10.4.9). With FHSS systems, the same pattern of separation also holds for the effects of coding, with coding being potentially more effective against narrowband interference than against wideband interference.

For an FHSS system operating in narrowband interference, as we explained earlier, an appropriately chosen error-correcting code can be highly effective in eliminating the errors in signal hops affected by the interferer. The general expression for symbol-error probability at the output of a t -error-correcting decoder is given by

$$P_s \leq \sum_{k=t+1}^n \binom{n}{k} P^k (1-P)^{(n-k)} \quad (10.4.22)$$

For an FHSS system in narrowband interference, coding is made most effective by making the block length or constraint length sufficiently long to span many hops. In this way, the hops affected by interference are a small fraction of the set of hops spanned by the code, and with a judicious choice of the error-correction limit t , most or all of the errors can be eliminated. In a fast-FHSS system, the interference affects

only one bit out of N , and therefore a good choice of block-coding parameters is to have $t/n = 1/N$, where n is the code block length. In slow frequency hopping we need an interleaver to disperse the clusters of errors that occur when the signal lands on the hop frequency with interference. With a sufficiently long interleaver the code performance in slow hopping will approach that found in fast hopping. However, interleaving involves buffering operations at both ends of the link, and this creates an added source of time delay in the delivered data stream. For digital voice service, the time delay must be carefully controlled, and this places a practical limit on the amount of interleaving that can be done. For a data service, time delay is not so critical, and longer interleaving buffers can be used.

It is useful to observe that in an FHSS system, the use of coding to combat narrow-band interference has much similarity to the use of diversity combining on a fading channel. The redundancy of the error-correction code is similar to the multiplicity of branches in a diversity system. The code effectively combines a small number of erroneous bits with a larger number of error-free bits and uses the mathematical structure of the code in a decoding algorithm to identify and correct the erroneous bits. Similarly, the diversity combiner brings together one or a few low-SNR branches with several higher-SNR branches and applies a combining algorithm that allows the higher-SNR branches to prevail in forming the decision on the information symbol. Here we have one hop with low and many hops with high SIR values. We saw in Section 8.3 that the error probability improvements provided by diversity is an exponential function of the diversity order D , and thus with a limited number of diversity branches we can improve the performance by several orders of magnitude. With the use of appropriate coding, the channel error rate given by Eq. (10.4.8) can be brought down to an extremely low postdecoding error rate, corresponding to a substantial lowering of the SIR required.

In all cases other than that of an FHSS system in narrowband interference, coding will provide more modest reductions in the minimum signal-to-interference power ratio required for the proper operation of the system, at the cost of either reducing the information rate or increasing the system bandwidth. In general, for all these other cases the channel error rate without coding is given by $P_e(\gamma_b)$, where γ_b is the uncoded SIR and P_e is the appropriate error probability equation for the given modulation technique. Let us say that we use an (n, k) block code to encode the information bits for transmission and that the Hamming weights of the codewords are $\{W_m\}$, $m = 1, 2, 3, \dots, M$, where $M = 2^k$ is the total number of codewords in the code. Let $m = 1$ designate the all-zeros codeword and let us assume that this codeword is transmitted. (We could assume transmission of any of the M codewords, but the discussion is made simpler by assuming an all-zeros codeword.) Now consider an ideal decoder at the receiving end, one that decodes the received sequence of n symbols by correlating the sequence with a list of the M possible codewords and selecting the codeword with the highest correlation. This is exactly the same as processing the received sequence with M matched filters, each filter corresponding to one possible codeword. Given this scheme, a decoding error occurs if the received sequence has higher correlation with any one of the $M - 1$ incorrect codewords than it does with the all-zeros codeword actually transmitted in our example. The probability of erroneously deciding in favor of the m th codeword is simply

$$P_m = P_e(\gamma_b R_c W_m)$$

where P_e denotes the probability of a binary decision error, $R_c = k/n$ is the rate of the block code, and W_m is the Hamming weight of the m th codeword. The overall word-error probability for this decoding procedure is difficult to compute precisely, but it can be bounded by taking the union bound of the $M - 1$ possible error events, which we write as

$$P_B \leq \sum_{m=2}^M P_e(\gamma_b R_c W_m) \quad (10.4.23)$$

In the limit of high SNR, the decoding error events will be dominated by false decodings to minimum-distance codewords, and if the minimum distance of the code is $d_{\min}$, the word-error probability becomes

$$P_B \simeq A_{d,\min} P_e(\gamma_b R_c d_{\min}), \quad \gamma_b \gg 1$$

where $A_{d,\min}$ denotes the number of minimum-weight codewords in the code. Similarly, for a convolutional code with minimum free distance d_{free} , and assuming soft-decision Viterbi decoding, we have a union bound on the probability of an information symbol decision error given by

$$P_C \leq \sum_{d=d_{\text{free}}}^{\infty} b_d P_e(\gamma_b R_c d) \quad (10.4.24)$$

where b_d is the number of code trellis paths of weight d times the number of corresponding output bit errors caused by decoding to an incorrect weight- d path. These equations are essentially the same as those derived for conventional nonspread modulation over additive noise channels. For a fixed source information rate, coding will improve the interference margin $I_{\text{av}}/P_{\text{av}}$ by a factor of $R_c d_{\min}$ (or $R_c d_{\text{free}}$). For a fixed transmission bandwidth, coding will improve the interference margin by a factor of $d_{\min}$.

10.5 PERFORMANCE OF CDMA SYSTEMS

With the growing interest in integration of voice, data, and imagery traffic in wireless networks, CDMA became the wireless access method of choice for third-generation systems. Fundamentally, integration of various types of traffic is readily accomplished in a CDMA environment because coexistence in such an environment does not require specific coordination among user terminals. In principle, CDMA can accommodate various wireless users with different bandwidth requirements, switching methods, and technical characteristics, without any need for coordination. Of course, because each user signal contributes to the interference seen by other users, power control techniques are essential in the efficient operation of a CDMA system. In practice, CDMA was not used in WLAN products because with an efficient channel access technique such as carrier-sense multiple access with collision detection (CSMA/CD), adequate numbers of users are supported in the unlicensed frequency bands allocated. However, CDMA systems became dominant for PCS and cellular telephone applications. For these voice-oriented networks, the volume of traffic is extremely high and the expensive licensed bands available have to be used as efficiently as possible in order

to support the large numbers of users. The third-generation standards for the CDMA cellular networks are written to include the integration of packet- and circuit-switched data as well. The flexibility of CDMA has made this approach the leading candidate for third-generation multimedia wireless networks. The two common CDMA techniques are direct-sequence CDMA (DS-CDMA) and frequency-hopping CDMA (FH-CDMA), which are constructed on the DSSS and FHSS techniques, respectively. The DS-CDMA technique has received by far the greater attention in ongoing developments of new systems, and we address a number of technical issues related to this technique in the remainder of Section 10.5. These issues include M -ary orthogonal coding used in IEEE 802.11b and IS-95 standards, interference cancellation, multiuser detection, and multi-carrier CDMA that are used to enhance the capacity and increase the flexibility of the CDMA systems for third-generation cellular networks and beyond. In Section 10.5.5 we discuss briefly the FH-CDMA technique used in a Bluetooth ad hoc network.

10.5.1 Direct-Sequence CDMA

In DS-CDMA, each user's PN sequence is generated by a unique code, and this allows receivers to distinguish among different users' signals. In this way, multiple users can transmit simultaneously using the chip bandwidth R_c , and their signals can be separated using their unique PN codes. As in any DSSS system, the receiver multiplies the received signal by a replica of the PN chip sequence used at the transmitter, and it integrates or otherwise filters over a bit interval. In the CDMA system, the receiver uses a PN code unique to the desired user signal, and the signals coded with other users' codes simply appear as wideband noise and are thus reduced by the DS processing gain.

CDMA with Ideal Power Control. Here we calculate the signal-to-interference-and-noise ratio (SINR) for signals received at the base station of a CDMA cellular system in terms of the number of users sharing the channel simultaneously. This is different from the SNR, which is the ratio of the signal power to background additive noise and does not include interference as a source of noise. Having determined the SINR, we can relate it to the error rate requirements and other factors involved in the design of digital cellular systems. Our basic assumption here is one of ideal power control. That is, we assume that the power of each user's transmitter is controlled by the central base station, so that precisely the same power is received from every terminal, and the power level is constant with time. We further assume a DS-CDMA system in which M users are sending information to the base station simultaneously and the bandwidth of the shared channel is $W = NR_b = N/T_b$, where N is the system processing gain. One of the users is the "target" mobile station; the other $M - 1$ users are wideband interferers to the target station. At the base station the total interference power from all interferers is $(M - 1)E_b R_b$. The power of the additive noise at the base station is $N_0 W$, and therefore the total interference plus noise at the base station is

$$I_{av} = N_0 W + (M - 1)E_b R_b$$

The power received from the target receiver is

$$P_{av} = E_b R_b$$

and the SINR is therefore given by

$$\gamma = \frac{W}{R_b} \frac{P_{av}}{I_{av}} = \frac{E_b/N_0}{1 + [(M-1)/N](E_b/N_0)} \quad (10.5.1)$$

If there is no interferer and only one user, $M = 1$ and $\gamma = E_b/N_0$. If the number of users in the system is large and the same power is received from all terminals, the interference noise dominates the background noise. Then the SINR is approximated by

$$\gamma \simeq \frac{N}{M-1} \quad (10.5.2)$$

and for some desired level of SINR η , the number of users is given by

$$M = 1 + \frac{N}{\gamma} \quad (10.5.3)$$

The bandwidth efficiency η is defined as the aggregate data rate of the users, normalized by the system bandwidth, which we write as

$$\eta = \frac{MR_b}{W} = \frac{M}{N} = \frac{1}{N} + \frac{1}{\gamma} \quad (10.5.4)$$

The required value of γ depends on the modulation and coding techniques used in the system, together with the users' requirement on acceptable error rate. In this analysis we have assumed that the same signal power is received from every user terminal. In the radio environment, of course, the received power fluctuates due to fading. However, if we extend our definition of ideal power control to mean *ideal instantaneous power control*, the power fluctuations caused by fading are eliminated and performance analysis is identical to that for a nonfading channel.

Example 10.19: Number of Users and Error Rate For BPSK modulation, $\gamma = 10$ dB will provide an error rate of 10^{-5} , which is a reasonable error rate for many applications using uncoded data. With a processing gain of $N \simeq 100$, the number of simultaneous users and the bandwidth efficiency, calculated from Eqs. (10.5.3) and (10.5.4), are 11 and 0.11, respectively. For $\gamma \simeq 7$ dB, the error rate is around 10^{-3} , an acceptable error rate for telephone-quality digital voice communications. The number of users and the bandwidth efficiency in this case will increase to 21 and 0.21, respectively.

In general, for evolving digital cellular systems the first terms in Eqs. (10.5.3) and (10.5.4) are much smaller than the second terms and the two equations can be further simplified as

$$M = \frac{N}{\gamma} = \frac{W/R_b}{E_b/I_0} \quad (10.5.5)$$

and

$$\eta = \frac{1}{\gamma} = \frac{1}{E_b/I_0} \quad (10.5.6)$$

Example 10.20: Ideal Bandwidth Efficiency Using QPSK modulation and convolutional coding, the existing CDMA digital cellular systems require $3 \text{ dB} < \gamma < 9 \text{ dB}$. Using Eq. (10.5.5) with $N \simeq 100$, the number of users will be in the range $12 < M < 50$ and the bandwidth efficiency will be in the range $0.12 < \eta < 0.5$.

Practical Considerations. In the practical design of digital cellular systems, three other parameters affect the number of users that can be supported by the system as well as the bandwidth efficiency of the system. These are the number of sectors in each base station antenna, the activity factor, and the interference increase factor. These parameters are quantified as factors used in the calculation of the number of simultaneous users that the CDMA system can support. The use of sectored antennas is an important factor in maximizing bandwidth efficiency. Antenna sectorization reduces the overall interference, increasing the allowable number of simultaneous users by a sectorization gain factor which we denote by G_A . With ideal sectorization, the users in one sector of a base station antenna do not interfere with users operating in other sectors, and $G_A = N_{\text{sect}}$, where N_{sect} is the number of sectors in the antenna pattern. In practice, antenna patterns cannot be designed with ideal sectorization, and due to multipath reflections, users in general communicate in more than one sector. Three-sector base station antennas are commonly used in cellular systems, and a typical value of the sectorization gain factor is $G_A = 2.5$ (4 dB). The activity interference reduction factor G_v is the ratio of the total connection time to the active talkspurt time. On average, in a two-way conversation, each user talks 50% of the time. The short pauses in the flow of natural speech reduce the voice activity factor further, to about 40% of the connection time in each direction. As a result, the typical number used for G_v is 2.5 (4 dB). The interference increase factor H_0 accounts for users in other cells in the CDMA system. Because all neighboring cells in a CDMA cellular network operate at the same frequency, they will cause additional interference. The interference is relatively small, due to the processing gain of the system; a value of $H_0 = 1.6$ (2 dB) is commonly used in the industry.

Incorporating these three factors as a correction to Eq. (10.5.5), the number of simultaneous calls in a CDMA cell is approximated by

$$M \simeq \frac{N}{\gamma} \frac{G_A G_v}{H_0} \quad (10.5.7)$$

Example 10.21: Effects of Practical Considerations If we continue Example 10.20 with the new correction factor included in Eq. (10.5.7), the range for the number of simultaneous users becomes $50 < M < 200$. The range of associated bandwidth efficiencies is $0.5 < \eta < 2.0$. To find the required bandwidth, let us assume that the system is carrying digitized voice encoded at $R_b = 7 \text{ kb/s}$. The occupied bandwidth is $W = 700 \text{ kHz}$ and the bandwidth per active user is between 3.5 and 14 kHz, which should be compared with the 30 kHz per channel used in existing FM-modulated analog AMPS systems.

Example 10.22: Performance of IS-95 For the IS-95 CDMA using each carrier with $W = 1.25 \text{ MHz}$ and a data rate per user of $R = 9600 \text{ bits/s}$, for $\gamma = 4$ (6 dB) and a performance improvement factor of $K = 4$, we have

$$N = \frac{W}{R} = 128$$

and

$$M = \frac{N}{\gamma} \frac{G_A G_v}{H_0} = 108$$

users per cell.

Practical Power Control. Our earlier discussion presented an ideal situation, which can be approached as a practical matter when power control is used on the link from a mobile station to a base station (usually termed the *uplink*). Power control is an essential function in any DS-CDMA system serving mobile users, because of the near-far problem. On the uplink each signal is received with a different amount of path loss, due to (1) variations in distances of the mobiles from the base site and (2) the statistical variability of the path loss. Therefore, if all the mobiles transmit with the same signal power, their signals will arrive at the base station with widely different power levels. As a result, a few mobile stations whose signals arrive with much less path loss than the other stations will overpower the remaining mobiles. This is readily seen from Eq. (10.5.1), which is written for the case where all signals arrive with equal power $E_b R_b$.

Example 10.23: Nonideal Power Control If one user's signal arrives with a power level 10 dB higher than the others, the interference from this user is equivalent to the interference of 10 other users. Therefore, the number of simultaneous users the system is able to support is reduced by nine. The SINR for this user, however, is 10 dB higher than for the others. In principle, it is feasible to design a CDMA system in such a way that user channels with different SINR requirements are overlaid (e.g., to provide for simultaneous operation of different types of services having different BER requirements). This might be done by controlling the power to different levels on different user channels.

The solution to this problem is to control the power transmitted by each mobile station so that the same power level is received at the base station from each mobile. This requires that the base station provide continuous feedback to each mobile station so that the mobile can adjust its power level dynamically. In a practical mobile communication system, it may be necessary to be able to adjust power levels over a range as wide as 80 dB.

Any practical power control algorithm has a limit on the range of signal variations that it can accommodate. Furthermore, the power-level adjustments are performed in discrete steps, and in the case of mobile and cellular systems, the speed of adjustment might not be adequate for all vehicular speeds. As a result, the power is not perfectly controlled and we can model the deviation from the ideal control signal as a random variable. The lognormal distribution is a natural model for smaller variations in the received power, measured in decibels. In the digital cellular industry a lognormal distribution has been used for this purpose and the modeling results have shown close agreement with the results of measurements [Gil91, Vit93, Pad94].

To calculate the allowable number of CDMA users by the approach that we outlined, we need to determine the required SINR averaged over the lognormal fading distribution. This approach involves integration of the error-rate functions (either exponential or erfc) with respect to the lognormal distribution, which is not analytically feasible.

An analytical approximation to the solution is provided in [Vit93, Pad94], in which the allowable number of users is approximated by

$$M = \frac{N}{e^{(\beta\sigma)^2/2 + \beta\mu}} \quad (10.5.8)$$

where $\beta = 0.1 \ln 10$ and μ and σ are the mean and variance of the lognormal power control error, respectively. Details of this derivation are provided in Problem 11 at the end of the chapter. As compared with the case of perfect power control given by Eq. (10.5.5), γ , the minimum required SIR, is now replaced by an exponential. Generally, γ is smaller than the exponential term in the denominator of Eq. (10.5.8). The ratio of the exponential term to γ represents the reduction in the number of allowable users, due to the use of nonideal practical power control.

It should be noted that although fluctuations in the received power due to the near-far problem have a serious negative impact on the capacity of a DS-CDMA system, the power fluctuations will in fact increase the throughput of contention-based wireless access methods such as ALOHA and CSMA. We explore this point in greater detail in Chapter 11, where we discuss network topologies and network access methods. The results presented above are based on the assumption that the signal amplitude received from any path is a Rayleigh-distributed random variable and that we have a fixed number of signal paths on any user channel. As we saw in Chapter 6 when we discussed channel simulation and modeling, these assumptions are not necessarily valid, and in most cases the empirical data indicate a lognormal amplitude distribution and modified Poisson arrival times. The most accurate way of predicting system performance is to use the results of channel impulse response measurements to calculate the error rate directly. This prediction method was investigated in [Cha93] using impulse responses measured on indoor radio channels; the impulse response measurements are those described in Chapter 3. Figure 10.27 shows a set of performance prediction curves obtained for a CDMA system operating over channels in five different manufacturing areas.

10.5.2 *M*-Ary Orthogonal Codes

M-ary orthogonal codes are used in different context in wireless networks; in this section we provide a summary of these applications. In CDMA links (e.g., a forward or downlink IS-95 channel), the orthogonal codes are used to separate different user channels. This setting also provides for flexibility to integrate user traffic with different data rates. In other CDMA links (e.g., a reverse or uplink IS-95 channel), *M*-ary orthogonal codes are used by individual users to improve the quality of the received signal at the base stations. Orthogonal codes can also be used to increasing the bandwidth efficiency of a DSSS system [Pah87]. In IEEE 802.11b WLANs, *M*-ary orthogonal codes are used by individual users to increase the 2-Mb/s data rate of the DSSS IEEE 802.11 to 11 Mb/s. In HDR cellular networks all 64 forward channels of the IS-95 are assigned to a single user, to provide data rates on the order of 2.4 Mb/s. With this approach, each user has a set of orthogonal sequences representing a set of symbols for transmission. The set of orthogonal sequences, or *orthogonal code*, will typically be a Hadamard or Walsh code, discussed in Chapter 8.

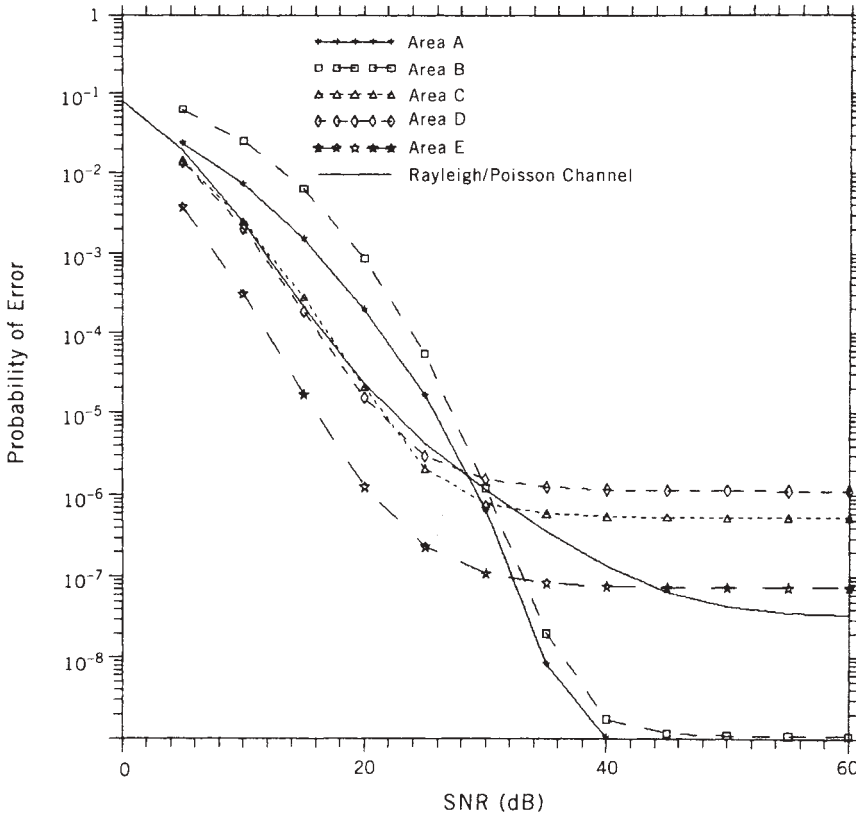


FIGURE 10.27 Performance prediction curves obtained for a CDMA system with one active transmitter operating over the five manufacturing floors described in Chapter 3. A Rayleigh fading model and a Poisson arrival model are assumed. There are two code sequences per user, with $N = 255$ chips per code sequence [Cha93].

Example 10.24: Walsh Codes in IS-95 Walsh codes are based on the Hadamard matrix, which is defined by the recursion

$$H_{2N} = \begin{bmatrix} H_N & H_N \\ H_N & \overline{H_N} \end{bmatrix}$$

Figure 10.28 shows a Hadamard matrix of orders 2, 4, and 8. The rows of this matrix provide for a set of orthogonal codes referred to as *Walsh codes*. Therefore, the first 8-Walsh code is $W_0 = [1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1]$ and the last 8-Walsh code is $W_7 = [1 \ 0 \ 0 \ 1 \ 0 \ 1 \ 1 \ 0]$. In general, Walsh codes of any order are orthogonal,

$$W_i W_j = \begin{cases} N, & i = j \\ 0, & i \neq j \end{cases}$$

because for each pair, half of the bits are different and the other half are the same. Walsh codes are very popular in CDMA systems. The 64-Walsh codes are used in the

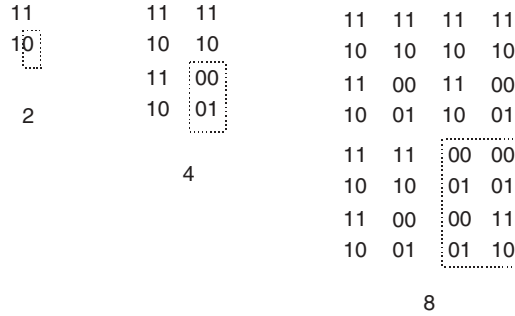


FIGURE 10.28 Hadamard matrix of orders 2, 4, and 8.

IS-95 forward channel (from base station to mobile) as the code to separate different users. The same codes are also used in the IS-95 reverse channel (from mobile to base) for M -ary orthogonal coding to improve the performance against the noise and interference.

A stream of information bits is segmented into groups and each group represents a nonbinary information symbol, which is associated with a particular transmitted code sequence. If there are n bits per group, one of a set of $K = 2^n$ sequences is transmitted in each symbol interval. The received signal is correlated with a set of $K = 2^n$ matched filters, each matched to the code sequence of one symbol. The correlator outputs are compared, and the symbol associated with the largest output is declared to be the transmitted symbol.

Example 10.25: CCK Codes in IEEE 802.11b Another popular M -ary coding technique is *complementary code keying* (CCK) used in IEEE 802.11b WLANs to support a data rate of 11 Mb/s on a channel that transmits 11-MS/s QPSK waveforms. A simplified block diagram of the basic principles of CCK is shown in Fig. 10.29. The 11-Mb/s input data stream is grouped into 8-bit symbols at 11 Mb/s/8 = 1.375 MS/s. The encoder maps each 8-bit symbol into eight four-phase coded symbols that are transmitted serially using a QPSK four-phase modulator. At the receiver, each eight received complex waveforms from the QPSK demodulator are grouped in a block and sent to the decoder to find the closest 8-bit symbol associated with the demodulated eight four-phase signals. The chip rate and occupied bandwidth of this system are the same as that of the original 2-Mb/s IEEE 802.11 DSSS standard, but the data rate is increased to 11 Mb/s.

The 256 coded symbols are selected from the $4^8 = 65,536$ available alternatives so that they are orthogonal to one another. The following equation, recommended by the 802.11b committee, gives the mapping rule for generation of the codes:

$$c = \{e^{j(\varphi_1 + \varphi_2 + \varphi_3 + \varphi_4)}, e^{j(\varphi_1 + \varphi_3 + \varphi_4)}, e^{j(\varphi_1 + \varphi_2 + \varphi_4)}, -e^{j(\varphi_1 + \varphi_4)}, e^{j(\varphi_1 + \varphi_2 + \varphi_3)}, e^{j(\varphi_1 + \varphi_3)}, -e^{j(\varphi_1 + \varphi_2)}, e^{j\varphi_1}\}$$

The eight input bits are grouped into four 2-bit complex four-phase symbols. The resulting four phases, $\varphi_1, \varphi_2, \varphi_3$, and φ_4 , are inserted in to the equation above to find the eight complex-coded waveforms that are serially modulated by the QPSK modem.

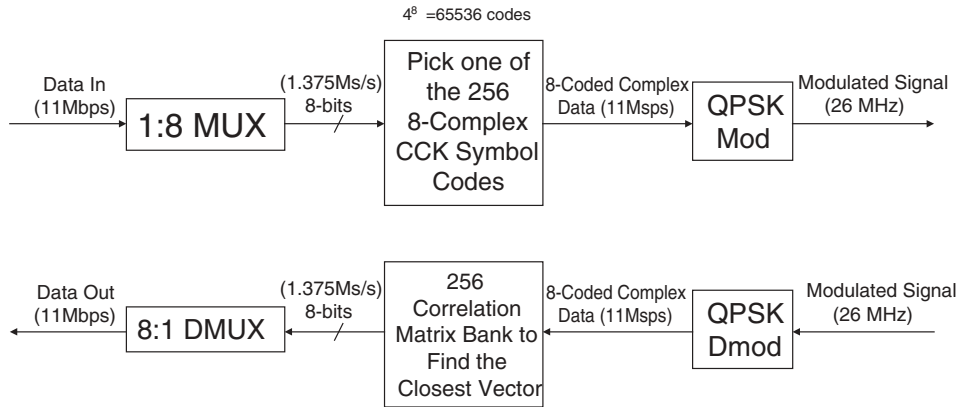


FIGURE 10.29 General concept of CCK M -ary coding used in the IEEE 802.11b.

The CCK codes are not strictly orthogonal, but the performance is as close as one can get to a perfect orthogonal code such as the Walsh codes. The novelty of this code is that it fits perfectly in the 802.11 standard, which uses 11 chips for spreading.

Example 10.26: RAKE Receiver for M -ary Codes To increase the data transmission rate in a CDMA network, it is possible to assign multiple parallel channels to a single user. A multipath condition destroys the orthogonality of these channels, creating interchannel interference (ICI), which degrades the performance of RAKE receivers. Figure 10.30 [Lat98b] illustrates degradation in the bit error probability of a RAKE receiver in a two-path Rayleigh fading channel as a function of the number of parallel channels. Different curves demonstrate the degradations for different spreading factors, and the SNR = 20 dB. As discussed before, instead of using multiple parallel channels, one can use smaller spreading factor to increase the data rate. When we examine Fig. 10.30, we observe that for a given error-rate requirement (e.g., 10^{-2}), we have multiple choices, such as a spreading factor of 16 and three parallel channels, a spreading factor of 32 and six parallel channel, or a spreading factor of 64 and 12 parallel channels. For a given spreading code, an increase in the number of parallel channels results in an increase in the ICI and additional performance degradation. Using an equalizer reduces the harmful effects of multipath and the resulting ICI [Lat98b].

Example 10.27: Practical Implementation The data rate of a traditional IS-95 CDMA network from the base station to mobile stations is 9600bits/s, the spreading code is 128, and the chip rate is $9600\text{bits/s} \times 128 \text{ chips/bit} = 1.2288 \text{ Mc/s}$. This system supports 64 voice users that are separated by 64 orthogonal Walsh codes of length 64. To support high-data-rate (HDR) systems, all 64 Walsh codes are used for a single data user. Figure 10.31 illustrates the general concept of the IS-95 CDMA and the single HDR evolution (1xEV), which uses all 64 CDMA codes for one user and allows all users access through a TDMA scheme. If BPSK were used for all channels, the maximum achievable data rates would be $64 \times 9600 \text{ bits/s} = 614.4 \text{ Mb/s}$. To increase the data transmission rate further, 1xEV uses multisymbol modulations to replace the existing BPSK modulation. The highest data rate in this system is achieved by a 16-QAM modulation with 4 bits per symbol, resulting in $614.4 \text{ Mb/s} \times 4 = 2.4576 \text{ Mb/s}$.

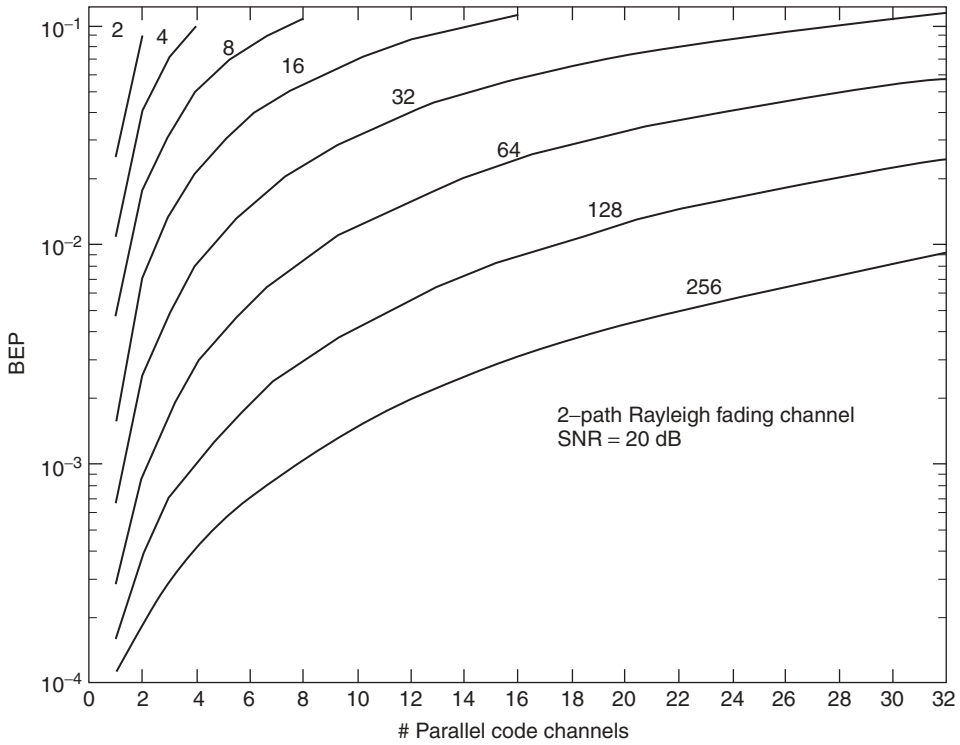


FIGURE 10.30 Performance of a RAKE receiver in a two-path Rayleigh fading channel for different numbers of parallel codes, a variety of spreading factors, and an $SNR = 20$ dB [Lat98b].

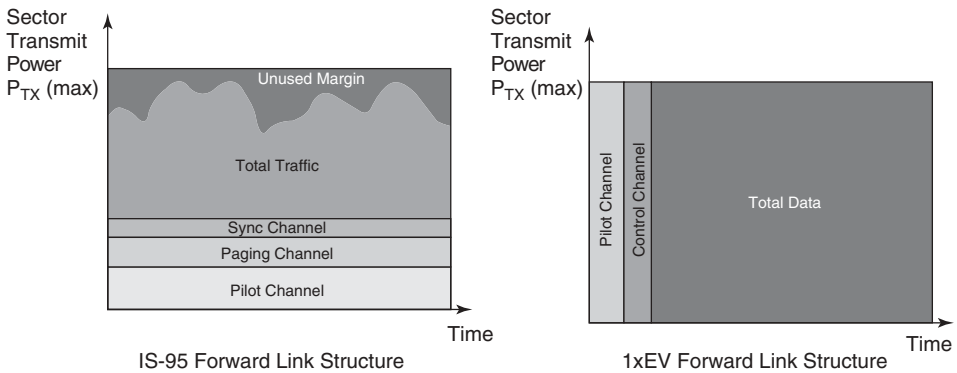


FIGURE 10.31 Distribution of transmitted power in IS-95 CDMA and 1xEV HDR forward channels using TDMA for data communications [Qua01].

[Ben00, Qua01]. As shown in Table 10.1, the 1xEV HDR system supports 12 data rates for packet-data transmission. Depending on the received SNR by the mobile station, the base station, adjusts the number of symbols of the modulation technique, number of bits per packet, and the number of times the packet is repeated (that is

TABLE 10.1 Specification of Different Data Rates in the 1xEV HDR System [Qua01]

Data Rate (kb/s)	Physical Layer Parameters											
	38.4	76.8	153.6	307.2	307.2	614.4	614.4	921.6	1228.8	1228.8	1843.2	2457.6
Modulation type	QPSK	QPSK	QPSK	QPSK	QPSK	QPSK	QPSK	8PSK	QPSK	16QAM	8PSK	16QAM
Bits per encoder packet	1024	1024	1024	1024	2048	1024	2048	3072	2048	4096	3072	4096
Code rate	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
Encoder packet duration (ms)	26.67	13.33	6.67	3.33	6.67	1.67	3.33	3.33	1.67	3.33	1.67	1.67
Number of slots	16	8	4	2	4	1	2	2	1	2	1	1

Source: [Qua01].

equivalent to changing the effective processing gain). The minimum data transmission rate is 38.4 kb/s, which is 64 times lower than the highest rate and is achieved by sending fourfold shorter packets in a 16-fold longer time interval. To support this data rate, a QPSK modulation is used. When the mobile is in close proximity to the base station, the highest data rate with 16-QAM modulation and minimal coding is used as the mobile moves away from the base station, the number of symbols in the signal constellation is reduced, and the coding for error protection is increased—to achieve the lowest data rate.

In the decision process, in addition to multiuser and multipath interference we have an additional form of interference introduced in the decision process when we compare the outputs of the correlators. Assuming a simple one-path channel, perfect power control, and negligible additive noise, the SIR is given by

$$\gamma = \frac{N}{M - 1 + K - 1} \quad (10.5.9)$$

where $M - 1$ represents the noise from other users and $K - 1$ represents the noise from the outputs of the correlators other than the one corresponding to the correct symbol. The allowable number of users is then given by

$$M = \frac{N}{\gamma} - K + 2 \quad (10.5.10)$$

which indicates that the number of simultaneous users is reduced by the use of orthogonal coding. However, we should note that the data rate of the individual users, and consequently, the bandwidth efficiency of the system, are improved by a factor $\log_2 K$, as shown by

$$\eta = \frac{MR_b}{W} = \frac{M \log_2 KR_s}{W} = \frac{M \log_2 K}{N} \quad (10.5.11)$$

where R_s is the symbol transmission rate.

TABLE 10.2 Bandwidth Efficiency for CDMA Systems Operating Over Fading Multipath Channels with $N = 256$, $K = 2 - 64$, $E_b/N_0 = 30$ dB, Two Orders of Explicit Diversity, Different Numbers of Symbols, and Different Orders of Implicit Diversity

K	Two Paths	Four Paths	Seven Paths
2	0.0429	0.0664	0.0742
4	0.0702	0.1171	0.1327
8	0.0937	0.1524	0.1868
16	0.1094	0.1875	0.2343
32	0.1171	0.2148	0.2734
64	0.1409	0.2577	0.3047

Source: [Pah90a].

Example 10.28: Multisymbol CDMA Consider a system in which we have $N = 100$, a required minimum SINR of $\gamma = 2$, and an orthogonal code having $K = 2$ symbols. For this case we have $M = 50$ allowable users at data rate R_b . Now consider a second case in which we have an orthogonal code with $K = 16$ symbols. Here the allowable number of users decreases to $M = 36$, but the data rate for each user increases by a factor $\log_2 16 = 4$. This means that each user in the second case can accommodate four users of the first case (e.g., in a TDMA format) and the equivalent number of users with the rates of the first case is $36 \times 4 = 144$, which is almost three times the users supportable (50) in the first case. The disadvantage of the orthogonal signaling scheme is the complexity of the receiver design. In this example we need 16 receiver correlators per user channel rather than only one needed in the simplest system design.

Table 10.2 gives the bandwidth efficiency for different numbers K of code sequences per user, in a system with code length 256 chips and two orders of explicit diversity. The results are given for two-, four-, and seven-path multipath channels. The maximum bandwidth efficiency of 0.3047 is found for the largest code set ($K = 64$) and maximum multipath spread (seven paths). This efficiency will increase if the order of explicit diversity is increased or if the constraint on the probability of error is relaxed from 10^{-4} . For two-path channels the bandwidth efficiency increases by approximately a factor of 3 as the number of code sequences per user increases from 2 to 64. For four- and seven-path channels, the corresponding increase is approximately a factor of 4. Table 10.3 shows the effect of increasing code length N for $K = 16$ orthogonal sequences and second-order explicit diversity. For a fixed bandwidth, as the code length increases, the transmission rate and consequently the bandwidth efficiency decrease. On the other hand, an increase in the code length reduces the interference from other users, which improves the bandwidth efficiency of the system. Considering both effects, Table 10.3 shows that for any number of signal paths, the bandwidth efficiency decreases slightly as the code length is increased. In practice, the two major drawbacks for implementation of orthogonal coding systems are (1) the task of finding a large set of orthogonal code sequences and (2) the complexity of the receivers.

10.5.3 Interference Cancellation and Multiuser Detection

Multiuser detection (MUD) refers to the detection of data from multiple users when each user is a source of interference for the others. This crosstalk interference arises in

TABLE 10.3 Bandwidth Efficiency for CDMA Systems Operating Over Fading Multipath Channels with $K = 16$, $E_b/N_0 = 30$ dB, Two Orders of Explicit Diversity, Different Lengths of Spread-Spectrum Code, and Different Orders of Implicit Diversity

N	$L = 2$	$L = 4$	$L = 7$
32	0.125	0.25	0.25
64	0.125	0.25	0.25
128	0.125	0.2188	0.25
256	0.1094	0.1875	0.2343
512	0.1016	0.1875	0.2188
1024	0.0977	0.1797	0.2148
2048	0.0938	0.1777	0.2129

Source: [Pah90a].

CDMA systems that use nonorthogonal spreading codes, such as IS-95 reverse channel, or CDMA systems that use orthogonally multiplexed codes, such as CDMA forward channels using Walsh codes, passing through severe multipath conditions. Therefore, in addition to the ISI and IPI interferences discussed in the analysis of the performance of RAKE receivers in Section 10.3, in a CDMA environment, receivers are also perturbed by multiple access interference (MAI). Similar to adaptive equalizers or other adaptive filters used in telecommunications, the basic principle of MUD systems is to use the characteristics of the crosstalk channel and the detected symbols or a reference signal to reduce interference and improve system performance. Significant research in the past two decade has shown that a variety of MUD techniques can provide for significant performance improvements.

With an optimal maximum-likelihood or maximum a posteriori probability receiver and perfect channel measurement, one can eliminate interference completely. Since practical implementation of such systems is quite complex, considerable amount of research has been devoted to discover more practical solutions. This research can be divided into those using linear minimum mean-squared error (LMMSE) algorithms [Ver98], iterative or successive algorithms [Hoo03], and low-complexity maximum likelihood techniques [Lin04]. The LMMSE algorithms use a linear filter similar to a multichannel equalizer to eliminate crosstalk among different users. Iterative algorithms make use of tentative channel-symbol decisions to provide feedback to improve the performance optimal or linear algorithms. Detailed description of MUD algorithms is beyond the scope of this book; one may refer to the classical source [Due95]. In general, nonlinear interference cancellation MUD algorithms are implemented at the base station, where the receiver can afford implementation of complex algorithms and the quality of the signal received is worse than that at the mobile terminal. In mobile terminals simple equalizer-like techniques are employed. We continue this discussion with an example demonstrating the performance expected using a MUD algorithm.

Example 10.29: MUD in a Forward Channel In the forward channel of CDMA systems, the base station uses orthogonal codes to separate messages for different users. However, due to the multipath nature of the channel, the orthogonality of the received signal is not preserved, and in addition to ISI and IPI, the receiver is distorted with MAI. In this example we consider the performance of three receivers: a traditional RAKE receiver, a LMMSE–RAKE receiver that passes the received signal through a

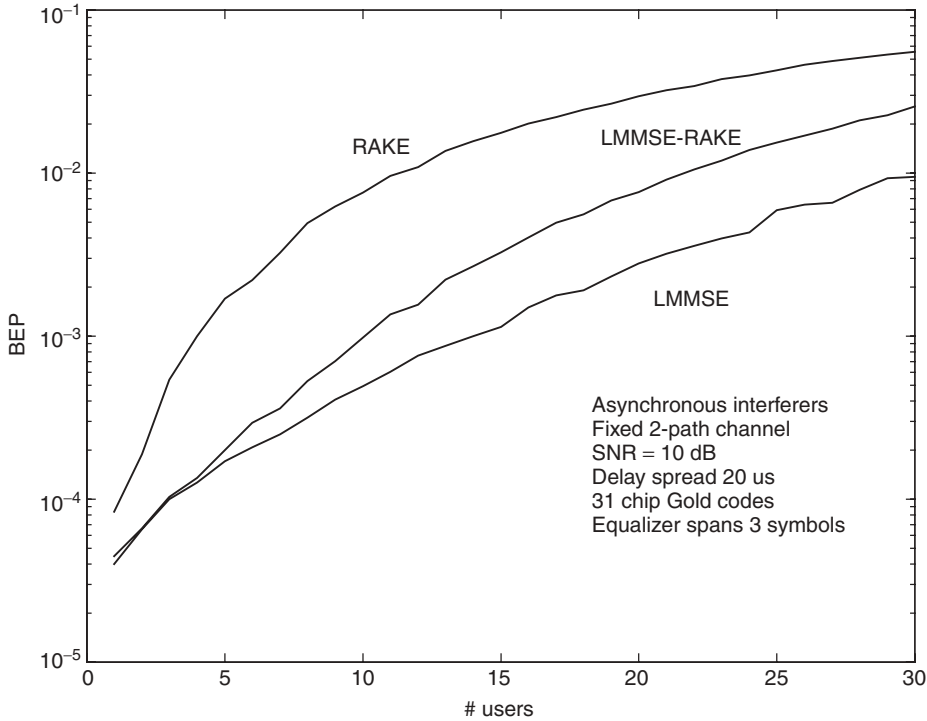


FIGURE 10.32 Performance of RAKE, LMMSE-RAKE, and LMMSE equalizer receivers in a multiuser environment [Lat98a].

multidimensional filter for crosstalk cancellation, and a three-tap linear equalizer that compensates multipath to reduce all the interferences. Figure 10.32 shows the results of simulations comparing the performance of the three receivers in a two-path fading channel. As the number of users increases, the bit-error probability increases for all cases. For an error rate of 10^{-3} , a LMMSE-RAKE multiuser detector can increase the number of users from five to 10. The performance of a LMMSE-RAKE receiver is close to that of an optimal LMMSE equalizer.

In general, as we discussed before, it is preferred to use an equalizer in the mobile because it is less complex and unlike a LMMSE multiuser detector, does not detect other users' information. The performance of a practical implementation of equalizers for WCDMA terminals is discussed in [Hoo03]. We provide some of those results as an example.

Example 10.30: Equalization Versus RAKE Figure 10.33 provides a comparison among performances of an approximated LMMSE equalizer with an infinite number of taps, a practical equalizer with a finite number of taps, and a RAKE receiver. The figure shows the SINR averaged over channel fading against the geometry factor. The geometry factor is used to describe the relative distance of the terminal with respect to the desired base station and neighboring base stations. It is defined as a ratio of the total received signal power from the desired base station (including all physical

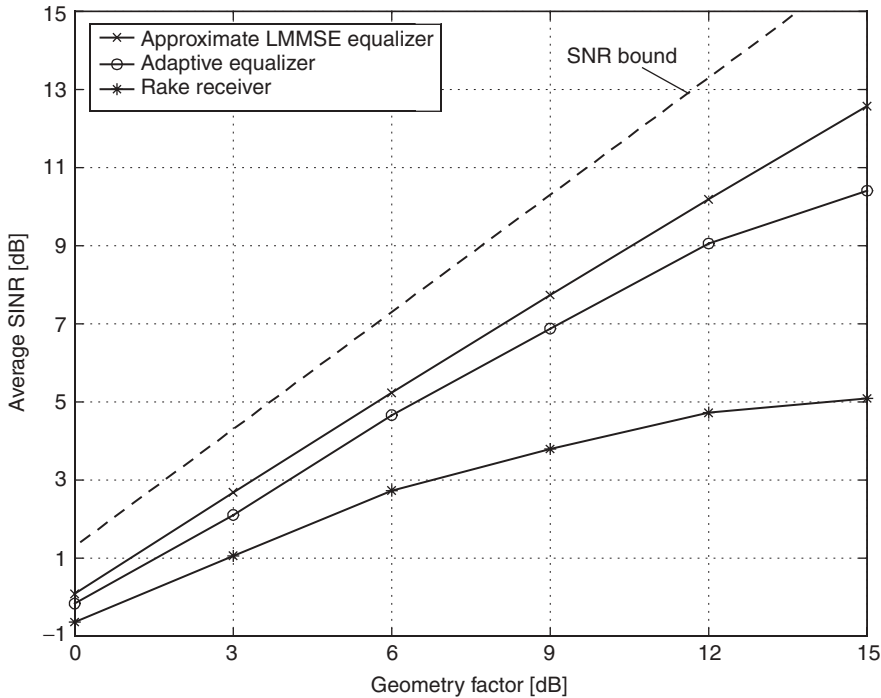


FIGURE 10.33 Performance of an approximated LMMSE equalizer, a practical equalizer, and a RAKE receiver [Hoo03].

channels) to the total received powers of all other base stations and noise. The dashed line in this figure shows the maximum achievable SNR for the channel. A theoretical infinite tap linear equalizer can achieve a performance very close to the bounds of an ideal channel with no impairments, and it outperforms a RAKE receiver in particular at high values of the geometry factor when a mobile is close to the boundaries of a cell. The figure also illustrates that implementation of such an equalizer in a mobile terminal using finite taps is also practical and is more effective than a RAKE receiver.

Results presented in Examples 10.29 and 10.30 are very strong because they are counterintuitive. In all of our analysis in a single-user multipath channel we showed that a RAKE receiver takes advantage of the in-band multipath diversity to provide optimal reception of a signal. However, the last two examples indicate that in a multiuser environment, in which channels are separated by orthogonal codes, a RAKE receiver is no longer the optimal receiver, and we are better off if we equalize the received signal to eliminate the multipath before we detect all users.

10.5.4 Multicarrier CDMA

Considering the popularity of CDMA techniques in third-generation wireless networks and OFDM in WLANs, a number of researchers have considered using multicarrier CDMA (MC-CDMA) for future cellular networks. As we discussed in Chapter 9, OFDM implementation of multicarrier modulation provides for a bandwidth-efficient

and relatively simple implementation of the receiver. The typical bandwidth of IEEE 802.11a and g OFDM systems is 20 MHz, they operate in indoor areas where the rms delay spread is about 100 ns, and there is no simultaneous use of channel bandwidth as in CDMA systems. The bandwidth of the IS-95 is 1.25 MHz, it is designed for outdoor applications with rms delay spreads that can reach 5 μ s in urban areas, and the bandwidth is shared among users using CDMA access. Therefore, receivers designed for optimal performance outdoors do not provide optimal performance in indoor areas.

Example 10.31: Multicarrier and Multipath Spread The chip duration of the IS-95 is $1/1.25 \text{ MHz} = 900 \text{ ns} \ll 5 \mu\text{s}$. As a result, in urban areas a RAKE receiver can resolve several paths to take advantage of the in-band time diversity of the signal and thus create optimal performance. The performance of the same system is not optimal in indoor areas because with an rms delay spread of about 100 ns, the receiver cannot resolve the multipath to take advantage of the internal diversity achieved by using a RAKE receiver. As indoor applications for wireless networks increase, one expects that future systems will provide a more balanced performance in indoor and outdoor applications.

The next-generation cellular networks are expected to support integrated multimedia services supporting traditional telephone service as well as high-speed Internet access in indoor and outdoor areas. Therefore, it is desirable to have a flexible system with optimal performance indoors and outdoors and the capability to support a variety of data services with significant nonsymmetric applications such as Internet access. The bandwidth of third-generation CDMA systems is several times greater than that of the IS-95 but still cannot provide significant diversity for indoor areas. Some researchers have studied the use of MC-CDMA to take advantage of the frequency diversity in different carriers and to provide the flexibility for supporting different data rates.

Example 10.32: Multicarrier and Data Rate Assume that the bandwidth of a MC-CDMA system is 20 MHz and the delay spread is between 100 ns and 4 μ s. If we set the chip duration for each carrier of the multicarrier system to $T_c = 4 \mu\text{s}$, we will not have multipath in any circumstances, so the need for a RAKE receiver is eliminated. Then the bandwidth of each carrier is $1/T_c = \frac{1}{4} \mu\text{s} = 250 \text{ kHz}$, and with a chip rate of 250 kc/s, we can support 80 carriers in our 20-MHz bandwidth. Depending on the medium access control, a variety of CDMA systems can be designed. Due to reduction in multipath arrivals, ISI, IPI, and MCI are eliminated and a multiuser detector receiver can provide a close-to-ideal performance. Similar to the IEEE 802.11, multisymbol modulations can be accommodated for high-speed data applications of up to several tens of Mb/s in short distances from the base stations.

Although MC-CDMA has a number of positive features suitable for future systems, similar to OFDM, symbol synchronization in MC-CDMA systems is challenging, and the system is sensitive to frequency offset caused by Doppler shift. In addition, cellular systems are expected to evolve around the existing systems, and dramatic changes may not be desirable. These issues have been studied by a number of researchers in the past decade. For more details on various options for implementation of MC-CDMA, readers are referred to [Sou96, Har97, Har99, Han03, Yan03]. For the effects of MC on the overall capacity of CDMA networks, one may refer to [Dea98].

10.5.5 MIMO CDMA

As we showed in Example 10.27, to support multimedia services in CDMA downlink channels, HDR services employ multisymbol modulation and M -ary orthogonal coding to achieve 2.4 Mb/s. To increase system capacity further and to enhance the quality of service, MIMO technology is being used for the next generation of CDMA networks. As discussed earlier, the performance of MIMO systems is sensitive to channel conditions. The rich scattering environment of urban and indoor areas is conducive to high-performance MIMO systems. In Section 9.8 we showed how high-performance MIMO can increase the data rate of WLANs in indoor areas. In cellular networks the conditions in urban channels give rise to less correlated fading among antenna elements, which is suitable for MIMO. As a result, using MIMO to improve the performance of third-generation CDMA systems is under consideration by standardization groups. Several MIMO CDMA techniques have been proposed for performance enhancement of third-generation systems. Table 10.4 shows summary results of one of these proposals, based on link-level simulations of a combination of V-BLAST and spreading code reuse operating over flat fading. The table gives the peak data rates achieved by the downlink shared channel using different MIMO techniques in the 2-GHz band with 5-MHz carrier spacing. Using a 4×4 MIMO, the data rate is double that of a conventional SISO channel.

10.5.6 Frequency-Hopping CDMA

The FH-CDMA technique is perhaps the easiest to visualize. In each time slot, users are assigned to distinct frequency bands, but in each successive time slot, the frequency assignments are permuted. For example, in time slot 1, user 1 occupies band 1, user 2 occupies band 2, and user 3 occupies band 3. In time slot 2, user 1 hops to band 3, user 2 hops to band 1, and so on. The modulation scheme used in an FH-CDMA system will typically be nonbinary FSK.

Let us now assume that we have M FHSS systems with independent random patterns occupying a bandwidth W , each with a data rate of R_b and a bandwidth expansion factor $N = W/R_b$. The error-rate performance of the FH-CDMA system is the same as the performance of an FHSS system in random narrowband interference. The symbol error probability without coding is given by Eq. (10.4.8) and with coding is determined from Eq. (10.4.22). As we have done earlier, let us designate one of the M users as the target user and consider the effect of the $M - 1$ interfering user signals on the target

TABLE 10.4 Peak Data Rate of MIMO for 3GPP Standard

(M, N)	Transmitter Technique	Code Rate	Modulation	Rate/Substream (kb/s)	Number of Substreams	Data Rate (Mb/s)
(1, 1)	Conventional	$\frac{3}{4}$	64QAM	540	20	10.8
(2, 2)	MIMO	$\frac{3}{4}$	16QAM	360	40	14.4
(2, 2)	MIMO	$\frac{3}{4}$	QPSK	180	80	14.4
(4, 4)	MIMO	$\frac{1}{2}$	8PSK	540	80	21.6

Source: [Ges03].

receiver. The probability that exactly k hops of the $M - 1$ simultaneous users hit the hop used by the target receiver is given by

$$P_{k\text{-hit}} = \binom{M-1}{k} P^k (1-P)^{M-1-k} \quad (10.5.12)$$

where $P = 1/N$ is the probability of having a hop at a specific hop frequency. The target receiver will have no error on a hop if no other user hits that hop frequency at the same time. This is obtained by evaluating Eq. (10.5.12) at $k = 0$, which yields

$$P_{0\text{ hit}} = \binom{M-1}{0} P^0 (1-P)^{M-1} = (1-P)^{M-1} \quad (10.5.13)$$

The probability of a hit is $1 - P_{0\text{ hit}}$ and the probability of error given a hit is $\frac{1}{2}$, which yields an overall error probability

$$P_{s,\text{FHCDMA}} = \frac{1}{2}(1 - P_{0\text{ hit}}) = \frac{1}{2} \left[1 - \left(1 - \frac{1}{N} \right)^{M-1} \right] \quad (10.5.14)$$

For a single user, $M = 1$, and the probability of error is zero because we neglect additive noise in this analysis. For two users $M = 2$ and we have $P_s = 1/2N$, which is consistent with the derivations for FHSS operation in random narrowband interference.

If power control is not used, or if the power control technique is unable to keep the power from all the receivers at precisely the same level, the effects are not as serious as in a DS-CDMA system. If the users' signals arrive with different power levels, there will be instances in which we have two hops at the same frequency but the difference between the received powers is large enough that one of the signals will survive. Analysis of this phenomenon is similar to the analysis of the effects of capture in contention-based access methods operating on fading channels, a topic we discuss in Chapter 11. Theoretically, the performance of an FS-CDMA system with received power fluctuations is expected to be slightly better than the predictions given above. However, we should note that with large fluctuations in the power we need to increase the transmitted power to achieve a negligible error rate in the absence of a hit. In wireless networks where battery-powered terminals are to be used, an increase in the transmitted power is an unattractive design choice, and for that reason power control techniques should be applied. In frequency-selective fading, when hops of different users' signals coincide or when a hop lands in a deep fade, errors will occur. In this situation the performance analysis must include the effects of faded hops as well as hits from other users.

In some applications, such as GSM, the FHSS technique is combined with TDMA. In these systems the hops are coordinated so that there are no hits among different users' signals. A system of this type, referred to as *FHSS/TDMA*, can be considered as an FD-CDMA system with nonrandom orthogonal codes. The capacity in this case is the same as the capacity of a TDMA system, with the number of user slots equal to the number of hop frequencies. With this approach, several packets are transmitted at each hop and the advantage is that only a portion of the packets for each user will suffer from the frequency-selective fading characteristics of the channel. For voice users, loss of a fraction of the packets generally does not cause a significant degradation in the

quality of the reconstructed voice. For data users, the packets lost in a hop exposed to frequency-selective fading will be retransmitted, very likely at hop frequencies where the channel is not faded. Given these considerations, the voice- and data-oriented wireless information network industries are both evaluating frequency-hopping systems for possible future application.

10.5.7 Time-Hopping CDMA

Another approach using spread-spectrum technology is time-hopping spread spectrum (THSS). This technique is used as one of the choices for the implementation of UWB systems. In THSS, similar to DSSS, time is slotted into chip intervals. However, within each chip the location of a very narrow pulse is used to encode the polarity of the transmitted chip. A sequence of chips spread with a PN code forms a transmitted information bit. If this system is observed in time, it appears like a very narrow pulse (ultrawideband) hopping in time, which is why it is called THSS. In a multiuser environment, different users adopt different PN sequences as their pattern of time hopping to form a CDMA system. A more detailed description, and examples of THSS and its form of CDMA multiuser operation, are provided in Chapter 12, where we discuss the specifics of emerging UWB systems.

QUESTIONS

- (a) Name five major advantages of spread-spectrum technology for wireless communications.
- (b) What were the major reasons for adopting CDMA technology in second- and third-generation cellular networks?
- (c) What are the differences between the FHSS used in Bluetooth and the first FHSS IEEE 802.11?
- (d) Why was spread spectrum adopted in the first IEEE 802.11 WLAN standard ?
- (e) What is the simplest method of implementing the correlator in a spread-spectrum receiver?
- (f) Is coding effective in DSSS systems operating over slow flat-fading channels? Explain.
- (g) Explain how coding can be used to improve the performance of FHSS systems operating over slowly time-varying fading multipath channels.
- (h) Why is slow frequency hopping used as an option in the GSM standard?
- (i) Is coding effective in DSSS systems operating over frequency-selective fading channels? Explain.
- (j) What is the difference between the Barker code used in the DSSS IEEE 802.11 standard and LFSR codes of length 7 or 15? Why were Barker codes selected instead of LFSR codes of length 8 or 16?
Hint: Note that the minimum processing gain defined for ISM bands was 10, and LANs always strive for higher data rates.

- (k) What are the advantages of a RAKE receiver in a frequency-selective fading multipath channel? What happens to these advantages when the channel is flat fading?
- (l) How does a matched filter using a SAW device work?
- (m) What are IPI and ISI, and how do they affect the performance of a RAKE receiver?
- (n) Name three generic methods for sending a training sequence for measurement of the channel impulse response used for maximal ratio combining of in-band diversity for RAKE receivers?
- (o) Why does Bluetooth's interference with the DSSS IEEE 802.11 cause higher packet loss than its interference with FHSS IEEE 802.11? Explain.
- (p) Why does Bluetooth have to be closer to DSSS devices than to FHSS IEEE 802.11 devices to interfere with them?
- (q) Name two major advantages of using power control in multiuser environments.
- (r) Can a DSSS WLAN operate efficiently without power control? Explain and give an example of a DSSS standard that does not mandate power control.
- (s) Can an FHSS/CDMA system operate efficiently without power control? Give an example of a FHSS/CDMA system that does not mandate power control.
- (t) What is the advantage of using a sectored antenna in a DSSS CDMA system?
- (u) How are Walsh codes used in IS-95?
- (v) What is the CCK code, and which standard is using it?
- (w) What is the difference between the multiuser detection algorithms used in forward (downlink) and reverse (uplink) CDMA channels?
- (x) Why has multicarrier CDMA attracted attention in third-generation cellular networks?
- (y) How do HDR services support 2.4 Mb/s in a CDMA network designed for 9600 bits/s per voice user?

PROBLEMS

1. The transmission bandwidth of a TV station is 5.5 MHz, and the minimum acceptable received signal level is 30 dB above the background noise. We want to transmit a BPSK/FHSS digital broadcast signal in the same band using the same transmitter antenna.
 - (a) What is the maximum transmission power for the digital broadcasting service that allows the performance of the TV station to remain acceptable?
 - (b) What is the minimum processing gain of the FHSS system if the maximum error rate of the digital broadcasting system is to stay at 10^{-5} ?
2. Repeat Problem 1 assuming that BPSK/DSSS is used. Neglect the effects of multipath.
3. (a) Sketch all eight of the 8-bit Walsh sequences.

- (b) Sketch the autocorrelation function (ACF) of all eight sequences.
- (c) Sketch the cross-correlation function (CCF) of the first and second sequences.
- 4. (a) Give the Barker code used as the PN sequence in IEEE 802.11 and draw its autocorrelation function (ACF). Assume that rectangular pulses with amplitude A and duration T are used as the chip waveform.
- (b) Draw ACF of an M -sequence of length 15 given by $[1 \ -1 \ -1 \ 1 \ 1 \ 1 \ -1 \ 1 \ 1 \ 1 \ -1 \ -1 \ -1 \ -1 \ 1]$.
- 5. (a) For the input sequence $\{01110010\}$ to an 802.11b CCK modulator, give the eight-symbol transmitted complex sequence.
- (b) Repeat part (a) for the input sequence $\{110110010\}$ and examine the orthogonality of the two encoded complex sequences.
- 6. The pulse shape

$$g(t) = A \sin^2 \frac{\pi t}{T}, \quad 0 < t < T$$

is used for shaping the chips of a DSSS system.

- (a) Show that

$$A = \sqrt{\frac{16E_c}{3T_c}}$$

where E_c is the energy in the pulse shape.

- (b) Determine the dc $G(0)$ value of the waveform.
- (c) Compare the signal-to-narrowband interference ratio of a DSSS system that uses these pulses for pulse shaping with a DSSS system that uses a rectangular waveform for pulse shaping.
- 7. We have two 100-mW BPSK/DSSS WLAN mobile terminals running streaming applications in the 2.4-GHz ISM band at the same center frequency. However, the two mobile terminals communicate with different access points (APs) in the same area. At the target AP, one of the mobile terminals communicates as an information source and the other terminal acts as a source of wideband interference. The received power from the information source and that received from the interference source are independent of one another, and they vary according to the relative location of the two transmitters with respect to the target receiver.
 - (a) If the processing gain of both systems is 10 dB and the received power from each transmitter is the same, what is the error rate at the receiver for the target transmitter?
 - (b) Repeat part (a) if the distance of the target mobile terminal from the target AP is such that the received signal strength from the interfering terminal is double that of the power of the target mobile terminal.
 - (c) Repeat parts (a) and (b) if the processing gain was 0 dB (no spreading and only BPSK modulation).
- 8. In this problem we address the near-far problem in DSSS CDMA systems. Assume that we have two 100-mW BPSK DSSS mobile terminals with processing gains of 20 dB communicating with the same base station in the 1.9-GHz PCS bands, at the same center frequency, and using two different PN sequences. One of the mobile

terminals communicates as an information source, and the other terminal acts as a source of wideband interference. There is no power control, and the received power from the information source and the interference source are independent of one another. Therefore, the received signal strength from the two terminals varies according to the relative location of the two transmitters with respect to the target receiver.

- (a) Give the average SIR as a function of the distance between the two terminals and the base station.
 - (b) Neglecting the background noise and assuming that the SIR forms an exponential random variable (Rayleigh amplitude fading), give an equation for calculation of the average error rate for the target terminal.
 - (c) Plot the average probability of error versus the ratio of the distance of the target terminal and the interfering terminal. At what relative distance ratio does the interfering terminal push the average error rate below 10^{-2} ? Assume that the distance power gradient in the area of operation is 3.
9. (a) Neglecting the frequency spectrum used for control channels, what is the maximum number of two-way IS-95 CDMA voice channels that can fit inside the frequencies allocated to the AMPS system? Assume a minimum required E_b/N_0 value of 6 dB and include the effects of antenna sectorization, voice activity, and extra CDMA interference.
- (b) Repeat part (a) for a wideband CDMA system that uses the entire AMPS band in each direction as the bandwidth of its CDMA system.
- (c) Considering the difference between the number of users supported in the two systems described in parts (a) and (b), explain why standards may migrate from original Qualcomm CDMA to a wideband CDMA system.
10. Derive an equation that gives the probability of outage versus threshold error rate of a binary DPSK-modulated CDMA system in flat fading. Assume that the effects of additive noise are negligible.
11. In this chapter we introduced a method to calculate the number of simultaneous users in a DS-SS system with perfect power control. In reality the power cannot be controlled perfectly and the SIR for user i ,

$$\gamma_i = \frac{E_{bi}}{I_0} = \frac{W}{R} \frac{P_i}{N_0 W + \sum_{j=1}^{M-1} P_j} \approx \frac{W}{R} \frac{P_i}{N_0 W + \sum_{j=1}^M P_j}$$

forms a lognormal distribution function with mean and variance of μ and σ decibels, respectively. In this equation, $P_i = E_{bi} R$ is the received power for user i and M is the number of simultaneous users in the system.

- (a) If we define the total SIR as

$$x = \frac{R}{W} \sum_{i=1}^M \gamma_{bi}$$

show that

$$E(x) = \frac{R}{W} M e^{(\beta\sigma)^2/2 + \beta\mu}$$

and

$$\text{Var}(x) = \left(\frac{R}{W}\right)^2 M e^{(\beta\sigma)^2 + 2\beta\mu} [e^{(\beta\sigma)^2} - 2]$$

where $\beta = \ln(10)/10$.

- (b) Show that the received total power-to-noise ratio is given by

$$\gamma = 1 + \sum_{i=1}^M \frac{P_i}{N_0 W} = \frac{1}{1-x}$$

- (c) If we define the received signal above noise level in decibels by $Z = 10 \log_{10} \gamma$, show that the commulative distribution function of Z is given by

$$\text{CDF} : f_Z(z) = f \int_{-\infty}^{[1-e^{\beta-E(x)}]/\sqrt{\text{Var}(x)}} e^{-y^2/2} dy$$

Let m be the median of Z and show that

$$E(x) = 1 - 10^{-m/10}$$

- (d) Using the results of parts (b) and (c), show that the following relationship exists between the median of the received signal above the noise background and the simultaneous number of users:

$$M = (1 - 10^{-m/10}) \frac{W/R}{e^{(\beta\sigma)^2/2 + \beta\mu}}$$

- (e) Compare the results of part (d), assuming that the background noise is negligible ($m \rightarrow \infty$), with the result of derivations with perfect power control analyzed in this chapter. Assume that $\mu = 7.9$ dB and $\sigma = 2.4$ dB [Pad94].

PART IV

SYSTEMS ASPECTS

In Part I we provided an overview of wireless systems. In Parts II and III we provided details of channel characterization and modem design technologies. In this part we address issues related to system engineering and provide examples of technologies and systems adopted for the emerging wireless sensor and ad hoc network industries.

Chapter 11: Topology, Medium Access, and Performance

In this chapter we turn our attention from links to networks, and we examine the topologies, access methods, and methods used for performance evaluation of wireless networks. We discuss the basic topologies used in wireless networks, with particular attention to the design of cellular topologies. We discuss centrally controlled frequency-, time-, and code-division medium-access methods originally designed for voice-oriented circuit-switched networks, as well as ALOHA contention-based medium-access methods originally designed for packet-switched wireless data communications. We also discuss metrics used for performance evaluation of voice- and data-oriented wireless networks.

Chapter 12: Ultrawideband Communications

In this chapter we provide an overview of UWB techniques and describe the emerging technologies for the UWB frequency bands. We first describe issues that distinguish UWB radio signal propagation, and then discuss adjustments in the wideband channel models (introduced in Chapter 6) that are needed to accommodate the special characteristics of UWB transmission. Finally, we describe the implementation of impulse radio, direct-sequence UWB, and multiband OFDM techniques in UWB bands.

Chapter 13: RF Location Sensing

Radio-frequency location-sensing techniques are the enabling technologies for the implementation of next-generation location-aware wireless networks. To lay the groundwork for an understanding of these techniques, we present the methods used for ranging, positioning, and modeling the behavior of location sensors. Ranging methods include time of arrival, angle of arrival, and received signal strength measurement algorithms. The algorithms employ either traditional least-squares techniques as used in global positioning systems, or pattern recognition techniques that take advantage of calibration measurements and building layout.

Chapter 14: Wireless Optical Networks

This chapter is devoted to the principles of infrared (IR) communications. IR technology became dominant in low-speed remote control applications and later attracted considerable attention for high-speed WLAN applications. Both diffused and directed beam IR communications have been the subject of considerable research in the past three decades. Therefore, understanding of principles of IR communications is a valuable asset for those studying wireless networks. The chapter begins with a discussion of the incentives for using IR and of issues related to its implementation and eye safety. Then we address the behavior of the channel and applied modulation and medium-access control techniques.

Chapter 15: Systems and Standards

In this chapter we describe the digital wireless systems that have been achieving the greatest commercial success in the wireless industry. We categorize the systems as either high-power wide-area systems or low-power local-area systems. Today, the dominant wide-area, high-power wireless systems are based on either GSM and its derivative standards, or on code-division multiple-access cellular standards. Deployed WLAN systems are essentially all based on the IEEE 802.11 family of protocol standards. Here we describe the principal characteristics of these systems and provide a brief historical overview of some of the earlier digital wireless systems that were developed and implemented but that have been supplanted or overtaken by the systems now predominating in the marketplace.

11

TOPOLOGY, MEDIUM ACCESS, AND PERFORMANCE

- 11.1 Introduction
- 11.2 Topologies for Local Networks
 - 11.2.1 Comparison of Local Network Topologies
- 11.3 Cellular Topology for Wide-Area Networks
 - 11.3.1 Network Planning for Cellular FDMA and TDMA
 - 11.3.2 Network Planning for CDMA Cellular
 - 11.3.3 Planning for Large-Scale WLANs
- 11.4 Centrally Controlled Assigned Access Methods
 - 11.4.1 Frequency-Division Multiple Access
 - 11.4.2 Time-Division Multiple Access
 - 11.4.3 Time-Division Duplex
 - 11.4.4 Comparison of FDMA and TDMA
 - 11.4.5 Code-Division Multiple Access
 - 11.4.6 Traffic Engineering for Circuit-Switched Controlled Access
 - 11.4.7 Packet-Switched Data in Central Control Access
 - 11.4.8 High Data Rates in Cellular Networks
- 11.5 Distributed Contention-Based Access Control
 - 11.5.1 ALOHA-Based Access Control
 - 11.5.2 CSMA-Based Access Control
 - 11.5.3 Wireless Access and Radio Channel Characteristics
- Questions
- Problems
- Project
 - Project 1. Performance of IEEE 802.11b/g WLANs

11.1 INTRODUCTION

In Chapters 3 through 10 we concentrated on the characteristics of individual point-to-point wireless links and the technologies that are used for digital communication over these links. In this chapter we turn our attention from links to networks, and we

examine both the topologies of networks and the wireless access methods that are used to enable a number of users to communicate with one another across the network.

Just as the nodes in a wired computer network can be connected in a variety of ways, the terminals in a wireless network can interconnect using a variety of topologies. The reader familiar with the configurations of wired communication networks such as the public telephone network, Internet, and wired local area networks (LANs) will find a few similarities, but also several important differences, between wireless and wired networks. The differences are due to the fact that wireless communications is fundamentally a broadcast medium. One consequence of this characteristic is that since terminals are not connected by wired circuits, transmitted message can in principle be received by an arbitrary (and perhaps unknown) number of other users. At the same time, given the absence of wired connections and the uncertainties of signal propagation, one cannot always be guaranteed a link from every transmitting terminal to every intended receiving terminal. Another consequence of the broadcast nature of wireless communications is that any user, in transmitting a message to any other user, is utilizing a scarce resource (some portion of the system bandwidth, perhaps the entire system bandwidth) for the duration of the message. Therefore, means have to be provided for the fair and efficient utilization of available bandwidth. Yet another consequence of the characteristics of wireless communications is that transmitted signal power becomes an important parameter in the operation of the system. Enough signal power must be used to achieve reliable communication from terminal to terminal, but care must also be given to the life of the battery in a mobile terminal and to avoid excessive interference with other networks operating in the same frequency band. These considerations have led to the use of a variety of topologies and medium access control techniques in wireless networks, the focal point of this chapter.

Topologies used in broadband local- and wide-area wireless networks have different characteristics and considerations. There are two popular types of topologies for broadband wireless local networks: centralized or infrastructure and ad hoc or peer-to-peer. In this chapter we describe these topologies and compare them to one another. Wireless networks covering wide areas use cellular topology that employs resource (frequency or code) reuse strategies. We describe network planning and capacity assessment techniques for cellular telephone and WLAN networks. As is the case in any area of system design, there are advantages and disadvantages with each configuration, and these are outlined and discussed.

There are two classes of medium access control used in wireless networks: centrally controlled assigned access and contention access. The basic *centrally controlled access techniques* are FDMA, TDMA, and CDMA, which were originally designed for second-generation cellular telephone and circuit-switched applications. However, with the growth in popularity of the Internet in the mid-1990s and the emergence of third-generation standards, modifications were made to these access methods to accommodate high-speed packet-switched data. *Contention access methods*, designed for bursty data networking in a packet-switched environment, are the dominant access method for WLANs but are also used for random access to wide-area cellular networks. Contention-based techniques can be divided into *ALOHA-based*, used primarily in wide-area networking, and *CSMA-based access methods*, the choice for IEEE 802.11 WLAN access. We provide a general description of these access methods, show how we can evaluate their performances, and analyze the effects of radio transmission characteristics on their performance.

An ultimate requirement in designing a wireless network is to allow users to communicate at rates comparable to the rates achievable in wired networks for the same category of application. This implies rates close to broadband Internet for mobile data applications and rates close to wired-LAN rates for WLANs. To approach or reach these goals, efficient channel access methods must be provided, and we find that different access methods are appropriate for use with different network topologies. In Section 11.2 we discuss the basic topologies used in local wireless communications networks, and in Section 11.3 we give particular attention to the design of cellular topologies. In Sections 11.4 to 11.6 we provide detailed descriptions of various access methods used in wireless networks. Section 11.4 is devoted to centrally controlled medium access methods originally designed for voice-oriented circuit-switched networks. In Section 11.5 we provide insight into contention medium access methods originally designed for packet-switched wireless data communications. In Section 11.6 we analyze the effect of fading on the performance of these access methods.

11.2 TOPOLOGIES FOR LOCAL NETWORKS

Two fundamental types of topologies are used in wireless local networks: (1) centralized or infrastructure network and (2) adhoc or peer-to-peer network, both shown in Fig. 11.1. In the *centralized configuration* (Fig. 11.1a), station 1 serves as the *access point* or *hub* of the network, and the user stations are located at the ends of the connections. Any communication from one user station to another (i.e., between peers) goes through the hub. The hub station controls the user stations and monitors what each station is transmitting. Thus, the hub station is involved in managing access to the network's allocated bandwidth by user stations. There is no provision in the centralized topology for direct peer-to-peer communication. The centralized topology is a common one in telephone networks, where all the subscribers in a local exchange area are connected through the central office switch. The star topology for WLANs is another example of a centralized configuration. Cellular mobile telephone systems use a centralized network configuration to serve the mobile terminals currently operating within the coverage area of any single cell base station. In the WLAN industry, the IEEE 802.11 standard uses centralized configuration.

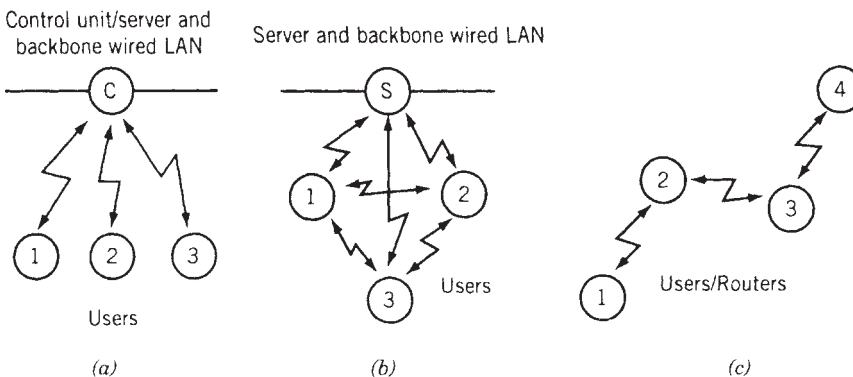


FIGURE 11.1 Wireless network topologies: (a) centralized network; (b) distributed network; (c) multihop meshed network. Circles denote wireless terminals, and lines denote communication paths.

Figure 11.1*b* and *c* include two variations of the *peer-to-peer network topology*. Figure 11.1*b* shows a *fully connected network*, in which, as the name implies, every user terminal has the functional capability to communicate directly with any other user terminal. In some systems, where users may be distributed over a wide area, any user terminal may be able to reach only a portion of the other users in the network, due to signal blockage or transmitter power limitations. In this situation, user terminals will have to cooperate in carrying messages across the network between widely separated stations. Networks designed to function this way, called *multihop meshed networks*, are shown in Fig. 11.1*c*. A fully connected meshed network is practical in an application where connectivity can be assured between every pair of terminals in the network, regardless of the site in which the network is installed. This is often feasible in WLAN applications, and, in fact, IEEE 802.11 also uses the fully connected network configuration. The multihop network configuration was used in military tactical networks, where reliable communication must be provided under unpredictable propagation conditions and over widely varying geographic areas.

11.2.1 Comparison of Local Network Topologies

There are a number of issues related to connectivity among user terminals in a network that bear on the choice of a topology for the network. We discuss briefly several key issues in relation to the two types of network topology.

Centralized or Infrastructure Topology. One important advantage of the centralized topology is that the network can be designed to operate with relatively efficient use of signal transmission power. For example, in contrast to a fully connected peer-to-peer network, user stations in a centralized network can reach stations at twice the distance with the same signal power. A related advantage is that the hub station can be placed in a location that is optimized with respect to the set of user stations. For example, in a WLAN application, the central hub or access point (AP) might be placed in an appropriate location to facilitate unobstructed propagation between user stations and the hub. For the IEEE 802.11, an AP is often installed on a ceiling or is located high on a wall. In the centralized configuration, the central hub also provides a natural point at which connection is made to a backbone network. Another advantage of this topology is that the user terminals can be made functionally simple, whereas the more sophisticated control functions are concentrated in a single access point or base station.

Power control is an effective means of minimizing the radiated power of individual user terminals, and this, in turn, conserves battery power and controls interference. Power control is essential in code-division multiple-access (CDMA) networks, where it is a key ingredient in achieving high levels of bandwidth efficiency. Power control also increases the capacity of time-division multiple-access (TDMA) cellular networks by minimizing co-channel interference from other cells. In applications where battery-operated portable units are to be used, power control serves the important requirement of maximizing battery life. In WLANs, the control of battery life is provided by implementation of the sleep mode. Terminals in sleep mode wake up periodically to see if there is any information available for them at the access point. If power control is to be employed in a radio network, a centralized configuration is necessary. The central unit in a network also provides a home for other centralized control functions, such as provision of a common timing reference among the terminals.

One disadvantage of a centralized topology is the presence of a single failure point. If the central control module fails, the entire network is disabled. This is in particular important in application environments such as battlefields or disaster-recovery operations, where the enemy or the instigator of the disaster may target the access point. Another less important disadvantage is the delay characteristic of the network. Because all user-to-user communications go through the hub, the store-and-forward delay is twice that of a fully connected peer-to-peer network. Another measure of communication efficiency is channel occupancy, measured in hertz-seconds (bandwidth $\times$ time). This parameter is gaining favor as a measure of channel occupancy, quantifying the fair time sharing of a common bandwidth among multiple users. This measure represents the volume of frequency that is occupied for the duration of the transmission. If we use this measure, we see that occupancy with a centralized network is twice that of a fully connected peer-to-peer network, because each message is broadcast twice.

An additional disadvantage of the centralized topology is that it does not offer the functional flexibility needed to deal with unpredictable propagation environments; nor can it cover wide areas where user-to-user connections can exceed the range of a single link in the network. The centralized topology is not suitable for multihop ad hoc networks, where users might want to use laptops for file transfer or multitask sharing, without networking arrangements having already been made.

Fully Connected Ad Hoc Topology. An important advantage of the fully connected topology is that it has no single point of failure. Another advantage is that peer-to-peer messages do not suffer the store-and-forward delay of centralized topology, and thus time delay and channel occupancy as measured by hertz-seconds are both halved. Given that no routing functions need be implemented in any of the stations, the complexity of equipment in this design can be minimized. Connection to a backbone network is provided by an additional terminal acting as a server. The server also acts as a bridge or gateway to convert communication protocols from the WLAN to the backbone network. If many of the network nodes are equipped with this connection capability, implementation complexity and cost become issues. A major disadvantage of a fully connected peer-to-peer network is that performance will degrade, or enhanced levels of transmitter power will have to be provided, in operation across large networks. A related issue is the presence of a *near-far problem*, owing to the fact that transmitters that need to operate at a high power level can interfere with unintended receivers in close proximity to the transmitting station. However, the fully connected topology offers an attractive alternative for small networks where reliable connectivity can be ensured between all pairs of user stations. The peer-to-peer ad hoc network topology is one of the two optional topologies of the IEEE 802.11 standard and is also featured in some voice-oriented services, such the Japanese PHS system, introduced in the early 1990s. The ad hoc peer-to-peer topology should prove attractive for applications where a limited number of low-power personal computing or communication devices, such as laptops or wireless phones, can be networked without prior arrangements. The nodes need not be complex, and they can be networked in the absence of a central unit or server.

Multihop Ad Hoc Topology. A major advantage of the multihop topology is power efficiency, which derives from the fact that message transmission between widely separated users is accomplished with multiple shorter hops. In some applications, such as military tactical communication over a wide operational area, multihop networks

provide the only practical approach to achieving reliable connectivity among mobile users. Multihop topologies, which first found important applications in military radio and public safety communications networks, have attracted considerable attention in recent years. One pioneering example of a multihop network design for military application was the U.S. Army Mobile Subscriber Equipment (MSE) network introduced in the mid-1980s [Sch84, Li87]. Multihop networks using repeater stations were also used in the land-mobile radio industry, where, for example, networks are required to serve municipal and state public safety organizations over wide geographic areas. In the mid-1990s, HIPERLAN standardization considered using multihop topology for WLANs, and in the early 2000s, multihop ad hoc networks became desirable for WLAN and WPAN in commercial and military applications. Multihop mesh networks are aiming at extending WLAN coverage for outdoor applications. Multihop protocol also provides for flexible self-healing networks.

As with a fully connected network, in multihop ad hoc networks, connection to a backbone or other networks is provided by equipping one or more nodes with the appropriate connection capability. One disadvantage of multihop topology is the added complexity needed in user terminals to implement efficient message routing and control algorithms. As a result, implementation of location-aware and power-efficient routing algorithms for ad hoc networks became a major research area in the early 2000s. A further disadvantage of multihop ad hoc networks is the accumulated store-and-forward delay incurred by multiple hops connecting widely separated users. Associated with the store-and-forward capability is a considerable amount of transmission overhead carried with transmitted messages.

The main obstacles in the implementation of reliable ad hoc networks can be divided into two categories: issues regarding the topology control and issues related to data transmission. For reliable topology control, one needs to answer a number of questions, such as how to discover neighbors and their locations, how to establish a link and a cluster with the neighbors, and how to control sleeping modes used for power control for individual terminals among all the nodes. Data transmission in ad hoc networks faces another set of challenging questions, including how to route the message from the source to the destination with minimal traffic load and power consumption, how to maintain a reasonably accurate location of the terminals, and how to handle broadcast and selective broadcast of messages in a dynamically changing topology. These issues become more complicated when the network is expected to provide a different quality of service to users. As a result, with the recent interest in ad hoc networks for military and commercial applications, a number of researchers have been engaged in this area.

11.3 CELLULAR TOPOLOGY FOR WIDE-AREA NETWORKS

Another important network topology is that of cellular networks employing frequency reuse. This was the network architecture employed in cellular mobile telephone networks, personal communication networks, mobile data networks, and more recently, is spreading into WLANs. This network configuration is particularly well suited to serving large numbers of mobile users operating over a much wider geographic area than coverage of an individual base station or access point. In cellular systems, the large service area is divided into smaller areas, each served by a fixed cell site. The cell sites are distributed in an approximately regular geometric pattern to cover the entire

service area with the level of signal to interference needed for acceptable service quality. Each cell in effect is a centralized network, with all communications to and from a mobile user in the cell area passing through and controlled by the cell base station or access point. In cellular networks, however, the base station (BS) or access point (AP) is connected through a wired backbone network to other network elements, providing for implementation of sophisticated network control and management algorithms. In a typical voice-oriented cellular network, as we will see in Chapter 15, BSs are wire-networked with base station controllers (BSCs) and mobile switching centers (MSCs). The BSCs and MSCs control call establishment, manage cell-to-cell handoffs as mobile users move about the system, and provide a connection to the public switched telephone network. Creating a hierarchy for network management and control facilitates the scalability of the network and reduces the complexity of the BS or AP. As described in Section 11.2, the WLAN standards originally designed for local coverage supported only centralized and ad hoc topologies. More recently, with rapid growth of the popularity of WLANs, a need developed to scale up the coverage area size of typical networks for corporate and outdoor applications, where a number of APs have to coordinate to provide reliable service in a wide area. In response to this need, a number of companies are designing network management and control elements for the wired backbone of WLANs to allow wider coverage and more coordinated operation of a large-scale network using numerous simple APs.

Cellular networks are designed with frequency reuse so as to maximize the overall capacity attainable with a given set of frequency channel allocations. A typical cellular frequency reuse scheme is shown in Fig. 11.2. In the figure, the cells are organized into seven-cell clusters, with the cells in a cluster designated as A to G. Seven different sets of frequency channels are used in each cluster, one set in each cell. At a sufficient distance from any cell, the same set of frequencies is used simultaneously.

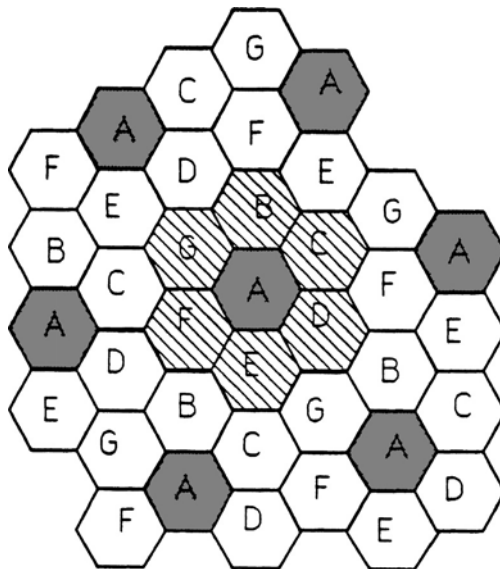


FIGURE 11.2 Typical cellular frequency reuse pattern, with seven-cell clusters. Cells with the same-letter designation use the same set of frequencies simultaneously.

The number of cells in each cluster is called the *frequency reuse factor* of the cellular system. Cells using the same frequency are called *co-channel cells*. It can be shown that the relation among the frequency reuse factor K , distance between two co-channel cells D , and radius of a cell R is given by

$$\frac{D}{R} = \sqrt{3K} \quad (11.3.1)$$

in which K can only take values from $i^2 + ij + j^2$, where i and j are integers (i.e., $K = 1, 3, 4, 7, 9, 12, 13, 19, \dots$). In other words, only for these values of K can we find a repeatable pattern for the clusters, which is essential for the implementation of a cellular topology. Therefore, the designer of a cellular topology has to find the smallest value for K among the set given above that provides an acceptable level of interference for the transmission techniques adopted in that system.

Using a cellular design scheme, the overall system capacity can, in principle, be made as large as desired by steadily reducing the area of each cell while controlling power levels to avoid *co-channel interference* (i.e., interference to other users operating in another cell on the same frequency channel). The existing cellular systems are evolving into a hierarchical cellular architecture in which *macrocells* have radius on the order of several kilometers to cover major travel routes and wide urban and suburban areas, *microcells* have coverage on the order of a few hundred meters to cover the streets in densely populated downtown areas, and *picocells* cover several rooms in an indoor areas. In the WPAN industry, networks operating in picocells are referred to as *piconets*. Another addition to the hierarchical cellular architecture are *megacells*, used primarily in satellite networks, covering nationwide areas with ranges of hundreds of kilometers.

11.3.1 Network Planning for Cellular FDMA and TDMA

Next, we examine some of the basic factors determining communication performance in cellular networks. The fundamental consideration is that communication performance in cellular networks is interference limited rather than noise limited. That is, cellular networks are generally designed in such a way that additive noise is low enough that the performance of any receiver is limited by the level of interference to the desired signal. In general, the interference is a combination of co-channel interference and multipath or intersymbol interference. If adaptive equalizers are used, the interference is primarily co-channel interference. The sketches in Fig. 11.3 depict three nodes in a network with the desired signal propagating from node 3 to node 2, and an interfering co-channel signal propagating from node 1 to node 2. In Fig. 11.3a, both signals arrive at the receiver without obstruction. In Fig. 11.3b, the intended signal is attenuated by an obstruction, whereas the interfering signal arrives without obstruction. The intended and interfering transmitters are located at distances R and D , respectively, from the receiver. Now, in a manner similar to Example 10.15, the carrier-to-interference ratio C/I is found using the power–distance relationship discussed in Chapter 4; that is,

$$\gamma_I = \frac{C}{I} = \frac{\text{power in the desired signal}}{\text{power in the interfering signal}} = \frac{A_d R^{-\alpha_d}}{A_i D^{-\alpha_i}}$$

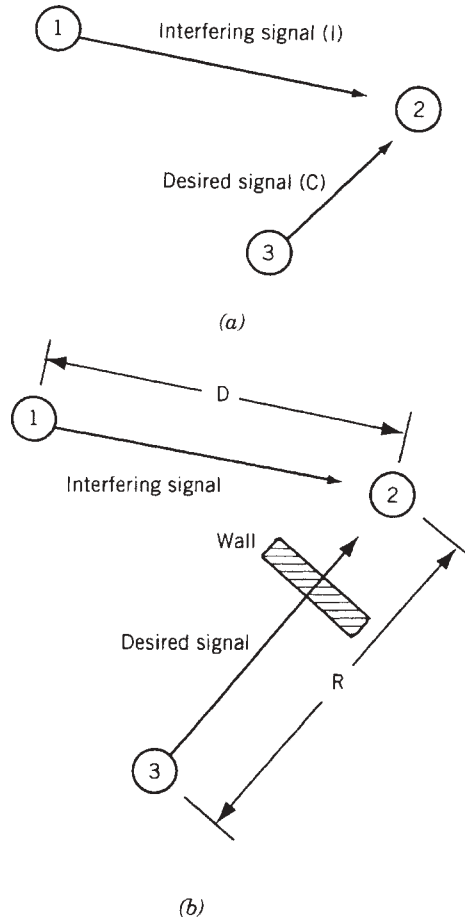


FIGURE 11.3 Signal reception with co-channel interference: (a) desired and interfering signals; (b) the desired signal is attenuated by an obstruction, while the interfering signal is unobstructed.

where α_d and α_i are power–distance-law gradients on the desired and interfering paths, respectively, and A_d and A_i are the received powers at 1 m from the desired and interfering terminals, respectively. If the attenuation factors and the received powers at 1 m are the same along the two paths and equal to α , the carrier-to-interference ratio is simply

$$\gamma_I = \frac{C}{I} = \left(\frac{D}{R} \right)^\alpha \quad (11.3.2)$$

The level of C/I that the intended receiver requires for acceptable performance is a function of the system design. The required C/I level in traditional second- and third-generation cellular networks is around 10 dB, and its exact value will depend on the modulation and coding techniques adopted and the bit-error-rate (BER) quality needed for the intended application. In designing a cellular layout, we take the distance R to be the radius of a cell: that is, the maximum range for the mobile user relative to the cellular base station. We then take D to be the spacing between cell sites operating

at the same frequency, which would represent the average distance from an interfering transmitter in that cell. It can be seen from Eq. (11.3.2) that because C/I depends on the ratio of the two distances, D and R can be scaled down proportionally, allowing greater reuse of the available frequencies and correspondingly greater overall system capacity. Another factor affecting C/I in cellular networks is the number of interferers with the same distance from the target cell, N_I . For example, in Figure 11.2 there are six A cells with the same distance from the central A cells interfering with that channel ($N_I = 6$). Therefore, for a cellular network, C/I is given by

$$\gamma_{C/I} = \frac{1}{N_I} \left(\frac{D}{R} \right)^\alpha \quad (11.3.3)$$

Example 11.1: Carrier-to-Interference Noise in an AMPS System In this example we calculate the carrier-to-interference noise for AMPS systems operating in suburban areas using a frequency reuse factor of $K = 7$. Using Eq. (11.3.1) for this system, we will have $D/R = \sqrt{3 \times 7} = 4.58$. As we showed in Example 3.2, the distance–power gradient for macrocell areas with two dominated paths is $\alpha = 4$. Using omnidirectional antennas, as shown in Fig. 11.2, we have $N_I = 6$. If we substitute these values in Eq. (11.3.3), C/I for this system is $\gamma_{C/I} = \frac{1}{6}(\sqrt{3K})^4 = 73.5$ (18.7 dB). Therefore, in the corners of the cell when the mobile station (MS) is farthest from the base station (BS), the received signal-to-interference ratio is above 18 dB, which provides an acceptable quality of the voice for analog cellular telephone users.

Example 11.2: Carrier-to-Noise Requirement in TDMA Digital Cellular In the deployment of digital cellular networks such as GSM, a frequency reuse factor of 3 is often used. Following the same method as in Example 11.1, the carrier-to-noise ratio for suburban areas is $\gamma_{C/I} = (1/N_I)(\sqrt{3K})^4 = 13.5$ (11.3 dB). Considering Fig. 7.10 for QPSK modulation, we can have an error rate of 10^{-6} for an SNR of around 10.5 dB. Therefore, 11.3 dB provides a very reasonable performance when we use this approximation. In practice, these systems often use three-sectored antennas, which reduces interference from the co-channel to $N_I = 2$. For these values, C/I is increased to $\gamma_{C/I} = (1/N_I)(\sqrt{3K})^4 = 45.5$ (16.6 dB), allowing approximately 6 dB more for other impairments, caused by such practical considerations as imperfect modem design and imperfect installations.

Capacity of FDMA and TDMA Cellular Networks. In FDMA and TDMA networks, the available bandwidth to the network is divided into a number of carriers. In FDMA systems such as first-generation analog cellular networks, each carrier carries one user channel. In TDMA networks similar to the FDMA, the total available bandwidth is divided into a number of carriers, but each carrier carries several user channels. The general approach for cellular planning is to divide the number of carriers by the frequency reuse factor to determine the number of carriers per cell. The number of simultaneous users per cell or the capacity per cell is then determined by multiplying the number of carriers per cell by the number of users per carrier. Therefore, the capacity per cell in the TDMA channels is given by

$$M = \frac{W}{W_c} \frac{N_u}{K} \quad (11.3.4)$$

in which W is the total available bandwidth to the network provider, W_c the bandwidth per carrier, N_u the number of simultaneous users per carrier, and K the frequency-reuse factor of the network. For FDMA networks the same equation holds, with $N_u = 1$. Apparently, the larger the frequency-reuse factor, the lower the number of channels per cell and consequently, the lower the capacity of the cellular network. Therefore, it is desirable to keep the frequency-reuse factor as low as possible.

Example 11.3: Comparison of AMPS, IS-136, and IS-95 In Example 10.22 we showed that for the IS-95 CDMA, using a carrier bandwidth of $W = 1.25$ MHz, the number of simultaneous users per cell can be approximated by 108. For an IS-136 system with a carrier bandwidth of $W_c = 30$ kHz, number of users per carrier $N_u = 3$, and frequency-reuse factor $K = 4$, each $W = 1.25$ MHz of bandwidth (equivalent to one IS-95 carrier) provides for

$$M = \frac{W}{W_c} \frac{N_u}{K} = 31.25$$

users per cell. For an AMPS analog system with a carrier bandwidth of $W_c = 30$ kHz and a frequency-reuse factor of $K = 7$ (commonly used in these systems), each $W = 1.25$ MHz of bandwidth provides for

$$M = \frac{W}{W_c} \frac{1}{K} = 6$$

users per cell. Therefore, use of IS-136 increases the capacity relative to AMPS more than fivefold. A CDMA system can potentially increase this performance approximately another threefold. In practice, the IS-95 standard has 64 channels, approximately twice that of a typical IS-136 system.

11.3.2 Network Planning for CDMA Cellular

The CDMA spread-spectrum cellular systems reuse the same frequencies in adjacent cells. The base stations in a CDMA cellular system cover designated areas and are connected using a wired backbone network. These setups can be considered as cellular with a frequency-reuse factor of 1. Figure 11.4 shows the frequency-reuse concept for frequency-division multiple-access (FDMA) and CDMA cellular systems. With

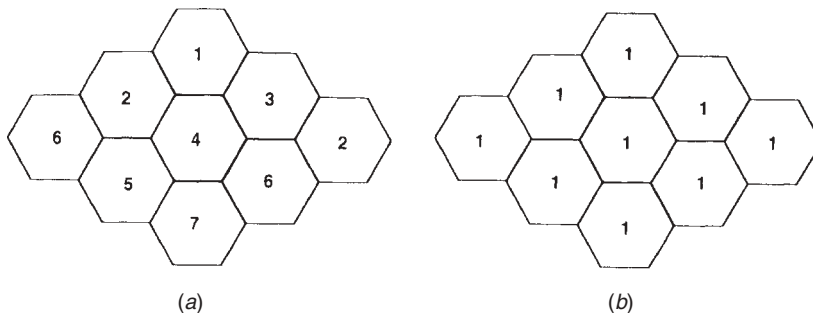


FIGURE 11.4 Cellular frequency reuse patterns: (a) FDMA reuse; (b) CDMA reuse.

FDMA, with a frequency-reuse factor 7, the available frequency band is divided into seven subbands used in different cells, as we described earlier. With CDMA, all cells use the entire band and the frequency-reuse factor is 1.

Path-loss models and coverage calculations for FDMA and TDMA systems are very similar to the path-loss models used for CDMA systems. However, system aspect issues related to the deployment of CDMA systems are different from those for FDMA and TDMA networks. All cellular networks are interference-limited systems; however, the sources of interference and the way they affect the performance of a system are different. In FDMA systems the main sources of interference are co-channel interference (CCI) and adjacent channel interference (ACI). The ACI is handled by avoiding assignment of the neighboring carriers to the same cell, leaving CCI as the main source of interference. In TDMA systems, in addition to CCI and ACI, we have intersymbol interference (ISI). By handling ACI with frequency assignment and ISI with adaptive equalization, CCI remains the main source of interference. Therefore, the dominant source of interference for both FDMA and TDMA systems remains CCI, which we actually used in the derivation of capacity in these networks, provided in Eq. (11.3.3).

Capacity of CDMA Cellular Networks. In FDMA and TDMA systems, to adjust the quality of the received signal-to-interference ratio for users, we change the frequency-reuse factor. In DSSS networks used in second- and third-generation systems, the frequency-reuse factor is always 1 and cannot be adjusted. In addition, the dominant noise involved in calculation of the capacity of the network is no longer CCI. As we described in Chapter 10, in CDMA systems, in addition to CCI, ACI, and ISI, we have multiuser interference (MUI) and interpath interference (IPI) plus the impact of imperfect power control, which disturbs the balanced distribution of interference among all the terminals. For simplified calculation of the capacity of the CDMA in Eq. (10.5.5), we assumed the dominant source of the noise to be MUI. In these calculations we have not taken into account the regions between two cells or sectors of an antenna, where soft handoff is taking place and one user occupies two channels and produces interference in two cells. Then, in Eq. (10.5.7),

$$M \simeq \frac{N}{\gamma} \frac{G_A G_V}{H_0}$$

we introduced three practical considerations affecting performance, one of which, H_0 , reflects the negative effects of CCI from neighboring cells and across the sectors of a single antenna. Equation (10.5.8) used empirical data to provide a more accurate account of the situation. Although Eq. (10.5.7) was helpful in Examples 10.22 and 11.3 to compare the capacity of CDMA systems to TDMA and FDMA, it cannot differentiate the performance of CDMA and WCDMA because it neglects the effects of ISI and IPI. To further clarify the situation and the effects of IPI and ISI, we provided some examples in Sections 10.5.2 to 10.5.4. In summary, CDMA network planning is far more complicated than FDMA and TDMA network planning, and this gets even more complicated when we integrate the bursty data packets with multirate multisymbol modulations with traditional voice applications.

11.3.3 Planning for Large-Scale WLANs

When WLAN standardization activities began in the late 1980s, the main issue related to installation was the selection of topology, not large-scale cellular deployment,

because WLANs were thought of as an extension to wired LANs, one that avoids wiring difficulties. With the popularity of Internet access beginning in the mid-1990s, growth of the WLAN industry, and thoughts on integrating WLANs into third-generation cellular systems, more attention was paid to large-scale WLAN deployment for wide-area coverage. Since the commercially successful cellular telephone networks were deployed very carefully based on relatively accurate channel models for path loss, research efforts in the systematic deployment of WLANs in large areas using automated coverage prediction software attracted attention at that time [For95]. In practice, WLANs were installed in small-area applications, such as a residence or small shop, randomly by the user in the most convenient location to connect the backbone Internet connection points to a cable or DSL modem. In large-area applications, such as wireless mobile access inside a large office building or warehouse area, or outdoor deployments for wireless Internet access, WLANs were deployed in grid formation. Access points were installed every 20 to 30 m inside a building in convenient locations such as corridors and large open areas where most traffic was expected. In outdoor applications, APs were installed in wider grids on top of utility poles, high on the outside walls of multifloor buildings, or on the roofs of buildings to optimize coverage.

Figure 11.5 illustrates basic idea behind user deployment, grid installation, and coverage optimization methods for deployment of WLANs. Figure 11.6 shows a practical example for user and grid installation of WLANs in a row of shops in a single-floor shopping mall. The location of each AP is denoted by an asterisk and a dashed rounded area shows the coverage of that AP on the floor of the mall. Shop owners install user-deployed APs inside shops, producing irregular coverage on the mall floor. A grid installation was deployed by the building owner or independent service provider in the hallways, where it provides excellent coverage with a large overlap between adjacent cells and a very low outage probability. The basic difference between user deployment and grid installation is that in a grid installation some primitive network planning by visual inspection, measurement of the RSS in selected locations, or study of the construction drawings for the building is done to provide better coordination among the APs. In user deployments the AP positions are decided based primarily on convenience of installation. An increase in infrastructure density clearly reduces the impact of user

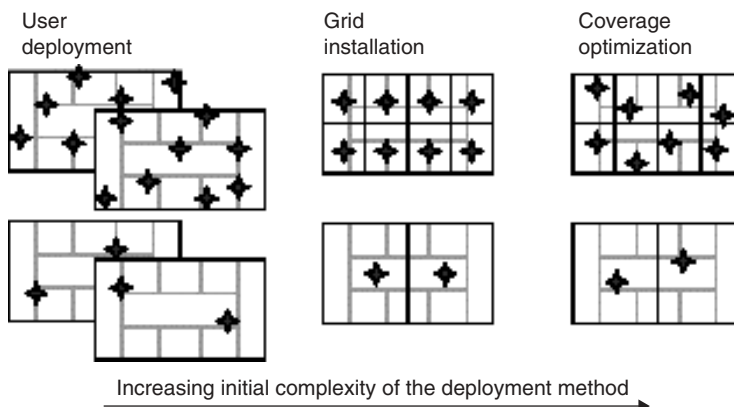


FIGURE 11.5 Various deployment strategies for WLANs [Unb02].

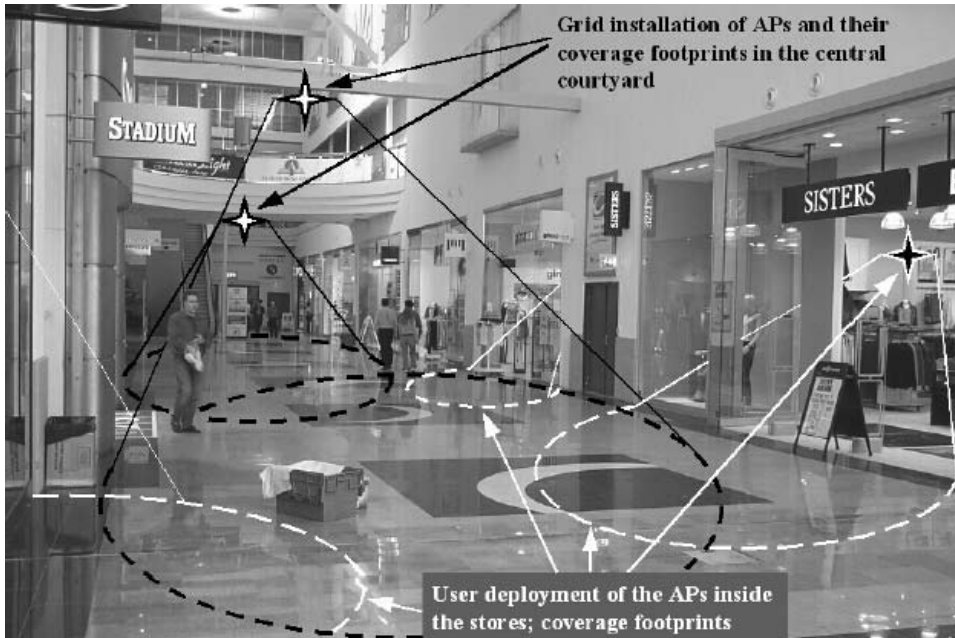


FIGURE 11.6 Example of user installation and grid installation in a shopping area [Unb02].

deployment on system performance. Figure 11.7 shows an outdoor grid installation in a street in a typical downtown or industrial area. Here we need specialized technicians and coordination with the city and with utility companies for the installation of APs on the tops of lampposts. Extensive wiring would also be needed to connect the APs to a backbone network. Physical installation using a coverage optimization technique will look very similar to a grid installation.

Before we compare the performance of the three approaches for the large-scale installation of WLANs, we show the differences among issues related to the deployment of WLANs and cellular telephone networks. WLANs operate in unlicensed bands, whereas cellular telephone networks use licensed bands. As we showed in Section 11.3, network capacity in FDMA and TDMA depends on the frequency-reuse factor of the deployment, which is determined by calculation of interference from neighboring cells. For CDMA networks we also had a parameter in the calculation of capacity that was related to interference from neighboring cells, which again relates capacity to interference. All these calculations are based on the assumption that the band is licensed to one operator, which technically means that the network planner had control over interference. WLANs operate in unlicensed bands in which a network planner does not have control over interference. Network managers on a university or corporate campus often restrict students or employees from use of WLANs other than those installed by the university or corporation, to control interference, which does not comply with government regulations on using these bands. For optimal deployment, cellular networks use relatively accurate statistical coverage prediction models such as the Akumura–Hata model discussed in Chapter 4. Statistical channel models for coverage in indoor areas are much less accurate, which poses a challenge in the analysis of coverage of WLANs unless we resort to labor-intensive empirical

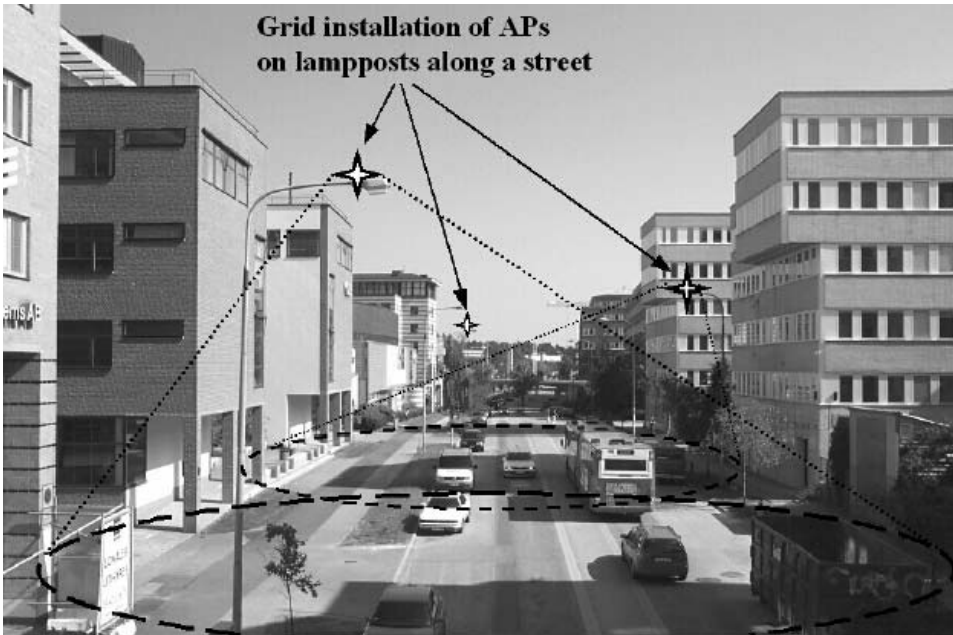


FIGURE 11.7 Grid installation along a street in a downtown area [Unb02].

measurements or ray-tracing algorithm. APs for WLANs are very inexpensive, around a few hundred dollars at the time of this writing, and do not require expensive antenna towers and site landscape for installation. BSs are orders of magnitude more expensive than APs, and for large-area coverage in macrocells, they require expensive towers and site landscapes. Radio resource management in WLANs is much simpler because WLANs have a limited number of nonoverlapping frequency bands (e.g., three for the popular IEEE 802.11b); cellular networks have an order-of-magnitude more channels to handle. In addition, traditional WLAN data traffic occurs in bursts, is nonsymmetric, and is location- and time-selective; therefore, medium access control is not performed centrally through radio resource management techniques. Mobility management for cellular networks is much more complex than for WLANs because most popular WLAN applications are quasistationary, whereas cellular networks are designed to support a user inside a vehicle. As a result of all these differences, an 802.11 AP is simple and connects to the Internet backbone directly, whereas a cellular BS is connected to the PSTN using a hierarchy that includes a base station controller (BSC) and a mobile switching center (MSC). The BSC and MSC are needed because in cellular networks we have more complex radio resource management, mobility management, and connection management techniques. Cells are larger, voice connection needs more quality control, and cellular telephones have greater mobility.

It is true that at the time of this writing the dominant bands for use by WLANs are the 2.4-GHz ISM bands. However, as the popularity of WLANs increases, many researchers envision that at some point the industry will migrate to higher frequencies, where wider portions of bandwidth are available. The path of migration is first to 5 GHz and then to tens of gigahertz and perhaps to the infrared domain. Cellular planning for higher frequencies would involve other challenges. In some situations,

illustrated by Fig. 11.3*b*, the desired signal is attenuated by obstacles, whereas an interfering signal arrives unobstructed and there is consequent degradation in C/I . This situation may prevent the design of a logical layout for the cells in an indoor area. However, if the signal is contained in one room and does not penetrate the walls of the room, the walls can be used to define the boundaries of the cells. This situation exists for infrared and microwave (above 20 GHz) WLANs, wherein each room constitutes one cell of a network. A more quantitative example should clarify this situation.

Example 11.4: WLAN Coverage in Various Environments A systematic comparative performance evaluation of the random, grid, and coverage optimization techniques for deployment of WLANs in office buildings, shopping centers, and campus areas at 5-, 17-, and 60-GHz frequencies is provided in [Unb02, Unb03]. In this work ray-tracing software is used to generate channel profiles in various environments. Figure 11.8 shows sample building layouts for an office, a shopping mall, and a campus area used by the ray-tracing software. Considering different AP densities, the received SNR in different locations is calculated to form the cumulative distribution function of the SNR for various conditions. Figure 11.9 shows two sets of cumulative distribution functions at 5 GHz in a campus environment. Each set consists of user deployment, grid installation, and optimal network planning installation for 0.1 and 0.15 AP per 1000-m² area. Table 11.1 provides the 10th percentile points for the density function of the SNR in different environments and 5, 17, and 60 GHz. Since 17 GHz has coverage difficulties in shopping areas and 60 GHz is not suitable even for 17-GHz operation in open areas, these values are not included in the table.

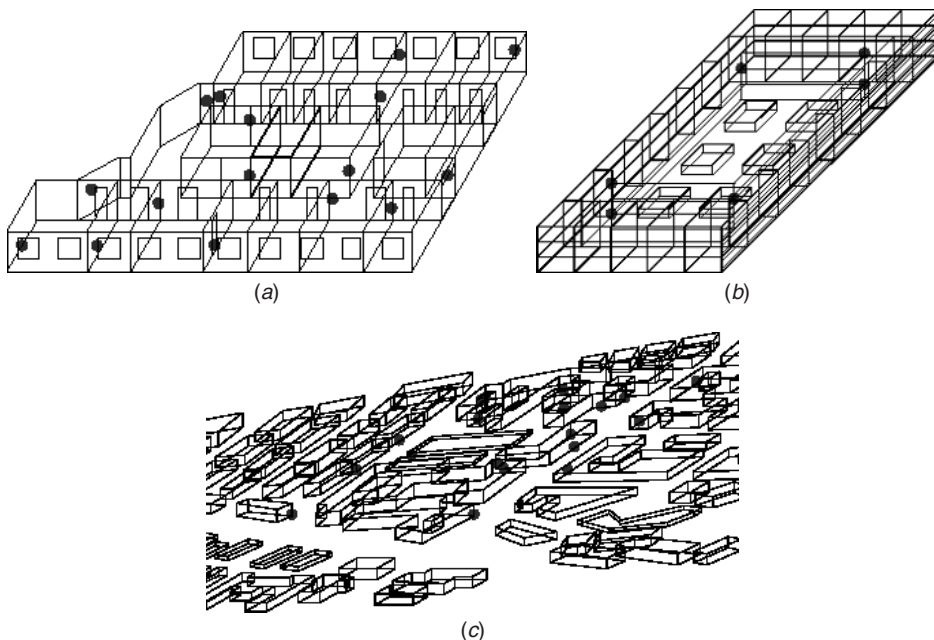


FIGURE 11.8 Typical building layouts used for the performance analysis of WLAN using ray-tracing software in (a) an office (b) a shopping mall, and (c) a campus area [Unb02].

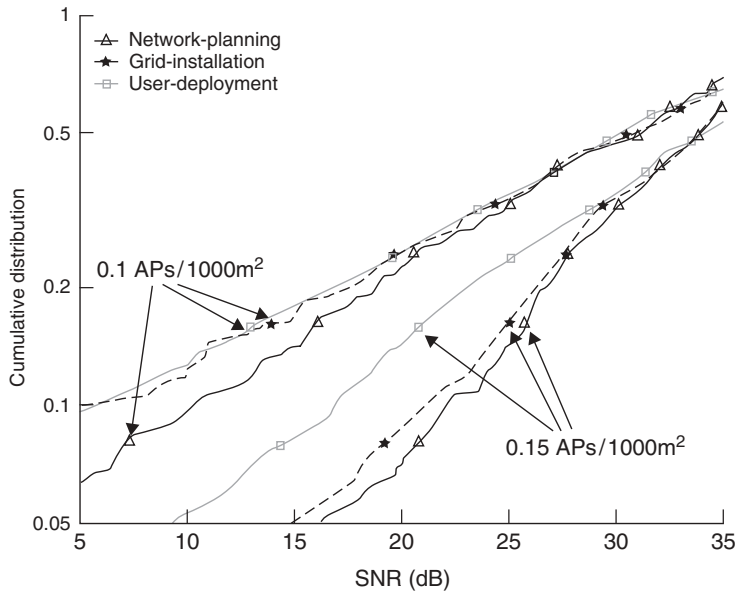


FIGURE 11.9 Cumulative distribution function of the received SNR at 5 GHz [Unb02].

TABLE 11.1 Required 10th Percentile SNR in Various Environments and Frequencies

Deployment Method	f_0 (GHz)	Minimum AP Density (APs/ 10^3 m ²)	10th Percentile SNR (dB)
Office environment			
User deployment	5	1.85	14.6
Grid installation	5	1.85	17.5
Coverage optimization	5	1.85	21.2
User deployment	17	14.8	16.1
Grid installation	17	7.4	16.9
Coverage optimization	17	7.4	19.9
User deployment	60	74	18.0
Grid installation	60	52	15.2
Coverage optimization	60	52	14.9
Shopping mall environment			
User deployment	5	0.5	21.3
Grid installation	5	0.17	19.0
Coverage optimization	5	0.17	22.2
User deployment	17	6	16.2
Grid installation	17	3	16.0
Coverage optimization	17	3	18.5
Campus environment			
User deployment	5	0.15	14.2
Grid installation	5	0.15	21.0
Coverage optimization	5	0.15	22.0

In general, office buildings provide a simple environment for WLAN deployment. For 2.4- and 5-GHz systems, provided that the signal easily penetrates the walls, the coverage is relatively unproblematic and a few APs, installed opportunistically in convenient areas such as corridors, sitting areas, and large lecture halls, typically cover the floor of a building. At 17 and 60 GHz, propagation is limited primarily to LOS, and essentially one AP per room is needed, provided that the size of the room is larger than the coverage of the AP. Since the cell size is very small and more bandwidth is available at higher frequencies, the throughput per user can be increased well beyond that achieved at 2.4 or 5 GHz. This additional throughput is at the expense of higher infrastructure cost. In office areas, deployment methodology is uncritical in this type of environment, and performing sophisticated network planning or coverage optimization does not result in substantial performance gain. In such high frequencies, deployment must be dense, and user deployment can, in fact, outperform grid installation or coverage optimization. In a campus environment, coverage with an acceptable infrastructure density can be provided only if a 2.4- or 5-GHz system is used. The small cell sizes at 17 and 60 GHz basically preclude outdoor installation of such networks. The deployment method does, in fact, have a significant influence on the system capacity in a campus environment, and some form of network planning should preferably be performed. A shopping mall is a problematic and very complex environment, as to both coverage and interference. The potentially very large user populations in such environments not only cause strong shadowing, but also result in a relatively low average throughput per user. The strong fragmentation of the environment and the fact that the network layout needs to be done in three dimensions (the shopping mall comprises multiple floors) makes it very difficult to devise a reasonably good network plan that provides adequate coverage. Coverage optimization, and in some cases grid installation, can improve performance somewhat. However, the capacity is limited primarily by intense interference, which also fluctuates heavily due to the strong fragmentation of the environment. The potential improvement that could be achieved by using sophisticated network planning is, however, limited.

Capacity of WLAN Cellular Networks. As a mobile user moves inside the coverage area of a cell, the received SNR changes as a random variable. Figure 11.9 illustrates the statistical behavior of the SNR as a mobile terminal is located in different locations within a campus environment. For traditional circuit-switched voice applications, each user has a single data rate independent of its received SNR, and we calculate the capacity using Eq. (11.4.3) or (10.5.7), in which capacity means the number of available voice channels in a cell or in the entire network, which is a function of the total available bandwidth. All modern wireless data applications are multirate systems for which the data rate is adjusted to the received SNR. Examples of multirate WLAN and mobile data systems are IEEE 802.11a and g (see Example 9.5) and HDR services (see Example 10.27). In multirate systems each user can operate at one of the multiple choices of data rates according to the value of its received SNR. In a single-user environment the data rate of a user is the average of all data rates at which it operates while moving in the area. In a multiple-user environment, the average data rate per user is also a function of the medium access control technique and the number of users in the area. In the remainder of this section we provide a framework for understanding the capacity of a wireless data network regardless of its medium access control. The details of medium access control methods are described in subsequent sections.

In wireless data applications, for short distances to an access point or base station, a multirate terminal has its highest data rate. As the distance increases, the RSS and consequently, SNR decrease until a point that the necessary SNR for the highest data rate is not available and the modem is switched to the next lower data rate. As the distance continues to increase, the data rate continues to fall to lower rates, until the signal strength falls below the coverage of the AP or BS. In a cellular network with comprehensive coverage, we expect that when the signal falls below the lowest threshold, there is another AP or BS to which to connect. Therefore, if we consider an area that is covered by a wireless data service, the data rate available to the user has a spatial distribution in which associated to any location we have one of the available multirates of the system. In other words, the data rate in a random location in the area of the coverage forms a discrete random variable. One way to define a capacity for this multirate system with a statistical data rate is to define the *spatial capacity* as the average of the data rates observed by a user located randomly in the coverage area. With this definition the spatial capacity will be given by

$$R_{av} = \sum_{n=1}^N p_n R_n \quad (11.3.5)$$

in which R_{av} is the average spatial data rate, R_n one of the available multirates, and p_n the probability of occurrence of that data rate, which is the ratio of the areas in which we have that specific data rate to the overall area of the coverage of the AP or BS.

Example 11.5: Spatial Capacity of IEEE 802.11b IEEE 802.11b supports four data rates: 11, 5.5, 2, and 1 Mb/s. In a semiopen indoor area, these data rates will cover up to distances of 50, 70, 90, and 115 m, respectively. Figure 11.10a shows the area of coverage and circles around the AP that belong to different data rates. If a terminal

Data Rate (Mbps)	Coverage Distance (m)	Area of Coverage (m ²)	$p_n = \frac{A_i}{\pi D_4^2}$
$R_1 = 11$	$D_1 = 50$	$A_1 = 7850$	0.19
$R_2 = 5.5$	$D_2 = 70$	$A_2 = 7536$	0.18
$R_3 = 2$	$D_3 = 90$	$A_3 = 10048$	0.24
$R_4 = 1$	$D_4 = 115$	$A_4 = 16092$	0.39

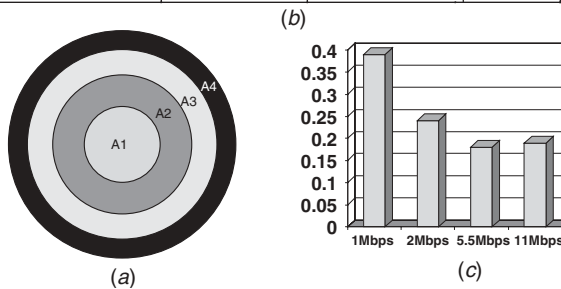


FIGURE 11.10 Data rates and coverage areas for the IEEE 802.11b in a semiopen indoor area: (a) coverage area for various data rates; (b) calculation of probabilities for data rates; (c) probability density function of the data rates.

in located randomly in the area of coverage, the probability of being in each area is given by the ratio of the area for the specific data rate to the total coverage area:

$$p_n = \frac{A_i}{\pi D_4^2}$$

Figure 11.10*b* shows the data rate, distance coverage, cylindrical area, and the probability of having a certain data rate. Figure 11.10*c* shows the probability density function of the data rates calculated from the ratio of the coverage area for a data rate and the overall coverage area. If we substitute the data rates and their probabilities from the density function in Fig 11.10 into Eq. (11.3.5), the average data rate or spatial capacity of the AP is

$$R_{av} = 0.19 \times 11 + 0.18 \times 5.5 + 0.24 \times 2 + 0.39 \times 1 = 2.584 \text{ Mb/s}$$

That is well below the expected 11 Mb/s.

The scenario above provides a worst-case condition for cellular wireless data networks. In actual deployment we always have overlapping cells, and overlaps are at low data rates, reducing their contribution to the average data rate. The upper bound for spatial capacity in cellular deployment is achieved when adjacent APs are close enough that the user always observes the highest data rate (11 Mb/s in Example 11.5). For the popular grid installation shown in Fig. 11.11, the minimum distance between access points (the size of the grid) to support the maximum spatial diversity is $D_g = \sqrt{2} D_1$, in which D_1 is the coverage of the maximum data rate (50 m in Example 11.5). For this situation the probability density function of the data rates for the system is only an impulse at 11 Mb/s with a probability of 1. As the grid distance between access

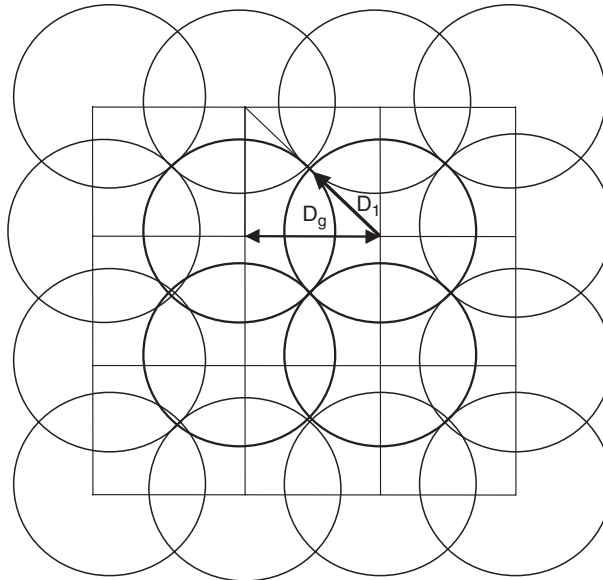


FIGURE 11.11 Grid deployment with optimal spatial capacity.

points increases beyond $D_g = \sqrt{2} D_1$ the probability density function of the data rates starts to include lower data rates, resulting in decreases in the spatially averaged data rate. As we said earlier, if we have no coverage gap, the minimum spatial throughput cannot be less than the spatial throughput of a single access point (2.584 Mb/s for 802.11b in Example 11.5). In our discussions in this chapter, we addressed single-floor indoor environments; for discussions on multifloor environments, the reader may refer to [Hil01].

11.4 CENTRALLY CONTROLLED ASSIGNED ACCESS METHODS

From the preceding discussion of network topologies and the ways in which they facilitate communication between user stations, it should be apparent to the reader that channel-access methods and methods for sharing a channel among multiple users are essential ingredients in achieving efficient operation and good performance in a wireless network. Users in a wireless network seldom have need to access a channel for a long period of time. Thus, schemes are needed for providing multiuser access, usually referred to simply as *multiple access*, to the frequency, time, and spatial resources of a network in an orderly manner and in a way that minimizes transmission overhead while maximizing overall network capacity. In this and the following sections we outline the principal categories of channel access methods and describe the primary characteristics of each method. There are two major categories of channel-access methods for the wireless communications environment: centrally controlled assigned access and contention random access. Access techniques in both categories are commonly used for wired communications networks as well, but here we discuss access techniques from the perspective of their use in wireless networks. In the centrally controlled assigned access methods described in this section, a central controller, a base station, or an access point assigns a fixed allocation of channel resources (frequency, time, or space, or any combination of them) in a predetermined basis to a single user. We describe briefly the three basic centrally controlled access methods (FDMA, TDMA and CDMA) and draw some comparisons among them. We describe them here in their simplest forms. However, these schemes can and often are implemented with various multiuser access algorithms to form random-access schemes. We discuss random-access schemes separately in Section 11.5.

11.4.1 Frequency-Division Multiple Access

The frequency-division multiple-access (FDMA) technique is built on the well-known frequency-division multiplexing (FDM) scheme for combining nonoverlapping user channels for transmission as a wider-bandwidth signal. FDMA, shown in Fig. 11.12, is the simplest and oldest form of multiplexing and has long been used in the telephone and commercial radio and television broadcasting industries. In fixed-assignment FDMA, a fixed subchannel is assigned to a user terminal and is retained until released by the user. At the receiving end, the user terminal filters the designated channel out of the composite signal. FDMA was used in the first generation of cellular mobile telephone systems and in VHF and UHF land-mobile radio systems. It was also the most common form of multiple access in satellite networks [Bha81]. This access method is very efficient and simple if the user has a steady flow of information to be sent, as in

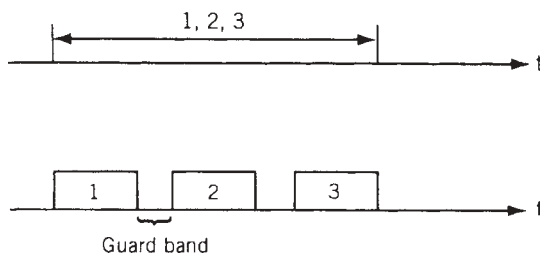


FIGURE 11.12 Fixed-assignment FDMA format.

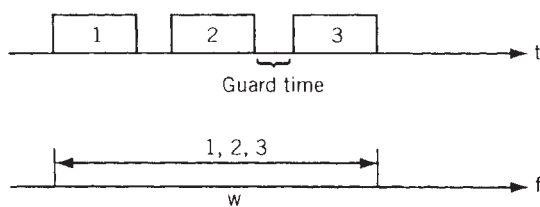


FIGURE 11.13 Fixed-assignment TDMA format.

circuit-switched voice traffic, but can be very inefficient if the user data are sporadic in nature, as is the case with bursty computer data or short-message traffic.

11.4.2 Time-Division Multiple Access

In time-division multiplexing (TDM), a digital stream is apportioned among multiple users by making a deterministic allocation to each user of time intervals called *time slots*. The time slots, one or more for each user plus associated overhead bits, are typically organized into frames, as illustrated in Fig. 11.13. The time-division multiple-access (TDMA) scheme is built on a TDM transmission format. In fixed-assignment or circuit-switched TDMA, used for cellular voice applications, a central controller serves to assign users to time slots, and an assigned time slot is held by a user until the user releases it. At the receiving end, a user station synchronizes to the TDMA signal frame and extracts the time slot designated for that user. As is the case with FDMA, this access method makes efficient use of channel resources with simple channel assignment techniques when a user has a steady flow of data to be transmitted over a circuit-switched channel but needs more complex channel allocation algorithms and communication protocols to support bursty data sources for packet switching. The TDMA concept was developed in the late 1960s for use in digital satellite communication systems [Sk188] and first became operational commercially in the mid-1970s [Kwa75]. The fixed-assignment form of TDMA is seldom used in satellite systems, and more flexible multiple-access methods are used instead. However, TDMA is the most popular access method used in second-generation digital cellular systems in North America (IS-136), Europe (GSM), and Japan (JDC). As currently defined, these systems, which we describe further in Chapter 15, operate in a fixed-assignment manner for digital voice service and for circuit-switched data and facsimile services. At the time of this writing, GSM TDMA systems have the most subscribers worldwide for cellular networks.

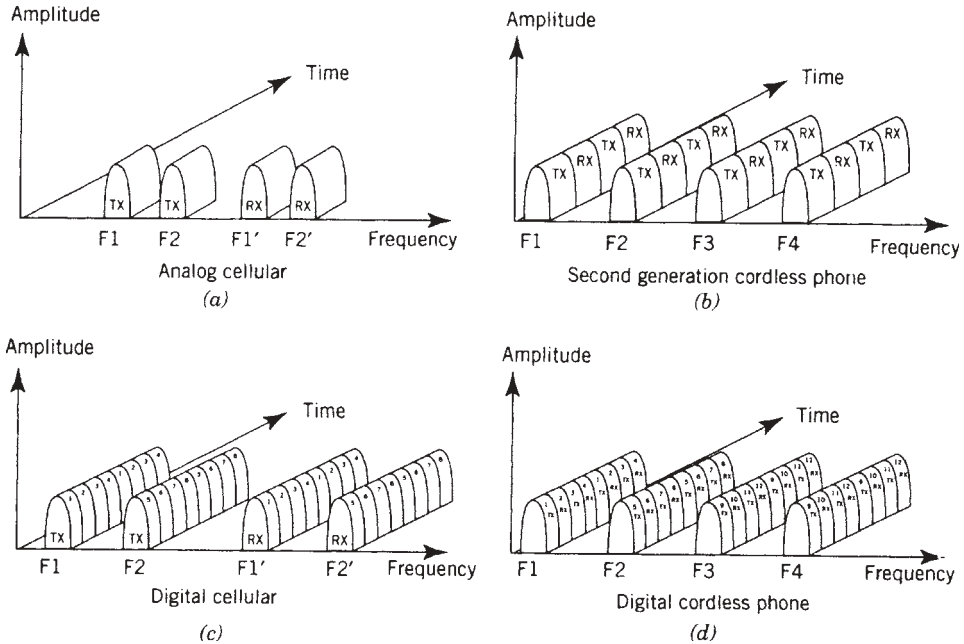


FIGURE 11.14 Multiple-access techniques: (a) FDMA; (b) TDD/FDMA; (c) TDMA/FDMA; (d) TDMA/TDD/FDMA.

11.4.3 Time-Division Duplex

A form of FDMA and TDMA being used in some wireless systems is time-division duplex (TDD), in which alternating time slots on the same carrier signal are assigned to the forward and reverse directions of communication. The European cordless telephone standard, CT-2, used TDD/FDMA, and the European DECT system for low-power local-area voice applications used the TDD/TDMA format.

Figure 11.14 shows FDMA, TDD/FDMA, TDMA/FDMA, and TDMA/TDD/FDMA access methods in time, frequency, and amplitude dimensions. The advantage of TDD is that the reciprocity of the channel allows for exact open-loop power control and simultaneous synchronization of the forward and reverse channels, because only one frequency is needed for duplex operation. The TDD technique is used in systems intended for low-power local-area communications, where interference must be carefully controlled, and low complexity and low electronic power consumption in portable devices are very important.

11.4.4 Comparison of FDMA and TDMA

Comparisons between FDMA and TDMA access methods involve a number of performance and implementation issues. Some of the issues have greater or less importance, depending on the type of system in which the access method is to be employed. Here we describe briefly some of the major points of comparison and comment on the choices that might be made in certain applications.

Format Flexibility. Compared with FDMA, a major advantage of the TDMA format is its flexibility. Because of the fully digital format and the flexibility of buffering and multiplexing functions, time-slot assignments among the multiple users are readily adjusted to provide different access rates for different users. This feature can be particularly useful in a network where a combination of services at different rates is to be supported. This architecture is therefore well suited to supporting multimedia applications built on a hierarchy of data rates. The TDMA format is also advantageous where a system design is to evolve over time from one multiplexing format to another. This was the case, for example, in the planned two-phase development of the IS-136 digital cellular system. In the first phase of implementation of the system, three user channels were supported in each six-slot TDMA frame, with each user channel comprising two slots. In the second phase of IS-136 evolution, when reduced-bit-rate vocoders were available, up to six users (one time slot per user) were planned to support each TDMA frame. Multislot assignment is also available to provide a wider choice of data rates per user [EIA94a, EIA94b]. An additional advantage of the TDMA format is that it lends itself more readily to integration of digital voice and data services. As we show in Chapter 15, GPRS and EDGE systems both integrate packet-switched data into the circuit-switched GSM interface. It is significantly more difficult to provide the various forms of flexibility, system evolution, and services integration in an FDMA system, particularly if channel splitting must be implemented.

Fading, Diversity, and Related Issues. The difference in bandwidth between an individual FDMA user channel and the composite multiuser TDMA channel is very important in multipath fading environments. In a typical FDMA system, the channel bandwidth is usually smaller than the coherence bandwidth of the transmission channel and there is no need to use an adaptive equalizer at the receiver. At the same time, this situation removes the opportunity for the implicit frequency diversity gains which are achievable when signal bandwidth approaches the coherence bandwidth. With the TDMA format, in many instances the composite TDMA signal bandwidth is close to the coherence bandwidth of the channel and an adaptive equalizer is needed. That is a benefit that provides a form of implicit frequency diversity, mitigating the effects of frequency-selective fading. Equalization of the channel requires the transmission of a training sequence as a part of each transmitted packet. The length of the packet should be much longer than the length of the training sequence, so that only a small fraction of transmission time is devoted to training of the equalizer. On the other hand, the length of the packet should be short enough so that between the two consecutive transmissions of training sequences the channel remains almost stationary. GSM packets carry 26 training bits in the middle of each information burst, which comprises 148 bits (not including the guard time).

Bit-Rate Capability and Message Delay. From a broad perspective, it might appear that FDMA and TDMA should provide the same capability for carrying information from a set of users over a network. In fact, this is true with respect to bit-rate capability if for simplicity we neglect all overhead elements, such as guard bands in FDMA and guard times in TDMA, and we neglect the signal-to-noise requirement for proper operation of the systems. That is, assuming M users each with the same source data-rate, and assuming that the overall data-rate capability for the system is the same R value for either system, a data rate of R/M is available to each user in either

system. However, as we showed in Example 11.3, the capacity of IS-136 is more than fivefold than that of AMPS systems. This was due to two facts: (1) In AMPS, each user occupies 30 kHz of bandwidth to provide an acceptable quality of service (in IS-136, each user occupied one-third of that bandwidth), and (2) the SNR requirement is lower for IS-136, allowing the system to use smaller frequency-reuse factors and to assign more channels per cell. In addition, if the measure of performance is the average delay encountered by a message packet, TDMA again performs better than FDMA. For example, it is a simple matter to show that for M users generating data at a constant uniform rate of R/M bits/s, and given that FDMA and TDMA systems each transmit a packet of N bits every T seconds, the average packet delay (packet waiting time before transmission plus packet transmission time) is

$$D_{\text{FDMA}} = T \quad (11.4.1)$$

for the FDMA system and

$$D_{\text{TDMA}} = \frac{T}{2} \left(1 - \frac{1}{M}\right) + \frac{T}{M} = D_{\text{FDMA}} - \frac{T}{2} \left(1 - \frac{1}{M}\right) \quad (11.4.2)$$

for the TDMA system [Sk188]. Therefore, for two or more users, TDMA is superior to FDMA with respect to average packet delay. Note that for large numbers of users, the difference in packet delay is approximately $T/2$, where T is the time interval between transmitted packets. Although this result applies to a case of constant deterministic rates for the M sources, the smaller average delays for TDMA are exhibited for any independent message arrival process [Rub79, Ham86].

Amplifier Back-off. The FDMA system requires that multiple modulated carriers be added together to form the transmitted signal. This addition of carriers produces amplitude modulation in the composite signal feeding the transmitting power amplifier. In some systems this makes it necessary to reduce the input to the amplifier from its maximum drive level in order to control intermodulation distortion. In satellite repeaters, this procedure, referred to as *input back-off*, is an important factor in maximizing the power efficiency of the repeater. This issue as it relates to overall capacity was a major factor in the adoption of TDMA in satellite systems. However, this is not necessarily as important in terrestrial communication systems, because a mobile terminal will be transmitting only one channel at a time. The situation for the base station in a centralized network is similar to that for a satellite repeater, but the power consumption in base station electronics is not a major design issue.

Spurious Interference. In transmission environments where spurious narrowband interference is a problem, the FDMA format, with a single user channel per carrier, has an advantage in that a narrowband interferer can impair the performance of only one user channel. With the wider-bandwidth TDMA format, a single narrowband interfering signal can affect the performance of all user channels in the TDMA stream.

11.4.5 Code-Division Multiple Access

It is useful to think of the FDMA and TDMA access methods as distinct methods for apportioning the time-frequency signal domain of the system among multiple users. In

FDMA, the system signal domain is sliced into frequency bands, perhaps with guard bands left between the user bands. In centrally controlled assignment access, each user has use of an assigned band during the entire connection time. In TDMA, the system signal domain is sliced into time slots, perhaps with guard-time intervals left between slots or frames, and each user occupies the entire signal bandwidth, but only during each of his assigned slot times. Thus, the user channels are cleanly separated in either the frequency or the time domain. The third access method, code-division multiple access (CDMA), can be viewed as a hybrid combination of FDMA and TDMA. With CDMA, multiple users operate simultaneously over the entire bandwidth of the time-frequency signal domain, and the signals are kept separate by their distinct user-signal codes. As we discussed in Chapter 10, the CDMA method is built on spread-spectrum signaling. It should be noted, however, that the use of spread-spectrum signals does not necessarily imply the use of CDMA. Spread spectrum is used, however, to provide sufficient degrees of freedom to be able to separate the user signals in the time-frequency domain. As discussed in Chapter 10, the two common CDMA techniques are (1) direct-sequence CDMA (DS-CDMA) and (2) frequency-hopping CDMA (FH-CDMA), which are constructed on the DS and FH spread-spectrum techniques, respectively.

CDMA in Portable and Mobile Radio Networks. After more than a decade of intense debates in the cellular industry, DS-CDMA finally emerged as the favored medium access control technique for third-generation cellular networks. FH-CDMA was also implemented in the Bluetooth technology, which attracted tremendous attention for WPAN applications in the late 1990s. Much attention is being given to the adoption of CDMA methods in emerging new wireless networks. The advantages of the CDMA approach in these systems are as follows:

1. *Timing flexibility.* The chief advantage of CDMA relative to TDMA, as mentioned above, is that it can operate without timing coordination among the various simultaneously transmitted user signals. The separation of user signals is ensured by the design of the signal codes and is unaffected by transmission-time variations.
2. *Performance in frequency-selective fading.* In frequency-selective fading, one portion of the signal spectrum can be attenuated, whereas other portions are received without attenuation. In an FDMA system, a user signal transmitted in the attenuated portion of the spectrum will be degraded as long as the fade persists. However, in an FH-CDMA system, a user's signal will be attenuated only during the time interval when the signal hops into that part of the spectrum. Therefore, the FH-CDMA format will tend to distribute the effects of frequency-selective fading over all the users' signals. As we mentioned in Chapter 10, error-correction coding with interleaving can significantly improve performance by correcting the errors caused by signal fades.
3. *Interference resistance.* Because both forms of CDMA are constructed on spread-spectrum signals, they provide an inherent resistance to both intentional and unintentional interference. The performance of spread-spectrum communication techniques in interference is discussed in greater detail in Chapter 10.
4. *Communication privacy.* In CDMA systems, communication between a particular transmitter–receiver pair can be made private by use of a signal code that is

known only to that pair of stations. This technique is the basis of many secure communications systems used in military applications.

5. *System capacity.* CDMA provides more users per cell, as we discussed in Chapter 10.
6. *Soft handoff.* Because adjacent cells in a spread-spectrum cellular network use the same frequency, when a mobile moves from one cell to another, the handoff can be made “seamless” by the use of signal combining. When the mobile station approaches the boundary between cells, it communicates with both cells and combines the signals with a RAKE receiver. When a reliable link has been established with the new base station, the mobile stops communicating with the previous base station, and communication is fully established with the new base station. This technique is referred to as *soft handoff*.
7. *Soft capacity limit.* With CDMA there is no hard limit on the number of users that a system can support, as there is in an FDMA or TDMA system. Rather, each user is a source of noise for all other users, as the users occupy the same time and frequency space with distinct user signal codes. Thus, theoretically, we may arbitrarily increase the number of users at the expense of degradation in the performance seen by every user. As a practical matter, systems are designed to deny access by new users as the BER of the individual user signals approaches a prescribed threshold, around 10^{-3} . The number of users at which the threshold is reached will depend on the particular geographic distribution of mobiles being served by the system.
8. *Overlay.* The second-generation digital cellular systems are deployed in the same bands previously occupied by analog AMPS systems. The transition to digital service was accomplished gradually as digital mobile terminals and base stations replaced analog equipment. Spread-spectrum transmissions can overlay the existing analog systems and allow the two systems to coexist during the transition phase.
9. *Interference control with antenna sectorization.* Another advantage of spread-spectrum technology is that the sectored antennas used for interference control increase network capacity. The reduction of interference allows more users to operate simultaneously in the network.
10. *Time diversity.* As we discussed in Chapter 10, spread spectrum provides a means of combatting multipath through the use of RAKE receivers. As a result, a spread-spectrum system provides better signal coverage than standard radio systems by exploitation of multipath as a form of implicit time diversity.

At the heart of the debates for selection of CDMA for third-generation systems, the main concerns about the CDMA technique were:

1. *Implementation complexity.* Spread spectrum is a two-layer modulation technique requiring greater circuit complexity than that of conventional modulation schemes. This, in turn, could lead to higher electronic power consumption and higher weight and cost for the mobile terminals. Gradual improvements in low-power digital electronic technology resolved this concern over time. Indeed, since CDMA provides better coverage, its overall power consumption became one of its advantages.

2. *Power control.* As we discussed in Chapter 10, the capacity of a spread-spectrum CDMA system operating without power control is extremely limited. As a result, power control is essential for the practical operation of CDMA systems. In TDMA–FDMA networks, power control also reduces interference and improves the quality of received signals. With CDMA, power control is the key ingredient in maximizing the number of users that can operate simultaneously in the system. Another motivation for using power control is to conserve transmitted signal power, thereby increasing the battery-recharging cycle. Power control is performed by using either an open- or a closed-loop method. The open-loop method is based on the similarity of power loss in the forward and reverse directions. The mobile terminal keeps the sum of the transmitted and received power at a constant level (typically, -73 dBm). Thus, by constant monitoring of the received power, the transmitted power is, in turn, adjusted. In closed-loop power control the base station monitors the received power from the mobile and commands the mobile to adjust its power up or down one step at a time. In the Qualcomm CDMA system, adjustments are made 800 times per second (every 1.25 ms) and the step size is 1 dB. The tight power control needed for proper CDMA operation also helps in improvement of the battery life by avoiding transmission at higher powers when it is not necessary.
3. *Message delay.* In the CDMA channel, similar to the FDMA and unlike TDMA, the transmission rate of the system is the same as the transmission rate of the users. As a result, the time delay for packet transmission remains as in Eq. (11.4.1):

$$D_{\text{CDMA}} = D_{\text{FDMA}} = T \quad (11.4.3)$$

leaving basic TDMA with a better message delay characteristic than basic CDMA or FDMA.

Example 11.6: Legacy Qualcomm System The legacy example of the DS-SS-CDMA systems is the Qualcomm system used by the TIA Subcommittee TR45.5 as the basis from which it developed the IS-95 standard. This system operates in the existing cellular frequency bands, 824–849/869–894 MHz. The transmitter power for mobile stations ranges from 2.2 mW to 6 W. The modulation method for base station transmission is QPSK, whereas OQPSK, which has a more nearly constant envelope characteristic, is used for transmission from the mobile. The mobile-to-base signal structure includes a rate- $\frac{1}{3}$ $K = 9$ convolutional code, whereas the base-to-mobile channel uses a rate- $\frac{1}{2}$ $K = 9$ code. Variable-rate speech coding is used, with the speech coded at 8000 bits/s during talk spurts and at 1000 bits/s during listening intervals and pauses between talk spurts. This variable-rate scheme serves to minimize radio interference, which enhances the capacity of the CDMA system while minimizing battery drain. The chip rate in the Qualcomm system is 1.2288 Mb/s, and the receiver uses a three-tap RAKE receiver. The number of antenna sectors per base station is three.

11.4.6 Traffic Engineering for Circuit-Switched Controlled Access

In circuit-switched controlled-access networks, the performance of the network is measured by the blockage rate for initiated calls. The cellular service providers often design their networks so that the blockage rate at peak traffic is always below 2%. The

blockage rate is a function of the number of subscribers, the number of calls initiated, and the lengths of conversations. The Erlang B and Erlang C equations are used to relate the probability of blockage to the average rate of the arriving calls and the average length of a call. The Erlang B equation relates the probability of blockage $B(N, \rho)$ to the number of channels N and the normalized call density in units of channels, ρ . The Erlang B formula is given by

$$B(N, \rho) = \frac{\rho^N / N!}{\sum_i^N (\rho^i / i!)} \quad (11.4.4)$$

where $\rho = \lambda / \mu$, λ being the call arrival rate and μ the service rate of the calls. This equation assumes that the arrivals are Poisson and the service rate is exponential [Ber87]. Figure 11.15 shows a chart relating the probability of blockage $B(N, \rho)$ to N , the number of channels available in the system, and ρ , the normalized traffic density in units of channels.

Example 11.7: Number of Subscribers in a Cellular Network Assume that an IS-95 cellular provider installs 120 cells to cover a large area. The maximum number of user channels for each 1.25-MHz IS-95 carrier is 64. From Fig. 11.15 for 2% probability of blockage, we will have $\rho \simeq 50$. If we assume that the average telephone conversation is 5 minutes, $\mu = \frac{1}{5} \text{ min}^{-1}$ and the acceptable arrival rate of calls is $\lambda = \rho \times \mu = \frac{1}{5} \text{ min}^{-1} \times 50 \text{ erlangs} = 10 \text{ erlangs/min}$. If we assume that each user makes an average of two calls per hour, the system can accept $10 \text{ erlangs/min} / 2 \text{ erlangs/60 min} = 300$

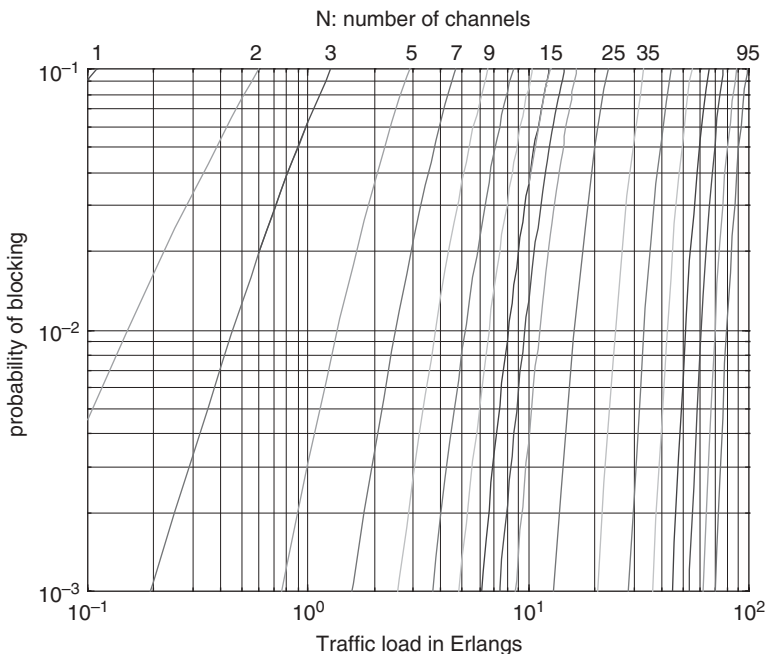


FIGURE 11.15 Erlang B chart showing the blocking probability as a function of number of channels and traffic offered.

subscribers/cell. Therefore, the total number of subscribers is 300 subscribers/cell $\times$ 120 cells = 36,000.

The Erlang C equation relates the queuing delay to the number of channels and the traffic density. This formula starts with the probability of having a delayed call:

$$P[\text{delay} > 0] = \frac{\rho^N}{\rho^N + N!(1 - \rho/N) \sum_{k=0}^{N-1} (\rho^k/k!)} \quad (11.4.5)$$

Figure 11.16 illustrates the relationship between the probability of delay, the number of channels N , and the normalized traffic per available channel, ρ . The second equation in this line is the probability of having a delay that is more than a given time t , which is given by

$$P[\text{delay} > t] = P[\text{delay} > 0]e^{-(N-\rho)\mu t} \quad (11.4.6)$$

indicating that the delay decays exponentially with the service rate, μ , multiplied by the average left over channels, $N - \rho$. The average delay is then given by the average of the exponential distribution:

$$D = P[\text{delay} > 0] \frac{1}{\mu(N - \rho)} \quad (11.4.7)$$

Example 11.8: Average Call Delay If we continue Example 11.7 for $N = 64$ and $\rho = 50$, from Fig. 11.16 we have $P[\text{delay} > t] = 0.1$. Substituting this value in Eq. 11.4.7, we will have

$$D = P[\text{delay} > 0] \frac{1}{\mu(N - \rho)} = \frac{0.1}{\frac{1}{5}(64 - 50)} = 0.036 \text{ min} = 2.14 \text{ s}$$

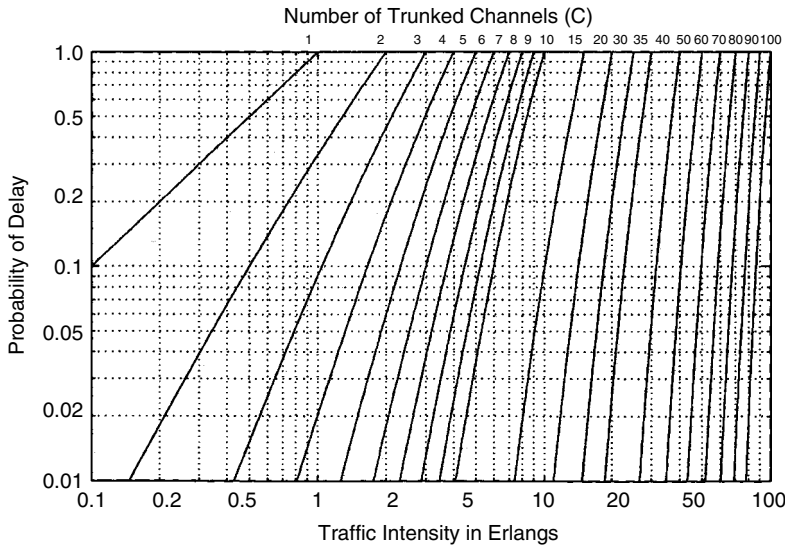


FIGURE 11.16 Erlang C chart relating the traffic offered to the number of channels and the probability of delay.

11.4.7 Packet-Switched Data in Central Control Access

Centrally controlled access methods such as FDMA, TDMA, and CDMA in first- and second-generation cellular networks were designed with primary emphasis on circuit-switched voice and data applications. High-speed packet-switched data applications became the focal point of third-generation cellular networks after the huge success of Internet technology starting in the mid-1990s. The issue of channel capacity for circuit-switched cellular networks that was the focal point for comparison of TDMA and CDMA networks in the early 1990s was complemented with the ease of support of higher speeds and integration of circuit- and packet-switched services in a multimedia wireless network. The main objective of emerging new wide- and local-area wireless networks became to provide for a physical layer modem technology that supports higher transmission speeds and a medium access control that supports more efficient integration of circuit- and packet-switched traffic. The increase in transmission data rate is achieved by using multisymbol modulation and multirate transmission. Several medium access control techniques became popular for implementation of these integrated systems. To provide a good understanding of the important issues related to the emerging integrated multimedia systems, we first discuss the benefits of integration and then provide examples of medium access methods for integrated environments. In the following section we address high data rates in TDMA and CDMA environments.

Capacity Benefits of Integration. In a circuit-switched network we assign a dedicated line to the user for the entire connection time. To maintain a certain probability of blockage $B(N, \rho)$, as we explained in Section 11.4.5, on average we have ρ channels attempting to access from the N channels available. Therefore, we often have a set of idle channels, and the average number of these idle channels, N_{idle} , is given by

$$N_{\text{idle}} = N - \rho(1 - B(N, \rho)) \quad (11.4.8)$$

in which the relation among $B(N, \rho)$, N , and ρ was described by the Erlang B formula given in Eq. (11.4.4). Assuming that each channel receives an equal fraction of the call load, the normalized average idle time of a channel is the ratio of average idle channels to the average occupied channels, given by [Bud97]

$$T_{\text{idle}} = \frac{N - \rho(1 - B(N, \rho))}{\rho(1 - B(N, \rho))} \quad (11.4.9)$$

For a typical holding time of a few minutes and a blocking probability of 2%, this period is long enough for the system to detect the availability and to send bursts of data.

Example 11.9: Average Idle Channels in CDMA For the $N = 64$ channel system considered in Examples 11.7, the average channel utilization for a blocking rate of 2% was $\rho = 50$. Using Eq. (11.4.8), we will have $N_{\text{idle}} = 15$. With 9600 bits/s allocated to each digital voice channel, on average we will have $9600 \text{ bits/s} \times 15 = 144 \text{ kb/s}$ of the idle transmission data rate available in the system, which can be exploited with a more intelligent medium access control. The average idle time is $T_{\text{idle}} = 0.31$, which on average allows 1.5-minutes of silence for 5-minute call durations.

If the system can take advantage of the silence periods in two-way telephone conversations, as we indicated in calculation of the capacity of CDMA systems in Eq. (10.5.7),

an additional throughput of up to 60% is available for data applications. The analysis above is for FDMA and CDMA types of access; more elaborate calculations for the TDMA channels [Zha90] provide similar conclusions, which indicate the benefit of integration of circuit- and packet-switched traffic.

Polling Technique. One way of imposing discipline on a network of independent users is to equip one station in the network as a controller that polls all other stations periodically to determine if they have data to transmit. This approach is used in Bluetooth to provide an integrated voice and data service. The controller station may have a polling list giving the order in which the terminals are polled. If the polled station has something to transmit, it starts sending. If not, a negative reply (or lack of any reply) is detected by the controller, which then polls the next terminal in the sequence. Polling requires the constant exchange of control message between the controller and the terminals, and it is efficient only if (1) the round-trip propagation delay is small, (2) the overhead due to polling messages is low, and (3) the user population is not large and dominated by bursty data. Several types of polling have been analyzed in [Kon74, Tob80a], and their analysis has been applied to the radio environment in [Tob76, Aca87]. The results show that as the population of users increases, thus containing more and more bursty users, the performance of polling degrades significantly. All of these conditions hold for Bluetooth, which supports up to seven live connections and provides three different quality levels for the voice connection and several symmetric and asymmetric data services.

Packet Reservation Multiple Access. Another example of a system for integrating voice and data services is the work done by David Goodman and his colleagues in developing the concept of *packet reservation multiple access* (PRMA) [Goo89]. PRMA is a method for transmitting, in a wireless environment, a variable mixture of voice packets and data packets. PRMA has been developed for use in centralized networks operating over short-range radio channels. Short propagation times are an important ingredient in providing acceptable delay characteristics for the voice service. Our description here closely follows [Goo89, Goo91b].

The transmission format in PRMA is organized into frames, each containing a fixed number of time slots. The frame rate is identical to the arrival rate of speech packets. The terminals identify each slot as either “reserved” or “available,” in accordance with a feedback message received from the base station at the end of the slot. In the next frame, only the user terminal that reserved the slot can use a reserved slot. Any terminal not holding a reservation, which has information to transmit, can use an available slot.

Terminals can send two types of information, referred to as *periodic* and *random*. Speech packets are always periodic. Data packets can be random if they are isolated, or they can be periodic if they are contained in a long unbroken stream of information. One bit in the packet header specifies the type of information in the packet. A terminal having periodic information to send starts transmitting in contention for the next available time slot. Upon detecting the first packet in the information burst successfully, the base station grants the sending terminal a reservation for exclusive use of the same time slot in the next frame. The terminal in effect “owns” that time slot in all succeeding frames as long as it has an unbroken stream of packets to send. After the end of the information burst, the terminal sends nothing in its reserved slot. This, in turn, causes the base station to transmit a negative acknowledgment (NACK) feedback message, indicating that the slot is once again available.

To transmit a packet, a terminal must verify two conditions. The current time slot must be available, and the terminal must have permission to transmit. Permission is granted according to the state of a pseudorandom-number generator, with permissions at different terminals being statistically independent. The terminal attempts to transmit the initial packet of a burst until the base station acknowledges successful reception of the packet or until the terminal discards the packet because it has been held too long. The maximum holding time, $D_{\max}$ seconds, is determined by delay constraints on speech communication and is a design parameter of the PRMA system. If the terminal drops the first packet of a burst, it continues to contend for a reservation to send subsequent packets. It drops additional packets as their holding times exceed the limit $D_{\max}$. Terminals with periodic data (as opposed to voice) packets send stored packets indefinitely while they contend for slot reservations (equivalent to setting $D_{\max} = \infty$). Thus, as a PRMA system becomes congested, both the speech packet dropping rate and the data packet delay increase. Modified versions of PRMA are used in GPRS and EDGE to integrate bursty traffic with the periodic voice.

11.4.8 High Data Rates in Cellular Networks

In the mid-1990s, TDMA and CDMA technologies were the dominant second-generation cellular networks, while third-generation cellular network specifications for IMT-2000 demanded data rates of 384 kb/s for coverage everywhere and a data rate of 2 Mb/s for local areas. Since each voice user produces coded speech at around 10 kb/s, it is obvious that a high-data-rate user needs to combine the time slots of several TDMA voice users or use many codes of a CDMA channel. As we described in Chapter 2, each carrier of the dominant TDMA system, GSM, carries a data stream of 270 kbits/s, and the chip rate of the popular CDMA system, IS-95, is 1.25 Mb/s. Therefore, even if we allocate the entire carrier to one user, we cannot satisfy the IMT-2000 basic requirement with the existing physical layers. The only way to stay with the same carrier spacing would be to use multisymbol and/or multicarrier modulation as was proposed the CDMA-2000. Otherwise, one would have needed wider bandwidth, as proposed in wideband CDMA of the pan-European UMTS standard. In the meanwhile, since the main objective in the new trends for wireless cellular networks was to support higher data rates, a number of high-data-rate system emerged as an evolution of the exiting legacy cellular networks. The most dominant of the TDMA networks were GPRS and EDGE packet-switching networks, and the dominant CDMA systems were the 1xEV (single-carrier evolution) and 3xEV (three-carrier) systems, which are also referred to as HDR (high-data-rate) systems, described in Example 10.27. The principles of operation of these systems are important through two considerations: to show (1) how one can achieve high data rates using a voice-oriented network designed for lower speeds per user, and (2) how we can assess the performance of such data networks for a mobile user. We discuss these two issues in the rest of this section using EDGE and HDR as our specific example.

As we described in Example 10.27, the medium access control of HDR is really a TDMA/CDMA scheme. All 64 codes usually assigned to separate voice users are allocated to one HDR data user, and on top of that the BPSK modulation used in the forward (downstream) channel is replaced with the optional QPSK, 8-PSK, and 16-QAM modulations to support the variety of data rates shown in the table referred to in Example 10.27. The GPRS and EDGE data services are based on the GSM infrastructure and air interface. Similar to HDR, GPRS and EDGE use multiple voice

user resources and multisymbol modulations to achieve higher data rates. To extend Example 10.27, we describe details of GPRS and EDGE in the following example.

Example 11.10: High Data Rates in EDGE and GPRS GPRS retains GMSK, the modulation technique used in GSM, and creates different data rates by assigning different number of time slots per user and changing the error-correcting code used for error control in formation of the packets for packet data transmission. GSM has eight slots per carrier; therefore, GPRS or EDGE can use different numbers of slots to support higher data rates. A single 200-kHz carrier in GSM has eight time slots, each capable of carrying data at 9.2, 13.55, 15.75, or 21.55 kb/s (if forward error correction is omitted completely). The raw data rate for GPRS can thus be as high as $8 \times 21.55 = 172.4$ kb/s. Table 11.2 shows three sets of four data rates for one-, four-, and eight-slot GPRS systems. The differences in the four categories of the data rates are made by using coding rates of $\frac{1}{2}$, $\frac{2}{3}$, $\frac{3}{4}$, and 1 for the formation of each packet. More details of the coding schemes are not our main concern, but they are available in [Cai97]. (Also see Section 15.2.) We observe data rates ranging from 9.2 to 172.4 kb/s supported by this system. As the distance of the mobile from the base station increases, the received signal strength decreases and the system falls to lower data rates. The main difference between the EDGE and GPRS is that the EDGE system uses an optional 8-PSK modulation with 3 bits/symbol, which can increase the maximum data rate another threefold, to 473.6 kb/s. Table 11.3 shows four of the nine data rates supported by the EDGE system for one, four, and eight slots. The details of the coding for the GPRS and EDGE systems are slightly different; therefore, the data

TABLE 11.2 GPRS Data Rates (kb/s) for Channel Coding Schemes 1 to 4

Channel Coding Scheme	Slot Combinations		
	1 Slot	4 Slots	8 Slots
CS1	9.2	36.8	73.6
CS2	13.55	54.2	108.4
CS3	15.75	63	126
CS4	21.55	86.2	172.4

TABLE 11.3 EDGE Data Rates (kb/s) Using 8PSK Modulation Versus Various Channel Coding Schemes and Slot Combinations

Channel Coding Scheme	Modulation	Slot Combinations		
		1 Slot	4 Slots	8 Slots
MCS1	GMSK	8.8	35.2	70.4
MCS4	GMSK	17.6	70.4	140.8
MCS5	8PSK	22.4	89.6	179.2
MCS9	8PSK	59.2	236.8	473.6

rates with the same modulation and the same number of slots do not match exactly. The details of different data rates and the performance of EDGE are available in [Yal02].

Capacity of Wireless Data Networks. In Section 11.3.3 we addressed the issue of spatial throughput for wireless data services with application to WLANs. The spatial throughput concept defined in Eq. (11.3.5) provides the expected or average data rate for a fixed terminal located randomly in the area of coverage of a wireless database station or access point. In Example 11.5 we calculated the spatial capacity of an IEEE 802.11b access point using the table of received signal levels for various data rates. In that example we have neglected the effects of fading and interference from neighboring cells. In that way we could simply calculate the probability density function of data rates based on the average expected spatial coverage, considering only one access point. A more realistic but elaborate approach for calculation of the spatial distribution function of the data rates for a multirate mobile data service is to simulate a cellular environment with interference and fading effects or to perform actual measurements of the carrier to interference and background noise ratio in the spatial area of interest and use that distribution to determine the statistics of the data rates. An example should clarify this discussion.

Example 11.11: Capacity of One-Slot EDGE in One Sector Figure 11.17 provides the probability distribution function, $f(\gamma_{I+N})$, of the received SINR, γ_{I+N} , of an EDGE¹ mobile terminal in one sector of a 10-MHz system with 48 carriers and a frequency-reuse factor of 4. Each base station has 12 carriers and each sector of three-sector antennas carries four carriers. The carrier-to-interference ratio for the mobile is set at 24 dB. Using this figure and the relation between different data rates and their required SNRs, a statistical distribution of the data rates can be obtained. Figure 11.18 shows the percentage of cell area coverage for one sector of a base station for all nine data rates supported by a one-slot EDGE system. The staircase function under

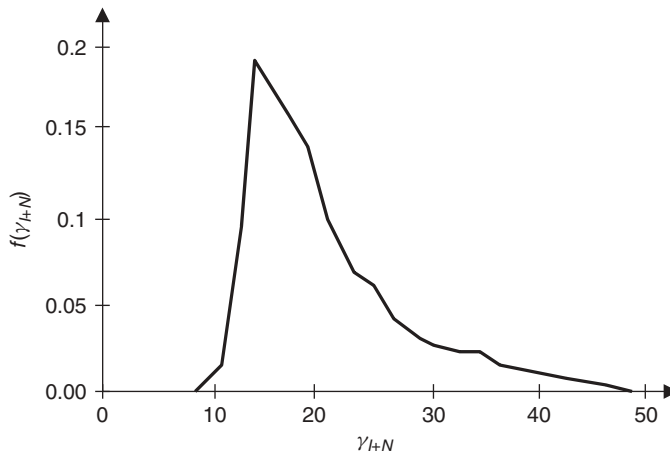


FIGURE 11.17 Probability distribution function of the carrier-to-interference plus noise ratio for EDGE [Yal02].

¹Note that distributions of the SINR for EDGE, GPRS, and GSM are all the same because they use the same bands and transmission powers.

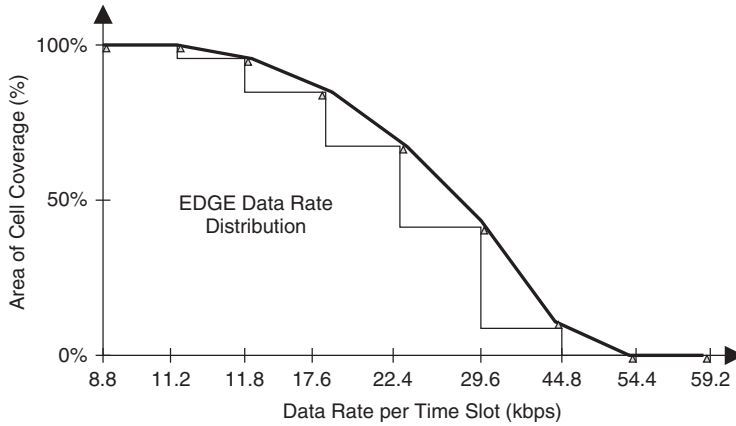


FIGURE 11.18 Percentage of the cell coverage area in a sector for EDGE data rates [Yal02].

the curve is the complementary cumulative distribution function of the data rates, whose derivative is the probability distribution function of the data rates (similar to Fig. 11.10c). If we use the data rates and distribution function of the data rates in Eq. (11.3.5), similar to Example 11.5, the average data rate of EDGE would be calculated at 31 kb/s per time slot in a sector. This capacity is related to one carrier, in one sector, using one-slot data transmission. The overall capacity of the network is given by $R_{\text{total}} = R_{\text{av}} \times N_{\text{carriers}} \times N_{\text{slots}}$. In our example we have four carriers per slot, and for maximum of eight slots per user we will have

$$R_{\text{total}} = 31 \text{ kb/s} \times 4 \text{ carriers/slot} \times 8 \text{ slots/user} = 992 \text{ kb/s}$$

for 10 MHz of bandwidth, which leaves us with a spectral efficiency of less than 10%.

Delay Analysis in Multirate Mobile Data. If we consider that the transmitted data are sent with packets, the spatial throughput that we discussed so far in this section and in Section 11.3.3, based on Eq. (11.3.5), assumes that the lengths of the transmitted packets are equal. Since we have N different data rates, the delay or latency (transmission time) of the packets sent with lower data rates is longer than transmitted with higher data rates. If we change the length of the packets in proportion to the inverse of the data rate, A/R_n , with A a constant, the average data rate given by Eq. (11.3.5) will change to [Ben00]

$$R_{\text{av}}^v = \frac{\sum_{n=1}^N P_n R_n (A/R_n)}{\sum_{n=1}^N P_n (A/R_n)} = \frac{1}{\sum_{n=1}^N P_n / R_n} \quad (11.4.10)$$

because now we send more information, inversely proportional to the data rate, at higher rates. Equation (11.3.4) is related to constant packet rate and packet delays that are proportional to the inverse of the data rate; in contrast, Eq. (11.4.10) provides for equal packet delays and variable packet lengths which are inversely proportional to the data rate. A numerical example should clarify this situation.

Example 11.12: Delay Throughput Relation in Multirate Assume a simple case of two data rates, $R_1 = 16$ kb/s and $R_2 = 1024$ kb/s with equal probability, $p_1 = p_2 = 0.5$. For a variable packet length we have the same latency, $L_1/L_2 = 1$, but from Eq. (11.4.10) the average throughput is

$$R_{av}^v = \frac{1}{\sum_{n=1}^N P_n/R_n} = \frac{1}{0.5/16 + 0.5/1024} = 31.51 \text{ kb/s}$$

For a variable packet length we have the same latency, $L_1/L_2 = 64$, but from Eq. (11.3.5), we have

$$R_{av} = \sum_{n=1}^N p_n R_n = 0.5 \times 16 + 0.5 \times 1024 = 520 \text{ kb/s}$$

This provides an instructive example for compromise between delay and latency. In general, it can be shown [Ben00] that for a multirate system with N rates, the maximum achievable data rate, R_{max}^v , as a function of latency ratio L_{max}/L_{min} is given by

$$R_{max}^v = \frac{\sum_{n=1}^{n_0} p_n + \sum_{n=n_0+1}^N \frac{p_n}{(L_{max}/L_{min})}}{\sum_{n=1}^{n_0} p_n/R_n + \sum_{n=n_0+1}^N \frac{(p_n/R_n)}{(L_{max}/L_{min})}} \quad (11.4.11)$$

where n_0 is the largest integer for which $R_n \leq R_{max}^v$.

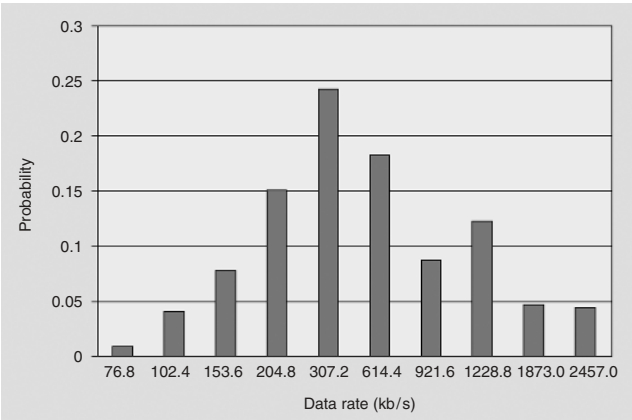
Example 11.13: Capacity of HDR Figure 11.19 shows the data rates supported by the HDR system described in Example 10.27. Figure 11.19a shows different data rates and the minimum SNR requirements for them. Figure 11.19b shows the probability density function of the data rates calculated from a distribution of SNR in a typical embedded sector of a three-sector cellular network. In Fig. 11.20, Eq. (11.4.11) with different latency ratios and the probability density function of Fig. 11.19 are used for calculation of the maximum capacity, R_{max}^v , in the coverage area of one sector. Considering the bandwidth per carrier of 1.24 MHz for the HDR 1xEV, the spectral efficiency of this system ranges from 30 to 60% for different latency ratios.

11.5 DISTRIBUTED CONTENTION-BASED ACCESS CONTROL

When each user has a steady flow of information to be transmitted (e.g., when we have digitized voiced traffic, data file transfer, facsimile transmission, or any type of steady data streaming), connection-based circuit-switched transmission provides an efficient method with low overhead but a poor ability to control the quality of service. If the information to be transmitted is intermittent or bursty in nature, connection-based circuit-switched data transmission can result in communication resources being wasted much of the time. Furthermore, in a system such as cellular mobile telephone, where subscribers pay for service as a function of channel connection time, connection-based circuit switching can be an expensive means of transmitting short messages. Thus,

Data rate (kb/s)	E_c/N_t (dB)
38.4	-12.5
76.8	-9.5
102.6	-8.5
153.6	-6.5
204.8	-5.7
307.2	-4.0
614.4	-1.0
921.6	1.3
1228.8	3.0
1843.2	7.2
2457.6	9.5

(a)



(b)

FIGURE 11.19 Multiple data rates in HDR: (a) SNR requirements for various data rates; (b) probability density function of the data rates [Ben00].

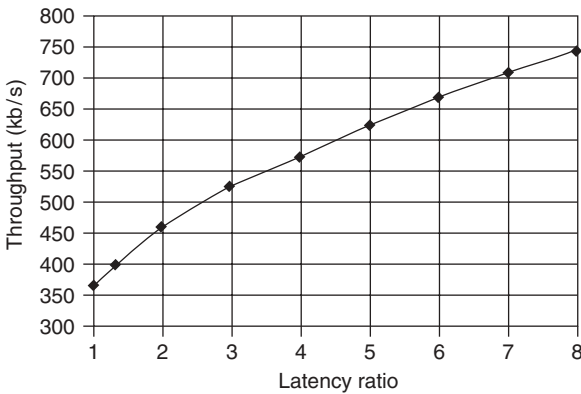


FIGURE 11.20 Maximum achievable data rate (capacity) versus latency ratio for the HDR system [Ben00].

connectionless packet switching has emerged as a more flexible and efficient way of managing transmission of short messages. As we have seen in Chapters 1 and 2, second- and third-generation wireless networks are integrated multimedia networks handling both circuit- and packet-switched data with the same medium access control.

In our discussion of centrally controlled assigned access methods, we noted that such methods use simple algorithms for resource allocation and make relatively efficient use of communication resources when each user has a steady flow of information to be transmitted. This would be well suited to circuit-switched digitized voiced traffic, data file transfer, facsimile transmission, or any other type of data streaming that demands management of quality of service. In centrally controlled assigned access, the user must communicate its needs for resources with the central station before actual assignment of the resources, and this communication imposes a fixed overhead for any assignment. Therefore, centrally controlled assigned access methods cannot be used to start a communication link for a flow of steady information, and they may not be efficient for short packet communication. Contention-based access consists of simple distributed access methods more suitable for packet-switched data. Here we discuss the two classes of the most commonly used contention-based random-access methods, ALOHA-based schemes and carrier-sense multiple-access (CSMA)-based techniques. As we will see later, ALOHA-based access methods are commonly integrated in centrally controlled assigned techniques to be used in wide-area wireless networks, where the propagation delay is relatively high. CSMA-based techniques are more popular in wireless local communication networks, which have smaller propagation delays for signal transmission. In contrast to fixed-assignment access schemes, contention-based random-access schemes provide each user station with varying degrees of freedom in gaining access to the network whenever information is to be sent. A natural consequence of randomness of user access is that there is contention among the users of the network for access to a channel, and this is manifested in collisions of contending transmissions. Therefore, these access schemes are sometimes called *contention-based schemes* or simply *contention schemes*. We begin by describing the simplest and least disciplined of these access methods, the ALOHA protocol.

11.5.1 ALOHA-Based Access Control

The original ALOHA protocol is sometimes called *pure ALOHA*, to distinguish it from subsequent enhancements of the original protocol. This protocol derives its name from the ALOHA system, an experimental wireless data communications network developed by Abramson and his colleagues at the University of Hawaii and first put into operation in 1971 [Abr70, Abr73, Kob77, Ray84, Leu86]. The initial system used ground-based UHF radios to connect computers on several of the island campuses with the university's main computer center on Oahu, by use of a random-access protocol which has become known as the *ALOHA protocol*. Later, the slotted ALOHA and reservation ALOHA access techniques became popular in a number of wireless networks for both circuit- and packet-switched data. We describe the principles of operation, performance, and practical applications of these techniques in the remainder of this section.

Pure ALOHA. The concept of pure ALOHA, shown in Fig. 11.21, is very simple: Users transmit whenever they have information to send. A user sends information in packets, each packet encoded with an error-detection code. Of course, because users

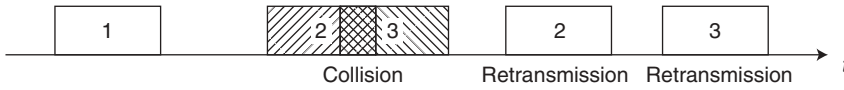


FIGURE 11.21 Collision mechanism in ALOHA.

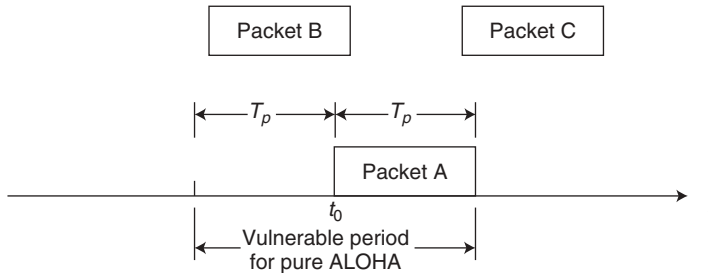


FIGURE 11.22 Packet collision in pure ALOHA. The vulnerable period is two times the packet interval.

transmit packets at arbitrary times, there will be collisions between packets whenever packet transmissions overlap by any amount of time, as indicated in Fig. 11.22. Thus, after sending a packet, the user waits a length of time equal to the round-trip delay for an acknowledgment (ACK) from the receiver. If no acknowledgment is received, the packet is assumed lost in a collision and is retransmitted with a randomly selected delay, to avoid repeated collisions. This is a very simple distributed medium access control method for bursty data packets that does not need passive or active coordination among the terminals and the base station.

Let us assume for simplicity that all packets have a standard length and each packet requires the same amount of time T_p for transmission. Now, referring to Fig. 11.22, consider the transmission of packet A beginning at time t_0 , and let us determine the interval of time during which packet A is vulnerable to collision. If another user starts the transmission of packet B between times $t_0 - T_p$ and t_0 , the end of packet B will collide with the beginning of packet A. This can occur if long propagation times make it impossible for the sender of packet A to know that the transmission of packet B had already begun. Similarly, if another user begins transmitting packet C between times t_0 and $t_0 + T_p$, the beginning of packet C will collide with the end of packet A. From this we can see that the vulnerable interval for packet A is $2T_p$, twice the packet transmission time. If two packets overlap by even the smallest amount of time, each packet will suffer one or more errors, which will be recognized at the receiver by the failure of the error-detection parity-check bits on each packet. The receiver will not be able to acknowledge receipt of either packet, and both will have to be retransmitted.

Let us now determine the channel *throughput* S , which we define as the average number of successful packet transmissions per time interval T_p . Also, let us assume an infinite population of users, and let G be the *traffic intensity* or total traffic “offered” to the channel, defined as the number of packet transmissions attempted per packet time T_p , including new packets as well as retransmissions of old packets. The standard unit of traffic flow is the *erlang*, named for the Danish mathematician A. K. Erlang, whose formulas for traffic engineering was discussed in Section 11.4.6. For our purposes here, we can define an erlang by thinking of the channel time being segmented into intervals

of T_p seconds each; then a traffic flow of one packet per T_p seconds has a value of 1 erlang. By our definition of the throughput S , we see that S cannot exceed 1 erlang without collisions, and thus we expect the throughput to be bounded as $0 < S < 1$. We note further that if the offered traffic load is very low, $G \simeq 0$, there will be very few collisions and in turn very few retransmissions, so we expect that $S \simeq G$ at low traffic load. At very high traffic loads, we expect a large number of collisions and consequent retransmissions, so that we will have $S \ll G$ and S will eventually decrease toward 0.

To calculate throughput S as a function of the traffic load G , we make the standard assumption for the traffic model, in which the probability that k packets are generated during a given packet time obeys a Poisson distribution with a mean of G packets per packet time; that is,

$$P(k) = \frac{G^k e^{-G}}{k!} \quad (11.5.1)$$

The throughput S is simply the traffic load G times the probability that a packet transmission is successful, which we write as $S = GP_0$, where we define P_0 as the probability of no collision, which from our previous discussion is the probability that no other packet is generated during the vulnerable interval of two packet times. Using Eq. (11.5.1), the average rate of packet arrivals in two packet slots is $2G$, and therefore the probability that zero packets are generated in an interval that is two packet times long is $P_0 = e^{-2G}$, and thus the throughput is $S = Ge^{-2G}$, which is plotted in Fig. 11.23. The maximum throughput occurs at traffic load $G = 0.5$, where $S = 1/2e$, which is about 0.184. This means that the best channel utilization achievable with the pure ALOHA protocol is only about 18%. The Poisson arrival model, which assumes an infinite population of users, generally provides a good approximation for a network serving large numbers of terminals. For a finite population, a more accurate model is provided by the binomial distribution [Gan87].

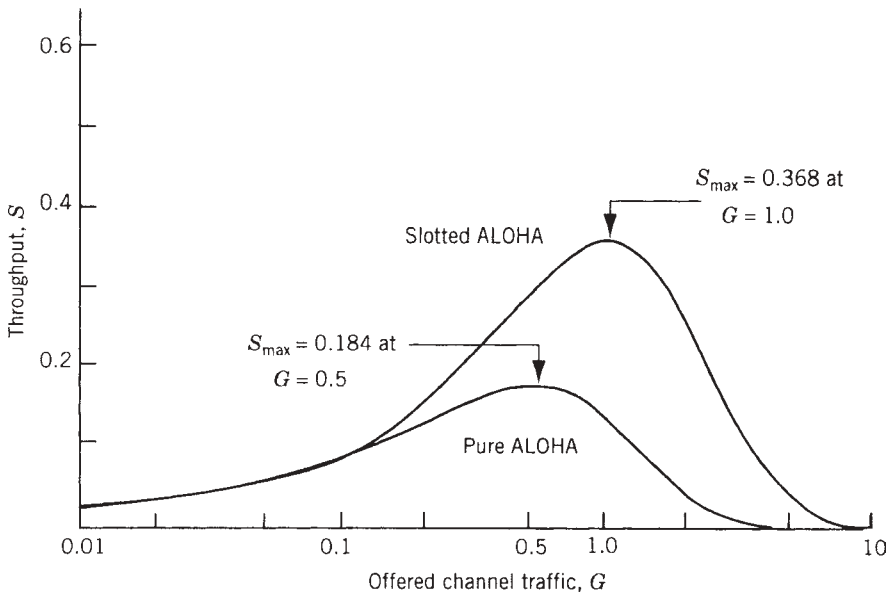


FIGURE 11.23 Throughput versus traffic load for ALOHA protocols.

As we noted above, the traffic offered, G , is the aggregate of all traffic on the network, including newly generated packets as well as retransmitted old packets.

Example 11.14: Relation Between Data Rate and Throughput As a practical example, assume that we have a base station or an access point that supports a maximum transmission rate of 10 Mb/s and serves a large set of mobile user terminals with the pure ALOHA protocol. If only one terminal is activated, it can support a 10-Mb/s streaming application. If all terminals actively transmit, the terminals contending for access to the central module can altogether succeed in getting at most 1.84 Mb/s of information through the network. At that peak the total traffic from the terminals is 5 Mb/s, which is composed of 1.84 Mb/s of successfully delivered packets (some mixture of new and old packets) and 3.16 Mb/s of packets doomed to collide with one another.

As we can see in Example 11.14, the data rate per user in pure ALOHA is a variable function, depending on the number of users sharing the channel and their traffic load. Therefore, in contrast with centrally controlled access methods, pure ALOHA (and in general, all distributed contention access methods) lacks the structure to support the quality of service (QoS) in the medium access control, and one needs to resort to higher layers of protocol to implement QoS.

Slotted ALOHA. To increase the efficiency of the ALOHA protocol, the *slotted ALOHA* scheme was proposed [Rob75]. In this scheme, shown in Fig. 11.24, the transmission time is divided into time slots, with each slot exactly equal to a packet transmission time. All the users are then synchronized to these time slots, so that when a user terminal generates a packet of data, the packet is held and transmitted in the next time slot. The synchronization is usually accomplished by transmitting a periodic beacon signal from one designated station in the network, usually the base station or the access point. With this scheme of synchronized time slots, the interval of vulnerability to collision for any packet is reduced to one packet time from two packet times in pure ALOHA.

Because the interval of vulnerability is reduced by a factor of 2, the probability of no collision, using Eq. (11.5.1), is now $P_0 = e^{-G}$, and therefore the throughput for slotted ALOHA is $S = Ge^{-G}$, which is plotted in Fig. 11.23 along with the throughput for pure ALOHA. The throughput for slotted ALOHA peaks at $G = 1.0$, where $S = 1/e$ or about 0.368, double that of pure ALOHA. Note that at the point of peak throughput, 37% of the slots are carrying successfully transmitted packets, while the same fraction of slots (37%) are empty (this is the probability of having no packets in a time slot, with $G = 1$), and therefore about 26% are in collision. If we try to operate at higher traffic loads, the numbers of empty and successful slots both decrease and the number of collisions increases rapidly. By any reasonable measure, this cannot be considered a very efficient scheme. Somewhat higher levels of throughput are attainable with ALOHA schemes in the presence of signal capture effects; we examine capture effects in Section 11.6. But first we describe access methods that provide higher levels of throughput efficiency by imposing greater discipline on access to the channel.

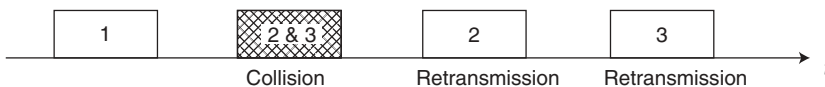


FIGURE 11.24 Collision mechanism in slotted ALOHA (no partial collision).

Slotted ALOHA is commonly used in digital cellular networks for registration of the terminals. When a mobile terminal is turned on, first it passively synchronizes itself with the base station, using special channels used for synchronization purposes. Then, using the slotted ALOHA protocol in a channel usually referred to as a random access channel, its request is sent to establish a connection and register to the network. Only after this initial random access connection does the centrally controlled access mechanism of the cellular network take over the medium access control. The slotted ALOHA protocol is also used in satellite communications networks.

Stability Considerations. In using any contention-based multiple-access technique, there are important issues as to the stability of operation of the network. The fundamental question here is: What user data rates can be supported by the network in stable operation over an acceptable finite period of time? As it happens, this is usually a difficult question to answer for any networks, except for those using the simplest of the random access protocols. It is beyond the purpose of this book to delve into this issue in mathematical detail, but we provide a brief overview of the topic.

In the preceding derivations of throughput functions S versus G for the pure and slotted ALOHA protocols, an important assumption of statistical equilibrium was made in modeling the offered traffic as a Poisson process with a fixed average arrival rate. This model assumes a steady-state condition. That is, the offered traffic, which is a combination of new data packets and retransmissions of earlier packets suffering collisions, remains constant at the value G . However, an examination of Fig. 11.23 will reveal that this assumption may not always be valid.

For example, let us examine the slotted ALOHA throughput curve in Fig. 11.23 and consider some traffic load value $G_1 > 1.0$, that is, an operating point somewhere to the right of the peak of the curve. Let us assume that the network is initially in stable operation at the point selected and then G increases somewhat, due to a statistical fluctuation in the offered load. In response to this increase in G , the throughput S will decrease, as indicated by the shape of the S versus G curve. This reduction in throughput means that there are fewer successful packet transmissions and more collisions. This, in turn, means that the number of retransmissions increases, further increasing both the backlog of messages to be transmitted and the traffic load G , in turn decreasing the throughput S . The end result of this situation is that the operating point keeps moving to the right and the throughput eventually goes to zero. Referred to as *channel saturation*, this is an unavoidable result of operating at values of G greater than the optimum value, 1.0.

The curves plotted in Fig. 11.23 were derived with an assumption of an infinite number of users, a condition that greatly simplifies the derivation of the throughput expressions. The reader might well ask if the unstable operation described above is simply a mathematical artifact arising from an infinite-population assumption. The answer is that with an infinite number of users, all contention-based protocols are inherently unstable, whereas with finite numbers of users, a network may be either stable or unstable, subject to design choices. We discuss these two points in order.

Let us examine the slotted ALOHA curve in Fig. 11.23 once more, now considering an operating point somewhere to the left of the peak of the curve. Again we assume initial operation at some steady load value, say G_2 , and assume a sudden small increase in G , one that keeps the value of G well within the range $0 < G < 1.0$. Because of the shape of the curve, this increase in G results in an increase in throughput S , which

decreases the backlog of messages to be retransmitted and consequently, reduces G toward its initial steady-state value, G_2 . Thus, the system can operate stably around operating points in the region $0 < G < 1.0$ for some length of time. However, if a short-term increase in the traffic load is enough to move the operating point to a point to the right of the peak where the throughput is lower than the initial operating point, unstable operation will ensue and the throughput will eventually go to zero. A key element of this scenario is that with an infinite number of users, even a very large increase in the backlog of messages waiting to be transmitted does not diminish the number of new data packets offered for transmission. This is a heuristic explanation of the statement that contention-based protocols are inherently unstable with infinite user populations. However, the situation is different with a finite number of users, as we describe next.

With a finite number of users, a short-term increase in the number of collisions causes some number of user stations to go into the *blocked state*, thereby reducing the flow of new data packets, while the packets that were unsuccessful previously are cleared through the network by retransmission. This does not mean that having a finite number of user terminals ensures the stability of the network. However, with a finite number of terminals, the retransmission strategy can be designed to ensure stable operation. The key idea here is that stability can be ensured by providing a *backoff algorithm* that spreads retransmissions out over an interval of time sufficient to ensure that short-term increases in traffic load do not trigger a decrease in throughput.

A number of authors have carried out detailed analyses of stability characteristics of contention-based multiple-access protocols. For any given protocol, the analysis begins with derivation of the delay versus throughput characteristic, which incorporates the assumed backoff algorithm for management of retransmissions. For some of the simpler multiple-access protocols, this relationship can be found mathematically. In other cases, simplifying assumptions must be invoked or computer simulations used to develop the delay-throughput relationship. For a given access scheme and retransmission strategy, one then finds that the mean transmission time for a data packet increases as a monotone function of throughput S to a point S' , such that for levels of throughput equal to S' or greater, the average packet-transmission delay is infinite. To delve further into the delay and stability analyses for specific protocols is beyond the objective of this book. We simply conclude the discussion by noting that for networks with finite numbers of users, which of course is the situation of practical interest, stable network operation can always be ensured by proper choice of the buffer size and the retransmission of backoff parameters.

Detailed analyses of the performance of ALOHA schemes, including stability considerations, may be found in the papers of Lam and Kleinrock [Lam74, Kle75b, Lam75] and in the book by Hammond and O'Reilly [Ham86], as well as in references cited therein.

Reservation ALOHA. The *reservation ALOHA* (R-ALOHA) schemes were devised for use on multiuser satellite systems, but they are equally applicable to terrestrial radio systems. There are several versions of the R-ALOHA scheme, each of which can be viewed as a combination of the slotted-ALOHA and TDM protocols, with the apportioning of transmission time to each protocol being varied in response to the traffic demand. Here we describe briefly a scheme due to Roberts [Rob73]. Let us visualize the time axis as divided into fixed-length frames, with the frame length

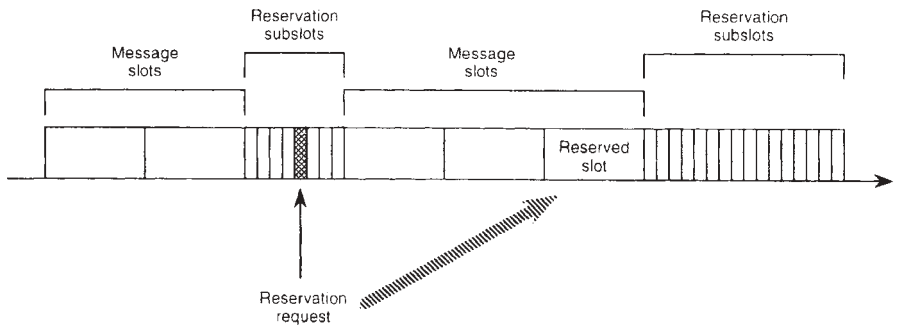


FIGURE 11.25 Reservation ALOHA.

chosen to be longer than the longest propagation delay in the network. Each frame is then divided into equal-length slots, some of which are designated as *message slots*, whereas some are subdivided further into short *reservation subslots*. The mixture of message slots and reservation subslots is different for each of two modes of operation, as we now describe.

Consider Fig. 11.25. In the *unreserved mode*, there are no message slots, and every slot is composed of reservation subslots. User stations with data to send transmit short reservation requests in the reservation subslots using the slotted ALOHA protocol. After transmitting a reservation request, a user waits for the intended recipient to return a positive acknowledgment and a slot assignment. If the reservation request packet does not suffer a collision, the reservation acknowledgment returned advises the sending station when to send its first packet. (The reservation request may have asked for a single slot or multiple slots, up to some design limit.) The system then switches to the *reserved mode*.

In the reserved mode, one assigned slot in each frame is composed of reservation subslots, and all the remaining slots in the frame are available for use as message slots, on a reservation basis. A sending station that has been granted a reservation sends its packets in successive message slots, skipping over the designated reservation slots when they are encountered. Because all reservation exchanges are heard by all stations in the network, each sending station knows which message slots to skip over before starting its own transmission. This access scheme can be viewed as a kind of flexible TDMA in which the contention-based reservation exchanges are confined to the relatively short reservation subslots, while the message slots are shared among stations with data to send in an orderly noninterfering manner. The choice of the number of reservation subslots relative to the number of message slots is a design trade-off issue. The number of subslots should be small enough to keep the transmission overhead low, but large enough to accommodate the expected number of reservation requests. Given the throughput characteristic of slotted ALOHA, a reasonable design choice is to provide about three (i.e., $1/0.36$) reservation subslots per message slot.

It is interesting to note that there is no centralized control function in the R-ALOHA scheme; instead, control of the system is distributed among all the stations in the network. To make this procedure work properly, each station maintains information on the queue of outstanding reservations for all the other stations as well as for the slots at which its own reservations begin. When the queue length drops to zero, the system returns to the unreserved mode, in which all the slots revert to reservation subslots.

Other versions of the R-ALOHA scheme have been developed by Crowther et al. [Cro73] and by Binder [Bin75]. A description of several of the R-ALOHA schemes may be found in [Tan88] and [Tob80a]. A multiple-access scheme similar to R-ALOHA, called *packet reservation multiple access* (PRMA), was discussed earlier in the chapter.

Example 11.15: Dynamic Slotted ALOHA A version of reservation ALOHA called dynamic slotted ALOHA, used in the Mibotex mobile data network, was introduced in the mid-1980s by Ericsson and Swedish Telecomm. Dynamic slotted ALOHA operates in the full-duplex scenario, in which forward and reverse links operate in separate channels. Figure 11.26 shows a typical example of the operation of this access method by illustrating all packets communicated between a base station and three mobile users. In this example, mobile 1 transmits a status report that consists of a short packet that can be accommodated in one random slot. Mobile 2 has a longer message, which requires a reservation. First, mobile 2 sends a request for transmission and then transmits the actual message. The base station controls the entire process, and at the same time in this example, needs to send some messages to mobile 3. Operation is initiated from the base station by transmitting a <FREE> signal. This signal announces to all mobile users that six short slots are available for contention, the number of contention slots being determined by the size of the message intended for mobile 3. Mobile 1 and mobile 2 each select one of the six slots randomly, with mobile 1 picking slot 1 and mobile 2 picking slot 4. If the same slot is used by the two mobiles, we will have a collision and the process will be repeated in the next free slot. In this example, however, we have no collision and the base station will receive the request for transmission from mobile 2 and the status report from mobile 1 while it is sending its message to mobile 3. After completion of reception by mobile 3, this mobile station sends an acknowledgment to the base station and the base station sends an acknowledgment to mobile 1 confirming reception of the status report and an acknowledgment to mobile 2 to grant the following time slot for its message to be sent to the base station. While mobile 2 is sending its message to the base station, the base station will send another message to mobile 3. At the end of this period, the base station sends an acknowledgment to mobile 2 and mobile 3 sends its acknowledgment to the base station. At this stage the transaction cycle among all terminals is completed and another transaction cycle is initiated by the base using another free-slot announcement.

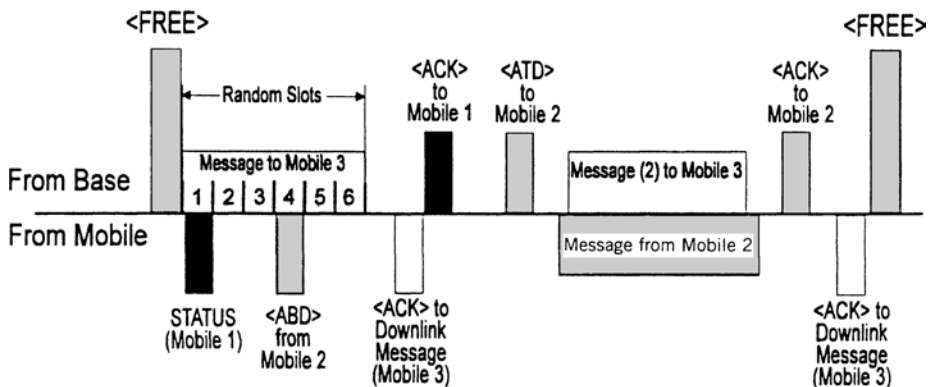


FIGURE 11.26 Example of operation of dynamic slotted ALOHA used in Mobitex.

Example 11.16: R-ALOHA in GPRS/EDGE Services As we explained in Example 11.10, GPRS and EDGE provide a variety of data rates over the GSM bands using one, four, or eight slots of the TDMA access. The medium access control of these systems uses the TDMA slots of GSM to implement a reservation ALOHA access for packet-switched data. For uplink data transmission, in the contention phase a slotted-ALOHA random-access technique is used to transmit reservation requests to the BS on the TDM slots. The BS then transmits a notification to the MS, indicating the channel allocation for an uplink transmission. Finally, the MS transfers data on the allocated slots without contention. On the downlink, the BS transmits a notification to the MS indicating the channel allocation for downlink transmission of data to the MS. The MS will monitor the channels indicated, and the transfer occurs without contention.

11.5.2 CSMA-Based Access Control

The relative inefficiency of the ALOHA schemes lies in the fact that users take no account of what other users are doing when they attempt to transmit data packets, and this leads to a high rate of packet collisions. The collision problem can be dealt with only by adjusting the traffic load offered for maximum throughput, which in turn has the effect of leaving a significant portion of the channel transmission time unused. Much better use can be made of channel resources if a user station listens to the channel before attempting to transmit a packet. This technique, shown in Fig. 11.27, is the basis of several protocols termed *carrier-sense multiple access* (CSMA), sometimes called the *listen-before-talk protocol*. Here we describe several common protocols of this type, which differ primarily in the rules that are followed by a station with data ready to send after sensing the state of the channel. Detailed descriptions and analyses of CSMA systems may be found in papers by Kleinrock, Tobagi, and others [Kle74, Kle75c, Tob75, Tak85, Soh87, Tak87]. Variations of CSMA are used in applications with low transmission delays. The CSMA technique has been widely used in both wired and wireless LANs. A version of CSMA called *data-sense multiple access* (DSMA) was used in ARDIS, CDPD, and TETRA wireless data networks. Another version, called *CSMA with collision detection* (CSMA/CD), is used in legacy IEEE 802.3 Ethernet, the dominant wired LAN standard. The most popular version of CSMA in wireless networking is *CSMA with collision avoidance* (CSMA/CA), which is the most popular of the three medium access techniques recommended by the IEEE 802.11 WLAN standard.

Basic CSMA Techniques. The basic implementation of CSMA has several alternatives for transmission of a packet after a user senses the channel. The simplest implementation of CSMA is one in which each user terminal with data to transmit first listens to the channel to determine if other users are transmitting. If the channel is busy, the user terminal listens continuously, waiting until the channel becomes free and then sends a data packet immediately. This protocol is called *1-persistent CSMA*, signifying its

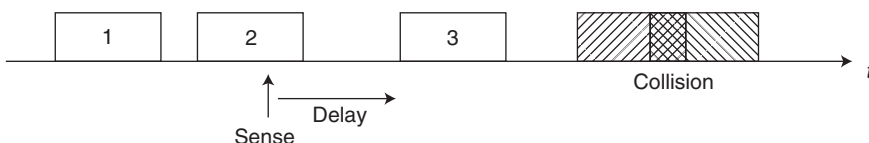


FIGURE 11.27 Collision mechanism in CSMA.

transmission strategy, which is to transmit with probability 1 as soon as the channel is available. After sending the packet, the user station waits for an ACK, and if none is received in a specified amount of time, the user will wait a random amount of time and then resume listening to the channel. When the channel is again sensed to be idle, the packet is retransmitted immediately. Clearly, the objective of this strategy is to avoid collisions with other user packets. However, propagation delays can have a significant effect on CSMA protocols; and with delays, collisions occur despite the discipline of waiting for the channel to be clear. For example, let us say that stations A and B are widely separated and that station A begins to transmit. Until station A's transmission reaches station B, the latter will sense a quiet channel and will be free to start its transmission, which will result in a collision. Even without propagation delay, there can be collisions. Suppose, for example, that stations A and B both have data ready to send and are sensing the channel, which is occupied by a transmission from station C. As soon as station C stops transmitting, the first two stations will begin transmitting simultaneously, again resulting in a collision. Although this protocol does not completely eliminate collisions, and its performance depends on propagation delay, its throughput characteristic for realistic ranges of propagation delay is appreciably better than that of either pure or slotted ALOHA. We summarize the throughput calculations later after describing other versions of the CSMA scheme.

The 1-persistent CSMA scheme also has a slotted form, in which user stations are synchronized and all transmissions, whether initial transmissions or retransmissions, are synchronized to the time slots [Kle75c, Tob75]. The throughput for this scheme is given later. Carrier sensing for the implementation of CSMA can be nonpersistent. With *nonpersistent CSMA*, a user station does not sense the channel continuously while it is busy. Instead, after sensing the busy condition, it waits a randomly selected interval of time before sensing again. As with 1-persistent CSMA, a user with data to send begins transmitting immediately when the channel is sensed to be idle. But here the randomized waiting times between channel sensings eliminate most of the collisions that would result from multiple users transmitting simultaneously upon sensing the transition from busy to idle condition. This leads to throughput values much higher than 1-persistent CSMA at high traffic loads and maximum throughput values of 80% or higher given realistic propagation delays. However, at low traffic loads, the throughput of nonpersistent CSMA is somewhat poorer than that of 1-persistent CSMA, because the waiting strategy is of no benefit when few users are trying to transmit. As with 1-persistent CSMA, there is also a slotted version of nonpersistent CSMA, which is especially beneficial at higher traffic loads [Kle75c].

A third approach for implementation of the basic CSMA is *p-persistent CSMA* protocol, which is a generalization of the 1-persistent CSMA scheme, applicable to slotted channels. The slot length is typically chosen to be the maximum propagation delay. When a station has data to send, it senses the channel; and if the channel is sensed idle, it transmits with probability p . With probability $q = 1 - p$, the station defers action to the next slot, where it senses the channel again. If that slot is idle, the station transmits with probability p or defers again with probability q . This procedure is repeated until either the frame has been transmitted or the channel is sensed to be busy. When the channel is detected busy, the station then senses the channel continuously; and when it becomes free, it restarts the procedure discussed above. If the station senses the channel to be busy initially, it simply waits one slot and applies the procedure described above.

The analysis of throughput for this scheme is rather tedious, and it is not presented here. We simply note that for low-to-intermediate values of propagation delay and with parameter p optimized, the throughput of p -persistent CSMA lies between that of slotted and unslotted nonpersistent CSMA, whereas for long propagation delays, its throughput somewhat exceeds that of either of the other two. Plots of throughput for a few selected cases are given in [Tan88], and detailed analyses are provided in [Kle75c, Tob75, Tak85, Tak87].

Throughput of Basic CSMA Techniques. Table 11.4 summarizes the throughput expressions for the ALOHA and the 1-persistent and nonpersistent CSMA protocols, including the slotted and unslotted versions of each. The expressions for p -persistent protocols are very involved and are not included here. The interested reader should refer to [Kle75c, Tob75, Tak85], where derivations of the other CSMA expressions can also be found. The expressions in the table are also derived in [Ham86, Kei89].

In Table 11.4, throughput S represents the number of packets delivered successfully per packet transmission time T_p , and G is the traffic load in packets offered per packet time. The four CSMA throughput expressions include the normalized delay parameter a , defined as $a = \tau/T_p$, where τ is the propagation delay. The parameter a corresponds to the time interval, normalized to packet duration, during which a transmitted packet can suffer a collision in the CSMA schemes. (Recall that when one station's transmitted packet reaches another station and is sensed there, the second station is inhibited from transmitting.) The propagation delay (3.33 $\mu\text{s}/\text{km}$) is generally much smaller than the packet transmission time, and thus values of a on the order of 0.01 are usually of interest.

Figure 11.28 shows plots of throughput S versus the traffic load G offered for the six protocols listed in Table 11.4, with normalized propagation delay $a = 0.01$. (For $a < 0.01$, the throughput curves for the slotted and unslotted versions of 1-persistent CSMA are essentially indistinguishable.) It can be seen from the figure that for low levels of traffic offered, the persistent protocols provide the best throughput, but at higher load levels the nonpersistent protocols are by far the best. It can also be seen that the slotted nonpersistent CSMA protocol has a peak throughput almost twice that of persistent CSMA schemes.

TABLE 11.4 Throughput Expressions for Two ALOHA and Three CSMA Protocols

Protocol	Throughput
Pure ALOHA	$S = Ge^{-2G}$
Slotted ALOHA	$S = Ge^{-G}$
Unslotted 1-persistent CSMA	$S = \frac{G[1 + G + aG(1 + G + aG/2)]e^{-G(1+2a)}}{G(1 + 2a) - (1 - e^{-aG}) + (1 + aG)e^{-G(1+a)}}$
Slotted 1-persistent CSMA	$S = \frac{G(1 + a - e^{-aG})e^{-G(1+a)}}{(1 + a)(1 - e^{-aG}) + ae^{-G(1+a)}}$
Unslotted nonpersistent CSMA	$S = \frac{Ge^{-aG}}{G(1 + 2a) + e^{-aG}}$
Slotted nonpersistent CSMA	$S = \frac{aGe^{-aG}}{1 - e^{-aG} + a}$

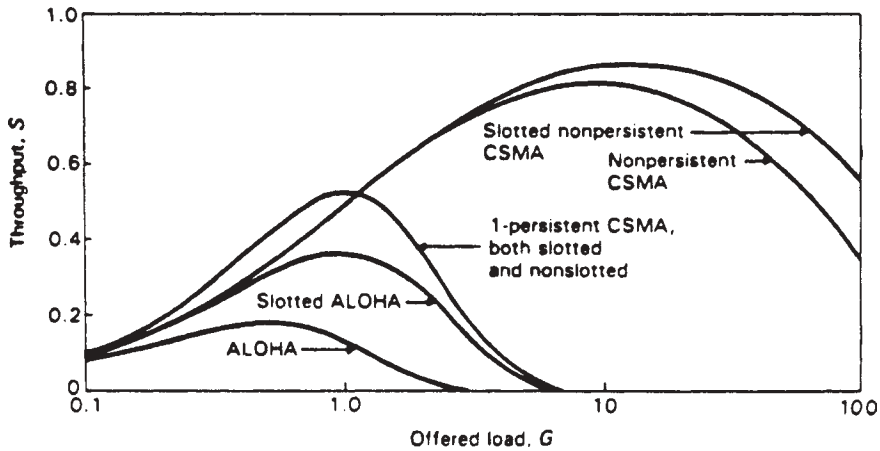


FIGURE 11.28 Comparison of throughput versus traffic load for ALOHA and CSMA protocols. Normalized propagation delay is $a = 0.01$. (From [Ham86] © 1986 Addison-Wesley; reprinted with permission.)

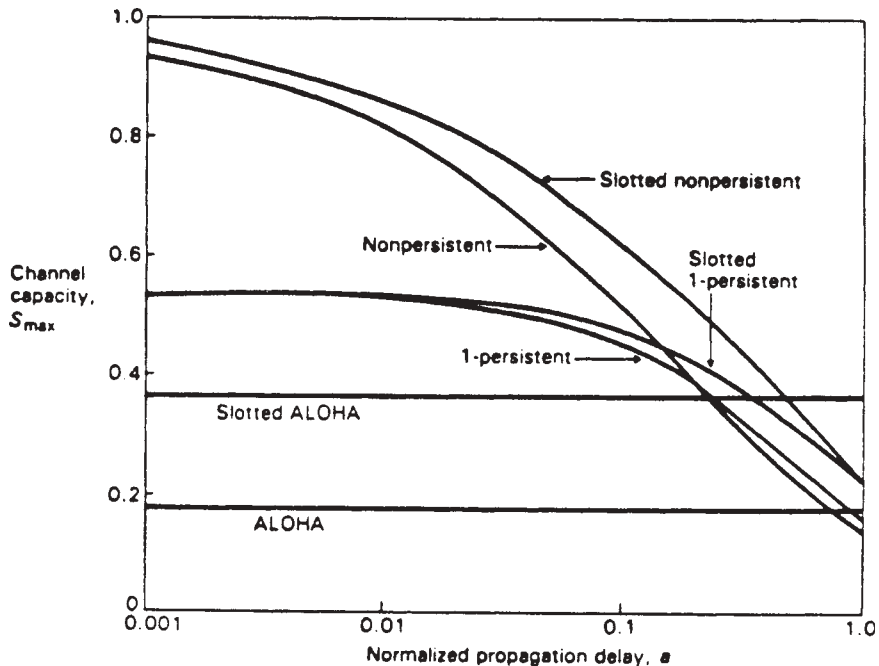


FIGURE 11.29 Capacity or peak throughput versus normalized propagation delay for ALOHA and CSMA protocols. (From [Ham86] © 1986 Addison-Wesley; reprinted with permission.)

The equations in Table 11.4 can also be used to calculate capacity, which is defined as the peak value $S_{\max}$ of throughput over the entire range of traffic load G offered [Ham86]. Curves of capacity versus normalized propagation delay are plotted in Fig. 11.29 for the same set of ALOHA and CSMA schemes. The curves show that for each type of protocol, the capacity has a distinctive behavior as a function of normalized propagation delay a .

For the ALOHA protocols, capacity is independent of a and is the largest of all the protocols compared when a is large. The reason for this is that with long propagation delays relative to packet transmission time, the channel state information arrives too late to be used effectively in reducing collisions. In other words, there is a relatively large time interval in which a sender's packet is vulnerable to collision because the presence of a potentially colliding packet has not yet been sensed.

The curves in Fig. 11.29 also show that the capacity of 1-persistent CSMA is less sensitive than nonpersistent CSMA to normalized propagation delay for small a . However, for small a , nonpersistent CSMA yields a larger capacity than does 1-persistent CSMA, although the situation reverses as a approaches the range 0.3 to 0.5 [Ham86].

CSMA with Busy-Tone Signaling. In wireless networks, it cannot always be assumed that all user terminals are within range and line of sight of each other. Typically, in a radio network, two terminals can each be within the range of some intended third terminal but out of range of each other, separated by excessive distance or by some physical obstacle that makes direct communication between the two terminals impossible. This situation is referred to as the *hidden terminal problem*. Consider as an example a packet radio network with centralized topology in which the central station is positioned so as to have line-of-sight communication with all the other stations but some of the stations cannot communicate directly with certain other stations. This is a likely situation for radio networks covering wide geographic areas in which hilly terrain blocks some groups of user terminals from other groups. If a CSMA protocol is being used in the network, it will successfully prevent collisions among the users of one group but will fail to prevent collisions among users in groups hidden from one another. In this case, given that there is a centralized station in range of all the users, the hidden terminal problem can be solved using a technique called *busy-tone multiple access* (BTMA).

In the BTMA scheme, the system bandwidth is divided into two channels: a *message channel* and a *busy-tone channel*. Whenever the central station senses signal energy on the message channel, it transmits a simple busy-tone signal (e.g., a sinusoid) on the busy-tone channel, and this is in principle detectable by all the user stations. With this method, a user station with a packet ready to send first senses the busy-tone channel to determine if the network is occupied. The procedure the user station then follows in transmission of the packet depends on the particular version of CSMA being used in the network, and any of the various CSMA schemes might be chosen. The BTMA scheme has been used in a number of packet radio networks having multihop topology. In this topology, when a terminal senses a signal on the channel, it turns on its busy tone. In other words, as a terminal detects that some user is on the message channel, it sounds the alarm on the busy-tone channel in an attempt to inform other users. In a military tactical situation (e.g., where mobile units must stay in communication with each other at all times), the BTMA scheme is the solution of choice.

Exact analysis of the BTMA scheme, particularly for a multihop network, involves several issues that do not arise in evaluations of conventional CSMA techniques. These issues revolve around the requirement for reliable detection of a busy-tone signal in a narrowband busy-tone channel. The parameters involved are the busy-tone detection-time window, the false-alarm probability, and the fraction of system bandwidth devoted to busy-tone signaling. (The situation is rather less complicated for a centralized network, where the throughput can be approximated by that of CSMA [Zdu89].) What is

interesting to note is that even if there are no hidden terminals, BTMA, with parameters properly chosen, can provide better throughput than that of other CSMA schemes. A detailed analysis of the nonpersistent version of the BTMA scheme is given by Tobagi and Kleinrock [Tob75]. A brief summary of BTMA is given in [Tob80a], where a number of packet communication protocols are discussed and compared.

Data-Sense Multiple Access. Digital or data-sense multiple access (DSMA) is very popular in full-duplex wireless data communication networks such as CDPD, ARDIS, and TETRA. In these networks, communication from the mobile to the base station and from the base to the mobile station are performed in separate frequency channels using different access methods. The link from the base to the mobile is referred to as the *forward channel or downlink* and uses a version of TDMA in which the messages for different mobile stations occupy nonoverlapping time slots. In the downlink, power control, synchronization, and other centralized control signals are also time-division-multiplexed with the information packets. The communication channel from the mobile to the base station, which is referred to as the *reverse channel or uplink*, uses a version of CSMA called DSMA. Interleaved among other signals broadcast on the forward channel, the base station transmits a *busy-idle* bit in each time frame to report the status of the reverse channel. A mobile station will check this *flag* bit before transmission. If it indicates that the channel is idle, the mobile station will proceed to send its packet in the following time slot. As soon as the transmission is started, the base station will switch the flag bit to the busy state until the transmission from the mobile terminal is completed.

With DSMA, we cannot eliminate collisions completely, and thus we have to devise a retransmission strategy similar to that for CSMA. The announcement of the occupancy of the reverse channel from the base station is very similar to the busy-tone multiple-access (BTMA) protocol used in a centralized topology, with the difference that with BTMA, a separate channel is dedicated to announcements on the status of the reverse channel, whereas with DSMA only one bit in each frame of the TDMA forward channel is devoted to the reverse channel status report.

CSMA with Collision Detection. Persistent and nonpersistent CSMA protocols improve on the ALOHA schemes by ensuring that no user station begins to transmit when it senses that the channel is occupied. One might well ask if a sending station can make further use of channel state information after its own transmission is under way. We shall briefly describe a technique that does precisely this. The technique is called *carrier-sense multiple access with collision detection* (CSMA/CD) and is sometimes referred to as the *listen-while-talk protocol*. The CSMA/CD technique can be used in nonpersistent, 1-persistent, or *p*-persistent variations of CSMA, each with a slotted or unslotted version. The defining feature of CSMA/CD is that it provides for detection of a collision shortly after its onset, and each transmitter involved in a collision stops transmission as soon as it senses a collision. In this way, colliding packets can be aborted promptly, minimizing the channel transmission time occupied by transmissions destined to be unsuccessful.

In the operation of CSMA/CD, if the channel is sensed to be idle or busy, a user station with a packet to send first follows the procedure dictated by the particular protocol variation in use. However, each station uses hardware that not only monitors the channel before transmission, but also monitors while transmitting. Unlike CSMA, which requires an acknowledgment (or lack of an acknowledgment) to learn

of the status of a packet collision, CSMA/CD requires no such feedback information, because the collision detection mechanism is built into the transmitter. If a collision is detected, the transmission is aborted immediately, a jamming signal is transmitted, and a retransmission back-off procedure is initiated, just as in CSMA. The jamming signal is not an essential feature of CSMA/CD, but it is used in many implementations of this access method. The jamming signal serves to force consensus among users as to the state of the network, in that it ensures that all other stations know of the collision and go into back-off condition. As is the case with any random-access scheme, proper design of the back-off algorithm is an important element in ensuring stable operation of the network.

The delay and throughput characteristics of CSMA/CD have been analyzed by Lam [Lam80] and others [Tob80b, Coy83, Med83, Cho85, Coy85, Hey86, Tas86]. Throughput and stability analyses of CSMA/CD may also be found in texts by Hammond and O'Reilly [Ham86] and Keiser [Kei89].

The initial development of CSMA/CD was done at Xerox in the early 1970s for application in local-area networks [Met76b]. Further development work was done in a joint effort by Digital Equipment Corporation, Intel, and Xerox, leading to the detailed specification for Ethernet, one of the first commercially available LAN products [Dig82]. Although the name *Ethernet* is often used, somewhat inaccurately, as a protocol designation, it in fact refers to a specification encompassing a set of products produced by several manufacturers and endorsed by preliminary IEEE standards. The IEEE 802.3 CSMA/CD standard for LANs is based on the Ethernet specification and is nearly identical to it [ANS85, ANS88]. The Ethernet specification and the IEEE 802.3 standard specify use of the 1-persistent version of CSMA/CD. The back-off algorithm recommended by IEEE 802.3 is referred to as the *exponential back-off algorithm*. In this algorithm, when a collision occurs, retransmission is scheduled with a probability of $\frac{1}{2}$ after a time slot that spans twice the maximum propagation delay allowed between the two terminals (512 μ s). A time slot that spans twice the maximum propagation delay is selected to ensure that in the worst-case scenario, the terminal will be able to detect the collision. If a second collision occurs, the terminals reattempt with a probability of $\frac{1}{4}$, which is half of the previous retransmission probability. If collision persists, the terminal continues reducing its retransmission probability by half up to 10 times, and after that it continues with the same probability six more times. If no transmission is possible after 16 attempts, it reports to the terminal that the network is congested and transmission is stopped. The disadvantage of this procedure is that the packets arriving later have a higher chance of surviving the collision, which results in an unfair first come–last served environment.

Although collision detection is performed easily on a wired network simply by sensing voltage levels, such a simple scheme is not readily applied to wireless channels. One thing that can be done is to have the transmitting station demodulate the channel signal and compare the resulting information with its own transmitted information. Disagreements can be taken as an indicator of collisions, and the packet can be aborted immediately. However, on a wireless channel the transmitting terminal's own signal dominates all other signals received in its vicinity, and thus the receiver may fail to recognize the collision and simply retrieve its own signal. To avoid this situation, the station's transmitting antenna pattern should be different from its receiving pattern. Arranging this situation is not convenient in radio terminals because it requires directional antennas and expensive front-end amplifiers for both transmitter and receiver.

Therefore, CSMA/CD has not been adopted for the WLANs using radio technology, and detecting collisions with an acknowledgment scheme is the approach typically taken in radio LANs. However, CSMA/CD has been used in several infrared LANs, where both transmission and reception are inherently directional and the design of the receiver front end is inexpensive. In direct-beam IR (DBIR) LANs, the transmitted optical signal is narrowly focused by design. In diffused IR (DFIR) LANs, the radiation is made directional by packaging of the LEDs. The field of view of the receiver photodiode provides the directionality for the received signal in both DFIR and DBIR LANs. In such an environment, detection of a station's own transmitted signal is readily feasible.

CSMA with Collision Avoidance. Since carrier detection in a radio environment is challenging, the approach called *CSMA with collision avoidance* (CSMA/CA) has been adopted by the IEEE 802.11 WLAN standard as its contention access method for infrastructure and ad hoc network topologies. As shown in Fig. 11.30, in the CSMA/CA of IEEE 802.11, we have three time periods: interframe spacing (IFS), contention window (CW), and a back-off counter time. The CW intervals are used for contention and transmission of the packet frames. The IFS is used as an interval between two CW intervals. The back-off counter is used to organize the back-off procedure for transmission of packets to avoid collision. Whenever a terminal has a packet to transmit, it senses the channel. If the channel is idle, it transmits the packet. If the channel is busy, it waits for completion of the packet transmission on the air and runs a random counter to determine a random length for backoff. After completion of the transmission of the packet in the air, the terminal waits for an IFS. Then it starts its counter to count down while listening to the channel. If another user's packet appears on the air before completion of the countdown, the terminal freezes the counter and waits for completion of that packet. If the countdown is completed without the appearance of another packet on the air, the packet is transmitted. The method of operation is best described by an example.

Example 11.17: CSMA/CA in the IEEE 802.11 Figure 11.31 provides an example of operation of the IEEE 802.11 CSMA/CA for five stations (A, B, C, D, and E). Station A has a frame on the air when stations B, C, and D sense the channel and find it busy. Each of the three stations runs its random-number generator to select a backoff time. Station C, followed by D and B, draws the smallest number. All three terminals persist in sensing the channel and defer their transmissions until transmission of the frame from terminal A is completed. After completion, all three terminals wait for the IFS period and start their counters. As soon as the first terminal, station C in this example, finishes counting its waiting time, it starts transmission of its frame. The other two terminals, B and D, each freeze their counter to the value that they have reached at the start of transmission for terminal C. During transmission of the frame

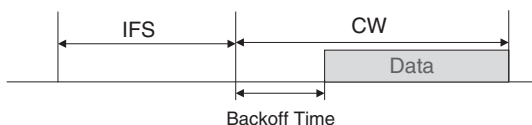


FIGURE 11.30 Basic time intervals in the CSMA/CA adopted by IEEE 802.11.

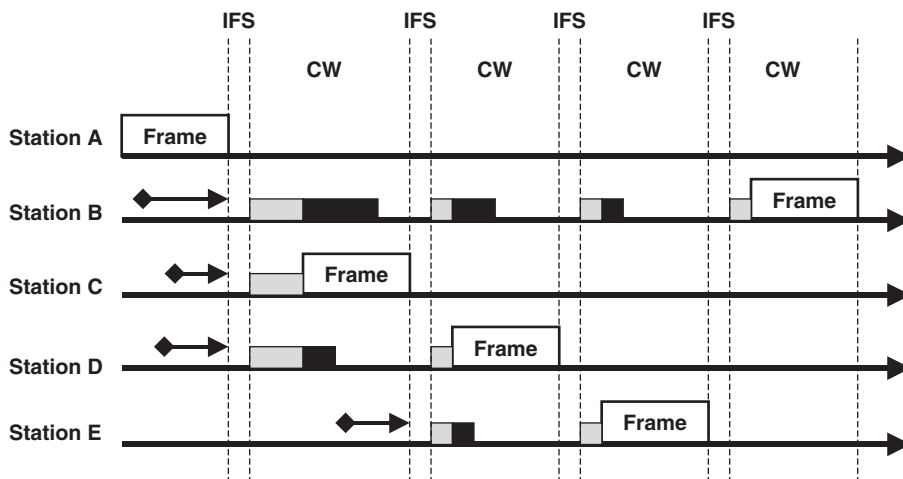


FIGURE 11.31 CSMA/CA operation for five terminals.

from station C, station E senses the channel, runs its own random-number generator, which in this case ends up with a number larger than the remainder of D and smaller than the remainder of B, and defers its transmission until after the completion of station C's frame. In the same manner as the previous instance, all terminals wait for IFS and start their counters. Station D runs out of its random waiting time earlier and transmits its own packet. Stations B and E freeze their counters and wait for completion of the frame transmission from terminal D and the IFS period after that before they start running down their counters. The counter for terminal E runs down to zero earlier, and this terminal sends its frame while B freezes its counter. After the IFS period following completion of the frame from station E, the counter in station B counts down to zero before it sends its own frame.

The advantage of this back-off strategy over the exponential back off used in IEEE 802.3 is that the collision detection procedure is eliminated, and the waiting time is fairly distributed in a way that, on average, a first come–first served policy is enforced.

Throughput of CSMA/CA. The exact calculation of the throughput of CSMA/CA is difficult because it is a function of the data rate, length of the packet, average duration of the back-off counter, and length of IFS. Therefore, computer simulations are good tools to use in gaining some insight into the operation of this protocol.

Example 11.18: Simulated Throughput of CSMA/CA in the IEEE 802.11

Figure 11.32 shows the simulated throughput of CSMA/CA for packet lengths of 64 and 576 bytes (8 bits). The overhead per packet is 66 bytes, which includes the medium access, logical link control, and physical layer overhead bits. In addition, CSMA/CA operates with 16-byte acknowledgement packets in the medium access control and interframe spacing, during which no information is transmitted. These are added together, and in this simulation, for long and short packets we have a maximum throughput of less than 50% and about 20%, respectively. The CSMA/CD used in IEEE 802.11 does not have acknowledgements and overhead is less, which brings the

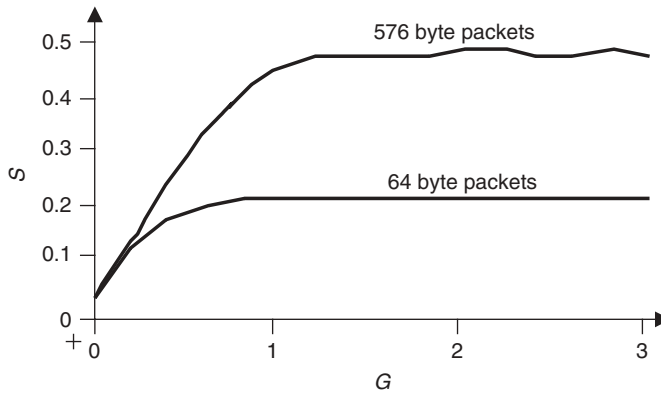


FIGURE 11.32 Simulated performance of the IEEE 802.11 CSMA/CA algorithm for a 2-Mb/s transmission rate and packet lengths of 4.6 kb and 512 bits.

throughput for long packets close to 90%. Fast acknowledgment in the medium access control helps the medium access to cope with the unreliability of the wireless medium caused by multipath fading and shadowing.

In wide-area mobile data networks, as we showed in several examples in Section 11.4.8, centralized medium access control using TDMA/FDD, which is well-suited to the implementation of QoS, is very popular. However, the access method for IEEE 802.11 WLANs is CSMA/CA. To explain the difference between throughput of the TDMA-based medium access control techniques used in the HIPERLAN-2 standard and the CSMA/CA used in IEEE 802.11, we provide an example.

Example 11.19: CSMA/CA Versus TDMA/TDD The HIPERLAN2 standard uses a TDMA/TDD protocol for medium access control over the same physical layer of the IEEE 802.11a. In this access method, communication for uplink and downlink is based on 2-ms time slots. These are divided in three parts: two parts used for uplink and downlink information and one part for control signaling, which allows random access and other control functions. The uplink and downlink parts of the time slot carry short and long packets of length 9 and 54 bytes. Figure 11.33 compares the simulated performance of the CSMA/CA used in the IEEE 802.11a and the TDMA/TDD used in the HIPERLAN2 standards for five users. As shown in Fig. 9.30, the OFDM modulation recommended for IEEE 802.11a/HIPERLAN2 multirate modems originally employed seven data rates: 6, 9, 12, 18, 27, 36, and 54 Mb/s. As the data rate increases ninefold, from 6 Mb/s to 54 Mb/s, the duration of the packets in time decreases ninefold. As we observed in Example 11.18, ninefold changes in packet duration reduces the throughput to less than one-half in a CSMA environment, making this protocol less efficient at higher data rates. The overhead of the TDMA/TDD is a fixed fraction of the transmission rate, which results in much better performance at higher data rates.

As we discussed in Section 11.4.7, TDMA/TDD is also more efficient for implementation of QoS. The question then raised by a casual reader not familiar with the evolutionary progress in systems engineering would be why CSMA/CA is the dominant medium access control for WLANs. CSMA/CA was a wireless version of CSMA/CD,

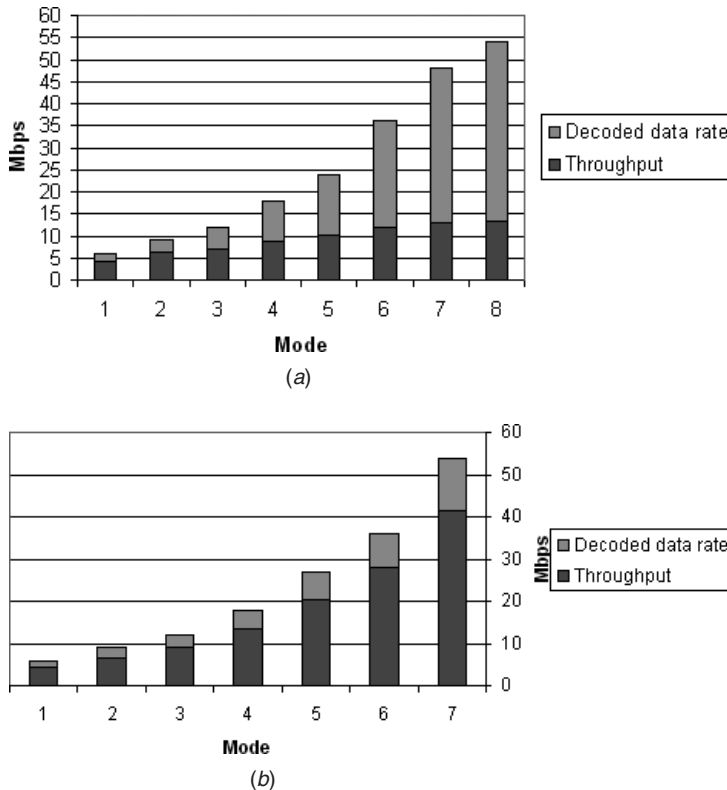


FIGURE 11.33 Throughput of the CSMA/CA used in (a) IEEE 802.11a versus (b) TDMA/TDD used in HIPERLAN2 for data rates of 6-, 9-, 12-, 18-, 27-, 36-, and 54-Mb/s transmission rates.

the dominant medium access control for wired Ethernet, and it appeared in the market early in the evolution of the WLAN industry. Resorting to a new medium access control creates challenges in providing backward compatibility of the devices, that have not attracted adequate attention and timely solutions. Emerging IEEE 802.15.3 WPANs, however, are considering centrally controlled medium access control techniques, and we provide practical examples for those systems in Chapter 12.

QoS in IEEE 802.11 Medium Access. Contention-based protocols evolved for bursty connection-based data applications. Since there is no mechanism for delay control, these access methods are not well suited for time-bounded telephone conversations, for which the QoS is controlled by keeping the delay and packet loss rate under certain levels (around 100 ms for delay and 1% for packet loss). In contention access methods such as CSMA/CA used with packet-switched data networks, there is no mechanism to ensure the delay, and indeed, packets arrive with a jitter added to the fixed propagation delay. In telephone applications, the received voice packets are played as a steady stream for the user at the encoding rate of the voice. To compensate for the jitter and provide for a steady stream of voice packets (i.e., packet-switched voice applications), a jitter-compensating buffer is used at the receiver. The input to the buffer is the received packets, with jitter caused by the contention access and store-and-forward operation in

the routers used in the packet-switching network. The output of this jitter-compensating buffer is a steady stream of voice packets at the rate of encoded speech. The total delay encountered by a received packet is the propagation delay plus the random jitter delay. The total delay for the packets at the output of the buffer is the propagation delay plus delay associated with the fixed length of the jitter compensation buffer. Since the packets are played in a stream of time slots at a fixed rate, when the delay of the arriving packet is more than the propagation delay plus the buffer queuing delay, the packet does not make it to the appropriate slot in the stream, and it gets discarded. The longer the buffer length (larger delay), the lower is the packet loss rate. In contention-based access methods, another parameter affecting the performance is the number of simultaneous voice users. The larger the number of users, the higher is the arrival delay or packet loss. Therefore, for a voice-over contention-based access and packet-switching environment, it is useful to generate curves relating buffer size, number of users, and packet loss.

The relationship among the packet-loss rate, P_{lost} , the delay of the jitter compensation buffer, D_{BUF} , and the probability density function of the jitter packets, $f_J(j)$, is very simple and is given by

$$P_{\text{loss}} = \int_{D_{\text{BUF}}}^{\infty} f_J(j) dj \quad (11.5.2)$$

This equation simply states that the probability of losing a packet is the same as the probability of a packet arriving late for the stream. The jitter in this equation is the delay of the received packet minus the propagation delay, and when this delay is more than the delay associated with the length of the buffer, the packet is late for the stream and is discarded. We conclude this discussion with a numerical example of such curves.

Example 11.20: Number of Voice Users in CSMA/CA To determine the relationship between the jitter compensation delay and the packet loss rate in CSMA/CA, a testbed was developed in [Fei00]. In the testbed an infrastructure for IEEE 802.11 WLAN operation using an AP and a number of laptops is used for measurement of the probability density function of the delay jitter in a voice-over-IP application. A set of probability densities functions for 1800 transmitted packets for different lengths of delay compensation buffer and number of users is measured in the testbed. Using the delay jitter densities and Eq. (11.5.2), the set of curves shown in Fig 11.34 is generated to illustrate the relation between packet loss rate and the length of the jitter compensating buffer for various numbers. The figure shows the experimental results for up to five mobile terminals connected to an AP. These curves can then be used to determine the required buffer length to assure a certain level of QoS when we have different numbers of users. If we fix the acceptable packet loss rate at 1%, the minimum buffer length is 0.5 and 5.5 ms for one and five users, respectively. The details of algorithms for implementation of the testbed and the results of OPNET simulation for a large number of voice users are available in [Fei99].

The application software used for the communication of time-bounded information, such as telephone conversations or audio/video streaming, uses the basic principles described in Example 11.20 and variable buffer lengths to optimize the QoS for time-bounded information. This approach for enforcing a QoS is based on the basic

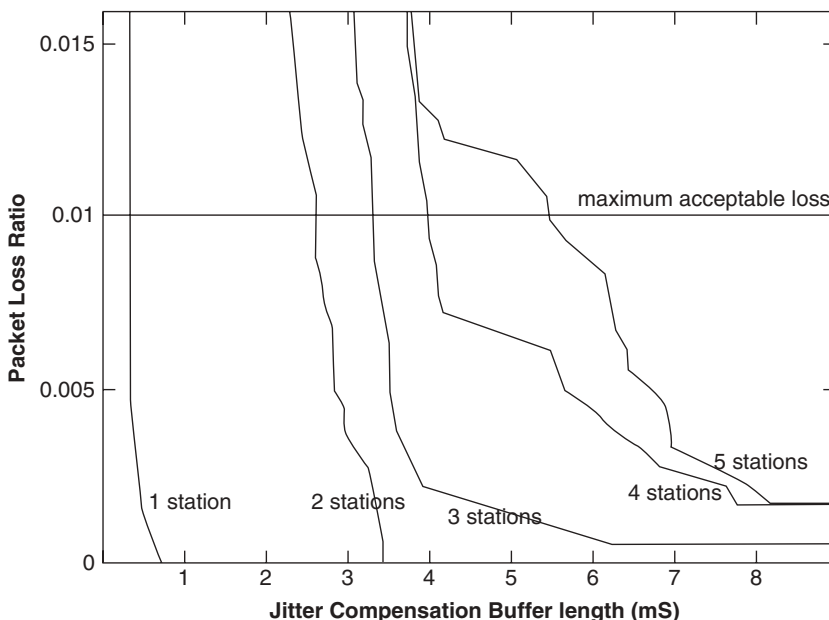


FIGURE 11.34 Packet loss versus jitter compensation buffer length for CSMA/CA used in IEEE 802.11.

CSMA/CA contention access of the IEEE 802.11 medium access control. Two other medium access control mechanisms recommended by the IEEE 802.11 are useful for QoS support: variable interframe spacing and the point coordination function (PCF). In the general description of the CSMA/CA of the IEEE 802.11 shown in Fig. 11.30 and the example in Fig. 11.31, we introduced only one IFS. In the original IEEE 802.11, there are three interframe spacings: the longest distributed coordination function IFS (DIFS), the medium-sized PCF interframe spacing (PIFS), and the short IFS (SIFS). After completion of a transmission on the air, a packet waiting for the next available time can choose one of the three IFSs before it starts transmission. Obviously, packets using SIFS can get on air faster than those using PIFS, and those using DIFS have the lowest priority. In this way, the medium access has an embedded three-level priority mechanism. Packets with the highest priority, such as acknowledgment packets, use SIFS; time-bounded information uses PIFS; and general contention data packets use DIFS. The IEEE 802.11 subcommunity on QoS is considering increasing the number of interframe spaces to provide more flexibility to control QoS at the medium access level.

The PCF is a separate medium access mechanism recommended by IEEE 802.11. This mechanism is not implemented in today's popular commercial products, but the recommendation exists. This mechanism is indeed a centrally controlled mechanism that allows integration of time-bounded and bursty data. In the PCF the access point captures control of the transmissions and allows terminals with time-bounded information to have a pseudoperiodic access for the duration of their packets while other terminals are denied access. Figure 11.35 illustrates the basic operation of the PCF. Time-bounded data are transmitted periodically, while other terminals are denied access. During idle times other terminals send random bursts through contention using

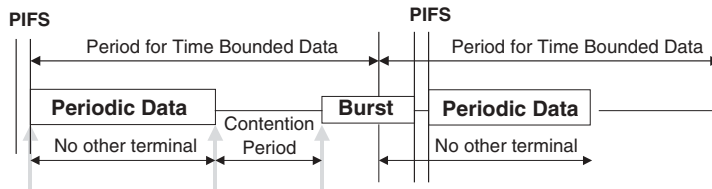


FIGURE 11.35 Basic operation of the PCF mechanism in IEEE 802.11 medium access.

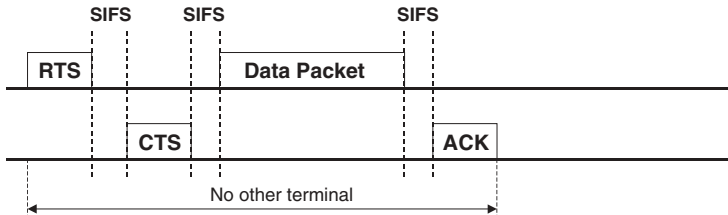


FIGURE 11.36 RTS/CTS mechanism in IEEE 802.11 medium access.

CSMA/CA. If a burst of data started in the idle periods stretches over the starting time of the periodic data (see Fig. 11.35), other terminals are stopped at the beginning of the periodic transmission time, but the transmission of the periodic data itself is deferred to after completion of transmission of the contention packet burst. Pseudoperiodic transmission ensures steady access. The length of the periodic data packets is variable, supporting a variable rate encoding for audio/video applications.

In addition to CSMA/CA and PCF, the IEEE 802.11 standard specifies a third alternative to medium access control, the *request to send, clear to send (RTS/CTS) mechanism*. This mechanism was originally designed to address hidden terminals having difficulty accessing the medium, and operates similarly to the polling technique described in Section 11.4.7. As shown in Fig 11.36, for RTS/CTS access the terminal starts by transmitting a short RTS packet to the destination and informs all other terminals not to transmit anymore. The destination terminal responds with another short packet, CTS, allowing the source terminal to transmit its data packet. After completion of the data transfer, the destination terminal sends its acknowledgment packet, which informs other terminals that the medium is available again. If we compare the performance of the RTS/CTS in a multirate WLAN environment with those of CSMA/CD and TDMA/TDD (e.g., those depicted in Fig. 11.33), the RTS/CTS at its best can perform slightly better than CSMA/CA, but it is still much less efficient than TDMA/TDD at high data rates.

11.5.3 Wireless Access and Radio Channel Characteristics

As we discussed earlier, the received power in a radio communication system is subject to extensive power fluctuations caused by the near–far effect as well as by the locations and movements of terminals and the fading characteristics of the medium. In point-to-point communications the extensive power fluctuations caused by the movements of terminals may cause temporary degradation of a communication link or may even hide a terminal completely from direct communication with another terminal. The effects

of power fluctuation in a multiuser environment depend largely on the multiple-access method employed in the system. In FDMA/TDMA cellular systems, power fluctuations will increase adjacent-channel and co-channel interference. The additional interference results in degradation of the quality of the communication link, but the throughput or capacity of the channel remains unaffected. As we discussed in Chapter 10, in CDMA systems, the capacity of the channel is limited by interference from other terminals that share the same time and frequency space. Therefore, if power fluctuations in a CDMA system are not controlled, the high power levels received from some terminals will sharply limit the number of users than can be on the system, resulting in drastic reduction in the overall capacity of the system. The quantitative analysis of capacity in the presence of multipath fading was discussed in Chapter 10. Power control is the traditional method for counteracting the effects of power fluctuations in multiuser radio systems. Power control will serve to minimize the interference in systems using fixed-assignment multiple-access methods, resulting in improvement in the quality of the received signal in FDMA/TDMA and improvement in capacity in CDMA systems. The superior performance offered by CDMA relative to other access methods cannot be realized unless effective power control is implemented. For FDMA/TDMA systems, power control is an added feature that improves communication quality. Another attractive feature of power control for any portable or mobile network is the saving in electronic power consumption, which is important for battery-powered applications.

In networks using contention access methods, the difference in power levels received from different stations is sometimes helpful. If there is a large difference in the power levels received from two terminals and data packets from the two terminals collide, one of the packets will survive the collision and only one packet will be destroyed, resulting in an increase in the throughput of the network. In reality, the throughput from the terminal with higher power will increase, while the throughput from the other terminals will remain the same. This phenomenon, which we discuss below, is referred to as *capture*.

Capture Effects in Contention-Based Protocols. In the analyses of throughput given earlier it was assumed that if two packets appear in the channel with any amount of overlap in their transmission intervals, the resulting collision renders both packets unusable. Terminals sending the colliding packets then each wait a randomly selected amount of time and retransmit. The throughput expressions for the ALOHA and CSMA protocols, summarized in Table 11.4, were all derived with the assumption that no packets survive collisions. Sometimes, collision of two packets in a radio channel may not destroy both packets. Because of signal fading, packets from different transmitting stations can arrive with very different power levels, and the strongest packet may survive a collision. In some situations, the differences in received signal power levels may simply be due to large differences in the length of transmission paths in the network, referred to as the *near-far effect*. This situation, in which one of the packets involved in a collision is received successfully, is termed *capture*, and the surviving packet is said to have captured the packets with which it collided. The possibility of a received packet surviving a collision depends on channel characteristics as well as the design of the modulation and coding scheme, the average received SNR, and the length of the packets. This is usually quantified by determining, for a given receiver, the minimum power ratio that one arriving packet must have relative to other colliding packets in order that it can be received successfully. This power ratio is referred to as the *capture parameter* or *capture ratio* of the receiver.

Capture effects in a random-access protocol lead to increased levels of throughput relative to results derived without accounting for capture. Many studies of capture effects on contention-based protocols have been reported, many dealing with ALOHA systems [Rob75, Met76a, Abr77, Kah78, Fra80, Kup82, Kle84, Nam84a, Nam84b, Nel84, Arn87, Gan87, Goo87, Hab88, Pra88, Zha89, Zha92] and others with CSMA systems [Zdu89, Zha91]. In [Gan87, Goo87] the analysis focuses on the near-far effects. The ratio of powers received from two terminals is formed and compared with a capture parameter, which takes path loss into account as a function of locations of transmitting terminals relative to the receiver. If the power ratio exceeds the capture ratio, the packet with the higher received power survives the collision. The difference in the received power levels is due solely to the different distances of user terminals from the central station, the terminals assumed to be uniformly distributed over the coverage area. The studies reported in [Arn87, Pra88] deal with Rayleigh-fading and lognormal-fading channels, respectively, and also consider a general spatial distribution for radio terminals. Other investigations have in addition considered the effects of signal design in the analysis of capture. An analysis for a system with uniformly distributed terminals using ALOHA on indoor radio channels is given in [Hab88], where the performance of modulation and coding in fast Rayleigh fading are taken into account. Capture discussions in the remainder of this section closely follow Zhang and Pahlavan [Zha91, Zha92], where slotted ALOHA and CSMA are treated with general distributions of terminals and assumptions of fast and slow Rayleigh fading.

Performance of a Slotted ALOHA System in Capture. Let us consider a system configured as an ideal slotted ALOHA network with a base station located in the center and terminals distributed around it with a given distribution. We shall assume that the system has negligible propagation delay, perfect acknowledgments from the receiver, and an infinite number of terminals.

For a slot length T_p and an average packet generation rate from all the terminals λ , the average number of packets arriving in a slot is $G = \lambda T_p$. If the arrival process is Poisson, the probability $P(k)$ that k packets arrive in a slot is calculated from Eq. (11.5.1).

At the beginning of every slot, we assume that the terminals generate a total of $k + 1$ packets. For this analysis, one of these packets is randomly chosen to be the *test packet*, which is phase-locked to the receiver, and the other k packets are considered to be interference to the test packet. Defining $P_C(k)$ as the probability that the test packet captures the k interfering packets, the average throughput of the system associated with this probability of capture is the average number of packets received successfully per time slot, given by

$$S = \sum_{i=1}^{\infty} P(k+1)P_C(k) \quad (11.5.3)$$

In conventional analysis of slotted ALOHA, it is assumed that in each collision all colliding packets are destroyed, and a packet survives only if it is received without collision. This implies that $P_C(0) = 1$ and $P_C(k) = 0$ for $k \geq 1$, which yields

$$S = Ge^{-G} \quad (11.5.4)$$

for conventional slotted ALOHA. Equation (11.5.4) is the same as before and provides an absolute lower bound for the performance, if transmission errors are neglected, and

is usually referred to as the *case with no capture*. In reality there are always transmission errors, and $P_C(0)$ is not exactly equal to 1. However, with packet error rates in the range 10^{-3} to 10^{-5} , $P_C(0)$ can be assumed approximately equal to 1.

In the presence of capture, some of the packets involved in a collision will survive. In an ideal situation, one packet survives all collisions with k interfering packets. This case, for which $P_C(k) = 1$ for all values of k , is referred to as *perfect capture*. Substituting $P_C(k) = 1$ into Eq. (11.5.3), perfect capture provides an upper bound on the throughput of slotted ALOHA with capture, given by

$$S = 1 - e^{-G} \quad (11.5.5)$$

For large values of G , the throughput approaches 1 and the channel is fully utilized in this ideal case of perfect capture in every time slot. The results presented in the remainder of this section are for practical cases, and the performance results in each instance will fall between the two bounds provided in Eqs. (11.5.4) and (11.5.5).

In general, the probability of capture is a function of modulation and coding, distribution of user terminals, signal-to-noise ratio, and packet length. Calculations of this sort are beyond the scope of this book. Zhang and Pahlavan [Zha92] provide an accurate method of analysis and two bounds on a generalized probability of capture on slow Rayleigh-fading channels. The results in [Zha91] are derived for ring- and bell-shaped distributions of user terminals. The ring distribution is equivalent to an assumption that the same average power is received from each terminal, which models a system having *average power control* but no means of tracking the instantaneous power fluctuations caused by multipath fading. The bell-shaped distribution assumes that the distance power gradient is 4 and the normalized distance r between transmitter and receiver obeys a distribution of the form

$$\rho(r) = 2re^{-(\pi/4)r^4}$$

Using the capture model developed earlier, $P_C(k)$ is obtained by taking the average value of $P_C(k|a_0, \mathbf{G})$, which is defined as the probability of capturing a packet given a_0 and $\mathbf{G}_k = [g_1, \dots, g_k]$, where a_0 is the amplitude of bits in the test packet and the $\{g_i\}$ define the interference from other packets. Given the probability density functions of a_0 and g_i , the average probability of capture for the test packet $P_C(k)$ can be obtained from the following k -dimensional integral:

$$P_C(k) = \int_0^\infty da_0 \int_{-\infty}^{+\infty} dg_1 \cdots \int_{-\infty}^{+\infty} dg_k f_{A_0}(a_0) f_{G_1}(g_1) \cdots f_{G_k}(g_k) P_C(k|a_0, \mathbf{G}_k) \quad (11.5.6)$$

Let us assume that each packet of L bits is protected by a BCH block code with block length L capable of correcting up to t bit errors per block. (For an uncoded system, $t = 0$). Then $P_C(k|a_0, \mathbf{G}_k)$ is given by

$$P_C(k|a_0, \mathbf{G}_k) = \sum_{i=0}^L \binom{L}{i} (1 - \bar{P}_b)^{L-i} \bar{P}_b^i$$

where $\bar{P}_b$ is given by

$$\bar{P}_b = \frac{1}{2^k} \sum_{\alpha_k = \pm 1} \frac{1}{2} \operatorname{erfc} \left(\frac{\alpha_0 + \sum_{i=1}^k \alpha_i g_i}{\sqrt{N_0}} \right)$$

Equation (11.5.6) provides an exact solution for calculation of the probability of capture for a given modulation and coding technique. The probability distribution function of the test bit amplitude, $f_{A_0}(a_0)$, and that of the interfering bit amplitudes, $f_{G_i}(g_i)$, are functions of the distance–power gradient and the distribution of terminals. These distribution functions for ring- and bell-distributed terminals are available in [Zha92]. Calculation of the probability of capture from Eq. (11.5.6) is rather tedious, and simpler upper and lower bounds are introduced in [Zha92]. Figure 11.37 shows upper and lower bounds on the probability of capture for slotted ALOHA, together with an exact calculation and the results of Monte Carlo computer simulations, all plotted versus the number k of interference packets [Zha92]. The exact calculation and simulations were done for binary phase shift keying (BPSK) modulation, SNR = 20 dB, packet size $L = 16$ bits, and the bell-shaped terminal distribution given earlier. The SNR is defined here for a terminal located at the median distance from the central station. The simulation results show close agreement with the exact calculation.

To determine the throughput for an average traffic intensity of G , we use Eq. (11.5.3) with $P_C(k)$ and $P(k)$ calculated from Eq. (11.5.6) and Eq. (11.5.1), respectively. For calculation of the upper and lower bounds we use the appropriate equations for the $P_C(k)$ provided in [Zha92] rather than Eq. (11.5.6).

Figure 11.38 shows throughput versus traffic intensity for a slotted ALOHA system operating under the same conditions as in Figure 11.37: BPSK modulation at SNR = 20 dB, Rayleigh fading, packet length $L = 16$ bits, and the bell-shaped distribution of terminals. The figure gives five curves, consisting of two sets of bounds and an exact calculation. The “absolute” upper and lower bounds are the cases of no capture and perfect capture, as given by Eqs. (11.5.4) and (11.5.5), respectively. The upper bound

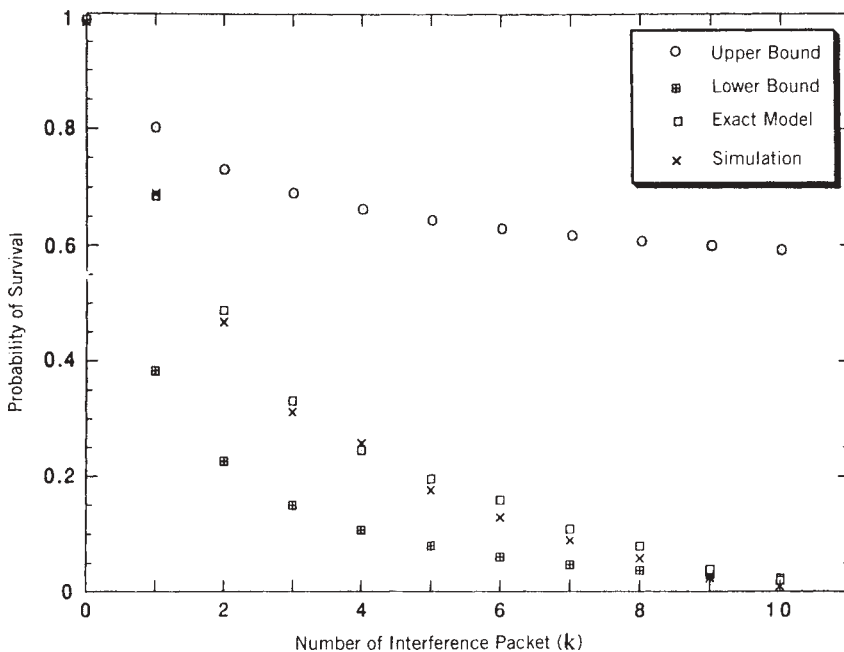


FIGURE 11.37 Probability of capture using calculation and Monte Carlo simulation. $L = 16$ bits, SNR = 20 dB, the modulation is BPSK, and the distribution is bell-shaped.

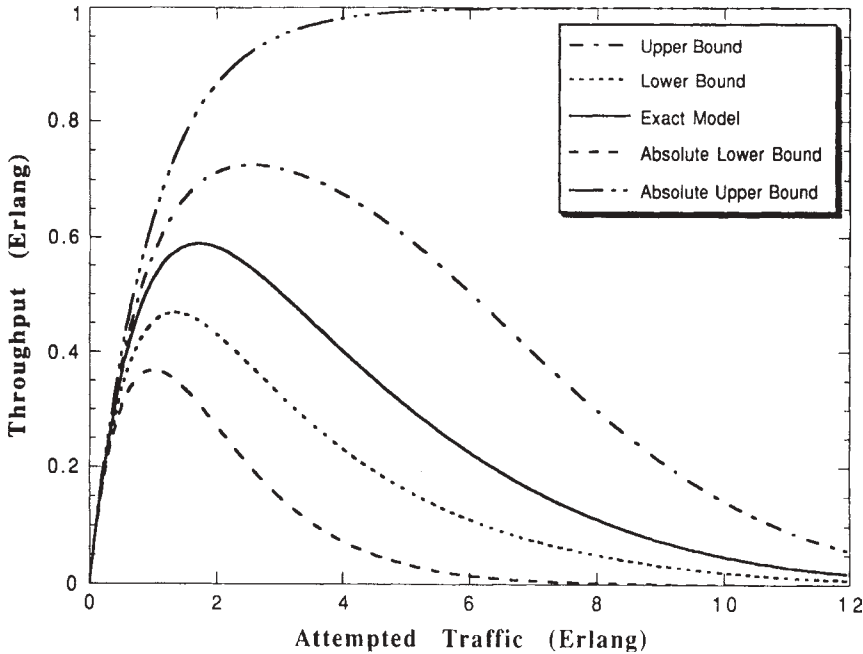


FIGURE 11.38 Throughput versus attempted traffic for a slotted ALOHA system operating at SNR = 20 dB in slow Rayleigh fading: Absolute upper and lower bounds, and upper and lower bounds and an exact calculation derived for packet length $L = 16$ bits and a bell-shaped distribution of the terminals. The modulation is BPSK.

on throughput is calculated by assuming that the interfering bit patterns have the same pattern as that of the test packet. The lower bound is calculated by assuming that signal fading affects each bit in a packet independent of any other bit in the packet. This is essentially equivalent to assuming extremely fast fading and provides a lower bound on the throughput for slotted ALOHA with capture, because the probability of capture is lower than that of slow fading, where all the bits in a packet fade in unison. As we discussed earlier, the details for calculation of these bounds is given in [Zha92]. The curve for the exact calculation shows throughput approaching 60% in Rayleigh fading, a significant improvement over the 36% throughput achievable without capture.

It is also of interest to consider how slotted ALOHA throughput is affected by variations in other system parameters. Figure 11.39 shows exact calculations of the relationship between throughput S and the attempted traffic G for BPSK modulation and packet length $L = 16$, given different levels of SNR in Rayleigh fading. Results are shown for both bell- and ring-shaped distributions of terminals. With the ring distribution, user terminals are located at a fixed distance from a central station, as we stated earlier. The greater effect of variations in SNR is found for the ring distribution, which shows an approximate 18% drop in peak throughput when the SNR is reduced from 25 dB to 10 dB. Because of the wider variations in power received from terminals in the bell-shaped distribution, that distribution is less sensitive to SNR changes. The performance with the bell-shaped distribution, where no power control of any sort is assumed, is better than with the ring distribution, with its assumed average-power

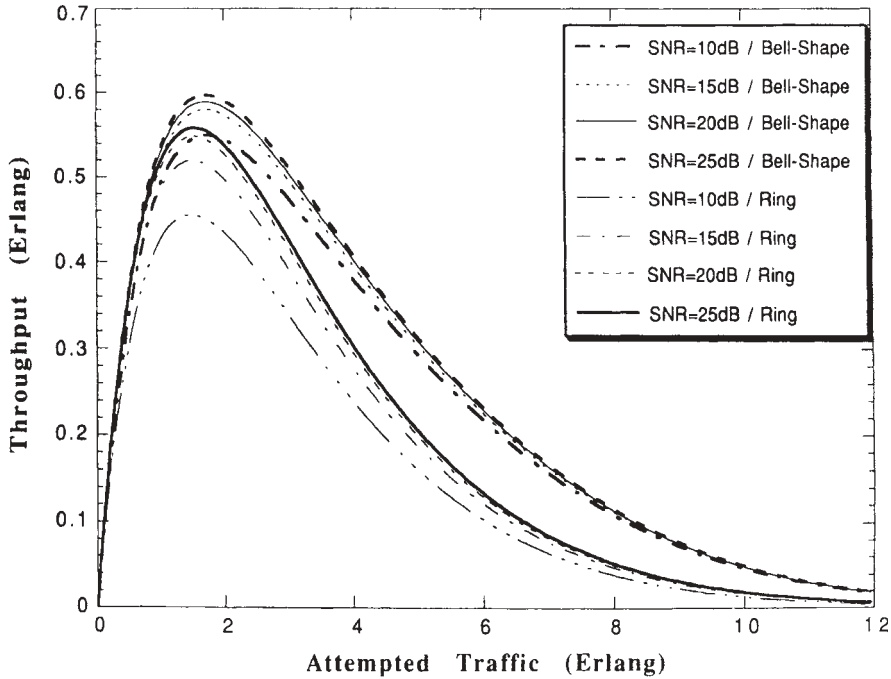


FIGURE 11.39 Exact calculations of throughput versus attempted traffic for slotted ALOHA in Rayleigh fading at various levels of SNR, assuming two distributions of terminals. For both bell- and ring-shaped distributions, the modulation is BPSK and the packet length is $L = 16$ bits.

control. The technique of instantaneous power control provides the same received power from each terminal, and the case of zero capture gives the closest approximation to this case. As one might expect, a greater degree of control on received power leads to lower probability of capture and, in turn, lower achievable throughput.

Figure 11.40 shows the relationship between S and G for slotted ALOHA with different packet lengths and with both the bell- and ring-shaped distributions of terminals. The greater variation of throughput with packet length is observed for the ring-shaped distribution, where the throughput with 64-bit packets is about 8% lower than that with 16-bit packets.

Figure 11.41 shows the effect of different choices of modulation method on throughput versus traffic intensity. Again, slotted ALOHA in Rayleigh fading is assumed, the packet length is $L = 16$, and $\text{SNR} = 20$ dB. The modulation schemes are PSK, FSK, and NCFSK, and bell- and ring-shaped distributions are both considered. The greater effect of modulation choice on throughput is observed for the ring distribution, where the maximum throughput with PSK modulation is about 5% higher than the maximum throughput with NCFSK.

Figure 11.42 shows the effect on throughput resulting from the use of error-correction coding with slotted ALOHA in slow fading. Results are shown for two cases, in which a 64-bit information packet is encoded with a 71- or 127-bit BCH block code. Transmission of uncoded 64-bit packets is included for comparison. The modulation is PSK, the channel SNR is 20 dB, and a bell-shaped distribution of terminals is assumed. The (71,64) code can correct one error in any code block, while the (127,64) code can correct up to 10

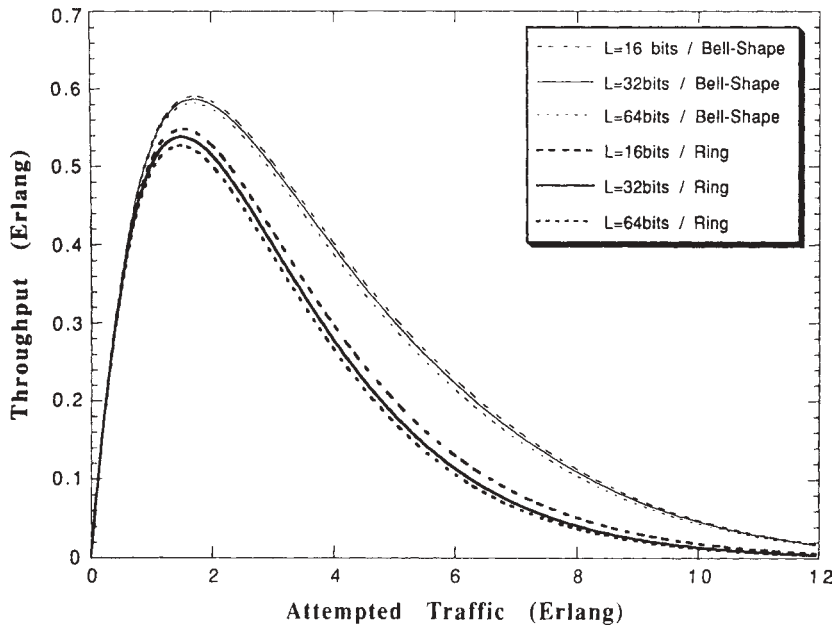


FIGURE 11.40 Relationship between S and G for various packet lengths and with both bell- and ring-shaped distributions of terminals. The greater variation of throughput with packet length is observed for the ring-shaped distribution, where the throughput with 64-bit packets is about 8% lower than with 16-bit packets. SNR = 20 dB, and the modulation is BPSK.

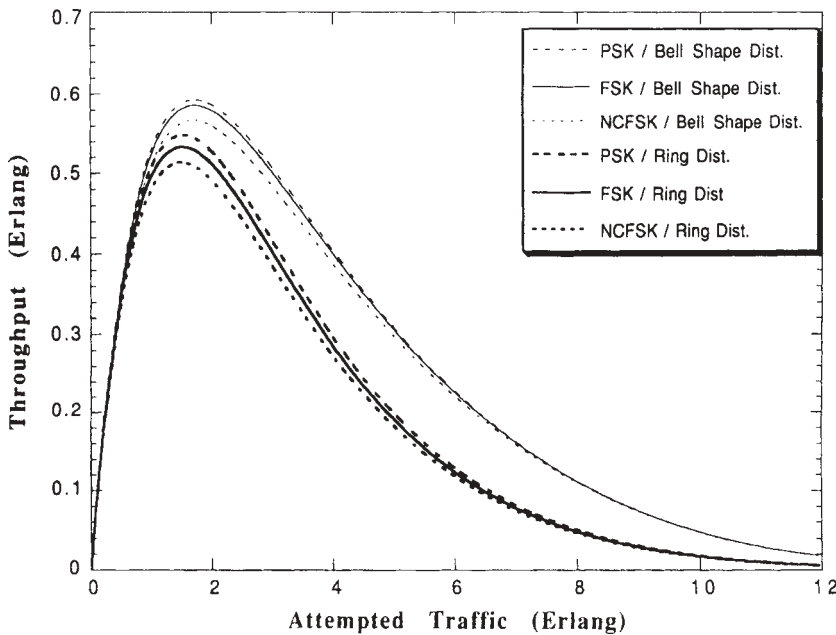


FIGURE 11.41 Effect of various choices of modulation method on throughput versus traffic offered. Slotted ALOHA in Rayleigh fading is assumed, the packet length is $L = 16$, and SNR = 20 dB. The modulation schemes are PSK, FSK, and NCFSK, and both bell- and ring-shaped distributions are shown.

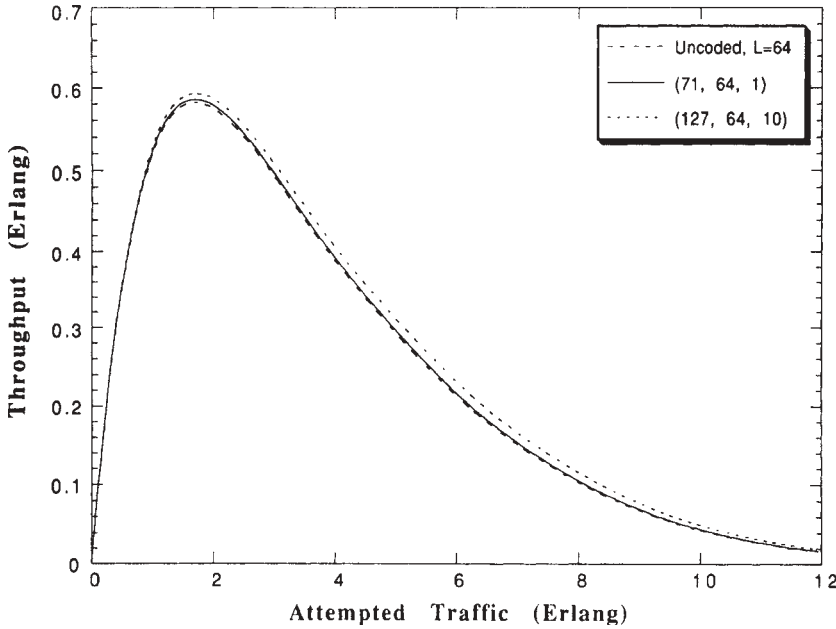


FIGURE 11.42 Effects on throughput resulting from the use of error-correction coding. Results are shown for two cases, in which a 64-bit information packet is encoded with a 71- or 127-bit BCH block code. Transmission of uncoded 64-bit packets is included for comparison. The modulation is BPSK, the channel SNR is 20 dB, and a bell-shaped distribution of terminals is assumed.

errors per block. It can be seen in the figure that the use of error-correction coding has very little effect on the achievable throughput in slow Rayleigh fading. The reason for this is that when the fade durations are long relative to the code block length, errors within a code block are highly correlated and thus bursty, rendering the error-correction coding ineffective [Lev76, Eav77]. Figure 11.43 is similar to Fig. 11.42, but now with 16-bit packets coded into length-31 or length-63 BCH code blocks. Again, error-correction coding provides a minimum benefit in slow Rayleigh fading. We can conclude from Figs. 11.38 to 11.43 that the major influence on throughput for slotted ALOHA in slow Rayleigh fading comes from the spatial distribution of user terminals in the network or, equivalently, from the choice of a power-control scheme.

Throughput of a CSMA Channel and the Capture Model. With $P_C(k)$ the probability that the test packet survives with k interfering packets, the throughput of a nonpersistent CSMA system with perfect acknowledgments on channels with capture is given by Zdunek et al. [Zdu89] as

$$S = \frac{\sum_{k=0}^{\infty} (1/k!) P_C(k) e^{-aG} (aG)^k}{1 + 2a + e^{-aG}/G}$$

where G is the average traffic intensity during the duration of a packet and a is the ratio of the maximum transmission delay to the duration of a packet. This equation

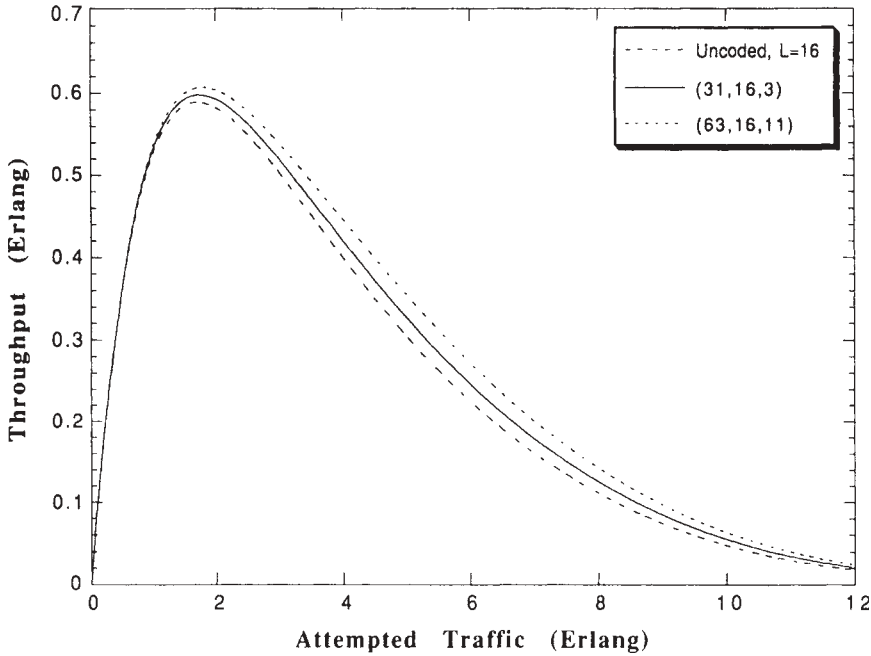


FIGURE 11.43 Effects on throughput with 16-bit packets coded into length-31 or length-63 BCH code blocks. The modulation is BPSK, the channel SNR is 20 dB, and a bell-shaped distribution of terminals is assumed.

is valid for versions of CSMA such as BTMA [Tob75] which deal with the *hidden terminal problem*. The capture probability $P_C(k)$ can be calculated as described earlier.

The average probability of capture $P_C(k)$ can be determined by Monte Carlo integration of Eq. (11.5.6) using the probability density functions $f_{A_0}(a_0)$ and $f_{G_i}(g_i)$ calculated for a bell-shaped distribution function of the terminals. Figure 11.44 shows the relationship between the average throughput S and the traffic intensity G for BPSK modulation with SNR = 20 dB and packet lengths of 16, 64, and 640 bits for both non-persistent CSMA and slotted ALOHA. The curves show the effects of packet length on achievable throughput and assume that $a = 0.01$ for all three lengths. Also shown for comparison are the curves for conventional nonpersistent CSMA and slotted ALOHA without capture. With capture, the maximum throughput of CSMA with packet length 16 bits is 0.88 erlang, which is 0.065 erlang more than the case without capture. The maximum throughput for slotted ALOHA with the same packet length is 0.591 erlang, which is 0.231 erlang higher than the case without capture. In slow-fading channels, if the terminal generating the test packet is in a “good” location, the interference from other packets is small and all the bits of the test packet survive the collision. In contrast, for a test packet originating from a terminal in a “bad” location, all the bits are subject to a high probability of error and the packet does not survive the collision. As a result, the system shows minimal sensitivity to the choice of packet length, which is consistent with our assumption of slow fading.

Figure 11.45 shows the effect of error-correction coding on the throughput of CSMA in slow fading for 64-bit information packets BCH encoded into 71- and 127-bit packets, respectively. The case of uncoded 64-bit packets is included in the figure for

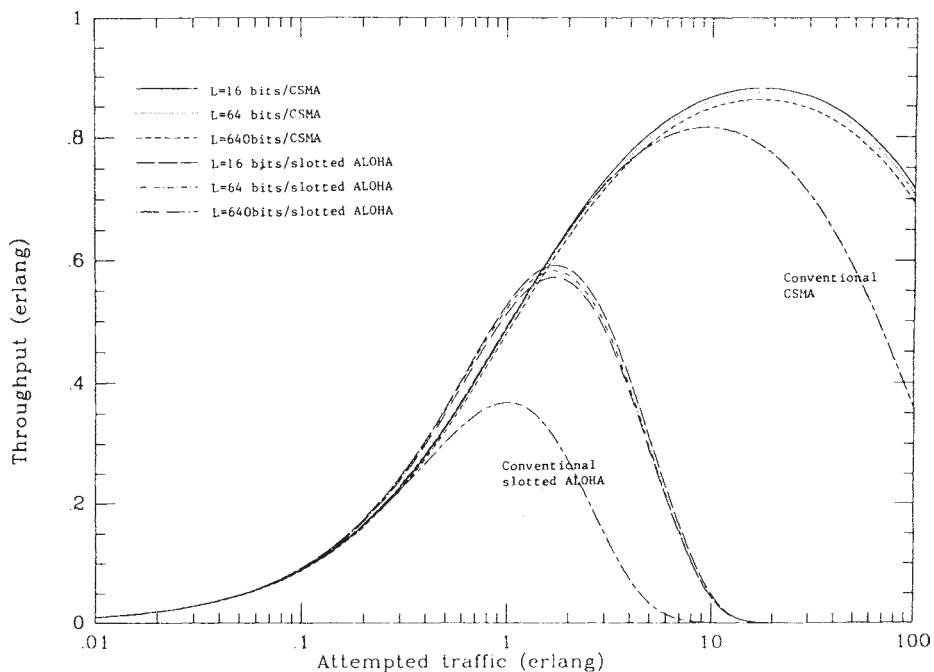


FIGURE 11.44 Effects of packet length on throughput for CSMA and slotted ALOHA with capture. The modulation is BPSK and SNR = 20 dB.

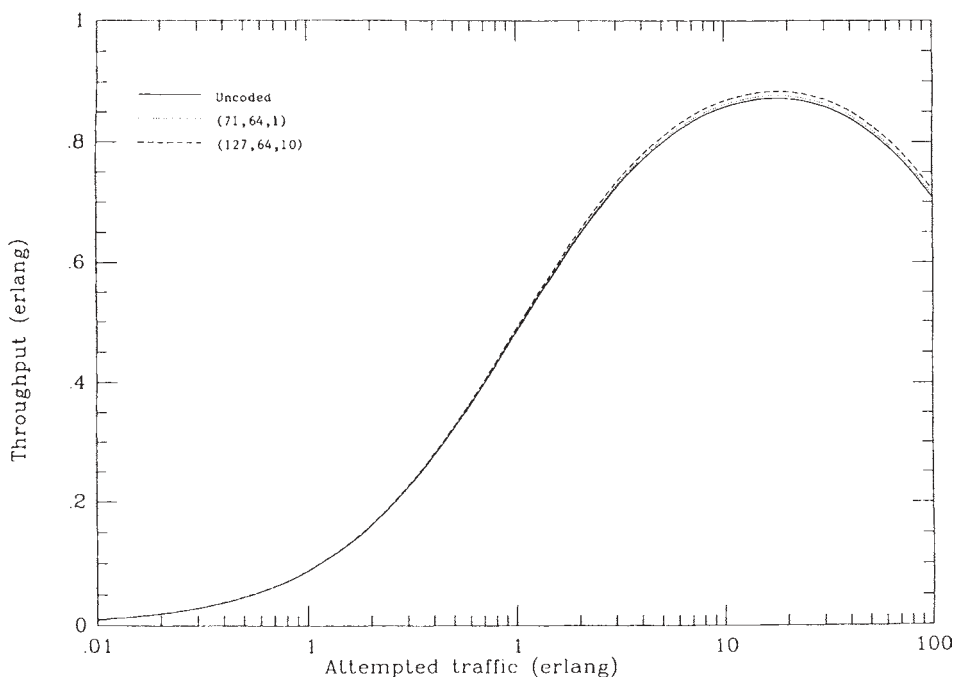


FIGURE 11.45 Effects of coding on throughput of CSMA. The modulation is BPSK and SNR = 20 dB.

comparison. The results show that error-correction coding is of negligible value in improving CSMA throughput in slow fading. The reason for this is essentially the same as given in the previous paragraph in explaining the lack of sensitivity of throughput to packet length. In simple terms, the presence of slow fading causes the bits in the test packet to be received as “all good” or “all bad” relative to interfering packets, and coding has little effect in such circumstances. Coding can be effective only when bit errors in a block are essentially independent, as occurs in additive Gaussian noise or perhaps in extremely rapid fading.

In summary, the performance results given in Figs. 11.44 and 11.45 have shown that the capture effect results in only about 0.06-erlang improvement in the maximum throughput of a CSMA packet radio network operating in slow fading. This improvement is considerably less than the improvement of approximately 0.2 erlang found for slotted ALOHA. The throughput has also been shown to have minimal sensitivity to the length of the packet and the complexity of the coding technique.

Delay Versus Throughput. The delay versus throughput characteristic for CSMA can be analyzed using an approximation suggested in [Kle75c]. This approximation assumes that acknowledgment of a received packet is always correct and that all channel delays other than those defined below are ignored. The expression for the average delay of the CSMA [Zha91] is

$$D = \left(\frac{G}{S} - 1 \right) (1 + 2\alpha + \delta) + G - \frac{H}{S} \delta + 1 + \alpha$$

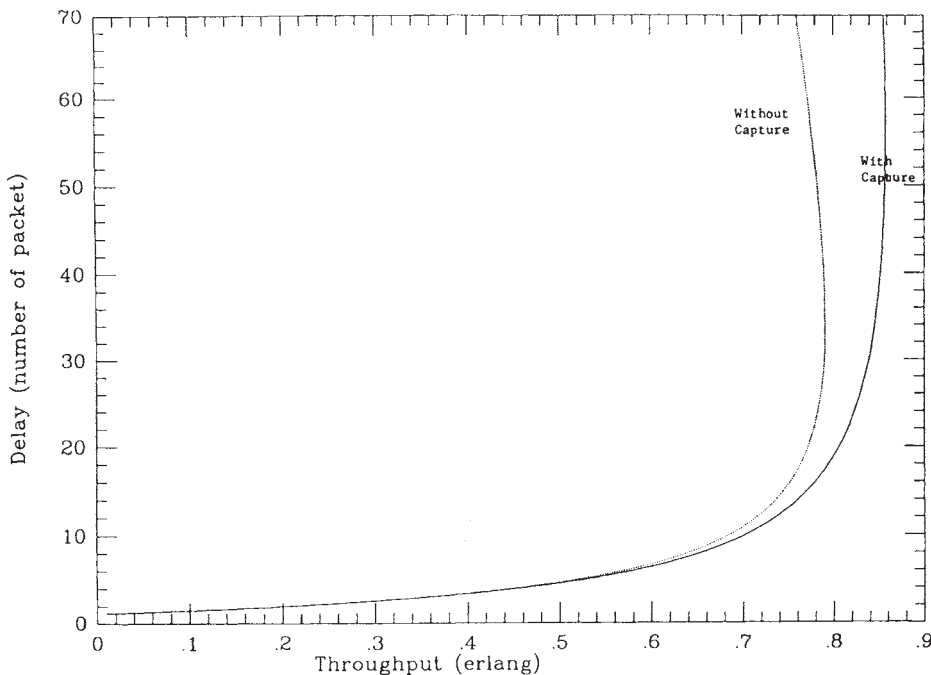


FIGURE 11.46 Delay versus throughput of CSMA. The modulation is BPSK and SNR = 20 dB.

where α is the normalized propagation delay; $H = (1 + \alpha G)/(1 + 2\alpha + e^{-\alpha G}/G)$, which is the number of actual transmissions; and δ is the average retransmission delay.

Figure 11.46 shows delay versus throughput for the CSMA system with and without capture for a 640-bit packet. The average retransmission delay is six packet lengths. As expected, the capture effect improves the throughput and reduces the delay at high traffic load.

QUESTIONS

- (a) Why is the use of error-correction coding more prevalent in voice-oriented networks than in data-oriented networks?
- (b) Why is error detection important in a packet data system?
- (c) What are the most popular access methods for voice-oriented cellular networks? Why do data-oriented WLANs use a different access method?
- (d) What are the practical differences between fully connected ad hoc networks and multihop ad hoc networks?
- (e) How do service providers control the level of interference in wide-area cellular networks operating in licensed bands? How can we control interference in unlicensed bands?
- (f) How can we define the capacity of a multirate wireless data network?
- (g) Why do we measure the performance of a voice-oriented network with the probability of blockage and the performance of a data-oriented network with throughput-delay characteristics?
- (h) What is the main obstacle to utilizing CSMA/CD in radio channels?
- (i) Explain the difference between the reservation ALOHA and TDMA access methods.
- (j) Compare the throughput of reservation ALOHA with that of pure ALOHA. Explain the difference.
- (k) Name an advantage for integration of voice and data on wirelines that does not hold for wireless media.
- (l) Explain how GPRS and EDGE integrates data service with GSM cellular voice service.
- (m) Why do real-time packet voice systems not use acknowledgment packets?
- (n) Why is CDMA not used in wireless LANs?
- (o) Why does the use of sectorized antennas at the base station increase the capacity of a CDMA digital cellular network? What is the effect of antenna sectorization on the performance of a TDMA cellular system?
- (p) Name two methods of overlaying data service onto cellular telephone networks. Give an example for each method.

- (q) Explain the purpose of having a delay-compensation buffer in voice-over-packet data network applications. How does the length of the buffer relate to the probability of packet loss?

PROBLEMS

1. Figure 11.2 shows a typical cellular frequency reuse scheme with hexagonal cells. In this topology the frequency reuse factor is $K = 7$ (seven frequency sets, A to G, are assigned to the cells).
 - (a) Assign the frequency sets for hexagonal cells with frequency reuse factors of $K = 3, 4$, and 12 .
 - (b) If the minimum distance between the centers of two cells using the same frequency sets is represented by D and the maximum distance from the center of a cell to any of its corners is represented by R , give the D/R ratios for $K = 3, 4, 7$, and 12 .
 - (c) Give an equation for calculation of the signal-to-interference ratio $\gamma_{C/I}$ in decibels versus D/R for an open urban area with a distance power gradient of 4. Assume that we have omnidirectional antennas.
2. We provide a wireless public phone with five lines to a ferry crossing between Helsinki and Stockholm carrying 120 passengers, where on the average each passenger makes a 3-minute telephone call every 2 hours.
 - (a) What is the probability that a passenger approaches the telephones and that none of the five lines are available.
 - (b) What is the average delay for a passenger to get access to the telephone?
 - (c) What is the probability of having a passenger waiting more than 3 minutes for access to a telephone?
 - (d) What would be the average delay if the ferry had 240 passengers?
3. We want to use a GSM system with sectorized antennas ($K = 3$) to replace an existing AMPS system ($K = 7$) with the same cell sites. In the existing AMPS system the service provider supports 395 duplex voice channels.
 - (a) Determine the number of voice channels per cell for the AMPS system.
 - (b) Determine the number of voice channels per cell for the GSM system.
 - (c) Repeat part (b) assuming a W-CDMA system with the bandwidth of 12.5 MHz for each direction. Assume an SNR requirement of 4 (6 dB) and include the effects of antenna sectorization (2.75), voice activity (2), and extra CDMA interference (1.67).
4. (a) Assume that we want to support a 19.2-kb/s data service with a maximum required error rate of 10^{-3} over a wideband CDMA (WCDMA) system. What minimum chip rate and bandwidth are needed to support 100 simultaneous users with one carrier of this WCDMA system? Assume an antenna sectorization gain of 3, a voice activity factor of 60%, and an extra CDMA interference of 1.67.
 - (b) What would be the bandwidth requirement in part (a) if the number of users were changed twofold?

- (c) What would be the bandwidth requirement in part (a) if the data-rate requirement were increased to 192 kb/s?
 - (d) What would be the bandwidth requirement in part (a) if the error-rate requirement were changed to 10^{-4} ?
5. Using the throughput formulas for the ALOHA and slotted-ALOHA protocols, show that the maximum values of the throughput for these protocols are $1/2e$ and $1/e$, respectively.
 6. A number of terminals use the pure ALOHA protocol to transmit to a central control station over a shared 100-Mb/s channel. Each terminal transmits a 1000-bit packet, on average, every minute.
 - (a) What is the maximum number of terminals that the channel will support?
 - (b) Repeat part (a) for the slotted ALOHA protocol.
 - (c) Repeat part (a) for both protocols for packet lengths of 2000 and 4000 bits.
 - (d) Describe how the results change if the transmission rate of the channel increases to 200 Mb/s.
 7. A small slotted-ALOHA network has k user stations, each of which transmits with probability $1/k$ (original transmissions and retransmissions combined) during any slot. Calculate the channel throughput as a function of k . Evaluate the throughput for $k = 2, 4, 8$, and 16 and for the limiting case of $k \rightarrow \infty$.
 8. A large network (assume an infinite population of user terminals) operates with the slotted-ALOHA protocol. Each station waits an average of W slots before retransmission after a collision. Calculate and plot the delay versus throughput curves for this network for $W = 2, 4, 8$, and 16.
 9. (a) Sketch the throughput versus offered traffic G for a mobile data network using slotted 1-persistent CSMA protocol. The packets are 40 ms long and the radius of coverage of each BS is 10 km. Assume the radio propagation speed is 300,000 km/s and use the worst delay for calculation of parameter a .
 - (b) Repeat part (a) for the slotted-ALOHA protocol.
 - (c) Repeat part (a) for the nonpersistent-CSMA protocol.
 - (d) Repeat part (a) for a WLAN with access point coverage of 100 m.
 - (e) Repeat part (a) for a satellite link with a distance of 20,000 km from Earth.
 10. In a slotted-CDMA packet data network we assign an orthogonal code of length 16 to each user, and the modulation technique is DPSK.
 - (a) Give the received signal-to-noise power ratio as a function of M , the number of packets arriving simultaneously. Assume that perfect power control is applied to the system and that the received signal-to-noise power ratio in the absence of interference is 10 dB.
 - (b) Determine the SNR per bit for M simultaneously arriving packets and give the probability of error per bit.
 - (c) Give an expression for calculation of the probability of capture $P_C(M)$ of a packet with a length of L bits as a function of M . Give the numerical values of $P_C(1)$, $P_C(2)$, and $P_C(10)$ for $L = 10$.
 - (d) Give an expression for the calculation of throughput as a function of offered traffic G if the packets arrive with a Poisson distribution.

11. (a) Use the Erlang B equation to calculate the number of slots required for a TDMA system with a probability of blockage of 2% and $\rho = \lambda/\mu = 0.5$.
 (b) Repeat part (a) for a probability of blockage of 20% and 0.2%.
 (c) Repeat part (a) for $\rho = 0.9$.
12. In the U.S. AMPS system, the total available band for reverse channel (uplink) is 12.5 MHz, and the frequency reuse factor of $K = 7$ is used with three sector antennas at the base stations.
 (a) What are the number of channels per sector? What is the number of users (Erlang capacity) per sector for a 2% probability of blockage?
 (b) Repeat part (a) for a digital cellular system in which three TDMA users operate simultaneously in each analog channel.
13. If the maximum number of calls per hour in a sector of a cell is 5000 and the average calling time is 1.76 min, how many radio channels are needed if the acceptable blocking probability is 2%?
14. A TDMA/TDD PCS system accommodates 50 simultaneous users (100 time slots) with a speech coding rate of 32 kb/s. If we use the idle times to integrate the voice and data for the same number of voice users, what is the data throughput? Use the approximate equations in Section 11.4.7 and call parameters of Example 11.7.
15. In a slotted-ALOHA wireless network the received signal amplitude from each terminal forms a Rayleigh-distributed random variable with an average received power of P . We examine the throughput of the network using a simple power-based capture model. In the model, if there is a collision of $n + 1$ packets, the target packet will be captured if its instantaneous power, p_s , is at least z times more than the total power of the n interfering packets, p_n . In other words, the target packet is destroyed if $\gamma = p_s/p_n < z$, where z is the capture threshold and γ is the S/I ratio at the collision.
 (a) Show that the probability density functions of the received power from the target user and the interference power from other terminals are given by

$$f_{P_s}(p_s) = \frac{1}{P} e^{-p_s/P}$$

and

$$f_{P_n}(p_n) = \frac{1}{P} \frac{(p_n/P)^{n-1}}{(n-1)!} e^{-p_n/P}$$

- (b) Show that the probability density function and probability distribution function of the S/I ratio are given by

$$f_{\Gamma}(\gamma) = n(\gamma + 1)^{-n-1}$$

and

$$F_{\Gamma}(\gamma) = 1 - \left(\frac{1}{\gamma + 1} \right)^n$$

- (c) Show that if the arrivals of the packets obey a Poisson distribution, the throughput of the system is given by

$$S = G \left[1 - \sum_{n=1}^{\infty} \frac{G^n}{n!} e^{-G} F_{\Gamma}(z) \right] = G e^{-Gz/(z+1)}$$

where G is the traffic offered.

- (d) Use MathCAD or MatLAB to sketch S versus G for capture parameter values of $z = 0, 3, 6, 20$, and ∞ (no-capture) dB.

PROJECT

Project 1. Performance of IEEE 802.11b/g WLANs

In Project 1 in Chapter 4 we found the relation between the data rate and coverage of 802.11b/g WLANs. Calculation of the real throughput of a WLAN involves calculation of overhead for packet formation, overhead for medium access, and the physical data rate that changes according to the RSS, which is a random variable. Several performance measurement software packages are available in the market. Some of these packages run a TCP/IP application over the channel and display the measured throughput. Similar to calculation of coverage one can use these tools to measure and model the real throughput of a WLAN in different indoor or outdoor environments. The same group of undergraduate students involved in Project 1 in Chapter 4 [Bha03] used the commercially available Chariot performance monitoring tool for performance analysis of IEEE 802.11b/g WLANs. In this project we use their measurement and modeling efforts to further analyze the performance behavior of the IEEE 802.11b/g technologies.

- (a) Figure P11.1 shows the overhead for packet formation and applications using TCP packets. Each TCP packet can have a length of up to 65,495 bytes, which should be fragmented to fit the maximum MAC packet of 2312 bytes. The TCP/IP header is 40 bytes, the 802.2 LLC/SNAP header is 8 bytes, and the 802.11 MAC and PLCP headers and synchronization preamble are 34 and 24 bytes, respectively. The TCP ACK is a TCP header with no application data and the MAC ACK is a 14 byte MPDU. Assuming SIFS and DIFS intervals of 10 μ sec and 50 μ sec, respectively, determine the application throughput of 802.11b for data rates of 11, 5.5, 2, and 1 Mb/s for data packets of length 100 and 1000 bytes.
- (b) In reality the throughput is a function of the channel characteristics and it fluctuates in time. Figure P11.2 shows a typical application throughput of an 802.11b terminal in a one-minute observation time. Due to the channel fading and other imperfections, this throughput varies in time as we measure it given a certain distance between the transmitter and the receiver. As the distance between the transmitter and the receiver increases and the RSS reduces, thus average throughput also reduces. The undergraduate group in WPI measured this throughput at different distances on the third floor of the AK Laboratories in the same locations

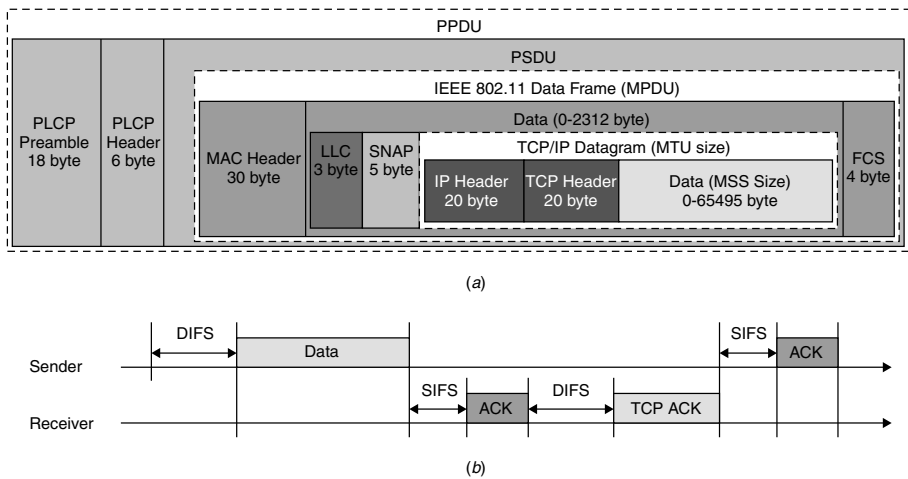


FIGURE P11.1 Packet transmission in the IEEE 802.11b. (a) Overheads for the formation of a packet. (b) Overheads for successful transmission of a TCP packet.

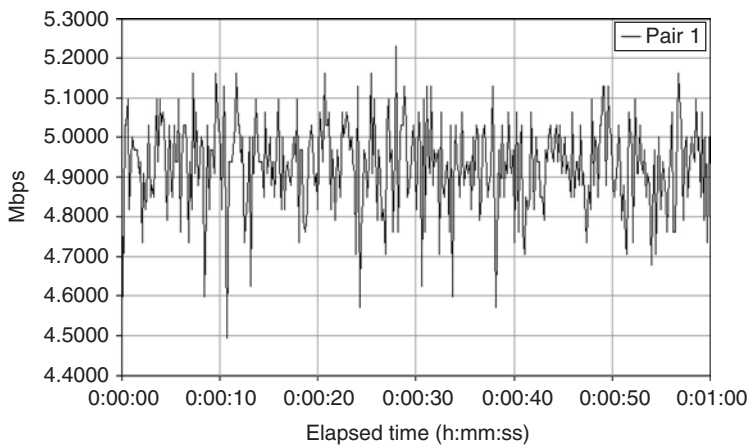


FIGURE P11.2 Throughput variations in one location for 802.11b.

used for modeling of the distance–power relation, as shown in Fig. P4.1. In their investigations they observed that the throughput of the IEEE 802.11b showed a linear decay with the increase in the distance while the throughput of IEEE 802.11g reduced with the logarithm of the distance. They fitted these values with two equations revealing the relation between the average throughput and the distance; it is shown in Table P11.1. Plot the throughput versus distance curves of IEEE 802.11b and g in the experiment performed by the undergraduate students and determine $R_{b/g}$, the distance for which the throughput of 802.11b or g calculated from the experimental equations approaches zero. Also, determine the distance at which average effective throughput of 802.11g falls below that of 802.11b.

TABLE P11.1 Relation Between Average Throughput and Distance for 802.11b/g

	Throughput Function $S(r)$ (Mb/s)
802.11b	$S_b(r) = -0.1669r + 5.3431$ For $0 < r < R_b$
802.11g	$S_g(r) = -22.344 \log_{10} r + 25.484$ For $1 < r < R_g$

- (c) Show that we can find the average spatial throughput of a user randomly walking about in the coverage area of an access point by

$$\overline{S_u} = \frac{2 \int_0^{R_{b/g}} r S(r) dr}{R_{b/g}^2} [\text{Mb/s}]$$

where $R_{b/g}$ is the distance for which the throughput of the 802.11b or g calculated from the experimental equations approaches zero.

- (d) Use the above equations and the experimental value of the $S(r)$ observed by the undergraduate students, given in Table P11.1, to calculate the average spatial throughput in the area of coverage for IEEE 802.11b and g access points.

TABLE P11.2 Predicted Coverage for 802.11b in Third Floor of AK

Data Rate (Mb/s)	Testbed Environment (m)	Signal Level (dBm)
11	48.70	−82
5.5	69.78	−87
2	93.06	−91
1	115.48	−94

TABLE P11.3 Predicted Coverage for 802.11g in Third Floor of AK

Data Rate (Mb/s)	Testbed Distance (m)	Signal Level (dBm)
54	23.71	−72
48	23.71	−72
36	25.48	−73
24	33.98	−77
18	42.17	−80
12	48.70	−82
9	56.23	−84
6	86.60	−90

- (e) In Project 1 of Chapter 4 we predicted the coverage of IEEE 802.11b/g in the third floor of AK building based on the results of measurements reported by the undergraduate group (see Tables P11.2 and P11.3). Use these results to calculate the average spatial transmission data rate in the area of coverage of an 802.11b/g AP. Compare the results of average spatial transmission rate in the coverage area with the average spatial throughput calculations given in part (c) and discuss the discrepancies.

12

ULTRAWIDEBAND COMMUNICATIONS

- 12.1 Introduction
- 12.2 UWB Channel Characteristics
 - 12.2.1 Path-Loss Modeling
 - 12.2.2 Modeling of Multipath Behavior
- 12.3 Impulse Radio and Time-Hopping Access
 - 12.3.1 Pulse Shape and Antenna
 - 12.3.2 Time-Hopping Spread-Spectrum Transmission
 - 12.3.3 Receiver for a Multiuser Environment
- 12.4 Direct-Sequence UWB
- 12.5 Multiband OFDM
 - Questions
 - Problems

12.1 INTRODUCTION

In the late 1990s and early 2000s, Ultrawideband (UWB) communications began to attract considerable attention as a method of transmission for short-range wireless ad hoc networking in both commercial and military applications. The first technology used for UWB communications was the impulse radio [Win00], which was based on direct baseband pulse transmission without using carrier modulation. The basic idea for pulse transmission is the same as channel measurement using direct pulse transmission, described in Section 5.2.1. A very narrow pulse is transmitted and after all multipath components arrived from different directions the next pulse is transmitted. In channel measurement we repeat the same pulse to observe the behavior of different paths. In data communication we modulate the transmitted pulse with information. The maximum bandwidth used for the measurement systems described in Chapter 5 is around 200 MHz, producing 5-ns pulses, but in impulse radio, wider bandwidths are used. One of the major advantages of impulse radio is that periodic transmission of narrow pulses can be implemented with a low power expenditure. Because of the low-power characteristic, as discussed in Chapter 14, pulse transmission techniques are also popular in the design of remote-control devices.

Pulse transmission has been used for military radar applications since the 1950s [Tay01] and the first patent for a communication application issued in 1973 [Ros73, Gha04]. The new wave of interest in UWB for military applications began with DARPA projects such as BodyLAN and the small unit operation situation awareness system (SUO/SAS). In these projects, low-power small-radio technology was needed to support broadband multirate wireless communications with accurate positioning in indoor and urban areas where traditional global positioning systems (GPSs) do not perform satisfactorily.

In 1997 the IEEE 802.15 wireless personal area networking (WPAN) standardization committee began to define specifications for BodyLAN and eventually adopted Bluetooth as its standard specification, 802.15.1. The same standardization group considered the UWB solution within the IEEE 802.15.3a subcommittee, which focuses on defining standard specifications for WPANs operating in the unlicensed bands at 3.1 to 10.6 GHz released by FCC in 2002. Several options have been evaluated for UWB communications, among which are the historical impulse radio technology, as well as the direct-sequence (DS-UWB) and multiband OFDM (MB-OFDM) systems, all of which we describe here. MB-OFDM and DS-UWB are being considered as the standard specifications for IEEE 802.15.3a.

Analog cellular system occupied a bandwidth of less than 30 kHz, and the total allocated bandwidth for each direction of communication was around 25 MHz. Second-generation systems generally occupy the same bands as first-generation systems, and additional tens of megahertz were allocated as PCS bands, which are used by network operators to support second-generation technologies. The successful GSM system occupies 200 kHz per carrier and second-generation cdmaOne occupies 1.25 MHz per carrier. Third-generation systems occupy less than 10 MHz of bandwidth per RF carrier, and the total available bandwidth is on the order of 100 MHz. The IEEE 802.11 WLANs occupy a maximum bandwidth of 26 MHz per carrier, and the largest bandwidths used in the 5-GHz domain are around several hundreds of megahertz. The UWB system specifications occupy a bandwidth on the order of gigahertz and the total available band is around several gigahertz. Figure 12.1 illustrates the relative bandwidth of cellular, PCS, WLAN, and UWB systems. Cellular and PCS systems have much higher power spectral density and occupy relatively smaller bandwidths. WLANs have larger allocated bandwidth and a much lower power spectral density than those of cellular systems. The UWB systems have the largest bandwidth and lowest power spectral density.

As we discussed in Chapter 7, multirate modems are very popular for broadband wireless data applications such as WLANs and mobile data applications. At the time of this writing, maximum data rates for third-generation mobile data systems are around 2 Mb/s, and the WLAN industry considers 100-Mb/s systems as the latest technology in its domain. Mobile data services provide a comprehensive coverage similar to cellular networks, and WLANs are intended to cover up to 100 m for local applications. All the candidate technologies for these systems use multirate modems, in which the highest supported data rate is supported in the vicinity of the base station or access point, and as we increase distance the data rate gradually drops. The IEEE 802.15.3a WPAN industry aims at hundreds of Mb/s in the vicinity of the access point for applications such as wireless USB connection to computers, and it is expected that the UWB technologies will support these applications. As the distance from the base station increases the data rate naturally decreases to support other applications, such as video transfer requiring

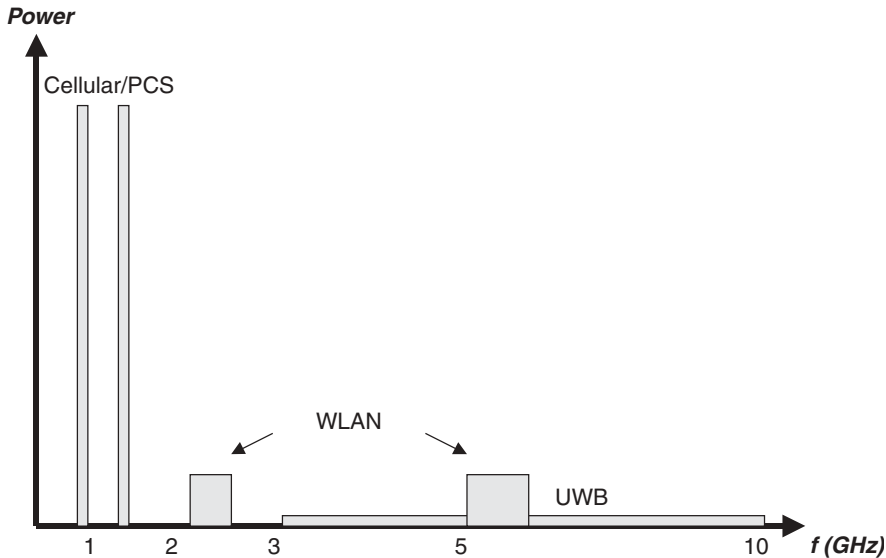


FIGURE 12.1 Relative power limits and the bandwidths available for cellular/PCS, WLAN, and UWB.

tens of Mb/s. The major considerations for implementation of UWB technology are low-power operation and low-cost implementation.

At this writing the only band available for UWB is the one announced by the FCC. Figure 12.2 shows the spectral mask mandated by the FCC for unlicensed UWB communications. The figure also identifies the frequency spectrum considered by the IEEE 802.15.3a proposal. Original petitions for the UWB band covered the entire spectrum and were intended to be used for impulse radio. However, after a number of controversial debates concerning interference with other low-power devices, in particular GPS, the FCC decided to specify this mask, which reduces the radiation in the spectrum between 0.96 and 3.1 GHz, so as to facilitate harmless coexistence with popular existing systems. As the UWB unlicensed bands emerged, other technologies, such as CDMA and OFDM, which do not use impulse transmission, appeared in these frequency bands. An analogous situation existed for the unlicensed ISM bands, which were originally released for use with spread-spectrum technology but today support operation of other technologies, such as OFDM.

Since the 1985 release of the ISM bands as the first unlicensed bands, we are gaining appreciation of the importance of spectrum regulation in the development of new technologies, and we are also learning how to utilize the frequency bands more wisely. To restrict the interference from devices transmitting more than 1 mW of power, the original ISM band regulations enforced spread-spectrum technology, which was foreseen to provide an effective interference-controlling mechanism. Later, in the early 1990s, a group called WINForum proposed an *etiquette* for unlicensed operation, which was applied to the unlicensed PCS bands [Ste94]. The interference issues were also addressed in U-NII band regulations and as we noted above, in the UWB regulations. The IEEE 802.15.2 initiative was focused specifically on concern about the interference between 802.11 and Bluetooth, as we discuss in Chapter 10. Today, some researchers are examining the question of how to share the licensed

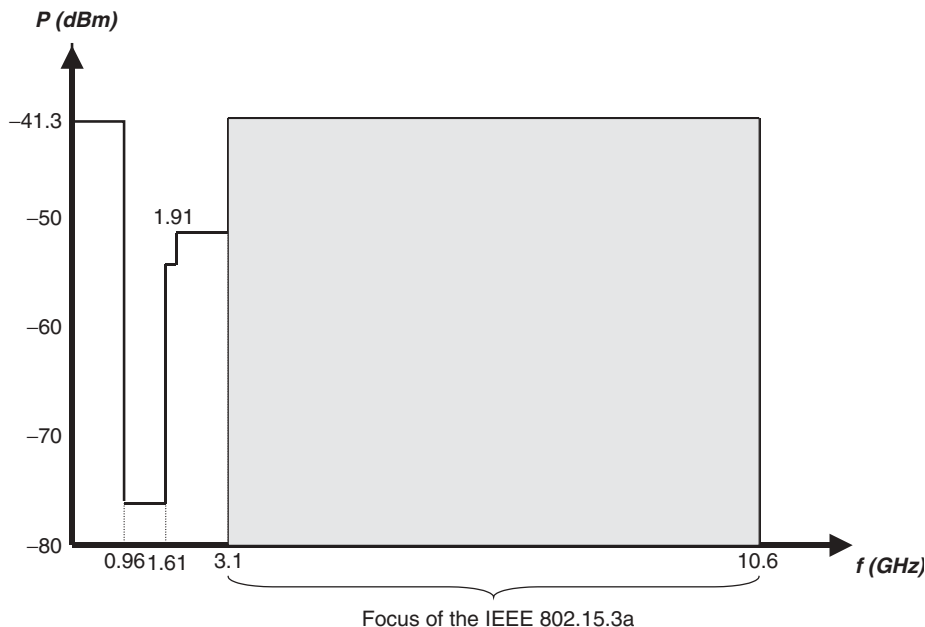


FIGURE 12.2 FCC spectral mask for unlicensed UWB communications.

bands among various service providers. If several service providers share their bands in a given geographical area, the spectrum available for all of them becomes wider, allowing achievement of higher data rates for data burst communications. In addition, some researchers are working on how we can develop a mechanism for fair-shared ownership of the licensed bands. Others are investigating interference-conscious design of the next generation of wireless networks, in which the main objective of the design is to minimize the overall radiation. This approach can not only reduce interference with other systems but also minimize concern over health issues. More details on UWB communications are available in [Gha04, Opp04] and references cited therein.

12.2 UWB CHANNEL CHARACTERISTICS

As we discussed in Section 12.1, the concept of UWB began with impulse transmission development, and today the name is used for any technology that occupies the UWB unlicensed bands. Therefore, for characterization of the UWB channels, we can apply the general wideband channel measurement and modeling techniques discussed in Chapters 5 and 6 to new sets of measurements collected in the UWB channels. Indeed, both time- and frequency-domain approaches discussed in Chapters 5 and 6 have been adopted for UWB channel measurement and modeling [Cas02, Gha03]. The characteristics of the UWB signal are examined, with system very similar to measurement systems described in Sections 5.2.1 and 5.3.1.

12.2.1 Path-Loss Modeling

In UWB propagation the signal bandwidth spans several gigahertz, and one may reasonably ask how, given such a wide bandwidth, the received signal strength should

be modeled. In earlier chapters we developed path-loss models based on narrowband measurements, and we utilized them in wideband applications as well. As shown in Eq. (3.2.1), received signal strength is an inverse function of the frequency, and therefore the received signal strength at the lower end of the UWB system spectrum can be much higher than that found at the higher end of the spectrum. At this point one might question whether we can still apply narrowband path-loss modeling techniques to UWB radio propagation. Assuming a perfect transmitter amplifier, a perfect antenna with the same gain at all frequencies of an UWB system, and free-space propagation, from Eq. (3.2.1) the received signal strength as a function of frequency is given by

$$P_r(f) = P_t G_t G_r \left(\frac{\lambda}{4\pi d} \right)^2 = \frac{P_t G_t G_r c^2}{(4\pi d)^2 f^2}$$

where P_t is the average transmitter power spectral density. Then the average transmitted power of the UWB system in the bandwidth W is $P_{t-\text{ave}} = P_t(f)W$, and the average power received around the center frequency f_c is given by

$$\begin{aligned} P_{r-\text{ave}}^{\text{UWB}} &= \int_{f_c-W/2}^{f_c+W/2} P_r(f) df = \frac{P_{t-\text{ave}} G_t G_r c^2}{W (4\pi d)^2} \left(\frac{1}{f_c - W/2} - \frac{1}{f_c + W/2} \right) \\ &= \frac{P_{t-\text{ave}} G_t G_r c^2}{(4\pi d)^2 f_c^2} \left[\frac{1}{1 - (W/2f_c)^2} \right] \end{aligned}$$

where the first term is the received signal power in a narrowband system with the same average transmitted signal power as in the UWB system. Defining

$$P_{r-\text{ave}}^{\text{NB}} = \frac{P_{t-\text{ave}} G_t G_r c^2}{(4\pi d)^2 f_c^2}$$

as the received signal power for the equivalent narrowband system, we have

$$P_{r-\text{ave}}^{\text{UWB}} = P_{r-\text{ave}}^{\text{NB}} \frac{1}{1 - (W/2f_c)^2} \quad (12.2.1)$$

which describes the relationship between the bandwidth and received signal power of a UWB system and its equivalent narrowband system. Considering the FCC regulations described in Figs. 12.1 and 12.2, the maximum allowed bandwidth of a UWB system with uniform spectral density can occupy the frequency range 3.1 to 10.6 GHz, for which $W = 7.5$ GHz and $f_c = 6.85$ GHz. Substituting these values in Eq. (12.1.1), we observe that the narrowband system differs from the UWB system by only 1.5 dB. From this discussion we can conclude that narrowband models for path loss describing the received power with one value rather than a function of frequency can be used to approximate the path loss for a UWB system [Sol01, Foe02]. This observation significantly reduces the complexity of empirical channel measurement and modeling techniques for UWB systems. Stated simply, we can apply to UWB systems all measurement and modeling techniques described in Chapters 4 to 6.

Figure 12.3 illustrates a scatter plot of the received signal power for a UWB system for different distances and the best one-gradient fit (similar to Fig. 4.3) and partitioned two-gradient fit (similar to Section 4.2.3). A frequency-domain measurement system, shown in Fig. 4.8, having UWB cone antennas is used for measurement of the channel frequency response in 3 to 6-GHz bands. The site of the measurements is the first floor

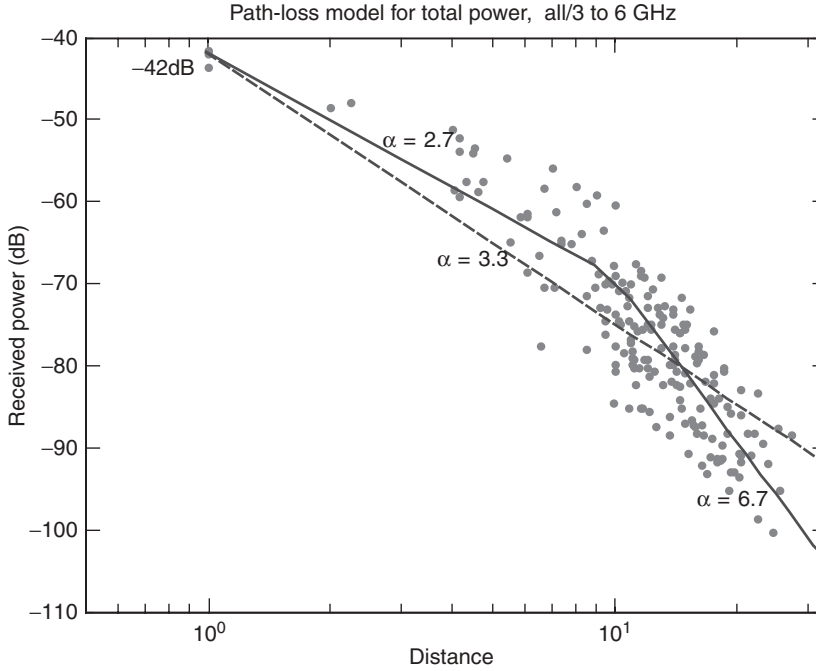


FIGURE 12.3 Scatter plot of the received power versus distance and the best first single-slope and two-segment partitioned path-loss models.

of the Atwater Kent Laboratories at Worcester Polytechnic Institute. The modeling used a theoretical path loss of 42 dB in the first meter for all cases in the figure. Similar to Eq. (4.2.1), the single-gradient path-loss model is given by

$$L_P = 42 + 33 \log_{10} d \quad (12.2.2)$$

The two-gradient partitioned model, described in Section 4.2.3, is described by

$$L_P = 42 + \begin{cases} 2.7 \log_{10} d, & d \leq 10 \text{ m} \\ 27 + 67 \log_{10} \frac{d}{10}, & d > 10 \text{ m} \end{cases} \quad (12.2.3)$$

Similar results are also reported in [Cas02]. More extensive measurements in residential and commercial buildings are reported in [Gha03a, Gha03b].

12.2.2 Modeling of Multipath Behavior

Figure 12.4 shows the frequency and time responses of a sample UWB channel with 3 GHz of bandwidth centered at 4.5 GHz, where the LOS path is obstructed. This measurement was taken using cone antennas, shown in Fig. 5.27b, and the frequency-domain system shown in Fig. 4.5. Comparable measurements for another pair of obstructed-LOS situations with 200 MHz of bandwidth are shown in Fig. 5.14. Since the bandwidth for Fig. 12.4 is 15 times larger than the bandwidth for Fig. 5.14, the

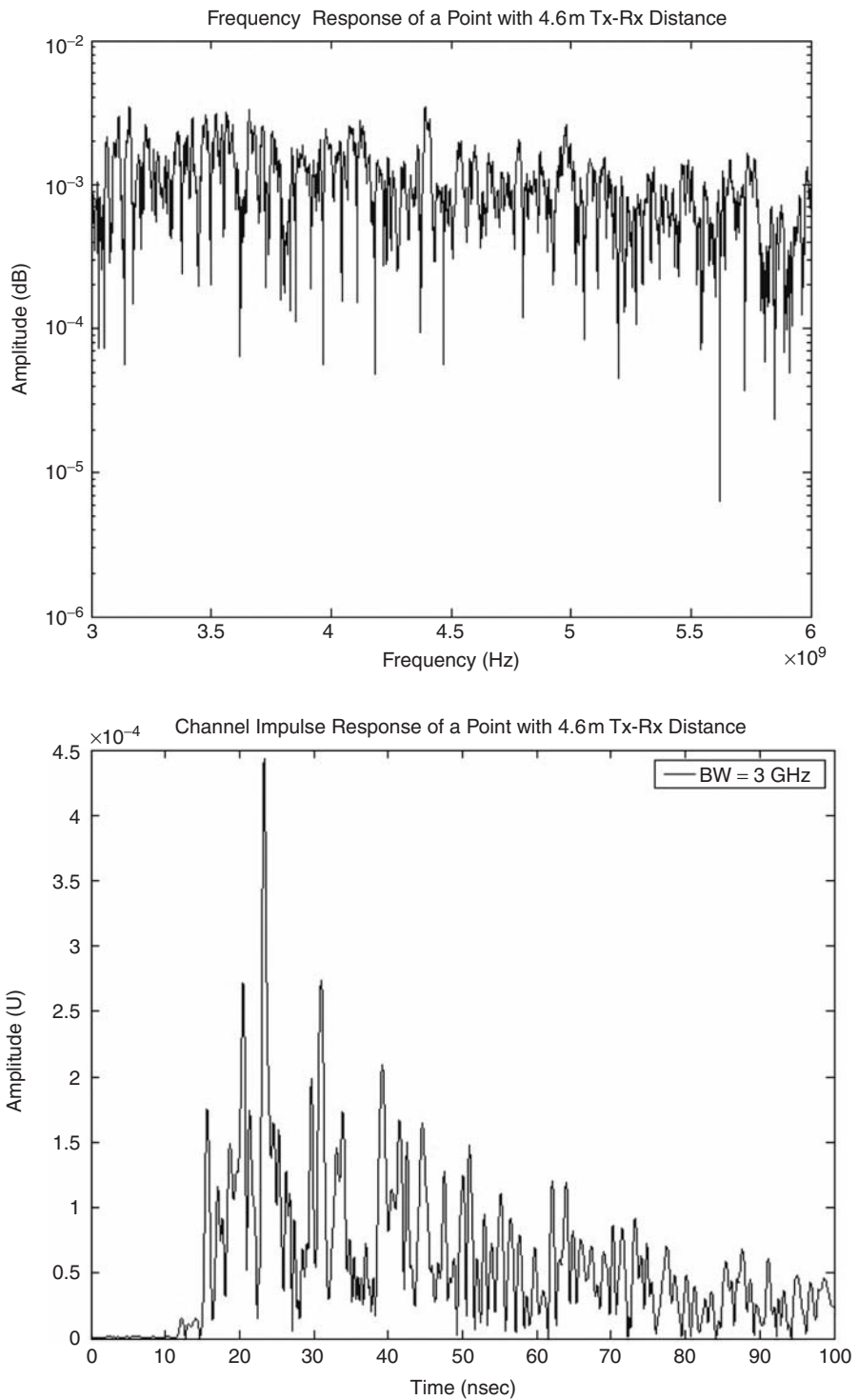


FIGURE 12.4 Frequency and time responses of an UWB system with 3 GHz of bandwidth in an obstructed LOS situation.

resolution of the UWB system is 15 times lower and many more paths are measured with the UWB system. An increase in the number of paths does not change the total power and rms delay spread of the channel but has an impact on the arrival rate and the statistics of fluctuations on the paths. In general, one can assume that for ultrabandwidth, all paths are isolated and the amplitude of each path is fixed. As the terminal is moved or people move about in the space between the transmitter and receiver, the amplitudes of the paths change due to shadow fading, which as we discussed in Section 4.2, exhibits a lognormal distribution. As the bandwidth becomes narrower, neighboring paths combine their amplitudes and phases, much as in Fig. 3.3. This results in rapid fluctuation of the composite signal amplitude as a result of the rapid phase changes in each path, producing fast multipath or Rayleigh fading. One of the main physical characteristics of UWB transmission is the stability of the received signal. This stability, which is due to its ultrawideband characteristics, isolates the paths, avoids the multipath Rayleigh fading, and exhibits only lognormal shadow fading. This observation leads to the conclusion that existing channel models, such as Saleh–Valenzuela’s model, described in Section 6.2.2, which assumes that individual paths are Rayleigh distributed, are not valid for UWB channel modeling and need to be modified. In the following paragraphs we describe the latest modification to the Saleh–Valenzuela model considered by the IEEE 802.15 WPAN community. This modified model is based on measurements performed at Intel [Foe02] and is also referred to as *Intel’s model*. During the standardization activities in the IEEE 802.15 committee, the IEEE 802.11 model, Saleh–Valenzuela’s model (Section 6.2.2), and the modified Poisson or Δ – K model (Section 6.2.3) were examined with the same set of data, and it was found that Intel’s model best fit the measured data reported in [Foe02].

Intel’s Model for UWB Multipath Propagation. This model begins with the framework of the double-cluster Saleh–Valenzuela channel impulse response given in Eq. (6.2.10), changes the statistical model for the amplitude β_{kl} and phase ϕ_{kl} , and adjusts the arrival rates of the clusters and paths within the clusters. The amplitude of the signal, β_{kl} , is assumed to be lognormal, and the phase, ϕ_{kl} , is modeled as a binary number taking only two values, 0 and 180° , with equal probability. To form the lognormal distribution, it is assumed that $20 \log_{10} \beta_{kl}$ has a normal distribution function with mean and variance of μ_{kl} and σ^2 , respectively. To show that the lognormal is the best fit for the amplitudes, a curve-fitting process similar to Fig. 4.13¹ was applied to the measured data. The mean of the lognormal distribution is given by

$$\mu_{kl} = \frac{10 \ln(\overline{\beta^2(0, 0)}) - 10T_l/\Gamma - 10\tau_{k,l}/\gamma}{\ln(10)} - \frac{\sigma^2 \ln(10)}{20}$$

in which $\overline{\beta^2(0, 0)}$ is the average received power in the first path, $\tau_{0l} = T_l$ is the arrival delay of the l th cluster, τ_{kl} is the delay of the k th path in the l th cluster, and Γ and γ are the cluster and ray decay factors, respectively. At this writing, the IEEE 802.15 WPAN community is using this framework to design models for different environments. Table 12.1 shows three-sample obstructed-LOS (OLOS) and two-sample LOS model parameters being studied by the committee.

¹Note that even for 200 MHz of bandwidth, as shown in Fig. 6.14, the amplitudes of the paths fit the lognormal distribution than the Rayleigh distribution.

TABLE 12.1 Five Sets of Parameters for Simulation of UWB Channel Behavior

	OLOS	OLOS	OLOS	LOS	LOS
Channel characteristics					
Mean excess delay, τ_m (ns)	17	22	27	3	4
RMS delay, τ_{rms} (ns)	15	20	25	5	9
Model parameters					
Λ (ns^{-1}) cluster arrival rate	1/11	1/14	1/15	1/22	1/60
λ (ns^{-1}) ray/path arrival rate	1/0.35	1/0.33	1/0.32	1/0.94	1/0.5
Γ cluster decay factor	16	22	30	7.6	16
γ ray decay factor	8.5	10	10	0.94	1.6
σ (dB) standard deviation of lognormal fading	4.8	4.8	4.8	4.8	4.8

Figure 12.5 shows two-sample simulated channel impulse responses using the IEEE 802.15 WPAN committee's recommended channel models. In this figure each path is represented by an impulse. To make it look more like an actual measurement, one needs to convolve this impulse response with the impulse response of a pulse (e.g., a raised-cosine pulse) having the bandwidth of the actual UWB measurement or communication system.

12.3 IMPULSE RADIO AND TIME-HOPPING ACCESS

As we noted earlier, the new wave of interest in UWB communication began with the idea of impulse radio transmission. With this technique, a pulse of very short duration (on the order of a few tenths of a nanosecond) and low power (a high duty cycle of several hundreds of nanoseconds) is used for information transmission. The spectrum of this pulse obviously occupies a very wide band (several gigahertz), and this is the reason it was given the name UWB. The spectral height of the UWB signal is very low because the low transmission power is spread over a very wide bandwidth. Therefore, similar to spread-spectrum signals, UWB signals can also be designed to coexist with other radio systems already in use. Since the bandwidths of UWB systems are much wider than those of spread-spectrum systems, they can readily be overlaid onto existing systems without causing interference to those systems. At the same time, since the bandwidth is ultrawide, the UWB system has a very low power spectral density and it can overlay many existing systems with different transmission requirements. As we discussed in Chapter 3, signal fading is caused by the overlap of signals received on different paths. Because of its ultrawide bandwidth, the UWB signal isolates (resolves) multipath components, resulting in stable received signal power with minimal multipath fading effects.

12.3.1 Pulse Shape and Antenna

Implementation of the impulse-radio UWB signal transmitter does not involve modulation, and if designed carefully, the transmitter can be much simpler than in traditional narrowband or wideband spread-spectrum systems. Designers of impulse radio UWB systems chose a transmission waveform that (1) has such a high bandwidth and signal processing gain that interference from existing systems into the impulse system is negligible; (2) its spectral height is comparable to background noise so that the

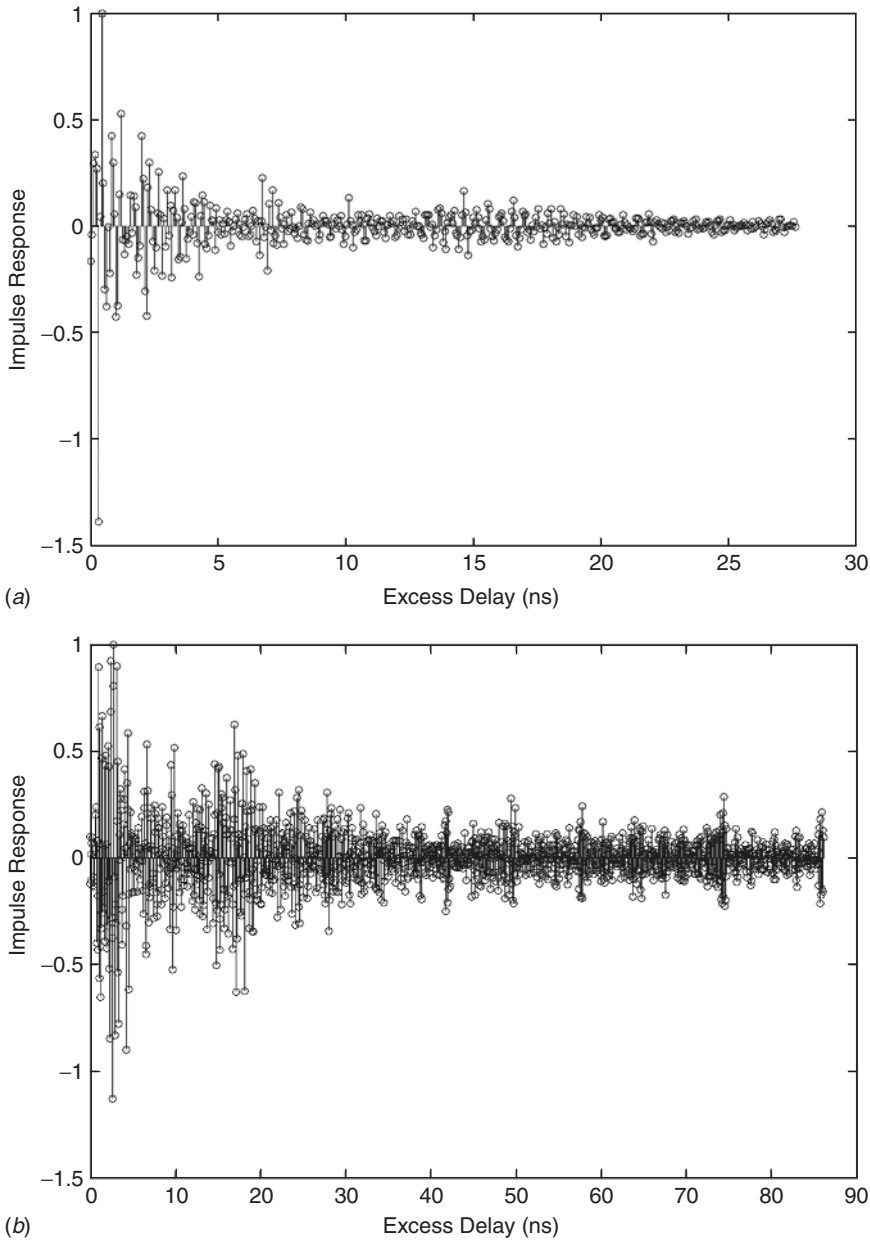


FIGURE 12.5 Sample channel simulations using IEEE 802.15 WPAN recommended model (a) LOS; (b) OLOS.

FCC allows it to coexist with the systems already in place; (3) its implementation is relatively simple; and (4) the spectrum looks like a passband signal that has no dc component.

Example 12.1: UWB Pulse Shape in Impulse Radio Time Domain Corporation (TDC) was a leading company engaged in the design of impulse radio UWB

systems [TDCweb]. The pulse shape used in their systems is a monocycle, described mathematically by

$$v(t) = 6A\sqrt{\frac{e\pi}{3}}\frac{t}{\tau}e^{-6\pi(t/\tau)^2} \quad (12.3.1)$$

where A represents the peak amplitude of the pulse, τ is a constant determining the width of the pulse, and t is the time variable. The spectrum of the monocycle pulse in the frequency domain is given by

$$V(f) = -j\frac{2f}{3f_c^2}\sqrt{\frac{e\pi}{2}}e^{\pi/6(f/f_c)^2} \quad (12.3.2)$$

where $f_c = 1/\tau$ is the center frequency of the pulse. Figure 12.6a shows a typical graph of the pulse for $\tau = 0.5$ ns associated with a center frequency of $f_c = 2$ GHz. The half power (3-dB) bandwidth of the pulse occupies about 2 GHz (Fig. 12.6b). The pulses are transmitted periodically and the information is encoded in the location of the pulses, implementing a pulse position modulation (PPM) system. Figure 12.6c and d illustrate pulse repetition and its effects on the spectrum of the signal. The typical pulse width used for this implementation is between 0.2 and 1.5 ns and the time interval between consecutive pulses ranges from 25 to 100 ns.

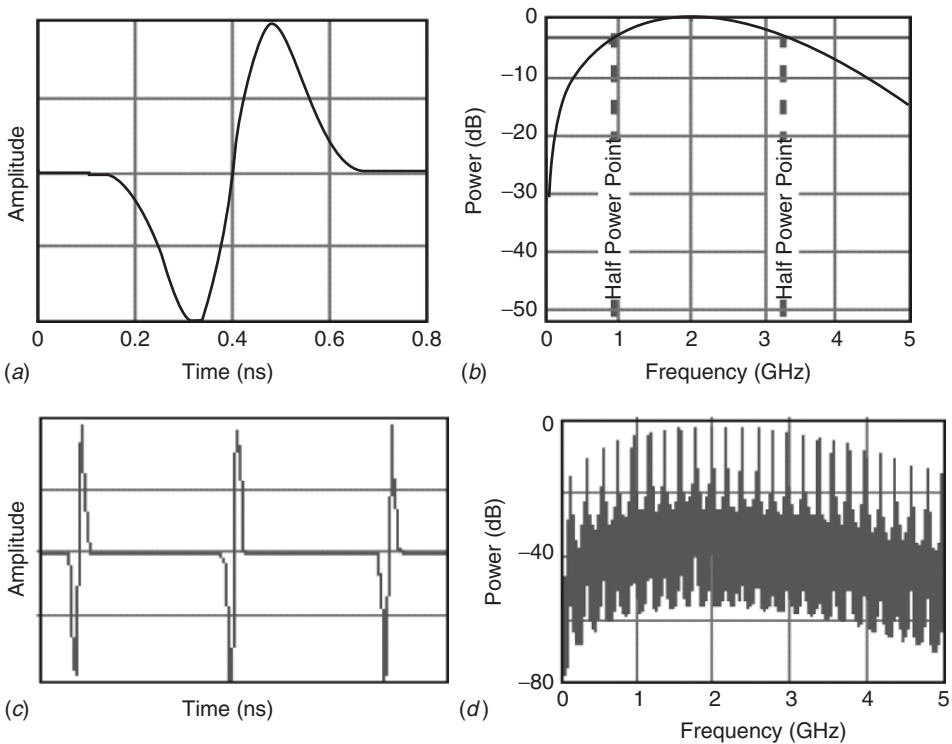


FIGURE 12.6 (a) Original UWB pulse used in pioneering impulse radio system designed by Time Domain Corporation; (b) the spectrum of the pulse; (c) repetition of the pulses; (d) effects of repetition on the spectrum of the signal [TDCweb].

One of the challenges in the design of UWB system is the choice of antennas. This is, in particular, more challenging for the original impulse-radio systems, which covered a spectrum ranging from very low frequencies up to a few gigahertz. The antennas designed for these bands need to have a flat response across a wide spectrum, and they also have to be compact. Since these conditions are difficult to maintain, the antennas used for impulse-radio transmission actually change the shape of the transmitted pulse.

Example 12.2: Pulse Shape and Antennas in Impulse Radio The pulse shape introduced in Example 12.6 is, in fact, a received pulse shape. Figure 12.7a represents the

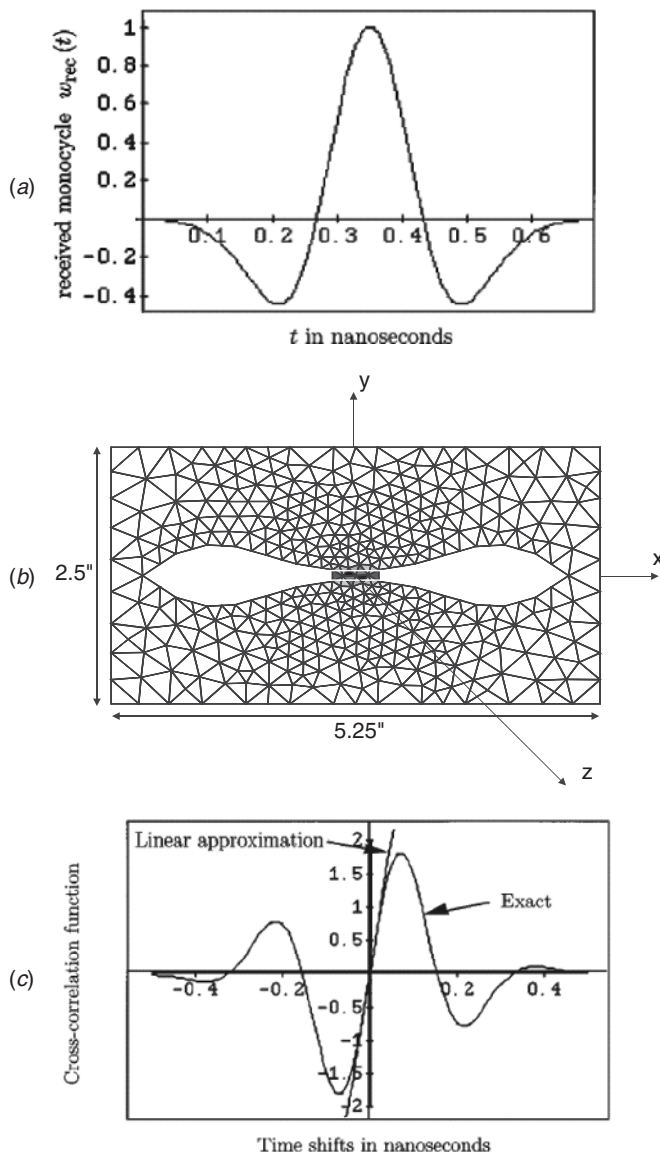


FIGURE 12.7 (a) Transmitted impulse; (b) bowtie UWB antenna; (c) received signal after antenna. (From [Win98] © IEEE.)

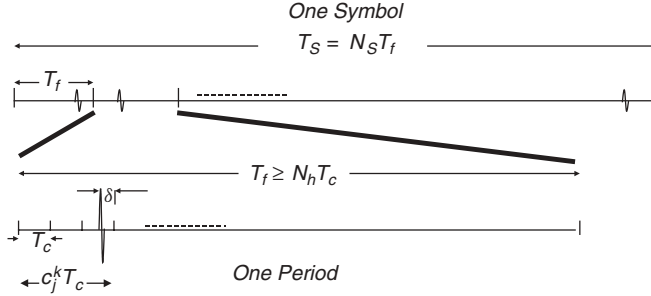


FIGURE 12.8 Transmission format of the time-hopping impulse radio system: relation among transmitted symbol, periods inside the symbol, chip time, and pulse position in a period.

actual transmitted pulse. When this pulse is applied to the bowtie antenna shown in Fig. 12.7b, the pulse propagates with distortion and the received signal will appear as shown in Fig. 12.7c. The antennas behave as filters, and even in free space, a differentiation of the pulse occurs as the wave radiates [Win98].

12.3.2 Time-Hopping Spread-Spectrum Transmission

The PPM modulation of the impulse radio is combined with time-hopping spread-spectrum (THSS) medium access control to support a multiuser environment. Figure 12.8 clarifies the operation of this transmission technique. The basic transmitted waveform, $w_{tr}(t)$, is repeated once every T_f second. Each period is divided into N_h shorter time slots, each referred to as the chip time, T_c , where $N_h T_c \leq T_f$. A random number generator of length N_s generates $\{c_j^k\}$, which is used to determine the delay $c_j^k T_c$ associated with the location of the pulse for that period. An additional shift of δ either in the positive or negative direction (according to the transmitted information bit) is added to place the pulse before or after the random delay for the period given by $c_j^k T_c$. To transmit an information bit, as shown in Fig. 12.8, N_s periods, each with a new random value of $\{c_j^k\}$, are sent. Thus, the transmitted signal from the k th user is represented by

$$x_{tr}^k(t^k) = \sum_{j=-\infty}^{\infty} w_{tr}(t^k - jT_f - c_j^k T_c - \delta d_{\lfloor j/N_s \rfloor}^k) \quad (12.3.3)$$

where $d_j = \{0, 1\}$ is the information digit that remains the same for N_s periods. Since we send 1 bit every $N_s T_f$ seconds, the data rate is $R_s = 1/N_s T_f$ bits per second.

Example 12.3: Data Rate of a Time-Hopping System For typical values of $T_f = 100$ ns and $N_s = 500$, the data rate is $R_s = 20$ kb/s, and for a maximum chip duration of $T_c = 0.2$ ns, we can take $\delta = 0.1$ ns.

Figure 12.9 illustrates the difference between narrowband, DSSS, and UWB-THSS systems. In the DSSS system the data stream is divided into chips that have smaller duration; therefore, with the same transmitted power and data rate we will have lower-power spectral density and a bandwidth expansion equal to the processing gain of the DSSS system. In UWB-THSS impulse-radio systems, transmitted pulses occupy a fraction of a chip, causing another expansion in the bandwidth and consequent reduction in the power spectral density.

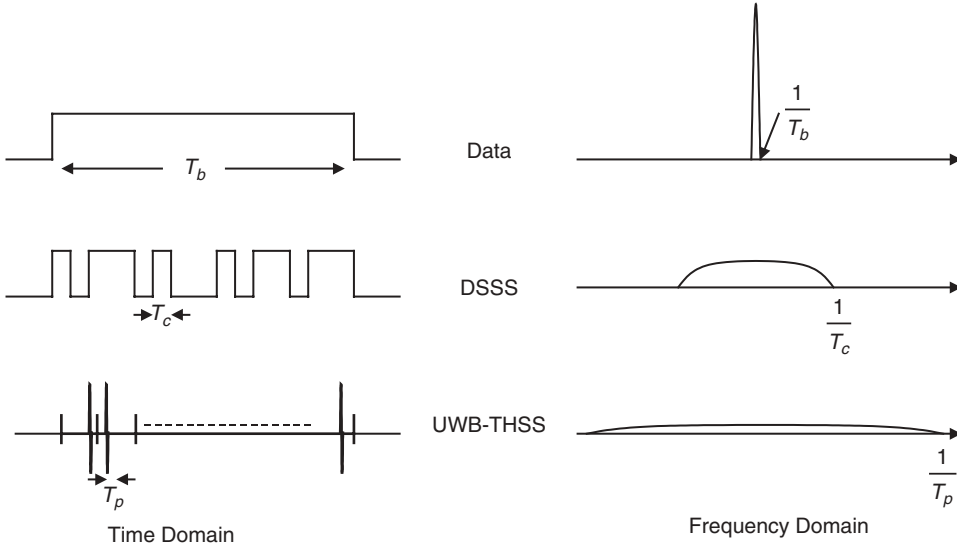


FIGURE 12.9 Comparison between the time and frequency characteristics of narrowband data, DSSS, and UWB-THSS.

12.3.3 Receiver for a Multiuser Environment

If we assume N_u users sharing an impulse-radio system using THSS, the received signal will be given by

$$r(t) = \sum_{j=-\infty}^{\infty} A_k x_{\text{rec}}^k(t - \tau_k) + \eta(t) \quad (12.3.4)$$

in which A_k represents the amplitude of the signal received from the k th transmitter, the random variable τ_k represents the time delay between the k th transmitter and the receiver, and $\eta(t)$ is the background noise at the correlator input. The received waveform is similar to the transmitted waveform given in Eq. (12.3.3), with the difference that here the received waveform is passed through the antenna filter (see Example 12.2). Therefore, we have

$$x_{\text{rec}}^k(t^k) = \sum_{j=-\infty}^{\infty} w_{\text{rec}}(t^k - jT_f - c_j^k T_c - \delta d_{\lfloor j/N_s \rfloor}^k)$$

in which w_{rec} is the received waveform that includes the effects of the antennas. Each user follows a pattern of time hops known to the receiver designed for that particular transmitter, as well as the amplitude and delay associated with that communication link. Using this information, the receiver has to decide on the transmitted information bit which was encoded in the delay δ . Whenever the transmitted information bit is “1,” the additional delay for information is zero, and for information bit “0,” an additional delay of δ is added to the signal.

Example 12.4: Multiuser Receiver for UWB-THSS Figure 12.10 shows the block diagram of a receiver for impulse-radio systems that was suggested in [Win98]. First

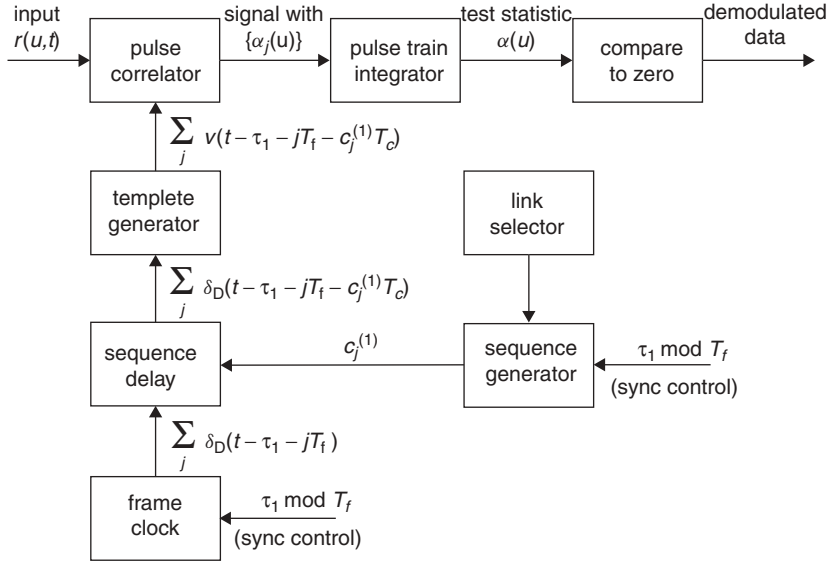


FIGURE 12.10 Impulse radio THSS receiver for the first user, suggested in [Win00]; $\delta_D(t)$ is the Dirac delta function.

the sequence generator, frame clock, and sequence delay blocks generate a sequence of impulses synchronized with the transmitter. Then for each of these impulses, the receiver generates two nonoverlapping pulses, one with zero delay (as if a “0” were sent) and positive sign, and one with δ delay (as if a “1” were transmitted) and a negative sign. This waveform representing two pulses is described mathematically by

$$v(t) = w_{\text{rec}}(t) - w_{\text{rec}}(t - \delta)$$

We correlate this template of the transmitted symbol with the received signal, using two pulses rather than one pulse per time period. If the received pulses had no delay, associated with transmission of a “0,” the first waveform is squared, the negative pulse drops out, and after integration a positive number survives. If a “1” was transmitted, the first positive pulses drops out and negative multiplication accumulates after the integration. Thus, the decision variable is given by

$$\alpha = \hat{d}_k = \sum_{j=0}^{N_s-1} \int_{\tau_1+jT_f}^{\tau_1+(j+1)T_f} r(t)v(t - \tau_k - jT_f - c_j^k T_c) dt$$

and the decisions are made using the following rule:

$$\begin{cases} \alpha > 0 \Leftrightarrow d_k = 0 \\ \alpha \leq 0 \Leftrightarrow d_k = 1 \end{cases}$$

12.4 DIRECT-SEQUENCE UWB

In the actual operation of impulse radio, the UWB pulses interfered with many existing systems, including cellular, WLAN, and GPS systems [Ham02]. In particular, GPS systems operating with very low signal strength were vulnerable to UWB interference. As

a result of these concerns, the FCC decided to mandate the frequency template shown in Fig. 12.1. This decision in year 2003 redirected the attention of the IEEE 802.15.3a UWB standardization activities toward the 3.1- to 10.6-GHz band, and consequently, impulse radio using the lower-frequency bands shown in Fig. 12.7 lost support in the standardization committee. Two leading 802.15 proposals brought forward after the 2003 FCC announcement are known as direct-sequence UWB (DS-UWB) and multi-band OFDM (MB-OFDM), the latter promoted by the MBOA alliance. The DS-UWB technique is addressed in the remainder of this section, and the MB-OFDM technique is described in the next section. The basic coverage cell in the WPAN industry, referred to as a *picocell*, has a nominal coverage range of about 10 m. A network operating within that range is referred to as a *piconet*. Different WPAN technologies support different numbers of overlapping piconets. For example, Bluetooth, the first WPAN standard under IEEE 802.15.1, supports seven overlapping piconets. The new UWB proposals consider multiple bands that can be combined with medium access control to support larger numbers of overlapping piconets.

The DS-UWB system uses the DSSS technique, which emerged successfully as the PHY layer of choice in third-generation cellular networks. This technique employs BPSK and QPSK modulation and a medium access control that combines FDM, TDM, and CDM. In the DS-UWB system, as shown in Fig. 12.11, the 3.1- to 10.6-GHz band is divided into a low band from 3.1 to 4.9 GHz and an optional high band from 6.2 to 9.7 GHz. The bandwidth of the high band is twice the bandwidth of the low band, resulting in shorter time-domain pulses in the high band. The 4.9- to 6.1-GHz band is purposely neglected to avoid interference with IEEE 802.11a devices operating in the 5-GHz U-NII bands. Each piconet of the DS-UWB operates in one of the two bands, and piconets in the same band are separated by code-division multiplexing using ternary multiple biorthogonal keying (M-BOK) spreading codes of length 24 or 32.

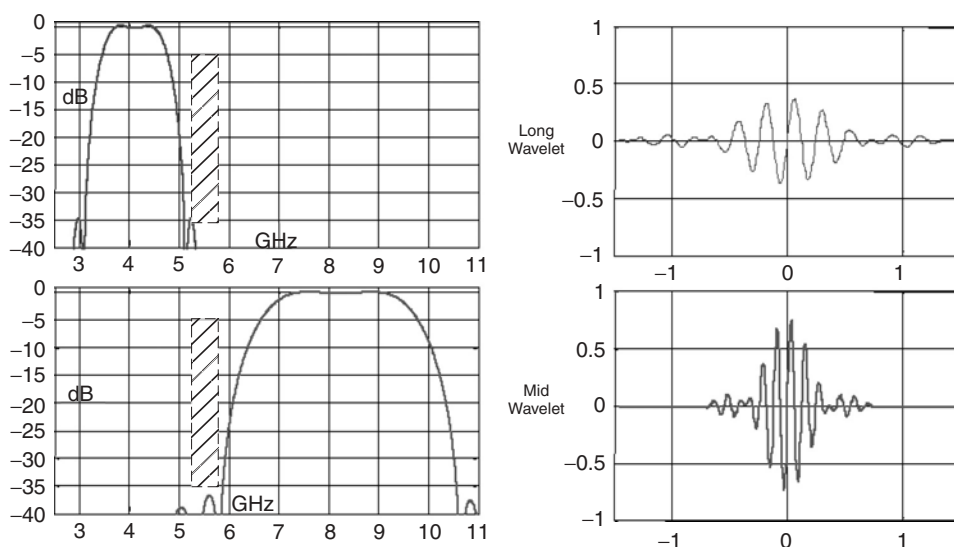


FIGURE 12.11 Frequency and time response of the two basic channels in the DS-UWB proposal [Koh04].

Example 12.5: Data M-BOK Codes Table 12.2 shows the M-BOK ternary codes of length 24 used in the DS-UWB system with BPSK. Each chip can take three values $\{-1, 0, 1\}$. For $M = 2$ the first row or its reverse are used to identify two symbols, and each incoming information bit is mapped to one of the two symbols. For $M = 4$ the first two rows and their reverses are used to form four symbols. The incoming information bits are grouped into two bits, and one of the four symbols is selected for each information digit. For $M = 8$ all four rows and their inverses are used to form 8-MBOK codes. The incoming information bits are grouped into 3-bit blocks, each selecting one of the 8-BOK codes. For more details on the longer codes, one may refer to [Wel03]. Figure 12.12 shows the comparative performance of 2-, 8-, and 16-BOK codes. At low error rates 8- and 16-BOK need approximately 1.0 and 2.5 dB more signal power, respectively, to perform as well as 2-BOK. With this additional power they can increase the data rate three and four times, respectively.

Variable data rates in DS-UWB are achieved by changing the processing gain and switching between BPSK and QPSK modulation. Lower data rates with higher processing gains can cover larger areas. The chip rates in the low and high bands are 1.368 and 2.736 Gc/s, respectively. The basic symbol transmission rate for the low band with spreading factors of 24 and 32 are $(1.368 \text{ Gc/s})/(24 \text{ chips}) = 57$ and

TABLE 12.2 M-BOK Ternary Codes of Length 24 for $M = 2, 4$, and 8^a

Code																								
1	-1	1	-1	-1	1	-1	-1	1	-1	0	-1	0	-1	-1	1	1	1	-1	1	1	1	-1	-1	-1
2	0	-1	-1	0	1	-1	-1	1	-1	-1	1	1	1	1	-1	-1	1	-1	1	-1	1	1	1	1
3	-1	-1	-1	-1	1	-1	1	-1	1	-1	-1	1	-1	-1	1	-1	-1	1	1	0	-1	0	1	1
4	0	-1	1	1	1	-1	-1	-1	-1	-1	-1	1	-1	1	-1	0	1	-1	1	1	-1	-1	1	1

^a2-BOK uses code 1; 4-BOK uses codes 1 and 2; 8-BOK uses codes 1 to 4.

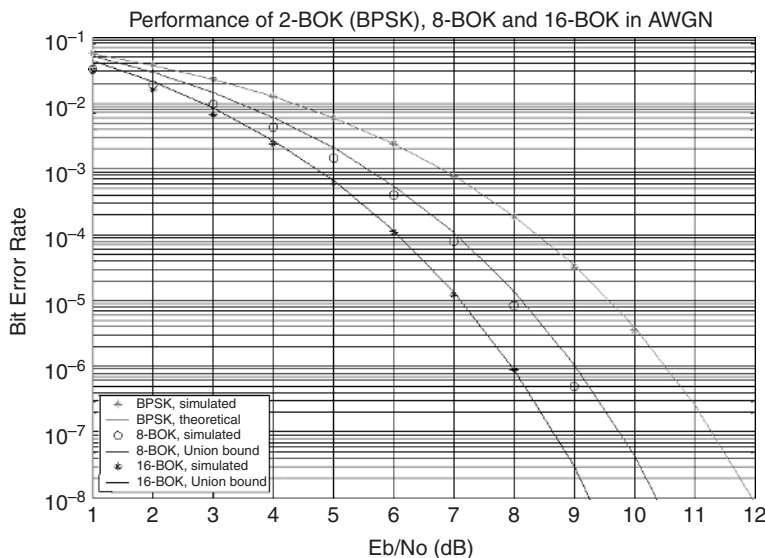


FIGURE 12.12 Theoretical and simulated performance of the M-BOK codes in AWGN [Wel03].

$(1.368 \text{ Gc/s})/(32 \text{ chips}) = 42.75 \text{ MS/s}$. Similarly, high-band operation supports two basic data rates, 114 and 85.5 MS/s. Different data rates are derived from these basic rates and three different coding schemes. The first coding option is a convolutional code with coding rate $R = 0.50$. The second coding option is a (55,63) Reed–Solomon code with rate $R = 0.87$. The third option is a concatenated code that uses both a rate- $\frac{1}{2}$ convolutional and a (55,63) Reed–Solomon code, resulting in an overall code rate $R = 0.44$.

Example 12.6: Data Rate Calculation in DS-UWB For the 25-Mb/s data rate in Table 12.3, we have two codes, representing bits 0 and 1 and a BPSK modulation (no quadrature phase). Therefore, each point in the constellation is identified by 1 bit and the uncoded transmission data rate is 57 Mb/s, the same as the symbol transmission rate. The coded data rate is $57 \text{ Mb/s} \times 0.44 = 25 \text{ Mb/s}$. For the 200-Mb/s rate, we have four symbols, resulting in 2 bits per symbol and a QPSK modulation that has another 2 bits per symbol. Therefore, in the overall scheme we have 4 bits per symbol. With a symbol rate of 114 MS/s and a coding rate of $R = 0.44$, we have

$$114 \text{ Ms/s} \times 4 \text{ bits/s} \times 0.44 = 200.64 \text{ Mb/s}$$

As we noted earlier, the medium access control combines FDM, TDM, and CDM techniques. The FDM scheme chooses one of the two operating frequency bands to control the interference. The CDM uses ternary code sets $(\pm 1, 0)$ with 2,4,8-BOK with length 24 and 64-BOK with length 32. Four CDMA codes within each frequency band are used to further separate the piconets by providing for implementation of logical channels. Within each piconet TDM separates different users. The following example is provided to aid the understanding of this complex medium access method.

Example 12.7: Medium Access Control in DS-UWB Figure 12.13 illustrates an example using FDM, CDM, and TDM access in the DS-UWB scheme in a multi-room indoor area. Each set of piconets is separated by FDM between the low band (LB) and the high band (HB). Within each set using the same band, the piconets are separated by CDM codes into channels A, B, and C. Then, within each cell, users are separated by TDMA using a central scheduling system similar to those used in other TDMA systems, such as GSM and HIPERLAN2.

TABLE 12.3 Various Data Rates Supported by DS-UWB

Information Data Rate (Mb/s)	Constellation	Symbol Rate	Quadrature	FEC Rate, R
25	2-BOK	57	No	0.44
50	2-BOK	114	No	0.44
114	4-BOK	114	No	0.50
112	8-BOK	85.5	No	0.44
200	4-BOK	114	Yes	0.44
224	64-BOK	85.5	No	0.44
450	64-BOK	85.5	Yes	0.44
900	64-BOK	85.5	Yes	0.87

Source: [Koh04].

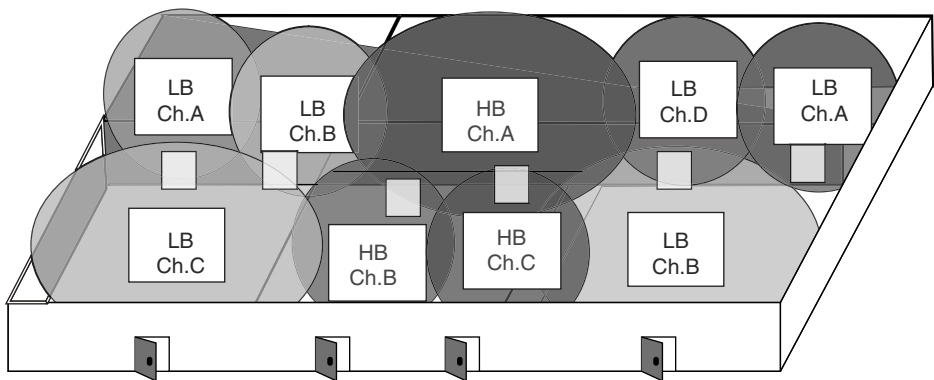


FIGURE 12.13 Distribution of piconets in the FDM/CDM/TDM access used in the DS-UWB proposal for IEEE 802.15.3a (© IEEE).

12.5 MULTIBAND OFDM

At this writing, the second and leading proposal for the IEEE 802.15.3a is multiband OFDM, the MB-OFDM technology developed by the MBOA Alliance. The MB-OFDM system [Sho03] uses the OFDM technique, which emerged as the technology of choice for IEEE 802.11 WLAN standards operating in the U-NII 2.4- and 5-GHz unlicensed bands, and in the UWB 3.1- to 10.6-GHz unlicensed bands. Following this approach, the spectrum is divided into 15 bands each of width 528 MHz. In each band a 128-point OFDM system using QPSK modulation is implemented to limit the required precision of mathematical operations and make digital implementation at ultrahigh sampling rates feasible. The medium access control is time-frequency multiple access (TFMA), which combines the time- and frequency-diversity benefits of FHSS and DSSS into one MAC technique.

Example 12.8: TFMA in MB-OFDM Figure 12.14 [Sho03] illustrates the basic operation of TFMA for a spread-spectrum code of length 7. The chip duration is 3.79 ns,

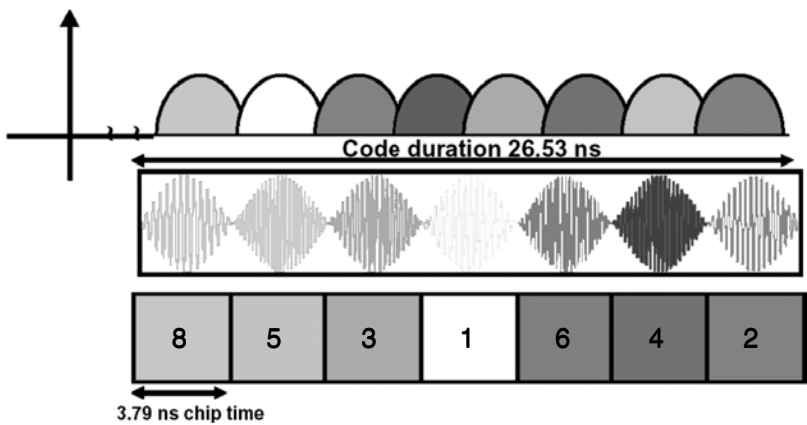


FIGURE 12.14 TFMA operation using a length 7 time frequency code. Each chip of a spread-spectrum code is transmitted in a different frequency channel [Sho03].

resulting in a code duration of $3.79 \times 7 = 26.53$ ns. Similar to DSSS, each symbol is divided into 7-chip periods to provide time diversity. Similar to FHSS, each chip is transmitted on a different frequency channel to provide frequency diversity.

Figure 12.15 gives an overview of the MB-OFDM proposal. The 15 bands in the 3.1- 10.6-GHz unlicensed UWB spectrum are divided into five groups of 528-MHz bands. Group 1 is the most desirable because group 2 interferes with U-NII bands and IEEE 802.11a devices, and higher groups have smaller coverage areas. Each physical piconet is implemented in a band group and several logical piconets share a band group using different TFMA codes. Groups 1 to 4 have four time-frequency codes and group 5 has two TF codes for logical channel separation. In this manner this proposal can accommodate 18 piconets in the entire UWB spectrum, and four of these piconets implemented in band group 1 are the most popular of all.

Table 12.4 provides the four patterns of TF codes used in groups to 4 and the two patterns for group 5. Groups 1 to 4 have three different frequencies, and the length of the time sequence is 6, in which each band is used twice per symbol. Group 5 has two bands, a code length of 4, and each band is used twice during each symbol transmission. With this technique, as shown in Table 12.4, neighboring piconets have two collisions per code length. Multirate data communication in this system is supported by adjusting the spread of the pulses in time. The same time sequence is spread by a factor of 2 or 4 to increase the symbol transmission rate by 2 or 4. The following example illustrates this technique.

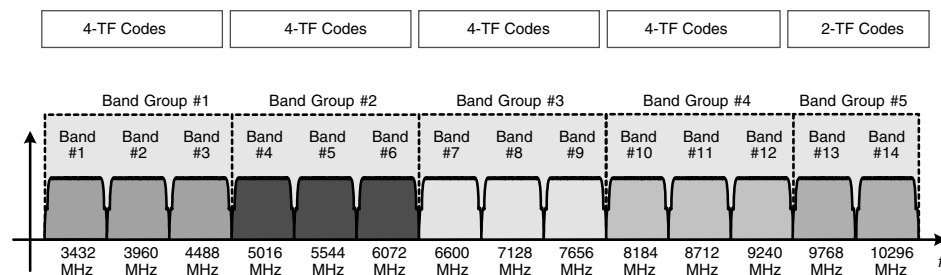


FIGURE 12.15 Frequency bands, groups of frequencies, and TFMA codes within each group of the MB-OFDM approach to UWB communications in 3.1- to 10.6-GHz unlicensed UWB bands proposed to the IEEE 802.15.3a WPAN standard.

TABLE 12.4 TF Codes Recommended by MBOA for the MC-OFDM System

Band Groups	Preamble Pattern	TF Code Length	Time Frequency Code					
1 to 4	1	6	1	2	3	1	2	3
	2	6	1	3	2	1	3	2
	3	6	1	1	2	2	3	3
	4	6	1	1	3	3	2	2
5	1	4	1	2	1	2		
	2	4	1	1	2	2		

Source: [Wel03].

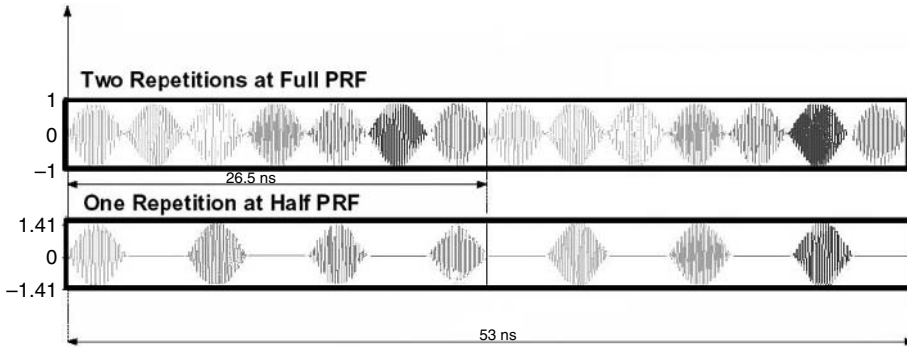


FIGURE 12.16 Control of data rate using spreading time. The data rate of the lower spreading code shown at the top is twice that of the lower code spread in twice as much time.

Example 12.9: Variable Data Rate in TFMA Figure 12.16 shows the TFMA system with seven bands, discussed in Example 12.8. The top of the figure is the same as in Fig. 12.14 with a full pulse rate frame (PRF) in which pulses are transmitted in sequence one after another. The lower part of the figure shows a transmission with a half PRF, where pulses are transmitted every other chip interval. The data rate of the scheme shown in the lower part of the figure is half of the data rate of the scheme shown in the upper part of the figure.

Figure 12.17 shows a general block diagram of the OFDM system proposed by the MBOA Alliance. This system uses a 242.4-ns information length with a 60.6-ns prefix for multipath protection, and a 9.5-ns guard interval to provide time for switching between bands, for a total symbol duration of 312.5 ns. From the 128 tones or carriers, 100 are data tones used to transmit information, 12 are pilot tones used for carrier and phase tracking, 10 are guard tones (previously called *dummy tones*), and 6 are NULL tones. Table 12.5 provides specifications for all mandatory and optional data rates supported by this system. A simple example illustrates how this table can be read.

Example 12.10: Variable Data Rate in TFMA Since the symbol duration is 312.5 ns and the modulation is QPSK with 2 bits/symbol, the basic transmission rate for all data rates is

$$1/312.5 \text{ (ns/symbol)} \times 2 \text{ (bits/symbol)} \times 100 \text{ (carriers)} = 640 \text{ Mb/s}$$

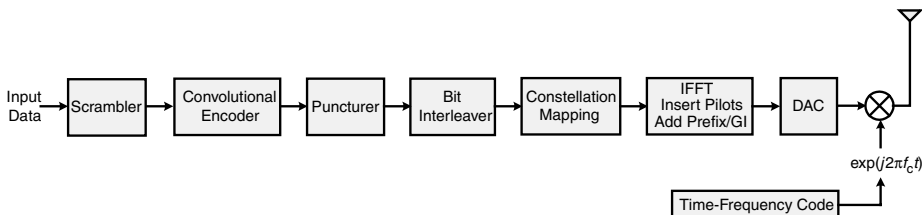


FIGURE 12.17 General block diagram of the OFDM modulation system recommended for the MC-OFDM proposal [Wei03].

TABLE 12.5 Specification of Various TF Codes Recommended by MBOA for the MC-OFDM System

	Information Data Rate (Mb/s)						
	55 ^a	80 ^b	110 ^a	160 ^b	200 ^a	320 ^b	480 ^b
Modulation/ constellation	OFDM/QPSK	OFDM/QPSK	OFDM/QPSK	OFDM/QPSK	OFDM/QPSK	OFDM/QPSK	OFDM/QPSK
FFT size	128	128	128	128	128	128	128
Coding rate ($K = 7$), R	11/32	1/2	11/32	1/2	5/8	1/2	3/4
Spreading rate	4	4	2	2	2	1	1
Data tones	100	100	100	100	100	100	100
Information length (ns)	242.4	242.4	242.4	242.4	242.4	242.4	242.4
Cyclic prefix (ns)	60.6	60.6	60.6	60.6	60.6	60.6	60.6
Guard interval (ns)	9.5	9.5	9.5	9.5	9.5	9.5	9.5
Symbol length (ns)	312.5	312.5	312.5	312.5	312.5	312.5	312.5
Channel bit rate (Mb/s)	640	640	640	640	640	640	640

Source: [Wei03].

^a Mandatory.

^b Optional.

TABLE 12.6 Simulated Performance of the MB-OFDM System in Four Different Band Groups^a

	AWGN	CM1	CM2	CM3	CM4
Range (m)					
110 Mb/s	20.5	11.4	10.7	11.5	10.9
200 Mb/s	14.1	6.9	6.3	6.8	4.7
480 Mb/s	7.8	2.9	2.6	N/A	N/A

Source: [Wel03].

^aThe coverage is compared with the coverage in an AWGN.

The lowest data rate in the table is 55 Mb/s, which is obtained by a coding rate of 11/32 with a spreading rate (repetition for variable data rate support) of 4. Therefore, the data rate is

$$640 \text{ Mb/s} \times 11/32 \text{ (bits/coded bits)} / 4 \text{ (spreading rate)} = 55 \text{ Mb/s}$$

Table 12.6 shows the range of coverage for the 100-, 200-, and 480-Mb/s systems in AWGN and the first four groups of bands for the MB-OFDM proposal. The 480-Mb/s option in band groups 3 and 4 has limited coverage, and in the first two bands has range less than 3 m. This distance is suitable for WPAN applications such as connecting a device to the USB port of a computer or laptop. The 110-Mb/s option covers about 10 m, which is the desirable coverage range for IEEE 802.15 devices. The 200-Mb/s option covers more than 6 m in the first three groups and less than 5 m in the fourth group. The performance criterion for this simulation was a packet error rate of 8% in 90% of locations. Simulations include several practical factors, such as losses due to front-end filtering and conversions, multipath degradation, channel estimation errors, carrier tracking, and packet acquisition. More detailed analysis of the performance, effects of interference with IEEE 802.11 and other devices, and specification for implementation of MB-OFDM systems may be found in [MBO04].

QUESTIONS

- What is the difference between the direct pulse transmission technique used for channel measurement in Chapter 5 and the pulse transmission technique used for UWB communications?
- Which IEEE standardization group is involved in UWB communication, and what other popular wireless standard has been developed with that group?
- Compare Bluetooth and UWB as two solutions for WPAN applications.
- Name two popular antenna systems used for UWB communications and compare them for practical applications in personal communications.
- Which ISM band overlaps with the UWB bands, and which IEEE WLAN standard uses that band?
- What is a natural breakpoint for modeling the path loss for an UWB device that is located inside a building?

- (g) What are the data rates, frequency bands, and MAC and PHY layers of the DS-UWB proposal for 802.15.3a?
- (h) What are the data rates, frequency bands, and MAC and PHY layers of the MB-OFDM proposal for 802.15.3a?
- (i) What is the difference between DSSS and UWB time hopping?
- (j) What is the difference between bandwidths and signal spreading techniques of the DSSS and UWB time-hopping technique?
- (k) Give the frequency band, modulation technique, chip rate, code length, and number of code words in the constellation for the 900-Mb/s DS-UWB proposal for 802.15.3a.
- (l) Why in the DS-UWB system is the band between 4.9 and 6.2 GHz not used?
- (m) Give the range of the frequency bands, number of channels, and the range of data rates that are supported by 802.15.3a DS-UWB proposal.
- (n) Give the range of the frequency bands, number of channels, and the range of data rates that are supported by the 802.15.3a MB-OFDM proposal.
- (o) What is the MB-OFDM proposal for the IEEE 802.15.3a, how does it relate to the UWB, and what is its medium access control?
- (p) How many different piconets can be made in the MB-OFDM proposal for 802.15.3a? Explain how you got that number. If we avoid frequency bands for 802.11a, how many piconets are left?

PROBLEMS

1. Assume that a UWB device with a bandwidth of 1 GHz operates in a band that includes the entire IEEE 802.11a band. If both devices attempt to communicate with an access point supporting both technologies, and they are located at the same close distance from the access point for which the distance–power gradient is 2:
 - (a) What is the signal-to-interference ratio for the received IEEE 802.11a signal at the access point?
 - (b) What is the signal-to-interference ratio for the received UWB signal at the access point?
2. (a) Assuming that the starting frequency of a band is given by f_{st} , use Eq. (12.2.1) to derive the received UWB power to the NB power ratio as a function of bandwidth W and the starting frequency.
 - (b) Using the equation derived in part (a), plot the power ratio in decibels versus $0 < f_{st} < 5$ GHz for $W = 5$ GHz.
 - (c) Repeat part (b) for $W = 10$ GHz.
 - (d) Using the results of parts (b) and (c), discuss the sensitivity of the flatness of the spectrum to the starting frequency and bandwidth of the system.
3. (a) Using Eq. (12.3.3), sketch the two symbols used for binary communications using the time-hopping UWB pulse transmission technique for $T_f = 100$ ns,

$N_s = 7$, $T_c = 10$ ns, $\delta = 1$ ns, and for the time-hopping sequence, $\{c_j\}$. Use the length 8 LFSR code described in Example 10.1.

- (b) What is the data rate of this system?
 - (c) Use the receiver described in Fig. 12.10 to show how binary waveforms generated in part (a) can be detected.
 - (d) Calculate the energy per bit received after processing at the receiver, and give an equation for calculation of the error rate in terms of the received SNR per bit.
4. (a) Give all 8-BOK codes of length 24. Show that the first code is orthogonal to the second and third codes.
- (b) Calculate the two basic data rates of the DS-UWB for 24 and 32 spreading factors in the high band.
5. (a) Show that for all TF hopping patterns for the MB-OFDM shown in Table 12.4, we have two collisions per code.
- (b) Repeat Example 12.9 for all data rates given in Table 12.5.

13

RF LOCATION SENSING

- 13.1 Introduction
 - 13.1.1 Technical Aspects of RF Positioning
- 13.2 RF Location-Sensing Techniques
 - 13.2.1 TOA Techniques for Indoor and Urban Ranging
 - 13.2.2 RSS Techniques Metrics
 - 13.2.3 Direction-Based Techniques
- 13.3 Modeling The Behavior of RF Sensors
 - 13.3.1 Models of the Behavior of AOA
 - 13.3.2 Behavior of RSS Sensors
 - 13.3.3 Behavior of TOA Sensors
- 13.4 Wireless Positioning Algorithms
 - 13.4.1 Traditional Techniques
 - 13.4.2 Pattern Recognition Techniques
- Questions
- Problems

13.1 INTRODUCTION

The problem of locating mobile radios originated with military operations during World War II, when it was critical to locate soldiers in emergency situations. About 20 years later, during the Vietnam conflict, the U.S. Department of Defense launched a series of global positioning system (GPS) satellites to support military operations in combat areas. In 1990, the signals from GPS satellites were made accessible to the private sector for commercial applications such as fleet management, navigation, and emergency assistance. Today, GPS technology is widely available in the civilian market for personal navigation applications. GPS receivers are designed to determine the locations of boats, planes, or mobile vehicles in open areas such as waterways, skyways, and highways. However, the accuracy of GPS positioning is significantly impaired in urban and indoor areas, where received signals can suffer from multipath effects. As we discussed in Chapter 10, the IS-95 CDMA standard uses GPS-derived

location and time information to synchronize its base stations. A fuller description of GPS system is beyond the scope of this book, but the interested reader may find much information in the open literature [Kap96].

In 1996 the FCC introduced regulations requiring wireless service providers to be able to locate mobile callers in emergency situations with specified accuracy: 100-m accuracy 67% of the time. Such emergency service is called E-911 in the United States and E-112 in many other countries. In a manner similar to the release of the ISM bands and subsequent emergence of the WLAN industry, the FCC mandate for E-911 services quickly gave rise to the development of the wireless geolocation industry. The service and E911 industries were spurred on by the growing cellular subscriber base and the inability of traditional 911 services to support reliable caller identification and location. In addition to the E-911 services, work on the underlying problem of locating a mobile user within a wireless network has created interest in a number of other location-based applications, such as child and elder tracking, driving directions, mobile worker management, asset tracking, and electronic location confinement. Although GPS-based location is reliable and accurate in outdoor situations, it does not provide satisfactory performance in indoor and urban areas with extensive multipath conditions. Furthermore, GPS technology is not easily integrated into cellular telephone infrastructure and mobile devices. Therefore, other technologies have been developed to implement the E-911 mandate [Caf98, McG02]. These technologies include GPS-assisted techniques, a variety of time-of-arrival (TOA) techniques, angle-of-arrival (AOA) techniques, and technologies using measurements of received signal strength (RSS). GPS-assisted systems perform well in rural areas but require the addition of GPS receiver functionality to cellular phones. A variety of TOA, time differential (TDOA), and extension of time differential (EOTD) techniques require special location-measurement hardware integrated in the base stations and in some cases accurate synchronization between the mobile terminals and base stations. In contrast with those approaches, RSS systems provide a lower-cost solution that can avoid additional hardware installation but does require incorporating training functions into the system.

As an alternative to TOA/TDOA or enhanced observed time difference (EOTD) solutions using geometric triangulation with more accurate TOA metrics, in the early 2000s a group of E-911 service designers began working with statistical radio channel propagation modeling experts to develop RSS databases for urban-area positioning. The RSS-based technologies do not require additional hardware; therefore, they are less costly and less complex to implement and manage. As we saw in Chapter 4, due to the complexity of radio-wave propagation in urban areas, there are some parameters, such as the distance–power gradient, whose true values cannot be derived from the underlying theory. To calculate these parameters for specific areas, we need empirical signal data. One challenge for these systems is to create and update a database for measurements and to incorporate more complex channel models that can use the data efficiently. The accuracy of this new wave of RSS-based systems is within 20 to 50 m [McG02]. Although these urban positioning systems use the infrastructure of the cellular system, there are other novel techniques for taking advantage of the RSS information gathered from numerous WLAN APs installed at scattered locations in commercial and residential areas. To train the WLAN-based systems, a mobile terminal is operated in a target urban area to measure the signal strengths of the APs and to associate them with their IP addresses and their physical locations in an electronic map. Using the electronic map with identified locations and with expected RSS levels

associated with specific IP addresses, these systems can locate a mobile wireless device. Since coverage area of a WLAN is much smaller than that of a cellular base station, these systems can provide more accurate location estimates than those derived using cellular infrastructure.

In the late 1990s, at about the same time that E-911 technologies were emerging, another initiative for accurate indoor geolocation began independently, motivated by a variety of envisioned applications for indoor location sensing in commercial, public safety, and military settings [Kos00, Pot00, Pah02]. In commercial applications for residences and nursing homes, there is an increasing need for indoor location-sensing systems to track people with special needs, the elderly, and children who are away from visual supervision. Other applications include systems to assist the sight-impaired, to locate instrumentation and other equipment in hospitals, to locate surgical equipment in an operating room, and to locate specific items in warehouses. In public safety and military applications, indoor location-sensing systems are needed to track inmates in prisons and to guide policemen, firefighters, and soldiers in accomplishing their missions inside buildings. More recently, location sensing has found applications in location-based handoffs in wireless networks [Pah00], location-based adhoc network routing [Ko98, Jai01], and location-based authentication and security [Sma02]. These and other applications have stimulated interest in modeling the propagation environment to assess the accuracy of various sensing techniques [Pah98, Kri99], as well as in developing novel technologies to implement the systems [Bah00a, Bah00b, Fon01]. We are already seeing implementations of the first generation of indoor positioning products using a variety of technologies [Wer98, Roo02a, Roo02b]. To help the growth of this emerging industry, there is a need to develop a scientific framework and foundation for design and performance evaluation of such systems.

The most popular emerging indoor location and tracking systems operate using information about RF signal properties in large indoor areas [Pah02a, Pah02b]. However, location-sensing using multiple cameras [Foc02] and ultrasound [Pri00] are also becoming popular for line-of-sight (LOS) applications within a room. One can also think of more general applications, such as locating unwanted chemical or biological substances using distributed sensors or locating radioactive material using sensors reporting the presence of such material. Although our interest in this book is RF sensing, the framework that we discuss for performance evaluation is applicable to any indoor positioning and tracking system. In RF-based sensing systems the target object includes a trackable component such as an RF tag, a cell phone, or a WLAN network card. The position-sensing system locates the target based on the signal radiated from the device.

There is analogy between the operation of mobile data services versus WLANs on the one hand, and wireless E-911 versus indoor geolocation systems on the other. Mobile data services provide lower data rates with comprehensive coverage over a large area, while WLANs provide higher-speed data communication for local spot coverage. E-911 services are expected to provide low-accuracy positioning (around 100 m) with comprehensive coverage, whereas indoor geolocation systems are expected to provide high-accuracy local positioning (about a few meters). As with mobile data and WLANs, the basic concepts and technologies used in E-911 services and indoor geolocation are very similar. In the remainder of this chapter we discuss the underlying technical aspects for these technologies, with primary emphasis on indoor geolocation. We describe the system concepts used in the two industries and discuss the characteristics

of radio propagation for geolocation in multipath environments. Then we introduce algorithms used in wireless geolocation and methods for performance evaluation of these systems.

13.1.1 Technical Aspects of RF Positioning

In this section we provide an overview of the wireless geolocation systems used for RF positioning. Figure 13.1 illustrates the functional block diagram of a wireless geolocation system. The main elements of the system are a number of location-sensing devices that provide metrics related to the relative position of a mobile terminal (MT) with respect to a known reference point (RP), a positioning algorithm that processes metrics reported by location-sensing elements to estimate the location coordinates of MT, and a display system that depicts the location of the MT to the users. The location metrics may indicate the approximate arrival direction of the signal or the approximate distance between the MT and the RP. The angle of arrival (AOA) is the most common metric used in direction-based systems. The received signal strength (RSS), carrier signal phase of arrival (POA), and time of arrival (TOA) of the received signal are the metrics used for estimation of the distance. As the measurement of metrics becomes less reliable, the complexity of the position algorithm increases. The display system can simply show the coordinates of the MT or may identify the relative location of the MT in the layout of an area. This display system could be a software residing in a private PC or a mobile locating unit, a locally accessible software in a local area network, or a universally accessible service on the Web. Obviously, as the horizon of the accessibility of the information increases, design of the display system becomes more complex.

There are two basic approaches to designing a wireless geolocation system. The first approach is to develop a signaling system and a network infrastructure of location sensors focused primarily for geolocation application. The second approach is to use an existing wireless network infrastructure to locate an MT. The advantage of the first approach is that the physical specifications, and consequently, the quality of the location-sensing results, are under the control of the designer. The MT can be designed as a very small wearable tag or sticker, and the density of sensor infrastructure can be adjusted to the accuracy required for the location-finding application. The advantage of the second approach is that it avoids expensive and time-consuming deployment

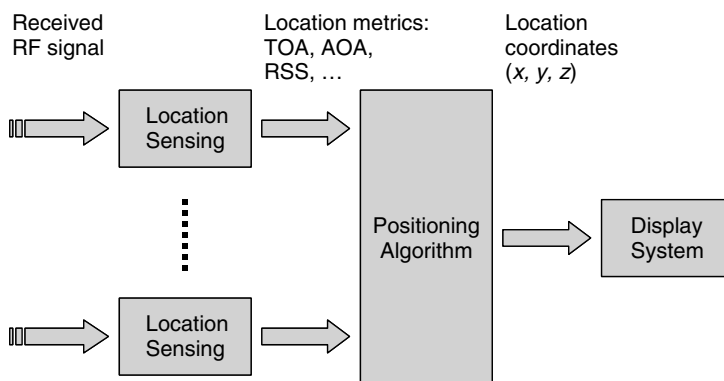


FIGURE 13.1 Functional block diagram of a wireless geolocation system.

of infrastructure. These systems, however, need to use more intelligent algorithms to compensate for the low accuracy of the metrics measured. Both approaches have their own markets, and design work for both technologies has been pursued in the past few years.

In this chapter we provide an overview of the application and models for behavior of RF sensors, discuss algorithms used to process the results of sensors in location finding, and provide a framework for performance evaluation of positioning systems.

13.2 RF LOCATION-SENSING TECHNIQUES

RF location sensors operating in different environment measure RSS, AOA, POA, TOA, and the signature of the delay power profile as location metrics [Pah02a, Pah02b]. Indoor and urban radio channels suffer from severe multipath propagation and heavy shadow-fading conditions, so these measurements are far from accurate in many instances. In general, measurements of POA and AOA in large indoor and urban areas provide very unreliable results, and sensors used for indoor positioning normally sense the more accurate TOA or the easy-to-measure RSS. However, dramatically large errors due to undesirable multipath conditions also occur in TOA estimation. To estimate the TOA in indoor areas accurately, we need to resort to different and more complex signaling formats, frequencies of operation, and signal processing techniques that can resolve the problems. The signature of the delay power profile, available at RAKE receivers in third-generation systems provides a combination of TOA and RSS of the dominant paths that can provide more accurate metrics for positioning. We start our discussion on metrics with TOA estimation techniques.

13.2.1 TOA Techniques for Indoor and Urban Ranging

TOA-based systems measure the distance based on an estimate of signal propagation delay (i.e., TOA), between a transmitter and a receiver since in free space or in air, radio signals travel at the constant speed of light. The TOA can be measured either by measuring the phase of the received narrowband carrier signal or by direct measurement of the arrival time of a wideband narrow pulse. The wideband pulses for measuring TOA can be generated either directly or by using spread-spectrum technology. In this section we present these techniques in three classes: narrowband, wideband, and ultrawideband techniques.

The behavior of the TOA sensors in indoor multipath propagation is highly sensitive to the bandwidth of the sensor. The UWB systems, which exploit bandwidths in excess of 1 GHz, have attracted considerable attention as a means of measuring accurate TOA for indoor geolocation applications [Fon01]. Due to the high attenuation associated with the use of a high-frequency carrier and recent FCC regulations on UWB, these systems are typically focused on 3.1 to 10.6-GHz unlicensed bands. With the results of propagation measurement in a typical modern office building, it has been shown that the UWB signal does not suffer multipath fading, which is desirable for accurate TOA estimation in indoor areas. However, similar to other TOA systems, UWB systems cannot avoid undetected direct path (UDP) problems [Pah98], they have limited coverage, and their actual deployment requires compliance with FCC regulations. The main concern of the FCC authorities is the interference of UWB devices, among other licensed services, to the GPS systems that operate approximately at the

1.5-GHz frequency band. To achieve high resolution with limited bandwidth, super-resolution techniques can be applied [Li03a,b]. This approach uses spectral estimation techniques to refine the resolution in a bandlimited situation.

Narrowband TOA Phase Metric. The received signal phase of the carrier is one possible geolocation metric. In this technique, shown in Fig. 13.2a, the narrowband phase difference between received and transmitted carrier signals is used to measure the distance between two points. The phase of a received carrier signal, ϕ , and the TOA of the signal, τ , are related by $\tau = \phi / \omega_c$, where ω_c is the carrier frequency in rad/sec. It is well known that differential GPS (DGPS) using a measured reference carrier phase at the receiver improves the location accuracy of the traditional GPS from about 20 m to within 1 m [Kap96]. One problem associated with phase measurements lies in the ambiguity, shown in Fig. 13.2a, resulting from the periodic property (with period 2π) of the signal phase, while as we see in the next two sections, the standard DSSS and UWB measurements are unambiguous. Consequently, in DGPS, the ambiguous carrier phase measurement is used to fine-tune the DSSS measurements. In GPS systems, a complementary Kalman filter is used to combine the low-noise ambiguous carrier phase measurements and the unambiguous but noisier TOA measurements [Kap96]. Unlike the GPS/DGPS, where the DLOS signal path is always assumed to be present, the severe multipath condition of the indoor and urban geolocation environment causes substantial errors in the phase measurements. When a narrowband carrier signal is transmitted in a multipath environment, as shown in Fig. 13.2b, the composite received carrier signal is the sum of a number of carriers arriving along different paths of the same frequency but different amplitude and phase. The frequency of the composite received signal remains unchanged, but the phase can differ substantially from that of the DLOS signal, resulting in substantial distance measurement error. The immediate

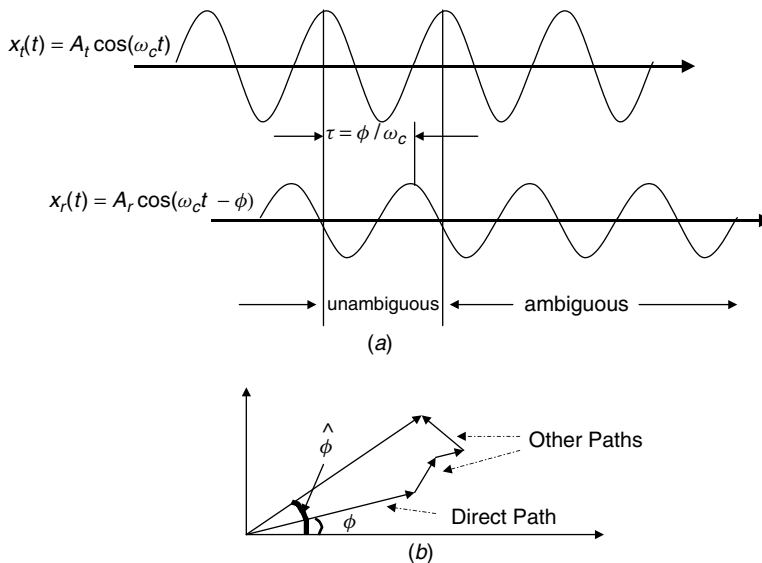


FIGURE 13.2 (a) Narrowband TOA measurement using the phase of the received signal and the ambiguity in the measurement; (b) phasor diagram for multipath arrival.

conclusion is that phase-based distance measurement using narrowband carrier signal cannot provide an accurate estimate of distance in a heavy multipath environment.

Wideband TOA Metrics. Direct-sequence spread-spectrum (DSSS) wideband signals have been used in GPS and other ranging systems for many years. In such a system, similar to the wideband measurement systems described in Section 5.2.2, a signal coded by a known pseudorandom (PN) sequence is transmitted. Then a receiver cross-correlates the received signal with a locally generated PN sequence using a sliding correlator, shown in Figs. 5.4 and 5.5, or a matched filter described in Fig. 10.10. The distance between the transmitter and the receiver is determined from the arrival time of the first correlation peak. Because of the processing gain of the correlation process at the receiver, the DSSS ranging systems perform much better than competing systems in suppressing interference from other radio systems operating in the same frequency band. Figure 13.3 shows the basic concepts involved in wideband TOA measurements using the arrival time of the first path in a typical indoor multipath environment [Ala03a,b]. In this figure the direct path (DP) is represented by the first path, which is also the strongest path. The location of this path is the expected value of the TOA. Other paths with a number of reflections and transitions arrive after the DP with lower amplitudes. These paths would have been observed at the receiver if the bandwidth of the system was infinity. In practice, bandwidth is limited, and as we described in Section 5.2.2, the received signal will be a number of pulses whose amplitude and arrival time are the same as those of impulses, but they have a pulse shape, and the addition of all these pulse shapes forms the received signal, which we refer to as the *channel profile*. A common practice to estimate the location of the DP is to estimate that with the peak of the first path, which is the estimated TOA. In a single-path environment, the actual expected and estimated direct paths are the same. In multipath conditions, however, as shown in Fig. 13.3, the peak of the channel profile gets shifted from the expected TOA resulting in a TOA estimation error caused by

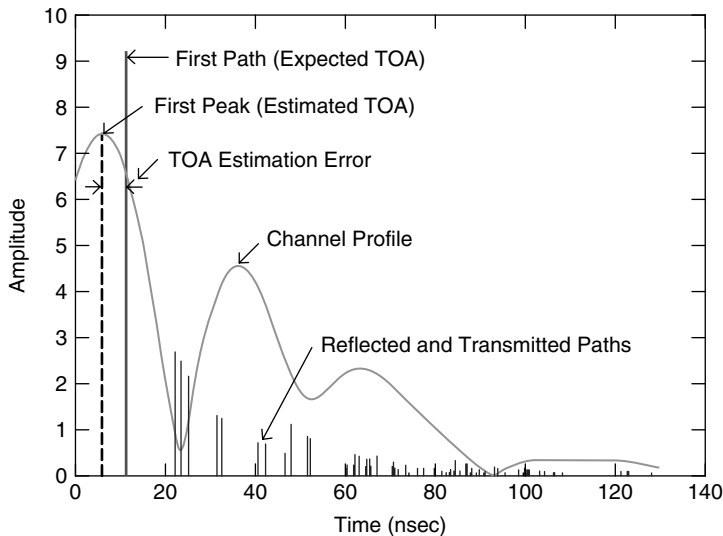


FIGURE 13.3 Parameters involved in wideband TOA measurement using the arrival of the first path directly connecting the transmitter and the receiver.

the multipath conditions. We refer to the distance error caused by erroneous estimates of the TOA as the distance measurement error. For a given multipath condition, we expect that as we increase the bandwidth, the distance measurement error is reduced.

Superresolution for TOA Estimation. Due to the scarcity of the available bandwidth in practice, in some indoor geolocation applications, the DSSS ranging systems cannot provide adequate accuracy. On the other hand, it is always desirable to achieve higher-ranging accuracy using the same bandwidth. Inspired by high-resolution spectrum estimation techniques, a number of researchers have studied superresolution techniques [Pal91, Li03a,b]. Figure 13.4 illustrates the usefulness of the superresolution technique for a sample measured indoor channel. In this figure the MUSIC algorithm is used as an example of superresolution techniques. In the first of the other two techniques, the frequency-domain channel response is converted directly to the time domain using inverse Fourier transform (IFT), and then the arrival time of the DLOS is detected. The second technique uses traditional cross-correlation techniques with DSSS signals (DSSS/xcorr). The superresolution technique can determine the TOA with much higher resolution from the frequency channel response.

Ultrawideband (UWB) Approach. As mentioned earlier, the signal bandwidth is one of the key factors that affects TOA estimation accuracy in multipath propagation environments: The larger the bandwidth, the higher the ranging accuracy. In Chapter 12 we observed that these systems can exploit bandwidths in excess of several gigahertz. Naturally, these systems have attracted considerable attention as a means of measuring accurate TOA for indoor geolocation applications. As shown in Figs. 12.1 and 12.2, FCC regulations allow an unlicensed flat frequency bands at 3.1 to 10.6 GHz, for which there are two proposals: the DS-UWB and MB-OFDM. The larger band in DS-UWB, shown in Fig. 12.11, occupies a 2.7-GHz band, and each channel of the MB-OFDM, shown in Fig. 12.15, are 528 MHz wide. Figure 13.5a illustrates a typical measurement of the channel impulse response in an office area with 500 MHz of bandwidth, resembling the MB-OFDM channels, and Fig. 13.5b provides the same profile with 3-GHz bandwidth, resembling the wider channel of the DS-UWB. The expected delay of arrival, depicting the actual distance between the transmitter and the receiver, is 40.5 ns, and the estimated arrival with the 500-MHz and 3-GHz bands

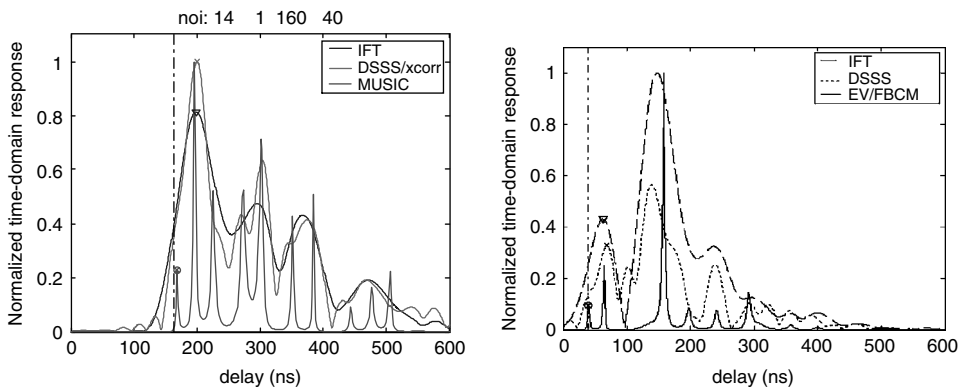


FIGURE 13.4 Effectiveness of the superresolution algorithm to resolve multipath components [Li03a,b].

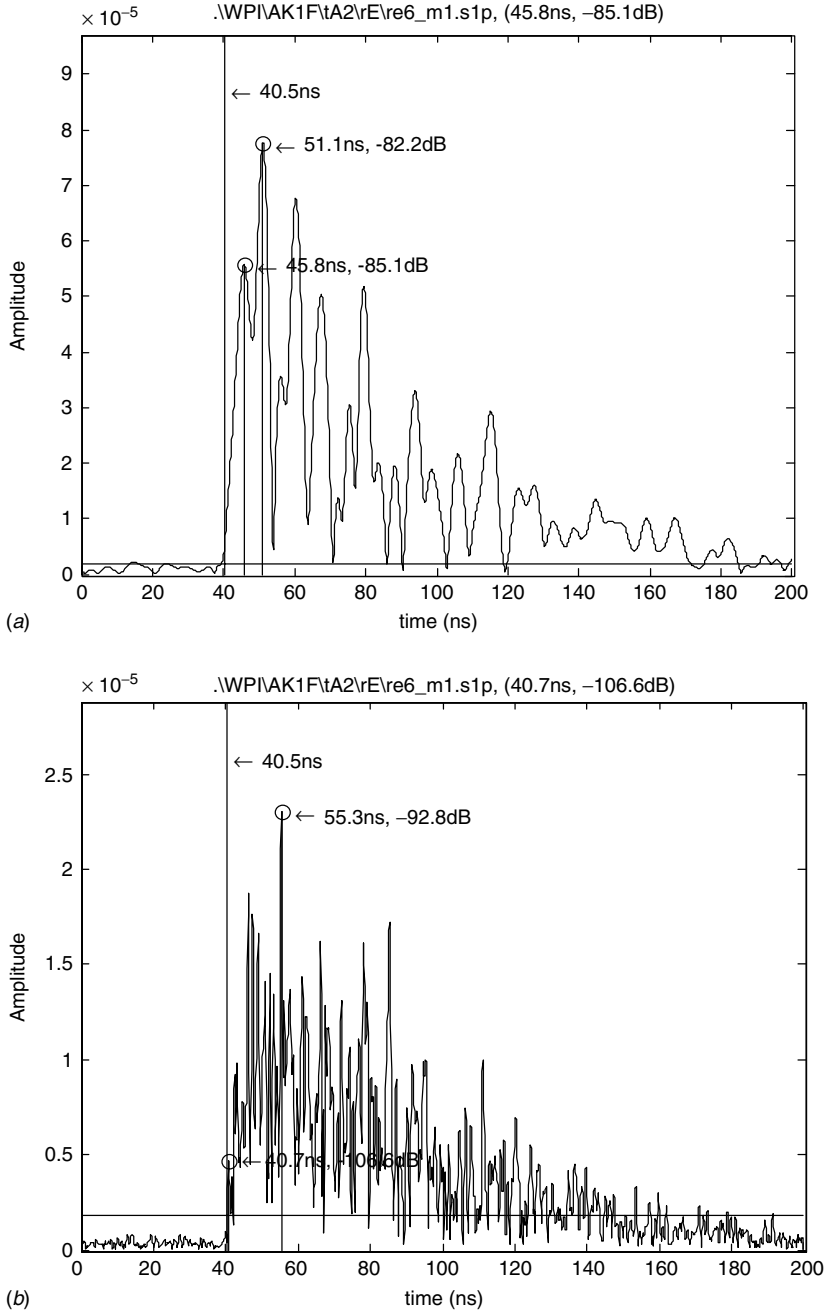


FIGURE 13.5 Typical UWB channel profile for bandwidths of (a) 500 MHz and (b) 3 GHz.

are 45.5 and 40.7 ns, respectively. The 5- and 0.2-ns errors in TOA estimates result in 1.67- and 7-cm errors, respectively, clearly illustrating the higher precision of the 3-GHz bandwidth. A closer look at the figure reveals that the amplitude of the first path in the 500-MHz band is larger than that of 3 GHz of bandwidth. This is due to

the fact that with the 500-MHz band, several of the paths in the 3-GHz band combine to form one path; therefore, the amplitudes of the paths in 500 MHz are statistically larger than those of the 3-GHz band received signal. In other words, one expects that coverage of the 500-MHz system will be larger than that of the 3-GHz UWB system because its paths are stronger. A natural question arises at this point: What happens if the first path is not detected? We refer to these cases as *undetected direct path* (UDP), and we treat them in the next section.

Undetected Direct Path in TOA Estimation. As a mobile terminal moves away from a base station, the strength of the direct path (DP) and the total received signal power decay exponentially. In an OLOS environment, when the DP goes below the threshold and still other paths are detectable, the receiver assumes the first path in the profile to be the DP, and this mistake causes a substantial error in wideband TOA measurements. We refer to this situation as the undetected direct path condition [Pah98]. Figure 13.6 shows the occurrence of the UDP scenario from the results of ray tracing for a transmitted pulse with a bandwidth of 200 MHz. Since the difference between the strength of the strongest path and the first path is more than the dynamic range (the range of detectable signal level below the strongest path) of the receiver, we have a clear UDP in which the first path is detected and declared as the DP, resulting in a 5.23-m measurement error. Figure 13.7 illustrates two cases of UDP measurements for UWB measurements using 500-MHz and 3-GHz bandwidths (this figure should be compared with Fig. 13.5 for DDP). We have errors of 13.5 ns (3 m) and 20.4 ns (6.8 m) for 500-MHz and 3-GHz bandwidths, respectively. As we mentioned earlier, with the 500-MHz band, several paths for 3-GHz bands combine and the overall profiles have larger path strengths. For the example in Fig. 13.7, some of the early paths for the 3-GHz bandwidth (Fig. 13.7*b*) that were under the threshold have been combined in the 500-MHz band (Fig. 13.7*a*), and the resulting paths have crossed the threshold, resulting in a smaller TOA measurement error.

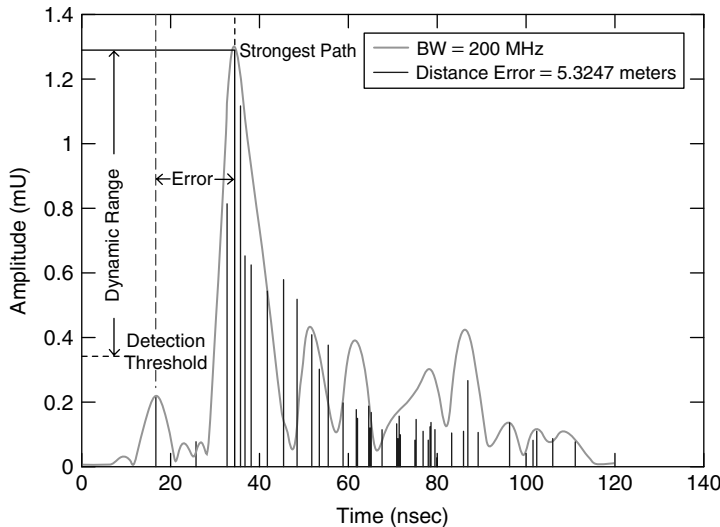


FIGURE 13.6 UDP multipath condition from results of ray-tracing simulation and a channel profile with 200 MHz.

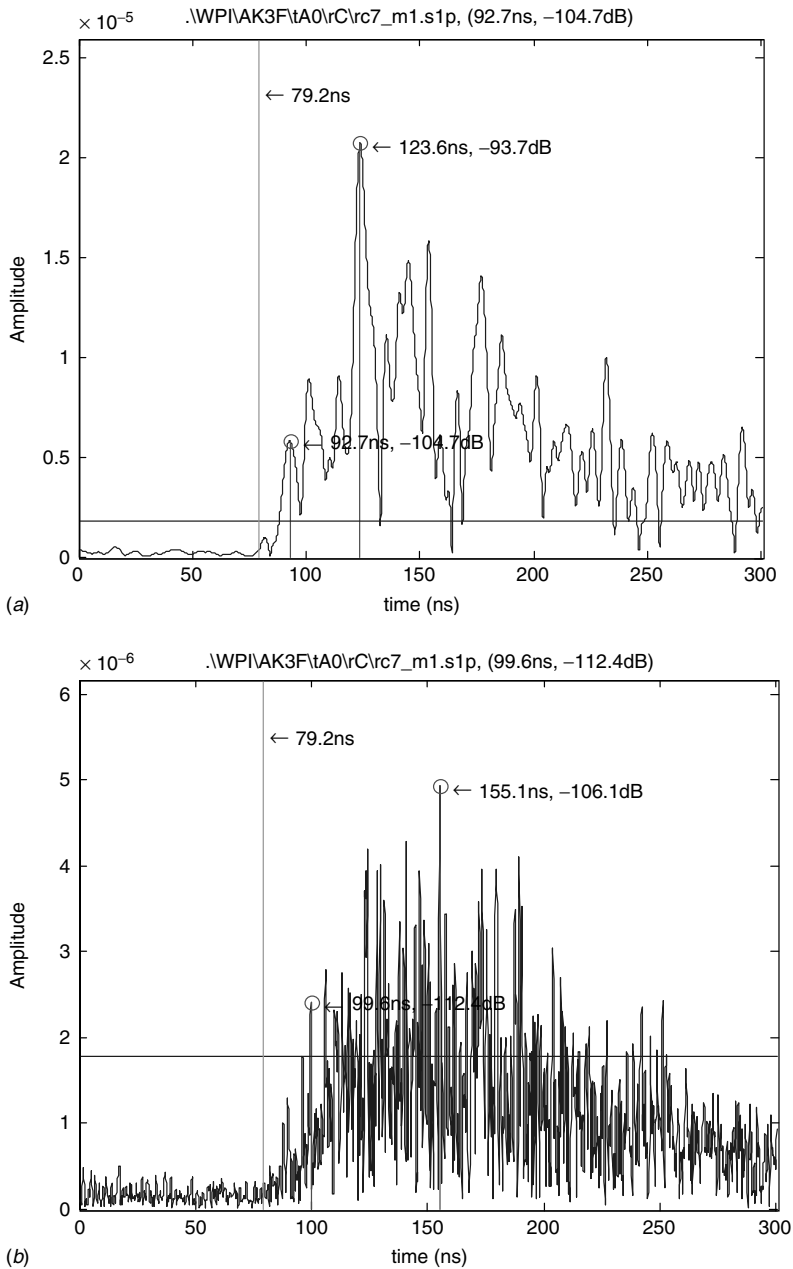


FIGURE 13.7 Sample UDP for UWB measurements for bandwidths of (a) 500 MHz and (b) 3 GHz.

Earlier in this chapter we stated that one of the limitations in the accuracy of the wideband TOA system is caused by restrictions in bandwidth. In narrower bandwidths, paths close to DP combine, resulting in a shift in the location of the estimated DP (the first peak of the profile). This problem is resolved when we increase the

bandwidth. In this section we showed that UDP conditions cause large errors in wideband TOA measurement, which at certain points increase when we increase the bandwidth. With the release of UWB bands we have adequate bandwidth for very accurate distance measurement, and the main challenge for the implementation of accurate wideband TOA systems is to find a remedy for UDP conditions.

13.2.2 RSS Techniques Metrics

As the distance between the transmitter and the receiver increases, the RSS decreases according to the path-loss models described in Chapter 4. Therefore, in the same way that we use the TOA measurements to calculate the distance between a transmitter and a receiver, we can use the measured RSS to calculate the distance. Of course, measuring the RSS is much simpler than measurement of the TOA. In cellular and WLAN networks the RSS is calculated for power control and handoff applications, and it is already available to the system without any change to the terminals or infrastructure of the system. As a result, RSS systems have attracted considerable attention for urban and indoor geolocation systems.

In TOA systems the relation between distance and TOA is unique and deterministic; therefore, the challenge is to find a method for accurate measurement of the TOA. Direct measurement of the distance from the RSS is unreliable because of the wide variety in statistical path-loss models and the large standard deviations in the errors associated with these models due to shadow-fading effects. To make RSS systems more reliable, we need to build certain intelligence in the system to recognize the area by previous calibration measurements using complex building-specific models such as ray tracing or complex pattern recognition algorithms for location finding [Pah02a]. Therefore, the complexity of RSS-based systems is in the processing of unreliable RSS reports.

13.2.3 Direction-Based Techniques

Angle-of-arrival (AOA) metrics can be used with simple location finding with a minimum of two reference points (RFs), as shown in Fig. 13.8. The location sensor measures the direction of received signals (i.e., angle of arrival) from the target transmitter using directional antennas or antenna arrays. If the accuracy of the direction measurement is $\pm\theta_s$, AOA measurement at the receiver will restrict the transmitter position around the line-of-sight (LOS) signal path with an angular spread of $2\theta_s$. AOA measurements at two reference point will provide a position fix as illustrated in Fig. 13.8. We can clearly observe that given the accuracy of AOA measurement, the

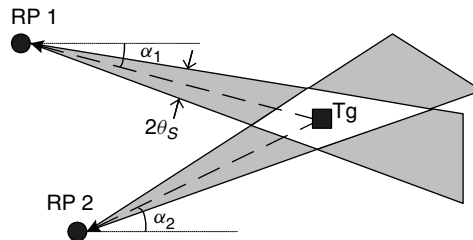


FIGURE 13.8 AOA technique for geolocation.

accuracy of the position estimation depends on the transmitter position with respect to the receivers. When the transmitter lies between the two receivers, AOA measurements will not be able to provide a position fix. As a result, more than two receivers are normally needed to improve the location accuracy. For macrocellular environment where the primary scatters are located around the transmitter and far away from the receivers, the AOA method can provide acceptable location accuracy. But dramatically large location errors will occur if the LOS signal path is blocked and the AOA of a reflected or scattered signal component is used for estimation. In an indoor environment, the LOS signal path is usually blocked by surrounding objects or walls. Thus, the AOA method will not be usable as the only metric for an indoor geolocation system. Although this is a feasible option in next-generation cellular systems, where smart antennas are expected to be deployed to increase capacity, it is in general not a good solution for low-cost applications, in particular in indoor areas.

13.3 MODELING THE BEHAVIOR OF RF SENSORS

As the number of applications for RF location sensing increases and in response to that a number of different technologies are developed, we need a framework for the comparative performance evaluation of these technologies. Any performance evaluation technique requires statistical models for the behavior of the sensors that describe the deviation of the measured metrics from the expected value if the sensor was operating under ideal conditions. These models are needed to relate the performance of a system to the density and deployment strategy of the sensors. The behavior of different sensors using TOA, RSS, AOA, or a signature of the delay power profile is quite different; therefore, we need to develop separate models for different metrics.

With extensive multipath conditions in indoor and urban areas, the overall channel impulse response for a given location of the transmitter and receiver is represented by

$$h(\theta, \tau, t, d) = \sum_{i=1}^L \beta_i^d(t) e^{j\phi_i^d(t)} \delta[t - \tau_i^d(t), \theta - \theta_i(t)] \quad (13.3.1)$$

where β, ϕ, θ , and τ represent the amplitude, phase, angle, and delay of arrival of each path traveling between the transmitter and the receiver, and d is the distance between them.

13.3.1 Models of the Behavior of AOA

RF sensors using AOA measure and estimate the angle of arrival of the direct path, $\hat{\theta}_1(t)$, for each reference point to calculate the location fix, as shown in Fig. 13.8. The only model addressing the AOA is Spencer's model, described in Section 6.2.5. A variation of this model is considered for the IEEE 802.11n wideband channel model. In this model, the arrival of clusters and of rays within the clusters is modeled as Poisson process and the angle of arrival for rays and clusters are modeled with Laplacian distribution. A novel measurement of the AOA and limited empirical data related to an indoor area is available in [Tin01]. As we mentioned earlier, since AOA in urban and indoor geolocation systems has not yet proven reasonable cost-effectiveness compared with the more popular RSS- and TOA-based systems, they have not attracted the amount of research work necessary for development of a model of channel behavior.

Further research in this area would be helpful to determine limitations on the performance of directional antennas for indoor and urban area location-finding applications. Existing models, such as [Spe00] or [Tin01], are developed and verified for communication applications. Further research is needed to verify the accuracy of these models for the analysis of the behavior and design of algorithms for AOA positioning systems.

13.3.2 Behavior of RSS Sensors

Received signal strength RF sensors process the received signal to determine the average RSS and use it to estimate, $\hat{d}$, the distance between the target object and the location sensor. The average RSS at a given distance is given by

$$\text{RSS}_d = \sum_{i=1}^L \overline{|\beta_i^d(t)|^2} \quad (13.3.2)$$

where β_i is the amplitude of an arriving path defined in Eq. (13.3.1). As we discussed in Chapter 4, measurement of the average RSS is independent of the bandwidth of the measurement device; therefore, the measured distance using RSS is independent of the bandwidth. In wideband measurements the effects of multipath fading are averaged over the spectrum of the signal, and we average that by measuring the strength of each arriving path and using it in Eq. (13.3.2). For narrowband systems, where we have only one arriving pulse with fluctuating amplitude according to the multipath fading characteristics, we need to average the signal over a longer period to make sure that the multipath fading is averaged out.

To calculate the distance between the transmitter and the receiver, we use the average RSS and a distance–power relationship to determine $\hat{d}$, the distance between the target object and the location sensor. If we define distance measurement error as the difference between the measured and actual values of distance, $\varepsilon_d = \hat{d} - d$, this error in RSS systems is independent of the bandwidth of the system. Measurement of the RSS is relatively simple and accurate, but the relation between the measured RSS and the distance is relatively complex and diversified. Therefore, the accuracy of RSS techniques depends on the accuracy of the model used for estimation of the RSS.

A number of statistical models for relating RSS to the distance between the transmitter and the receiver in indoor areas, developed for wireless communication applications, are presented in Chapter 4. These models can be used for analysis of the behavior of RSS sensors. As we discussed in Section 4.2, the common principle behind all statistical models for the calculation of RSS in a distance d is given by

$$\text{RSS}_d = 10 \log_{10} P_r = 10 \log_{10} P_t - 10\alpha \log_{10} d + X \quad (13.3.3)$$

where P_t is the transmitted power, d is the distance between the transmitter and the receiver, and α is the distance–power gradient of the environment. The random variable X is a lognormal distributed random variable representing shadow fading. Since a location sensor using RSS does not know the exact value of α , and the shadow-fading element X is a random variable anyway, distances calculated from these models are not very reliable. As a result, RSS sensors are either used in applications where accuracy is not a prime concern or are used with pattern-recognition algorithms that need substantial calibration measurements.

As we showed in Section 6.5, ray-tracing algorithms provide a much more reliable estimation of the received power by using the layout of the building. Therefore, they can be used to improve the performance of sensors using RSS. Ray-tracing algorithms are computationally extensive and alternatives to them are the geometrical statistical models. The advantage of geometrical statistical modeling and its ability to model site specificity while eliminating the complexity of ray-tracing computations are described in [Has98, Has02]. One of the pioneering applications of ray tracing for indoor positioning and intruder detection is published in [Hat04].

13.3.3 Behavior of TOA Sensors

Measurements of RSS are independent of bandwidth of the system and can be taken with simple receivers, but models relating RSS to distance are inaccurate and complex. In TOA systems, TOA measurement requires more complex receivers, measurement accuracy depends on system bandwidth. As we described in Section 13.2.1, TOA measurement suffers from UDP conditions for which large measurement error occurs even with UWB systems. However, the relation between the estimated TOA and the sensor is simple and deterministic. A TOA sensor estimates the distance from $\hat{d}_w = c\hat{\tau}_{1,w}$, where c is the speed of light and $\hat{\tau}_{1,w}$, given in Eq. (13.3.1), is an estimate of the TOA of the DP. As we described in Section 13.2.1, estimates of TOA are obtained by detecting the first peak of the received signal, and this value is a function of bandwidth and the occurrence of UDP conditions. The distance error is then defined as

$$\varepsilon_{d,w} = \hat{d}_w - d \quad (13.3.4)$$

where d is the actual distance between the sensor and target object and $\hat{d}_w = c\hat{\tau}_{1,w}$ is an estimate of the distance obtained from measurement of the first peak of the received channel profile for a given bandwidth.

To analyze the behavior of TOA-based positioning systems in severe multipath conditions, we need models for the distance measurements error. Similar to RSS positioning systems, the first natural solution that comes to mind is the use of existing multipath arrival models developed for telecommunication application. However, such multipath arrival models, described in Chapter 6, have not paid specific attention to the arrival time of the DP. In all these models the first path is assumed to be the DP, but this assumption neglects the existence of UDP conditions, which are one of the major causes of unpredicted large errors in TOA-based positioning systems.

Example 13.1: TOA Estimation and the Saleh–Valenzuela Model Figure 13.9 presents the distance measurement error using TOA estimates derived from the Saleh–Valenzuela model and actual measurements of the channel impulse response for 500-MHz and 3-GHz bandwidths in a small indoor area. At 3 GHz, Fig. 13.9a, the distance measurement error using the model is always zero because pulses are narrow enough to isolate the first path from following paths and the model assumes that DP is the first path and that it always exists. However, due to UDP conditions, the actual measurement error may be as high as 6 m and differs significantly from the results of simulation. Figure 13.9b represents similar results for a bandwidth of 500 MHz. With a 500-MHz bandwidth, paths close to the first path combine with that path and result in a shift in the peak detected and in a distance measurement error. However, the UDP conditions remain unaccounted for, and again the results of modeling and empirical measurements do not fit.

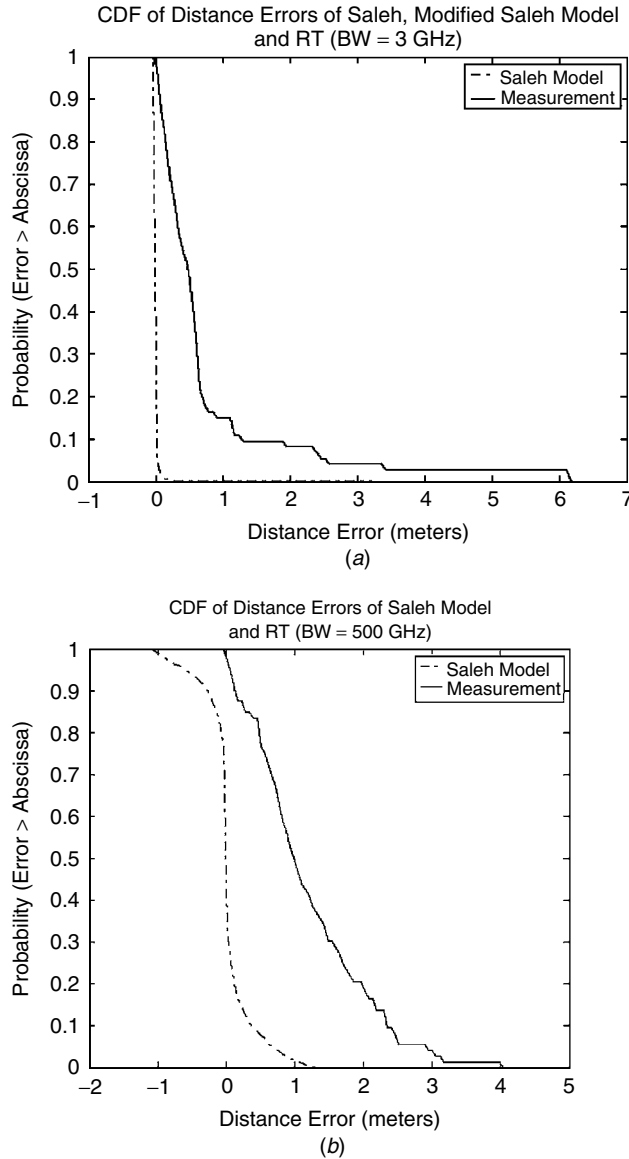


FIGURE 13.9 Comparison of the CDF of distance measurement error simulated by the Saleh–Valenzuela model and the empirical data for bandwidths of (a) 3 GHz and (b) 500 MHz.

The Saleh–Valenzuela model assumes random Poisson arrival, which allows arrival of paths very close to the first path; other models, such as GSM or JTC, provide even more optimistic results for TOA measurements. Therefore, there is a need for new models for geolocation in severe multipath conditions in indoor and urban areas. There are two approaches to this problem. The simpler approach is to model distance measurement error directly. A more challenging approach is to design a multipath arrival model that fits both communication and positioning applications. We provide two examples of these approaches.

Example 13.2: Direct Modeling of Distance Measurement Errors The first statistical model for the behavior of TOA sensors in indoor area is a primitive model reported in [Ala03a, Ala03b]. This model uses the results of measurement-calibrated ray-tracing software to develop a statistical model for the distance measurement error and to relate that to the bandwidth of the sensor. Since the behavior is distinctly different in the LOS and OLOS areas, two separate models are developed for the two scenarios. Since the error is proportional to the distance, this model normalizes the distance error with the actual distance:

$$\gamma_w = \frac{\varepsilon_{d,w}}{d} \quad (13.3.5)$$

In this way, representation of the estimated distance is represented by

$$\hat{d}_w = d(1 + \gamma_w) \quad (13.3.6)$$

and the modeling of the error reduces to the modeling of the statistical properties of the normalized distance measurement error as a function of bandwidth.

This model divides the environment into line-of-sight (LOS) and obstructed LOS (OLOS) areas, and for a TOA sensor with bandwidth w determines the distribution function of $\gamma_{L,w}$ and $\gamma_{O,w}$, the normalized distance error in the LOS and OLOS environments, respectively. The distribution function of the normalized distance error for the LOS, where DP is often the strongest path, is modeled with a Gaussian distribution:

$$f(\gamma_{L,w}) = \frac{1}{\sqrt{2\pi}\sigma_{L,w}} e^{-\gamma_{L,w}^2/2\sigma_{L,w}^2} \quad (13.3.7)$$

In OLOS areas, the error distribution is modeled as a combination of a Gaussian and an exponential distribution, given by

$$f(\gamma_{O,w}) = W_G \frac{1}{\sqrt{2\pi}\sigma_{O,w}} e^{-\gamma_{O,w}^2/2\sigma_{O,w}^2} + W_{\text{exp}} \lambda e^{-\lambda\gamma_{O,w}} u(\gamma_{O,w}) \quad (13.3.8)$$

Where $\lambda = 5.3$, $W_G = 0.79$, and $W_{\text{exp}} = 0.21$, and $\sigma_{L,w}$ and $\sigma_{O,w}$ are determined from

$$\sigma_{L,w} = A_L \left(\frac{1}{w} - m_L \right)^2 + B_L \quad (13.3.9a)$$

and

$$\sigma_{O,w} = A_O \left(\frac{1}{w} - m_O \right)^2 + B_O \quad (13.3.9b)$$

where $A_L = 52691$, $A_O = 9052$, $B_L = 0.43$, $B_O = 2.6$, $m_L = 0.0001$, and $m_O = 0.16$ [Ala03a, Ala03b]. We assume that the cause of additional exponentially distributed error in the OLOS scenarios is the existence of the UDP phenomenon described earlier. Figure 13.10 shows the fitness of the results of modeling with the actual errors measured in the building using the complementary CDF functions of distributions. Very large values of error in OLOS areas, even with ultrawideband and high SNRs, illustrate the harmful effects of UDP conditions. Comparing Fig. 13.10 with Fig. 13.9 reveals the significance of the model developed in [Ala03a, Ala03b].

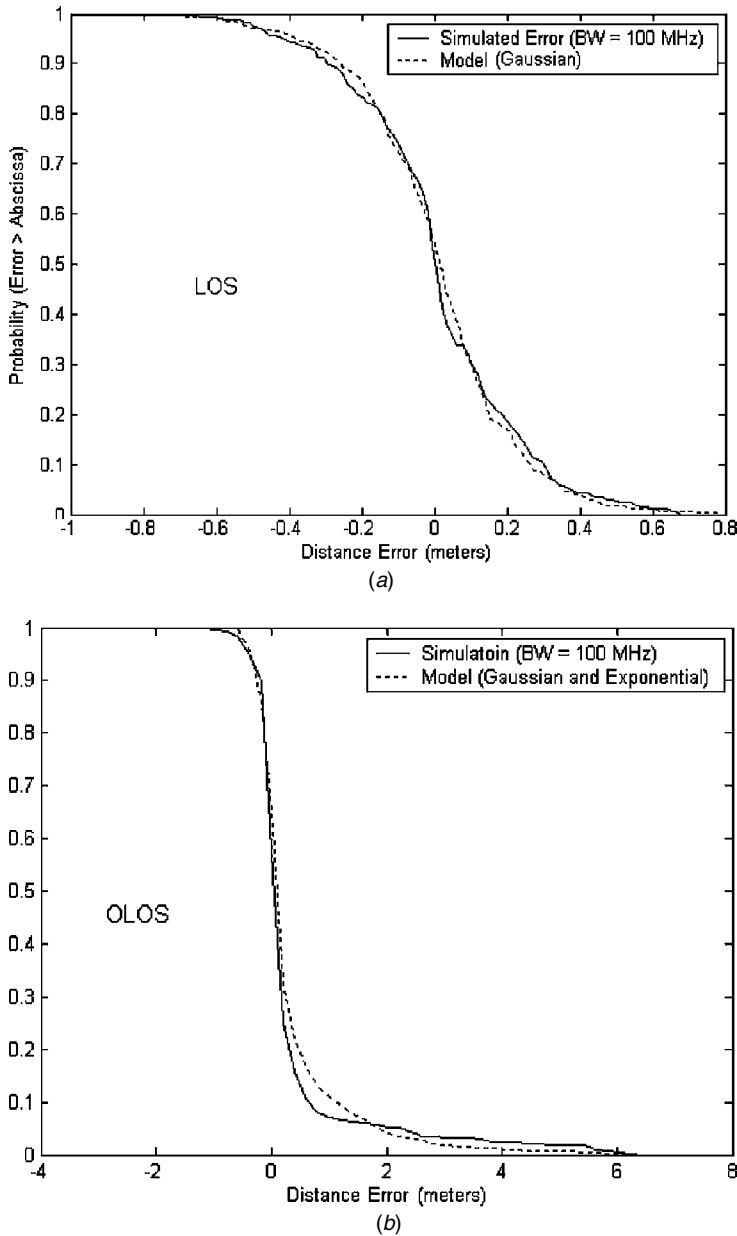


FIGURE 13.10 Comparison between complement CDFs of simulated and model results: (a) LOS; (b) OLOS [Ala03a, Ala03b].

In Chapter 5 we showed a number of multipath channel models for telecommunication applications. As shown in Fig. 13.11, in these models the channel impulse response is modeled as a few paths, and each path has a specific power and a random distribution function such as Rayleigh or Rician. The amplitude and distribution functions of variations of these multipath components are designed to ensure that the rms multipath spread of the model fits with the rms multipath spread of the empirical

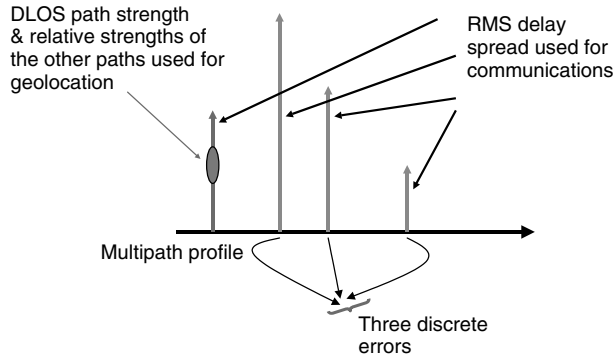


FIGURE 13.11 Multipath modeling requirements for communication and geolocation applications.

data collected in a particular environment. Multipath models designed for geolocation applications need to emphasize the distance measurement error caused by UDP conditions and limitations on the bandwidth. Assuming adequate bandwidth to isolate all paths, the only source of error is the UDP conditions that occur when the DP is overshadowed by the strongest path. Consider the four-path channel model shown in Fig. 13.11 and imagine that all paths are random variables changing their values. As the first path takes small values below the acceptable threshold of detection, one of the other three paths is detected as the first path, causing a distance measurement error proportional to the difference between the arrival time of the detected path and the expected arrival time of the DP. Having three paths be mistaken by the first path, we can have three discrete values for the distance measurement error. This situation does not reflect the real empirical observation, in which distance measurement error can take any value. A more reasonable approach to solving this problem is to assume random arrivals for the paths. However, as shown in Fig. 13.9, existing models such as Saleh–Valenzuela do not provide acceptable results. In fact, there is no reason for existing multipath models to produce good results for geolocation applications because they are designed to produce good statistics for the rms delay spread, not the distance measurement error caused by errors in calculating the TOA of the DP. Therefore, under ideal circumstances, it is desirable to design a multipath profile that produces a good fit for rms delay spread and distance measurement errors. The following example provides a simple two-path model to address this issue.

Example 13.3: Integrated Modeling of the Multipath Profile Figure 13.12a shows a simple two-path model that works for both communication- and TOA-based geolocation applications. Using an iterative approach, this model has fitted the results to the statistics of the rms delay spread for communication applications and the distance measurement error for geolocation applications to the results of extensive ray tracing in the Atwater Kent Laboratories at Worcester Polytechnic Institute. The amplitudes of the two paths are assumed to be Rayleigh distributed, and the distance between the two paths is assumed to have an exponential distribution. The average amplitude of the first path and the average arrival rate of the second path are assumed fixed, and the average amplitude of the second path is adjusted to fit the distance measurement error curves while maintaining an average rms delay spread. Figure 13.12b compares

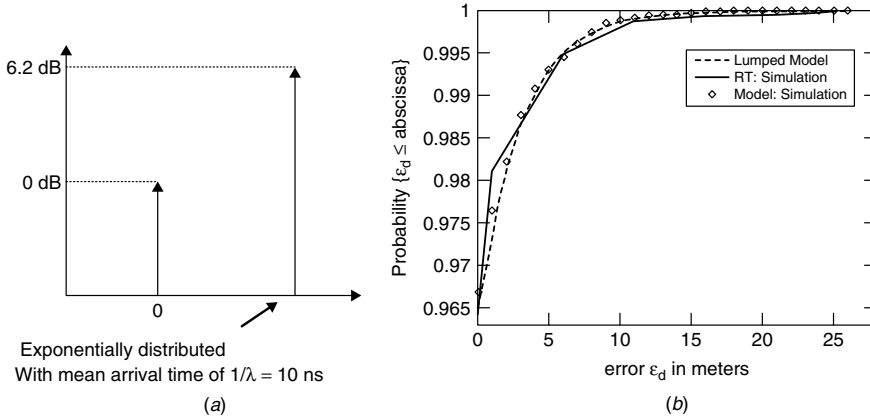


FIGURE 13.12 (a) Simple two-path model for integrate communication and geolocation application. (b) CDF of error from the model and ray-tracing simulations [Kri99a,b].

the results of simulation and analysis with the results of actual ray tracing used for calculation of the model parameters.

13.4 WIRELESS POSITIONING ALGORITHMS

Algorithms with well-defined properties are available for satellite-based GPS systems. There are least-squares algorithms and maximum-likelihood algorithms; there are algorithms based on a single snapshot of the measurements and those using measurement history. There are various kinds of sequential filters, including formulations that adaptively estimate some unknown parameters of the noise processes [Kap96, Mis02]. The GPSs, in particular, have focused a great deal of attention on positioning algorithms based on TOA, with considerable success. GPSs can provide positioning accuracy ranging from tens of meters to centimeters in real time, depending on a user's resources [Mis02]. In essence, these techniques are readily applicable to indoor location-sensing systems. However, indoor location sensing involves quasistationary applications and a number of unreliable reference points for which existing GPS algorithms, designed for mobile systems with a few reliable reference points, do not provide the optimum solution.

The measurement accuracy of location metrics in indoor and urban areas depends on the location-sensing technology and harshness of multipath conditions in the environment. Due to imperfect implementation of location-sensing techniques, lack of bandwidth, and the complexity of the multipath indoor radio propagation channel, there are always varying errors associated with measurements of location metrics. To achieve high positional accuracy when the measurements of location metrics are unreliable, the errors encountered in the measurement process have to be mitigated in the positioning process. In the next two subsections we discuss traditional positioning algorithms used with reliable measurements of location metrics and more intelligent pattern recognition techniques that can be used to improve positioning performance when the measurements of location metrics are unreliable.

13.4.1 Traditional Techniques

As we discussed earlier, TOA and RSS metrics are the most popular in wireless positioning systems for urban and indoor areas. TOA metrics provides for a more accurate measure of distance but may need additional infrastructure. The RSS is an easier metrics to measure and integrates well with the existing infrastructure, but it is less reliable and often needs more complex algorithms and additional calibration procedure. Positioning algorithms using AOA metrics are often simple and use the principles shown in Fig. 13.8. Therefore, our discussion of the algorithms will emphasize TOA and RSS systems.

With reliable TOA-based distance measurements using UWB technology, simple geometrical triangulation methods can be used to find the location of a tag or a mobile terminal [Fon01]. Due to estimation errors of distances at RP receivers caused by inaccurate TOA measurement, the geometrical triangulation technique can only provide a region of uncertainty, instead of a single-position fix, for the estimated location of the mobile terminal. To obtain an estimate of location coordinates in the presence of measurement errors of location metrics, a variety of direct and iterative statistical positioning algorithms have been developed to solve the problem by formulating it into a set of nonlinear iterative equations.

In some wireless geolocation applications, the purpose of the positioning systems is to provide visualization of the possible mobile locations instead of an estimate of the location coordinates. On the other hand, positional accuracy is not constant across the area of coverage, and poor geometry of the relative positions of the mobile terminal and RP can lead to high geometric dilution of precision. The output of statistical methods is an estimate of mobile location coordinates, and changes in the shape of the region of uncertainty are not revealed by this method. When both region-of-uncertainty information and an estimate of the location are needed, both geometric and statistical triangulation algorithms are used [Pah02].

Example 13.4: LS and RGWH Algorithms for Indoor Geolocation In [Kan04a] the performance of least squares (LS) and residual weighted (RGWH) LS algorithms in a square area with four RPs in the corners of the room is evaluated. The LS algorithm is a simple traditional gradient algorithm used in basic GPS systems which iteratively minimizes the square of the error in estimation of the position. For two-dimensional applications, considered in this example, it needs a minimum of three RPs. With more RPs the algorithm is expected to provide more accurate positioning. The RGWH, originally designed for cellular positioning [Che99a, Che99b], is another version of the LS algorithm, which calculates all possible solutions for positioning, in our case all possible 3RF combinations and the 4RF solution. The final estimate of the location is made by weighted averaging of all the estimates. The weighting factor is the inverse of the residual error. Thus less reliable positioning with higher residual error is counted with less emphasis. Using a random-number generator, a uniformly distributed location in the room is selected. Using the distance measurement errors model provided in Example 13.2, the estimated distances from the four RPs when a TOA system is used are determined. These distances are then used by LS or RGWH algorithms to determine the estimated location of the terminal in the area. The difference between the estimated location and the actual randomly selected location is used as the positioning error. By repeating this experiment, the statistics of the positioning error associated

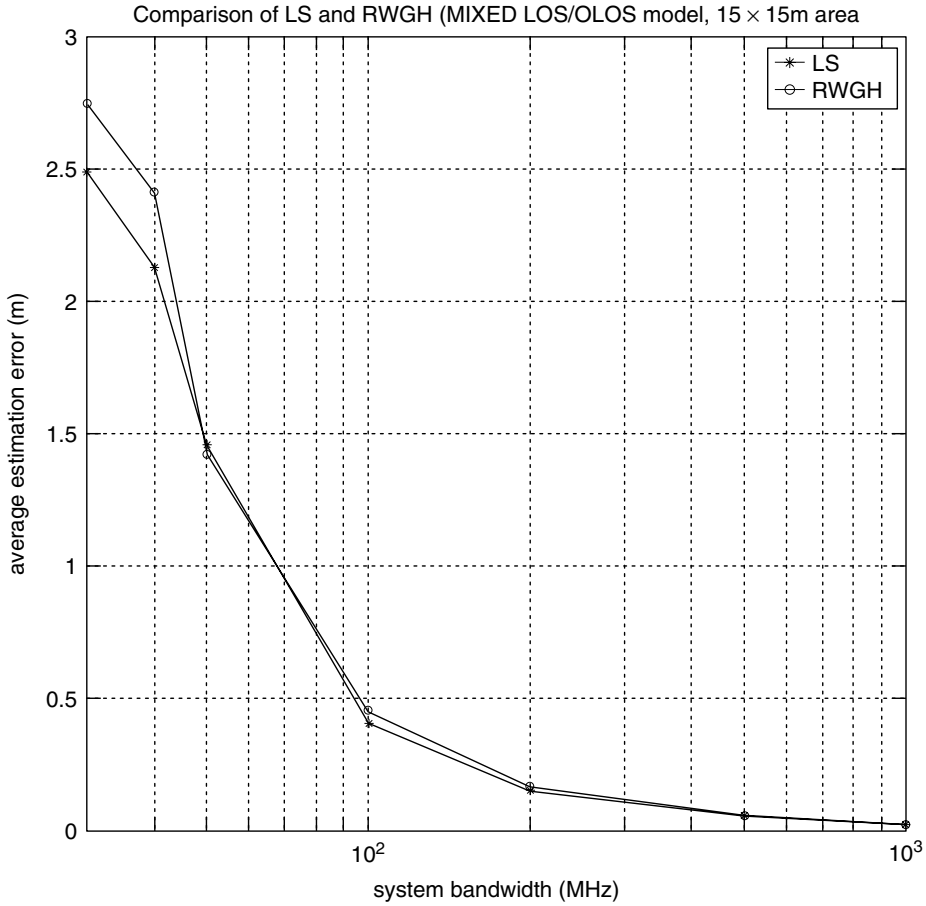


FIGURE 13.13 Average estimated positioning error versus bandwidth of a TOA system for the LS and RWGH algorithms in a 15- × 15-m room with mixed LOS and OLOS conditions [Kan04a].

with each of the algorithms is determined. Figure 13.13 shows a comparison of the performance of the two algorithms. RGWH provides a slightly better performance in lower bandwidths.

Example 13.5: CN-TOAG Algorithm for TOA Indoor Geolocation A closest-neighbor TOA grid (CN-TOAG) is presented in [Kan04b]. This algorithm takes advantage of the fact that for any given point in an area covered by a number of RPs, we know the exact value of the TOA expected from all the RPs. Consider a mobile location in an area and its measured distance from RFs to be d_i ($1 \leq i \leq N$), where N is the number of RPs that the mobile can see. Let $\mathbf{D}$ represent the vector of TOA range measurements from RPs, and let $\mathbf{Z}$ represent the vector of the TOA-based range measurements expected at a certain point, $\mathbf{r} = (x, y)$, in that area. We call $\mathbf{Z}$ the *range signature* associated with point $\mathbf{r}$. An estimate of the user's location, $\hat{\mathbf{r}}$, can be obtained by finding that point $\mathbf{r}$ where $\mathbf{Z}$ most closely approximates $\mathbf{D}$. If we define an error

function, $e(\mathbf{r}) = e(x, y)$, as

$$e(\mathbf{r}) = e(x, y) = \|\mathbf{D} - \mathbf{Z}(\mathbf{r})\| = \|\mathbf{D} - \mathbf{Z}(x, y)\| \quad (13.4.1)$$

where $\|\cdot\|$ represents the vector norm, the estimated location of the mobile, $\hat{\mathbf{r}}$, can then be obtained by finding the point (x, y) that minimizes Eq. (13.4.1). This point can be found by using the gradient relation,

$$\nabla e(x, y) = \mathbf{0} \quad (13.4.2)$$

In other words,

$$\hat{\mathbf{r}} = \text{argmin}(\|\mathbf{D} - \mathbf{Z}(\mathbf{r})\|) \quad (13.4.3)$$

where argmin refers to the value of the location coordinates that minimizes the error function. However, due to the complexity of the function in Eq. (13.4.1), it is not possible to find an analytical solution to this problem. The CN-TOAG algorithm solves this problem numerically. A grid of points in the area of the coverage of RPs is formed in which neighboring points are separated in each dimension with a distance h . For each point of the grid identified as (x_i, y_j) , a unique range signature, $\mathbf{Z}(x_i, y_j)$, is associated. It must be noted that the range signature associated with each grid point is exact, since it is based on straightforward geometric calculations. The CN-TOAG algorithm calculates

$$e_{ij} = e(x_i, y_j) = \|\mathbf{D} - \mathbf{Z}(x_i, y_j)\| \quad (13.4.4)$$

for every point in the grid and estimated location, $\hat{\mathbf{r}}_{ij} = (\hat{x}_i, \hat{y}_j)$, as the location with the minimum e_{ij} . As can be gathered from the preceding discussion, the granularity of the TOA grid, as embodied in the parameter h , is a major determinant of performance for this algorithm.

Figure 13.14 compares the CN-TOAG algorithm with the LS and RWGH algorithms in a $20\text{-m} \times 20\text{-m}$ room, using a system bandwidth of 500 MHz. The model for distance measurement error and the scenario for location of the RPs and the mobile are the same as in Example 13.4. As shown in this figure, CN-TOAG can achieve a better performance than the LS algorithm for our system scenario, provided that $h < 8.5$ m. We can see from Fig. 13.14 that whereas the LS algorithm has an RMSE of about 4.5 m, CN-TOAG has an RMSE value that can go down as much as 2.75 m (which implies a 38% improvement in estimation accuracy). We also observe from this figure that CN-TOAG performs better than the RWGH algorithm, provided that $h < 6.5$ m. Specifically, we see that the RMSE can go down from 3.14 m (for RWGH) to as much as 2.75 m in the case of CN-TOAG, reflecting a 12% improvement in performance.

13.4.2 Pattern Recognition Techniques

For indoor and urban geolocation applications, building floor plans and road maps are now accessible as electronic documents. The availability of electronic maps is one of the features that can be exploited in positioning algorithms. For example, while tracking a mobile terminal in buildings with the aid of the building floorplan, situations involving crossing walls or jumping through floors can easily be identified and eliminated. Another unique feature of having an electronic map is that we can make

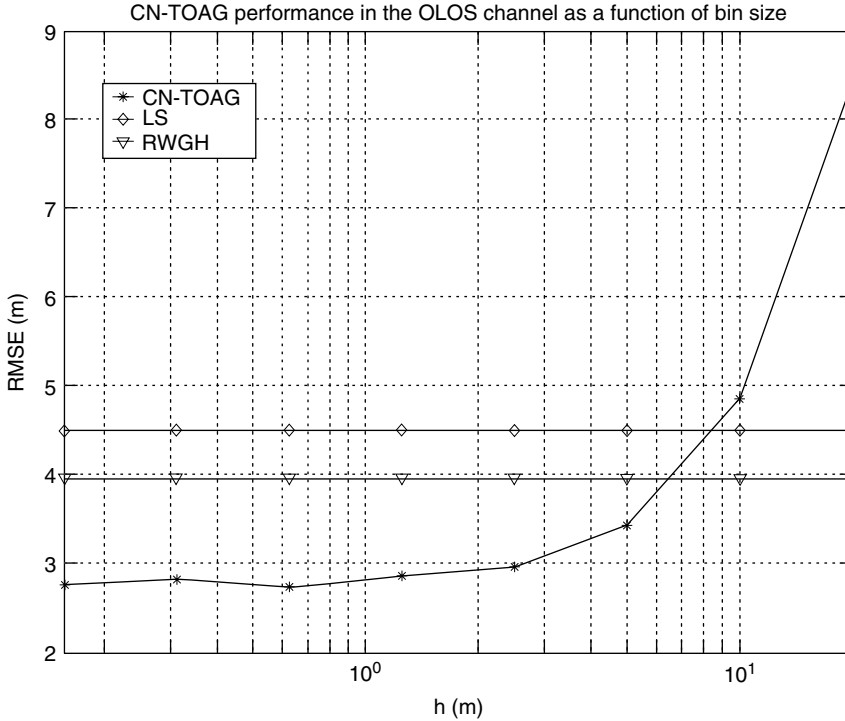


FIGURE 13.14 RMSE for CN-TOA with different grid sizes and its comparison with the LS and RWGH algorithms [Kan04b].

pre-measurements that associate the observation of the metrics with the specific location to be used as a reference for future location estimations. Careful pre-measurements can significantly reduce measurement errors of location metrics caused by propagation conditions. As a result, the pre-measurement-based location pattern recognition (also called *location fingerprinting*) technique is gaining increasing attention for indoor applications. On the other hand, in many of indoor and urban positioning applications, such as finding needed equipment or locating patients in critical condition, a mobile terminal is used in a quasi-stationary situation. For these situations, pattern recognition algorithms work better than traditional triangulation and Kalman filter-based tracking techniques.

The basic operation of pattern recognition positioning algorithms is simple. Each building and urban area is unique in its signal propagation characteristics; each spot would have a unique signature in terms of RSS, TOA, and/or AOA observed from different location sensors in the area. A pattern recognition system determines the unique pattern features or location signature of the area of interest in a training process, and this knowledge is used to develop rules for recognition. The metrics associated with the features could be extracted from actual measurements or by using models that accurately predict the expected values of the metrics. The challenge for such algorithms is to distinguish locations with a similar signature and the computational complexity for storage and processing of the database when we are dealing with a large area. To build a signature database, a terminal is carried through the service

area to collect the desired location-sensing metrics from all location-sensing elements. The service area is divided into nonoverlapping zones or grids, and the algorithm analyzes the received signal patterns and compiles a unique signature for each zone. For quasistationary applications, the simplest method for pattern recognition is the nearest-neighbor process. In this method, the Euclidean distance is calculated between the measured metrics, RSS, TOA, or AOA, and all entities in the signature database. The location estimate is determined to be the one associated with the minimum Euclidean distance [Bah00a, Bah00b]. The following example provides an experiment similar to the one in [Bah00a, Bah00b] but in a different building.

Example 13.6: Performance of the Nearest-Neighbor Method Figure 13.15a presents a partial layout of the Telecommunications Laboratory and Centre for Wireless Communications at the University of Oulu in Finland. The locations of four 802.11b access points (APs) and 31 measurement locations along a long corridor, with about 2 m of separation between adjacent points, are illustrated. A mobile terminal is carried along the corridor and the RSS are measured at each location. Figure 13.15b shows the measured RSS values at all four APs as the terminal travels from the right corner close to AP-I to the end of the vertical corridor after AP-IV. When the nearest-neighbor pattern recognition method is applied to the measurement data, the standard deviation of the positioning error is 2.4 m, and at about 80% of locations, the positional error is less than 3 m.

The nearest- or closest-neighbor algorithm is more popular for RSS-based systems. However, as we showed in Example 13.5, it is also applicable to TOA systems in which we do not need to measure the metrics at each reference point because the expected

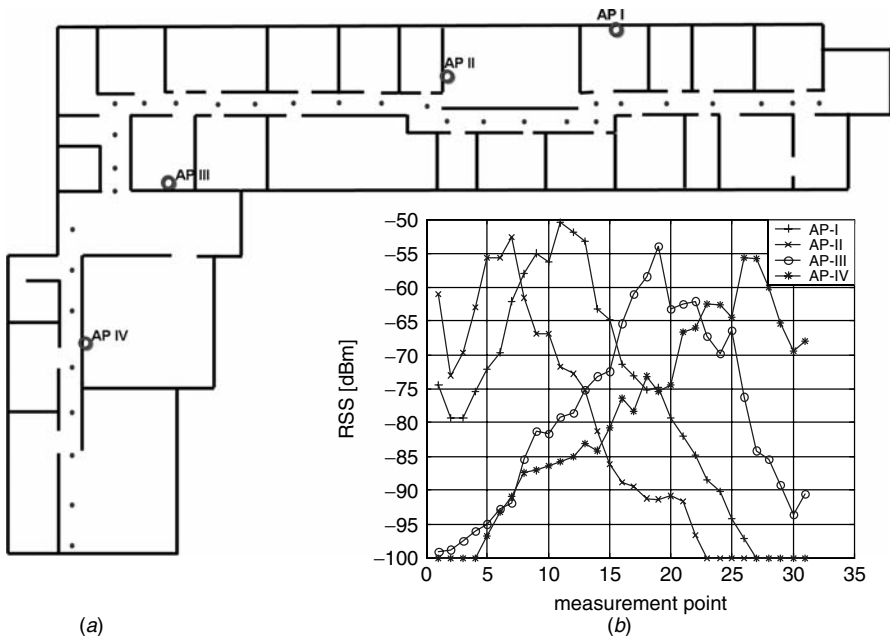


FIGURE 13.15 (a) Building layout and location of WLAN APs at the first floor of the CWC Laboratory at the University of Oulu, Finland. (b) RSS signature in various locations [Pah02b].

value of TOA is calculated from the distance between the RP and the mobile. It is also possible to calculate RSS from the distance between the RP and the mobile by using path-loss models and avoid time-consuming measurements. Since path-loss models relating distance to RSS are statistical, RSS metrics are not reliable and RSS measurement has proven to provide a more practical solution [Bah00a, Bah00b]. This argument leads to the point that one may think of using statistical models with measurements to improve performance. This approach can reduce the number of measurements or increase the accuracy with the same number of measurements. In this approach, referred to as a *statistical approach*, we assume that RSS in any training location (x_i, y_i) is a random variable p with a density function $f[p/(x_i, y_i)]$ that is drawn from the samples of measurements performed at that location with certain assumptions. Then for a new measurement, $\text{RSS} = P$ at an unknown location (x, y) , we can calculate the probability of having that measurement taken in the specific training points (x_i, y_i) :

$$p_i = f[p/(x_i, y_i)]$$

If we have N training points associated with every measured RSS at an unknown location, we will have N locations (x_i, y_i) and N probabilities p_i for the unknown location to be one of the training points. Then our actual estimate of the unknown location (x, y) is the statistical average over all measured training locations:

$$(\hat{x}, \hat{y}) = \sum_{i=1}^N p_i(x_i, y_i) \quad (13.4.6)$$

Different approaches can be used to define $f[p/(x_i, y_i)]$, the distribution function of the RSS around any measured training point; therefore, various statistical approaches can be realized.

Example 13.7: Statistical Approach to Positioning Pioneering work using a statistical approach for indoor positioning is reported in [Roo02a]. In this work, two approaches, the kernel and the histogram methods, were applied to define the distribution function. In the *kernel method* it is assumed that the RSS at each measurement forms a Gaussian distribution. Therefore, with L access points and M measurements in each location (x_i, y_i) , the probability density function of the RSS from all access points is given by

$$f[\bar{p}/(x_i, y_i)] = \frac{1}{M} \sum_{j=1}^M G_L(\bar{P}_j, \sigma) \quad (13.4.7)$$

where $\bar{p}$ is an L -dimensional random variable representing RSS measurements from L access points around location (x_i, y_i) , $\bar{P}_j$ is the L -dimensional vector of the j th measurement of the RSS at training point (x_i, y_i) from all access points, and $G_L(\bar{P}, \sigma)$ is an L -dimensional Gaussian random variable with mean $\bar{P}$ and variance σ . For any unknown location the $\bar{P}$ measured is placed in Eq. (13.4.7) to determine p_i , the probability that an unknown point is located at (x_i, y_i) . The results are then placed in Eq. (13.4.6) to estimate the location.

In the *histogram method*, multiple measurements from the j th access point in one location are used to form a piecewise linear probability density $f_j[p_j/(x_i, y_i)]$. Then Eq. (13.4.5) is used to calculate p_{ij} , the probability that the location is (x_i, y_i) when

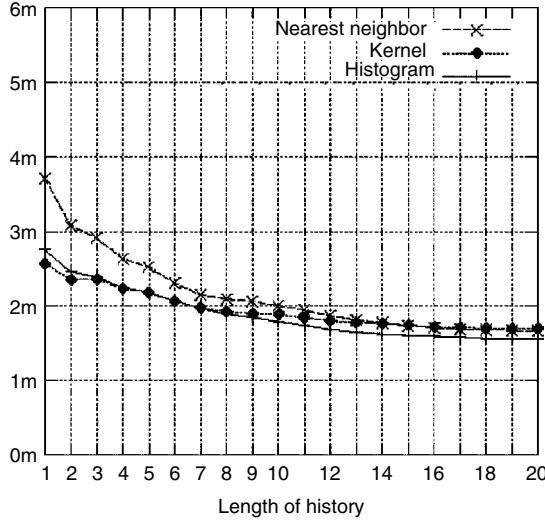


FIGURE 13.16 Comparison of the performance of nearest-neighbor and statistical (kernel and histogram) methods [Roo02a].

only the j th access point is considered. Since the RSS measurements from different access points are independent, the total probability

$$p_i = \prod_{j=1}^L p_{ij}$$

and the rest remains the same as before. Figure 13.16 provides a comparison of the nearest-neighbor, kernel, and histogram techniques. The test area is a 16×40 m office with concrete, wood, and glass structures with 10 access points (eight around the perimeter and two in the center). The training data comprise 155 points using a 2-m grid, and at each grid (calibration) point, 40 RSS observations are recorded. The test data are collected independently using a similar 2-m grid, selecting the test points to be as far as possible from the calibration points. At each of the 120 test points, 20 observations are gathered. The vertical axis is the average distance measurement error and the horizontal axis represents the number of measurements in each location. In general, all techniques provide good results, but statistical methods provide greater improvement with shorter histories.

The algorithms presented in Examples 13.6 and 13.7 do not take advantage of the channel modeling knowledge presented in Chapters 3 to 6. In principle, if we have a reliable channel model, we can avoid measurements and train the algorithms with predicted values of the RSS calculated from the channel model. We have statistical and building-specific models for the behavior of radio propagation. The statistical models described in Chapter 4 are less accurate and need extensive empirical measurements to calculate their parameters for a specific area. If we do extensive measurements to determine an accurate model, we might as well use these measurements directly in positioning applications. In Chapter 6, however, we showed that ray-tracing algorithms provide a close estimate of the RSS (see, e.g., Fig. 6.43). Ray-tracing algorithms require

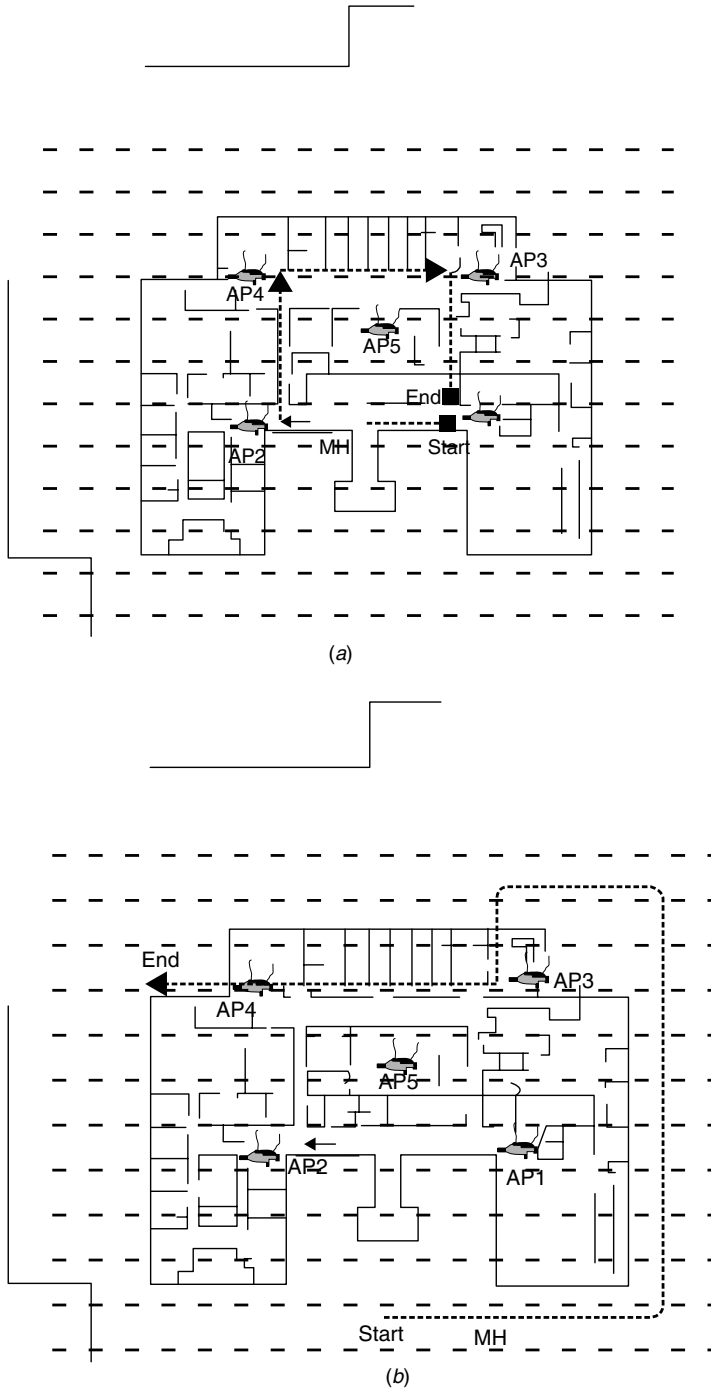


FIGURE 13.17 Layout of the first floor of AKL with outside walls from other buildings, location of the APs, and the grid used for ray tracing: (a) inside path used for positioning experiment; (b) outside path used for intruder detection.

an electronic map of the area, and in most wireless positioning applications we do have an electronic map to show the location of the terminal in the area. The following example provides some preliminary results in that area.

Example 13.8: Indoor Positioning with Ray Tracing In [Hat04], the two-dimensional ray-tracing software described in Section 6.5.1 is used for positioning and intruder detection experiments on the first floor of the Atwater Kent Laboratories (AKL) at Worcester Polytechnic Institute. Figure 13.17 is a layout of the building and surrounding walls, the paths used for positioning experiments (Fig. 13.17a), the path used for intruder detection experiments (Fig. 13.17b), the location of five access points, and the area covered by the grid. The ray-tracing software generates a vector of five RSSs from the access points (APs) for nodes of the grid. Using slight variations of the nearest-neighbor method, the location of the mobile host is calculated. For the positioning application, the estimated location is used to calculate the distance measurement error. The purpose of the intruder detection algorithm is to use the received signal strength from all APs to determine whether the mobile host is inside or outside the building. The intruder detection algorithm uses the estimated location to find out if it is inside or outside the building and compares that with its actual location to determine the statistics of intruder detection. Figure 13.18a shows the cumulative distribution function of the distance measurement error in the AKL building for various grid sizes. Compared with the results of Examples 13.6 and 13.7, the AKL building is larger; it has a lot of brick, concrete, and metal in its construction; and we have fewer APs. For the grid size of 1 m, in 60% of locations, distance measurement errors are less than 5 m. Figure 13.18b shows the effects of the number of APs. If we reduce the APs to two, the 60% locations' error rises from 5 m to about 13 m. Figure 13.19 shows the intruder detection probability in the path shown in Fig. 13.17b for a different number of APs. Although the results are not very consistent (e.g., sometimes we get larger errors with a smaller grid), the overall conclusion is that with a grid size of 1 m and three to five APs, we can detect the intruder in 90% of the attempts. These are results from static calculations; by adding more intelligence to the algorithm and adjusting the location of the APs, these statistics can be improved.

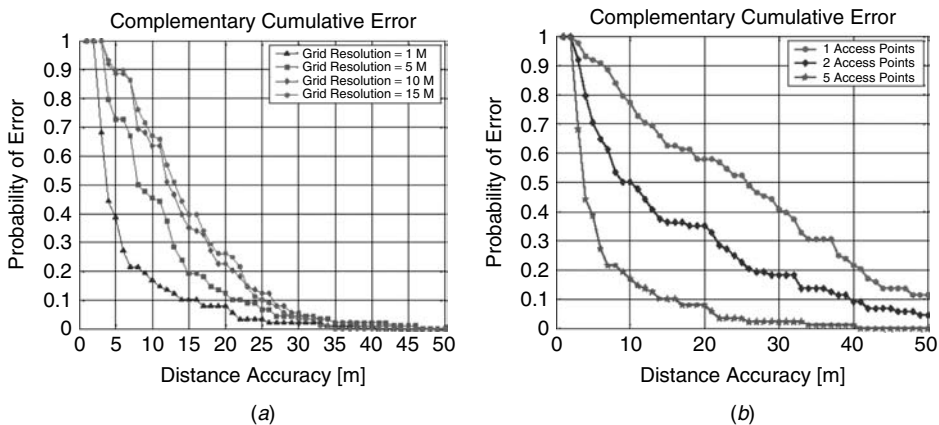


FIGURE 13.18 Statistics of the distance measurement errors for positioning using two-dimensional ray tracing: (a) effects of grid size; (b) effects of number of APs [Hat04].

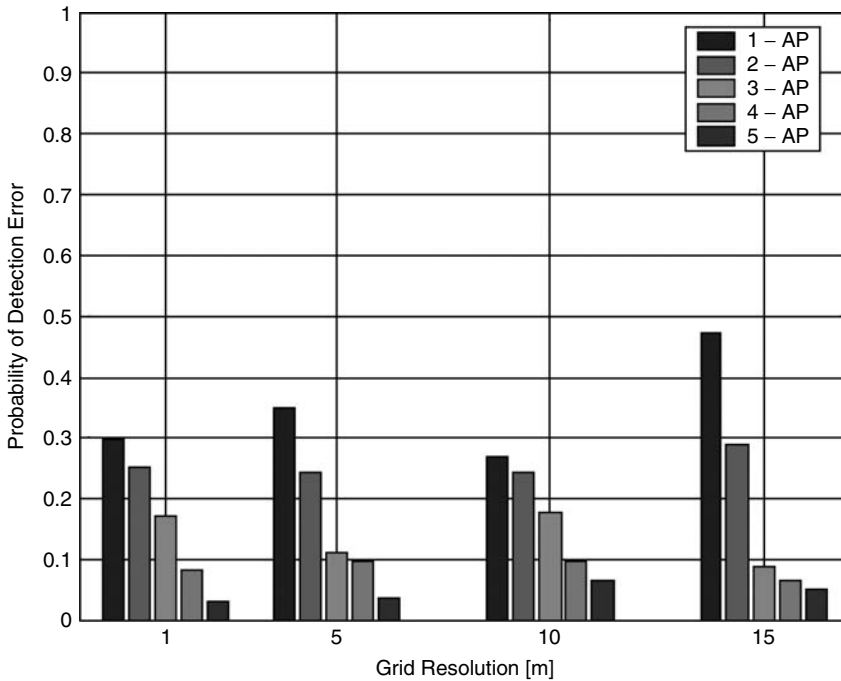


FIGURE 13.19 Statistics of the probability of intruder detection error for different numbers of APs and grid size [Hat04].

In diversity channels, rather than taking the total power in each location, we may measure the signature of the RSS in each diversity branch to add to the accuracy of the system. When these measurements are already available in the system, performance enhancement can be implemented at low cost. For example, in DSSS CDMA systems it is possible to use the measured time and signal strength of all fingers of the RAKE receiver in place of RSS to improve the positioning performance.

QUESTIONS

- What was the motivation for research activities in indoor and urban area positioning?
- What is the difference between E-911 services and GPS?
- What technologies are being considered for the implementation of E-911 services?
- What are the two most popular indoor positioning systems for commercial and military applications, and what is the difference between these technologies?
- Why is measurement of the distance based on the phase of the received carrier of a signal not reliable in extensive multipath conditions?
- Why is measurement of the distance based on angle of arrival of a signal not reliable in extensive multipath conditions?

- (g) Explain the difference between the standard use of spectral estimation algorithms for time-domain problems and the use of these algorithms for superresolution TOA estimation.
- (h) Explain why in Fig. 13.5 the peak of channel measurements at 500 MHz is higher than the peak of the signal in 3-GHz measurements.
- (i) What are the two primary causes of error in TOA measurements in multipath-rich indoor areas?
- (j) Explain why UDP conditions occur and how they cause large distance measurement errors.
- (k) Why are channel models developed for telecommunication applications not well suited for indoor geolocation applications?
- (l) Why does Saleh–Valenzuela’s model not provide a good model for distance measurement errors?
- (m) Why are weighted least-squares algorithms most effective in multipath-rich environments?
- (n) Compare the least-squares and pattern recognition algorithms used in wireless positioning in terms of need for training sequence, computational complexity of the algorithm, and accuracy of the results.
- (o) Why are RSS-based pattern recognition algorithms popular in the IEEE 802.11 WLAN community?
- (p) Why is TOA-based positioning popular in the IEEE 802.15.3 WPAN community?

PROBLEMS

1. Two base stations (BS1 and BS2) are located at (0,0) and (200, 200), and a mobile terminal is located at (150, 100). BS1 measures the angle of arrival with respect to the horizontal axis to be 45° , and BS2 measures the angle to be -135° . Determine the location of the terminal and the value of the distance measurement error.
2. Three access points are located at AP1, AP2, and AP3 are located at (0, 0), (50, 0), and (0, 50), and a mobile terminal (MT) is located at (20,10).
 - (a) What are the expected distances between the MT and each of the APs?
 - (b) Assume that the distances from the MT to BS1, BS2, and BS3 are measured as 18, 35, and 38. What are the three ranging errors?
 - (c) Draw three circles corresponding to the distances measured from the base stations and determine the location of the mobile terminal. What is the distance measurement error for this system?

14

WIRELESS OPTICAL NETWORKS

- 14.1 Introduction
- 14.2 Implementation
- 14.3 Eye Safety
- 14.4 IR Channel Characterization and Data-Rate Limitations
 - 14.4.1 Optical Propagation and Multipath Effects
 - 14.4.2 Effects of Ambient Light
 - 14.4.3 Photodiode Capacitance and Noise Enhancement
- 14.5 Modulation Techniques for Optical Communications
 - 14.5.1 Analog Modulation Techniques
 - 14.5.2 Pulse Modulation Techniques
 - 14.5.3 Digital Modulation Techniques
 - 14.5.4 Baseband PCM
- 14.6 Multiple Access and Data Rate Questions

14.1 INTRODUCTION

Optical communication technology has several features that are well suited to wireless indoor applications. Infrared (IR) transmitters and receivers can be built at relatively low cost and with the small size and low power consumption suitable for battery-supported operation. Optical modems can be built less expensively than radio-frequency (RF) equipment, at costs comparable to those for wired connections. IR transmissions will not interfere with existing RF systems, and IR systems do not fall under any regulations of the FCC. Because IR signals do not penetrate walls, these systems provide a considerable degree of privacy simply by confinement of transmissions within an office or other work area. The only way for IR signals to be detected outside the installation area is through windows, which can easily be covered with curtains or shades. The confinement of IR signals by walls also allows concurrent use of similar systems in neighboring offices without mutual interference. Therefore, in a cellular network architecture, all units can be identical—in contrast with RF systems, in which

the operating frequencies of neighboring cells must be different. In an IR network, terminals within a cell communicate with a node or “satellite” installed on the ceiling, and these nodes are interconnected to the rest of the network with wires, cables, radio, or fiber optics lines [Gfe82, Yen85]. The cell might be a small office or a section of a large open office, depending on the architecture of the building. Figure 14.1 shows two typical setups for optical networks. The first setup is a manufacturing floor divided into subareas covered by satellites. The second setup comprises three areas covered by separate satellites. The satellites in either example are connected by a wired backbone network to a central control station. Wireless optical networks are sensitive to shadowing, and therefore coverage will usually suffer to some extent from blind corners or areas with outage. In these instances, coverage can be extended by installing passive or active reflectors.

IR communication technology is dominant in the low-speed remote control market [Cia82] and is also finding applications in cordless phones, wireless keyboards for personal computers, and wireless local area networks (WLANs). The major drawbacks

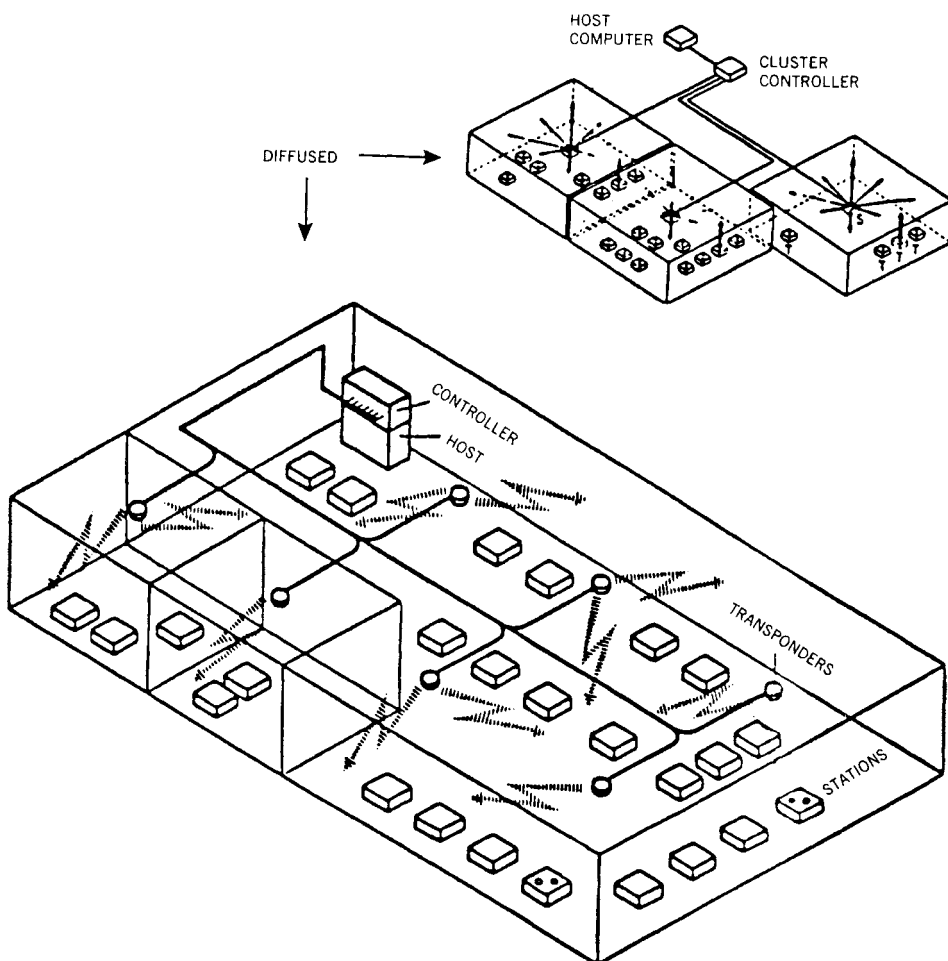


FIGURE 14.1 Typical setups for IR networks. (From [Gfe79a] © IEEE.)

in the use of IR technology in wireless networks are its data-rate limitations, extensive power fluctuation, and susceptibility to interference from ambient light. In particular, there is a lack of a simple and reliable multiple-access technique for IR systems capable of counteracting the effects of ambient light and objects moving near the transmitter or receiver.

14.2 IMPLEMENTATION

In the past two decades, most developments in wireless optical communications have concentrated on diffused IR radiation (DFIR) technology. Figure 14.2 shows a typical DFIR network configuration. The primary advantage of this mode of transmission is that it does not require a direct line of sight between the transmitter and the receiver [Gfe79b]. Instead, the receiver can collect a transmitted signal by way of reflections from walls, ceiling, and other objects in a work area. Therefore, network installation does not require precise alignment to establish a communication link, which facilitates the portability of user terminals. As a result, DFIR networks are well suited for applications requiring portability, such as cordless phones or networking of laptop or pen-pad computers.

The disadvantages of DFIR transmission are as follows:

1. It consumes relatively high power to cover the entire installation area.
2. The data rate is limited, due to the effects of multipath.
3. Diffusion of IR signals elevates the risk of eye exposure.
4. In simultaneous two-way communications, each receiver collects its own transmitted reflections, which are stronger than the transmitted signal from the other end of the connection.

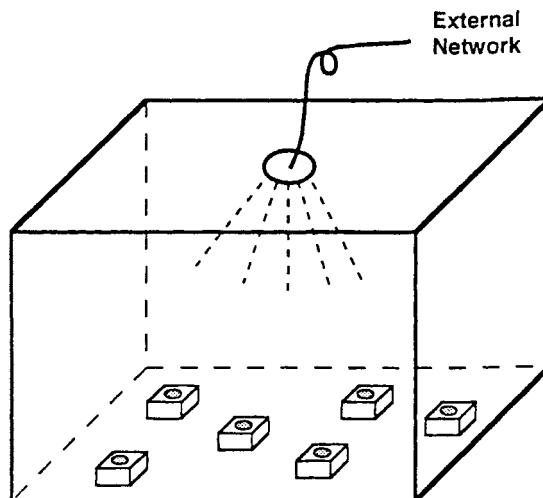


FIGURE 14.2 Diffuse infrared network configuration. Light is diffusely reflected and scattered in the room from a wide-angle source or a diffusing spot formed on a reflector. It is received by wide-acceptance-angle photodetector portable units located anywhere in the room. The advantages of this configuration are that there are no alignment requirements and that there is freedom of movement.

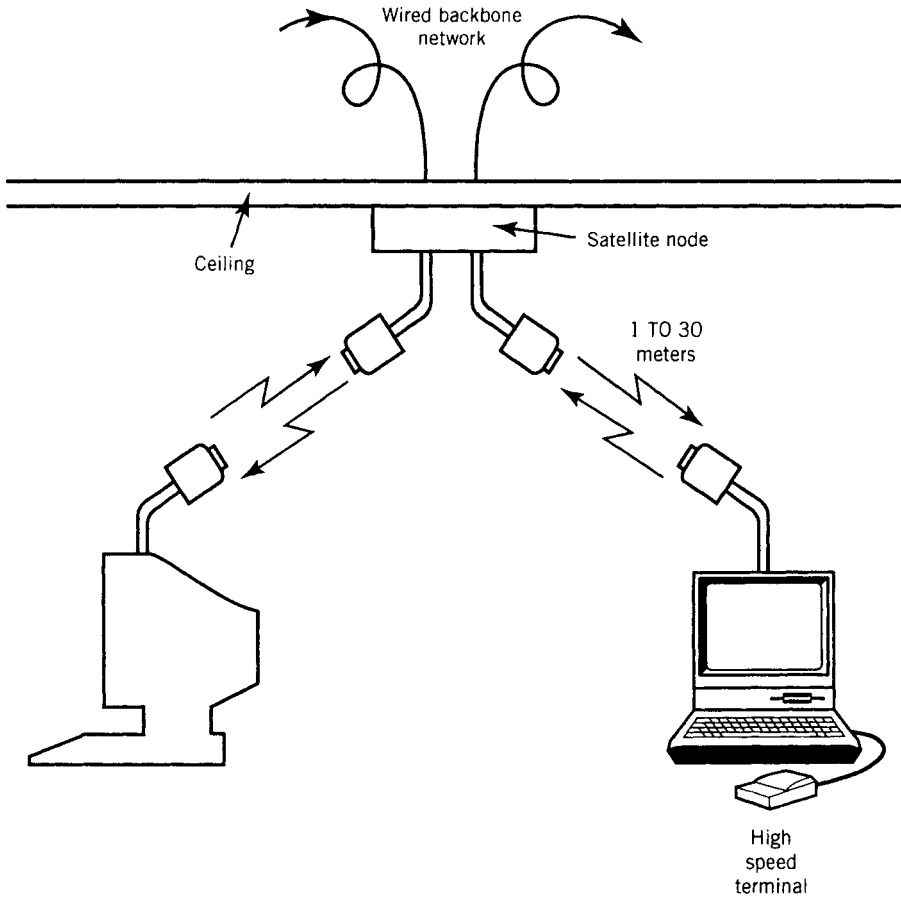


FIGURE 14.3 Directed-beam infrared network configuration. Two collimated IR beams connect a terminal to the network, one for the uplink and one for the downlink.

An alternative to diffused transmission is directed-beam IR (DBIR) communication, which has been investigated for application to wireless information networking [Yen85]. In the WLAN industry, this technology has not been as widely adopted as the DFIR technology, but some products using this technology have emerged in the market. Figure 14.3 shows a typical DBIR network. The transmitted radiation pattern is aimed in the direction of the receiver. The advantage of DBIR transmission is that it requires less optical power for communication, does not suffer from extensive multipath propagation, and can handle bidirectional communications better than can diffused radiation. As a result, higher data rates and larger areas can be covered using this method of transmission. The disadvantages are (1) the need for alignment of transmitters and receivers, and (2) signal interruption caused by shadowing. This transmission method is used in applications where the terminals are in fixed locations during the transmission interval. In typical installations, the transmitters are placed on high posts to avoid shadowing. In such installations, care must be taken to minimize direct eye exposure to the directed beams.

Semiconductor laser diodes and light-emitting diodes (LEDs) are used as the radiating elements in wireless optical networks. A semiconductor laser emits a narrow optical

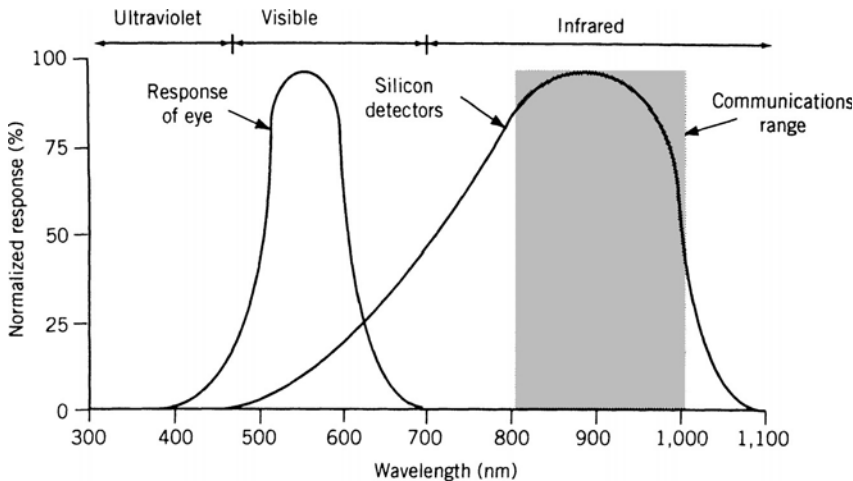


FIGURE 14.4 Responses of the human eye and of silicon detectors. The shaded portion of the figure shows the range of wavelengths for IR communication systems.

beam, and for diffused transmission applications a diffusing lens or some other optical device must be used to increase the coverage area. The more expensive semiconductor laser diodes have a more linear electrical-to-optical power conversion characteristic, and they provide higher radiated optical power. The radiating element plays a role similar to that of the last-stage power amplifier in a radio transmitter. Linearity of this element provides a more flexible environment for employing a variety of modulation techniques. For example, multi-amplitude or multifrequency modulation requires use of a linear transmitter. At the receiver, an avalanche photodiode (APD) or a *p*-intrinsic-*n* (PIN) photodiode is used to convert the received optical power to an electrical signal. These diodes are analogous to front-end RF circuits in radio modems. The more expensive APD has a higher internal gain, which supports the higher signal-to-noise ratios needed for wideband communication links. Figure 14.4 shows the response of the human eye and of silicon detectors and LEDs.

The choice of optical wavelength is another implementation issue. In addition to extremely inexpensive LEDs and PIN diodes, inexpensive gallium arsenide (GaAs) laser diodes and low-noise silicon APDs at 850 nm are available in the market. The decreased safety hazards of optical radiation at longer wavelengths allow higher transmitted power, which facilitates higher transmission rates. Disadvantages are that devices at these frequencies are more expensive and noisier than today's implementations. Today's IR technology operates primarily at about 900 nm, but the evolving technology includes 1.5- μm devices.

14.3 EYE SAFETY

The human eye acts similarly to a camera, focusing the energy density of the incident light onto the retina by a factor of 100,000 or more. As a result, the maximum permissible exposure (MPE) levels are quite small. The outer layer of the eye, the cornea, acts as an optical bandpass filter, passing wavelengths ranging roughly from 0.4 to 1.4 μm . Light energy at wavelengths outside this range is absorbed by the cornea and

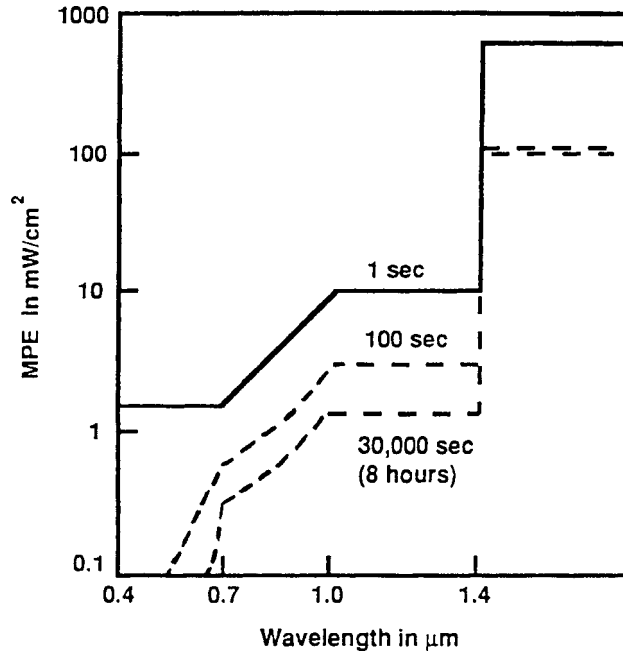


FIGURE 14.5 Levels of maximum permissible exposure. The sharp increase in MPE for $\lambda > 1.4 \mu\text{m}$ is due to the filtering property of the eye's cornea. (From [Chu87] © IEEE.)

does not pass through to the retina. Because of this, wavelengths of $1.5 \mu\text{m}$ and higher are relatively safe for the human eye. The low cost of GaAs laser diodes and silicon APDs, which operate at $0.85 \mu\text{m}$, make them popular devices, but care must be taken in systems using them not to exceed acceptable MPE levels. The acceptable levels of MPE are specified by standards organizations. Figure 14.5 shows a typical curve representing acceptable MPE levels for various wavelengths.

14.4 IR CHANNEL CHARACTERIZATION AND DATA-RATE LIMITATIONS

There are three principal limitations for IR communications, caused by (1) interference from ambient light, (2) multipath channel characteristics in the case of diffused transmission, and (3) the transient response of IR devices. The rise and fall times of inexpensive LEDs limit the data rate to about 1 Mb/s. Explanation of the other two limitations requires more detail discussion, which follows.

14.4.1 Optical Propagation and Multipath Effects

As implied in the preceding discussion, IR communication in an indoor environment makes use of signals arriving at a receiver by a multiplicity of paths, and therefore the channel is a multipath channel. The basis for calculation of optical signal propagation is the *Lambertian law*. The Lambert equation

$$P(\theta) = \frac{n+1}{2\pi} P_t \cos^n \theta, \quad -90^\circ < \theta < 90^\circ$$

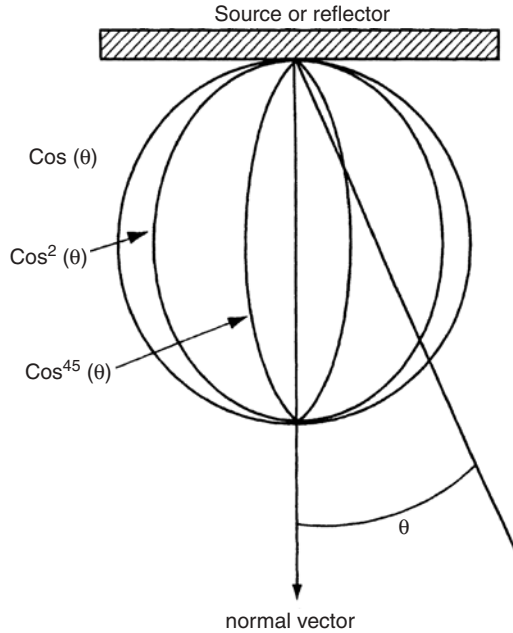


FIGURE 14.6 Lambertian radiation intensity pattern. The equation for the generalized Lambertian pattern is $P(\theta) = [(n + 1)/2\pi]P_T \cos^n \theta$, where $-90^\circ < \theta < 90^\circ$.

gives the power per unit solid angle $P(\theta)$ received from an IR source or a reflector with total transmitted or reflected power P_T , where n is the *mode order* representing the directionality of the source. Figure 14.6 shows the radiation pattern for different values of n . In the simulation of optical propagation, the ceiling and walls in an indoor area are modeled as diffuse Lambertian reflectors with $n = 1$. To determine the reflected power, in a manner similar to ray tracing in radio propagation modeling, a reflection coefficient ρ is included in the calculation to represent the ratio of the total power reflected into the hemisphere to the power incident on the surface. The reflection coefficient ρ is a positive number less than 1. A typical value of ρ for plaster walls and most ceiling materials is in the range 0.7 to 0.85, while for painted, wooden, and textile surfaces it varies from 0.4 to 0.9.

Figure 14.7 shows a simple room with a transmitter and a receiver, a model used for simulation of optical radiation in [Bar91a]. The receiver has a photosensitive area A_R with a field of view (FOV) determined by the packaging of the photosensitive diode. The power received from an area radiating or reflecting power W is given by

$$P_R = W A_R \sin^2(\text{FOV}), \quad 0^\circ < \text{FOV} < 90^\circ$$

As indicated by this equation, the received power P_R is independent of the position and angular orientation of the photodetector relative to the radiating element if the radiating element covers the entire FOV. In practice, the FOV is typically reduced by device packaging to some angle less than 90° , and this constrains the application of diffused IR systems in open environments. However, in indoor areas the photosensitive diode can absorb power reflected from the walls, ceiling, and other objects located within the FOV.

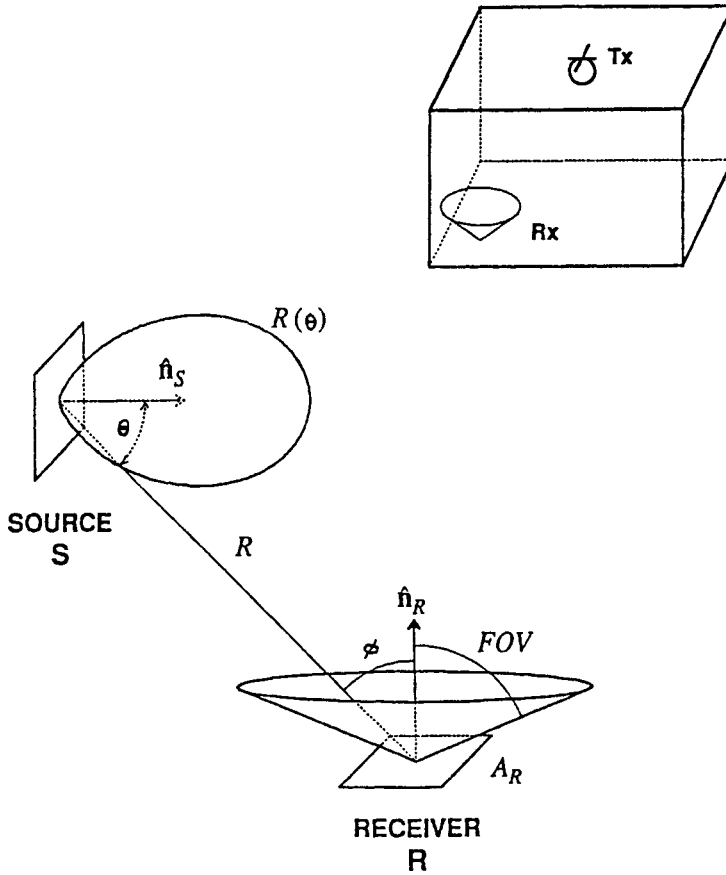


FIGURE 14.7 IR transmitter and receiver patterns. (From [Bar91a] © IEEE.)

To simulate channel impulse response, the signal received in the FOV forms the line-of-sight (LOS) power. The power from the first-order reflection is obtained by calculating the incident and reflected power over all the areas of the walls and ceiling. For higher-order reflections, reflections from the floor are also included. Figure 14.8 shows the result of simulation of the channel impulse responses for different orders of reflections in a $5 \times 5 \times 3$ -m room [Bar91a]. The reflection coefficient is assumed to be 0.85 for the walls and the ceiling and 0.3 for the floor. The only active radiating element is a Lambertian point at the center of the ceiling pointing downward. The receiver is assumed to have a FOV of 85° and is located at coordinates (0.5 m, 1 m, 0 m) in the three-dimensional model. Figure 14.9 shows the overall impulse response and the associated frequency response of the channel. The existence of a strong LOS signal is represented by an impulse. The four subsequent peaks represent reflections from the four walls. The overall power fluctuation in the entire 8-MHz bandwidth is less than 8 dB. The deep fades on the order of 30 dB typically observed on radio channels are not evident here.

There are some important differences between the simulations of radio and optical channels. In simulation of a radio channel, some of the signal energy incident on a wall is assumed to pass through the wall and it has to be traced, whereas for optical

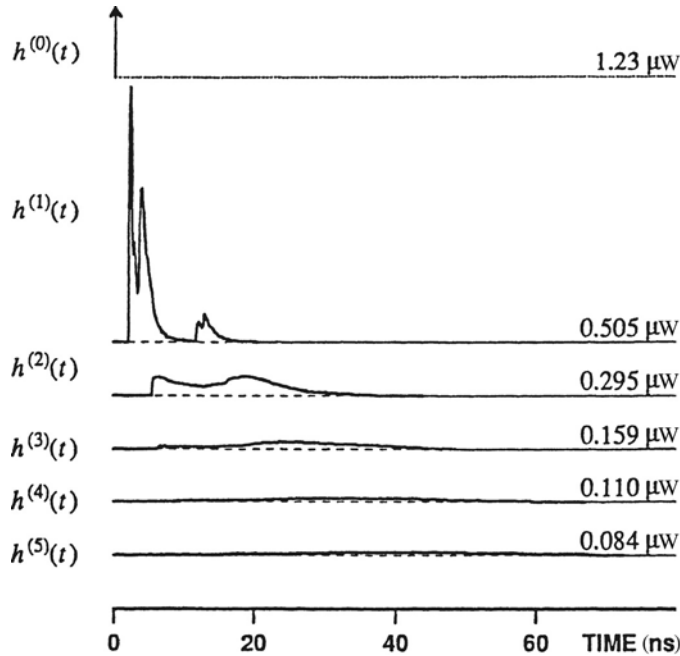


FIGURE 14.8 IR impulse response and a number of reflections. (From [Bar91a] © IEEE.)

transmission there is no penetration through walls or other structures. In simulation of the optical channel, the entire surface of a wall or ceiling is used for calculation of the received power, whereas in ray tracing of radio propagation, only the paths whose reflections pass through the receiver are included in the calculation. Therefore, ray tracing is less computationally complex. The signals arriving from different paths in optical propagation simply add in their power levels, whereas in radio propagation modeling, the amplitudes and phases of the arriving signal paths are added together vectorially. The characteristics of the radio channel may change drastically with movement of the antenna by a fraction of a wavelength. However, the power received by a photodetector remains almost unchanged as we move the antenna. Figure 14.10 shows the received power in an IR simulation. This figure should be compared with Fig. 3.8 for the received power in a radio channel. Signal shadowing is much more evident on optical wireless channels. The optical signal with a FOV is analogous to a radio signal observed in one sector of a multisector antenna.

The multipath on an optical channel causes time dispersion of the transmitted symbol, and the resulting intersymbol interference limits the maximum digital transmission rate, just as it does on a radio channel. As in radio propagation, as room dimensions become larger, the multipath spread increases and the maximum supportable bit rate decreases. The theoretical limit on the data rate is proportional to the bandwidth of the channel frequency response, which is in turn related to the number of reflections included in the model. Figure 14.11 shows the 3-dB width of the frequency response for different numbers of reflections. As the number of reflections increases, the bandwidth decreases, lowering the supportable transmission rate. The theoretical limit for the transmission rate with only first-order reflections is 260 Mb/s [Gfe79b]. Inclusion of up to fifth-order reflections will reduce this number to 85 Mb/s [Bar91a]. Therefore,

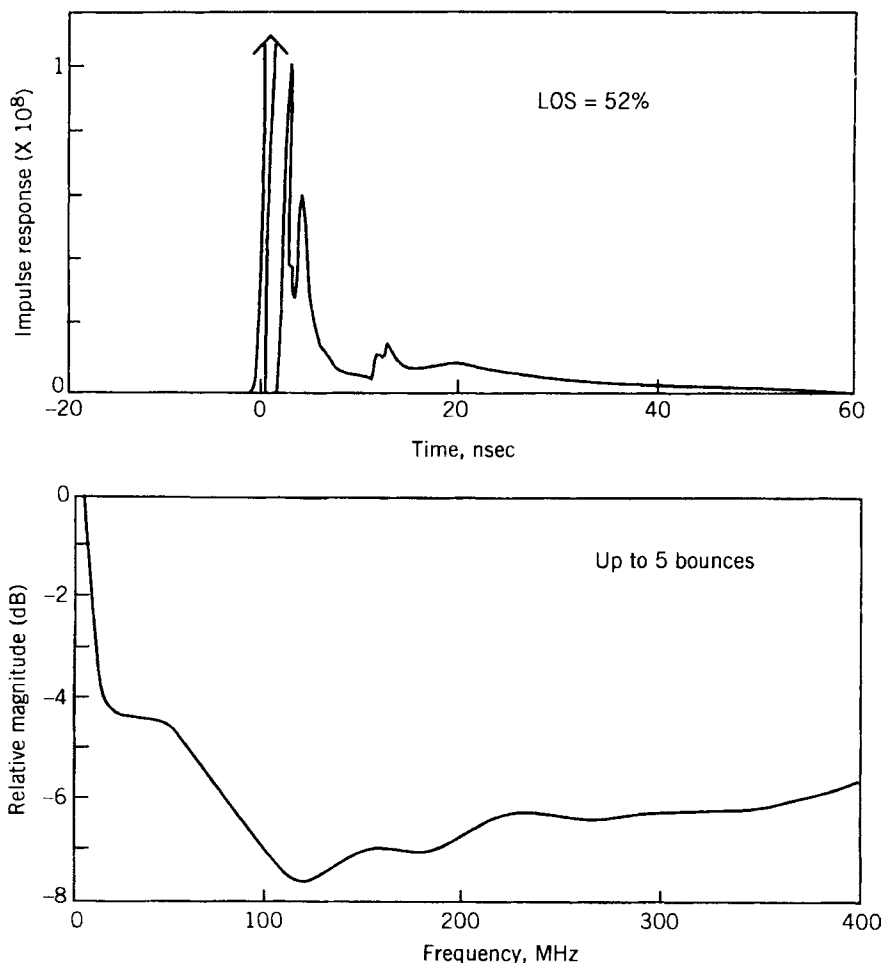
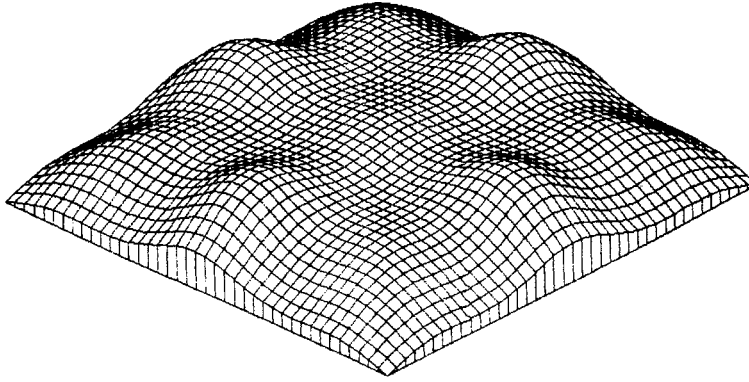


FIGURE 14.9 IR impulse response and corresponding frequency response. (From [Bar91a] © IEEE.)

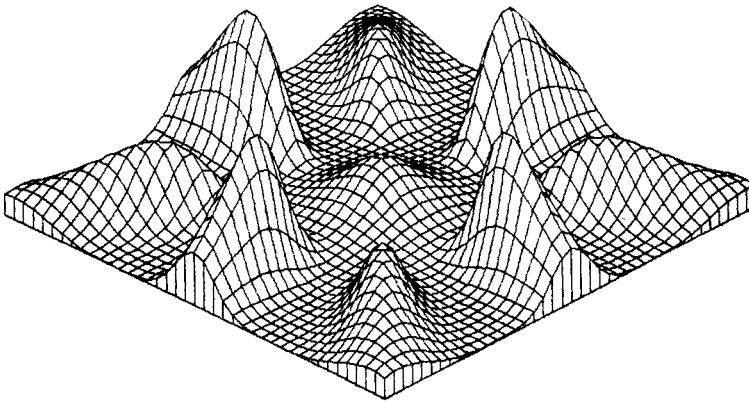
for a room of length 10 m, one expects an achievable transmission rate of 26 Mb/s if only first-order reflections are considered, and 8.5 Mb/s if reflections up to fifth order are included in the simulation.

14.4.2 Effects of Ambient Light

Ambient light is the main source of noise in wireless optical communications. The IR content of ambient light interferes with the reception of IR radiation and, if extensive, can overload the receiver photodiode and drive it beyond its operating point. The current in the photodetector due to the ambient light is modeled as Gaussian shot noise plus a strong direct-current (dc) component. As long as the dc component does not saturate the photodiode, its effect is negligible, and therefore it is the shot noise that determines the quality of the received signal. Three sources of ambient light—daylight, incandescent illumination, and fluorescent lamps—potentially interfere with IR communications. Figure 14.12 shows the power spectral density of the shot



POWER RECEIVED AS FUNCTION OF TERMINALS



RADIANT EMITTANCE PRODUCED BY 9 LED'S

FIGURE 14.10 Power received in an IR channel simulation.

noise component of these three ambient light sources over a range of wavelengths from 0.4 to 1.4 μm . Incandescent light, being rich in long-wavelength (red) light, has the worst effect because its spectral peak (around 0.7 μm) overlaps that of GaAs diodes.

Daylight contains less IR radiation than incandescent light does, but if sunlight falls directly onto a receiver lens, whether indoors or outdoors, the dc component can saturate the photosensitive diode, preventing proper operation of the receiver. Fluorescent light, the predominant type of illumination in office and professional buildings, contains a relatively small amount of IR radiation. However, unlike sunlight and incandescent light, the optical interference generated by a fluorescent light follows the pattern of the line voltage, causing a strong 120-Hz interference baseband signal rich in harmonics reaching as high as 50 kHz [Ank80]. Figure 14.13 shows the spectrum of harmonics of the interfering signal from a fluorescent bulb.

One way of avoiding this dominant source of interference is to modulate the transmitted signal onto a carrier frequency which is at least several hundred kilohertz. Figure 14.14 shows the maximum theoretical transmission speed of pulse code modulation (PCM) systems for different room sizes in the presence of fluorescent

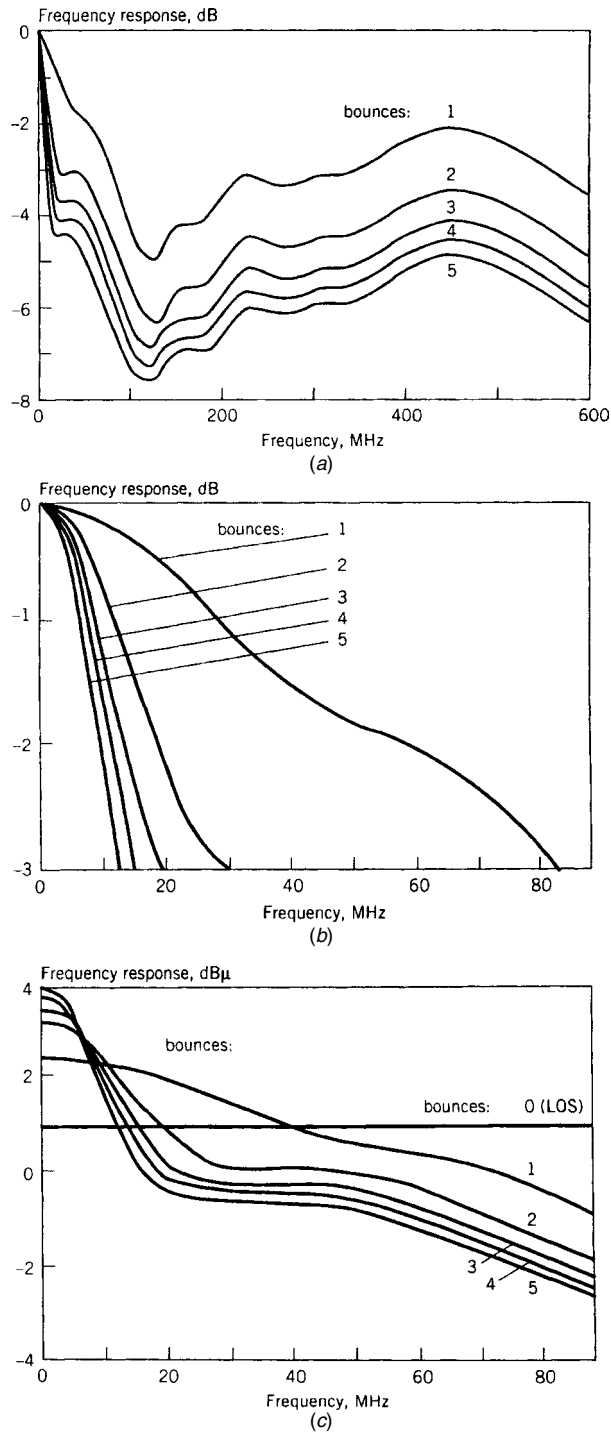


FIGURE 14.11 Frequency responses for different numbers of reflections. In (a) the range is 600 MHz and in (b) the range is 90 MHz, with both curves normalized to unity dc gain to highlight the -3 -dB bandwidths. In (c) the range is 90 MHz but the curves are not normalized, and one can see the higher dc gain provided by multiple reflections. The curves are labeled with the highest number of considered reflections.

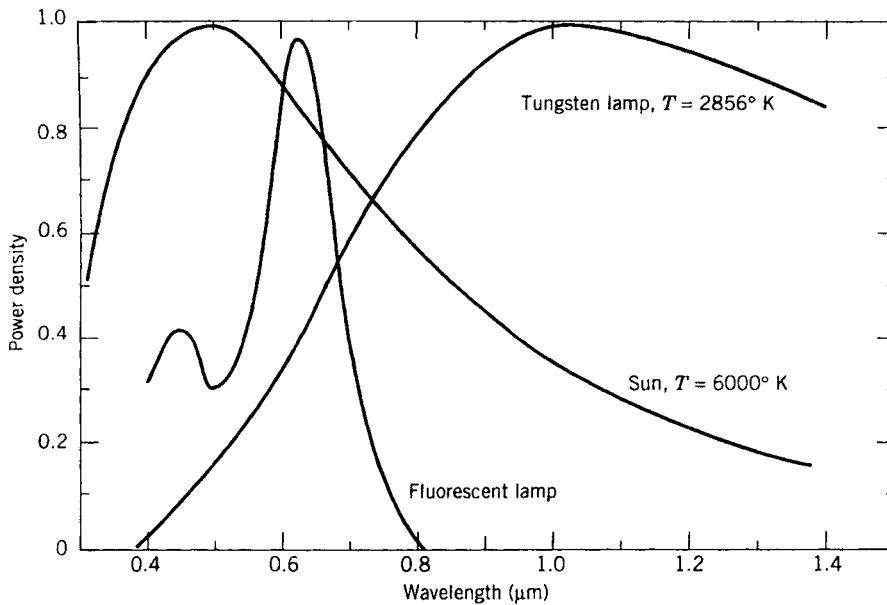


FIGURE 14.12 Power spectral densities of sources of ambient light and the spectral center of the GaAs diodes used for diffused infrared communications. (From [Gfe79b] © IEEE.)

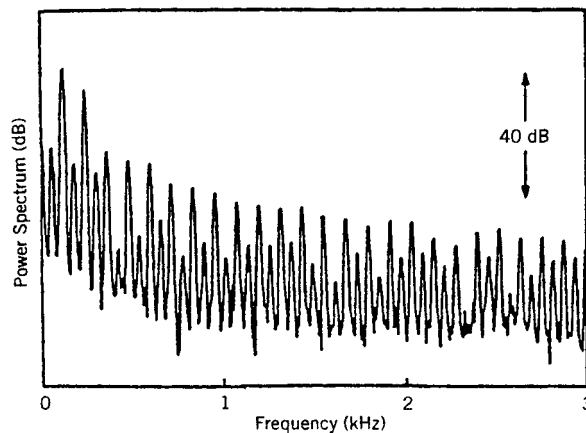


FIGURE 14.13 Harmonics contained in fluorescent light. The spectral components appear at multiples of the line frequency, extending to over 100 kHz. (From [Bar91b] © IEEE.)

light, daylight, and multipath distortion for a probability of error equal to 10^{-9} . The noise from the ambient light can be suppressed by an optical filter in front of the photodetector. Figure 14.15 shows a typical response for such a filter. These filters are implemented with a stack of dielectric slabs. One drawback of mounting a filter of this sort in front of the photodiode is that it will reduce the FOV of the receiver.

14.4.3 Photodiode Capacitance and Noise Enhancement

The light incident on the surface of a photodetector causes electrical charges on the surface which are discharged in the subsequent preamplifier. The photodetectors appear

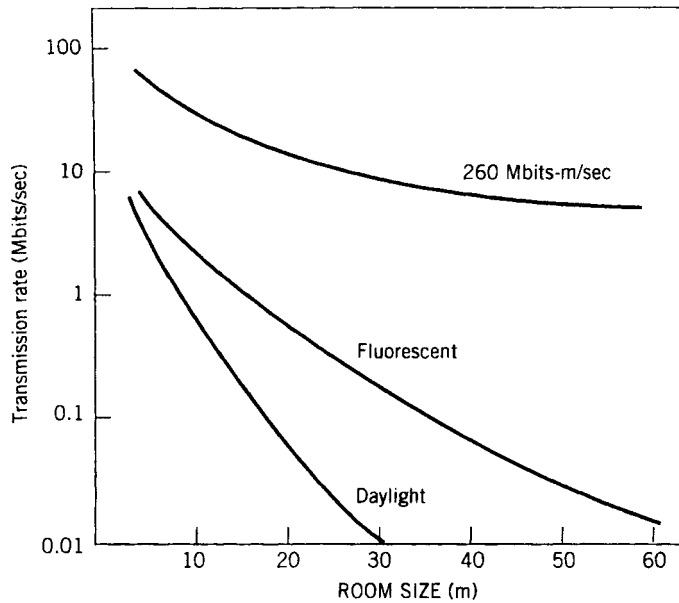


FIGURE 14.14 Maximum theoretical transmission rates in megabits per second versus room size in meters. Optical power, 1 W; receiver area, 1 cm²; $\lambda = 950$ nm. (From [Gfe79b] © IEEE.)

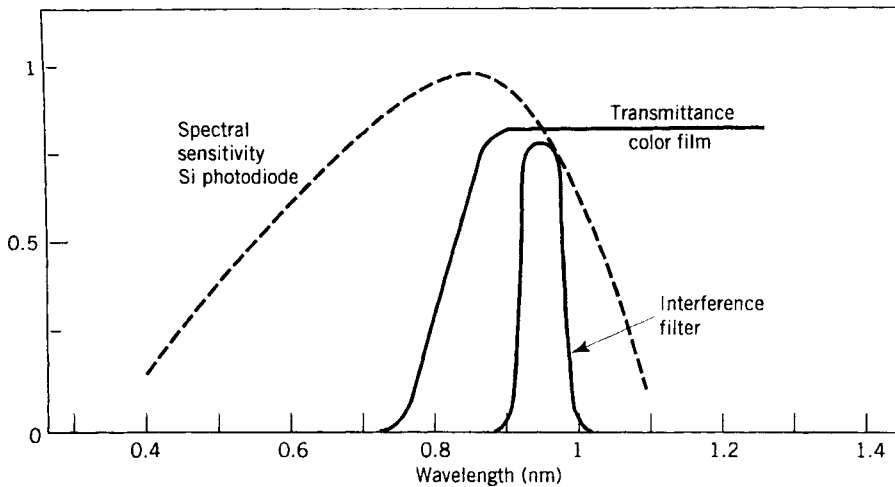


FIGURE 14.15 Noise reduction using an optical filter. Visible light can be blocked using developed unexposed color film. (From [Gfe79b] © IEEE.)

to the preamplifier as an input capacitance or first-order low-pass filter. Let us assume that the incident light is intensity-modulated by a sinusoid of frequency f and that we are measuring the SNR at the output of the preamplifier. If the only source of noise is the thermal noise of the preamplifier, the measured SNR will decrease as the modulation frequency increases. In other words, the capacitance of the photodetector attenuates the high-frequency components of the received signal but has no effect on the noise subsequently introduced in the preamplifier. However, referred to the input

signal, the effective noise component increases as the modulation frequency increases. When a first-order low-pass filter is followed by white Gaussian noise, the equivalent noise referred back to the input of the low-pass filter has a power spectrum that is quadratic in frequency and therefore is referred to as f^2 noise [Bar91b]. To collect sufficient signal power, it would appear that we should use a photodetector with a large area. However, an increase in the area of the photodetector increases the value of the capacitance, which in turn reduces the bandwidth of the low-pass filter and thus worsens the noise penalty.

To counter the effect of f^2 noise, we divide the required photodetector surface area into n smaller areas, each smaller photodetector having its own preamplifier. In this way we reduce the effect of f^2 noise at the expense of increasing the effect of thermal noise. The optimal number of photodetectors is determined by striking a balance between f^2 noise and thermal noise. More detailed discussions of f^2 noise and methods for counteracting its effects are available in [Gfe79a, Bar91b].

14.5 MODULATION TECHNIQUES FOR OPTICAL COMMUNICATIONS

The most common applications of IR communications in the past two decades include remote control, hearing aids, wireless audio systems, cordless phones, wireless keyboards, and WLANs. Various modulation and multiple access techniques have been examined by various manufacturers for use in these systems. In this section we provide a brief summary of these techniques.

In optical communications, modulation is ordinarily performed in two stages, as indicated in Fig. 14.16. The message signal first modulates a carrier frequency, and the resulting signal then modulates the emitted optical light. The second or optical stage of modulation can be performed using intensity (amplitude), frequency (wavelength), or phase modulation. With *intensity modulation* the amplitude of the emitted light is varied in accordance with the voltage of the information signal, and the implementation is straightforward. In *wavelength modulation* the information is impressed on the optical signal by modulating the amplitude of the signal at different wavelengths. The implementation of wavelength modulation requires a bank of expensive narrowband optical filters, and this generally makes the technique impractical. *Coherent optical modulation*, in which the phase of the optical signal is modulated directly, has not been considered for wireless optical communications because it is an extremely expensive approach.

The most typical form of second-stage or optical modulation is intensity modulation, in which the voltage of a diode is changed in accordance with the signal amplitude.

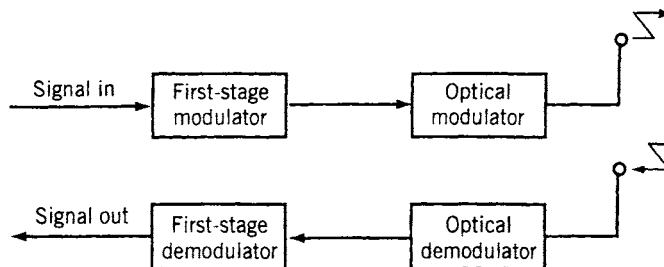


FIGURE 14.16 Block diagram of a two-stage modulator.

At the receiver the detector is a photodiode generating a voltage proportional to the light incident on the photodetector plate. A variety of modulation techniques are used for the first stage of modulation in different applications, and in the remainder of this section we describe several of the commonly used techniques.

14.5.1 Analog Modulation Techniques

Analog modulation techniques are typically used for the first stage of modulation in wireless optical communications for audio applications. One motivation for modulating the audio signal is to move the transmitted signal spectrum away from the lower end of the band, where there is much interference from the harmonics of fluorescent light. Another motivation for using modulation in wireless optical audio systems is to provide a convenient means of supporting multiuser applications using frequency-division multiplexing (FDM). Although FDM is commonly used in radio systems, it is not widely used in wireless optical systems.

Amplitude modulation is generally not suitable for wireless optical communications because of its relatively poor resistance to noise. FM is the standard choice for the first stage of modulation in an analog audio communications system. The frequency-modulated signal is used in the second stage to modulate the intensity of the optical signal, as shown in Fig. 14.17. A carrier frequency of 95 kHz with a frequency deviation of 50 kHz is commonly used in single-channel wireless optical audio systems [Ank80]. The carrier frequency can readily be increased up to 500 kHz. In a multichannel environment, a nine-channel FDM system has been developed for simultaneous interpretation in multilingual conferences and exhibitions [Ank80].

14.5.2 Pulse Modulation Techniques

Pulse modulation techniques include pulse amplitude modulation (PAM), pulse duration modulation (PDM), and pulse position modulation (PPM). PAM is a form of amplitude modulation and therefore is not suitable for optical wireless media because of relatively poor performance in additive noise. PPM is the most popular modulation technique for wireless optical communications. For analog signal applications such as cordless phones the audio signal to be transmitted is sampled and the position of the transmitted pulse is adjusted in accordance with the sample amplitude [Bra80]. Figure 14.18 illustrates the principle of operation of PPM for a sampled analog signal. For transmission of digital messages, the pulse position is adjusted in accordance with the value of the transmitted digit.

For low-speed data, the remote control and wireless keyboard applications are very similar. In wireless keyboards, the transmitted alphabets are larger and the transmission rates are higher than are needed for remote control functions. For remote control, various modulation techniques have been examined by different manufacturers. The most commonly used modulation technique is simple on-off keying (OOK), which we discuss later. For wireless keyboards, a form of digital PPM is used in which the IR pulse is transmitted at the beginning or at the middle of the bit interval to represent transmission of a 1 or a 0. A 40-MHz carrier frequency is virtually a standard for both applications [LaR84], and in either application the data rate is below 2400 bits/s. In typical practice, several consecutive narrow pulses are used to represent one modulation symbol. This technique saves transmitted signal power, thus increasing battery life.

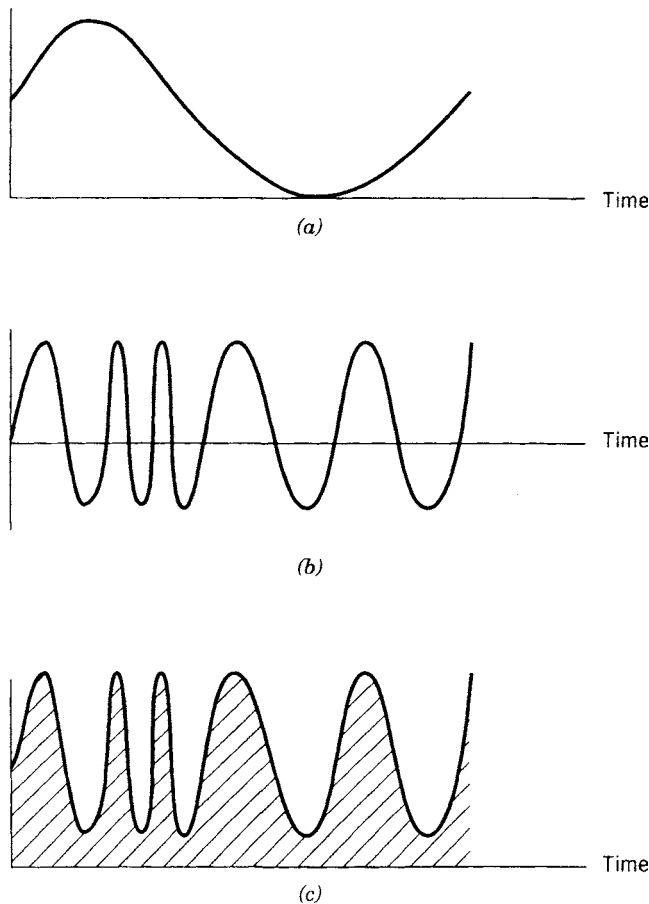


FIGURE 14.17 Example of analog AM/FM modulation. The FM-modulated audio signal is used to modulate the intensity of the IR light: (a) message amplitude; (b) FM-modulated signal; (c) intensity of the light.

Figure 14.19 shows a biphase data stream for two PPM systems used in wireless keyboard applications. In one of the PPM systems, three short pulses are used to represent each modulation symbol; in the other, nine are used. A disadvantage of transmitting multiple pulses per modulation symbol is that the transmission bandwidth is much wider than the information bit rate. In WLAN applications, where very high data rates are desirable, this is not practically feasible. In addition, the wide bandwidth requirement introduces significant f^2 noise into the system, which in turn restricts the area of coverage, leaving this method suitable only for short-range applications. However, the PPM approach has the advantage of low power consumption, which translates into long battery life.

14.5.3 Digital Modulation Techniques

Binary phase shift keying (PSK) and frequency shift keying (FSK) have been examined for use in high-speed wireless optical data communication networks. Figure 14.20 shows a typical binary PSK system for optical wireless communications. The

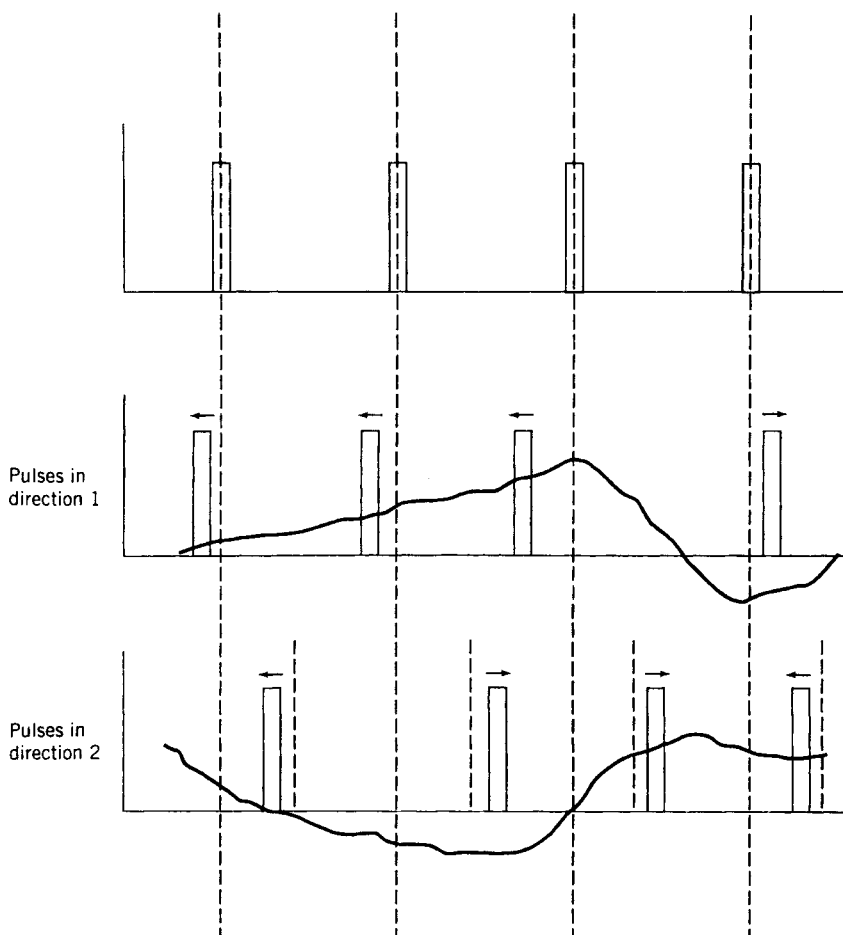


FIGURE 14.18 Pulse position modulation for bidirectional analog transmission in a cordless phone system.

phase-modulated signal is used to intensity-modulate the emitted optical radiation. An experimental 64-kb/s binary PSK system is reported in [Gfe79a]. This system uses a carrier frequency of 256 kHz. The receiver has two phase-locked loops for carrier and bit synchronization. The high carrier frequency eliminates the effects of ambient light from fluorescent lamps. The receiver operates with a low error rate of 10^{-7} in a 380-lux ambient light environment [Gfe79b]; but during the 100-ms switching time of a lamp, the error rate increases to 10^{-3} .

An FSK experimental system using 200- and 400-kHz carrier frequencies for down-link and uplink, respectively, is reported in [Gfe81]. Carrier synchronization in a multipath environment would add to the complexity of the design, and therefore differentially encoded PSK and noncoherent FSK are preferable. To achieve higher data rates, multiphase PSK and multifrequency FSK are the natural design choices. However, these systems require higher SNRs for proper operation.

Another approach to increasing the data rate is to use multicarrier modulation. Optical multicarrier systems are popular in video distribution applications [Bar91b].

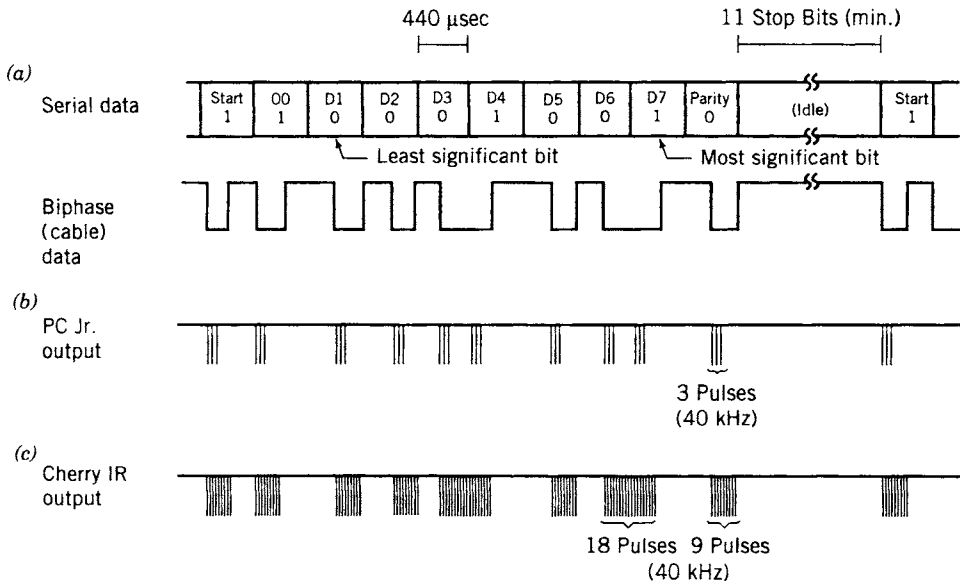


FIGURE 14.19 PPM for a wireless keyboard application.

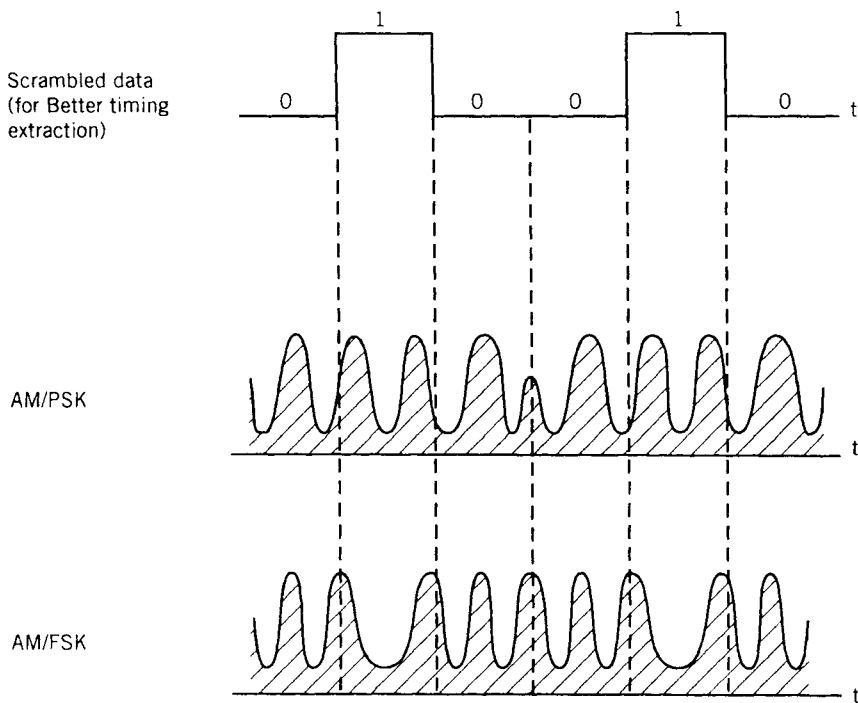


FIGURE 14.20 AM/BPSK modulation in optical wireless communications.

In a multicarrier system, the incoming data bits are divided into several interleaved data streams, each modulated onto a separate carrier frequency. The number of carriers should be chosen so that the data rate in each subcarrier is low enough that the intersymbol interference (ISI) caused by multipath is negligible for individual

carrier signals. Multicarrier communication over a frequency-selective channel offers advantages similar to those described for frequency-hopping spread-spectrum systems (Chapter 9). Coding can be applied across bits on different carriers so that the errors on a carrier falling into a deep fade can be recovered from the bits received correctly on the other carriers. Another approach to improving the performance is to adjust the power or the data rate of the carrier affected by the frequency-selective fading. Multicarrier communication requires linearity in the transmitter final stage and the receiver front end. Otherwise, harmonic distortion will cause interference among the carriers. Another difficulty with the multicarrier systems is that they require high peak transmission power.

14.5.4 Baseband PCM

The simplest, yet most effective modulation technique for high-speed wireless optical communications is baseband pulse code modulation (PCM). PCM-coded bits are modulated directly onto the emitted optical signal. To transmit one of the binary symbols, the light is emitted with the highest power, and for the other symbol nothing is transmitted. This is equivalent to OOK for the second stage of modulation. This method is suitable for high-speed data communications. To avoid interference from fluorescent lights, the incoming data stream is Manchester-coded, which reduces the low-frequency portion of the transmitted spectrum. Figure 14.21 represents a typical Manchester-coded baseband PCM system.

The spectrum of the Manchester-coded data is shown in Fig. 14.22. Although the Manchester-coded signal has the advantage of reduced frequency components at lower frequencies, it has the disadvantage that it doubles the transmission rate and consequently the required transmission bandwidth. As the data rate is increased, the main lobe of the spectrum shifts farther away from the low frequencies affected by ambient light. With the increase in data rate, the ISI caused by multipath will affect the performance of the system. An experimental PCM system using diffused optical propagation and

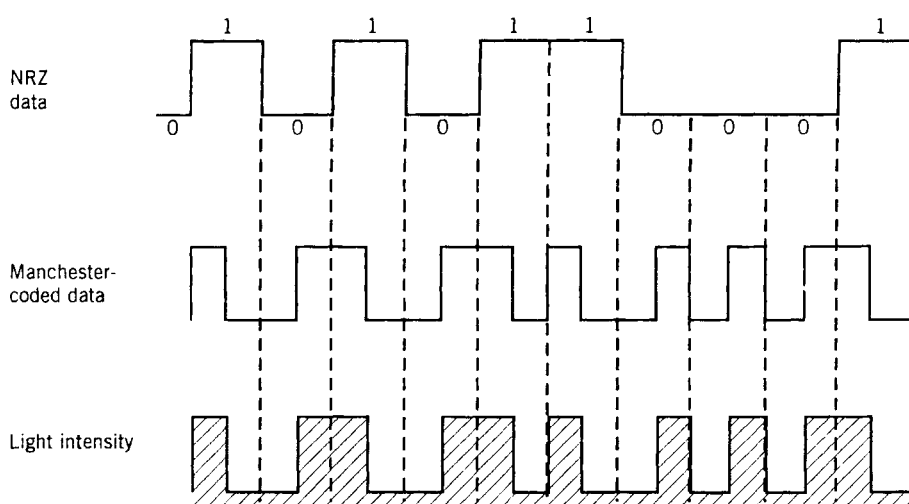


FIGURE 14.21 Manchester-coded pulse code modulation.

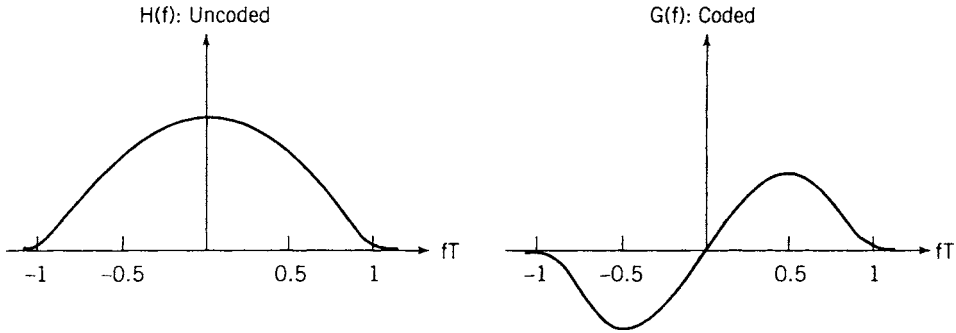


FIGURE 14.22 Spectrum of the Manchester-coded signal.

having a data rate of 125 kb/s was first reported in [Gfe79b]. Using the latest fast LEDs and PIN diodes, data rates on the order of 1 to 2 Mb/s are feasible with base-band PCM modulation. As with applications to radio channels, the DFE has been studied for mitigating the effects of multipath on high-speed wireless optical communication channels [Bar91b]. As shown in Fig. 14.9, the impulse response of the optical wireless channel has a strong LOS component followed by multipath arising from reflections. The DFE in this case does not need a forward equalizer, making implementation simpler and more cost-efficient. Using DFE technology and direct optical propagation, very-high-speed optical links can be designed.

Table 14.1 [Bar91b] shows the details of several experimental systems that have been reported in the literature. Commercial products with data rates of 1 to 2 Mb/s providing coverage of an indoor area for a portable computer are available in the market. Indoor coverage is aided by active (receive–transmit unit) or passive (mirror) reflectors. For lower data rates on the order of 19.2 kb/s, the received SNR is adequate to cover a large room without the need for a reflector.

14.6 MULTIPLE ACCESS AND DATA RATE

WLANs or wireless PBX systems are intended to support many users in a multiple-access environment, and there are requirements to provide much higher information transmission rates than those discussed above. Therefore, more work has to be done to address these applications. Experimental data systems have been implemented with DFIR, using PCM at a data rate of 125 kb/s [Gfe79b]. That system showed poor performance in the vicinity of ambient light, due to interference at low optical frequencies. To avoid this problem, a PSK system with a data rate of 64 kb/s was examined in a continuation of the same development. In a multiple-access environment, an experimental system was developed that operated with FSK modulation and a carrier-sense multiple-access (CSMA) protocol [Gfe79a]. (We discuss multiuser access protocols in Chapter 11.) In that experiment the uplink and downlink were separated in frequency, operating at 200- and 400-kHz center frequencies, respectively, and data rates up to 100 kb/s were supported in this system. In this experiment, transponders were installed in the ceiling. Communication between terminals and transponders used the IR links, and transponders were connected by fixed wiring to the controller. The wired portion

TABLE 14.1 Published Accounts of Wireless Infrared LANs^a

Date	Organization	Bit Rate	Directionality	Duplex MUX	Multiple Access	Subcarrier Frequency	Modulation	Wavelength	Power	Area
1979	IBM Zurich	64 kb/s	Diffuse	Subcarrier	CSMA/CD	200/400-kHz	Subcarrier	950 nm	100 mW	0.67 cm ²
1981		125 kb/s				Baseband	BPSK			
1983	Fujitsu	19.2 kb/s	LOS, diffuse	Subcarrier		1/1.5 MHz	Manchester Subcarrier	880 nm	15 mW	1 cm ²
1985	Fujitsu	48 kb/s	Narrow up, wide down	Subcarrier		1/1.5 MHz	FSK Subcarrier	880 nm		
1984	Hitachi	250 kb/s	Narrow up, wide down	Subcarrier	CSMA/CD polling	>10 MHz	BPSK Subcarrier		300 mW	
1985	HP Labs	1 Mb/s	Narrow LOS	Wavelength	CSMA/CD	Baseband	FSK	660–880 nm	165 mW	1 cm ²
1986	Motorola	50 kb/s	Widebeam LOS		TDMA	Baseband	RZ OOK	950 nm	16 mW	7.6 mm ²
1987	Bell Labs	45 Mb/s	Narrow beam LOS			Baseband	800 nm	1 mW		
1988	Matsushita	19.2 kb/s	Narrow up, wide down	Subcarrier	Spatial	OOK 650/950 kHz	Subcarrier FSK	880 nm		

Source: [Bar91b] © IEEE.

^a *Directionality* refers to a system's dependence on accurate alignment between transmitter and receiver; diffuse systems have no directionality, whereas line-of-sight (LOS) narrowbeam systems are highly directional. Duplex MUX specifies how interference between uplink and downlink is avoided. *Power* refers to the total optical power of the base station, and *area* refers to the total area of the receiver photodetectors.

of the network in such a configuration does not have to be changed when terminals are relocated; therefore, terminals can be relocated at minimal cost. None of these systems has been marketed successfully. As can be seen in Table 14.1, subsequent developments of IR WLANs have typically incorporated forms of CSMA as the multiple-access protocol, although time-division multiple access (TDMA) and spatial separation have also been used. The current direction of research in IR WLANs is toward higher-order data rates (tens of Mb/s), using directed-beam optical transmission.

QUESTIONS

- (a) Name three major advantages and three major disadvantages of IR communication for WIN systems.
- (b) What is the dominant mechanism in optical wave propagation?
- (c) Describe the field of view of an IR-photosensitive diode.
- (d) What are the major sources of ambient light affecting IR transmissions?
- (e) Why does the spectrum of fluorescent light contain a 120-Hz interference component and its harmonics?
- (f) What is f^2 noise, and how we can counter its effects?
- (g) Why are analog signals modulated onto a carrier before modulation onto IR light?
- (h) Explain why FM modulation is used for analog modulation of an audio signal onto an IR signal. What carrier frequency and frequency deviations are typically used in practice?
- (i) In analog modulation onto IR signals, how high can the carrier frequency be made?
- (j) What modulation techniques are typically used in digital IR systems?
- (k) In lower-speed IR data communication systems, why are several narrow pulses used to represent an IR pulse? Is this a bandwidth-efficient solution? Explain. Why is this method not applied to higher-speed IR communications systems?
- (l) Explain how the baseband modulation techniques counter the effects of ambient light at low frequencies.
- (m) What are the advantages and disadvantages of DBIR transmission compared with DFIR transmission?
- (n) What types of multiple-access methods are used in IR wireless LANs?
- (o) What access method is used in multichannel analog IR systems?
- (p) What is the maximum data rate that can be supported with IR transmission?

15

SYSTEMS AND STANDARDS

- 15.1 Introduction
- 15.2 GSM, GPRS, and EDGE
 - 15.2.1 Global System for Mobile Communication
 - 15.2.2 General Packet Radio Service
 - 15.2.3 Enhanced Data for Global Evolution
- 15.3 CDMA and HDR
 - 15.3.1 CDMA Cellular
 - 15.3.2 HDR
- 15.4 Other Historical Systems
 - 15.4.1 CT2 Telepoint
 - 15.4.2 DECT
 - 15.4.3 U.S. Digital Cellular
 - 15.4.4 JDC/PDC
- 15.5 Wireless LANs
- 15.6 Speech Coding in Wireless Systems
- Questions

15.1 INTRODUCTION

In this chapter we describe the digital wireless systems that have been achieving the greatest commercial success in the wireless industry. We categorize the systems as either high-power wide-areas systems or low-power local-area systems. The most successful systems are those based on industry-sponsored interoperability standards. As pointed out in Chapter 2, standards are of great importance for systems that are intended to serve large numbers of users who require portability and mobility over wide areas, with cellular telephone systems being the best example. In the early development of wireless local area networks (WLANs), systems providing users with wireless access over a well-defined area, standards were initially thought to be unimportant, and a number of proprietary designs emerged. However, as the popularity of *mobile computing*

grew, the WLAN industry and the computer industry recognized the potential benefits to be gained for both their industries by the adoption of WLAN standards, and this consensus led to the IEEE standardization of WLAN protocols. As a consequence, the WLAN industry has experienced a tremendous growth in the use of wireless access for both desktop and laptop computers to networks not only in business, academic, and home environments, but also to increasingly popular public access points (“hot spots”) installed in many of the venues visited by traveling businesspeople and professionals.

Today, the dominant wide-area, high-power wireless systems are based on either GSM and its derivative standards, or on CDMA cellular standards. Deployed WLAN systems are essentially all based on the IEEE 802.11 family of protocol standards. In this chapter we describe the principal characteristics of these systems and provide a brief historical overview of some of the earlier digital wireless systems that were developed and implemented but have been supplanted or overtaken by the systems now predominating in the marketplace.

15.2 GSM, GPRS, AND EDGE

Here we discuss the GSM cellular system and its derivatives, the general packet radio system (GPRS) and enhanced data for global evolution (EDGE). GPRS builds on the GSM architecture, providing a true packet-switched data service over the GSM network. EDGE, in turn, builds on GSM and GPRS by adding higher-rate data services, both circuit- and packet-switched.

15.2.1 Global System for Mobile Communication

The pan-European standard for digital cellular telephony, called the Global System for Mobile Communication (GSM), was motivated by the need for a common mobile standard throughout Europe, where six incompatible analog cellular systems had been in use in different geographic areas. At the start of work on GSM in the early 1980s, spectrum was reallocated throughout much of Europe and surrounding regions so that a completely new system, based on digital technology, could be developed. A good overview of the considerations leading to the GSM standard is given in [Hau94]. Since 1989, the GSM standardization activities have been carried out under the aegis of the European Technical Standards Institute (ETSI).

The Phase 1 GSM900 specifications were finalized in 1990, and the first GSM networks went into operation in 1991. By the end of 1992, some 13 networks in seven countries were on the air, providing customers with cellular voice service. By the end of 1993, GSM operators had put roaming agreements into place, and 32 networks in 18 countries were on the air. GSM data services were launched in 1994, and by the end of that year, 69 GSM networks were operational in 43 countries [Ebe01]. Since its introduction, GSM has enjoyed rapidly growing adoption throughout Europe and elsewhere in the world, particularly in the Asia-Pacific region. As of early 2005, the GSM Alliance estimates that there are about 1.3 billion GSM users in more than 200 countries across the world (see the GSM World Web site, <http://www.gsmworld.com/>).

In principle, the GSM standard can be implemented in any frequency band. The initial frequency allocation for GSM, termed the *900 band*, comprises two 25-MHz subbands: 890 to 915 MHz for uplink (mobile to base) communication and 935 to

960 MHz for downlink (base to mobile) communications. While work on Phase 1 GSM was in progress, the GSM standardization group began considering GSM variants for use in other frequency bands. In 1991, specifications were finalized for the *Digital Communication System 1880* (DCS 1800), which uses the frequency subbands 1710 to 1785 MHz for uplink communication and 1805 to 1880 MHz for downlink communication. Note that the DCS 1800 system is allocated three times the bandwidth allocated to the primary GSM cellular system. In the UK and some other countries, DCS 1800 is also known as the *personal communication network* (PCN).

In 1995, as part of the GSM Phase 2 standards, GSM 1900 was adopted, and that same year GSM 1900 service was launched in the United States, where it is also referred to as the *personal communication system* (PCS), and where the FCC refers to the assigned band, 1850 to 1990 MHz, as the *PCS band*. GSM 1900 networks use the frequency subbands 1850 to 1910 MHz for uplink communication and 1930 to 1990 MHz for downlink communications. Two other variants of the GSM standard, GSM 400 and GSM 850, have also been standardized but have not yet been widely deployed [GSMweb]. GSM standardization is an ongoing evolutionary process, intended to meet market demands for new capabilities and services as they emerge. The current phase of standardization activity, termed Phase 2+, is addressing a number of issues, including enhanced data services, improved voice-coding techniques, and call-processing capabilities. Two of the enhancements that have already emerged from the Phase 2+ effort, GPRS and EDGE, are described in later paragraphs.

The primary GSM 900 standard and its variants differ essentially only in their frequency allocations, and the channel structure and transmission format are the same for each. In GSM 900, the 25-MHz band used in either direction is divided into 124 channels, with carriers spaced 200 kHz apart. Each cell site in a GSM system has a fixed number of channels (two-way), ranging from one to usually not more than 15. The cells range in size from 1 to several kilometers.

The GSM network reference architecture is shown in Fig. 15.1. The three basic network elements essential for cellular wireless access to the wired public network are mobile terminals, base stations, and switching centers. Each base station provides wireless communication to and from mobile terminals in its cell coverage area, and the switching centers provide connections from the base stations to the wired public network, including both the PSTN and the ISDN. In addition to providing wireless-to-wired network connections, the switches control the assignment of radio channels to mobile terminals and manage the handoff of calls from one cell site to another as mobile users move through the service area.

GSM also provides support to *roamers*, users who move from one service system to another. The management of intersystem roaming is implemented using two databases, shown in Fig. 15.1, called the *home location register* (HLR) and the *visitor location register* (VLR). GSM provides personal mobility by use of a *subscriber identity module* (SIM), which carries the personal number assigned to the user. Mobile users identify themselves to the network by insertion of a SIM into a mobile terminal. In this way, users may access a GSM network from arbitrary locations, including public GSM terminals. The various functions concerned with handoff of mobile calls from cell to cell or from one service system to another, or access from remote locations, are all grouped under the name *mobility management*. Good descriptions of the GSM mobility management functions can be found in [Rah93, Hei99, Ebe01]. Figure 15.1 also includes another database, the *equipment identity register* (EIR). This database holds

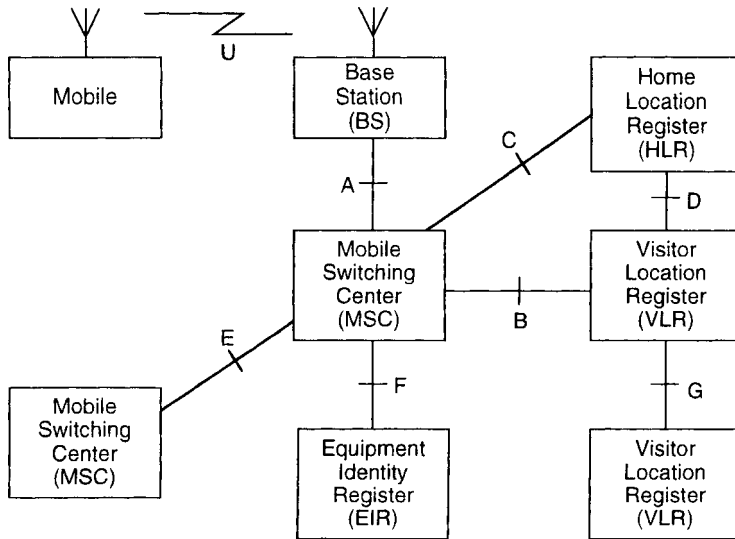


FIGURE 15.1 GSM network architecture.

subscribers' equipment identities, used for identification of unauthorized subscriber equipment, which is dealt with by denial of service.

All of the network elements and databases shown in Fig. 15.1 are contained in all of second-generation cellular systems. However, in GSM, all the interfaces labeled A to G and U are defined by GSM standards. This gives the cellular operators flexibility in configuring their individual systems and in procuring equipment from a variety of manufacturers.

In GSM, each frequency channel supports multiple user channels with a TDMA signal format. Each 200-kHz FDM channel uses an aggregate bit rate of 270.833 kb/s, carried over the radio interface using GMSK modulation with a bandwidth–time product of 0.3. Use of a transmitted bit rate as high as 270 kb/s requires the implementation of adaptive equalization techniques to deal with channel multipath, and GSM specifications require that equipment be built to accommodate time dispersion up to 16 μ s. The GSM standard provides for the optional use of slow frequency hopping, at a rate of 217.6 hops/s, as a means of reducing other-user interference. Frequency hopping has been implemented in some GSM service networks, and practical experience thus far indicates that performance improvements can be achieved using this feature.

The 270-kb/s data stream in each FDM channel is divided into eight fixed-assignment TDMA channels termed *time slots* or *logical channels*. These logical channels are organized into a hierarchical frame structure that provides each mobile terminal with a two-way traffic channel and a separate two-way control channel. The numbering of time slots is offset between the two directions on the radio link to prevent a mobile terminal from transmitting and receiving at the same time. This use of offset time slots is sometimes referred to as *time-division duplexing* (TDD).

The GSM frame hierarchy depicted in Fig. 15.2 ranges from a bit interval of 3.69 μ s to a hyperframe (used for initializing an encryption algorithm) of length 3 h 28 min 53.76 s. The basic building block of the frame hierarchy is a 4.615- μ s frame, shown in Fig. 15.3. Each frame comprises eight 577- μ s time slots, and the composition of a time

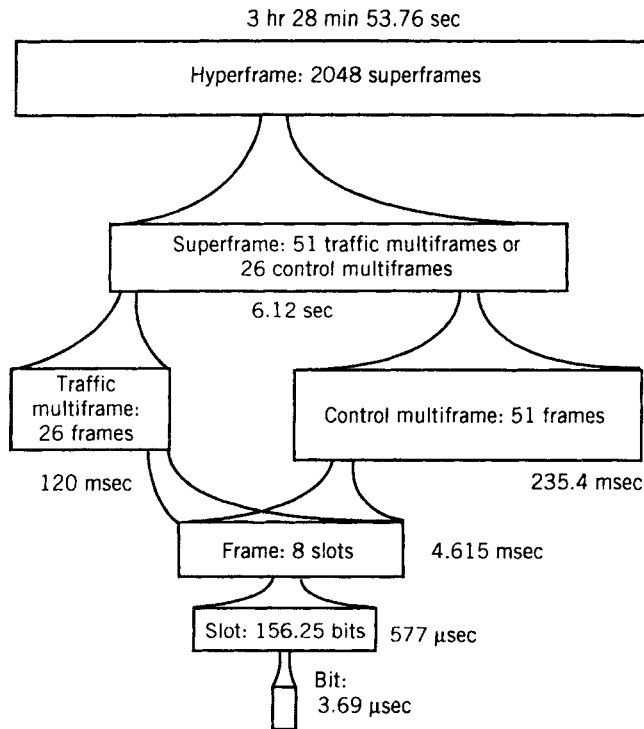


FIGURE 15.2 GSM frame hierarchy.

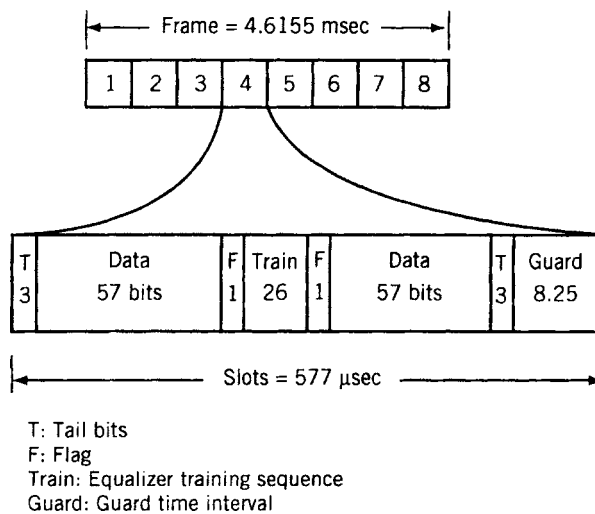


FIGURE 15.3 GSM frame and time-slot structure.

slot is shown in the expanded portion of the figure. The time-slot interval is equivalent to the transmission time for about 156.25 bits, although only 148 bits are actually transmitted in each slot interval. The remaining time, 8.25 bits of time duration or about 30.5 μ s, is guard time, in which no signal is transmitted, to prevent overlapping

of signal bursts arriving at a base station from different mobile terminals. Each slot carries two data bursts, each burst consisting of 57 bits of user information and a flag bit to identify the 57 bits as either digitized speech or other information. The remaining bits are a 26-bit equalizer training sequence in the middle of the slot and two sets of “tail bits,” all logical zeros. The tail bits are used in the convolutional decoding of the 57-bit segments of user data.

The next level in the hierarchy is a GSM multiframe, shown in Fig. 15.4. Every 120-ms multiframe is composed of 26 frames, each containing eight time slots. In each multiframe, 24 frames carry user information, while two frames carry system control information in *associated control channels*. The GSM specification provides for two types of traffic channels, *half-rate* and *full-rate*. All GSM service networks implement full-rate traffic channels, each channel utilizing all 24 user-information frames in each multiframe. From Figs. 15.3 and 15.4 we see that one full-rate traffic channel utilizing one slot in each of the 24 frames in a 120-ms multiframe has a data rate of $24 \times 2 \times 57 / 0.120 = 22,800$ bits/s. The full-rate speech coder specified for GSM uses linear predictive coding with regular pulse excitation and long-term prediction (RPE-LTP) [Var88]. This speech coder has a data rate of 13 kb/s, and the addition of error-detection and error-correction coding brings the transmission rate up to 22.8 kb/s. As an alternative to digitized speech, a full-rate traffic channel can support data service at user rates of 2.4, 4.8 and 9.6 kb/s, employing transmission rates of 3.6, 6.0 and 12.0 kb/s, respectively.

The half-rate traffic channel specification, designed to improve spectrum utilization by allocating only 12 frames per multiframe to user channels, is used to support the GSM half-rate speech coder, the design of which was finalized early in 1995 [Ste01]. The GSM half-rate codec uses the vector-sum excited linear prediction (VSELP) algorithm. The VSELP algorithm is an analysis-by-synthesis coding technique and belongs to the class of speech coding algorithms known as CELP (code excited linear prediction) [GSM89]. The half-rate speech coder operates at a rate of 6.5 b/s, and channel coding brings the transmitted data rate up to 11.4 kb/s. The half-rate traffic channels can also be used to support data services at user data rates of 7.0, 4.8, and 2.4 kb/s. Whereas full-rate traffic channels are implemented in all GSM service networks, half-rate channels and services have been implemented selectively at the option of individual service providers.

As noted earlier, two frames in each 120-ms multiframe are allocated to control channels. In the full-rate implementation of GSM, only one of the frames, the one near

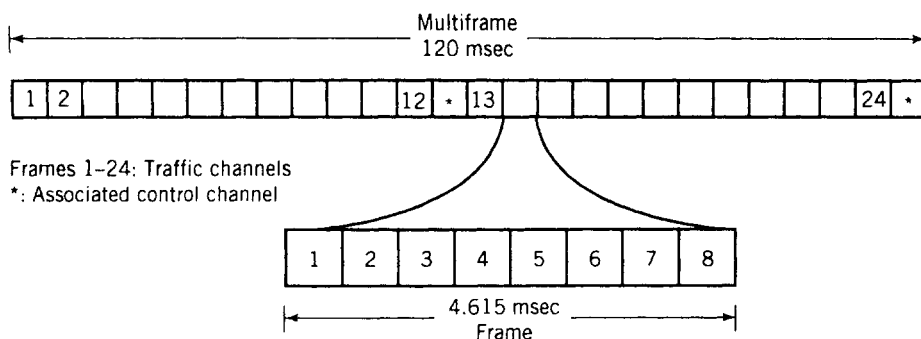


FIGURE 15.4 GSM multiframe and frame structure.

the center of the multiframe, is used. It provides eight *slow-associated control channels* (SACCHs), one for each logical traffic channel, to carry link control information between mobile terminals and base stations. In the half-rate GSM implementation, the last frame in each multiframe is used to provide the additional eight SACCH channels. The transmission format also provides for establishing a *fast associated control channel* (FACCH) on demand within any traffic channel. This is done by usurping one or both 57-bit data blocks within a slot and by indicating this using the flag bits in each slot time. The FACCH is used by either the mobile terminal or the base station for conveying signal quality information and other information used in managing call handoffs.

In the GSM frame hierarchy (Fig. 15.2), the eight-slot frames may be organized into control multiframes rather than traffic multiframes. Control multiframes, which we do not discuss in detail, are used to establish several types of signaling and control channels used for system access, call setup, synchronization, and other system control functions. Either traffic or control multiframes are grouped into superframes, which are in turn grouped into hyperframes. A detailed description of the GSM frame structure, including the specified allocations of bits in the frame hierarchy, may be found in the GSM recommendations [Rah93] and in several texts, including [Mou92, Red95, Meh97, Hei99, Ste01, Hal03].

Phase 2+ of the GSM development introduced *high-speed circuit-switched data* (HSCSD), in which a mobile station can access up to four time slots in each TDMA frame, thereby achieving data rates up to 57.6 kb/s. Another important Phase 2+ element in the evolution of data services in GSM is discussed next.

15.2.2 General Packet Radio Service

The GSM full-rate data services and HSCSD are all circuit-switched services, which means that when a user invokes a data service, a data traffic channel is assigned and held in place until a call is terminated. This corresponds essentially to using a dial-up modem in the public telephone network. The early GSM networks included provision for wireless access to public-switched data networks (PSDNs); however, this form of access simply carried packet-formatted data traffic over a circuit-switched traffic channel for the full duration of a call connection. In cases of bursty traffic, such as Internet traffic, this form of access can result in very inefficient use of the wireless spectrum. Clearly, the use of true packet-switched data service on the GSM radio channels can provide better utilization of wireless network capacity.

The central concept of true packet-switched data transmission is that the network is accessed only when packets are being transmitted and is then released for use by other packet transmissions. See the discussion of multiple-access methods in Chapter 11. In a wireless service network such as GSM, a single logical channel can be used to support a number of packet data calls, the channel being allocated when needed for sending one user's data packets, then released for allocation to another user's data packets, and so on. Such a capability is useful for many user applications not requiring a continuous bidirectional flow of information. Recognizing the transmission efficiencies achievable with true packet-switched data service, the GSM standardization group incorporated this form of service into the GSM Phase 2+ specifications, and termed the service *general packet radio service* (GPRS). GPRS, launched in year 2000, provides true packet-switched data services in a GSM network and greatly improves

wireless access to packet data networks. GPRS provides access to IP-based networks such as the Internet and corporate intranets, as well as legacy X.25 networks [Ebe01, Ste01]. GPRS supports point-to-point (PTP) and point-to-multipoint (PTM) services as well as *short message service* (SMS).

Stated simply, GPRS overlays true packet data service onto a GSM network in a way that maintains strict separation between a radio subsystem and the network system, with the advantage that no modifications to MSCs are required [Ste01]. GPRS provisioning does require some enhancements and modifications of GSM infrastructure and mobile stations (MSs). In GPRS, the allocation of the eight time slots on one RF channel is flexible. One to eight time slots in a TDMA frame can be assigned to the service, and uplink and downlink time slots are allocated separately. The capacity of an RF channel can be shared dynamically between circuit-switched and packet data services. GPRS provides data rates up to 171.2 kb/s, this upper limit achieved by assigning all eight time slots at 21.4 kb/s each, with no error-correction coding.

The GPRS architecture and its interfaces are shown in Fig. 15.5. Integration of GPRS into the standard GSM architecture (see Fig. 15.1) requires the addition of two new network elements, the *gateway GPRS support node* (GGSN) and the *serving GPRS support node* (SGSN). The GGSN provides an interface between an external packet data network, such as the Internet, and the GSM network supporting GPRS service. The GGSN stores information needed to route incoming data packets to the SGSN that is serving a particular mobile station. An SGSN transfers data packets to and from the mobile stations within its service area. It performs packet routing and transfer, attachment and detachment of mobile stations, authentication functions, and logical link management. Associated with each SGSN is a location register that stores location information and user profiles for all users registered with that SGSN. In general, a GGSN can interface to an external packet network for several SGSNs, and an SGSN may route its packets to different GGSNs [Ebe01, Ste01].

The GPRS standard defines three classes of mobile stations, which differ in their capability for simultaneous use of packet- and circuit-switched data GSM (speech, data, and SMS) services. Class A mobiles fully support simultaneous use of GPRS and conventional GSM services. Class B mobiles can register simultaneously for both categories of services but can use only one of the service types at a time. Class C mobiles can connect to the network for only one category of service at a time.

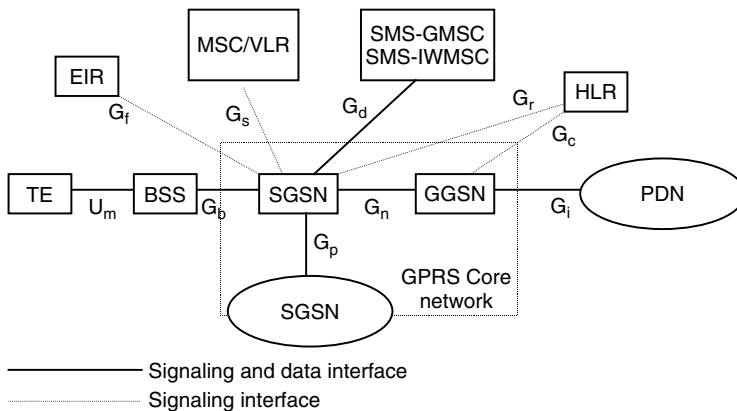


FIGURE 15.5 GPRS network architecture.

The transmission formatting for GPRS is based on much the same underlying structure as those found in GSM; however, there are important differences between the two. As in the GSM specification, each radio channel carries TDMA frames of duration 4.615 ms, each containing eight time slots. Also, as in GSM, *physical channels* are each defined by a combination of frequency channel and time slot for the uplink and downlink, and logical channels are mapped onto physical channels for carrying data traffic and for carrying signaling and control information. However, a number of new logical channels were defined for the GPRS standard. Also, many changes in channel structure were made relative to the channel structure defined for GSM circuit-switched services. These changes include a longer multiframe length (described below), new channel coding schemes, new power-control algorithms, and a *link adaptation* scheme designed to change the channel coding scheme according to conditions on the radio channel. It is beyond the scope of our treatment here to describe all of the details of the GPRS radio interface, but the reader may find these details in a number of recent texts [Ste01, Hal03, Seu03]. In the following paragraphs we briefly summarize the multiframe structure of a GPRS physical channel.

Figure 15.6 shows the structure of a 240-ms GPRS multiframe, which contains 52 TDMA frames, each having eight time slots. (A GPRS multiframe has the same duration as two GSM multiframes, each containing 26 TDMA frames; see Fig. 15.4.) The GPRS multiframe contains 12 blocks (B0 to B11) of four consecutive TDMA frames plus four idle frames. A physical channel is referred to as a *packet data channel* (PDCH), fully defined by a frequency channel together with one uplink time slot and one downlink time slot. On any given PDCH, blocks of four bursts, called *radio blocks*, form the logical channels, carrying either user data or signaling information. Each radio block consists of 456 bits. The packet data logical channels are mapped dynamically onto radio blocks B0 to B11 of the multiframe. The mapping can change from one radio block to another, controlled by parameters broadcast on a *packet broadcast control channel* (PBCCH). This dynamic mapping function allows the system to adapt to the traffic load by allocating or releasing resources as needed [Seu03].

An important feature of the GPRS specification is the provision of four *coding systems* (CSs), providing different levels of error protection. The lowest coding rate (greatest redundancy) is provided by CS-1 and the highest rate (no redundancy) by CS-4. A *link adaptation procedure* chooses the coding system in accordance with the measured radio transmission environment. The coding uses a *cyclic redundancy check* (CRC) *code* together with convolutional channel coding for CS-1 to CS-3. For CS-4, only CRC coding is used, providing error detection without error correction. Code puncturing (selective removal of parity bits) is applied to the convolutional encoding output to adjust the transmitted bit sequence to the radio block length. To illustrate the channel coding scheme, Fig. 15.7 shows the encoding of one radio block for coding systems CS-1 to CS-3.

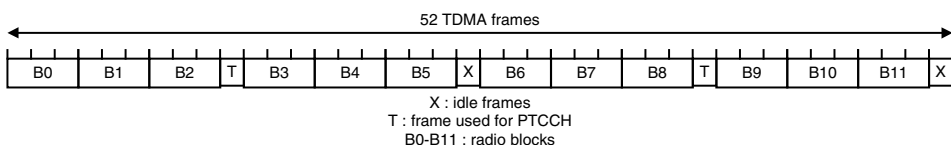


FIGURE 15.6 GPRS multiframe structure.

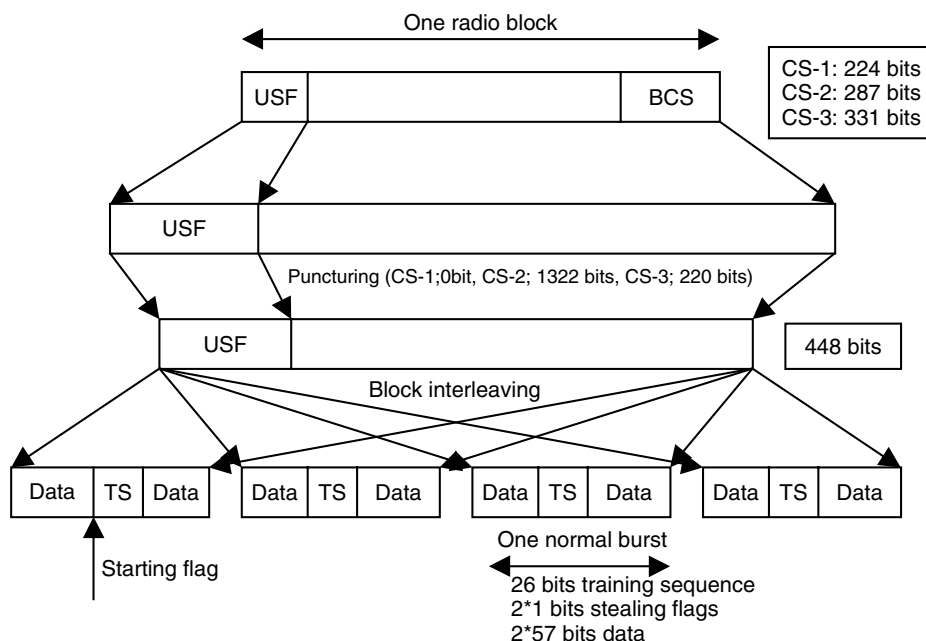


FIGURE 15.7 Radio block encoding for three GPRS coding schemes.

15.2.3 Enhanced Data for Global Evolution

Soon after the year 2000 introduction of GPRS, it was perceived by network operators that GSM networks would have to provide even higher data rates. The need for higher rates has grown out of the market demand for an expanding menu of services, such as multimedia transmission. In response to this market demand, the ETSI standardization organization defined a new family of data services, built on the existing structure of GPRS. This new family of data services was initially named enhanced data rates for GSM evolution, and subsequently renamed *enhanced data for global evolution* (EDGE). Although the primary motivation for EDGE development was enhancement of data services in GSM/GPRS networks, EDGE can also be introduced into networks built to the IS-136 (U.S. digital cellular) standard [Fur99]. In Europe, EDGE is considered a *2.5-generation* (2.5G) standard, providing a transition between second- and third-generation systems. As is the case with GPRS, a GSM network operator requires no new license to implement EDGE services in its network, since the 200-kHz RF channel organization of conventional GSM is reused with a different organization of logical channels within each RF channel. In this subsection we describe the implementation of EDGE in GSM/GPRS networks.

EDGE enhances data service performance over GPRS in two ways. First, it replaces the GMSK radio link modulation used in GSM with an 8-PSK modulation scheme capable of tripling the data rate on a single radio channel. Second, EDGE provides more reliable transmission of data using a *link adaption technique*, which dynamically chooses a modulation and coding scheme (MCS) in accordance with the current transmission conditions on the radio channel. The EDGE link adaptation mechanism is an enhanced version of the link adaptation mechanism used in GPRS.

EDGE provides two forms of enhanced data service for GSM networks: *enhanced circuit-switched data* (ECSD) for circuit-switched services, and *enhanced GPRS* (EGPRS) for packet-switched services. In each form of EDGE data service, there are provisions for combining logical channels (time slots) in the GSM transmission format to provide a wide menu of achievable data rates.

The major impact that EDGE has on the GSM/GPRS protocol structure is at the physical layer of the radio interface, and there is only minor impact on MAC and radio link control (RLC) layers [Hal03]. The EDGE RF specification and definition of burst structures are common to EGPRS and ECSD. A major modification in EGPRS, relative to GPRS, is the *link quality control* (LQC) function, which provides a wider array of link error-control schemes than exist in conventional GSM or in GPRS. The LQC function provides nine different modulation and coding schemes, MCS-1 to MCS-9, as well as related signaling and adaptation procedures for switching between MCSs in response to radio link conditions. Table 15.1 summarizes the nine MCSs used in EGPRS and shows the data throughput provided by each MCS. The EGPRS modulation and coding schemes are organized into three families, A to C, according to their RLC block sizes. For example, in family A, MCS-9 carries 1184 payload bits in two RLC data blocks during one EGPRS radio block consisting of four consecutive bursts. MCS-6 carries 592 payload bits in one RLC block, and MCS-3 carries 296 payload bits within the RLC block. In each family, the numbers of payload bits carried in the lower-rate coding schemes are submultiples of the payloads at higher rates. This feature enables efficient retransmission of negatively acknowledged RLC blocks when this functionality is needed following a sudden change in radio channel conditions. For a detailed discussion of EGPRS link adaptation and analysis of its performance, the reader is referred to [Hal03].

The ECSD service is built on the *high-speed circuit-switched data* (HSCSD) service standardized in the GSM Phase 2+ specification. Although the user data rates provided by ECSD (up to 64 kb/s) are not higher than those in HSCSD, the same data rates are achieved in ECSD by using smaller numbers of time slots and simpler implementation in the mobile stations. [Hal03] As was the case with HSCSD, ECSD provides both *transparent* (T) and *nontransparent* (NT) services. Transparent data service employs forward-error correction and is characterized by a fixed time delay in data delivery. Generally speaking, transparent service cannot assure a specified level of accuracy in delivered data. Nontransparent data service uses a *radio link protocol* (RLP), including

TABLE 15.1 EGPRS Modulation and Coding Parameters

Modulation and Coding Scheme	Code Rate	Modulation Type	Data Rate per Time Slot (kb/s)	MCS Family
MCS-9	1.00	8-PSK	59.2	A
MCS-8	0.92	8-PSK	54.4	A
MCS-7	0.76	8-PSK	44.8	B
MCS-6	0.49	8-PSK	29.6	A
MCS-5	0.37	8-PSK	22.4	B
MCS-4	1.00	GMSK	17.6	C
MCS-3	0.80	GMSK	14.8	A
MCS-2	0.66	GMSK	11.2	B
MCS-1	0.53	GMSK	8.8	C

TABLE 15.2 ECSD Service Data Rates

Coding Scheme ^a	Code Rate	Modulation Type	Gross Rate (kb/s)	Radio Interface Rate (kb/s)	User Rate (kb/s)
TCH/28.8 (NT/T)	0.419	8-PSK	69.2	29.0	28.8
TCH/32.0 (T)	0.462	8-PSK	69.2	32.0	32.0
TCH/43.2 (NT)	0.629	8-PSK	69.2	43.5	43.2

^aNT, nontransparent service; T, transparent service.

error detection and retransmission [termed *automatic repeat request* (ARQ)] to assure that the error rate in delivered data will be maintained below a specified level. The use of retransmission results in variable time delay in the delivered data. Table 15.2 shows the data rates provided by ECSD service in EDGE.

15.3 CDMA AND HDR

While the GSM standards group was developing the pan-European digital cellular standard, in North America, manufacturers and service providers, under the aegis of the Telecommunications Industry Association (TIA), were developing North American standards for digital cellular communications. In its initial work, the TIA settled on an FD/TDMA channel architecture with the 30-kHz carrier spacing of the legacy analog AMPS system, with TDMA used to multiplex multiple digital traffic channels onto each 30-kHz frequency channel. This system came to be known as the USDC TDMA standard, initially designated IS-54, evolving into IS-136. Subsequent to development of the USDC TDMA standard, a proposal was brought forward for another digital cellular system using CDMA technology [CTI89]. This initiative was based on a cellular system design developed by Qualcomm, Inc. [Gil91]. Proponents of the CDMA system projected that the system would provide a significantly larger capacity increase (over analog AMPS) than would be achieved with the USDC TDMA standard. Subsequently, the TIA established subcommittee TR45.5, which began developing a standard based on the Qualcomm CDMA design.

The resulting air-interface specification IS-95 was released in July 1993. A revised standard, IS-95A, was released in 1995 and first deployed in commercial service in 1996. The IS-95A standard provides the basis for many of the CDMA networks deployed throughout the world. In addition to digital voice service, the IS-95A CDMA networks provide circuit-switched data service at rates up to 14.4 kb/s. A subsequent revision of the standard, IS-95B, provided for packet-switched data services at rates up to 64 kb/s, termed *medium data rate* (MDR), in addition to voice service. Due to this increased data rate capability, IS-95B is referred to as a 2.5-generation technology. The IS-95B revision, deployed in 1999, also introduced improved handoff procedures that proved to be advantageous in congested CDMA markets [Lib99, Gar00a, Kim00, Tie01].

As the new digital cellular equipment became available, service providers in North America adopted both the TDMA and CDMA standards, some operators implementing one standard in a given market area and the other standard in another market area. However, as the cellular service market grew, subscribers and service providers showed

an increasing preference for the CDMA technology; better voice quality was typically cited as the deciding factor. Today, IS-95-based CDMA, usually referred to collectively as *cdmaOne*, is the dominant cellular technology in North America and has become a de facto worldwide standard as well. Also, *cdmaOne* has been accepted as the basis for cellular industry evolution to third-generation systems, including *cdma2000*. A key driver in the evolution toward third-generation systems is the requirement for data service rates significantly higher than those offered by second- and 2.5-generation systems. An important development proposed to meet this requirement is CDMA-based *high-data-rate* (HDR) service. In the following two subsections we describe the CDMA cellular system and HDR.

15.3.1 CDMA Cellular

In contrast with the FD/TDMA structure of the GSM and USDC standards, the CDMA cellular system uses a spread-spectrum signal with 1.2288-MHz spreading bandwidth, a frequency span equivalent to 41 AMPS channels. (The forward and reverse links are actually using two separate carrier frequencies spaced 45 MHz apart.) Clearly, this design did not offer USDC's advantage of selective replacement of analog AMPS channels. Instead, large blocks of AMPS analog channels were replaced at one time by the CDMA system.

Because in a CDMA system every user is a source of interference to every other user on the system, control of the mobile station power is a critical element of the system design. The CDMA system is a relatively sophisticated design, using several power-setting algorithms to optimize system operation [Sal91, Vit93]. In addition, the system uses (1) powerful error-correction coding with interleaving, (2) speech activity detection and variable-rate speech encoding, and (3) RAKE receiver techniques to maximize system capacity.

A simplified block diagram of the CDMA system is given by Fig. 15.8, which shows the signal-processing functions for transmission from base station to mobile unit. The underlying data rate for the system is 9600 bits/s, which represents the speech coder rate of 8550 bits/s augmented by error-correction coding which is tailored to the speech-coding technique. The speech coder specified in the IS-95 system is the Qualcomm code excited linear predictive (QCELP) coder. The QCELP coder actually detects speech activity and changes the data rate to lower values during quiet intervals, but 9600 bits/s is the maximum error-protected data rate. (A higher-rate speech coder, QCELP13, was introduced by Qualcomm in 1995. It encodes speech at 13.4 kb/s and tailored error coding brings the data rate up to 14.4 kb/s.)

The 9600-bit/s stream is segmented into 20-ms blocks and then encoded further with a $K = 9$, rate- $\frac{1}{2}$ convolutional code, bringing the data rate to 19.2 kb/s. (In the reverse direction, mobile to base, rate- $\frac{1}{3}$ convolutional coding is used, which brings the coded data rate up to 28.8 kb/s.) The convolutional encoding is followed by interleaving over each 20-ms interval for burst-error protection in the radio channel. The 19.2-kb/s stream is then modified by use of a *long code*, which serves as a privacy mask. The modified stream is then encoded for spread-spectrum transmission using Walsh codes [Pro89] of dimension 64. This produces a 64-fold spreading of the data stream, resulting in a transmitted bit rate of 1.2288 Mb/s. The structure of the Walsh code provides 64 orthogonal sequences, and one of the 64 sequences is assigned to a mobile station during call setup. In this way, 64 orthogonal "channels" are established by CDMA

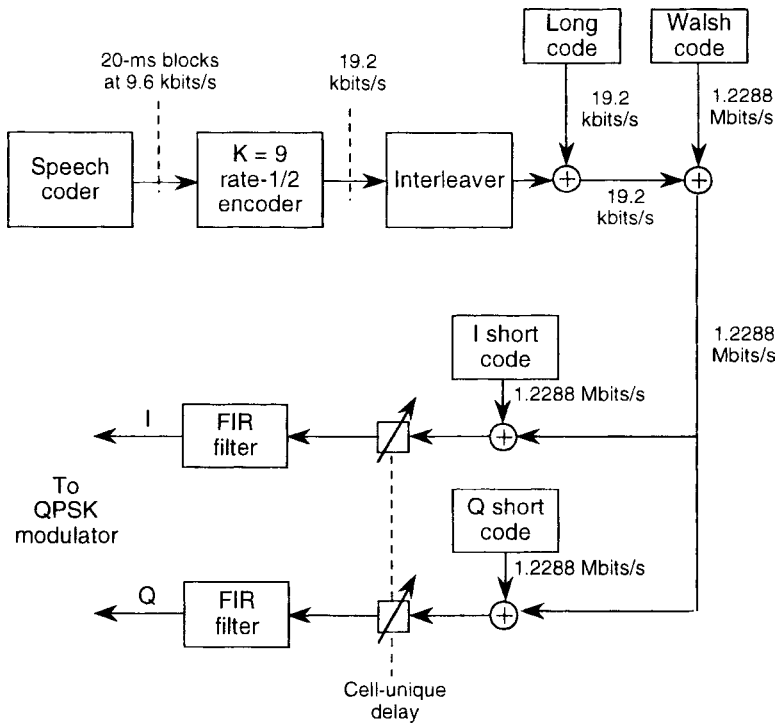


FIGURE 15.8 IS-95 CDMA cellular system (base to mobile).

encoding on the same carrier frequency. (The 64-kb/s data rate of IS-95B is achieved by aggregating up to eight different Walsh codes.) After Walsh encoding, the spread data stream is separated into I and Q streams, each of which is modified using a unique “short code.” The spread-spectrum stream is carried over the air interface with filtered QPSK modulation. In other cells the same carrier frequency is used again, with the same set of 64 Walsh codes, but each cell applies a unique time offset to the two short codes, which allows the CDMA channels in each cell to be uniquely identified. The same two short codes are used by all cell sites and all mobile stations.

When the signal is received, the short codes are removed, using the cell-specific time offset, and the channel-specific Walsh code is decoded by correlation decoding. The rate- $\frac{1}{2}$ convolutional code is then decoded, producing the received 9600-bit/s data stream. More detailed descriptions of the CDMA cellular system can be found in [Gil91, Vit95, Ste01].

15.3.2 HDR

The spread-spectrum structure of CDMA, together with its use of orthogonal Walsh channels, lends itself to the convenient implementation of multiple data rates and combinations of data and voice traffic in a given RF channel bandwidth. For these reasons, CDMA technology has formed the basis for development of standards for third-generation digital wireless systems. The principal objectives in the third-generation initiative are provision of substantially higher data rates than can be supported in second-generation

systems, and efficient support of multimedia communications. CDMA technology is highly suited to meeting these objectives.

The third-generation standard that has evolved directly from cdmaOne is termed cdma2000, and cdma2000 is in turn a component of a global standardization initiative called *International Mobile Telecommunication 2000* (IMT-2000), which will eventually integrate a wide range of digital terrestrial and satellite systems into one worldwide mobile system. Further details on the IMT-2000 initiative may be found in [Ste01]. In the IMT-2000 arena, a competitor to cdma2000 is *wideband CDMA* (W-CDMA), and the two families of standards have different spectrum requirements. Whereas cdma2000 in all its forms utilizes the 1.25-MHz channels of the cdmaOne system, W-CDMA requires a 5-MHz channel bandwidth. At this writing the cdma2000 technology appears to be gaining stronger support from network operators.

Many cdmaOne network operators are following a two-stage migration path to cdma2000. The first stage is implementation of cdma2000 1 × RTT (single-carrier radio transmission technology), which uses a 1.25-MHz CDMA channel to support packet-switched data service at speeds up to 144 kb/s and also doubles voice traffic capacity relative to cdmaOne networks. The second stage of migration is cdma2000 1 × EV-DO (single-carrier evolution, data optimized, or data only), which provides packet-switched data service at rates up to 2.4 Mb/s. The cdma2000 1 × RTT technology is usually categorized as 2.5 generation, since the achievable data rate is still significantly lower than objectives set for the third-generation. However, cdma2000 1 × EV-DO is unequivocally a first step of third-generation cellular technology.

The name *cdma2000 1 × EV-DO* is commonly used for the TIA/EIA Standard IS-856, which is based on Qualcomm's high-data-rate (HDR) technology [Ben00]. Thus, cdma2000 1 × EV-DO is also termed *high data rate* [Est00]. As the designation "1 ×" denotes, HDR operates in a single 1.25-MHz CDMA channel, and a network operator can implement HDR on any or all of the cell sites in the service network, with no change in cell site layout being required. HDR-modified base station equipment is required, however. While HDR operates in the 1.25-MHz channels of basic CDMA, it does not overlay HDR data traffic onto digital voice traffic on the same RF carrier frequency. Instead, it uses a pair of RF channels dedicated to HDR traffic. The dedicated-channel approach was adopted to optimize the 1 × EV-DO standard for data service. If the user requires voice service, a dual-mode mobile terminal is needed to access a separate RF channel. To emphasize this data-only aspect of HDR, terminals used for data-only applications are referred to as *access terminals* (ATs). Also, the IS-856 standard refers to base stations as *access networks* (ANs).

The forward HDR channel (AN to AT) signal is always transmitted at full power and a data rate control scheme, rather than power control, is used. The AN communicates with multiple ATs but with only one AT at a time, allocating transmission time to each AT on a demand and opportunity basis. As shown in Fig. 15.9, the HDR forward link is composed of slots of length 2048 chips, or 1.66 . . . ms at the channel chip rate of 1.2288 Mc/s [3GP00]. Groups of 16 slots are synchronized to system timing on even-second time ticks. The system uses GPS to maintain precise time synchronism.

Each time slot is divided into two half-slots, each of which is time-division multiplexed into MAC (128 chips), pilot (96 chips) and data (800 chips) fields. The data fields support either a forward traffic channel or a control channel, distinguishable at the AT by a preamble inserted into the data. Note that each pilot channel burst is

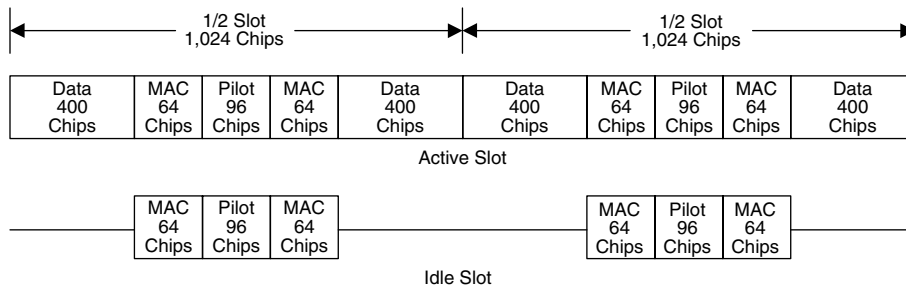


FIGURE 15.9 HDR forward link slot structure.

centered at the midpoint of a half-slot and between a pair of MAC channel bursts. The pilot channel bursts provide a reference in the AT for coherent demodulation of traffic and MAC channel signals and enable SNR estimation at the AT. All fields in a 2048-chip active slot are transmitted at the same power level.

The forward traffic channel is a packet-based variable-rate channel in which user data can be transmitted to an AT at data rates ranging from 38.4 kb/s to 2.4576 Mb/s. The transmission is organized into blocks called *physical layer packets*, and a packet may utilize 1 to 16 slots, depending on the data rate. When more than one slot is allocated, consecutive slots are separated by three intervening slots, which are occupied by slots of other physical layer packets transmitted to the same AT or to other ATs. For further details, the reader is referred to the 3GPP2 Web site.

To receive data, an AT measures SNR on the forward link pilot channel as often as every 1.66 . . . ms, that is, every slot time. Based on the SNR measurement, the AT requests a data rate by sending a data-rate control (DRC) message to the AN. Typically, the AN receives requests from a number of ATs and determines which ATs are to be served next, based on which terminals have reported the best SNR values. The DRC message indicates not only the data rate but also the modulation, Turbo-coding rate, preamble length, and the maximum number of time slots needed to transmit a physical layer packet to the AT. Table 15.3 shows the modulation parameters for a selection of HDR data rates on the forward link. The traffic channel utilizes 1600 data chips in every slot, as described earlier; however, in each physical layer packet, a number of chips is allocated to a preamble of length shown in the table.

TABLE 15.3 Selected HDR Data Rates on the Forward Link

Data Rate (kb/s)	Packet Length (bytes)	Code Rate	Modulation Type	Data Chips per Packet	Preamble Chips per Packet
38.4	128	1/5	QPSK	24,576	1,024
76.8	128	1/5	QPSK	12,288	512
153.6	128	1/5	QPSK	6,144	256
307.2	128	1/5	QPSK	3,072	128
614.4	128	1/5	QPSK	1,536	64
1,228.8	256	1/3	QPSK	1,536	64
1,843.2	384	1/3	8-PSK	1,536	64
2,457.6	512	1/3	16-QAM	1,536	64

The forward-channel user data stream is turbo-encoded, then scrambled with a PN sequence unique to each user. The data rate going to each user is adjusted by repetition of the channel-coded bits, lower rates using higher repetition factors. Code puncturing is also used to adjust some of the data rates. The coded data stream is block interleaved, then modulated using QPSK, 8-PSK, or 16-QAM, depending on the data rate, as indicated in Table 15.3. The resulting modulated data sequence has a chip rate of 1.2288 Mc/s, regardless of the data rate. The in-phase and quadrature channels are then time-demultiplexed into 16 separate streams, each running at 76.8 kc/s, each of which is covered by a 16-chip Walsh code sequence. The 16 Walsh-coded streams are then summed and spread by quadrature PN sequences, bandlimited, and upconverted. The resulting RF signal has the same characteristics as an IS-95 CDMA signal, thus allowing reuse of all analog and RF circuits developed for IS-95 (or cdmaOne) base stations [Ben00].

The control channel transmits broadcast messages and AT-directed messages at a data rate of 38.4 or 76.8 kb/s, using the same modulation parameters as given in Table 15.3 for traffic channels at those two rates. As noted earlier, control channel transmissions are distinguished from traffic channel transmissions by a preamble, which in a control channel is covered by a special biorthogonal cover sequence.

On the reverse HDR link (AT to AN), the data traffic channel can support five data rates ranging in doubling steps from 9.6 to 153.6 kb/s. The different data rates are implemented by varying the repetition factor. On the reverse link, there is no distinct control channel; instead, all signaling information is embedded within the data channel. Details of the HDR specifications can be found at the 3GPP2 Web site, http://www.3gpp2.org/Public_html/specs/.

15.4 OTHER HISTORICAL SYSTEMS

In the near-20-year history of evolution of digital wireless systems, there have been a number of developments that were either overtaken or supplanted by other systems and thus did not gain wide adoption in the wireless marketplace. In this section we describe briefly four of the more important of these developments. CT-2 and DECT originated as enhanced cordless phone systems, and the latter system found its marketplace position as a wireless PABX technology for office environments. USDC and PDC, TDMA-based cellular systems, were overtaken in the marketplace by CDMA technology.

15.4.1 CT2 Telepoint

CT2 is an early standard developed for second-generation cordless telephones. It began as a British initiative to provide digital wireless access to the public telephone network. The service is termed *telepoint service*. CT2 telepoint service was initiated commercially in the United Kingdom in September 1989, and subsequently in other European countries and the Far East. In European markets, CT2 was soon supplanted by GSM digital cellular service and by another digital cordless system, DECT. Consequently, in late 2001 the ECC made the decision to phase out CT2 license applications in the 900-MHz band and eventually to reallocate the CT2 frequencies to other radio services. However, CT2 is still in widespread use in Far East markets.

CT2 telepoint was the digital successor to the first generation of simple cordless telephones (termed CT-0 and CT-1), allowing use of a single handset in conjunction

with home, office, and public base stations. With this system, the user subscribes to public telepoint service and uses a CT2 phone to gain wireless connection to telepoint base stations, which are in turn connected to the switched telephone network. Charges are billed to the caller's home or office. Public base stations are located in airports, railway stations, and other heavily trafficked areas. The CT2 phone is able to make calls within a range of about 200 m of the base station. The CT2 phone user is able to initiate but not receive calls, and the system cannot provide call handoff to another base station.

The CT2 standard uses the 864- to 869-MHz band and provides 40 channels, with carrier frequencies spaced 100 kHz apart. The access method is frequency-division multiple access (FDMA); each carrier supports one call, with time-division duplex (TDD) for two-way conversation. The channel bit rate of CT2 is 72 kb/s, and its timing structure is relatively simple in comparison with wide-area digital cellular systems. The CT2 modulation technique is binary frequency shift keying (BFSK). With a channel spacing of 100 kHz, the bandwidth efficiency of CT2 is 0.72 bit/s per hertz. The speech-coding technique is standard adaptive differential pulse code modulation (ADPCM) operating at 32 kb/s [CCI84b].

15.4.2 DECT

The Digital European Cordless Telephone (DECT) standard was developed as a successor to CT2, differing significantly from CT2 in its level of sophistication and its range of services offered. DECT, developed by ETSI and finalized in mid-1992, is the first pan-European standard for digital cordless telephones. The DECT signal structure comprises 10 carriers, at 1.728-MHz spacing, each carrying 12 channels in a time-division multiple-access (TDMA) format. DECT, as does CT2, uses TDD to support a two-way conversation on the same carrier. However, the TDMA signal structure of DECT, in contrast with FDMA in CT2, provides power savings by being able to turn the signal off intermittently during a call when no information is being transmitted. DECT has been designed and specified to interwork with many other types of network, such as PSTN, ISDN, and GSM.

DECT is particularly suited for high-density, small-cell applications such as cordless private branch exchanges (PBXs) in a private organization that connect the telephones to the outgoing lines. A typical office, factory, or warehouse has many telephones distributed throughout the premises all connected by wires back to the on-premises switchboard, the PABX. The DECT wireless PABX provides an alternative that offers much greater flexibility: A small number of radio base stations are wired back to the PABX instead, and cordless handsets communicate, by radio, to these base stations. A key difference between DECT and CT2 systems is that DECT is sold as a complete system to a closed group of users rather than simply as a handset and service. Unlike CT2, DECT allows users to both receive and initiate calls and supports call handoff between base stations as well.

DECT operates in the frequency band 1.88 to 1.9 GHz, with 10 channels defined for 1881.792 to 1897.344 MHz with a carrier spacing of 1728 kHz. The DECT channel structure can be described as FDMA/TDMA/TDD. Each base station supports 12 duplex digital speech channels, and each time slot may occupy any of the DECT channels. The frame duration is 10 ms, with 5 ms allocated for each direction of communication. The transmitter transfers information in signal bursts on slots of duration

$10/24 = 0.417$ ms. With 480 bits/slot (including a 64-bit guard time), the total bit rate is 1.152 Mb/s. Each slot contains 64 bits for system control (C, P, Q, and M channels) and 320 bits for user information (I channel).

The modulation technique used in DECT is Gaussian minimum shift keying (GMSK), in common with GSM. The normalized bandwidth of the Gaussian filter is $BT = 0.5$, where B is the 3-dB bandwidth (in hertz) and T is the inverse of the bit rate. The bandwidth efficiency, 1.152 Mb/s in 1.728 MHz, is comparable to that of CT2. In common with CT2, the speech coder in DECT is ADPCM at 32 kb/s, transmitted without channel coding.

15.4.3 U.S. Digital Cellular

As noted earlier, in the late 1980s cellular network operators in the United States and Canada saw the need to replace their existing AMPS analog networks with a new cellular technology that would provide for significant growth in traffic capacity and would also provide new services and features for their cellular service customers. Thus two industry associations, the Electronic Industries Association and the Telecommunications Industry Association formed the EIA/TIA TR45.3 Cellular Standards Committee and charged it with developing a new standard for cellular telephone services. The natural choice for the new standard was a digital system, and the development of a new TDMA standard was accomplished in several steps. In 1990 the committee issued Interim Standard 54B (IS-54B), which defined a dual-mode system that allowed mobile phones to operate in digital mode in service areas where the new TDMA technology was installed but also to be able to operate in AMPS analog mode (FDMA) where the TDMA service had not yet been installed. The IS-54B standard was followed by the IS-136 standard, which built on the original IS-54B standard, adding enhancements, including digital control channels, enabling the addition of new services and features. The IS-54 and IS-136 systems have at different times been referred to collectively as *North American digital cellular* or *U.S. digital cellular* (USDC). In recent years, the name USDC has been in more prevalent use. A detailed description of the IS-136 system and its evolution from IS-54 may be found in [Har98].

As with GSM, the USDC radio channel structure is a combination of FDM and TDMA, with user traffic and control channels built on the logical channels provided by TDMA time slots. The designated frequency channels for this digital standard are the same as those in AMPS, with carriers spaced 30 kHz apart. The overall network architecture of the USDC system is much like that of GSM, with a cellular design supported by base stations and switching centers and also with roaming capability supported by home and visitor location registers. In USDC, each 30-kHz digital channel has a transmission rate of 48.6 kb/s. The 48.6-kb/s stream is divided into six TDMA channels of 8.1 kb/s each. In contrast with the GMSK modulation scheme used in GSM, the USDC standard uses $\pi/4$ -shift DQPSK, implemented with square-root RC filtering and rolloff parameter 0.35. The principal advantage of this modulation scheme is bandwidth efficiency. The channel transmission rate of 48.6 kb/s, taken together with a channel spacing of 30 kHz, yields channel utilization of 1.62 bits/s per hertz, a 20% advantage relative to GSM. The principal disadvantage of a linear modulation scheme is poorer power efficiency relative to constant-envelope modulation.

Following the development of the competing IS-95 CDMA standard and installation of the CDMA system in cellular service networks, the relative advantages of the CDMA

technology became increasingly apparent—higher traffic capacity limits and better quality of digital voice service—and the USDC technology was subsequently overtaken by the CDMA technology.

15.4.4 JDC/PDC

The *personal digital cellular* (PDC) system is a second-generation digital cellular system introduced in Japan in 1991 and deployed primarily in Japan. The system was originally named Japanese digital cellular, but the name was changed to personal digital cellular; the name Pacific digital cellular is also used for this system. PDC operates in the 800/900-MHz and 1.5-GHz bands. As in GSM and USDC, the radio channel structure for JDC is a combination of FDM and TDMA. For the 800/900-MHz band the uplink/downlink duplex offset is 130 MHz, whereas in the 1.5-GHz band the offset is 48 MHz. The carrier spacing is 25 kHz, which is compatible with the legacy analog system in Japan.

The air-interface standard for PDC is similar to that of USDC. The modulation method is $\pi/4$ -shift QPSK, with a rolloff factor of 0.5, and the channel bit rate is 42 kb/s. As in GSM and USDC, PDC provides full- and half-rate traffic channels. In full-rate mode there are three TDMA user channels per 20-ms frame; and in half-rate mode, six per frame. In the full-rate mode, there are 224 user data bits per 20-ms frame, for a transmitted bit rate of 11.2 kb/s, comprising coded voice bits and error-control coding. The full-rate speech coder is based on vector-sum excitation linear prediction (VSELP) operating at 6.7 kb/s. In the half-rate phase of implementation, the coded bit rate is 5.6 kb/s. More details of the PDC system may be found in [Nak90, Kin91].

The original PDC system supported circuit-switched data services, but a 28.8-kb/s packet data service (PDC-P) has also been introduced into the system. PDC has not been adopted to any significant extent outside Japan, and in Japan, as well as other Far East market areas, CDMA has become the preferred digital cellular technology.

15.5 WIRELESS LANs

In the early days of the WLAN industry, products were developed using several different wireless technologies, and many of those products were based on proprietary designs. However, the WLAN industry gradually shifted its attention to standards-based designs, giving support to the IEEE 802.11 standardization initiative [IEE99a]. Since the 1999 release of the original IEEE 802.11 standard, the WLAN industry has converged on the manufacture of interoperable WLAN products based on the evolving 802.11 family of standards [IEE99a, IEE99b]. Today, essentially all WLAN products are based on IEEE 802.11 standards, and in this section we discuss the architecture of 802.11-based networks and the basic protocol structure that underlies the several versions of the standard.

In addition to the IEEE 802.11 standards, there are other standards for WLANs, and an important example is the European HIPERLAN2 standard, developed by ETSI. The interested reader is referred to Chapter 2 in this book and to [Pah02a] and the ETSI HIPERLAN2 Web site, <http://portal.etsi.org/bran/kta/HIPERLAN/HIPERLAN2.asp>.

In Chapter 2 we briefly summarize the current principal versions of the IEEE 802.11 standards, 802.11b, 802.11a, and 802.11g. The original 802.11 standard provided a

maximum channel data rate of only 2 Mb/s, while the 802.11b standard supports a maximum data rate of 11 Mb/s. Subsequent versions of the standard, 802.11a and 802.11g, provide a range of data rates up to 54 Mb/s. All of the 802.11 standards operate in the ISM bands (902 to 908 MHz, 2.4 to 2.5 GHz, and 5.8 to 5.9 GHz). For more details on the evolution and parameters of the 802.11 standards, see Chapter 2.

The 802.11 standard provides a general PHY- and MAC-layer specification that can support any *connectionless* applications whose transport and network layers accommodate the IEEE 802.11 MAC layer. In connectionless operation, a communication path and its bandwidth are selected dynamically for each packet of exchanged information. Today, TCP/IP is the dominant transport/network layer protocol supporting all the ubiquitous connectionless applications, such as WWW access, e-mail, FTP, and Telnet, and the IEEE 802.11 standard is structured to allow TCP/IP to operate as efficiently over its MAC layer as it does over any wired-LAN MAC layer.

The IEEE 802.11 reference architecture defines two network topologies, (1) infrastructure network and (2) ad hoc network, illustrated in Fig. 15.10. In the infrastructure topology, wireless terminals are connected to a backbone network, typically an Ethernet LAN, which ties together various wired devices such as desktop computers, servers, and printers. Also attached to the backbone network are wireless *access points* (APs), each of which acts as a base station or bridge between the wireless terminals and the wired backbone. In the ad hoc topology, wireless terminals can communicate directly on a peer-to-peer basis. (See also the discussion of network topologies in Chapter 11.)

In IEEE 802.11 terminology, an AP provides access to *distribution services* over a wireless link. The *basic service area* (BSA), also called a *cell*, is the area covered by one AP. The *basic service set* (BSS) is a set of wireless terminals controlled by one AP. The distribution system (DS) is the fixed (wired) infrastructure used to connect a set of BSSs to create an *extended service set* (ESS). The IEEE 802.11 distribution services enable a wireless terminal to roam freely within an ESS and also allow an 802.11 WLAN to connect to the wired LAN infrastructure. The IEEE 802.11 standard also defines a *portal* as a logical point at which non-802.11 packets enter an ESS.

Figure 15.11 depicts the entities in the IEEE 802.11 protocol stack. The MAC layer is divided into the MAC sublayer and the MAC management sublayer. The MAC

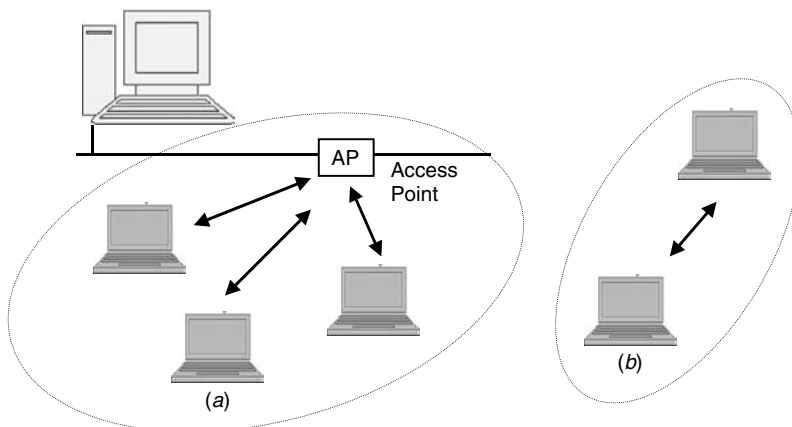


FIGURE 15.10 IEEE 802.11 network configurations: (a) infrastructure network; (b) ad hoc network.

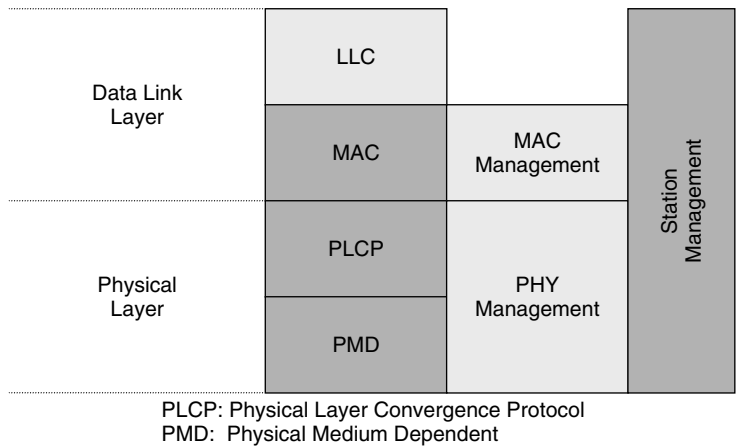


FIGURE 15.11 MAC and PHY layers in IEEE 802.11.

sublayer is responsible for the access mechanism and for fragmentation and reassembly of packets. The MAC-layer management sublayer is responsible for roaming within the ESS, power management, and association, dissociation, and reassociation processes for registration connection management. The PHY layer is divided into three sublayers: PHY-layer convergence protocol (PLCP), PHY medium-dependent (PMD) protocol, and PHY-layer management sublayer. The PLCP is responsible for carrier-sensing assessment and forming packets for various PHY layers. The PMD sublayer specifies the modulation and coding techniques to be used at the PHY layer, and PHY-layer management selects among the options available for each PHY layer. In addition, IEEE 802.11 defines a station management sublayer that is responsible for coordination of the interactions between MAC and PHY layers. More detailed treatments of the IEEE 802.11 protocol layer functionality can be found in [Gei99, O’Ha99, Kei02, Pah02a], and references cited therein.

The IEEE 802.11 standard currently defines three PHY-layer options: Frequency-hopping spread spectrum (FHSS), direct-sequence spread spectrum (DSSS), and infrared (IR). When the MAC protocol data units (MPDUs) arrive at the PLCP layer, a header is attached that is designed specifically for the PMD of the PHY layer chosen. The PLCP packet is then transmitted by the PMD according to the PHY-layer specification chosen. The standard defines three PLCP packet formats to prepare the MPDUs for transmission on the three respective PHY layers.

For illustrative purposes, Fig. 15.12 shows the composition of a PLCP frame for the FHSS PHY-layer option. The preamble enables the receiver to synchronize to the incoming signal properly before the actual frame contents arrive. The header field

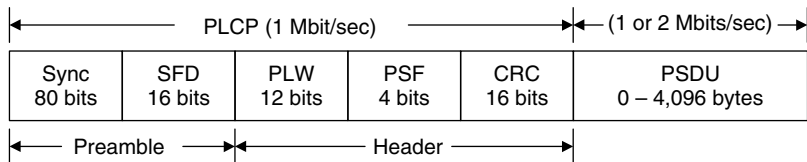


FIGURE 15.12 PLCP frame format for the FHSS PHY-layer option in IEEE 802.11.

provides information about the frame, and the PSDU (PLCP service data unit) is the MAC protocol data unit being sent. There are two data rates for transmission of the information: 1 and 2 Mb/s using binary and four-level Gaussian FSK (GFSK) modulation, respectively. The lower data rate provides a simpler signal reception condition for synchronization functions in the receiver. Thus, the PLCP header is always transmitted with the lower data rate using binary GFSK. Then the MPDU is transmitted using either of the two data rates. The remaining bits in the PLCP frame consist of the preamble and header. The preamble is an 80-bit sequence of alternating 0's and 1's that is used to extract the received clock for carrier and bit synchronization. The start of frame delimiter (SFD) is a specific pattern of 16 bits, as shown in the figure. The header has three fields. The 12-bit packet length width (PLW) gives the packet length, which can be up to 4 kB. The 4-bit packet-signaling field (PSF) gives the data rate in 0.5-Mb/s steps starting with 1 Mb/s. The 16-bit CRC frame check field protects the PLCP bits by being able to detect up to two errors in the PLCP header.

More detailed descriptions of the IEEE 802.11 frame formats at various layers of the protocol stack may be found in references cited earlier.

15.6 SPEECH CODING IN WIRELESS SYSTEMS

In any wireless system that provides digital voice service, there is need for a speech-coding process that will encode analog speech into a digital stream for transmission over the wireless medium and then reconstitute an analog speech signal that the listener will find acceptable. In systems designed to serve large numbers of users, such as cellular service networks, a proper balance must be struck between bandwidth requirements and the perceived quality of voice service. The high cost of spectrum licenses and infrastructure equipment creates the need for compressing the encoded speech signal to as low a bit rate as is feasible to maximize the number of simultaneous voice calls that the network can support. However, compressing to too low a bit rate can result in characteristics that the listener perceives as “artificial” or “synthetic.” Therefore, a great deal of research has been done in the field of speech coding, with emphasis in recent years on wireless applications, to develop speech coding and synthesis techniques that can operate at progressively lower bit rates without sacrificing voice quality. The interested reader may refer to a number of books [Fla72, Rab78, Jay84, Chu03] and survey articles [Fla79, Cro83, IEE88b, Jay90, IEE92] on this wide-ranging subject. Here we do not attempt to cover this subject in any great depth, but instead simply summarize speech-coding techniques that have been adopted into the leading standards-based wireless systems. Other treatments of speech coding for mobile radio applications can be found in a number of articles, including [Cox91, Ste93], and in several texts [Ste92, Ata93].

The key parameters of concern in the choice of a speech-coding technique for application in a wireless system are (1) the transmitted bit rate, (2) the delivered speech quality, (3) robustness in the presence of transmission errors, and (4) complexity of implementation. The relative importance of each of these parameters will vary from one category of system to another.

The transmitted bit rate for a speech-coding technique is an important element in determining the overall spectral efficiency for a wireless system. As already noted, this is a particularly important issue in the design of mobile and cellular systems. Thus,

these systems have adopted some of the most complex of the speech-coding algorithms, techniques that operate at bit rates of 13 kb/s and below. In some systems, where spectral efficiency is less important than voice quality and implementation complexity, 32-kb/s coders have been used. Table 15.4 shows the speech-coding techniques, with transmission rates, used in a number of second-generation cellular and cordless systems. There are two bit rates associated with a speech-coding technique: the basic bit rate of the speech coder, termed “uncoded” in the table, and aggregate bit rate, including error-protection coding applied to the coded voice bits, and repetition, when it is used.

In networks where spectral efficiency is judged to be less important than voice quality (this was the perception in the early days of PCS system development), higher-rate voice coders such as adaptive differential pulse code modulation (ADPCM) have been employed. Resistance to channel errors is also an important issue. Voice-coding algorithms may exhibit poor performance when the channel error rates are as high as 10^{-2} . For this reason, some error-control coding is applied to some of the most vulnerable bits in low-bit-rate voice coders. The voice compression scheme removes redundancy in the digitized voice signal, and the error-control scheme introduces structured redundancy, resulting in better received voice quality for a given aggregate data rate. The appropriate balance between speech compression and error-control coding will be different for each speech-coding technique.

The most prominent speech-coding techniques are *waveform coding* techniques such as pulse code modulation (PCM) and ADPCM; *model-based* techniques such as linear predictive coding (LPC), regular-pulse excitation (RPE), and code-excited linear prediction (CELP) techniques; and hybrid schemes. Waveform coding techniques offer relatively low complexity with very high voice quality, but the transmitted bit rates make them unattractive in wireless networks, where spectral efficiency is a high priority. LPC techniques can provide good voice quality with bit rates as low as 2400 bits/s, in contrast with the 64 kb/s rate of PCM, but at the cost of a high computational burden. The GSM system uses a version of RPE, called RPE-LTP, that has an acceptable implementation complexity and latency, operating at 13 kb/s with a speech frame duration of 20 ms. At still lower bit rates, the quality of RPE-LTP coded voice is judged inadequate, and CELP techniques are preferred. Vector-sum excitation linear prediction (VSELP) and Qualcomm’s CELP (QCELP) are used in the USDC and CDMA cellular standards, respectively, both operating at about 8 kb/s [Pah02a].

TABLE 15.4 Speech Coders in Selected Wireless Systems

System	Application	Voice Coder	Uncoded Rate (kb/s)	Overall Rate (kb/s)
GSM	European digital cellular	RPE-LTP	13	22.8
GSM	European digital cellular (half-rate)	VSELP	6.5	11.4
IS-136	U.S. digital cellular	VSELP	8	13
JDC	Japanese digital cellular	VSELP	8	13
IS-95	CDMA digital cellular	QCELP	9.6, 4.8, 2.4, 1.2	28.8, 19.2 ^a
DCS-1800	PCS in the United States	RPE-LTP	13	22.8
CT-2	European cordless	ADPCM	32	32
DECT	Cordless and WPBX	ADPCM	32	32

^aThe IS-95 overall rates include FEC coding and repetition.

QUESTIONS

- (a) How does the bandwidth efficiency of GSM compare with that of USDC?
- (b) Although many aspects of USDC are similar to those of GSM, the transmission bandwidth per TDMA channel is 30 kHz, more than six times smaller than that of GSM. Explain why this difference came about.
- (c) How does GSM interface to ISDN and the PSTN?
- (d) What are the basic blocks of a mobile radio network?
- (e) What is the advantage of defining standardized interfaces within a wireless network?
- (f) How many interfaces are defined by the GSM standard, and what are they?
- (g) Which system, GSM or USDC, is more vulnerable to multipath fading effects? Explain.
- (h) What are the frame size and the number of TDMA slots per frame in GSM, and how do they compare with the USDC format?
- (i) What are the differences among the speech coding rates in GSM, USDC, and JDC?
- (j) How many layers of coding are used in the Qualcomm CDMA system? What is the type and the purpose of each of these coding layers?
- (k) What is the spreading factor of the codes used in the Qualcomm CDMA system?
- (l) What is the maximum data rate supportable by each channel in the Qualcomm CDMA system?
- (m) Compare the speech coding rate, access method, and transmission power of the CT2 Telepoint service with the GSM digital cellular system.
- (n) From a user's point of view, what are the differences between CT2 service and GSM service?
- (o) Compare the speech coding rate, access method, transmission power, and the frequency of operation of the CT2 and DECT systems.
- (p) From a user's point of view, what are the differences between CT2 service and DECT service?
- (q) Discuss the differences between the DECT and DCS-900 systems.
- (r) Name two major applications for circuit-switched data services.
- (s) What is the most popular transmission technology for wireless LANs?
- (t) What is the major drawback in using spread spectrum for wireless LANs?
- (u) What transmission technologies are expected to evolve in unlicensed data PCS bands?
- (v) Name the two major technical specifications for wireless LANs.

- (w) What is the highest instantaneous data rate that can be provided to a mobile terminal if only a single time slot on a single GSM channel is available?
- (x) What is the highest instantaneous data rate that can be provided to a mobile terminal using IS-95 CDMA if four channels are assigned to a single user?
- (y) Explain the differences between circuit- and packet-switched data services.
- (z) Compare the advantages and disadvantages, from a user's perspective, of data services provided by a dedicated mobile-data network versus data services integrated with voice service in a digital cellular network.
- (aa) Discuss the principal technical differences among ECSDS, GPRS, and EDGE, all of which are overlaid onto GSM networks.
- (bb) Search the Web for technical information on the CDPD system, which was designed as a packet-data overlay to analog AMPS networks. Summarize the principal technical characteristics of CDPD.
- (cc) Explain the differences between waveform coding and model-based block coding as speech coding techniques.
- (dd) Search the Web for the latest information on the status of development of the IEEE 802.11n standard. What are the key technical issues and principal proposals that are being considered by that working group?
- (ee) Compare the PLCP frame compositions for all the PHY layer options in the IEEE802.11 specification.
- (ff) Suppose that in a particular geographic area, one cellular service network uses cdmaOne, while another uses the USDC standard. Explain how a call connection is supported between two customers, one in each network.

REFERENCES

- [3GPP00] 3GPP2 Project, cdma2000 High Rate Packet Data Air Interface Specification, Version 2, October 27, 2000.
- [Abr70] N. Abramson, The ALOHA system: another alternative for computer communications, *Proc. AFIPS, Fall Joint Computer Conference*, Vol. 37, Houston, TX, 1970, pp. 281–285.
- [Abr73] N. Abramson, The ALOHA system, in *Computer-Communication Networks*, N. Abramson and F. Kuo (eds.), Prentice Hall, Englewood Cliffs, NJ, 1973.
- [Abr77] N. Abramson, The throughput of packet broadcasting channels, *IEEE Trans. Commun.*, **COM-25**, 117–128 (1977).
- [Aca87] A. S. Acampora and J. H. Winters, A wireless network for wide-band indoor communications, *IEEE J. Sel. Areas Commun.*, **SAC-5**, 796–805 (1987).
- [Aka87] Y. Akaiwa and Y. Nagata, Highly efficient digital mobile communications with a linear modulation method, *IEEE J. Sel. Areas Commun.*, **SAC-5**, 890–895 (1987).
- [Ake88] D. Akerberg, Properties of a TDMA pico-cellular office communication system, *Proc. IEEE GLOBECOM '88*, Hollywood, FL, Dec. 1988, pp. 1343–1349.
- [Ala98] S. Alamouti, A simple transmit diversity technique for wireless communications, *IEEE J. Sel. Areas Commun.*, Oct. 1998, pp. 1451–1458.
- [Ala03a] B. Alavi and K. Pahlavan, Modeling of the distance error for indoor geolocation, *Proc. IEEE WCNC '03*, New Orleans, LA, Mar. 2003.
- [Ala03b] B. Alavi and K. Pahlavan, Bandwidth effect on distance error modeling for indoor geolocation, *Proc. IEEE PIMRC '03*, Beijing, China, Sept. 2003.
- [Ale82] S. E. Alexander, Radio propagation within buildings at 900 MHz, *IEE Electron. Lett.*, **18**, 913–914 (1982).
- [Ale83] S. E. Alexander, Characterizing buildings for propagation at 900 MHz, *IEE Electron. Lett.*, **19**, 860 (1983).
- [Ali94] M. H. Ali, A. S. Parker, and K. Pahlavan, Time and frequency domain modeling of wideband radio propagation for personal, mobile and indoor applications, *Proc. IEEE PIMRC '94*, The Hague, The Netherlands, 1994.
- [Ali98] M. Hassan-Ali, Using ray-tracing techniques in site-specific statistical modeling of indoor radio channels, Ph.D. dissertation, Worcester Polytechnic Institute, Worcester, MA, May 1998.
- [Ali02] M. H. Ali and K. Pahlavan, A new statistical model for site-specific indoor radio propagation prediction based on geometric optics and geometric probability, *IEEE J. Sel. Areas Commun.*, Special Issue on Wireless Networks, Jan. 2002.

- [Alo01] M.-S. Alouini and M. K. Simon, Performance of generalized selection combining over Weibull fading channels, *Proc. IEEE Vehicular Technology Conference*, Vol. 3, Atlantic City, NJ, Oct. 7–11, 2001, pp. 1735–1739.
- [Als03] N. Alsindi, Performance of TOA estimation algorithms in different indoor multipath conditions, M.S. thesis, Worcester Polytechnic Institute, Worcester, MA, May 2003.
- [Als04] N. Alsindi, X. Li, and K. Pahlavan, Performance of TOA estimation algorithms in different indoor multipath conditions, *Proc. IEEE WCNC '04*, New Orleans, LA, Apr. 2004.
- [Amo80] F. Amoroso, The bandwidth of digital data signals, *IEEE Commun. Mag.*, **18**(6), 13–24 (1980).
- [And93] H. R. Anderson, A ray-tracing propagation: model for digital broadcast systems in urban areas, *IEEE Trans. Broadcast.*, **AB-39**, 309–317 (1993).
- [And94] J. B. Anderson, T. S. Rappaport, and S. Yoshida, Propagation measurements and models for wireless communications channels, *IEEE Commun. Mag.*, Nov. 1994.
- [Ank80] H. A. Ankerman, Transmission of audio signals by infrared light carrier, *Soc. Motion Picture and Television Engrs. J.*, **89**, 834–837 (1980).
- [ANS85] ANSI/IEEE Std. 802.3-1985, *Carrier Sense Multiple Access with Collision Detection (CSMA/CD)*, 1985.
- [ANS88] ANSI/IEEE Std. 802.3a,b,c,e-1988, *Supplement to Carrier Sense Multiple Access with Collision Detection (CSMA/CD)*, 1988.
- [Arn87] J. Arnbak and W. Blitterswijk, Capacity of slotted ALOHA in Rayleigh-fading channels, *IEEE J. Sel. Areas Commun.*, **SAC-5**, 261–269 (1987).
- [Arn89] H. W. Arnold, R. L. Murray, and D. C. Cox, 815 MHz radio attenuation measured within two commercial buildings, *IEEE Trans. Antennas Propag.*, **AP-37**, 1335–1339 (1989).
- [Arr73] G. A. Arredondo, W. H. Chriss, and E. H. Walker, A multipath fading simulator for mobile radio, *IEEE Trans. Commun.*, **COM-21**, 1325–1328 (1973).
- [Ata93] B. S. Atal et al., *Speech and Audio Coding for Wireless and Networking Applications*, Kluwer, Norwell, MA, 1993.
- [Bah99] A. Bahai and B. R. Saltzberg, *Multicarrier Digital Communications: Theory and Applications of OFDM*, Kluwer, Norwell, MA, 1999.
- [Bah00a] P. Bahl and V. Padmanabhan, RADAR: an in-building RF-based user location and tracking system, *Proc. IEEE INFOCOM '00*, Israel, Mar. 2000.
- [Bah00b] P. Bahl, V. N. Padmanabhan, and A. Balachandran, *Enhancements to the RADAR User Location and Tracking System*, Tech. Rep. MSR-TR-00-12, Microsoft Research, Feb. 2000.
- [Bar86] P. J. Barry and A. G. Williamson, Radiowave propagation into and within a building at 927 MHz, *IEE Electron. Lett.*, **22**, 248–249 (1986).
- [Bar91a] J. R. Barry et al., Simulation of multipath impulse response for indoor diffuse optical channel, *Proc. IEEE Workshop on Wireless LANs*, Worcester, MA, 1991, pp. 81–87.
- [Bar91b] J. R. Barry et al., High-speed nondirective optical communication for wireless networks, *IEEE Network Mag.*, Nov. 1991, pp. 44–54.
- [Bel63a] P. A. Bello, Characterization of randomly time-variant linear channels, *IEEE Trans. Commun. Syst.*, **CS-11**, 360–393 (1963).
- [Bel63b] P. A. Bello, and B. D. Nelin, The effect of frequency selective fading on the binary error probabilities of incoherent and differentially coherent matched filter receivers, *IEEE Trans. Commun. Syst.*, **CS-11**, 170–186 (1963).
- [Bel65] P. A. Bello, Selective fading limitations of the Kathryn modem and some system design considerations, *IEEE Trans. Commun. Technol.*, **CT-13**, 320–333 (1965).
- [Bel69] P. A. Bello, A troposcatter channel model, *IEEE Trans. Commun. Technol.*, **CT-17**, 130–137 (1969).

- [Bel79b] C. A. Belfiore and J. H. Park, Jr., Decision-feedback equalization, *Proc. IEEE*, **67**, 1143–1156 (Aug. 1979).
- [Bel84] P. A. Bello and K. Pahlavan, Adaptive equalization for SQPSK and SQPR over frequency selective microwave LOS channels, *IEEE Trans. Commun.*, **COM-32**, 609–615 (1984); also presented at *IEEE ICC '82*.
- [Bel88] P. A. Bello, Performance of some RAKE modems over the non-disturbed wide band HF channel, *IEEE MILCOM '88 Conf. Rec.*, San Diego, CA, Oct. 1988, pp. 89–95.
- [Ben99] J. Beneat, K. Pahlavan, and P. Krishnamurthy, Radio channel characterization for indoor and urban geolocation at different frequencies, *Proc. IEEE PIMRC '99*, Osaka, Japan, Sept. 1999.
- [Ben00] P. Bender et al., CDMA/HDR: a bandwidth efficient high-speed wireless data service for nomadic users, *IEEE Commun. Mag.*, **38**(7), 70–77 (2000).
- [Ber87] D. Bertsekas and R. Gallager, *Data Networks*, Prentice Hall, Upper Saddle River, NJ, 1987.
- [Ber93] C. Berrou, A. Glavieux, and P. Thitimajshima, Near Shannon limit error-correcting coding and decoding: turbo codes, *Proc. International Conference on Communications*, Geneva, 1993, pp. 1064–1070.
- [Bha81] V. K. Bhargava et al., *Digital Communications by Satellite-Modulation, Multiple Access and Coding*, Wiley, New York, 1981.
- [Bha03] V. Bhatia, W. Y. Chen, and L. T. Metreud, Design of a technique for deployment of IEEE 802.11b/g WLANs, undergraduate MQP project, Worcester Polytechnic Institute, Worcester, MA, Fall 2003.
- [Big91] E. Biglieri et al., *Introduction to Trellis-Coded Modulation with Applications*, Macmillan, New York, 1991.
- [Bin75] R. Binder, A dynamic packet switching system for satellite broadcast channels, *Proc. IEEE ICC '75*, San Francisco, June 1975.
- [Bin88] J. Bingham, *The Theory and Practice of Modem Design*, Wiley, New York, 1988.
- [Bin90] J. A. C. Bingham, Multicarrier modulation for data transmission: an idea whose time has come, *IEEE Commun. Mag.*, **28**(5), 5–14 (1990).
- [Bla91] U. Black, *The V Series Recommendations*, McGraw-Hill, New York, 1991.
- [Bla03] R. Blahut, *Algebraic Codes for Data Transmission*, Cambridge University Press, Cambridge, 2003.
- [Blu00] Bluetooth Special Interest Group, Specifications of the Bluetooth System, Vol. 1v.1.1, “Core,” and Vol. 2v.1.0B “Profiles” (2000).
- [Bod84] D. Bodson, G. F. McClure, and S. R. McConoughey (eds.), *Land-Mobile Communications Engineering*, Selected Reprint Series, IEEE Press, New York, 1984.
- [Bra75] K. Brayer (ed.), *Data Communications via Fading Channels*, Selected Reprint Series, IEEE Press, Piscataway, NJ, 1975.
- [Bra80] E. Braun and S. Schon, A cordless infrared telephone, *Telecom Rep.*, **3**, 83–86 (1980).
- [Bro93] G. Bronson, K. Pahlavan, and H. Rotithor, Performance prediction of wireless LANs based on ray tracing algorithm, *Proc. PIMRC '93*, Yokohama, Japan, Sept. 1993, pp. 151–156.
- [Bud97] K. C. Budka, H. J. Jiang, and S. E. Sommars, Cellular packet data networks, *Bell Syst. Tech. J.*, Summer 1997.
- [Bul47] K. Bullington, Radio propagation above 30 megacycles, *Proc. IRE*, **35**, 1122–1136 (1947).
- [Bul77] K. Bullington, Radio propagation for vehicular communications, *IEEE Trans. Veh. Technol.*, **VT-26**, 295–308 (1977).
- [Bul87] R. J. C. Bultitude, Measurement, characterization and modeling of indoor 800/900 MHz radio channels for digital communications, *IEEE Commun. Mag.*, **25**(6), 5–12 (1987).

- [Bul89] R. J. C. Bultitude, S. A. Mahmoud, and W. A. Sullivan, A comparison of indoor radio propagation characteristics at 910 MHz and 1.75 GHz, *IEEE J. Sel. Areas Commun.*, **SAC-7**(1), 20–30 (1989).
- [Bur83] W. D. Burside and K. W. Burgener, High frequency scattering by thin lossless dielectric slab, *IEEE Trans. Antennas Propag.*, **AP-31**, 104–110 (1983).
- [Caf98] J. Caffery and G. Stuber, Subscriber location in CDMA cellular networks, *IEEE Trans. Veh. Technol.*, **VT-47**(2), 406–416 (1998).
- [Cai97] J. Cai and D. Goodman, General packet radio services in GSM, *IEEE Comput. Soc. Mag.*, Oct. 1997.
- [Cas90] E. Casas and C. Leung, A simple digital fading simulator for mobile radio, *IEEE Trans. Veh. Technol.*, **VT-39**, 205–212 (1990).
- [Cas93] G. Castagnoli, S. Brauer, and M. Hermann, Optimization of cyclic redundancy check codes with 24 and 32 parity bits, *IEEE Trans. Commun.*, **COM-41**, 883–892 (1993).
- [Cas02] D. Cassioli, M. Z. Win, and A. F. Molisch, The ultra-wide bandwidth indoor channel: from statistical model to simulations, *IEEE J. Sel. Areas Commun.*, **SAC-20**(6), 1247–1256 (2002).
- [CCI84a] CCITT Recommendation V.32, *A Family of 2-Wire, Duplex Modems Operating at Data Signalling Ranges of Up to 9600 Bit/s for Use on the General Switched Telephone Network and on Leased Telephone-Type Circuits*, Málaga, Torremolinos, Spain, 1984.
- [CCI84b] CCITT, *32-kb/s Adaptive Differential Pulse Code Modulation (ADPCM)*, Recommendation G.721, fascicle III.3, Málaga, Torremolinos, Spain, 1984.
- [CDP93] *Cellular Digital Packet Data Specification*, Release 1.0, Version 1.0, CDPD, Kirkland, WA, July 1993.
- [Cha72] D. Chase, A class of algorithms for decoding block codes with channel measurement information, *IEEE Trans. Inf. Theory*, **IT-18**, 170–182 (1972).
- [Cha75] D. Chase and P. A. Bello, A combined coding and modulation approach for high speed data transmission over troposcatter channel, *Proc. NTC '75*, New Orleans, LA, Dec. 1975, pp. 28–32.
- [Cha82] K. A. Chamberlain and R. J. Leubbers, An evaluation of Longley–Rice and GTD propagation models, *IEEE Trans. Antennas Propag.*, **AP-30**, 1093–1098 (1982).
- [Cha91] L. F. Chang, Throughput estimation of ARQ protocols for a Rayleigh fading channel using fade- and interfade-duration statistics, *IEEE Trans. Veh. Technol.*, **40**, 223–229 (1991).
- [Cha93] M. Chase and K. Pahlavan, Performance of DS-CDMA over measured indoor radio channels using random orthogonal codes, *IEEE Trans. Veh. Technol.*, **VT-42**, 617–624 (1993).
- [Che82] P. Chevillat and G. Ungerboeck, Optimum FIR transmitter and receiver filters for data transmission over band-limited channels, *IEEE Trans. Commun.*, **COM-30**, 1909–1915 (1982).
- [Che99a] P.-C. Chen, Mobile position location estimation in cellular systems, Ph.D. dissertation, WINLAB, Electrical and Computer Engineering Department, Rutgers University, New Brunswick, NJ, 1999.
- [Che99b] P.-C. Chen, A non-line-of-sight error mitigation algorithm in location estimation, *Proc. IEEE WCNC '99*, New Orleans, LA, 1999.
- [Cho85] G. Choudhury and S. S. Rappaport, Priority access schemes using CSMA-CD, *IEEE Trans. Commun.*, **COM-33**, 620–626 (1985).
- [Cho90] A. Chouly and H. Sari, Coding for multicarrier digital radio systems, *Proc. IEEE ICC '90*, Apr. 1990, pp. 1755–1759.
- [Chu87] T. S. Chu and M. J. Gans, High speed infrared local wireless communication, *IEEE Commun. Mag.*, **25**(8), 4–10 (1987).

- [Chu03] W. C. Chu, *Speech Coding Algorithms: Foundation and Evolution of Standardized Coders*, Wiley, New York, 2003.
- [Cia82] S. Ciarcia, Use infrared communication for remote control, *Byte*, Apr. 1982, pp. 40–49.
- [Cla68] R. H. Clarke, A statistical theory of mobile-radio reception, *Bell Syst. Tech. J.*, **47**, 957–1000 (1968).
- [Cla81] G. C. Clark, Jr., and J. B. Cain, *Error-Correction Coding for Digital Communications*, Plenum Press, New York, 1981.
- [Com84] R. A. Comroe and D. J. Costello, Jr., ARQ schemes for data transmission in mobile radio systems, *IEEE Trans. Veh. Technol.*, **VT-33**, 88–97 (1984).
- [Con78] W. J. Conner, ANTRC-170, a new digital troposcatter communication system, *Proc. IEEE ICC '78*, Toronto, 1978, pp. 40–43.
- [Coo78] G. R. Cooper and R. W. Nettleton, A spread spectrum technique for higher capacity mobile radio, *IEEE Trans. Veh. Technol.*, **VT-27**, 264–275 (1978).
- [COS86] CAST 207 TD(86)51-REV 3 (WG1), *Propagation on Channel Transfer Functions to Be Used in GSM Tests*, Sept. 1986.
- [Cox72] D. C. Cox, Delay Doppler characteristics of multipath propagation at 910 MHz in a suburban mobile radio environment, *IEEE Trans. Antennas Propag.*, **AP-20**, 625–635 (1972).
- [Cox91] R. V. Cox et al., Subband speech coding and matched convolutional channel coding for mobile radio channels, *IEEE Trans. Signal Proc.*, **39**, 1717–1731 (1991).
- [Cox93] D. C. Cox, 910 MHz urban radio propagation: multipath characteristics in New York City, *IEEE Trans. Veh. Technol.*, **VT-42**, 104–110 (1993).
- [Coy83] E. J. Coyle and B. Liu, Finite population CSMA/CD networks, *IEEE Trans. Commun.*, **COM-31**, 1247–1251 (1983).
- [Coy85] E. J. Coyle and B. Liu, A matrix representation of CSMA/CD networks, *IEEE Trans. Commun.*, **COM-33**, 53–64 (1985).
- [Cro73] W. Crowther, R. Rettberg, D. Walden, S. Ornstein, and F. Heart, A system for broadcast communication: reservation ALOHA,” *Proc. 6th Hawaii International Conference on Systems Science*, Jan. 1973, pp. 371–374.
- [Cro83] R. E. Crochiere and J. L. Flanagan, Current perspectives in digital speech, *IEEE Commun. Mag.*, **21**(1), 32–40 (1983).
- [CTI89] CTIA CDMA Digital Cellular Technology Open Forum, June 6, 1989.
- [Dha02] N. Al-Dhahir, C. Fragouli, A. Stamoulis, W. Younis, and R. Calderbank, Space-time processing for broadband wireless access, *IEEE Commun. Mag.*, **40**(9), 136–142 (2002).
- [Dea98] T. Dean, P. Fleming, and A. Stolyar, Estimates of multicarrier CDMA system capacity, *Soc. for. Comp. Sim., Proc. 1998 Winter Simulation Conference*, D. J. Medeiros, E. F. Watson, J. S. Carson, and M. S. Manivannan (eds.), Washington, DC.
- [Den96] L. R. Dennison, BodyLAN: A wearable personal network, Second Wireless Workshop on WLANS, Worcester, MA, 1996.
- [Des72] G. A. Deschamps, Ray techniques in electromagnetics, *Proc. IEEE*, **60**, 1022–1035 (1972).
- [Dev84] D. M. J. Devasirvatham, Time delay spread measurements of wideband radio systems within a building, *Electron. Lett.*, **20**, 50–51 (1984).
- [Dev87] D. M. J. Devasirvatham, A comparison of time delay spread and signal level measurements within two dissimilar office buildings, *IEEE Trans. Antennas Propag.*, **AP-35**, 319–324 (1987).
- [Dev91] D. M. J. Devasirvatham, Multi-frequency propagation measurement and models in a large metropolitan commercial building for personal communication, *Proc. PIMRC '91*, London, Sept. 1991, pp. 98–103.

- [Dig82] Digital Equipment, Intel, and Xerox Corporations, *The Ethernet: A Local Area Network: Data Link Layer and Physical Layer Specifications*, Version 2.0, Nov. 1982.
- [Div87] D. Divsilar and M. K. Simon, Trellis coded modulation for 4800–9600 bits/s transmission over a fading mobile satellite channel, *IEEE J. Sel. Areas Commun.*, **SAC-5**, 162–175 (1987).
- [Div88a] D. Divsilar and M. K. Simon, The design of trellis coded MPSK for fading channels: performance criteria, *IEEE Trans. Commun.*, **COM-36**, 1013–1022 (1988).
- [Div88b] D. Divsilar and M. K. Simon, The design of trellis coded MPSK for fading channels: set partitioning for optimum code design, *IEEE Trans. Commun.*, **COM-36**, 1004–1012 (1988).
- [Dix84] R. C. Dixon, *Spread Spectrum Systems*, 2nd ed., Wiley, New York, 1984.
- [Dos92] B. T. Doshi and P. K. Johri, Communication protocols for high speed packet networks, *Comput. Networks ISDN Syst.*, **24**, 243–273 (1992).
- [Dum94] L. Dumont, M. Fattouche, and G. Morrison, Super-resolution of multipath channels in a spread spectrum location system, *Electron. Lett.*, **30**, 1583–1584 (Sept. 1994).
- [Due95] A. Duel-Hallen, J. Holtzman, and Z. Zvonar, Multiuser detection for CDMA systems, *IEEE Pers. Commun.*, April 1995, pp. 46–58.
- [Eav77] R. E. Eaves and A. H. Levesque, Probability of block error for very slow Rayleigh fading in Gaussian noise, *IEEE Trans. Commun.*, **COM-25**, 368–374 (1977).
- [Eav87] J. L. Eaves and E. K. Reedy, *Principles of Modern Radar*, Van Nostrand Reinhold, New York, 1987.
- [Ebe01] J. Eberspacher et al., *GSM Switching, Services and Protocols*, 2nd ed., Wiley, New York, 2001.
- [EIA94a] EIA/TIA, *Cellular System Dual-Mode Mobile Station-Base Station Compatibility Standard*, IS-54, Rev. C, Electronic Industries Association, Washington, DC, June 1994.
- [EIA94b] EIA, *Service Description for IS-54 Data Services*, Electronic Industries Association, Washington, DC, June 1994.
- [Eln86] S. M. Elnoubi, Analysis of GMSK with discriminator detection in mobile radio channels, *IEEE Trans. Veh. Technol.*, **VT-35**, 71–76 (1986).
- [Enn98] Grey Ennis, *Impact of Bluetooth on 802.11 Direct Sequence*, Doc: IEEE P802.11-98/319, September 15, 1998.
- [Erc03] V. Erceg et al., *Indoor MIMO WLAN Channel Models*, IEEE 802.11-03/161r0, June 11, 2003.
- [Est00] E. Esteves, The high data rate evolution of the cdma2000 cellular system, in *Multiaccess, Mobility and Teletraffic for Wireless Communications*, Vol. 5, Kluwer, Norwell, MA, 2000, pp. 61–72.
- [ETS00] ETSI, *Universal Mobile Telecommunications System (UMTS): Multiplexing and Channel Coding (FDD)*, 3GPP TS 125.212 Version 3.4.0, Sept. 23, 2000, pp. 14–20.
- [Fal76a] D. D. Falconer, Analysis of a gradient algorithm for simultaneous passband equalization and carrier phase recovery, *Bell Syst. Tech. J.*, **55**, 409–428 (1976).
- [Fal76b] D. D. Falconer and F. R. Magee, Evaluation of decision feedback equalization and Viterbi algorithm detection for voiceband data transmission, part I, *IEEE Trans. Commun.*, **COM-24**, 1130–1139 (1976).
- [Fal76c] D. D. Falconer and F. R. Magee, Evaluation of decision feedback equalization and Viterbi algorithm detection for voiceband data transmission, part II, *IEEE Trans. Commun.*, **COM-24**, 1238–1245 (1976).
- [Fal85] D. D. Falconer et al., Comparison of DFE and MLSE receiver performance on HF channels, *IEEE Trans. Commun.*, **COM-33**, 484–486 (1985).

- [Fal96] A. Falsafi, K. Pahlavan, and G. Yang, Transmission techniques for radio LAN's—a comparative performance evaluation using ray tracing, *IEEE Journal on Selected Areas in Communications*, **14**(3), 477–491 (1996).
- [Fed94] G. Fedele, Error probability for diversity detection of binary signals over Nakagami fading channels, *Proc. IEEE PIMRC '94*, Vol. 2, The Hague, The Netherlands, Sept. 1994, pp. 607–611.
- [Feh87] K. Feher (ed.), *Advanced Digital Communications-Systems and Signal Processing Techniques*, Prentice Hall, Englewood Cliffs, NJ, 1987.
- [Feh91] K. Feher, MODEMS for emerging digital cellular-mobile radio system, *IEEE Trans. Veh. Technol.*, **VT-40**, 355–365 (1991).
- [Fei99] J. Feigin and K. Pahlavan, Measurement of characteristics of voice over IP in a wireless LAN environment *IEEE International Workshop on Mobile Multimedia Communications (MoMuC '99)*, 1999, pp. 236–240.
- [Fei00] J. Feigin, K. Pahlavan, and M. Ylianttila, Hardware-fitted modeling and simulation of VoIP over a wireless LAN, *52nd IEEE VTS Fall VTC, Vehicular Technology Conference*, Vol. 3, 2000, pp. 1431–1438.
- [Fer80] P. Ferert, Application of spread spectrum radio to wireless terminal communications, *Proc. NTC '80*, Houston, TX, Dec. 1980, pp. 244–248.
- [Fla72] J. L. Flanagan, *Speech Analysis, Synthesis and Perception*, Springer-Verlag, New York, 1972.
- [Fla79] J. L. Flanagan et al., Speech coding, *IEEE Trans. Commun.*, **COM-27**, 710–737 (1979).
- [Foc02] D. Focken and R. Stiefelhagen, Towards vision-based 3-D people tracking in a smart room, *Proc. ICMI*, Pittsburgh, PA, 2002.
- [Foe02] J. Foerster and Q. Li, *UWB Channel Modeling Contribution from Intel*, Working Group for Wireless Personal Area Networks (WPANs), IEEE P802.15-02/279r0-SG3a, June 24, 2002.
- [Fon01] R. Fontana, Advances in ultrawideband indoor geolocation systems, presented at the 3rd IEEE Workshop on WLAN, Boston, Sept. 2001.
- [For72] G. D. Forney, Jr., Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference, *IEEE Trans. Inf. Theory*, **IT-18**, 363–378 (1972).
- [For73] G. D. Forney, Jr., The Viterbi algorithm, *Proc. IEEE*, **61**, 268–278 (1973).
- [For84] G. D. Forney, Jr., et al., Efficient modulation for band-limited channels, *IEEE J. Sel. Areas Commun.*, **SAC-2**, 632–647 (1984).
- [For95] S. Fortune et al., WISE design of indoor wireless systems: practical computations and optimization, *IEEE Computational Science and Engineering*, **2**(1), 58–68 (1995).
- [Fos85] G. J. Foschini, Equalizing without alternating or detecting data, *Bell Syst. Tech. J.*, **64**, 1885–1911 (1985).
- [Fos99] G. J. Foschini et al., Simplified processing for high spectral efficiency wireless communication employing multi-element arrays, *IEEE J. Sel. Areas Commun.*, **SAC-17**(11) (1999).
- [Fra74] L. E. Franks and J. P. Bubrous, Statistical properties of timing jitter in a PAM timing recovery scheme, *IEEE Trans. Commun.*, **COM-22**, 913–920 (1974).
- [Fra80] L. A. Fratta and D. Sant, Some models of packet radio networks with capture, *Proc. 5th International Conference on Computer Communications*, Atlanta, GA, 1980, pp. 155–161.
- [Fre91a] T. A. Freeburg, Enabling technologies for wireless in-building network communications: four technical challenges, four solutions, *IEEE Commun. Mag.*, **29**(4), 58–64 (1991).
- [Fre91b] T. Freeburg, A new technology for high speed wireless local area networks, *Proc. IEEE Workshop on Wireless LANs*, Worcester, MA, May 1991, pp. 127–139.
- [FS93] *Proposed Federal Standard 1102 for Narrowband Digital Land Mobile Radio*, Jan. 1993.
- [Fur99] A. Furuskar et al., EDGE: enhanced data rates for GSM and TDMA/136 evolution, *IEEE Pers. Commun.*, **6**, 56–66 (1999).

- [Fus90] M. Fusco, FDTD algorithm in curvilinear coordinates, *IEEE Trans. Antennas Propag.*, **AP-38**, 76–88 (1990).
- [Gan87] R. Ganesh and K. Pahlavan, Effects of retransmission and capture for local area ALOHA systems, *Proc. Conference on Information Science Systems*, Baltimore, MD, 1987, pp. 272–273.
- [Gan89] R. Ganesh and K. Pahlavan, On the arrival of the paths in multipath fading indoor radio channels, *IEE Electron. Lett.*, **25**, 763–765 (1989).
- [Gan91a] R. Ganesh, Time domain measurements, modeling, and simulation of the indoor radio channel, Ph.D. dissertation, Worcester Polytechnic Institute, Worcester, MA, May 1991.
- [Gan91b] R. Ganesh and K. Pahlavan, Modeling of the indoor radio channel, *IEE Proc. I: Commun. Speech Vision*, **138**, 153–161 (1991).
- [Gan92] R. Ganesh and K. Pahlavan, Statistical characterization of a partitioned indoor radio channel, *IEE Proc. I: Commun. Speech Vision*, **139**, 539–545 (1992).
- [Gan93] R. Ganesh and K. Pahlavan, Statistics of short time and spatial variations measured in wideband indoor radio channels, *IEE Proc. H: Microwave Antennas Propag.*, **140**, 297–302 (1993).
- [Gar00] A. Leon-Garcia and I. Widjaja, *Communication Networks: Fundamental Concepts and Key Architectures*, McGraw-Hill, Boston, 2000.
- [Gei99] J. Geier, *Wireless LANs: Implementing Interoperable Networks*, Macmillan, New York, 1999.
- [Ges03] D. Gesbert, M. Shafi, D. Shiu, P. J. Smith, and A. Naguib, From theory to practice: an overview of MIMO space–time coded wireless systems, *IEEE J. Sel. Areas Commun.*, **SAC-21**(3), 281–302 (2003).
- [Gev89] J. Gevargiz, P. K. Das, and L. B. Milstein, Adaptive narrowband interference rejection in a DS spread spectrum intercept receiver using transform domain signal processing techniques, *IEEE Trans. Commun.*, **COM-37**, 1359–1366 (1989).
- [Gfe79a] F. R. Gfeller, H. R. Miller, and P. Vettiger, Infrared communications for in-house applications, *Proc. IEEE COMPCON '79*, Washington, DC, 1979, pp. 132–138.
- [Gfe79b] F. R. Gfeller and U. Bapst, Wireless in-house data communication via diffuse infrared radiation, *Proc. IEEE*, **67**, 1474–1486 (1979).
- [Gfe81] F. R. Gfeller, *Infranet: Infrared Microbroadcasting Network for In-House Data Communication*, IBM Res. Rep. RZ 1068 (#38619), Apr. 27, 1981.
- [Gfe82] F. R. Gfeller, Infrared microbroadcasting for in-house data communications, *IBM Tech. Discl. Bull.*, **24**, 4043–4046 (1982).
- [Gha03a] S. S. Ghassemzadeh, L. J. Greenstein, A. Kavèia, T. Sveinsson, and V. Tarokh, UWB indoor path loss model for residential and commercial buildings, *Proc. IEEE Vehicular Technology Conference*, Lake Buena Vista, FL, Fall 2003, pp. 3115–3119.
- [Gha03b] S. S. Ghassemzadeh, L. J. Greenstein, A. Kavèia, T. Sveinsson, V. Tarokh, UWB indoor delay profile model for residential and commercial buildings, *Proc. IEEE Vehicular Technology Conference*, Lake Buena Vista, FL, Fall 2003, pp. 3120–3125.
- [Gha04] M. Ghavami, L. B. Michael, and R. Kohno, *Ultra-wideband Signals and Systems in Communication Engineering*, Wiley, Hoboken, NJ, 2004.
- [Gil91] K. S. Gilhousen et al., On the capacity of a cellular CDMA system, *IEEE Trans. Veh. Technol.*, **VT-40**, 303–312 (1991).
- [Git73] R. D. Gitlin, E. Y. Ho, and J. E. Mazo, Passband equalization for differentially phase-modulated data signals, *Bell Syst. Tech. J.*, **52**, 219–238 (1973).
- [Git82] R. D. Gitlin, H. C. Meadors, and S. B. Weinstein, The tap-leakage algorithm: an algorithm for the stable operation of a digitally implemented, fractionally-spaced adaptive equalizer, *Bell Syst. Tech. J.*, **61** 1817–1839 (1982).

- [Git92] R. D. Gitlin, J. F. Hayes, and S. B. Weinstein, *Data Communications Principles*, Plenum Press, New York, 1992.
- [Gla89] A. S. Glassner, *An Introduction to Ray Tracing*, Academic Press, San Diego, CA, 1989.
- [God80] D. Godard, Self-recovering equalization and carrier tracking in two-dimensional data communication systems, *IEEE Trans. Commun.*, **COM-28**, 1867–1875 (1980).
- [Goo87] D. J. Goodman and A. A. M. Saleh, The near/far effect in local ALOHA radio communications, *IEEE Trans. Veh. Technol.*, **VT-36**, 19–27 (1987).
- [Goo89] D. J. Goodman, R. A. Valenzuela, K. T. Gayliard, and B. Ramamurthi, Packet reservation multiple access for local wireless communications, *IEEE Trans. Commun.*, **COM-37**, 885–889 (1989).
- [Goo90] S. H. Goode, H. L. Kazecki, and D. W. Dennis, A comparison of limiter-discriminator, delay and coherent detection for $\pi/4$ -QPSK, *Proc. 40th IEEE Vehicular Technology Conference*, Orlando, FL, May 1990, pp. 687–694.
- [Goo91b] D. J. Goodman and S. X. Wei, Efficiency of packet reservation multiple access, *IEEE Trans. Veh. Technol.*, **VT-40**, 170–176 (1991).
- [Grz75] C. J. Grzenda, D. R. Kern, and P. Monsen, Megabit digital troposcatter subsystem, *Proc. IEEE NTC '75*, New Orleans, LA, Dec. 1975, pp. 28–15 to 28–19.
- [GSM89] GSM Recommendation 06.10, *GSM Full Rate Speech Transcoding*, Tech. Rep. Version 3.2, ETSI/GSM, July 1989.
- [GSM91] GSM Recommendation 05.05, *Radio Transmission and Reception*, ETSI/PT 12, Jan. 1991.
- [GSMweb] <http://www.gsmworld.com/>.
- [Hab88] I. M. Habab, M. Kavehrad, and C.-E. W. Sundberg, ALOHA with capture over slow and fast fading radio channels with coding and diversity, *IEEE J. Sel. Areas Commun.*, **SAC-6**, 79–88 (1988).
- [Hal03] T. Halonen, J. Romero, and J. Meloro, *GSM, GPRS and EDGE Performance: Evolution Toward 3G/UMTS*, 2nd ed., Wiley, Hoboken, NJ, 2003.
- [Ham86] J. L. Hammond and P. J. P. O'Reilly, *Performance Analysis of Local Computer Networks*, Addison-Wesley, Reading, MA, 1986.
- [Ham02] M. Hämäläinen, V. Hovinen, R. Tesi, J. H. J. Iinatti, and M. Latvaaho, On the UWB system coexistence with GSM900, UMTS/WCDMA, and GPS, *IEEE J. Sel. Areas Commun.*, **SAC-20**(9), 1712–1721 (2002).
- [Han02] L. Hanzo, T. H. Liew, and B. L. Yeap, *Turbo Coding, Turbo Equalization and Space Time Coding*, Wiley, Hoboken, NJ, 2002.
- [Han03] L. Hanzo et al., *Single- and Multi-Carrier DS-CDMA: Multi-User Detection, Space-Time Spreading, Synchronisation, Standards and Networking*, IEEE Press, Piscataway, NJ, and Wiley, Hoboken, NJ, June 2003.
- [Har92] P. H. Harms, J. F. Lee, and R. Mittra, A study of the nonorthogonal FDTD method versus the conventional FDTD technique for computing resonant frequencies of cylindrical cavities, *IEEE Trans. Microwave Theory Tech.*, **MTT-40**, 741–746 (1992).
- [Har93] S. Hara et al., Multicarrier modulation techniques for broadband indoor wireless communications, *Proc. PIMRC '93*, Yokohama, Japan, Sept. 1993, pp. 132–136.
- [Har97] S. Hara and R. Prasad, Overview of multicarrier CDMA, *IEEE Commun. Mag.*, Dec. 1997, pp. 126–133.
- [Har98] L. J. Harte, A. D. Smith and C. A. Jacobs, *IS-136 TDMA Technology, Economics and Services*, Artech House, Norwood, MA, 1998.
- [Har99] S. Hara and R. Prasad, Design and performance of multicarrier CDMA system in frequency-selective Rayleigh fading channels, *IEEE Trans. Veh. Technol.*, **VT-48**(5), 1584–1595 (1999).

- [Has79] H. Hashemi, Simulation of the urban radio propagations, *IEEE Trans. Veh. Technol.*, **VT-28**, 213–225 (1979).
- [Has93a] H. Hashemi, The indoor radio propagation channel, *Proc. IEEE*, **81**, 943–968 (1993).
- [Has93b] H. Hashemi, Impulse response modeling of indoor radio propagation channels, *IEEE J. Sel. Areas Commun.*, **SAC-11**, 967–978 (1993).
- [Has98] M. Hassan-Ali, Using ray-tracing techniques in site-specific statistical modeling of indoor radio channels, Ph.D. dissertation, Worcester Polytechnic Institute, Worcester, MA, 1998 (supervised by the PI).
- [Has02] M. Hassan-Ali and K. Pahlavan, A new statistical model for site-specific indoor radio propagation prediction based on geometric optics and geometric probability, *IEEE J. Sel. Areas. Commun. Wireless*, **1**(1) (2002).
- [Hat80] M. Hata, Empirical formula for propagation loss in land-mobile radio services, *IEEE Trans. Veh. Technol.*, **VT-29**, 317–325 (1980).
- [Hat04] A. Hatami and K. Pahlavan, In-building intruder detection for WLAN access”, presented at the IEEE Aerospace and Electronic Systems Society Conference, PLANS, Monterey, CA, Apr. 2004.
- [Hau94] T. Haug, Overview of GSM: philosophy and results, *Int. J. Wireless Inf. Networks*, **1**, 7–16 (1994).
- [Hay91a] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, Upper Saddle River, NJ, 1991.
- [Hed97] R. Heddergott, B. H. Fleury, and U. P. Bernhard, Stochastic radio channel model for advanced indoor mobile communication systems, *Proc. IEEE PIMRC '97*, 1997.
- [Hee99] C. Heegard and S. B. Wicker, *Turbo Coding*, Kluwer, Norwell, MA, 1999.
- [Hei98] R. Heile, WPAN functional requirement, Doc. IEEE 802.11/98/58, January 22, 1998.
- [Hei99] G. Heine, *GSM Networks: Protocols, Terminology, and Implementation*, Artech House, Norwood, MA, 1999.
- [Hey86] D. P. Heyman, The effects of random message sizes on the performance of the CSMA/CD protocol, *IEEE Trans. Commun.*, **COM-34**, 547–553 (1986).
- [Hil01] A. Hills, Large scale wireless LAN design, *IEEE Commun. Mag.*, Nov. 2001, pp. 98–105.
- [Hir79] K. Hirade et al., Error-rate performance of digital FM with differential detection in land mobile radio channels, *IEEE Trans. Veh. Technol.*, **VT-28**, 204–212 (1979).
- [Hir84] M. Hirono, T. Miki, and K. Murota, Multilevel decision method for band-limited digital FM with limiter discriminator detection, *IEEE Trans. Veh. Technol.*, **VT-33**, 114–122 (1984).
- [Ho94] C. M. P. Ho, T. S. Rappaport, and M. P. Koushik, Antenna effects on indoor obstructed wireless channel and a deterministic image-based propagation model for in-building personal communication systems, *Int. J. Wireless Inf. Networks*, **1**, 61–76 (1994).
- [Hol83] R. Holland, Finite difference solutions of Maxwell’s equations in generalized nonorthogonal coordinates, *IEEE Trans. Nucl. Sci.*, **NS-30**, 4689–4691 (1983).
- [Hol92a] T. Holt, K. Pahlavan, and J. F. Lee, A graphical indoor radio channel simulator using 2D ray tracing, *Proc. PIMRC '92*, Boston, Oct. 1992, pp. 411–416.
- [Hol92b] T. Holt, A computer graphic package for indoor radio channel simulation using a 2D ray tracing algorithm, M.S. thesis, Worcester Polytechnic Institute, Worcester, MA, May 1992.
- [Hol92c] T. Holt, K. Pahlavan, and J. F. Lee, A computer package for indoor channel simulation using a 2D ray tracing algorithm, presented at the 17th IEEE Annual Conference on Local Computer Networks, Minneapolis, MN, Sept. 1992.
- [Hol01] H. Holter, On the capacity of the OFDM channel: a tutorial introduction, NORSIG’01, *Proc. IEEE Norwegian Symposium on Signal Processing*, Trondheim, Norway, Oct. 18–20,

- 2001, pp. 167–172; http://ilab0.ux.his.no/norsig/finalPapers/57.Capacity_of__1992001154555.pdf.
- [Hom00] HomeRF Special Interest Group (SIG) presentations (2000).
- [Hon92] W. Honcharenko, H. L. Bertoni, J. L. Dailing, J. Qian, and H. D. Yee, Mechanisms governing UHF propagation on single floors in modern office buildings, *IEEE Trans. Veh. Technol.*, **VT-41**, 496–504 (1992).
- [Hoo03] K. Hooli, Equalization in WCDMA terminals, Ph.D. dissertation, University of Oulu, Oulu, Finland, Dec. 2003.
- [How90a] S. Howard and K. Pahlavan, Doppler spread measurements of the indoor radio channel, *IEE Electr. Lett.*, **26**, 107–109 (1990).
- [How90b] S. J. Howard and K. Pahlavan, Autoregressive modeling of the indoor radio channel, *IEE Electr. Lett.*, **26**, 816–817 (1990).
- [How90c] S. J. Howard and K. Pahlavan, Measurement and analysis of the indoor radio channel in the frequency domain, *IEEE Trans. Instrum. Meas.*, **IM-39**, 751–755 (1990).
- [How91] S. J. Howard, Frequency domain characteristic and autoregressive modeling of the indoor radio channel, Ph.D. dissertation, Worcester Polytechnic Institute, Worcester, MA, May 1991.
- [How92] S. J. Howard and K. Pahlavan, Autoregressive modeling of wideband indoor radio propagation, *IEEE Trans. Commun.*, **COM-40**, 1540–1552 (1992).
- [IEE00] IEEE 802.11b standard groups document, IEEE 802.11-00/282r2, 2000.
- [IEE04] IEEE 802.11 standard group document, IEEE 802.11-03/940r4, May 2004.
- [IEE88a] IEEE Vehicular Technology Society, Special Issue on Mobile Radio Propagation, *IEEE Trans. Veh. Technol.*, **VT-37**, 3–72 (1988).
- [IEE88b] IEEE, Special issue on voice coding for communications, *IEEE J. Selected Areas in Commun.*, **6**, 225–456 (1988).
- [IEE92] IEEE, Special issue on speech and image coding, *IEEE J. Selected Areas in Commun.*, **10**, 793–976 (1992).
- [IEE99a] IEEE Working Group 802.11, *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, 1999.
- [IEE99b] IEEE Working Group 802.11a, *High Speed Physical Layer (PHY) in the 5-GHz Band*, 1999.
- [Ike91] F. Ikegami, T. Takeuchi, and S. Yoshida, Theoretical prediction of mean field strength for urban mobile radio, *IEEE Trans. Antennas Propag.*, **AP-39**, 229–302 (1991).
- [Iye93] V. Iyengar and J. Michaelides, *Performance Evaluations of RLPs (Radio Link Protocol) for TDMA Data Services*, TIA Contrib. TR45.3.2.5/93.03.30.10, Chicago, Mar. 30, 1993.
- [Jai01] R. Jain, A. Puri, and R. Sengupta, Geographical routing using partial information for wireless ad hoc networks, *IEEE Pers. Commun.*, Feb. 2001, pp. 48–57.
- [Jak74] W. C. Jakes (ed.), *Microwave Mobile Communications*, Wiley, New York, 1974.
- [Jay84] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*, Prentice Hall, Englewood Cliffs, NJ, 1984.
- [Jay90] N. S. Jayant, High-quality coding of telephone speech and wideband audio, *IEEE Commun. Mag.*, **28**(1), 10–20 (1990).
- [Jen65] F. A. Jenkins and H. E. White, *Fundamentals of Optics*, 4th ed., McGraw-Hill, New York, 1965.
- [Jer92] M. C. Jeruchim, P. Balaban, and S. Shanmugan, *Simulation of Communication Systems*, Plenum Press, New York, 1992.
- [Joh93] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing*, Prentice Hall, Upper Saddle River, NJ, 1993.

- [JTC94] Joint Technical Committee for PCS T1 R1P1.4, *Technical Report on RF Channel Characterization and System Deployment Modeling*, JTC(AIR)/94.09.23-065R6, Sept. 1994.
- [Kaa99] V.-P. Kaasila and A. Mämmelä, Bit error probability of a matched filter in a Rayleigh fading multipath channel in the presence of interpath and intersymbol interference, *IEEE Trans. Commun.*, **COM-47**(6), 809–812 (1999).
- [Kah78] R. E. Kahn, S. A. Gronemeyer, J. Burchfiel, and R. C. Kunzelman, Advances in packet radio technology, *Proc. IEEE*, **66**, 1468–1496 (1978).
- [Kam81] M. A. Kamil, Simulation of digital communication through urban/suburban multipath, Ph.D. dissertation, EECS Department, University of California, Berkeley, CA, June 1981.
- [Kan04a] M. Kanaan and K. Pahlavan, A comparison of wireless geolocation algorithms in the indoor environment, *Proc. IEEE WCNC '04*, New Orleans, LA, Apr. 2004.
- [Kan04b] M. Kanaan and K. Pahlavan, CN-TOA: a new algorithm for indoor geolocation, *Proc. IEEE PIMRC*, Sept. 2004.
- [Kap96] E. D. Kaplan, *Understanding GPS: Principles and Applications*, Artech House, Norwood, MA, 1996.
- [Kav81] M. Kavehrad, An accessing technique for information packet networks, *Proc. IEEE EASCON '81*, pp. 42–45.
- [Kav85] M. Kavehrad and P. McLane, Performance of low-complexity channel coding and diversity for spread spectrum in indoor wireless communication, *Bell Syst. Tech. J.*, **64**, 1927–1965 (1985).
- [Kav87] M. Kavehrad and P. J. McLane, Spread spectrum for indoor digital radio, *IEEE Commun. Mag.*, **25**(6), 32–40 (1987).
- [Kei89] G. E. Keiser, *Local Area Networks*, McGraw-Hill, New York, 1989.
- [Kei02] G. Keiser, *Local Area Networks*, 2nd ed., McGraw-Hill, New York, 2002.
- [Kim00] K. I. Kim (Ed.), *Handbook of CDMA System Design, Engineering, and Optimization*, Prentice Hall, Upper Saddle River, NJ (2000).
- [Kin91] K. Kinoshita, M. Kuramoto, and N. Nakajima, Development of a TDMA digital cellular system based on the Japanese standard, *Proc. 41st IEEE Vehicular Technology Conference*, St. Louis, MO, 1991, pp. 642–649.
- [Kle74] L. Kleinrock and F. A. Tobagi, Carrier sense multiple access for packet switched radio channels, *Proc. IEEE ICC '74*, Minneapolis, MN, June 1974, pp. 21B–7 to 21B-7.
- [Kle75b] L. Kleinrock and S. S. Lam, Packet switching in a multiaccess broadcast channel: performance evaluation, *IEEE Trans. Commun.*, **COM-23**, 410–423 (1975).
- [Kle75c] L. Kleinrock and F. A. Tobagi, Packet switching in radio channels, part I: carrier sense multiple-access modes and their throughput-delay characteristics, *IEEE Trans. Commun.*, **COM-23**, 1400–1416 (1975).
- [Kle84] L. Kleinrock, On a self-adjusting capability of random access networks, *IEEE Trans. Commun.*, **COM-32**, 40–47 (1984).
- [Ko98] Y. Ko and N. H. Vaidya, Location-aided routing (LAR) in mobile ad hoc networks, *Proc. ACM/IEEE MOBICom '98*, Dallas, TX, 1998.
- [Kob71a] H. Kobayashi, Application of Hestens–Stiefel algorithm to channel equalization, *Proc. IEEE ICC '71*, 1971, pp. 21.25–21.30.
- [Kob71b] H. Kobayashi, Simultaneous adaptive estimation and detection algorithm for carrier modulated data communication systems, *IEEE Trans. Commun. Technol.*, **CT-19**, 268–280 (1971).
- [Kob77] H. Kobayashi, Y. Onozato, and D. Huynh, An approximate method for design and analysis of an ALOHA system, *IEEE Trans. Commun.*, **COM-25**, 148–157 (1977).
- [Koh04] R. Kohno, M. Welborn, and M. McLaughlin, *DS-UWB Proposal, IEEE P802.15 Working Group for Wireless Personal Area Networks (WPANs)*, IEEE 802.15-04/140r2, Mar. 2004.

- [Kon74] A. G. Konheim and B. Meister, Waiting lines and times in a system with polling, *J. ACM*, **21**, 470–490 (1974).
- [Kos00] H. Koshima and J. Hoshen, Personal locator services emerge, *IEEE Spectrum*, Feb. 2000, pp. 41–48.
- [Kri99a] P. Krishnamurthy and K. Pahlavan, Distribution of range error and radio channel modeling for indoor geolocation applications, *Proceedings of PIMRC*, Kyoto, Japan, September 1999.
- [Kri99b] P. Krishnamurthy, Analysis and modeling of the wideband radio channel for indoor geolocation applications, Ph.D. dissertation, Worcester Polytechnic Institute, 1999.
- [Kuc85] A. Kucar and K. Feher, Performance of multi-level modulation systems in the presence of phase noise, *Proc. IEEE ICC '85*, Chicago, IL, June 1985.
- [Kup82] F. Kuperus and J. Arnbak, Packet radio in a Rayleigh channel, *Electr. Lett.*, **18**(12), 506–507 (1982).
- [Kwa75] R. K. Kwan, The Telesat TDMA field trial, *Proc. ICDSC-3*, Kyoto, Tokyo, 1975, pp. 135–143.
- [Lam74] S. S. Lam, Packet switching in a multi-access broadcast channel with application to satellite communications in a computer network, Ph.D. dissertation, Computer Science Department, University of California, Los Angeles, Mar. 1974.
- [Lam75] S. S. Lam and L. Kleinrock, Packet switching in a multiaccess broadcast channel: dynamic control procedures, *IEEE Trans. Commun.*, **COM-23**, 891–904 (1975).
- [Lam80] S. S. Lam, A carrier sense multiple access protocol for local networks, *Comput. Networks*, **4**, 21–32 (1980).
- [LaR84] B. LaReau, IR unit runs computer remotely, *Electron. Week*, Nov. 12, 1984, pp. 71–75.
- [Lat98a] M. Latva-aho, Advanced receivers for wideband CDMA systems, Ph.D. dissertation, University of Oulu, Oulu, Finland, 1998.
- [Lat98b] M. Latva-aho, Bit error probability analysis for FRAMES WCDMA downlink receivers, *IEEE Trans. Vehi. Technol.*, **VT-47**(4), 1119–1133 (1998).
- [Law91] M. C. Lawton, R. L. Davies and J. P. McGeehan, A ray launching method for the prediction of indoor radio channel characteristics, *PIMRC '91*, King's College, London, Sept. 1991, pp. 104–108.
- [Law92] M. C. Lawton and J. P. McGeehan, The application of GTD and ray launching techniques to channel modelling for cordless radio systems, *Proc. 42nd IEEE Vehicular Technology Conference*, Denver, CO, 1992, pp. 125–130.
- [Leb89] M. Lebherz et al., Calculation of broadcast coverage based on a digital terrain model, *Proc. IEEE ICAP '89*, 1989, pp. 355–359.
- [Leb92] M. Lebherz, W. Wiesbeck, and W. Krank, A versatile wave propagation model for the VHF/UHF range considering three-dimensional terrain, *IEEE Trans. Antennas Propag.*, **AP-40**, 1121–1131 (1992).
- [Lee82] W. C. Y. Lee, *Mobile Communications Engineering*, McGraw-Hill, New York, 1982.
- [Lee89] W. C. Y. Lee, *Mobile Cellular Telecommunications Systems*, McGraw-Hill, New York, 1989.
- [Lee93] J. F. Lee, Numerical solutions of TM scattering using an obliquely Cartesian finite difference time domain algorithm, *IEE Proc. H*, **140**(1), 23–28 (1993).
- [Leu86] V. C. M. Leung and R. W. Donaldson, Effects of channel errors on the delay throughput performance and capacity of ALOHA multiple access systems, *IEEE Trans. Commun.*, **COM-34**, 497–502 (1986).
- [Lev76] A. H. Levesque, Block-error distributions and error control code performance in slow Rayleigh fading, *Proc. NTC '76*, Houston, TX, 1976, pp. 24.4.1–24.4.4.

- [Lev93] A. H. Levesque and P. J. Kush, *Evaluation and Correction of the Fading Model Used in the Error Mask Generator Computer Program*, Contrib. TR45.3.2.5/93.02.08.06, TIA/TR45.3 Data Services Task Group, Feb. 1993.
- [L-Ga00] A. Leon-Garcia and Widjaja, *Communication Networks: Fundamental Concepts and Key Architectures*, McGraw-Hill, Boston, 2000.
- [Li87] V. O. K. Li, Multiple access communications networks, *IEEE Commun. Mag.*, **25**(6), 41–48 (1987).
- [Li03a] X. Li and K. Pahlavan, Super-resolution TOA estimation with diversity for indoor geolocation, *IEEE Trans. Wireless Commun.*, Dec. 2003.
- [Li03b] X. Li, Super-resolution TOA estimation with diversity for indoor geolocation, Thesis, Worcester Polytechnic Institute, Worcester, MA, May 2003.
- [Lia85] J. K. Liang, R. J. P. DeFigueiredo, and F. C. Lu, Designing of optimum Nyquist, partial response, N th band, and nonuniform tap spacing FIR digital filters using linear programming technique, *IEEE Trans. Circuits Syst.*, **CAS-32**, 386–392 (1985).
- [Lib99] J. C. Liberti, and T. S. Rappaport, *Smart Antennas for Wireless Communications: US-95 and Third Generation Applications*, Prentice Hall, Upper Saddle River, NJ (1999).
- [Lin73] W. C. Lindsey and M. K. Simon, *Telecommunication Systems Engineering*, Prentice Hall, Englewood Cliffs, NJ, 1973.
- [Lin83] S. Lin and D. J. Costello, Jr., *Error Control Coding: Fundamentals and Applications*, Prentice Hall, Englewood Cliffs, NJ, 1983.
- [Lin84] S. Lin, D. J. Costello, Jr., and M. J. Miller, Automatic-repeat-request error-control schemes, *IEEE Commun. Mag.*, **22**(12), 5–17 (1984).
- [Lin04] S. Lin and D. J. Costello, Jr., *Error Control Coding: Fundamentals and Applications*, 2nd ed., Prentice Hall, Upper Saddle River, NJ, 2004.
- [Liu89] C. L. Liu and K. Feher, Noncoherent detection of $\pi/4$ -shift QPSK systems in a CCI-AWGN combined interference environment, *Proc. 39th IEEE Vehicular Technology Conference*, San Francisco, May 1989, pp. 83–94.
- [Liu90] C. L. Liu and K. Feher, A new generation of Rayleigh fade compensated $\pi/4$ -QPSK coherent modems, *Proc. 40th IEEE Vehicular Technology Conference*, Orlando, FL, May 1990, pp. 482–486.
- [Liu92] R. Prasad and C.-Y. Liu, Throughput analysis of some mobile packet radio protocols in Rician fading channels, *IEE Proc. I: Commun. Speech Vision*, **139**(3), 297–302 (1992).
- [Lo92] T. Lo and J. Litva, Angles of arrival of indoor multipath, *Electron. Lett.*, **28**(18), 1687–1688 (Aug. 27, 1992).
- [Lo94] T. Lo, J. Litva, and H. Leung, A new approach for estimating indoor radio propagation characteristics, *IEEE Trans. Antennas Propagat.*, **AP-42**, 1369–1376 (Oct. 1994).
- [Lon68] A. G. Longley and P. L. Rice, *Prediction of Tropospheric Radio Transmission over Irregular Terrain: A Computer Method*, 1968, ESSA Tech. Rep. ERL 79-ITS 67, U.S. Government Printing Office, Washington, DC, July 1968.
- [Lor79] R. W. Lorenz, *Theoretical Distribution Functions of Multipath Fading in a Mobile Radio and Determination of Their Parameters by Measurement*, Deuseche Bundespost Forschungsinstitut, Technischer Bericht, FI 455 TBr 66, Mar. 1979 (in German).
- [Luc65] R. W. Lucky, Automatic equalization for digital communication, *Bell Syst. Tech. J.*, **44**, 547–588 (1965).
- [Luc66] R. W. Lucky, Techniques for adaptive equalization of digital communication, *Bell Syst. Tech. J.*, **45**, 255–286 (1966).
- [Luc68] R. W. Lucky, J. Salz, and E. J. Weldon, Jr., *Principles of Data Communication*, McGraw-Hill, New York, 1968.

- [Lue84a] R. J. Luebbers, Finite conductivity uniform GTD versus knife edge diffraction in prediction of propagation path loss, *IEEE Trans. Antennas Propag.*, **AP-32**, 70–76 (1984).
- [Lue84b] R. J. Luebbers, Propagation prediction for hilly terrain using GTD wedge diffraction, *IEEE Trans. Antennas Propag.*, **AP-32**, 951–955 (1984).
- [Lyo75] D. L. Lyon, Envelope-driven timing recovery in QAM and SQAM systems, *IEEE Trans. Commun.*, **COM-23**, 1327–1331 (1975).
- [Mac77] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*, North-Holland, Amsterdam, 1977.
- [Mag73] R. F. Magee and J. G. Proakis, Adaptive maximum likelihood sequence estimation for digital signaling in the presence of ISI, *IEEE Trans. Inf. Theory*, **IT-19**, 120–124 (1973).
- [Man00] D. Manolakis, V. Ingle, and S. Kogon, *Statistical and Adaptive Signal Processing*, McGraw-Hill, New York, 2000.
- [Mar85] M. J. Marcus, Recent U.S. regulatory decisions on civil use of spread spectrum, *Proc. IEEE GLOBECOM '85*, New Orleans, LA, Dec. 1985, pp. 16.6.1–16.6.3.
- [Mar87a] M. J. Marcus, Regulatory policy considerations for radio local area networks, *IEEE Commun. Mag.*, **25**(7), 95–99 (1987).
- [Mar87b] S. L. Marple, *Digital Spectral Analysis with Applications*, Prentice Hall, Englewood Cliffs, NJ, 1987.
- [Maz75] J. E. Mazo, Optimum timing phase for an infinite equalizer, *Bell Syst. Tech. J.*, **54**, 189–201 (1975).
- [MBO04] *MBOA Physical Layer Technical Specification*, May 2004.
- [McG02] J. McGeough, Wireless location positioning, based on signal propagation Data, White Paper, <http://www.wirelessdevnet.com/software/>, 2002.
- [McK91] J. W. McKown and R. L. Hamilton, Jr., Ray tracing as a design tool for radio networks, *IEEE Network Mag.*, **5**(6), 27–30 (1991).
- [McL88] P. J. McLane et al., PSK and DPSK trellis codes for fast fading, shadowed mobile satellite communication channels, *IEEE Trans. Commun.*, **COM-36**, 1242–1246 (1988).
- [Med83] J. S. Meditch and C. T. Lea, Stability and optimization of the CSMA and CSMA/CD channels, *IEEE Trans. Commun.*, **COM-31**, 763–774 (1983).
- [Med93] J. Meditz, Development of custom coded blocks in signal processing workstation (SPW) software for indoor radio propagation modeling, M.S. thesis, ECE Department, Worcester Polytechnic Institute, Worcester, MA, May 1993.
- [Meh97] A. Mehrota, *GSM System Engineering*, Artech House, Norwood, MA, 1997.
- [Met76a] J. J. Metzner, On improving utilization in ALOHA networks, *IEEE Trans. Commun.*, **COM-24**, 447–448 (1976).
- [Met76b] R. Metcalfe and D. Boggs, ETHERNET: distributed packet switching for local computer networks, *Commun. ACM*, **19**, 395–404 (1976).
- [Mic85] A. M. Michelson and A. H. Levesque, *Error-Control Techniques for Digital Communication*, Wiley-Interscience, New York, 1985.
- [Mil92] L. B. Milstein et al., On the feasibility of a CDMA overlay for personal communications networks, *IEEE J. Sel. Areas Commun.*, **SAC-10**, 655–668 (1992).
- [Mil98] C. Miller and J. Lee, BER expressions for differentially detected $\pi/4$ -DQPSK modulation, *IEEE Trans. Commun.*, **COM-46**, 71–81 (1998).
- [Mis02] P. Misra and P. Enge, *Global Positioning System: Signals, Measurements, and Performance*, Ganga-Jamuna Press, Lincoln, MA, 2002.
- [Moh89] M. L. Moher, TCMP-A modulation and coding strategy for Rician fading channels, *IEEE J. Sel. Areas Commun.*, **SAC-7**, 1347–1355 (1989).

- [Mon77] P. Monsen, Theoretical and measured performance of a DFE modem on a fading multipath channel, *IEEE Trans. Commun.*, **COM-25**, 1144–1153 (1977).
- [Mon84] P. Monsen, MMSE equalization of interference on fading diversity channels, *IEEE Trans. Commun.*, **COM-32**, 5–12 (1984).
- [Mor92] G. Morrison, M. Fattouch, and H. Zhohloul, Statistical analysis and autoregressive modeling of the indoor channel, *Proc. International Conference on Personal Communications*, Dallas, TX, Oct. 1992.
- [Mot88a] A. J. Motley and J. M. P. Keenan, Personal communication radio coverage in buildings at 900 MHz and 1700 MHz, *IEE Electr. Lett.*, **24**, 763–764 (1988).
- [Mot88b] A. J. Motley, Radio coverage in buildings, *Proc. National Communications Forum*, Chicago, Oct. 1988, pp. 1722–1730.
- [Mou92] M. Mouly and M.-B. Pautet, *The GSM System for Mobile Communications*, ISBN 2-9507190-0-7, 1992.
- [Mul73] K. H. Muller, A new approach to optimum pulse shaping in sampled systems using time-domain filtering, *Bell Syst. Tech. J.*, **52**, 723–729 (1973).
- [Mur81] K. Murota and K. Hirade, GMSK modulation for digital mobile radio telephony, *IEEE Trans. Commun.*, **COM-29**, 1044–1050 (1981).
- [Mus82] J. M. Musser and J. W. Daigle, Throuput analysis of an asynchronous CDMA system, *Proc. IEEE ICC '82*, Philadelphia, PA, 2F2.1–7.
- [Nag98] A. Naguib et al., A space-time coding modem for high-data-rate wireless communications, *IEEE J. Sel. Areas Commun.*, Oct. 1998, pp. 1459–1477.
- [Nak60] M. Nakagami, The m -distribution: a general formula of intensity distribution of rapid fading, in *Statistical Methods of Radio Wave Propagation*, Pergamon Press, Elmsford, NY, 1960.
- [Nak90] N. Nakajima et al., A system design for TDMA mobile radios, *Proc. 40th IEEE Vehicular Technology Conference*, Orlando, FL, May 1990, pp. 295–298.
- [Nam84a] C. Namislo, Analysis of mobile radio slotted ALOHA networks, *IEEE J. Sel. Areas Commun.*, **SAC-2**, 583–588 (1984).
- [Nam84b] C. Namislo, Analysis of mobile radio slotted ALOHA networks, *IEEE Trans. Veh. Technol.*, **VT-33**, 199–204 (1984).
- [Nel84] R. Nelson and L. Kleinrock, The spatial capacity of a slotted ALOHA multihop packet radio network with capture, *IEEE Trans. Commun.*, **COM-32**, 684–694 (1984).
- [Ogo81] S. Ogose and K. Murota, Differentially encoded GMSK with 2-bit differential detection, *Trans. IECE Japan*, **J64-B**, 248–254 (1981).
- [O'Ha99] B. O'Hara and A. Petrick, *IEEE 802.11 Handbook: A Designer's Companion*, IEEE Press, Piscataway, NJ, 1999.
- [Oku68] Y. Okumura et al., Field strength and its variability in VHF and UHF land-mobile service, *Rev. Electr. Commun. Lab.*, **16**, 825–873 (1968).
- [Opp04] I. Oppermann, M. Hamalainen, and J. Iinatti, *UWB Theory and Applications*, Wiley, Hoboken, NJ, 2004.
- [Pad94] R. Padovani, The capacity of the CDMA cellular: reverse link field test results, *Proc. International Zurich Seminar on Digital Communications*, Zurich, Switzerland, Mar. 1994, pp. 56–66.
- [Pah79] K. Pahlavan, Channel measurement for wideband digital communication over fading channels, Ph.D. dissertation, Worcester Polytechnic Institute, Worcester, MA, June 1979.
- [Pah80] K. Pahlavan and J. W. Matthews, Performance of channel measurement techniques over fading channels, *Proc. IEEE NTC '80*, Houston, TX, 1980, pp. 58.5.1–58.5.5.
- [Pah84] K. Pahlavan, A review of wireless in-house data communication systems, *Proc. IEEE Computer Networking Symposium*, Gaithersburg, MD, 1984 pp. 129–136.

- [Pah85a] K. Pahlavan, Comparison between the performance of QPSK, SQPSK, QPR, and SQPR systems over microwave LOS channels, *IEEE Trans. Commun.*, **COM-33**, 291–296 (1985).
- [Pah85b] K. Pahlavan, Wireless communications for office information networks, *IEEE Commun. Mag.*, **23**(6), 19–27 (1985).
- [Pah87] K. Pahlavan, Spread spectrum for wireless local networks, *Proc. IEEE PCCC*, Phoenix, AZ, 1987, pp. 215–219.
- [Pah88b] K. Pahlavan, Signal processing in telecommunications, Chapter 22 in C. H. Chen (ed.), *Handbook of Signal Processing*, Marcel Dekker, New York, 1988.
- [Pah88c] K. Pahlavan and J. L. Holsinger, Voice-band data communication modems: a historical review, 1919–1988, *IEEE Commun. Mag.*, **26**(1), 16–27 (1988).
- [Pah89] K. Pahlavan, R. Ganesh, and T. Hotaling, Multipath propagation measurements on manufacturing floors at 910 MHz, *IEE Electr. Lett.*, **25**, 225–227 (1989).
- [Pah90a] K. Pahlavan and M. Chase, Spread-spectrum multiple-access performance of orthogonal codes for indoor radio communications, *IEEE Trans. Commun.*, **COM-38**, 574–577 (1990).
- [Pah90b] K. Pahlavan and S. J. Howard, Statistical AR models for the frequency selective indoor radio channels, *IEE Electr. Lett.*, **26**, 1133–1135 (1990).
- [Pah90c] K. Pahlavan and J. W. Matthews, Performance of adaptive matched filter receivers over fading multipath channels, *IEEE Trans. Commun.*, **COM-38**, 2106–2113 (1990).
- [Pah93] K. Pahlavan, S. J. Howard, and T. A. Sexton, Decision feedback equalization of the indoor radio channel, *IEEE Trans. Commun.*, **COM-41**, 164–170 (1993).
- [Pah91] K. Pahlavan, Nonlinear quantization and multi-level/phase modulation and coding, *IEEE Trans. Commun.*, (1991).
- [Pah95] K. Pahlavan and A. Levesque, *Wireless Information Networks*, Wiley, New York, 1995.
- [Pah98] K. Pahlavan, P. Krishnamurthy, and J. Beneat, Wideband radio channel modeling for indoor geolocation applications, *IEEE Commun. Mag.*, **36**(4), 60–65 (1998).
- [Pah00] K. Pahlavan, P. Krishnamurthy, A. Hatami, M. Ylianttila, J. Mäkelä, R. Pichna, and J. Vallström, Handoff in hybrid mobile data networks (invited paper), *IEEE Pers. Commun. Mag.*, **7**(2), 34–47 (2000).
- [Pah02a] K. Pahlavan and P. Krishnamurthy, *Principles of Wireless Networks: A Unified Approach*, Prentice Hall, Upper Saddle River, NJ, 2002.
- [Pah02b] K. Pahlavan, X. Li, and J.-P. Mäkelä, Indoor geolocation science and technology, *IEEE Commun. Mag.*, **40**(2), 112–118 (2002).
- [Pal91] M. Pallas and G. Jourdain, Active high resolution time delay estimation for large BT signals, *IEEE Trans. Signal Process.*, **SP-39**(4), 781–788 (1991).
- [Pan02] A. Pandharipande, Principles of OFDM, *IEEE Potentials*, Apr./May 2002, pp. 16–19.
- [Par89] J. D. Parsons and J. G. Gardiner, *Mobile Communication Systems*, Blackie, Glasgow, 1989.
- [Pas79] S. Pasupathy, Minimum shift keying: a spectrally efficient modulation, *IEEE Commun. Mag.*, **17**(4), 14–22 (1979).
- [Pet72] W. W. Peterson and E. J. Weldon, Jr., *Error-Correcting Codes*, 2nd ed., MIT Press, Cambridge, MA, 1972.
- [Pot00] G. J. Pottie and W. J. Kaiser, Wireless integrated network sensors, *Commun. ACM*, **43**(5), 51–58 (2000).
- [Pra88] R. Prasad and J. Arnbak, Enhanced throughput in packet radio channels with shadowing, *IEE Electr. Lett.*, **24**, 986–988 (1988).
- [Pra02] G. S. Prabhu and P. M. Shankar, Simulation of flat fading using MATLAB for classroom instruction, *IEEE Trans. Educ.*, **45**(1), 19–25 (2002).

- [Pre91] W. H. Press et al., *Numerical Recipes in C*, Cambridge University Press, New York, 1991.
- [Pri58] R. Price and P. E. Green, A communication technique for multipath channels, *Proc. IRE*, **46**, 555–570 (1958).
- [Pri83] R. Price, Further notes and anecdotes on spread-spectrum origins, *IEEE Trans. Commun.*, COM-31, 85–97 (1983); reprinted in C. E. Cook et al. (eds.), *Spread-Spectrum Communications*, IEEE Press, New York, 1983.
- [Pri00] N. Priyantha, A. Chakraborty, and H. Balakrishnan, The cricket location-support system,” presented at the 6th ACM International Conference on Mobile Computing and Networking (ACM MOBICOM), Boston, MA, Aug. 2000.
- [Pro89] J. G. Proakis, *Digital Communications*, 2nd ed., McGraw-Hill, New York, 1989.
- [Pro01] J. G. Proakis, *Digital Communications*, 4th ed., McGraw-Hill, New York, 2001.
- [Qoa01] Qualcomm, Inc., 1x Evolution IS-856 TIA/EIA Standard, *Airlink Overview*, Rev. 7.2, Nov. 7, 2001.
- [Qur77] S. H. Qureshi and G. D. Forney, Performance and properties of a T/2 equalizer, *Proc. NTC '77*, Los Angeles, CA, Dec. 1977, pp. 11.1.1–11.1.14.
- [Qur85] S. H. Qureshi, Adaptive equalization, *Proc. IEEE*, **73**, 1349–1387 (1985).
- [Rab78] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice Hall, Englewood Cliffs, NJ, 1978.
- [Rah93] M. Rahnema, Overview of the GSM system and protocol architecture, *IEEE Commun. Mag.*, **31**(4), 92–100 (1993).
- [Rap89] T. S. Rappaport, Characterization of UHF multipath radio channels in factory buildings, *IEEE Trans. Antennas Propag.*, **AP-37**, 1058–1069 (1989).
- [Rap90] T. S. Rappaport, S. Y. Seidel, and R. Singh, 900 MHz multipath propagation measurements for U.S. digital cellular radio telephone, *IEEE Trans. Veh. Technol.*, **VT-39**, 132–139 (1990).
- [Rap91a] T. S. Rappaport, The wireless revolution, *IEEE Commun. Mag.*, **29**(11), 52–71 (1991).
- [Rap91b] T. S. Rappaport, S. Y. Seidel, and K. Takamizawa, Statistical channel impulse response models for factory and open plan building radio communication system design, *IEEE Trans. Commun.*, **COM-39**, 794–807 (1991).
- [Rap92] T. S. Rappaport and D. A. Hawbaker, A ray tracing technique to predict path loss and delay spread inside buildings, *Proc. IEEE GLOBECOM '92*, Orlando, FL, Dec. 1992, pp. 649–653.
- [Rap02] T. S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed., Prentice Hall, Upper Saddle River, NJ, 2002.
- [Ray81] D. Raychaudhuri, Performance analysis of random access packet-switched code division multiple access systems, *IEEE Trans. Commun.*, June 1981, pp. 895–901.
- [Ray84] D. Raychaudhuri, ALOHA with multipacket messages and ARQ-type retransmission protocols: throughput analysis, *IEEE Trans. Commun.*, **COM-32**, 148–154 (1984).
- [Ray92] D. Raychaudhuri and N. Wilson, Multimedia personal communication networks: system design issues, *Proc. 3rd WINLAB Workshop on Third Generation Wireless Information Networks*, Rutgers, NJ, 1992, pp. 259–288.
- [Red95] S. M. Redl, M. K. Weber, and M. W. Oliphant, *An Introduction to GSM*, Artech House, Norwood, MA, 1995.
- [Ric48] S. O. Rice, Statistical properties of a sine wave plus random noise, *Bell Syst. Tech. J.*, **27**, 109–157 (1948).
- [Rob73] L. Roberts, Dynamic allocation of satellite capacity through packet reservation, *Proc. AFIPS Conf.*, **42**, June 1973.

- [Rob75] L. G. Roberts, ALOHA packet system with and without slots and capture, *ACMSIG-COM Comput. Commun. Rev.*, **5**, 28–42 (1975).
- [Roh99] H. Rohling et al., Broadband OFDM radio transmission for multimedia applications, *Proc. IEEE*, **87**(10), 1778–1789 (1999).
- [Roo02a] T. Roos, P. Myllymaki, H. Tirri, P. Miskangas, and J. Sievanen, A probabilistic approach to WLAN user location estimation, *Int. J. Wireless Inf. Networks*, **9**(3) (2002).
- [Roo02b] T. Roos, P. Myllymaki, and H. Tirri, A statistical modeling approach to location estimation, *IEEE Trans. Mobile Comput.*, **MC-1**(1) (2002).
- [Ros73] G. F. Ross, Transmission and reception system for generating and receiving baseband duration pulse signals without distortion for short baseband communication systems, U.S. patent 3,728,632, Apr. 1973.
- [Ros92] J.-P. Rossi and A. J. Levy, A ray model for deterministic radio wave propagation in urban area, *Radio Sci.*, **27**, 971–979 (1992).
- [Ros93] J.-P. Rossi and A. J. Levy, Propagation analysis in cellular environment with the help of models using ray theory and GTD, *Proc. 43rd IEEE Vehicular Technology Conference*, Secaucus, NJ, 1993, pp. 253–256.
- [Ros97] J. P. Rossi, J. P. Barbot, and A. J. Levy, Theory and measurement of the angle of arrival and time delay of UHF radiowaves using a ring array, *IEEE Trans. Antennas Propag.*, **AP-45**, 876–884 (May 1997).
- [Rub79] I. Rubin, Message delays in FDMA and TDMA communication channels, *IEEE Trans. Commun.*, **COM-27**, 769–777 (1979).
- [Rus91] A. J. Rustako, Jr., N. Amitay, G. J. Owens, and R. S. Roman, Radio propagation at microwave frequencies for line-of-sight microcellular mobile and personal communications, *IEEE Trans. Veh. Technol.*, **VT-40**, 203–210 (1991).
- [Saa97] H. Saarnisaari, TLS-ESPRIT in a time delay estimation, *Proc. IEEE 47th Vehicular Technology Conference*, Phoenix, AZ, 1997, pp. 1619–1623.
- [Sal73] J. Salz, Optimum mean-square decision feedback equalization, *Bell Syst. Tech. J.*, **52**, 341–1373 (1973).
- [Sal82] A. C. Salazar and V. B. Lawrence, Design and implementation of transmitter and receiver filters with periodic coefficient nulls for digital systems, *Proc. IEEE ICASSP '82*, Paris, May 1982, pp. 306–310.
- [Sal87a] A. M. Saleh, A. J. Rustako, and R. S. Roman, Distributed antennas for indoor radio communications, *Proc. IEEE ICC '87*, Seattle, WA, 1987, pp. 3.3.1–3.3.5.
- [Sal87b] A. M. Saleh and R. A. Valenzuela, A statistical model for indoor multipath propagation, *IEEE J. Sel. Areas Commun.*, **SAC-5**, 128–137 (1987).
- [Sal91] A. Salamasi and K. S. Gilhousen, On the system design aspects of code division multiple access (CDMA) applied to digital cellular and personal communications networks, *Proc. 41st IEEE Vehicular Technology Conference*, St. Louis, MO, May 1991, pp. 57–62.
- [Sar86] H. Sari, L. Desperben, and S. Moridi, Minimum mean-square error timing recovery schemes for digital equalizers, *IEEE Trans. Commun.*, **COM-34**(7), 694–702 (1986).
- [Sas93] H. Sasaoka and Y. Omori, Multi-carrier 16QAM system in land mobile communications, *Proc. PIMRC '93*, Yokohama, Japan, Sept. 1993, pp. 84–88.
- [Sch66] M. Schwartz, W. R. Bennett, and S. Stein, *Communication Systems and Techniques*, McGraw-Hill, New York, 1966.
- [Sch82] R. A. Scholtz, The origins of spread-spectrum communications, *IEEE Trans. Commun.*, **COM-30**, 822–853 (1982).
- [Sch84] D. Schaum et al., MSE mobile subscriber equipment, *Army Commun.*, **9**(3), 6–22 (1984).

- [Sch89] C. Schlegel and D. J. Costello, Bandwidth-efficient coding for fading channels, *IEEE J. Sel. Areas Commun.*, **SAC-7**, 1356–1368 (1989).
- [Sch00] R. A. Scholtz and M. Z. Win, Ultra-wide bandwidth time-hopping spread-spectrum impulse radio for wireless multiple-access communications, *IEEE Trans. Comm.*, **48**, 69–689 (2000).
- [Sei92] S. Y. Seidel and T. S. Rappaport, 914 MHz path loss prediction models for wireless communications in multifloored buildings, *IEEE Trans. Antennas Propag.*, **AP-40**, 207–217 (1992).
- [Sie00] T. Siep, I. Gifford, R. Braley, and R. Heile, Paving the way for personal area network standards: An overview of the IEEE P802.15 working group for wireless personal area networks, *IEEE Pers. Commun.* Feb. 2000.
- [Ses91] N. Seshadri and Y.-K. Kim, *A Channel Simulator to Generate Error Masks with Channel State Information*, TIA Contribution TR45.3.2.3.2/91.10.22.06, TR45.3 Subcommittee on Digital Cellular Standards, Speech Working Group, Oct. 1991.
- [Seu03] E. Seurre, P. Savelli, and P.-J. Pietri, *EDGE for Mobile Internet*, Artech House, Norwood, MA, 2003.
- [Sex89a] T. A. Sexton and K. Pahlavan, Channel modeling and adaptive equalization of indoor radio channels, *IEEE J. Sel. Areas Commun.*, **SAC-7**, 114–121 (1989).
- [Sex89b] T. A. Sexton, Channel modeling and high speed transmission performance for the indoor radio channel, Ph.D. dissertation, Worcester Polytechnic Institute, Worcester, MA, Aug. 1989.
- [Sha48] C. E. Shannon, A Mathematical Theory of Communication, *Bell Syst. Tech. J.*, **27**, 379–423–,623–656 (1948); reprinted in book form with postscript by W. Weaver, University of Illinois Press, Urbana, IL, 1949.
- [Sha49] C. E. Shannon, Communication in the presence of noise, *Proc. IRE*, **37**, 10–21 (1949).
- [Sha59] C. E. Shannon, Probability of error for optimal codes in a Gaussian channel, *Bell Syst. Tech. J.*, **38**, 611–656 (1959).
- [Sha05] H. G. Shantz, A brief history of UWB antennas, *IEEE A&E Mag.*, Apr. 2005, pp. 22–26.
- [She75] A. R. Sherwood and I. A. Fantera, Multipath measurement over troposcatter paths with application to digital transmission, *Proc. NTC '75*, New Orleans, LA, Dec. 1975, pp. 28.1–28.5.
- [Sho03] G. Shor, TG3a-Wisair-CFP-Presentation, DS-UWB Proposal, IEEE P802.15 Working Group for Wireless Personal Area Networks (WPANs), IEEE 802.15-03/151r3, May 2003.
- [Sim84] M. K. Simon and C. C. Wang, Differential detection of Gaussian MSK in a mobile radio environment, *IEEE Trans. Veh. Technol.*, **VT-33**, 307–320 (1984).
- [Sim85] M. K. Simon et al., *Spread Spectrum Communication*, Computer Science Press, New York, 1985.
- [Skl88] B. Sklar, *Digital Communications: Principles and Applications*, Prentice Hall, Englewood Cliffs, NJ, 1988.
- [Skl97] B. Sklar, A primer on turbo concepts, *IEEE Commun. Mag.*, **35**(12), 94–102 (1997).
- [Skl01] B. Sklar, *Digital Communications: Fundamentals and Applications*, Prentice Hall, Upper Saddle River, NJ, 2001.
- [Sma02] A. Smailagic and D. Kogan, Location sensing and privacy in a context-aware computing environment, *IEEE Wireless Commun.*, Oct. 2002, pp. 10–17.
- [Soh87] K. Sohrawy, M. L. Molle, and A. N. Venetsanopoulos, Comments on “Throughput analysis for persistent CSMA systems,” *IEEE Trans. Commun.*, **COM-35**, 240–243 (1987).
- [Sol01] S. Soliman, Report of Qualcomm Incorporated, in the matter of revision of Part 15 of the Commission’s Rules Regarding Ultra-Wideband Transmissions Systems, ET Docket 98–153, Mar. 5, 2001.

- [Sou96] E. A. Sourour and M. Nakagawa, Performance of orthogonal multicarrier CDMA in a multipath fading channel, *IEEE Trans. Commun.*, **44**, 356–367 (Mar. 1996).
- [Spe00] Q. H. Spencer, B. D. Jeffs, M. A. Jensen, and A. L. Swindlehurst, Modeling the statistical time and angle of arrival characteristics of an indoor multipath channel, *IEEE J. Sel. Areas Commun.*, **SAC-18**(3), 347–360 (2000).
- [Ste92] R. Steele (ed.), *Mobile Radio Communications*, Pentech Press, London, 1992.
- [Ste93] R. Steele, Speech codecs for personal communications, *IEEE Commun. Mag.*, **31**(11), 76–83 (1993).
- [Ste94] D. G. Steer, Coexistence and access etiquette in the United States unlicensed PCS bands, *IEEE PCS Mag.*, **1**, 36–43 (1994).
- [Ste01] R. Steele, C.-C. Lee, and P. Gould, *GSM, cdmaOne and 3G Systems*, Wiley, Chichester, West Sussex, England, 2001.
- [Suz77] H. Suzuki, A statistical model for urban radio propagation: multipath characteristics in New York City, *IEEE Trans. Commun.*, **COM-25**, 673–680 (1977).
- [Taf75] A. Taflove and M. E. Morris, Numerical solution of steady-state electromagnetic scattering problems using the time-dependent Maxwell's equations, *IEEE Trans. Microwave Theory Tech.*, **MTT-23**, 623–630 (1975).
- [Tak85] H. Takagi and L. Kleinrock, Throughput analysis for persistent CSMA systems, *IEEE Trans. Commun.*, **COM-33**, 627–638 (1985).
- [Tak87] H. Takagi and L. Kleinrock, Correction to “Throughput analysis for persistent CSMA systems,” *IEEE Trans. Commun.*, **COM-35**, 243–245 (1987).
- [Tan88] A. S. Tannenbaum, *Computer Networks*, 2nd ed., Prentice Hall, Englewood Cliffs, NJ, 1988.
- [Tar98] V. Tarokh, N. Seshadri, and A. R. Calderbank, Space-time codes for high data rate wireless communications: performance criterion and code construction, *IEEE Trans. Inf. Theory*, Mar. 1998, pp. 744–765.
- [Tas86] S. Tasaka, Dynamic behavior of a CSMA-CD system with a finite population of buffered users, *IEEE Trans. Commun.*, **COM-34**, 576–586 (1986).
- [Tay01] J. D. Taylor (ed.), *Ultra-Wideband Radar Technology*, CRC Press, Boca Raton, FL, 2001.
- [TDCweb] Time Domain Corporation, <http://www.timedomain.com>.
- [Tel95] E. Telatar, *Capacity of Multiantenna Gaussian Channels*, AT&T Bell Laboratories, Tech. Memo., June 1995.
- [TIA92] Telecommunication Industry Association, Project PN-2759, *Cellular System Dual-Mode Mobile Station-Base Station Compatibility Standard*, IS-54, Rev. B, Jan. 9, 1992.
- [TIA93a] Telecommunication Industry Association, Project 25 PN-3124, *Common Air Interface*, May 19, 1993.
- [TIA93b] Telecommunication Industry Association, TR45.3.2.5, *Radio Link Protocol for Asynchronous Data Service*, 1993.
- [TIA94] Telecommunication Industry Association, Project PN-3123, *Async Data and Fax*, and Project PN-3306, *Radio Link Protocol 1*, both issued Nov. 14, 1994.
- [Tie01] E. G. Tiedemann, “CDMA2000-1X: New Capabilities for CDMA Networks,” *IEEE Vehicular Technology Newsletter*, **48**, No. 4, (2001).
- [Tin99a] R. Tingley and K. Pahlavan, A comparison of two techniques for parameter estimation on an indoor radio channel, *Proc. Wireless '99*, Calgary, Alberta, Canada, July 1999, pp. 12–14.
- [Tin99b] R. Tingley and K. Pahlavan, Measurement of the spatial and temporal propagation characteristics of an indoor radio channel using an antenna array, *Proc. IEEE PIMRC '99*, Kyoto, Japan, Sept. 10, 1999.

- [Tin00] R. Tingley, Time-space characteristics of indoor radio channel, Ph.D. dissertation, Worcester Polytechnic Institute, Worcester, MA, May 2000.
- [Tin01] R. Tingley and K. Pahlavan, Time-space measurement of indoor radio propagation, *IEEE Trans. Instrum. Meas.*, **IM-50**(1), 22–31 (2001).
- [Tob75] F. A. Tobagi and L. Kleinrock, Packet switching in radio channels, part II: the hidden-terminal problem in carrier sense multiple access and the busy-tone solution, *IEEE Trans. Commun.*, **COM-23**, 1417–1433 (1975).
- [Tob76] F. A. Tobagi and L. Kleinrock, Packet switching in radio channels, part III: polling and (dynamic) split channel reservation multiple access, *IEEE Trans. Commun.*, **COM-24**, 832–845 (1976).
- [Tob80a] F. A. Tobagi, Multiaccess protocols in packet communication systems, *IEEE Trans. Commun.*, **COM-28**, 468–488 (1980).
- [Tob80b] F. A. Tobagi and V. Hunt, Performance analysis of carrier sense multiple access with collision detection, *Comput. Networks*, **4**, 245–259 (1980).
- [Tur72] G. L. Turin et al., A statistical model of urban multipath propagation, *IEEE Trans. Veh. Technol.*, **VT-21**, 1–9 (1972).
- [Unb02] M. Unbehau, On the design and deployment of low-cost wireless infrastructure, Ph.D. dissertation, Royal Institute of Technology, Stockholm, Sweden, 2002.
- [Unb03] M. Unbehau and M. Kamenetsky, On the deployment of picocellular wireless infrastructure, *IEEE Wireless Commun. Mag.*, Dec. 2003, pp. 70–80.
- [Ung74] G. Ungerboeck, Adaptive maximum likelihood receiver for carrier modulated data transmission systems, *IEEE Trans. Commun.*, **COM-22**, 624–636 (1974).
- [Ung76] G. Ungerboeck, Fractional tap-spacing equalizer and consequences for clock recovery in data modems, *IEEE Trans. Commun.*, **COM-24**, 856–864 (1976).
- [Ung82] G. Ungerboeck, Channel coding with multilevel/phase signals, *IEEE Trans. Inf. Theory*, **IT-28**, 55–67 (1982).
- [Ung87] G. Ungerboeck, Trellis-coded modulation with redundant signal sets, *IEEE Commun. Mag.*, **25**(2), 5–21 (1987).
- [Var88] P. Vary et al., Speech codec for the European mobile radio system, *Proc. ICASSP '88*, Apr. 1988, pp. 227–230.
- [Ver98] S. Verdú, *Multisuser Detection*, Cambridge University Press, Cambridge, 1998.
- [Vit67] A. J. Viterbi, Error bounds for convolutional codes and an asymptotically optimum decoding algorithm, *IEEE Trans. Inf. Theory*, **IT-13**, 260–269 (1967).
- [Vit93] A. J. Viterbi and A. M. Viterbi, Erlang capacity of a power controlled CDMA system, *IEEE J. Sel. Areas Commun.*, **SAC-11**, 882–890 (1993).
- [Vit95] A. J. Viterbi, CDMA, *Principles of Spread Spectrum Communications*, Addison-Wesley, Reading, MA, 1995.
- [Wei71] S. B. Weinstein and P. M. Ebert, Data transmission by frequency-division multiplexing using the discrete Fourier transform, *IEEE Trans. Commun. Technol.*, **CT-19**, 628–634 (1971).
- [Wei84a] L.-F. Wei, Rotationally invariant convolutional channel coding with expanded signal space, part I: 180°, *IEEE J. Sel. Areas Commun.*, **SAC-2**, 659–671 (1984).
- [Wei84b] L.-F. Wei, Rotationally invariant convolutional channel coding with expanded signal space, part II: nonlinear codes, *IEEE J. Sel. Areas Commun.*, **SAC-2**, 672–686 (1984).
- [Wel03] M. Welborn, M. McLaughlin, P. Ceva, and R. Kohno, DS-UWB Proposal, IEEE P802.15 Working Group for Wireless Personal Area Networks (WPANs), IEEE 802.15-03/334r3, Sept. 2003.
- [Wer98] J. Werb and C. Lanzl, Designing a positioning system for finding things and people indoors, *IEEE Spectrum*, **35**(9), 71–78 (1998).

- [Wic94] S. B. Wicker and V. K. Bhargava, *Reed–Solomon Codes and Their Applications*, IEEE Press, Piscataway, NJ, 1994.
- [Wid66] B. Widrow, *Adaptive Filters, I: Fundamentals*, Tech. Rep. 6764–6, Stanford Electronics Laboratory, Stanford University, Stanford, CA, Dec. 1966.
- [Wil95] T. A. Wilkinson, T. Phipps, and S. K. Barton, A report on HIPERLAN standardization, *International Journal on Wireless Information Networks*, **2**, 99–120, (1995).
- [Win85] J. H. Winters and Y. S. Yeh, On the performance of wideband digital radio transmission within buildings using diversity, *Proc. IEEE GLOBECOM '85*, New Orleans, LA, 1985, pp. 32.5.1–32.5.6.
- [Win98] M. Win and R. Scholtz, On the performance of ultra-wide bandwidth signals in dense multipath environment, *IEEE Commun. Lett.*, **2**(2), 51–53 (1998).
- [Win00] M. Z. Win and R. A. Scholtz, Ultra-wide bandwidth time-hopping spread spectrum impulse radio for wireless multiple-access communications, *IEEE Trans. Commun.*, **COM-48**(4), 679–691 (2000).
- [Wol78] J. K. Wolf, Efficient maximum likelihood decoding of linear block codes using a trellis, *IEEE Trans. Inf. Theory*, **IT-24**, 76–81 (1978).
- [Wol82] J. K. Wolf, A. M. Michelson, and A. H. Levesque, On the probability of undetected error for linear block codes, *IEEE Trans. Commun.*, **COM-30**, 317–324 (1982).
- [Wol98] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel, *Proc. ISSSE '98*, Pisa, Italy.
- [Woz65] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*, Wiley, New York, 1965.
- [Yal02] R. Yallapragada, V. Kripalani, and A. Kripalani, EDGE: a technology assessment, presented at the IEEE International Conference on Personal Wireless Communications, Dec. 2002.
- [Yam91] H. Yamada, M. Ohmiya, Y. Ogawa, and K. Itoh, Superresolution techniques for time-domain measurements with a network analyzer, *IEEE Trans. Antennas Propag.*, **AP-39**, 177–183 (Feb. 1991).
- [Yan93a] G. Yang, S. Li, J. F. Lee, and K. Pahlavan, Computer simulation of indoor radio propagation, *Proc. PIMRC '93*, Yokohama, Japan, Sept. 1993.
- [Yan93b] G. Yang, K. Pahlavan, and J. F. Lee, A 3D propagation model with polarization characteristics in indoor radio channels, *Proc. IEEE GLOBECOM '93*, Houston, TX, Nov. 1993.
- [Yan94a] G. Yang and K. Pahlavan, Performance analysis of multicarrier modems in an office environment using 3D ray tracing, *Proc. IEEE GLOBECOM '94*, San Francisco, CA, Dec. 1994.
- [Yan94b] G. Yang, Performance evaluation of high speed wireless data systems using a 3D ray tracing algorithm, Ph.D. dissertation, Worcester Polytechnic Institute, Worcester, MA, June 1994.
- [Yan94c] G. Yang, K. Pahlavan, and T. J. Holt, Sector antenna and DFE modems for high speed indoor radio communications, *IEEE Trans. Veh. Technol.*, **VT-43**, 925–933 (1994).
- [Yan03] L.-L. Yang and L. Hanzo, Multicarrier DS-CDMA: a multiple access scheme for ubiquitous broadband wireless communications, *IEEE Commun. Mag.*, Oct. 2003, pp. 116–124.
- [Yee66] K. S. Yee, Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media, *IEEE Trans. Antennas Propag.*, **AP-14**, 302–307 (1966).
- [Yee93] N. Yee, J.-P. Linnartz, and G. Fettweis, Multi-carrier CDMA in indoor wireless radio networks, *Proc. PIMRC '93*, Yokohama, Japan, Sept. 1993, pp. 109–113.
- [Yeg91] P. Yegani and C. D. McGillem, A statistical model for the factory radio channel, *IEEE Trans. Commun.*, **COM-39**, 1445–1454 (1991).

- [Yen85] C. S. Yen and R. D. Crawford, The use of direct optical beams in wireless computer communications, *Proc. IEEE GLOBECOM '85*, New Orleans, LA, 1985, pp. 39.1.1–39.1.5.
- [Yon86a] A. Yongacoglu, D. Makrakis, and K. Feher, 1-Bit differential detection of GMSK with data-aided phase control, *Proc. IEEE ICC '86*, Toronto, 1986, pp. 57.8.1–57.8.5.
- [Yon86b] A. Yongacoglu et al., A new receiver for differential detection of GMSK, *Proc. IEEE GLOBECOM '86*, Houston, TX, Dec. 1986, pp. 1039–1044.
- [Yon88] A. Yongacoglu, D. Makrakis and K. Feher, Differential detection of GMSK using decision feedback, *IEEE Trans. Commun.*, **COM-36**, 641–649 (1988).
- [Zan03a] E. D. Zand, K. Pahlavan, and J. Beneat, Frequency domain measurement for indoor geolocation, *Proc. IEEE PIMRC '03*, Beijing, China, Sept. 2003.
- [Zan03b] E. Zand, Measurement of TOA using frequency domain techniques for indoor geolocation, M.S. thesis, Worcester Polytechnic Institute, Worcester, MA, May 2003.
- [Zdu89] K. J. Zdunek, D. Ucci, and J. LoCicero, Throughput of nonpersistent inhibit sense: multiple access with capture, *Electr. Lett.*, **25**, 30–32 (1989).
- [Zel04] A. V. Zelst and T. C. W. Schenk, Implementation of a MIMO OFDM based wireless LAN system, *IEEE Trans. Signal Process.*, **SP-52**(2) (2004).
- [Zen00] M. Zeng, A. Annamalai and V. K. Bhargava, Harmonization of global third-generation mobile systems, *IEEE Commun. Mag.*, **38**(12), 94–104 (2000).
- [Zha89] K. Zhang, K. Pahlavan, and R. Ganesh, Slotted ALOHA radio networks with PSK modulation in Rayleigh fading channels, *IEE Electr. Lett.*, **25**, 412–413 (1989).
- [Zha90] K. Zhang and K. Pahlavan, An integrated voice/data system for mobile indoor radio networks, *IEEE Trans. Veh. Technol.*, **VT-39**, 75–82 (1990).
- [Zha91] K. Zhang and K. Pahlavan, The effects of capture on CSMA local radio networks with BPSK modulation in Rayleigh fading channels, *IEEE MILCOM '91 Conf. Rec.*, McLean, VA, (1991), pp. 1172–1176.
- [Zha92] K. Zhang and K. Pahlavan, Relation between transmission and throughput of slotted ALOHA local packet radio networks, *IEEE Trans. Commun.*, **COM-40**, 577–583 (1992).
- [Zha96] Q. T. Zhang, Outage probability in cellular mobile radio due to Nakagami signal and interferers with arbitrary parameters, *IEEE Trans. Veh. Technol.*, **VT-45**(2), 364–372 (May 1996).
- [Zwi98] T. Zwick, C. Fischer, and W. Wiesbeck, A statistical channel model for indoor environments including angle of arrival, *Proc. IEEE Vehicular Technology Conference*, Ottawa, 1998.

INDEX

- Access methods:
 - centrally controlled, 521–537
 - contention-based, 537–571
 - distributed, 537–539
 - fixed-assignment, 521–537
 - for packet-switched data, 531–533
 - integrated, 531
- Adaptive:
 - channel measurement, 386–389
 - discrete matched filter (DMF), 389–390
 - equalization, 393–405
 - MLSE, 393
 - receivers, 378
- Ad hoc networking, 41–43, 46–47, 505–506, 609
- Adjacent channel interference (ACI), 314
- Advanced Mobile Phone Service (AMPS), 25, 30, 33, 510, 681
- ALOHA:
 - dynamic slotted, 546
 - pure, 539
 - reservation (R-ALOHA), 544
 - slotted, 542
 - stability, 543
- Ambient light, 648
- Antenna:
 - diversity, 405–407
 - leaky cable, 54
 - pattern, 408
 - sectorized, 405–409
- Antipodal signaling, 293
- Automatic gain control (AGC), 329
- Autoregressive (AR) modeling, 235–241, 278
 - Akaike information criteria, 236
 - all-pole model, 235
 - of indoor radio channel, 238–241
 - order selection criteria, 236
- Average probability of error:
 - for flat fading channels, 345–347
 - for MRC, 349–351
 - for selection diversity, 353
- Bandwidth efficiency, 287, 312–313
- Bandwidth efficiency of CDMA:
 - with ideal power control, 477–479
 - in IS-95, 479
 - in lognormal fading, 480–481
- Bit error rate (BER):
 - binary modulation techniques, 301–302
 - BPSK/DFE, 400–405
 - CDMA, 478–480, 481–482
 - complementary error function, 287
 - DSSS in fading, 481–482, 455
 - exponential equation, 287
 - FHSS in fading, 492–493
 - irreducible, 380
 - nonbinary modulation techniques, 302–307
- Bluetooth:
 - IEEE 802.11b interference, 466–467
 - initiative, 43
 - packet length, 440
 - specification, 48–49
- Broadband modems:
 - adaptive channel measurement, 386–389
 - adaptive discrete matched filter (ADMF), 389–393
 - adaptive equalization, 393–405, 425–428
 - adaptive MLSE, 393
 - comparisons, 421–423
 - discrete multipath fading model, 384–386
 - effects of frequency-selective fading, 380–384
 - implicit diversity, 384
 - MIMO in fading, 423–424
 - MIMO OFDM, 424–425
 - multiamplitude and multiphase, 419–420
 - multicarrier, 379, 411–418
 - multirate, 418–419
 - OFDM, 415–416, 420, 424–425, 432–434

- Broadband modems (*continued*)
 - sectorized antennas, 405–411
 - STC-MIMO, 423–424
- Busy-tone multiple access (BTMA), 551
- Capture effects, 561
 - in CSMA, 568–571
 - in slotted ALOHA, 562–568, 575–576
- Carrier-sense multiple access (CSMA), 547
 - l-persistent, 547
 - with busy-tone signaling, 551
 - with collision avoidance, 554
 - with collision detection, 552
 - in IEEE 802.11, 554–555
 - exponential backoff, 553
 - nonpersistent, 548
 - p-persistent, 548
 - throughput, table, 549
- Carrier-to-interference (C/I) ratio, 509–510
- CDMA:
 - advantages, 526–527
 - with antenna sectorization, 479
 - bandwidth efficiency, 477–481
 - cdmaOne, 675
 - cdma2000, 675, 677
 - coding in IS-95, 675
 - in digital cellular, 675–676
 - direct sequence, 477
 - frequency hopping, 492–494
 - in frequency selective fading, 526
 - high data rate (HDR), 676–679
 - implementation complexity, 527
 - in IMT-2000, 677
 - interference resistance, 526
 - IS-95 specification, 675–676
 - medium data rate, 674
 - MIMO CDMA, 492
 - multicarrier, 490–491
 - near-far problem, 480–481
 - orthogonal signaling, 481–483
 - overlay, 527
 - performance, 476–481
 - PN sequence, 477
 - in portable and mobile radio, 526–528
 - practical power control, 480
 - Qualcomm system, 528
 - sectorization gain factor, 479
 - soft capacity, 527
 - soft handoff, 527
 - time-hopping, 494
 - timing flexibility, 526
 - voice activity factor, 479
 - Walsh codes, 482
 - W-CDMA for IMT-2000, 39–40
- CDMA power control, 330, 477–481, 504, 528
 - ideal instantaneous, 477–478
 - practical, 480
- Cellular digital packet data (CDPD), 20
- Cellular telephone:
 - architecture, 6–7
 - evolution, 25–27
 - frequency allocations, 33–36
 - three generations, 32
- Centralized network, 503
- Channel capacity:
 - AWGN channel, 307
 - voice-band modems, 308
- Channel measurement, 386–389
 - MMSE estimate, 387
 - using cross-correlation, 387
 - using LMS algorithm, 387
- Channel modeling and simulation:
 - for AOA, 232–233
 - autoregressive, 235–241, 278
 - bell-shaped spectrum model, 133
 - Clarke-Jakes model, 129–131
 - COST-231 model, 109
 - FDTD, 85, 207, 262–263
 - flat-spectrum model, 131
 - GSM model, 212–214
 - IEEE 802.11b model, 216–218
 - JTC models, 214–215
 - for MIMO, 233–234
 - for MIMO, UWB and positioning, 85
 - Okumura-Hata model, 109
 - ray tracing, 245–261
 - Saleh-Valenzuela model, 218–219
 - wideband models, 211–219
- CNR or C/N, 287
- Co-channel interference, 508
- Coding:
 - ARQ, 359–360, 675
 - Barker, 447
 - BCH, 357
 - block, 354–356
 - CCK in IEEE802.11b, 483
 - convolutional, 356
 - CRC, 354
 - effects of fading, 360–363
 - FEC, 354
 - for FHSS systems, 474–476
 - for MIMO systems, 366–370, 492
 - for multiband OFDM, 600
 - Golay, 357
 - Hadamard, 358
 - Hamming, 357
 - hard- and soft-decision decoding, 355
 - interleaving, 361
 - in IS-95, 675
 - Reed-Solomon, 358
 - scrambling, 451
 - space-time (STC), 363–365
 - spreading, 445
 - Turbo, 358
 - Viterbi algorithm, 356

- Walsh codes, 358, 482, 676
- wireless applications, table, 357
- Coherence bandwidth, 75
- Contention schemes, 537
- Centrally controlled access:
 - polling, 532
 - PRMA, 532
- Cordless telephone, 6, 26–27, 30, 679
- CT2 telepoint, 679
- Data-oriented networks:
 - development issues, 20, 27–29
 - history, table, 28
 - systems and services, 5–6, 37–39
 - vs. voice-oriented networks, 5–6
- Data-sense multiple access (DSMA), 552
- DCS-1800, 25, 27
- Decision feedback equalizer, 395–396, 400–405
- DECT, 27, 680
- Delay power spectrum, 74, 75, 212, 267–270, 400
 - equivalent, 401
 - exponential, 391
 - in Saleh-Valenzuela model, 218–219
- Digital cellular:
 - beyond 3G, 41–43
 - CDMA, IS-95, 34, 675–676
 - cdmaOne, 675
 - cdma2000, 675, 677
 - EDGE, 672–674
 - GPRS, 669–672
 - GSM, 25, 34, 35, 664–669
 - HDR, 676
 - IMT-2000, 39–41, 677
 - Japanese (JDC), 34, 682
 - motivation, 20, 26
 - second-generation (2G), 34–37
 - table of standards, 34
 - third-generation (3G), 39–41
 - UMTS, 40
 - USDC, IS-54, IS-136, 25, 327, 681
 - W-CDMA for IMT-2000, 39–41
- Direct sequence spread spectrum (DSSS):
 - effect of fading, 449–451
 - effect of interference, 471–473
 - implementation, 444–449
 - interference suppression, 473–474
 - RAKE receiver, 451–455, 456, 459–461
 - SAW correlator, 448, 452
 - synchronization, 446
 - transform domain processing, 473
- Distance-power relationship:
 - in buildings, 96–97, 101–102, 105
 - in cellular networks, 509–510
 - in free space, 57
 - in mobile radio networks, 59–60
- Distribution functions:
 - lognormal, 123
 - Nakagami, 124
 - Rayleigh, 122
 - Rician, 123
 - Suzuki, 123
 - Weibull, 124
- Diversity:
 - antenna, 405–407
 - explicit, 454
 - frequency, 348
 - implicit, 384, 451
 - in-band, 384, 451
 - internal, 384, 451
 - polarization, 405
 - spatial, 405
 - time, 348, 451
- Diversity combining:
 - linear, 349
 - maximum-ratio (MRC), 349–351, 452
 - performance evaluations, 349–351
 - selection, 349, 353
- Doppler:
 - effect, 64
 - frequency, 65, 112
 - local movements, 64–66
 - rms bandwidth, 117
 - rms spread, 79, 80
 - shift, 65–66, 91, 112, 119
 - spectrum, 66, 78, 110
 - spread, 66, 79, 113, 114
- Doppler power spectrum, 66, 78
 - bell-shaped, 133
 - flat, 131
- Energy contrast ratio (E_b/N_0), 287
- Enhanced data for global evolution (EDGE), 672–674
 - capacity, example, 535–536
 - ECSD data rates, 674
 - EGPRS modulation and coding parameters, 673
 - enhanced circuit-switched data (ECSD), 673–674
 - enhanced GPRS (EGPRS), 673
 - link quality control (LQC), 673
 - use of R-ALOHA, 547
- Equalization:
 - adaptive, 395–405
 - bandpass, 396
 - blind, 396
 - decision feedback, 395–396, 400–405, 407
 - for indoor radio channel, 407–411
 - fractionally spaced, 395
 - linear transversal, 394–395
 - LMS algorithm, 398–400
 - tap gain divergence, 395
 - tap leakage algorithm, 395
 - vs. RAKE, 489
 - zero-forcing, 394
- Error-control coding:
 - ARQ, 359–360

Error-control coding (*continued*)

- block, 354–356
- convolutional, 356
- fading effects, 360–363
- Turbo, 358

Etiquette for unlicensed operation, 583

European Telecommunications Standards Institute (ETSI), 664, 672

Fading:

- average error rate, 344–345
- bell-shaped spectrum, 133
- Clarke model for mobile radio, 111–113
- Clarke-Jakes simulation model, 129–131
- complex Gaussian process, 73
- duration, 120–121
- flat spectrum, 131
- frequency-nonselective, 343
- frequency-selective, 77
- margin, 103–104
- multipath, 57–62
- probability of outage, 342, 345–346
- rate, 118
- Rayleigh, 95, 122, 343–347
- shadow fading or large-scale fading, 94, 102, 142
- simulation of fast fading, 139–140
- simulation using Clarke-Jakes model, 129–131
- simulation using filtered Gaussian noise, 127–129
- simulation using JTC model, 140–142

Finite difference time-domain (FDTD) modeling, 85, 207

- computation time, 264
- description, 262–263

Fixed-assignment access methods:

- CDMA, 502, 511–512, 525–528
- FDMA, 502, 508, 510–511, 521–522
- TDD, 523, 560
- TDMA, 502, 508–511, 522, 523–525

Frequency division multiplexing (FDM), 521

Frequency hopping spread spectrum (FHSS):

- in Bluetooth, 440, 466
- effect of fading, 442–443
- effect of coding, 474–476
- effect of interference, 465–470
- fast, 474
- in IEEE 802.11, 440
- in GSM, 443
- power control, 493
- slow, 474
- TDMA, 442, 493

Frequency reuse, 35, 507, 510, 512

General packet radio service (GPRS), 664, 669–672

- link adaptation, 671
- multiframe structure, 671

- network architecture, 670
- overlay onto GSM, 670
- radio block encoding, 671–672

Geolocation, *see* RF location sensing

Global positioning system (GPS), 607, 608

Global system for mobile communication (GSM):

- control channels, 668
- ECSD, 673–674
- EDGE, 672–673
- EGPRS, 673
- frame format, 666–669
- frequency allocation, 34, 664–665
- GPRS, 669–672
- HSCSD, 673
- logical channels, 666
- mobility management, 665
- network architecture, 665–666
- Phase 1 specifications, 664
- Phase 2 specifications, 665
- propagation models, 270–272
- radio interface, 666
- speech coder, 668, 686

Hidden terminal problem, 551, 569,

High data rate (HDR), 676–679

cdma2000 1 x EV-DO, 677

forward-link slot structure, 677–678

forward-link data rates, 678–679

turbo encoding, 679

High frequency (HF), 73

Home access networks, 45–46

IEEE 802.11 standards, 19–21, 28–29, 682–685

IEEE 802.15 WPAN working group, 47, 588, 596

IEEE 802.15.3a UWB working group, 596, 599

IMT-2000, 40, 677

Infrared (IR) signals:

- ambient lights, 648–651
- avalanche photodiode (APD), 643
- data-rate limitations, 644–653
- diffused IR (DFIR), 641, 658
- directed-beam IR (DBIR), 642
- field of view (FOV), 645
- maximum permissible exposure (MPE), 643
- PIN diode, 643
- propagation simulation, 646–648
- wavelength selection, 643
- WLANs, 655, 660

Infrared (IR) transmission:

- amplitude modulation, 654
- decision feedback equalizer, 659
- frequency division multiplexing (FDM), 654
- frequency shift keying (FSK), 656, 659
- maximum rate, 647–648
- multiamplitude modulation, 643
- multicarrier modulation, 656
- on-off keying (OOK), 654
- phase shift keying (PSK), 655, 659

- pulse amplitude modulation (PAM), 654
- pulse code modulation (PCM), 649
- pulse duration modulation (PDM), 654
- pulse position modulation (PPM), 654
- received power, 645–646
- wavelength modulation, 653
- Intersymbol interference, 333
- Ionospheric communication, 73
- IS-54 standard, 35, 681
- IS-95 standard, 35, 675–676
- IS-136 standard, 35, 681
- ISM bands, 28
- JDC/PDC, 682
- Joint Technical Committee (JTC), 103
 - channel model, 103–104, 272–274, 277
 - equivalent autoregressive model, 236
 - macrocell path loss model, 107
 - path loss model, 103, 107–108
 - PCS band path loss model, 103
- Lambertian law, 644–645
- Land mobile radio (LMR), 19, 313
- Least-mean-square (LMS) algorithm, 387, 398–400
- Macrocells, 107–110, 508
- Manchester-coded signal, 658
- Matched filter, 284–286, 339–340
- Maxwell's equations, 57, 85
- Measurement plans:
 - frequency dependence, 150
 - local or small-scale, 150, 165
 - partitioned, 150
 - spatial, 165
 - traffic-effects or temporal, 150
- Measurement results:
 - angle of arrival (AOA), 186–191
 - CDF of rms delay spread, 167–169, 177–178
 - coherence bandwidth, 178–179
 - distance-power gradient, 166
 - in frequency domain, 176–179, 181–183
 - in indoor areas, 174–178
 - received power, 170–171
 - spatial, 165–170
 - temporal variations, 169–171
 - time of arrival (TOA), 181–186
 - for UWB, 195–197
- Measurement systems:
 - comparison between, 179–180
 - discrete maximum likelihood, 194–195
 - DSSS as virtual pulse, 155–158
 - frequency domain, 171–176
 - for MIMO, 186, 189
 - MUSIC algorithm, 185
 - pulse transmission, 152–154
 - spatial filter periodogram, 191–194
 - superresolution algorithms, 183–186
 - time-domain, 151–170
 - TLS-ESPRIT, 184
 - using network analyzer, 172–173, 200
 - using sliding correlator, 154–165
- Microcell, 508
 - COST-231 model, 109–110
 - JTC path loss model, 107
 - Okumura-Hata model, 109–110
 - in urban environment, 259–261
- Minimum distance:
 - block code, 355
 - signal constellation, 289
- Mobile data services:
 - ARDIS, 20, 29, 37–38
 - CDPD, 20, 29
 - GPRS, 29
 - Mobitex, 20, 29, 37
 - overview, 37–38
 - RAM Mobile Data, 28
 - table of parameters, 37
 - TETRA, 37
- Mobile radio:
 - Clarke-Jakes simulation model, 129–131
 - isotropic scattering model, 129
- Mobile switching center (MSC), 10, 507
- Mobile telephone service evolution, 25–27
- Mobitex, 20, 29, 37
- Modeling, statistical, 208
- Modems, *see* Broadband modems; Narrowband modems
- Multicarrier CDMA, 490–491
- Multipath:
 - delay power spectrum, 74, 75
 - delay spread, 71, 74, 75, 94
 - excess delay spread, 71, 74
 - fading, 60–62, 73, 110
 - frequency-selective fading, 77
 - rms delay spread, 71–72, 77, 170
- Multiple input multiple output (MIMO) systems, 365
 - AOA simulation, 232
 - capacity limits, 369–370
 - channel modeling, 85–86
 - design of codes, 366
 - in frequency-selective fading, 423
 - MIMO CDMA, 492
 - MIMO OFDM, 424–425
 - practical considerations, 370
 - Spencer's model, 232–233
 - STC-MIMO, 423–424
- Multiuser detection (MUD), 487–489
- Narrowband:
 - interference canceller, 473
 - measurements, 99–102, 113–118, 119–120, 124–126

Narrowband (*continued*)

- multipath fading effects, 110–113, 118–122, 122–124, 129–133

Narrowband modems, 281

- 4-ary FSK, 293, 363
- carrier and timing recovery, 330–332
- carrier-to-noise ratio (CNR or C/N), 287
- constant envelope, 315
- digital FM, 315
- DPSK, 294
- error rate, 287–289, 301–307
- FSK, 291–293
- GMSK, 292, 320–324
- linear modulation, 315
- matched filtering, 284–286
- M -phase PSK, 299
- MSK, 292, 318–320
- multiphase, 299
- OOK, 289–291
- OQPSK, 317–318
- PAM, 294–296
- partial response signaling, 300
- $\pi/4$ -shift QPSK, 324–328
- power control, 329
- PSK, 293–294
- pulse shaping, 332–335
- QAM, 296–299
- QPR, 300
- Signal-to-noise ratio (SNR), 286–287
- TCM, 282, 300–301
- theoretical limits, 307–311
- voiceband, 308

Near-far effect, 505, 561

Network topologies:

- ad hoc, 505–506
- cellular, 506–508
- centralized, 504–505
- fully connected, 505
- multihop, 505, 505–506
- peer-to-peer, 504

North American digital cellular, *see* USDC

OSI seven-layer reference model, 13

Packet reservation multiple access (PRMA), 532–533

Path amplitude model:

- lognormal, 231
- Nakagami, 124
- Rayleigh, 224, 230, 343–347
- Rician, 123
- Suzuki, 123
- Weibull, 124

Path arrivals:

- clusters, 232
- GSM model, 212–213, 270–271

- IEEE 802.11b model, 216–218

- IEEE 802.11 MIMO model, 233–234
- interarrival delays, 224

- JTC model, 214–215, 272–274

- Poisson distribution model, 220–222

- Poisson model, modified, 222–224

- Saleh-Valenzuela model, 218–219, 233

- Spencer model for AOA, 232–233

- Suzuki model, 221

Path loss modeling:

- IEEE 802.15 WPAN models, 589–590

- Intel model for UWB, 588–589

- JTC, 103–108

- macrocell, 107–109

- microcell, 107–109

- mobile radio, 59–60

- partitioned indoor, 105–107

Personal communication services (PCS):

- comparison with cellular, 36

- CT-2, 27, 679

- DCS-1800, 27

- DECT, 680

- overview, 36–37

- table of standards, 36

Phase locked loop, 331

Physical operating environment:

- indoor commercial, 81

- indoor office, 62, 81

- indoor residential, 81, 83

- outdoor residential, 82

- outdoor urban high-rise, 82, 83

- outdoor urban/suburban low-rise, 82, 83

- table of Doppler shift ranges, 83

Picocells, 508

Polling, 532

Positioning, *see* RF location sensing

Probability of outage:

- for erfc error functions, 345

- for exponential error functions, 345

- for MRC, 351

Propagation in free space, 57

Pseudonoise (PN) sequence in DS-CDMA, 477

Pulse shaping, 332–335

Raised cosine filter, 333

RAKE receiver:

- in IS-95, 453

- performance, 453–464

- structure, 451–452

RAM mobile, *see* Mobitex

Random access methods:

- ALOHA-based, 539–547

- CSMA-based, 547–560

- DSMA, 547, 552

- throughput comparisons, 549–551

Ray tracing:

- computation time, 264

- diffraction coefficient, 253–254

- Fresnel integral, 254
- image method, 245–246
- ray shooting method, 246
- reflection coefficient, 247–250
- scattering, 254–255
- slab model, 251–253
- three dimensional, 259–261
- two dimensional, 62, 255–259
- uniform theory of diffraction, 253
- wall-transmission coefficient, 250–251
- Receiver:
 - adaptive DMF, 389–390
 - adaptive MLSE, 393
 - RAKE, 389
- Reflection coefficient, 247–250
- RF location sensing, 607
 - AOA techniques, 608, 618–619
 - functional block diagram, 610
 - GPS-assisted, 608
 - nearest-neighbor method, 631
 - overview of techniques, 607–610
 - pattern recognition techniques, 629–631
 - positioning algorithms, 626–636
 - RSS techniques, 618
 - statistical approach, 632
 - superresolution for TOA, 614
 - TDOA techniques, 608
 - TOA techniques, 611–618
 - ultrawideband approach, 614
 - undetected direct path (UDP), 616–618
- SAW correlator, 448
- Scattering function, 80
- Sectorized antenna, 405–411
- Shannon:
 - bandwidth-limited region, 309
 - capacity formula, 308–309
 - limit, 310
 - power-limited region, 309
- Signal constellation, 289
 - minimum distance, 289
 - symbol error rate, approximated, 289
- Signal-to-noise ratio (SNR):
 - per bit, 286
 - alternative interpretations, 286
 - and matched filtering, 284
- Sliding correlator, 155, 159, 162–164
- Source coding, 329
- Space-time coding (STC):
 - Alamouti codes, 364
 - description, 363–365
 - with MIMO, 365–366
- Speech coding in wireless systems, 329, 675, 685–686
- Spread spectrum:
 - anti-interference characteristics, 437
 - anti-multipath characteristics, 437
 - autocorrelation function, 445
 - bandwidth expansion factor, 445
 - in Bluetooth, 440, 466–467, 470–471, 477
 - CCK codes in IEEE 802.11b, 483
 - chips, 444
 - direct sequence (DSSS), 444–449
 - effects of coding, 474–476
 - frequency hopping (FHSS), 439–443
 - in frequency-selective fading, 437–438, 442–443
 - in GSM, 443
 - in IEEE 802.11, 440, 447
 - interference, 465–474
 - interference margin, 467
 - in IS-95, 446, 451, 479
 - in multipath fading, 449
 - M*-ary orthogonal codes, 481
 - M*-sequences, 439
 - overlay, 437, 465
 - processing gain, 454
 - pseudonoise (PN) sequence, 157, 439, 445
 - RAKE receiver, 449–464
 - spreading code, 445
 - spreading factor, 445, 458
 - Walsh codes in IS-95, 482–483
- Steepest descent algorithm, 388
- Superresolution algorithm, 183–186
- TDMA:
 - comparison with FDMA, 523–525
 - in digital cellular, 522
 - with FHSS, 442, 493
- Time-division duplex (TDD), 523, 560
- Time-division multiplexing (TDM), 522
- Timing recovery, 332
- Trans-European Trunked Radio (TETRA), 37–38
- Troposcatter, 73, 75–77, 391, 400
- Ultrahigh frequency (UHF), 73, 293
- Ultrawideband (UWB), 581
 - channel modeling, 85–86
 - direct-sequence (DS-UWB), 595–599
 - FCC spectral mask, 583
 - IEEE 802.15, 588
 - IEEE 802.15.3a, 596, 597
 - impulse radio, 581, 589–595
 - Intel model, 588
 - MBOA Alliance, 599
 - modeling of multipath behavior, 586–589
 - multiband OFDM (MB-OFDM), 599–603
 - multiuser environment, 594–595
 - overlay, 589
 - path-loss modeling, 584–586
 - pulse shape and antenna, 589–593
 - table of TF codes for MC-OFDM, 602
 - Time Domain Corporation, 590
 - TFMA, 599–601
 - time-hopping, 593–595

- Universal Mobile Telecommunications System (UMTS), 40, 359
- United States digital cellular (USDC), 327, 681
- Very high frequency (VHF), 293
- Voice-oriented networks:
 - chronology, table, 25
 - development issues, 20, 25–27
 - systems and services, 32–37
 - vs. data oriented networks, 5–6
- Wideband:
 - autoregressive (AR) modeling, 235–241
 - CDMA (W-CDMA), 39–40
 - channel simulation, 210–211
 - frequency-domain channel modeling, 234
 - measurements, 171–180
 - multipath fading effects, 66
 - statistical modeling, 208, 241
 - statistical models, comparison, 243
- Wireless Internet, 7
- Wireless local area network (WLAN):
 - comparisons, 38–39
 - development, 20–21, 43–44
 - general description, 38–39
 - HIPERLAN2, 38–39, 682
 - IEEE 802.11 standards, 19–21, 28–29, 38–39, 682–685
 - infrared, 639–643, 660
 - ISM bands, 28
 - public, 44
 - spread-spectrum, 684
 - Wi-Fi Alliance, 44
- Wireless personal area network (WPAN):
 - Bluetooth, 43, 47, 48–49
 - BodyLAN, 43
 - emergence, 43
 - HomeRF, 43, 48
 - IEEE 802.15 WPAN working group, 47
- Wireless PBX, 27
- Wireless technologies in standards, tables, 33–34
- WSSUS channel model, 72–81
 - coherence bandwidth, 75
 - coherence time, 80
 - correlation functions, diagram, 82
 - delay power spectrum, 74
 - Doppler power spectrum, 78
 - Doppler spread, 79
 - rms Doppler spread, 79–80
 - scattering function, 80–81
 - spaced-time, spaced-frequency correlation function, 78
 - troposcatter channel, 75–77

ABOUT THE AUTHORS



Kaveh Pahlavan is a Professor of Electrical and Computer Engineering, and Computer Science, and Director of the Center for Wireless Information Network Studies, Worcester Polytechnic Institute, Worcester, MA. He is also a Visiting Professor at the Telecommunication Laboratory and Center for Wireless Communications, University of Oulu, Finland. His area of research is location aware broadband sensor and ad hoc networks. He has contributed to numerous seminal technical and visionary publications in wireless office information networks, home networking, and indoor geolocation science and technology. He is the principal author of the *Wireless Information Networks* (with Allen Levesque), John Wiley & Sons, 1995, and *Principles of Wireless Networks—A Unified Approach* (with P. Krishnamurthy), Prentice Hall, 2002. Before joining WPI, he was the Director of Advanced Development at Infinite, Inc., Andover, MA, working on data communications. He started his career as an

Assistant Professor at Northeastern University, Boston, MA. He is the founder and Editor-in-Chief of the *International Journal on Wireless Information Networks*, the first journal in modern wireless networks established in 1994, and a member of the advisory board of the *IEEE Wireless Magazine*. He was the founder, program chairman, and organizer of the IEEE Wireless LAN Workshop, Worcester, MA, in 1991, 1996, and 2001; organizer and technical program chairman of the IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications, Boston, MA, in 1992 and 1998; co-chair of the International Workshop on Ultra Wideband Systems, Oulu, Finland, 2003; co-chair of the International Workshop on Wireless Adhoc Network, Oulu, Finland, 2004; and chairman of the IEEE International Conference on Mobile Adhoc Sensor Systems, 2005. He has also been selected as a member of the Committee on Evolution of Untethered Communication, U.S. National Research Council, 1997, and has led the US review team for the Finnish R&D Programs in Electronic and Telecommunication in 1999 and NETs project in 2003. For his contributions to wireless networks, he was named the Weston Hadden Professor of Electrical and Computer Engineering at WPI in 1993–1996, was elected a Fellow of the IEEE in 1996, and became a Fellow of Nokia in 1999. From May to December 2000, he was the first Fulbright-Nokia scholar at the University of Oulu, Finland. He has been a consultant to a number companies, including CNR, GTE Laboratories, Steinbrecher Corp., Simplex, Mercurry Computers, WINDATA, SieraComm, 3COM, and Codex/Motorola in Massachusetts; JPL, Savi Technologies, RadioLAN in California; Aironet in Ohio; United Technology Research Center in Connecticut; Honeywell in Arizona; Nokia, LK-Products, Elektrobit, TEKES, and Finnish Academy in Finland; and NTT in Japan. Because of his inspiring visionary publications and his international conference activities for the growth of the wireless LAN industry, he is referred to as one of the founding fathers of the wireless LAN industry. Details of his contributions to this field are available at www.cwins.wpi.edu.

Allen H. Levesque is an affiliated faculty member in the Electrical and Computer Engineering Department at Worcester Polytechnic Institute, where he has been teaching digital communications and conducting research in the Center for Wireless Information Network Studies. He is also an independent consulting engineer specializing in digital communications technologies, with particular emphasis on wireless communications. Dr. Levesque is a Fellow of the IEEE and a Registered Professional Engineer in the Commonwealth of Massachusetts. In early 1999, he retired from the position of Senior Staff Scientist at GTE Laboratories, where he had been responsible for coordinating GTE (now Verizon) research activities in wireless and broadband communications, as well as providing support to the company's business initiatives in cellular, wireless data, wireless local loop, and indoor wireless communications.



WILEY SERIES IN TELECOMMUNICATIONS AND SIGNAL PROCESSING

John G. Proakis, Editor
Northeastern University

Introduction to Digital Mobile Communications
Yoshihiko Akaiwa

Digital Telephony, 3rd Edition
John Bellamy

ADSL, VDSL, and Multicarrier Modulation
John A. C. Bingham

Biomedical Signal Processing and Signal Modeling
Eugene N. Bruce

Elements of Information Theory
Thomas M. Cover and Joy A. Thomas

Erbium-Doped Fiber Amplifiers: Device and System Developments
Emmanuel Desurvire

Fiber-Optic Systems for Telecommunications
Roger L. Freeman

Fundamentals of Telecommunications, 2nd Edition
Roger L. Freeman

Practical Data Communications, 2nd Edition
Roger L. Freeman

Radio System Design for Telecommunications, 2nd Edition
Roger L. Freeman

Telecommunication System Engineering, 4th Edition
Roger L. Freeman

Telecommunications Transmission Handbook, 4th Edition
Roger L. Freeman

Introduction to Communications Engineering, 2nd Edition
Robert M. Gagliardi

Optical Communications, 2nd Edition
Robert M. Gagliardi and Sherman Karp

Efficient Algorithms for MPEG Video Compression
Dzung Tien Hoang and Jeffrey Scott Vitter

Active Noise Control Systems: Algorithms and DSP Implementations
Sen M. Kuo and Dennis R. Morgan

Mobile Communications Design Fundamentals, 2nd Edition
William C. Y. Lee

Expert System Applications for Telecommunications
Jay Liebowitz

Polynomial Signal Processing
V. John Mathews and Giovanni L. Sicuranza

Digital Signal Estimation

Robert J. Mammone, Editor

Digital Communication Receivers: Synchronization, Channel Estimation, and Signal Processing

Heinrich Meyr, Marc Moeneclaey, and Stefan A. Fechtel

Synchronization in Digital Communications, Volume I

Heinrich Meyr and Gerd Ascheid

Business Earth Stations for Telecommunications

Walter L. Morgan and Denis Rouffet

Wireless Information Networks, 2nd Edition

Kaveh Pahlavan and Allen H. Levesque

Satellite Communications: The First Quarter Century of Service

David W. E. Rees

Fundamentals of Telecommunication Networks

Tarek N. Saadawi, Mostafa Ammar, with Ahmed El Hakeem

Analogue and Digital Microwave Links: Theory and Design

Carlos Salema

Microwave Radio Links: From Theory to Design

Carlos Salema

Meteor Burst Communications: Theory and Practice

Donald L. Schilling, Editor

Digital Communication over Fading Channels, 2nd Edition

Marvin K. Simon and Mohamed-Slim Alouini

Digital Signal Processing: A Computer Science Perspective

Jonathan (Y) Stein

Vector Space Projections: A Numerical Approach to Signal and Image Processing, Neural Nets, and Optics

Henry Stark and Yongyi Yang

Signaling in Telecommunication Networks

John G. van Bosse

Telecommunication Circuit Design, 2nd Edition

Patrick D. van der Puije

Worldwide Telecommunications Guide for the Business Manager

Walter H. Vignault