Principles of
MOBILE
COMMUNICATION
Second Edition

Gordon L. Stüber

阅读声明

Principles of
Mobile Communication
Second Edition

This page intentionally left blank

Principles of
Mobile Communication
Second Edition

Gordon L. Stüber
Georgia Institute of Technology
Atlanta, Georgia USA

Created in the United States of America

Contents

Preface

This book follows from my first edition and is intended to provide a thorough, up to date, treatment of wireless physical communications. The book is derived from a compilation of course material that I have taught in a graduate-level course on physical wireless communications at Georgia Tech over the past decade. This textbook differs from others on the subject by stressing mathematical modeling and analysis. My approach is to include detailed derivations from first principles. The text is intended to provide enough background material for the novice student enrolled in a graduate level course, while having enough advanced material to prime the more serious graduate students that would like to pursue research in the area. The book is intended to stress the *fundamentals* of mobile communications engineering that are important to *any* mobile communication system. I have therefore kept the description of existing and proposed wireless standards and systems to a minimum. The emphasis on fundamental issues should benefit not only to students taking formal instruction, but also practicing engineers who are likely to already have a detailed familiarity with the standards and are seeking to deepen their knowledge of the fundamentals and principles of this important field.

Chapter 1 begins with an overview that is intended to introduce a broad array of issues relating to wireless communications. Included is a description of various wireless systems and services, basic concepts of cellular frequency reuse, and the link budget for cellular radio systems.

Chapter 2 treats propagation modeling and was inspired by the excellent reference by Jakes. It begins with a summary of propagation models for narrow-band and wide-band multipath channels, and provides a discussion of channel simulation techniques that are useful for radio link analysis. It concludes with a discussion of shadowing and path loss models. Chapter 3 is a related chapter that provides a detailed treatment of co-channel interference, the primary impairment in high capacity cellular systems.

Chapter 4 covers the various types of modulation schemes that are used in mobile communication systems along with their spectral characteristics. Chapter 5 discusses the performance of digital signal on narrow-band flat fading channels with a variety of receiver structures, while Chapter 6 includes a treatment of antenna diversity techniques.

Chapter 7 provides an extensive treatment of digital signaling on the fading ISI channels that are typical of mid-band land mobile radio systems. The chapter begins with the characterization of ISI channels and goes on to discuss techniques for combating ISI based on symbol-by-symbol equalization and sequence estimation. The chapter concludes with a discussion of co-channel demodulation and co-channel interference cancellation.

Chapter 8 covers bandwidth efficient coding techniques. The chapter begins with a discussion of basic block and convolutional coding. It then goes on to a detailed discussion on the design and performance analysis of convolutional and trellis codes for additive white Gaussian noise channels, and interleaved flat fading channels. The chapter concludes with an introduction to Turbo coding.

Chapter 9 is devoted to spread spectrum techniques. The chapter begins with an introduction to direct sequence and frequency hop spread spectrum. This is followed by a detailed treatment of spreading sequences. Also included is a discussion of the effects of tone interference on direct sequence spread spectrum, and the RAKE receiver performance on wide-band channels. The chapter wraps up with a discussion of the error probability of direct sequence code division multiple access.

Chapter 10 considers TDMA cellular architectures. The chapter begins with a discussion of conventional TDMA systems and how they are evolved to meet traffic growth. This is followed by hierarchical overlay/underlay architectures. Finally, the chapter wraps up with macrodiversity TDMA architectures. Chapter 11 is the CDMA counterpart to Chapter 10 and considers issues that are relevant to cellular CDMA, such as capacity estimation and power control.

Chapter 10 covers the important problem of link quality evaluation and handoff initiation, and handoff performance, in cellular systems. Chapter 11 provides an overview of the various channel assignment techniques that have been proposed for FDMA and TDMA cellular systems.

The book contains far too much detail to be taught in a one-semester course. However, I believe that it can serve as a suitable text in most situations through the appropriate selection of material. My own preference for a one-semester course is to include the following in order: Chapter 1, Chapter 2, Sections 3.1 and 3.2, Chapter 4, Chapter 5, and Chapter 6. Then choose from Chapters 8 through 13 depending on my interest at the time.

I would like to acknowledge all those who have contributed to the preparation of this book. The reviewers Vijay Bhargava at the University of Victoria and Sanjiv Nanda at Lucent Technologies were very valuable in the early

stages of the first edition of this book. The subsequent review by Upamanyu Madhow at the University of Illinois and in particular the detailed review by Keith Chugg at the University of Southern California were highly useful for improving this book. I am grateful to my doctoral students, past and present, who have contributed significantly to this book. The contributions of Wern-Ho Sheen, Khalid Hamied, Mark Austin, Jeff (Lihbor) Yiin, Ming-Ju Ho, Li-Chun (Robert) Wang, Krishna Narayanan, Dukhyun Kim, Jinsoup Joung, and John (Yongchae) Kim are particularly noteworthy. Finally, I would like to thank BellSouth, GTE Labs, Motorola, Panasonic, Hitachi, Nortel, Korea Telecom, WiLAN, and the National Science Foundation, for sustaining my research efforts in wireless communications over the past 10 years. This research experience has in many cases lead to material that I brought to the classroom and have included in this book.

<div style="text-align: right">GORDON L. STÜBER</div>

This page intentionally left blank

To my parents
Beatrice and Lothar Stüber

This page intentionally left blank

Chapter 1

INTRODUCTION

Wireless systems and services have undergone a remarkable development, since the first cellular and cordless telephone systems were introduced in the early 1980s. First generation cellular and cordless telephone systems were based on analog FM technology and designed to carry narrow-band circuit switched voice services. Second generation cellular and cordless telephone systems were introduced in the early 1990s that use digital modulation, and offer improved spectral efficiency, and voice quality. However, these second generation systems are still used for narrow-band voice and data services. Third generation wireless systems, currently under development that offer substantially higher bit rates ranging from 9.6 kb/s for satellite users, 144 kb/s for vehicular users, 384 kb/s for pedestrian users to 2.048 Mb/s for indoor office environments. These systems are intended to provide voice, data, the more bandwidth intensive multimedia services, while satisfying more stringent availability and quality of service (QoS) requirements in all types environments. Fourth generation systems are also on the horizon that will provide broadband wireless access with asymmetric bit rates that approach 1 Gb/s.

Radio access systems are often distinguished by their coverage areas and bit rates, as shown in Fig. 1.1. **Mobile satellite** systems provide global coverage to mobile users, but with very low bit rates. **Land mobile radio systems** use terrestrial cellular and microcellular networks to provide wide area coverage to vehicular and pedestrian users. **Fixed wireless** access systems provide radio connectivity over a campus or neighborhood area to stationary users. Finally, **wireless local area networks** provide stationary in-building users with very high speed services.
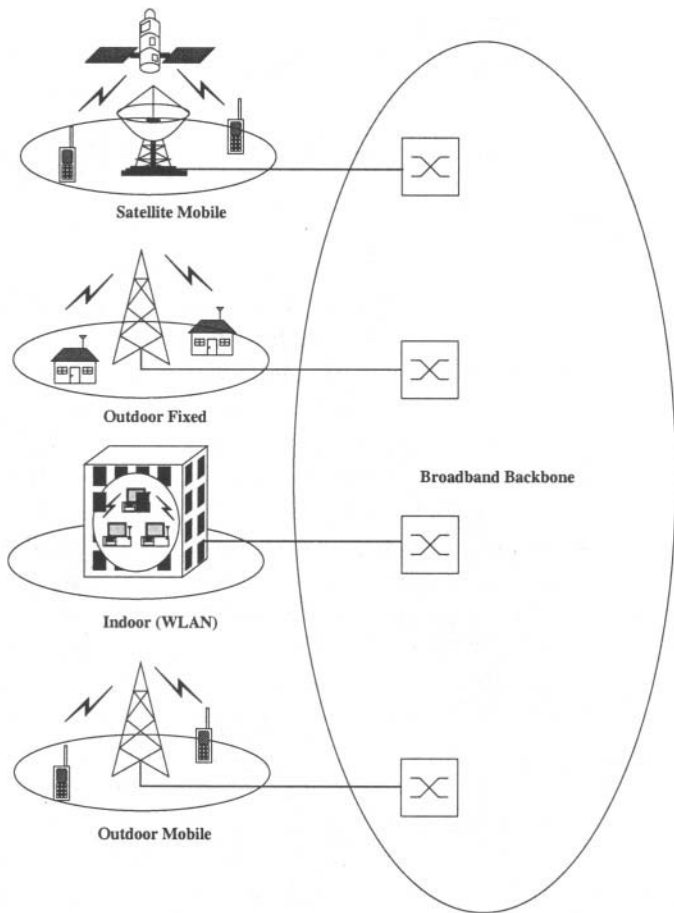
Figure 1.1. Wireless Services.

1.   WIRELESS SYSTEMS AND STANDARDS

1.1   FIRST GENERATION CELLULAR SYSTEMS

The early 1970s saw the emergence of the radio technology that was needed for the deployment of mobile radio systems in the 800/900 MHz band at a reasonable cost. In 1976, the World Allocation Radio Conference (WARC) approved frequency allocations for cellular telephones in the 800/900 MHz band, thus setting the stage for the commercial deployment of cellular systems. In the early 1980s, many countries deployed first generation cellular systems based on frequency division multiple access (FDMA) and analog FM technology. With FDMA there is a single channel per carrier. When a MS accesses the system two carriers (channels) are actually assigned, one for the forward (base-to-mobile) link and one for the reverse (mobile-to-base) link. Separation of the forward and reverse carrier frequencies is necessary to allow implementation of a **duplexer,** a complicated arrangement of filters that isolates the forward and reverse channels, thus preventing a radio transceiver from jamming itself.

In 1979, the first analog cellular system, the Nippon Telephone and Telegraph (NTT) system, became operational. In 1981, Ericsson Radio Systems AB fielded the Nordic Mobile Telephone (NMT) 900 system, and in 1983 AT&T fielded the Advanced Mobile Phone Service (AMPS) as a trial in Chicago. Several other first generation analog systems were also deployed in the early 1980s including TACS, ETACS, NMT 450, C-450, RTMS, and Radiocom 2000 in Europe, and JTACS/NTACS in Japan. The basic parameters of NTT, NMT, and AMPS are shown in Table 1.1. The NMT 900 system uses frequency interleaved carriers with a separation of 12.5 kHz such that overlapping carriers cannot be used with the same base station. In the NTT, NMT, and AMPS systems, a separation of 45 MHz is used between the transmit and receive frequencies, so as to implement the duplexer.

1.2   SECOND GENERATION CELLULAR SYSTEMS

Second generation digital cellular systems have been developed throughout the world. These include the GSM/DCS1800/PCS1900 standard in Europe, the PDC standard in Japan, and the IS 54-/136 and IS-95 standards in the United States. Parameters of the air interfaces of these standards are summarized in Tabs. 1.2 and 1.3, and a brief description of each follows.

1.2.1   GSM/DCS1800/PCS1900

European countries seen the deployment of incompatible first generation cellular systems that prevented roaming throughout Europe. As as result, the Conference of European Postal and Telecommunications Administrations (CEPT) established Groupe Speciale Mobile (GSM) in 1982 with the mandate of defining standards for future Pan-European cellular radio systems. The GSM

Feature	NTT	NMT	AMPS
Frequency Band RL/FL[a] (MHz)	925-940/870-885 915-918.5/860-863.5 922-925/867-870	890-915/917-950	824-849/869-894
Carrier Spacing (kHz)	25/6.25 6.25 6.25	12.5[b]	30
Number of Channels	600/2400 560 280	1999	832
Modulation	analog FM	analog FM	analog FM

[a] RL = reverse link, FL = forward link
[b] frequency interleaving using overlapping channels, where the channel spacing is half the nominal channel bandwidth.

Table 1.1.   First generation cellular standards

R	TCH	C	SW	C	TCH	R	G
3	57	1	26	1	57	3	8.25

0.577 ms   Time Slot

R   Guard time for burst transient response (Ramp time)
TCH   Traffic channel
C   Control bit
SW   Synchronization word
G   Guard bits
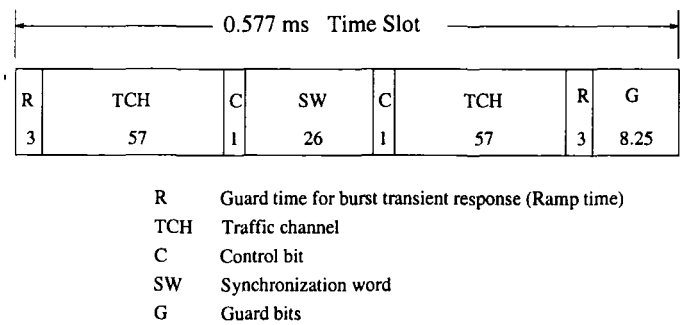
Figure 1.2.   Time slot format for GSM. Units are in bits.

system (now "Global System for Mobile Communications") was developed to operate in a new frequency allocation, and made improved quality, Pan-European roaming, and the support of data services its primary objectives.

GSM was deployed in late 1992 as the world's first digital cellular system. In its current version, GSM can support full-rate (8 slots/carrier) and half-rate (16 slots/carrier) operation, and provide various synchronous and asynchronous data services at 2.4, 4.8, and 9.6 kb/s that interface to voiceband modems (e.g., V.22bis or V.32) and ISDN. GSM uses TDMA with 200 kHz carrier spacings, eight channels per carrier with a time slot (or burst) duration of 0.577 ms, and Gaussian minimum shift keying (GMSK) with a raw bit rate of 270.8 kb/s. The time slot format of the GSM traffic channels is shown in Fig. 1.2.

Variants of GSM have also been developed to operate in higher frequency bands. In Europe, the Digital Cellular System 1800 (DCS 1800) was developed by ETSI as a standard for personal communication networks (PCNs). DCS 1800 is a derivative of the GSM system, but differs in a number of ways. First, DCS1800 operates in the 1710-1785 MHz (MS transmit) and 1805-1880 MHz (BS transmit) bands, whereas GSM operates in the 900 MHz band. Second, DCS 1800 is optimized for two classes of hand held portable units (rather than mobile units) with a peak power of 1 W and 250 mW, respectively. There are also some changes in the DCS 1800 standard to support overlays of macrocells and microcells.

GSM is deployed in North America as PCS 1900 and operates in the 1880-1990 MHz PCS bands. PCS 1900 is similar to DCS 1800, but with a few differences. One is the use of the ACELP EFR (Enhance Full Rate) vocoder that was developed for the North American market. GSM has been a phenomenal success. In late 1997, 66 million GSM subscribers were serviced by 256 network operators in 110 countries.

1.2.2   IS-54/136 AND IS-95

In North America the primary driver for second generation systems was the capacity limit felt by some AMPS operators in the largest US markets, e.g., New York, Chicago, Los Angeles. One of the key objectives established by the Cellular Telephone Industry Association (CTIA) was that any second generation cellular system must provide a 10-fold increase in capacity over AMPS. Furthermore, since AMPS was already deployed extensively throughout North America, it was desirable that any second generation cellular system be reverse compatible with AMPS. This eventually lead to the development of dual-mode transceivers.

While Europe seen the convergence to the GSM standard, North America seen the emergence of two second generation digital cellular standards, IS-54/136 and IS-95, based on time division multiple access (TDMA) and code division multiple access (CDMA) technology, respectively. The IS-54 standard was adopted in 1990, and specifies a new digital signaling scheme based on F/TDMA with 30 kHz carrier spacings and $\pi/4$ phase-shifted quadrature differential phase shift keyed ($\pi/4$-DQPSK) modulation with a raw bit rate of 48.6 kb/s [95]. IS-54 and IS-136 differ in the control channel; IS-54 uses an analog control channel, whereas IS-136 uses a digital control channel. The IS-54/136 air interface specifies 6 slots (or bursts) per frame, yielding 3 full rate channels or 6 half rate channels per carrier. The burst format for the IS-54/136 traffic channel is shown in Fig. 1.3. A straight forward deployment of IS-54/136 will offers 3 (6) times the cell capacity of AMPS for the full (half) rate systems, respectively. Additional capacity gains are also possible. IS-54/136 is now been deployed throughout North American and Indonesia.

6.67 ms Time Slot

Forward

SW	SACCH	TCH (FACCH)	CC	TCH (FACCH)	RSVD
28	12	130	12	130	12

6.67 ms Time Slot

Reverse

G	R	TCH (FACCH)	SW	TCH (FACCH)	SACCH	CC	TCH (FACCH)
6	6	16	28	122	12	12	122

G   Guard Bits
R   Guard Time for burst transient response (Ramp time)
SACCH Slow Associated Control Channel
FACCH Fast Associated Control Channel
SW   Synchronization Word
CC   Color Code
TCH   Traffic Channel
RSVD Reserved

Figure 1.3.   Burst format for IS-54/136 traffic channel. Units are in bits.

Just after the CTIA adopted IS-54 in 1990, another second generation digital cellular system was proposed by Qualcomm based on CDMA technology. In March 1992, CDMA was adopted as IS-95 [96]. With IS-95, the basic user data rate is 9.6 kb/s, which is spread by using PN sequence with a chip (clock) rate of 1.2288 Mchips/s (a processing gain of 128). The forward channel supports coherent detection by using a pilot channel (code) for channel estimation. Information on the forward link is encoded by using a rate-1/2 convolutional code, interleaved, spread by using one of 64 Walsh codes, and transmitted in 20 ms bursts. Each MS in a cell is assigned a different Walsh code, thus providing complete orthogonality under ideal channel conditions. Final spreading with a base-specific PN code of length 2^{15} is used to mitigate the multiple access interference to and from other cells. One of the major drawbacks of the IS-95 standard is that the coded downlink transmissions are not interleaved across bursts and, therefore, the signal is susceptible to fading.

CDMA systems are susceptible to the **near-far effect**, a phenomenon where MSs close into a BS will swamp out the signals from more distant MSs. For CDMA systems to work well, all signals must be received with the same power, a condition that is difficult to achieve in an erratic land mobile radio propagation environment. To combat the near-far effect, the IS-95 reverse link uses fast closed loop power control to compensate for fluctuations in received signal power due to variations in the radio link. The information on the reverse link is encoded by using a rate-1/3 convolutional code, interleaved, and mapped onto one of 64 Walsh codes. Unlike the forward channel that uses the Walsh codes for spreading, the reverse link uses the Walsh codes for 64-ary orthogonal

modulation. The BS receiver uses non-coherent detection, since no pilot signal is transmitted on the reverse link. Final spreading is achieved with a user-specific PN sequence of length $2^{42} - 1$. Both the BSs and the MSs use RAKE receivers to provide multipath diversity. A requirement of the IS-95 system is the need for **soft handoffs**, where the MS maintain can a radio link with multiple BSs in the boundary area between cells.

Ever since the introduction of IS-95, there has been a continued debate over the relative capacity of IS-54/136 and IS-95. Initial capacity claims for IS-95 were 40 times AMPS. However, current estimates are more conservative and experience from commercial deployments show a capacity that is 6 to 10 times AMPS.

1.2.3   PDC

In 1991, the Japanese Ministry of Posts and Telecommunications standardized Personal Digital Cellular (PDC). The air interface of PDC is similar in some respects to IS-54/136. PDC uses TDMA with 3 full rate (6 half rate) channels per carrier, 25 kHz carrier spacings, and $\pi/4$-DQPSK modulation with a raw bit rate of 42 kb/s. The burst format for the PDC traffic channels is shown in Fig. 1.4. Notice that the synchronization word is placed near the center of the PDC burst, whereas it is placed near the beginning of the IS-54/136 burst as shown in Fig. 1.3. This feature better enables the PDC receiver to track channel variations over the time slot. Another key feature of PDC standard is the inclusion MS antenna diversity. Like IS-54/136, PDC suffers from degraded performance under conditions of low delay spread due to the loss of multipath diversity. However, antenna diversity in the PDC MS receiver maintains spatial diversity under these conditions. More details on the PDC system can be found in the complete standard [280].

1.3   CORDLESS TELEPHONE SYSTEMS

Cordless telephones find several applications including domestic telephones, telepoint (cordless phone booth), wireless PABX (private access business exchange), and wireless local loops or radio drops. Similar to cellular telephones, first generation cordless telephones were based on analog FM technology. Since their introduction, cordless telephones gained high popularity. However, first generation cordless telephones have become victims of their own success; the voice quality was/is unacceptable in high-density subscriber areas. This lead to the development of second generation digital cordless telephones. In Europe two digital cordless telephone standards have been developed, CT2 and Digital European Cordless Telephone (DECT) [325]. In Canada a modification of CT2, called CT2+, has been developed, that offers two-way calling, roaming, and enhanced data service capabilities. In Japan, the Personal Handyphone

Forward	R 4	P 2	TCH (FACCH) 112	SW 20	CC 8	SF 1	SACCH (RCH) 15	TCH (FACCH) 112	G 6

6.67 ms Time Slot

Reverse	R 4	P 2	TCH (FACCH) 112	SW 20	CC 8	SF 1	SACCH (RCH) 21	TCH (FACCH) 112

6.67 ms Time Slot

R   Guard Time for burst transient response (Ramp time)
P   Preamble
TCH   Traffic Channel
FACCH   Fast Associated Control Channel
SW   Synchronization Word
CC   Color Code
SF   Steal Flag
SACCH   Slow Associated Control Channel
RCH   Housekeeping Bits
G   Guard Bits

Figure 1.4.   Time slot format for Japanese PDC. Units are in bits.

System (PHS) has been developed. The air interface parameters of various cordless phone standards are summarized in Tab. 1.4

1.4   THIRD GENERATION CELLULAR SYSTEMS

In March 1992, WARC approved a worldwide spectral allocation in support of IMT-2000 (International Mobile Telephone by the Year 2000) in the 1885-2200 MHz band. The IMT-2000 standard has been developed by the International Telecommunications Union Radio Communications (ITU-R) and Telecommunications (ITU-T) sectors. Various standards bodies around the world have provided inputs to the IMT-2000 standard definition. The vision of IMT-2000 is to provide ubiquitous wireless network that can support voice, multimedia and high-speed data communication. One of the main attributes of IMT-2000 is the introduction of wireless wide-band packet-switched data services for wireless access to Internet with speeds up to 2 Mb/s. The key principles of IMT-2000 are:

- *Terminal and personal mobility* with universal access and worldwide roaming through the use of portable terminals. Personal mobility will be facilitated through the use of personalized telephone numbers. The success of using a wireline telephone depends upon the knowledge of where a called party is located. As a result, 80% of the calls never reach the intended party. With personal communication services (PCS), intelligent networks (INs) will be employed to assume the burden of locating a called party, leaving the subscribers free to roam anywhere in the world.

Feature	GSM/DCS1800/PCS1900	IS-54/136
Frequency Band RL/FL (MHz)[a]	GSM: 890-915/ 935-960 DCS1800: 1710-1785/ 1805-1880 PCS1900: 1930-1990/ 1850-1910	824-829/ 869/894 1930-1990/ 1850-1910
Multiple Access	F/TDMA	F/TDMA
Carrier Spacing (kHz)	200	30
Modulation	GMSK	$\pi/4$-DQPSK
Baud Rate (kb/s)	270.833	48.6
Frame Size (ms)	4.615	40
Slots/Frame	8/16	3/6
Voice Coding (kb/s)	VSELP(HR 6.5) RPE-LTP (FR 13) ACELP (EFR 12.2)	VSELP (FR 7.95) ACELP (EFR 7.4) ACELP (12.2)
Channel Coding	Rate-1/2 CC	rate-1/2 CC
Frequency Hopping	yes	no
Handoff	hard	hard

[a] RL = reverse link, FL = forward link

Table 1.2. Second generation digital cellular standards

- *Expanded range of services* as implemented in two phases. Phase 1 will support circuit and packet switched multimedia with asymmetric user data rates up to 2 Mb/s. Phase 2 will provide user data rates up to 20 Mb/s. These capabilities will enable applications such as web browsing, file transfer, e-mail, and traveler information services, and multimedia services such as video conferencing. The minimum requirements on user data rate for both circuit and packet switched data in four different environments is as follows:

  Vehicular: 144 kb/s
  Pedestrian: 384 kb/s
  Indoor office: 2 Mb/s
  Satellite: 9.6 kb/s

- *Supplementary services* such as call waiting, caller ID, store and forward, etc.. Call management will become a necessity to ensure that PCS does not

Feature	PDC	IS-95
Frequency Band RL/FL (MHz)[a]	810-826/ 940-956 1429-1453/ 1477-1501	824-829/ 869-894 1930-1990/ 1850-1910
Multiple Access	F/TDMA	F/CDMA
Carrier Spacing (kHz)	25	1250
Modulation	$\pi/4$-DQPSK	QPSK
Baud Rate (kb/s)	42	1228.8 Mchips/s
Frame Size (ms)	20	20
Slots/Frame	3/6	1
Voice Coding (kb/s)	PSI-CELP (HR 3.45) VSELP (FR 6.7)	QCELP (8,4,2,1) RCELP (EVRC)
Channel Coding	rate-1/2 BCH	FL: rate-1/2 CC RL: rate-1/3 CC
Frequency Hopping	no	N/A
Handoff	hard	soft

[a] RL = reverse link, FL = forward link

Table 1.3. Second generation digital cellular standards

become a nuisance. That is, the subscribers must be able to control their availability for receiving calls.

- *Unified, seamless, infrastructure* that will unify diverse infrastructures such as paging, cellular, and satellite networks. In particular, the use of a common band for terrestrial and satellite networks.

- *Integration of mobile and wire-line networks* in attempt to achieve the strict QoS controls wire-line networks. Toll line voice quality is one example.

- *Service transparency* to provide the same services everywhere but with different data rates. International roaming is also desirable with a virtual home environment.

- *Spectral efficiency*, quality, flexibility, and overall cost improvement as a result of the utilization of advanced technologies.

The migration to third generation wireless systems presents some difficult challenges for wireless service providers including the following:

Feature	CT2	CT2+	DECT	PHS
Frequency Band (MHz)	864-868	944-948	1880-1900	1895-1918
Multiple Access	FDMA	F/TDMA	F/TDMA	F/TDMA
Duplexing	TDD	TDD	TDD	TDD
Carrier Spacing (kHz)	100	100	1728	300
Modulation	GFSK	GSFK	GFSK	π/4-DQPSK
Number of Carriers	40	40	10	77
Channels/Carrier	1	1	12	4
Bit Rate (kb/s)	72	72	1152	384
Speech Coding	ADPCM 32 kb/s	ADPCM 32 kb/s	ADPCM 32 kb/s	ADPCM 32 kb/s
Frame Size (ms)	2	2	10	5
Mean TX Power (mW)	5	5	10	10
Peak TX Power (mW)	10	10	250	80

Table 1.4.   Cordless telephone standards.

- *System revolution versus evolution.* A revolutionary approach provides the greatest flexibility. However, an evolutionary approach is more desirable because there are enormous infrastructure investments in legacy systems, and the maintenance of a large existing subscriber base requires a third generation system that is reverse compatible with the existing second generation systems.

- *Rapid and unpredictable growth* leads to difficulty in system planning. High spectral efficiency is essential to support large subscriber bases. Developing countries in particular are experiencing explosive growth due to the lack of a wired infrastructure.

- *Changing customer needs* requires a flexible solution.

- *Network management* for effective radio resource and mobility management with mixed services, billing, security, fraud prevention.

- *Mobile satellite systems* can make global spectral coordination very difficult.

Ten different multiple access schemes were originally proposed for IMT-2000. Two of these schemes are based on TDMA approaches, namely DECT and UWC-136. The remaining 8 proposals are based on wide-band CDMA,

Feature	UWC-136	UWC-136+	UWC-136HS vehicular GSM EDGE	UWC-136HS indoor
Multiple Access	TDMA	TDMA	TDMA	TDMA
Duplexing	FDD	FDD	FDD	FDD/TDD
Carrier Spacing (kHz)	30	30	200	1600
Modulation	π/4-DQPSK	CCH: π/4-DQPSK TXH: π/4-QPSK DTCH: 8-PSK	8-PSK GMSK	Q-O-QAM B-O-QAM
Frame Length (ms)	40	40	4.615	4.615
Slots/Frame	6	6	8	64@72μs 16@72μs
Bit Rate (kb/s)	48.6	72.9 (8-PSK) 48.6 (QPSK/ DQPSK)	812.5 (8-PSK) 270.8 (GMSK)	5200 (Q-O-QAM) 2600 (B-O-QAM) 2600 (B-O-QAM)

Table 1.5. Parameters of UWC-136.

referring to a CDMA system having a bandwidth of 5 MHz or more. The UWC-136 proposal is the 3G evolution of the IS-136 family of standards. Some parameters of the UWC-136 proposal are shown in Tab. 1.5. UWC-136 meets IMT-2000 requirements by using enhanced modulation techniques (IS-136+) and using a wider band 200 kHz carrier (UWC-136HS) for services that are not possible on the 30 kHz carrier. The UWC-136HS proposal is the same as EDGE (Enhanced Data for Global Evolution) which is an enhanced GSM air interface. EDGE is a system that is the convergence of the GSM and IS-136 family of standards.

Table 1.6 summarizes the parameters for the two remaining wide-band CDMA proposals to IMT-2000, namely W-CDMA and cdma2000. The common attributes of wide-band CDMA systems include the following:

- provision of multirate services
- packet data services

Feature	W-CDMA	cdma2000
Multiple Access	FDD: DS/CDMA TDD: T/CDMA	FDD: DS-CDMA TDD: T/CDMA
Chip Rate (Mcps)	FDD: 1.024/4.096 8.192/16.384 TDD: 4.096	1.2288/3.6864 7.3728/11.0593/ 14.7456
Carrier Spacing (MHz)	(1.25),5,10,20	1.25,5,10,15,20
Frame Length (ms)	10	20
Modulation	FDD: FL: QPSK RL: dual-channel QPSK TDD: FL&RL: QPSK	FL: QPSK RL: BPSK
Coding	rate-1/2, 1/3 $K = 9$ CC optional RS outer code	rate-1/2, 1/3, 1/4 $K = 9$ CC rate-1/2, 1/3, 1/4, $K = 4$ TC
Interleaving	inter/intraframe	intraframe
Spreading	FDD: FL: BPSK RL: QPSK TDD: FL,RL: QPSK	QPSK
Inter BS synchronization	asynchronous	synchronous

Table 1.6.   Parameters for W-CDMA and cdma2000.

- complex spreading
- a coherent uplink using a user dedicated pilot
- additional pilot channel in the downlink for beam forming
- seamless interfrequency handoff
- fast forward link power control
- optional multiuser detection

The major differences between the different system proposals center around the chip rate that is used, and synchronous (cdma2000) vs. asynchronous (W-CDMA) network operation.

Global spectral co-ordination is essential for the IMT-2000 concept. The spectral allocations of the regulatory agencies in major world markets are shown in Fig. 1.5. In the United States, the 1885-2200 MHz band allocated for IMT-2000 overlaps significantly with the 1850-1990 MHz band used to support PCS services, as shown in Fig. 1.6. Blocks A and B correspond to
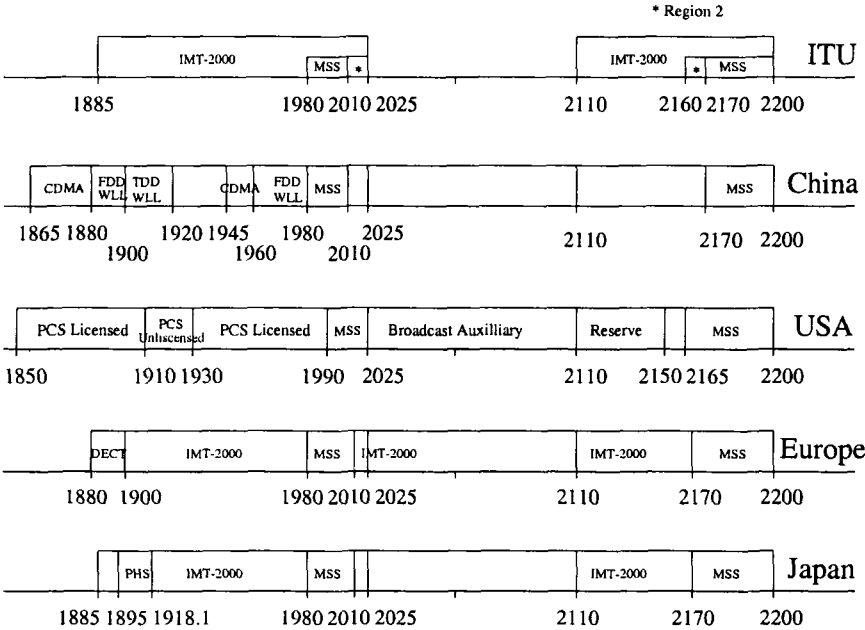
Figure 1.5. International frequency allocations.
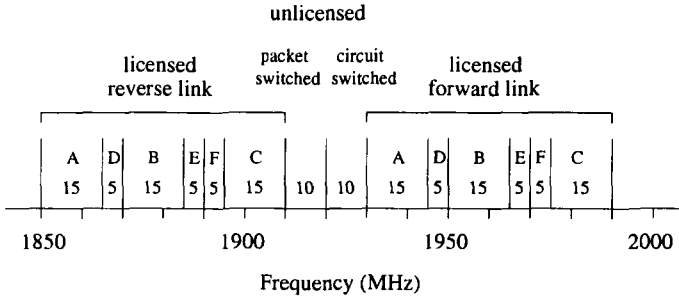
Figure 1.6. North American PCS frequency allocations.

major trading areas (MTAs) while blocks C through F correspond to basic trading areas (BTAs). There are 51 MTAs and 492 BTAs in the United States. In addition, 20 MHz of spectrum was reserved for unlicensed use according to FCC Part 15 rules. Of this 20 MHz, 10 MHz is for packet switched applications while 10 MHz is for circuit switched applications.

1.5   WIRELESS LANS AND AND PANS

A variety of wireless local area network (WLAN) and personal area network (WPAN) systems have been developed to operate in unlicensed bands. Tab. 1.7

Location	Frequency Range (GHz)	Maximum Output Power (mW or dBm)
North America	2.400-2.4835	1000 mW
Europe	2.400-2.4835	100 mW EIRP
Japan	2.471-2.497	10 mW
United States (UNII lower band)	5.150-5.250	minimum of 50 mW or 4 dBm $+ 10\log_{10}B$
United States (UNII middle band)	5.250-5.350	minimum of 250 mW or 11 dBm $+ 10\log_{10}B$
United States (UNII upper band)	5.725-5.825	minimum of 1000 mW or 17 dBm $+ 10\log_{10}B$

Table 1.7.   2.4 and 5 GHz bands for license exempt use. $B = $ -26 dB emission bandwidth in MHz.

lists the unlicensed bands that are used in various parts of the world. Until a few years ago, most of the WLAN systems that operated in unlicensed bands were based upon proprietary air interfaces and MAC protocols, without an open standard.

In 1997, the IEEE 802.11 standardization group established the first WLAN standard based to provide 1 and 2 Mb/s aggregate rates. IEEE 802.11 uses direct sequence spread spectrum modulation, an 11-bit Barker sequence for spreading, and either BPSK (1 Mb/s) or QPSK (2 Mb/s). Barker sequences are discussed in further detail in Chapter 8. In 1998, the IEEE 802.11b working group defined an enhanced air interface to provide 5.5 and 11 Mb/s aggregate data rates. The IEEE 802.11b air interface uses complementary code keying (CCK), which is described in further detail in Chapter 8. In 1998, IEEE 802.11a adopted orthogonal frequency division (OFDM) as the basis for their new 5 GHz standard, targeting a range of data rates ranging from 6 to 54 Mb/s. The principles of OFDM are discussed in Chapters 4 and 5. Following IEEE 802.11a, High-Performance LAN (HiperLAN/2) (Europe) and Multimedia Mobile Access Communication (MMAC) (Japan) adopted OFDM in their physical layer specifications. The parameters of the IEEE 802.11a OFDM standard are summarized in Tab. 1.8.

In 1999, the IEEE802.15 Working Group was created to develop a Wireless Personal Area Network (WPAN) standard. The Bluetooth specification has been proposed as one such WPAN standard [153]. Bluetooth is an ad hoc network that is based on Frequency Hop CDMA (FH-CDMA) and Gaussian frequency shift keying (GFSK) with a modulation index of 0.3. Bluetooth uses

Data Rate	6, 9, 12, 18, 24, 36, 48, 54 Mb/s
Modulation	BPSK, QPSK, 16-QAM, 64-QAM
Coding	1/2, 2/3, 3/4 CC
Number of subcarriers	52
Number of pilots	4
OFDM symbol duration	4 μs
Guard interval	800 ns
Subcarrier spacing	312.5 kHz
3 dB bandwidth	16.56 MHz
Channel spacing	20 MHz

Table 1.8.   Key parameters of the IEEE 802.11a OFDM standard, from [333].

a set of 79 hop carriers with a spacing of 1 MHz and a hop dwell time of 625 μs. A single FH channel supports a data rate of 1 Mb/s. Bluetooth uses either a very simple rate-1/3 3-bit repetition code or a simple rate-2/3 shortened Hamming code.

2.   FREQUENCY REUSE AND THE CELLULAR CONCEPT

A cellular telephone system has two basic functions; it must locate and track both active and inactive mobile stations (MSs), and it must always attempt to connect the active MSs to the best available base station(s) (BS(s)). The former task is the subject of user location updating and paging. The latter task requires the continuous evaluation of the radio link quality with the serving BS(s), and the radio link quality with alternate BSs. This monitoring is performed by a computer system that uses knowledge of the link quality evaluations, in addition to the system topology and traffic flow, to decide upon the best BS(s) to serve a particular MS.

A cellular telephone system uses low power (less than 1 watt) radio communication between a MS and a grid of BSs [213]. Movement of the MS, however, leads to highly erratic radio link quality, and careful monitoring and control are required to keep it acceptable. Evaluation of radio link quality is based upon a large number of criteria, but at the core is a statistical measurement process based on *a priori* knowledge of the expected radio channel characteristics. The time required to measure the radio link quality and the accuracy of the measurement depends on the local propagation characteristics. Time consuming

link quality measurements will limit the ability of the cellular system to react to degradations in link quality and compensate by changing the allocation of power and bandwidth resources. Conversely, if the link quality measurements can be made quickly, then the time required for the cellular system to process the link quality measurements, make decisions, and transmit desired changes to the network entities, including the MSs, will limit the adaptability of the cellular system. Limitations on the speed of measurement and control essentially determine overall link quality and the size and distribution of cells in modern cellular systems. The cell sizes, the ability radio links to withstand interference, and the ability of the cellular system to react to variations in traffic are the main factors that determine the spectral efficiency of a cellular system.

In cellular systems, the available spectrum is partitioned among the BSs, and a given frequency is reused at the closest possible distance that the radio link will allow. Smaller cells have a shorter distance between reused frequencies, and this results in an increased spectral efficiency and traffic carrying capacity. Dramatic improvement in spectral efficiency is the main reason for the interest in microcells. However, the microcellular propagation environment is highly erratic. Distributed resource allocation algorithms must be used to maintain high link quality.

The current trend is toward cellular systems that have high spectral efficiency and offer ubiquitous service coverage. These systems will require i) effective cellular architectures, ii) fast and accurate link quality measurements, iii) rapid control in all types of environments, iv) installation of BSs to provide radio coverage virtually everywhere, and v) power and bandwidth efficient air interface schemes that can mitigate the harsh effects of the propagation environment and tolerate high levels of noise and interference.

Cellular mobile radio systems that use TDMA and FDMA rely upon **frequency reuse**, where users in geographically separated cells simultaneously use the same carrier frequency. The cellular layout of a conventional macrocellular system is quite often described by a uniform grid of hexagonal cells or radio coverage zones. In practice the cells are not regular hexagons, but instead are distorted and overlapping areas. The hexagon is an ideal choice for representing macrocellular coverage areas, because it closely approximates a circle and offers a wide range of tesellating reuse cluster sizes. A tesellating reuse cluster of size N can be constructed if [258]

$$N = i^2 + ij + j^2 \qquad (1.1)$$

where i and j are non-negative integers, and $i \geq j$. It follows that the allowable cluster sizes are $N = 1, 3, 4, 7, 9, 12, \ldots$ Examples of 3-, 4-, and 7-cell reuse clusters are shown in Fig. 1.7. The reuse clusters are tesellated to form a frequency plan. A simplified 7-cell frequency reuse plan is shown in Fig. 1.8, where similarly marked cells use identical sets of carrier frequencies.
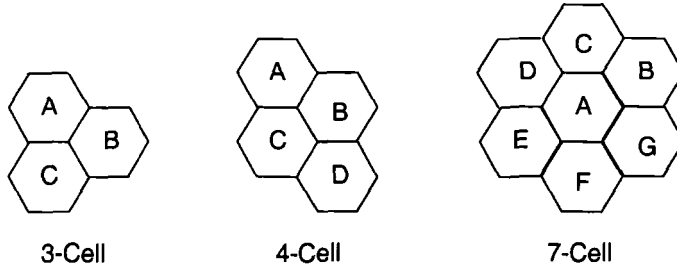
Figure 1.7. Commonly used cellular reuse clusters.
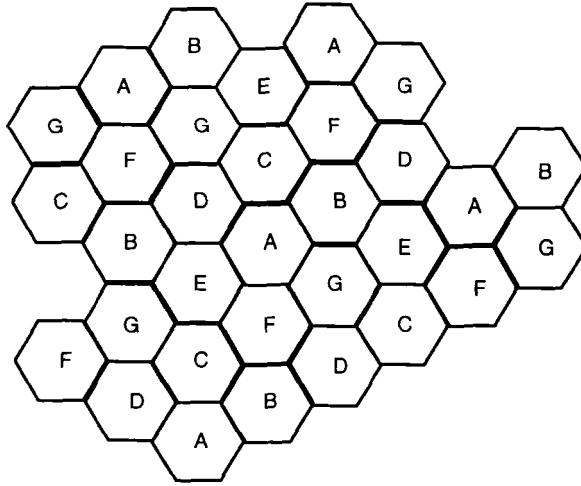
Figure 1.8. Macrocellular deployment using 7-cell reuse pattern.

The **co-channel reuse factor** D/R, is defined as the ratio of the co-channel reuse distance D between cells using the same set of carrier frequencies and the radius of the cells R[1]. For hexagonal cells, the reuse cluster size N and the co-channel reuse factor D/R are related by (see Problem 1.2)

$$D/R = \sqrt{3N} . \tag{1.2}$$

For microcellular systems with lower BS antenna heights, regular hexagons are no longer appropriate for approximating the radio coverage zones. Typical microcell BSs use an antenna height of about 15 m, well below the skyline of any buildings that might be present, and acceptable link quality can be obtained anywhere within 200-500 m of the BS. For microcells, the choice of cell shape

---

[1] For hexagonal cells, R is the distance from the center to the corner of a cell.

Figure 1.9.   Microcellular deployment along a highway with a 3-cell reuse pattern.
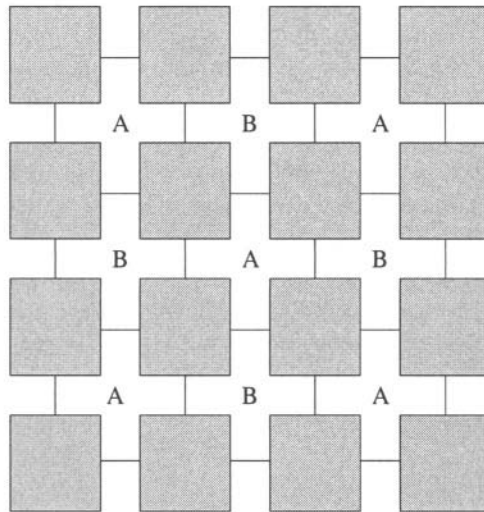
Figure 1.10.   Microcellular deployment in an urban canyon. Base stations are deployed at every intersection in a dense urban area with a 2-cell reuse pattern.

depends greatly upon the particular deployment. For example, the linear cells shown in Fig. 1.9 may provide a more accurate model **of highway microcells** that are deployed along a highway with directional antennas. In an area with urban canyons, the buildings act as wave guides to channel the signal energy along the street corridors. Fig. 1.10 shows a typical **Manhattan microcell** deployment that is often used to model microcells that are deployed in city centers.

3.   MOBILE RADIO PROPAGATION ENVIRONMENT

Radio signals generally propagate according to three mechanisms; **reflection**, **diffraction**, **and scattering**. Reflections arise when the plane waves are incident upon a surface with dimensions that are very large compared to the wavelength. Diffraction occurs according to Huygen's principle when there is an obstruction between the transmitter and receiver antennas, and secondary waves are generated behind the obstructing body. Scattering occurs when the plane waves are incident upon an object whose dimensions are on the order of a wavelength or less, and causes the energy to be redirected in many directions.

The relative importance of these three propagation mechanisms depends on the particular propagation scenario.

As a result of the above three mechanisms, macrocellular radio propagation can be roughly characterized by three nearly independent phenomenon; **path loss** variation with distance, slow log-normal **shadowing**, and fast **multipath-fading**. Each of these phenomenon is caused by a different underlying physical principle and each must be accounted for when designing and evaluating the performance of a cellular system. Multipath-fading results in rapid variations in the envelope of the received signal and is caused when plane waves arrive from many different directions with random phases and combine vectorially at the receiver antenna. Typically, the received envelope can vary by as much as 30 to 40 dB over a fraction of a wavelength due to constructive and destructive addition. Multipath also causes time dispersion, because the multiple replicas of the transmitted signal propagate over different transmission paths and reach the receiver antenna with different time delays. Time dispersion may require equalization in TDMA systems and RAKE reception in CDMA systems.

It is well known that the intensity of an electromagnetic wave in free space decays with the square of the radio path length, d, such that the received power at distance d is

$$\Omega_p(d) = \Omega_t k \left(\frac{\lambda_c}{4\pi d} \right)^2 \qquad (1.3)$$

where Ω_t is the transmitted power, λ_c is the wavelength, and k is a constant of proportionality. Although it may seem counter-intuitive, path loss is essential in high capacity cellular systems, the reason being that a rapid attenuation of signal strength with distance permits a small co-channel reuse distance and, therefore, a high spectral efficiency. The 800-900 MHz UHF band was chosen for first generation cellular systems, partly because of its relatively short range radio propagation characteristics. Of course if a large radio coverage area is desired, as is the case with low capacity emergency and dispatch communication systems (police, fire, etc..), then a small path loss is preferred. For this reason the VHF band is preferred for these applications which results in a smaller attenuation with distance.

Free space propagation does not apply in a mobile radio environment and the propagation path loss depends not only on the distance and wavelength, but also on the antenna heights of the MSs and the BSs, and the local terrain characteristics such as buildings and hills (in macrocells). The site specific nature of radio propagation makes the theoretical prediction of path loss difficult and there are no easy solutions. The simplest path loss model assumes that the received power is

$$\Omega_{p \text{ (dBm)}}(d) = \mu_{\Omega_{p \text{ (dBm)}}}(d_o) - 10\beta \log_{10}(d/d_o) + \epsilon_{\text{(dB)}} \text{ (dBm)} \qquad (1.4)$$

where $\mu_{\Omega_p\text{ (dBm)}}(d_o) = E[\Omega_p\text{ (dBm)}(d_o)]$ is the average received signal power (in dBm) at a known reference distance that is in the far field of the transmitting antenna. Typically, d_o is 1 km for macrocells, 100 m for outdoor microcells, and 1 m for indoor picocells. The value of $\mu_{\Omega_p\text{ (dBm)}}(d_o)$ will depend on the frequency, antenna heights and gains, and other factors. The parameter β is called the **path loss exponent** and is a key parameter that affects the spectral efficiency of a cellular system. This parameter is strongly dependent on the cell size and local terrain characteristics. The path loss exponent ranges from 3 to 4 for a typical urban macrocellular environment, and from 2 to 8 for a microcellular environment. Usually, the path loss exponents are determined by empirical measurements.

The parameter $\epsilon_{\text{(dB)}}$ in (1.4) is a zero-mean Gaussian random variable (in dB) that represents the error between the actual and estimated path loss. This statistical variation in $\Omega_p\text{ (dBm)}(d)$ is caused by shadowing. Shadows are generally modeled as being log-normally distributed, meaning that the probability density function of $\Omega_{\text{(dBm)}}(d)$ is

$$p_{\Omega_p\text{ (dBm)}(d)}(x) = \frac{1}{\sqrt{2\pi}\sigma_\Omega} \exp\left\{ -\frac{(x - \mu_{\Omega_p\text{ (dBm)}}(d))^2}{2\sigma_\Omega^2} \right\} \qquad (1.5)$$

where

$$\mu_{\Omega_p\text{ (dBm)}}(d) = \mu_{\Omega_p\text{ (dBm)}}(d_o) - 10\beta \log_{10}(d/d_o) \text{ (dBm)} . \qquad (1.6)$$

The parameter σ_Ω is the **shadow standard deviation**. A more accurate path loss model results in a smaller σ_Ω. For macrocells, σ_Ω typically ranges from 5 to 12 dB, with $\sigma_\Omega = 8$ dB being a typical value. Furthermore, σ_Ω has been observed to be nearly independent of the radio path length d. The received signal power in the absence of shadowing as defined by (1.6) is called the **area mean**, while the received signal power in the presence of shadowing as defined by (1.4) is called the **local mean**. Fig. 1.11 illustrates the above concepts by plotting the received signal strength as a function of the radio path length for both free space and a typical urban macrocellular environment.

4.   CO-CHANNEL INTERFERENCE AND NOISE

Frequency reuse in FDMA/TDMA cellular systems introduces **co-channel interference**, one of the major factors that limits the capacity of cellular systems. Co-channel interference arises when the same carrier frequency is used in different cells. In this case, the power density spectra of the desired and interfering signals completely overlap. Frequency reuse also introduces **adjacent channel interference**. This type of interference arises when neighboring cells use carrier frequencies that are spectrally adjacent to each other. In this
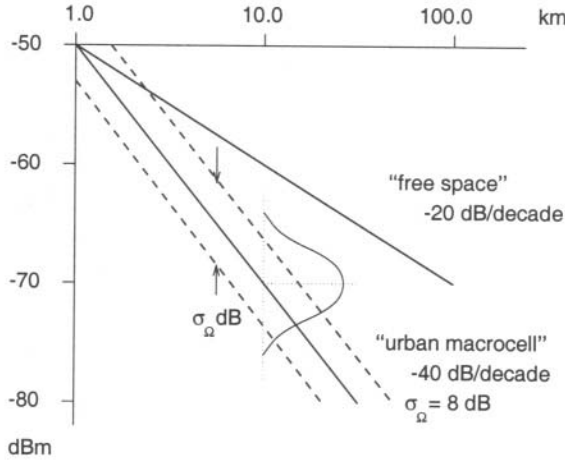
Figure 1.11. Path loss in free space and typical urban macrocellular environments; $\beta = 4$, $\sigma_\Omega = 8$ dB. The received signal strength in dBm at a distance of 10 km is Gaussian distributed with a mean of -70 dBm and a variance of σ_Ω^2 dB.

case the power density spectrum of the desired and interfering signals partially overlap.

Wireless radio links quite often exhibit a **threshold effect**, where the link quality is acceptable when both the carrier-to-noise ratio Γ and the carrier-to-interference ratio Λ exceed certain thresholds, denoted by Γ_{th} and Λ_{th}, respectively [115].[2] Otherwise, the link quality is unacceptable and an **outage** is said to occur. The thresholds Γ_{th} and Λ_{th} depend on many parameters of the radio link, including the particular modulation and coding scheme that is employed, the receiver structure, the measure of link quality, the propagation environment, the MS velocity, and other factors. Once the air interface is specified, the propagation environment determines whether or not an outage occurs. For fast moving MSs, path loss and shadowing determine the link quality once Γ_{th} and Λ_{th} have been specified. Conversely, for slow moving MSs, the link quality may also become unacceptable when the received signal envelope exhibits a deep fade due to multipath fading.

Here we introduce two types of outages. The first is the **thermal noise outage**, defined as

$$O_N = P(\Gamma < \Gamma_{th}) \tag{1.7}$$

and the second is the **co-channel interference outage**, defined as

$$O_I = P(\Lambda < \Lambda_{th}) \ . \tag{1.8}$$

---

[2]For the time being, the effect of adjacent channel interference will be neglected.

The overall outage due to both thermal noise and co-channel interference is

$$O = \mathrm{P}(\Gamma < \Gamma_{\mathrm{th}} \text{ or } \Lambda < \Lambda_{\mathrm{th}}) \ . \tag{1.9}$$

For lightly loaded cellular systems, thermal noise will dominate the performance. However, for heavily loaded cellular systems, thermal noise can be neglected in difference to the typically dominant effect of the co-channel interference.

5. RECEIVER SENSITIVITY AND LINK BUDGET

Receiver sensitivity refers to the ability of the receiver to detect radio signals in the presence of noise. This noise can arise from a variety of sources that are external to the system, such as atmospheric noise like lightning strikes, galactic noise, man made noise like automobile ignition noise, and thermal noise that is internal to the system.

The ratio of the desired carrier power to thermal noise power before detection is commonly called the carrier-to-noise ratio, Γ. The parameter Γ is a function of the communication link parameters, such as the transmitted power (or effective isotropic radiated power (EIRP)), path loss, receiver antenna gain, and the effective input-noise temperature of the receiving system. The formula that relates Γ to the link parameters is called the **link budget**. The link budget can be expressed in terms of the following parameters:

$$
\begin{aligned}
\Omega_t &= \text{transmitted carrier power} \\
G_{\mathrm{T}} &= \text{transmitter antenna gain} \\
L_p &= \text{path loss} \\
G_{\mathrm{R}} &= \text{receiver antenna gain} \\
\Omega_p &= \text{received signal power} \\
E_c &= \text{received energy per modulated symbol} \\
T_o &= \text{receiving system noise temperature in degrees Kelvin} \\
B_w &= \text{receiver noise bandwidth} \\
N_o &= \text{white noise power spectral density} \\
R_c &= \text{modulated symbol rate} \\
k &= 1.38 \times 10^{-23}\, Ws/K \ = \ \text{Boltzmann's constant} \\
F &= \text{noise figure, typically 5 to 6 dB} \\
L_{\mathrm{Rx}} &= \text{receiver implementation losses} \\
L_{\mathrm{I}} &= \text{losses due to system load (interference)} \\
M_{\mathrm{shad}} &= \text{shadow margin} \\
G_{\mathrm{HO}} &= \text{handoff gain} \\
S_{\mathrm{RX}} &= \text{receiver sensitivity}
\end{aligned}
$$

The effective received carrier power is

$$\Omega_p = \frac{\Omega_t G_T G_R}{L_{RX} L_p} \quad . \tag{1.10}$$

The total input noise power to the receiver is [116]

$$N = kT_o B_w F \quad . \tag{1.11}$$

The value of kT_o at room temperature of 17 °C (290 °K) is $kT_o = -174$ dBm/Hz. The received carrier-to-noise ratio defines the link budget

$$\Gamma = \frac{\Omega_p}{N} = \frac{\Omega_t G_T G_R}{kT_o B_w F L_{RX} L_p} \quad . \tag{1.12}$$

The carrier-to-noise ratio, Γ, and modulated symbol energy-to-noise ratio, E_c/N_o, are related as follows [116]

$$\frac{E_c}{N_o} = \Gamma \times \frac{B_w}{R_c} \quad . \tag{1.13}$$

Hence, we can rewrite the link budget as

$$\frac{E_c}{N_o} = \frac{\Omega_t G_T G_R}{kT_o R_c F L_{RX} L_p} \quad . \tag{1.14}$$

Converting to decibel units gives

$$E_c/N_{o(dB)} = \Omega_{t\,(dBm)} + G_{T\,(dB)} + G_{R\,(dB)} \tag{1.15}$$
$$-kT_{o(dBm)/Hz} - R_{c\,dBHz)} - F_{(dB)} - L_{RX\,(dB)} - L_{p\,(dB)} \quad .$$

The **receiver sensitivity** is defined as

$$S_{RX} = L_{RX} kT_o F(E_c/N_o) R_c \tag{1.16}$$

or converting to decibel units

$$S_{RX\,(dBm)} = L_{RX\,(dB)} + kT_{o(dBm)/Hz} + F_{(dB)} + E_c/N_{o(dB)} + R_{c\,(dBHz)} \quad . \tag{1.17}$$

In (1.17), all parameters are usually fixed except for E_c/N_o. To determined the receiver sensitivity we first find the minimum $E_c/N_{o(dB)}$ that will yield an acceptable link quality, and then substitute this value into (1.17). Then by substituting the resulting value for $S_{RX\,(dBm)}$ into (1.15) and solving for $L_{p\,(dB)}$ we obtain the **maximum allowable path loss**

$$L_{max\,(dB)} = \Omega_{t\,(dBm)} + G_{T\,(dB)} + G_{R\,(dB)} - S_{RX\,(dBm)} \quad . \tag{1.18}$$

Because we are interested in the link budget for cellular radio systems, there are three other very important link budget parameters; (i) the margin for system loading or **interference loading**, (ii) the **shadow margin**, and (iii) the **handoff gain**. The first two quantities will reduce the maximum allowable path loss, while the third increases it. There may be other factors, but they will apply equally to the various systems under consideration, so they are irrelevant when making *relative* comparisons between systems. However, they are important if we want to determine the *absolute* allowable path loss.

Interference Loading:.

System loading causes co-channel and adjacent channel interference. Hence, the cell radii in any cellular system will shrink and expand as the traffic load increases and decreases, respectively. This phenomenon is sometimes called **cell breathing**. If we wish to compare the relative coverage of different cellular systems as the subscriber load increases, then we must account for the increased traffic load by including an interference degradation margin in the link budget; otherwise, there will be very poor coverage near the planned cell boundaries. If the co-channel and adjacent channel interference is treated as white noise to a first approximation, then the effect is to increase the total input noise power to the receiver by a multiplicative factor of L_I. To account for the system loading interference degradation, we reduce the maximum allowable path loss in (1.18) by an amount equal to $L_{I\ (dB)}$, the interference margin. The required $L_{I\ (dB)}$ depends on the type of cellular system under consideration and the cell loading. CDMA systems typically require a higher interference margin than TDMA systems, because the signals of all users occupy the same bandwidth.

Shadow Margin and Handoff Gain:.

Suppose that a noise outage occurs whenever the received carrier-to-noise ratio $\Gamma = \Omega_{p\ (dB)(d)}/N < \Gamma_{th}$ or, equivalently, $\Omega_{p\ (dBm)}(d) < \Omega_{th\ (dBm)}$. The **edge noise outage probability** is defined as the probability that $\Omega_{p\ (dBm)}(R) < \Omega_{th\ (dBm)}$, where $d = R$ for a MS located on the cell edge. The **area noise outage probability** is defined as the probability that $\Omega_{p\ (dBm)}(d) < \Omega_{th\ (dBm)}$ when averaged over the entire cell area. To ensure a given edge or area outage probability we must introduce a **shadow margin**, M_{shad}, into the link budget.

The edge noise outage probability is

$$
\begin{aligned}
O_N(R) &= \mathrm{P}(\Omega_{p\ (dBm)}(R) < \Omega_{th\ (dBm)}) \\
&= \int_{-\infty}^{\Omega_{th(dBm)}} \frac{1}{\sqrt{2\pi}\sigma_\Omega} \exp\left\{ -\frac{(x - \mu_{\Omega_{p\ (dBm)}}(R))^2}{2\sigma_\Omega^2} \right\} dx
\end{aligned}
$$
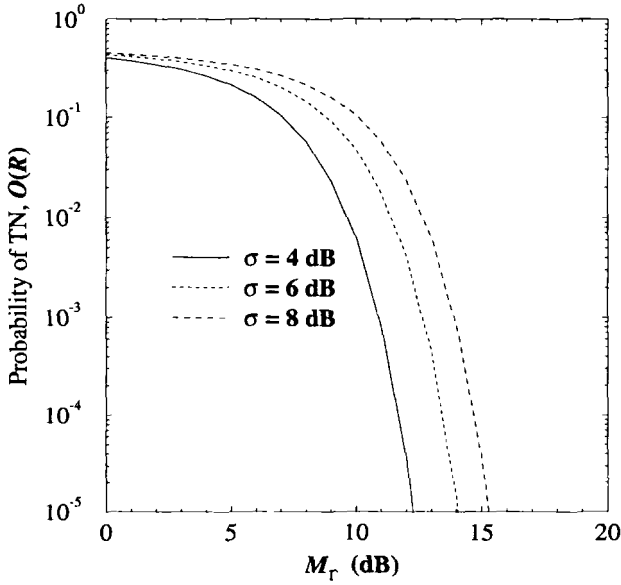
Figure 1.12. Edge noise outage probability against the shadow margin.

$$= Q\left(\frac{M_{\text{shad}}}{\sigma_\Omega}\right) \quad . \tag{1.19}$$

where

$$Q(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \quad . \tag{1.20}$$

and

$$M_{\text{shad}} = \mu_{\Omega_p \text{ (dBm)}}(R) - \Omega_{\text{th (dBm)}} \quad . \tag{1.21}$$

is the shadow margin. The edge noise outage probability, $O_N(R)$ is plotted against M_{shad} in Fig. 1.12 for various shadow standard deviations.

Example 1.1

Suppose that we wish to have $O_N(R) = 0.1$. To determine the required shadow margin, we choose M_{shad} so that the shaded area under the Gaussian density function in Fig. 1.13 is equal to 0.1. Hence, we solve

$$0.1 = Q\left(\frac{M_{\text{shad}}}{\sigma_\Omega}\right) \quad . \tag{1.22}$$

We have

$$\frac{M_{\text{shad}}}{\sigma_\Omega} = Q^{-1}(0.1) = 1.28 \tag{1.23}$$

For $\sigma_\Omega = 8$ dB, the required shadow margin is

$$M_{\text{shad}} = 1.28 \times 8 = 10.24 \text{ dB} \quad . \tag{1.24}$$
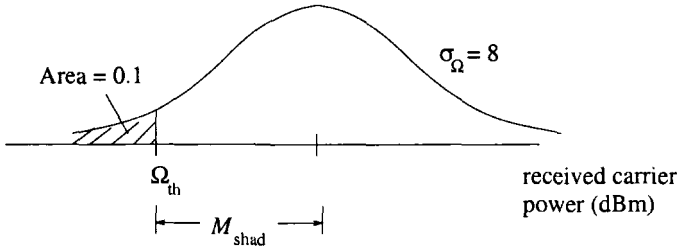
Figure 1.13.   Determining the required shadow margin.

To obtain a relationship between the edge and area noise outage probabilities, we need models for the propagation path loss and spatial distribution of MSs. For macrocells it is reasonable to assume that the MSs are uniformly distributed throughout the cell area. This assumption along with the path loss model in (1.6) yields an area noise outage probability [115]

$$
\begin{aligned}
O_N &= \frac{1}{\pi R^2} \int_0^R O(r) 2\pi r \, dr \\
&= Q(X) - \exp\left\{XY + Y^2/2\right\} Q(X+Y) \qquad (1.25)
\end{aligned}
$$

where

$$
X = \frac{M_{\text{shad}}}{\sigma_\Omega}, \qquad Y = \frac{2\sigma_\Omega}{\beta\xi} \qquad (1.26)
$$

where $\xi = 10/\ln 10$. The first term of this expression is equal to the edge noise outage probability, $O_N(R)$, while the second term is a correction factor.

The above argument applies to the case of a single isolated cell. For cellular systems where the geographical area is covered by multiple cells, the situation is more complex. As a MS moves from one cell to the next handoffs will be executed to maintain call continuity. Consider a MS that is located in the boundary area between two cells. Although the link to the serving BS may be shadowed and experience and outage, the link to an alternate BS may provide acceptable quality. Hence, at the boundary area between two cells, we obtain a diversity effect called **macrodiversity**. Handoffs take advantage of macrodiversity, and increase the maximum allowable path loss over the single cell case by an amount equal to the **handoff gain**, G_{HO}. There are a variety of handoff algorithms that are used in cellular systems. CDMA cellular systems such as IS-95 use **soft handoff**, while TDMA cellular systems such as GSM and DAMPS typically use **hard handoff**.

To illustrate the principle of handoff gain, consider a cluster of 7 cells; the target cell is in the center and surrounded by 6 other cells. By using

Monte Carlo simulation, we have calculated the area averaged noise outage probability for the target cell, assuming that the mobile station is uniformly distributed over the cell area. Our results assume correlated shadowing, where each of the six BSs surrounding target BS have a shadow correlation of 0.5 with the target BS, but the shadows are independent amongst the six surrounding BSs. Let $\Omega_{p,k \text{ (dBm)}}, k = 0, \ldots, 6$ denote the received signal strength for the target BS ($k = 0$) and the six surrounding BSs ($k = 1, \ldots, 6$). Three cases are considered; a single cell, soft handoffs and hard handoffs. For the single cell performance, no handoffs are used. With soft handoffs, the BS that provides the best link is always selected as the serving BS. If any BS results in a received signal power that is above the receiver threshold, $\Omega_{\text{th (dBm)}}$, then link quality is acceptable; otherwise an outage occurs.

With our hard handoff algorithm, the received signal power from the target BS is first determined. If it exceeds the receiver threshold, $\Omega_{\text{th (dBm)}}$, then the link quality is acceptable. Otherwise, the six surrounding BSs are tested for handoff candidacy. In order for a BS to be a handoff candidate, we must have $\Omega_{p,k \text{ (dBm)}} - \Omega_{p,0 \text{ (dBm)}} \geq H_{(\text{dB})}$ where $H_{(\text{dB})}$) is the handoff hysteresis. If any BS out of the handoff candidates results in a received signal power that is above the receiver threshold, $\Omega_{\text{th (dBm)}}$, then link quality is acceptable; otherwise an outage occurs.

The results are shown in Fig. 1.14, for $H_{(\text{dB})} = 6$ dB. Note that a 10% area noise outage probability (90% coverage) requires a shadow margin of 5.6 dB. With soft handoffs, the required shadow margin is 1.8 dB. The difference of 3.8 dB represents the soft handoff gain. The corresponding hard handoff gain is about 2.8 dB. Note that the soft handoff will always be greater than the hard handoff gain.

In summary, the maximum allowable path loss with the inclusion of the margins for shadowing and interference loading is

$$
\begin{aligned}
L_{\text{max (dB)}} \quad = \quad & \Omega_{t \text{ (dBm)}} + G_{T \text{ (dB)}} + G_{R \text{ (dB)}} - S_{\text{RX (dBm)}} \\
& - M_{\text{shad (dB)}} - L_{I \text{ (dB)}} + G_{\text{HO (dB)}} \ .
\end{aligned} \quad (1.27)
$$

6.   COVERAGE

Coverage refers to the number of base stations or cell sites that are required to "cover" or provide service to a given area with an acceptable grade of service. This is an important consideration when a cellular system is first deployed. Clearly the cellular system that requires the fewest number of cell sites to cover a given geographic area has an infrastructure cost advantage.

The number of cell sites that are required to cover a given area is determined by the maximum allowable path loss and the path loss characteristic. To compare the coverage of different cellular systems, we first determine the
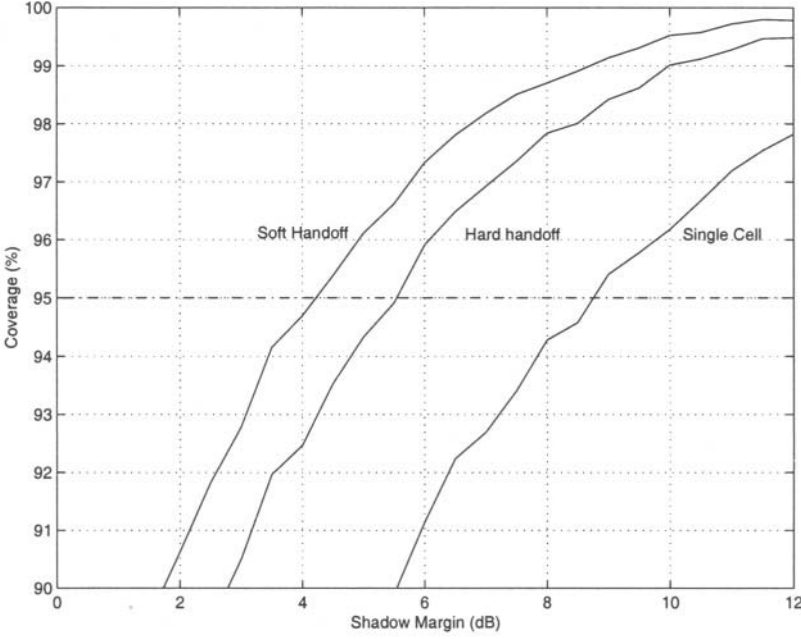
Figure 1.14. Required shadow margin with hard and soft handoffs and 95% coverage; handoff hysteresis $H_{(dB)} = 6$ dB.

maximum allowable path loss for the different systems by using a common quality criterion, i.e., the area averaged outage probability.

From (1.11), it is apparent that

$$L_{\text{max (dB)}} = C + 10\beta\log_{10}d_{\text{max}} \tag{1.28}$$

where d_{max} is the radio path length that corresponds to the maximum allowable path loss and C is some constant. The quantity d_{max} is equal to the radius of the cell. To provide good coverage it is desirable that d_{max} be as large as possible.

Once L_{max} has been determined for the various systems under consideration, the relative coverage advantages of different systems can be compared, assuming that all other factors are equal. As an example of how this is done, suppose that System 1 has $L_{\text{max (dB)}} = L_1$ and System 2 has $L_{\text{max (dB)}} = L_2$, with corresponding radio path lengths of d_1 and d_2, respectively. The difference in the maximum allowable path loss is related to the cell radii through the following relationship

$$
\begin{aligned}
L_1 - L_2 &= 10\beta\left(\log_{10}d_1 - \log_{10}d_2\right) \\
&= 10\beta\left(\log_{10}\frac{d_1}{d_2}\right) .
\end{aligned} \tag{1.29}
$$

Looking at things another way

$$\frac{d_1}{d_2} = 10^{(L_1 - L_2)/(10\beta)} \quad . \tag{1.30}$$

Since the area of a cell is equal to $A = \pi d^2$ (assuming a circular cell) the ratio of the cell areas is

$$\frac{A_1}{A_2} = \frac{d_1^2}{d_2^2} = \left(\frac{d_1}{d_2}\right)^2 \tag{1.31}$$

and, hence,

$$\frac{A_1}{A_2} = 10^{2(L_1 - L_2)/(10\beta)} \quad . \tag{1.32}$$

Suppose that A_{tot} is the total geographical area to be covered. Then the ratio of the required number of cell sites for Systems 1 and 2 is

$$\frac{N_1}{N_2} = \frac{A_{\text{tot}}/A_1}{A_{\text{tot}}/A_2} = \frac{A_2}{A_1} = 10^{-2(L_1 - L_2)/(10\beta)} \quad . \tag{1.33}$$

As an example, suppose that $\beta = 3.5$ and $L_1 - L_2 = 2$ dB. Then $N_2/N_1 = 1.30$. Hence, System 2 requires 30% more base stations to cover the same geographical area. In conclusion, a seemingly small difference in link budget translates into a large difference in infrastructure cost.

7.   SPECTRAL EFFICIENCY AND CAPACITY

Spectral efficiency is of paramount concern to cellular system operators. There are a variety of definitions for spectral efficiency, but an appropriate definition measures spectral efficiency in terms of the spatial traffic density per unit bandwidth. For a cellular system that consists of a deployment of uniform cells, the spectral efficiency can be expressed in terms of the following parameters:

$$
\begin{aligned}
G_c &= \quad \text{offered traffic per channel (Erlangs/channel)}\\
N_c &= \quad \text{number of channels per cell}\\
W_{\text{sys}} &= \quad \text{total system bandwidth (Hz)}\\
A &= \quad \text{area per cell (m}^2) \quad .
\end{aligned}
$$

One Erlang is the traffic intensity in a channel that is continuously occupied, so that a channel occupied for $x\%$ of the time carriers $x/100$ Erlangs. The spectral efficiency is defined as

$$\eta_S = \frac{N_c \cdot G_c}{W_{\text{sys}} \cdot A} \quad \text{Erlangs/m}^2/\text{Hz} \quad . \tag{1.34}$$

Suppose that the cellular deployment consists of N-cell reuse clusters. Then the number of channels per cell with FDMA is

$$N_c = \frac{W_{\text{sys}}}{W_c \cdot N} \qquad (1.35)$$

where W_c is the bandwidth per channel. If TDMA is used, then W_c is the bandwidth per carrier divided by the number of channels per carrier. The spectral efficiency can be written as the product of three efficiencies, viz.,

$$\begin{aligned}
\eta_S &= \frac{1}{W_c} \quad \cdot \quad \frac{1}{NA} \quad \cdot \quad \frac{G_c}{} \\
&= \eta_B \quad \cdot \quad \eta_C \quad \cdot \quad \eta_T \ .
\end{aligned} \qquad (1.36)$$

where

$$\begin{aligned}
\eta_B &= \text{bandwidth efficiency} \\
\eta_C &= \text{spatial efficiency} \\
\eta_T &= \text{trunking efficiency}
\end{aligned}$$

High bandwidth efficiency can be achieved by using low bit rate voice coding and bandwidth efficient signaling techniques.

Spatial Efficiency:.

High spatial efficiency can be achieved by i) minimizing the area per cell, and ii) minimizing the co-channel reuse distance. The first of these explains the intense interest in microcellular systems, where cell radii on the order of 200-500 m are used. The co-channel reuse distance D/R is minimized by i) controlling the generation of co-channel interference within the cellular system in the first place and, ii) minimizing the effect of the co-channel interference that is generated. The generated levels of co-channel interference can be controlled by using techniques such as cell sectoring, smart antennas, power control, discontinuous transmission, effective hand-off algorithms, macroscopic BS diversity, and others. The impact of co-channel interference on the radio link can be mitigated by using techniques such as interference cancellation, error control coding, antenna diversity, and others.

Consider the situation shown in Fig. 1.15, depicting the forward channel co-channel interference environment. The MS is at distance d_0 from the serving BS and at distances $d_k, k = 1, 2, \cdots, N_I$ from the first tier of $N_I = 6$ interfering co-channel BSs. If we let $\mathbf{d} = (d_0, d_1, \cdots, d_{N_I})$ denote the vector of distances at a particular MS location, then the downlink carrier-to-interference ratio as a function of $\mathbf{d}$ is

$$\Lambda_{(\text{dB})}(\mathbf{d}) = \Omega_{p\,(\text{dBm})}(d_0) - 10 \log_{10} \left\{ \sum_{k=1}^{N_I} 10^{\Omega_{p\,(\text{dBm})}(d_k)/10} \right\} \ . \qquad (1.37)$$
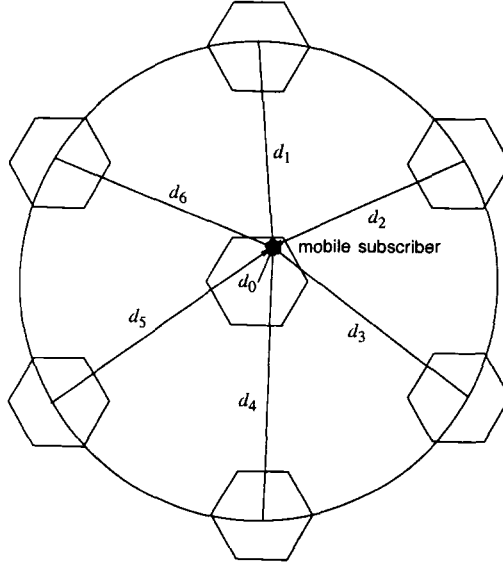
Figure 1.15. Co-channel interference on the forward channel at a desired MS. There are six interfering BSs.

At this point, we must account for the effect of handoffs. Consider, for example, soft handoffs. Let $\Lambda_{k\ (\text{dB})}(\mathbf{d}), k = 0, \ldots, M$ denote the carrier-to-interference ratio for serving BS and M surrounding BSs. Note that the vector $\mathbf{d}$ is different for each BS. With soft handoffs, the BS that provides the most robust link is always used so that the resulting carrier-to-interference ratio is

$$\Lambda_{(\text{dB})} = \max\{\Lambda_{0\ (\text{dB})}(\mathbf{d}), \Lambda_{1\ (\text{dB})}(\mathbf{d}), \ldots, \Lambda_{M\ (\text{dB})}(\mathbf{d})\} \qquad (1.38)$$

The area averaged probability co-channel interference outage is

$$O_I = \mathrm{P}\left(\Lambda_{(\text{dB})} < \Lambda_{\text{th(dB)}}\right) \ . \qquad (1.39)$$

where the calculation is performed by averaging the probability of co-channel interference outage over the random location of the MS within a reference cell.

Finally, Fig. 1.16 depicts the co-channel interference on the reverse channel at the serving BS. Note that the co-channel interference may not be exactly the same on the forward and reverse channels, because the vector $\mathbf{d}$ is different in each direction. This phenomenon is known as **link imbalance**.

Trunking Efficiency:.

High trunking efficiency can be achieved by using channel assignment schemes that maximize channel utilization. There is usually a trade-off be-
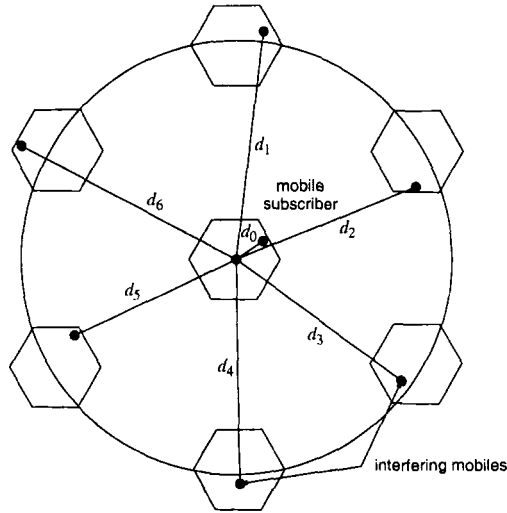
Figure 1.16. Co-channel interference on the reverse channel at a desired BS. There are six interfering MSs.

tween trunking efficiency (or offered traffic per channel) and grade of service in terms of new call and handoff blocking probabilities. Various fundamental formula were developed by Erlang, who laid the foundations of modern teletraffic theory. One of his most famous results is the **Erlang-B formula**, first derived in 1917, that gives the probability that a new call attempt will not find an available channel in a trunk of channels and is lost. Sometimes this policy is called the **blocked calls cleared** queueing discipline and it is widely used to model wireline telephone traffic. The Erlang-B formula is not really applicable to cellular systems, because it does not account for handoff traffic. Furthermore, the total offered traffic per cell is time-varying due to the spatial movement of the subscribers, whereas the offered traffic in the Erlang-B formula is assumed to be constant. Nevertheless, it provides useful insight. The Erlang-B formula is

$$B(\rho, m) = \frac{\rho^m}{m! \sum_{k=0}^{m} \frac{\rho^k}{k!}} \tag{1.40}$$

where m is the total number of channels in the trunk and $\rho = \lambda\mu$ is the total offered traffic (λ is the call arrival rate and μ is the mean call duration). The Erlang-B formula is derived under the assumption of an infinite subscriber population, Poisson call arrivals with rate λ calls/s, and exponentially distributed call durations with a mean call duration μ s/call.

Fig. 1.17 plots the blocking probability $B(\rho, m)$ as a function of the offered traffic per channel $G_c = \rho/m$. The benefit from **trunking** is obvious, since the offered traffic per channel, G_c, increases as the number of trunked channels
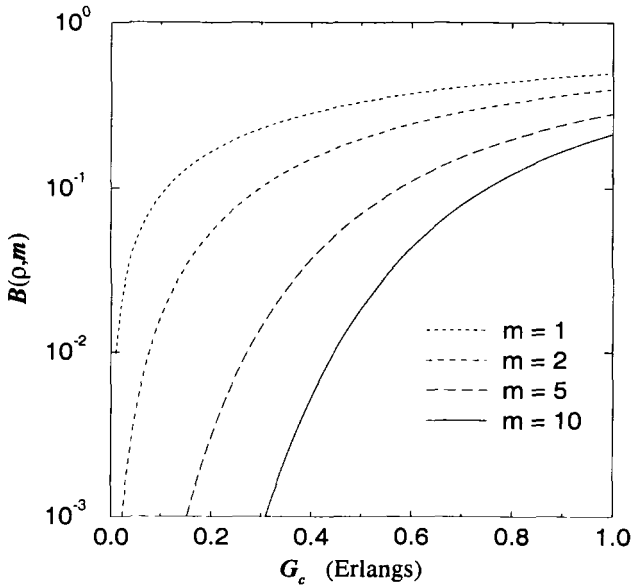
Figure 1.17. Erlang-B blocking probability $B(\rho, m)$ vs. offered traffic per channel $G_c = \rho/m$. Trunking is shown to improve the spectral efficiency.

increases, for any blocking probability. However, diminishing returns are obtained as the number of trunked channels becomes larger.

Capacity.

The capacity of a cellular system is often measured in terms of two quantities

1. the **cell capacity** or **sector capacity** is equal to the number of available voice channels per cell or cell sector.

2. the **cell Erlang capacity** is equal to the traffic carrying capacity of a cell (in Erlangs) for a specified call blocking probability.

Note that difference between spectral efficiency and Erlang capacity is that spectral efficiency accounts for the area per cell, A. If the area per cell is the same in two different cellular systems, then their relative spectral efficiencies and capacities will be the same.

Capacity comparisons between different cellular systems can be difficult, because the systems are often compared in different stages of their evolution and different deployment constraints. However, a fair comparison between suitably optimized digital cellular systems with out deployment constraints will probably show roughly equal capacities.

The cell capacity of FDMA (AMPS) and TDMA (PCS 1900, IS-54) cellular systems can be calculated in a fairly straight forward fashion once the allowable reuse cluster has been determined.

AMPS Capacity

Very often the capacity of 2nd and 3rd generation digital cellular systems (IS-54, IS-95, PCS 1900) are compared with the capacity of the analog AMPS system that is used in North America. Analog AMPS uses frequency division duplexing (FDD) with 30 kHz channels. In a bandwidth of 1.25 MHz (uplink or downlink only) there are $1250/30 \approx 42$ channels. Analog AMPS systems are typically deployed according to a 7/21 reuse pattern, i.e., there are 21 sectors in a reuse cluster. Hence, with analog AMPS there are 2 channels per sector. The corresponding cell capacity is 6.0 channels/cell. Likewise, in a bandwidth of 15 MHz, the sector capacity is 24 channels/sector.

PCS1900 Capacity

GSM systems in Europe were originally deployed *without frequency hop-ping*. A 4/12 reuse pattern was very common. For PCS 1900 *with frequency hopping*, a 3/9 reuse pattern may be possible. PCS 1900 has 8 channels that are time division multiplexed onto each carrier, and the carrier spacings are 200 kHz. Therefore, the bandwidth per channel is 25 kHz. In a bandwidth of 1.25 MHz (uplink or downlink only) there are $1250/25 = 50$ channels. Hence, there are $50/9 \approx 5$ channels per sector or $50/3 = 17$ channels/cell. Therefore, the cell capacity of PCS 1900 with a 3/9 reuse pattern is $17/6 = 2.8$ times the AMPS cell capacity.

IS-95 Capacity

The cell capacity of IS-95 has been the topic of debate for many years, because it cannot be determined in a straight forward manner. The capacity depends on a variety of complicated factors like (i) the propagation path loss exponent, (ii) the accuracy of the power control loop, and (iii) the geographical distribution of mobiles within a cell. To illustrate the difficulty in evaluating IS-95 CDMA capacity, consider the following simple example. Suppose there are N users in a cell; one desired user and $N - 1$ interfering users. Treating the co-channel signals as white Gaussian noise, the carrier-to-noise ratio is

$$\Gamma = \frac{1}{N-1} \qquad (1.41)$$

and the modulated symbol energy-to-noise ratio is

$$\begin{aligned} \frac{E_c}{N_o} &= \Gamma \times \frac{B_w}{R_c} \\ &= \frac{G}{N-1} \approx \frac{G}{N} \end{aligned}$$

where $G = B_w/R_c$. For a required $(E_c/N_o)_{\text{req}}$, the number of users that can be accommodated is

$$N \approx \frac{G}{(E_c/N_o)_{\text{req}}}$$

This figure is just the cell capacity. Now if $(E_c/N_o)_{\text{req}}$ is reduced by only 1 dB, i.e., a factor of 1.25 there is a 30% change in N, the cell capacity. Hence, the capacity of CDMA cellular systems is highly sensitive to the receiver sensitivity.

Problems

1.1. Show that the area noise outage probability is given by (1.25).

1.2. By using geometric arguments, show that the co-channel reuse factor for cellular deployments based on hexagonal cells is given by (1.2).

1.3. Consider a regular hexagonal cell deployment, where the MSs and BSs use omnidirectional antennas. Suppose that we are interested in the forward channel performance and consider only the first tier of co-channel interferers as shown in Fig. 1.15. Ignore the effects of shadowing and multipath fading, and assume that the propagation path loss is described by the inverse β law in (1.6).

    a) Determine the worst case carrier-to-interference ratio, Λ, as a function of the reuse cluster size N, for $\beta = 3, 3.5$, and 4.

    b) What is the minimum cluster size that is needed if the radio receivers have $\Lambda_{\text{th}} = 18$ dB?

    c) Referring to Fig. 1.16, repeat a) and b) for the reverse channel.

1.4. Whenever a mobile station crosses a cell boundary a handoff occurs to the target cell. However, a handoff will sometimes "fail" because there are no channels available in the target cell. One method to decrease the probability of handoff failure is to queue the handoff calls. A handoff call that does not find an idle channel in the target cell is allowed to remain in a queue for t_q seconds and is dropped from the queue, i.e., experience a **handoff failure**, if no channel becomes available in that time.

Suppose the queue is serviced using a "first come first served" discipline. If m is the total number of channels in the trunk and ρ is the total offered traffic, then the probability of queueing is given by the famous **Erlang-C formula**

$$C(\rho, m) = \frac{\rho^m}{\rho^m + m! \left(1 - \frac{\rho}{m}\right) \sum_{k=0}^{m-1} \frac{\rho^k}{k!}}$$

The probability that a queued call will have to wait more than t_q seconds in the queue is

$$P(W > t_q) = \exp\left\{-\frac{(m - \rho)t_q}{\mu}\right\}$$

where μ is the mean call duration. Assuming that $\mu = 120$ s and $t_q = 5$ s, plot the blocking probability against the normalized offered traffic per channel $G_c = \rho/m$, for $m = 5, 10, 15$.

1.5. Consider the worst case forward channel co-channel interference situation shown in Fig. 1.18 The path loss is described by the following simple model
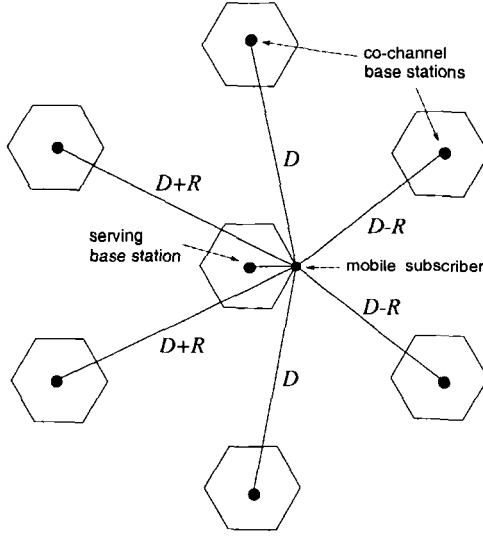
Figure 1.18.   Worst case co-channel interference on the forward channel.

$$\mu_{\Omega_p} = \frac{\Omega_t (h_b h_m)^2}{d^4}$$

where

$$
\begin{aligned}
\mu_{\Omega_p} &= \text{received power} \\
\Omega_t &= \text{transmitted power} \\
h_b &= \text{base station antenna height} \\
h_m &= \text{mobile station antenna height} \\
d &= \text{radio path length}
\end{aligned}
$$

a) Assuming that $h_b = 30$ m, $h_m = 1.5$ m, and all BS transmit powers are the same what is the worst case Λ for a cluster size $N = 4$?

b) Now suppose that the antenna height of the serving BS (in the center) is increased to 40 m while the other BS antenna heights remain at 30 m. This has the effect of enlarging the center cell. Assuming that we wish to maintain the same worst case Λ value obtained in part a), what is the new radius of the center cell?

c) Now suppose that the antenna height of one of the co-channel BSs is increased to 40 m while the antenna heights of the other BSs antenna heights, including the serving BS, remain at 30 m. This has the effect of shrinking the center cell and making it a non-regular hexagon. Assuming, again, that we wish to maintain the same worst case Λ value obtained in part a), what are the new dimensions of the center cell?

1.6. A cellular service provider uses a digital modulation scheme which can tolerate a worst-case signal-to-interference ratio of 15 dB.

a) Find the optimal cluster size N for the following cases;

  (i) omni-directional antennas

 (ii) 120° sectoring

(iii) 60° sectoring

Use path loss exponents of $\beta = 3$ and $\beta = 4$.

b) Assume that there are 200 traffic channels in the cellular system and that a blocked calls cleared queueing discipline is used with a target blocking probability of 1%. Further assume that each cell or sector has approximately the same number of channels, and the cells have uniform traffic loading. Ignore any handoff traffic. Determine the offered traffic load (per cell) in units of *Erlangs* and *calls per hour* for each of the cases in part (a).

1.7. Suppose that an urban area has three competing trunked mobile networks (systems A, B, and C) to provide cellular service. System A has 400 cells with 15 channels/cell, System B has 50 cells with 100 channels/cell, and System C has 100 cells with 60 channels/cell. Ignore handoff traffic and assume uniform cell traffic loading.

a) Plot the (Erlang-B) blocking probability, $B(\rho, m)$, for each system versus ρ.

b) Find the number of users that can be accommodated by each system for a blocking probability of 2% if the traffic loading offered by each user is 0.1 Erlangs.

Chapter 2

PROPAGATION MODELING

The design of spectrally efficient wireless communication systems requires a detailed understanding of the radio propagation environment. The characteristics of the radio channel vary greatly with the operating frequency, and the mode of propagation, e.g., line-of-sight (LoS) radio links, diffraction/scatter, and satellite links. In this book the emphasis is on land mobile radio channels that are typical of terrestrial cellular mobile radio systems, although many of the concepts will apply to other types of channels as well.

A typical cellular radio system consists of a collection of fixed base stations (BSs) that define the radio coverage areas or cells[1]. The height and placement of the BS antennas affects the proximity of local scatterers at the BS. In a macrocellular environment, the BS antennas are usually well elevated above the local terrain and relatively free of local scatterers. Typically, a non-line-of-sight (NLoS) radio propagation path will exist between a BS and mobile station (MS), because of natural and man-made objects that are situated between the BS and MS. As a consequence the radio waves must propagate via reflections, diffraction and scattering. At the MS, plane waves arrive from many different directions and with different delays, as shown in Fig. 2.1. This property is called **multipath propagation**. The multiple plane waves combine vectorially at the receiver antenna to produce a composite received signal.

The carrier wavelength used in UHF mobile radio applications typically ranges from 15 to 60 cm. Therefore, small changes in the differential propagation delays due to MS mobility will cause large changes in the phases of the individually arriving plane waves. Hence, the arriving plane waves arriving at the MS and BS antennas will experience constructive and destructive addition

---

[1]In military applications the BSs may be moving.
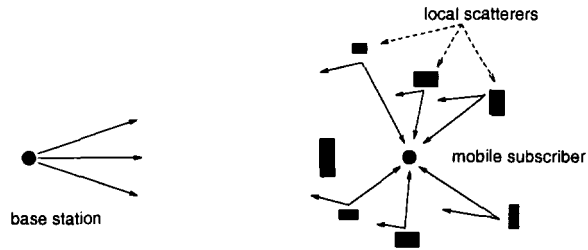
Figure 2.1. Typical macrocellular radio propagation environment.

depending on the location of the MS. If the MS is moving or there are changes in the scattering environment, then the spatial variations in the amplitude and phase of the composite received signal will manifest themselves as time variations, a phenomenon called **envelope fading**. As we will see later, the time rate of envelope fading depends on the velocity of the MS.

Radio channels are reciprocal in the sense that if a propagation path exists, it carries energy equally well in both directions. However, the spatial distribution of arriving plane waves may be significantly different in each direction. A MS in a typical macrocellular environment is usually surrounded by local scatterers so that the plane waves will arrive from many directions without a direct LoS component. Two-dimensional isotropic scattering where the arriving plane waves arrive in from all directions with equal probability is a very commonly used scattering model for the forward channel in a macrocellular system. For this type of scattering environment the received envelope is Rayleigh distributed at any time, and is said to exhibit Rayleigh fading.

The BSs in macrocells are relatively free from local scatterers so that the plane waves tend to arrive from one direction with a fairly small angle of arrival (AoA) spread as shown in Fig. 2.1. We will see later that these differences in the scattering environment for the forward and reverse channels cause differences in the spatial correlation properties of their respective faded envelopes.

In a microcellular environment, the BS antennas are often placed below the skyline of buildings and are surrounded by local scatterers, such that the plane waves will arrive at the BS with a larger AoA spread. Furthermore, a LoS path will sometimes exist between the MS and BS, while at others times there is no LoS path. Even in the absence of LoS propagation conditions, there often exists a dominant reflected or diffracted path between the MS and BS. The LoS or dominant reflected or diffracted path produces the **specular** component and the multitude of weaker secondary paths contribute to the **scatter** component of the received envelope. In this type of propagation environment, the received signal envelope still experiences fading. However, the presence of the specular component changes the received envelope distribution, and very often a Ricean

distributed envelope is assumed [148, 369, 359]. In this case the received envelope is said to exhibit Ricean fading.

If the envelope or squared-envelope is measured and averaged over a spatial distance of 20 to 30 wavelengths, the mean envelope or mean squared-envelope can be obtained. Sometimes, this quantity is called the **local mean** because it corresponds to the mean value a particular locality. Usually, the local mean will also experience slow variations over distances of several tens of wavelengths due to the presence of large terrain features such as buildings and hills. This phenomenon is known as shadow fading or **shadowing**. Experimental observations have confirmed that the shadow fades follow a log-normal distribution as in (1.5). This log-normal distribution applies to both macrocellular [188, 173] and microcellular environments [224, 226, 149].

If the local mean is averaged over sufficiently large spatial distances (to average over the shadows), the **area mean** is obtained. The area mean is the average signal strength that is received to/from a MS over a large area that lies at (approximately) the same distance from the BS. The area mean is directly related to the **path loss**, which predicts how the area mean varies with the distance between the BS and MS. Early studies by Okumura [253] and Hata [162] yielded empirical path loss models for urban, suburban, and rural areas that are accurate to within 1 dB for distances ranging from 1 to 20 km. These studies concentrated on macrocellular systems. More recent work has considered path loss prediction in microcells. The COST231 study [69] resulted in the COST231-Hata and COST231-Walfish-Ikegami models for urban microcellular path loss prediction.

The remainder of this chapter presents the fundamentals of radio propagation modeling and characterization. Section 1. introduces the mechanism of multipath-fading. Various properties of the faded envelope are then derived in Sections 1.1 through 1.5. Section 2. treats the statistical characterization of wide-band multipath-fading channels. Laboratory simulation of fading channels is covered in Section 3.. Shadowing models and simulation techniques are discussed in Section 4.. Finally, Section 5. treats theoretical and empirical models for path loss in macrocellular and microcellular systems.

1. FREQUENCY-NON-SELECTIVE (FLAT) MULTIPATH-FADING

In terrestrial cellular radio systems, the radio signals propagate in three dimensions. The signals that are transmitted by the BSs usually have vertical polarization. For vehicular applications, the MS antennas are also vertically polarized, while for portable applications tilting of the transmitter (or handset) antenna results in non-vertical polarization. Although it is important to account for polarization effects, we will assume that the transmitted signals are vertically polarized. Furthermore, we assume that the distance between the BS and MS
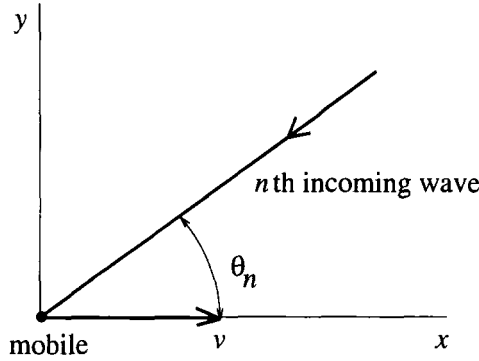
Figure 2.2. A typical plane wave incident on a MS receiver.

is sufficiently large so that the radio propagation environment can be modeled as two-dimensional.

Fig. 2.2 depicts a horizontal $x - y$ plane, where a MS is moving along the x-axis with velocity v. Vertical polarization is assumed throughout so that the electric field vector is aligned with the z-axis. The nth plane wave arrives at the MS antenna with an angle of incidence θ_n. The MS movement introduces a **Doppler** shift, or frequency shift, into the incident plane wave. The Doppler shift is given by

$$f_{D,n} = f_m \cos \theta_n \ \text{Hz} \qquad (2.1)$$

where $f_m = v/\lambda_c$ and λ_c is the wavelength of the arriving plane wave, and f_m is the maximum Doppler frequency occurring when $\theta_n = 0$. Plane waves arriving from the direction of motion will experience a positive Doppler shift, while those arriving opposite the direction of motion will experience a negative Doppler shift.

Consider the transmission of the band-pass signal

$$s(t) = \text{Re} \left[\tilde{s}(t) e^{j 2\pi f_c t} \right] \qquad (2.2)$$

where $\tilde{s}(t)$ is the complex envelope of the transmitted signal, f_c is the carrier frequency, and $\text{Re}[z]$ denotes the real part of z. If the channel is comprised of N propagation paths, then the noiseless received band-pass waveform is

$$r(t) = \text{Re} \left[\sum_{n=1}^{N} C_n e^{j 2\pi \left[(f_c + f_{D,n})(t - \tau_n) \right]} \tilde{s}(t - \tau_n) \right] \qquad (2.3)$$

where C_n and τ_n are the amplitude and time delay, respectively, associated with the nth propagation path. The magnitude C_n depends on the cross sectional area of the nth reflecting surface or the length of the nth diffracting edge.

Similar to (2.2), the received band-pass signal $r(t)$ has the form

$$r(t) = \text{Re}\left[\tilde{r}(t)e^{j2\pi f_c t}\right] \tag{2.4}$$

where the received complex envelope is

$$\tilde{r}(t) = \sum_{n=1}^{N} C_n e^{-j\phi_n(t)}\tilde{s}(t - \tau_n) \tag{2.5}$$

and

$$\phi_n(t) = 2\pi \left\{ (f_c + f_{D,n})\tau_n - f_{D,n}t \right\} \tag{2.6}$$

is the phase associated with the nth path. From (2.5), the channel can be modeled by a linear time-variant filter having the complex low-pass impulse response

$$g(t,\tau) = \sum_{n=1}^{N} C_n e^{-j\phi_n(t)}\delta(\tau - \tau_n) \tag{2.7}$$

where $g(\tau, t)$ is the channel response at time t due to an impulse applied at time $t - \tau$, and $\delta(\,\cdot\,)$ is the dirac delta function.

From (2.5) and (2.6), several interesting observations can be made. Since the carrier frequency f_c is very large, very small changes in the path delays τ_n will cause a large changes in the phases $\phi_n(t)$, due to the term $f_c\tau_n$. For example, a 900 MHz sinusoid has a wavelength of about 30 cm. Since, radio waves propagate at about 30 cm per nanosecond (ns), a path delay change of just 1 ns corresponds to one full wavelength (or 2π radians phase shift) in the 900 MHz sinusoid. At any time t, the random phases $\phi_n(t)$ may result in the constructive or destructive addition of the N multipath components. Multipath fading is primarily due to small variations in the path delays and, hence the received phases, of the multipath components that occur over small spatial distances.

If the differential path delays $\tau_i - \tau_j$ are small compared to the duration of a modulated symbol, then the τ_n in (2.7) are all approximately equal to $\hat{\tau}$. In this case, the channel impulse response has the form

$$g(t,\tau) = \sum_{n=1}^{N} C_n e^{-j\phi_n(t)}\delta(\tau - \hat{\tau}) = g(t)\delta(\tau - \hat{\tau}) \ . \tag{2.8}$$

However, since the carrier frequency is very high, small differences in the path delays will still correspond to large differences in the received phases $\phi_n(t)$. Therefore, the received signal still experiences fading. The corresponding channel transfer function is obtained by taking the Fourier transform of (2.8), giving

$$T(t,f) = g(t)e^{-j2\pi f\hat{\tau}} \ . \tag{2.9}$$

Since the amplitude response is $|T(t, f)| = g(t)$, all frequency components in the received signal are subject to the same complex gain $g(t)$. In this case the received signal is said to exhibit **flat fading**.

1.1 RECEIVED SIGNAL CORRELATION AND SPECTRUM

A flat fading channel can be characterized by assuming the transmission of an unmodulated carrier. Since $\tilde{s}(t) = 1$ in (2.5), the received band-pass signal in (2.4) can be expressed in the quadrature form

$$r(t) = g_I(t) \cos 2\pi f_c t - g_Q(t) \sin 2\pi f_c t \tag{2.10}$$

where

$$g_I(t) = \sum_{n=1}^{N} C_n \cos \phi_n(t) \tag{2.11}$$

$$g_Q(t) = \sum_{n=1}^{N} C_n \sin \phi_n(t) \tag{2.12}$$

are the inphase and quadrature components of the received band-pass signal. For large N, the central limit theorem can be invoked and $g_I(t)$ and $g_Q(t)$ can be treated as Gaussian random processes. Assuming that the band-pass process $r(t)$ is wide sense stationary, the autocorrelation of $r(t)$ is

$$
\begin{aligned}
\phi_{rr}(\tau) &= \mathrm{E}[r(t)r(t+\tau)] \\
&= \mathrm{E}[g_I(t)g_I(t+\tau)] \cos 2\pi f_c\tau - \mathrm{E}[g_Q(t)g_I(t+\tau)] \sin 2\pi f_c\tau \\
&= \phi_{g_I g_I}(\tau) \cos 2\pi f_c\tau - \phi_{g_Q g_I}(\tau) \sin 2\pi f_c\tau
\end{aligned}
\tag{2.13}
$$

where

$$\phi_{g_I g_I}(\tau) = \phi_{g_Q g_Q}(\tau) \tag{2.14}$$
$$\phi_{g_I g_Q}(\tau) = \phi_{g_Q g_I}(-\tau) . \tag{2.15}$$

It is reasonable to assume that the phases $\phi_n(t)$ and $\phi_m(t)$ are independent for $n \neq m$ since their associated delays and Doppler shifts are independent. Furthermore, the phases $\phi_n(t)$ can be assumed to be uniformly distributed over $[-\pi, \pi]$, since $f_c \tau_n \gg 1$. By using these properties, it is straightforward to obtain the autocorrelation $\phi_{g_I g_I}(\tau)$ from (2.1), (2.11) and (2.6) as follows:

$$
\begin{aligned}
\phi_{g_I g_I}(\tau) &= \mathrm{E}_{\tau,\theta}[g_I(t)g_I(t+\tau)] \\
&= \frac{\Omega_p}{2} \mathrm{E}_{\theta_i}[\cos 2\pi f_{D,n}\tau] \\
&= \frac{\Omega_p}{2} \mathrm{E}_{\theta}[\cos(2\pi f_m \tau \cos \theta)]
\end{aligned}
\tag{2.16}
$$

where

$$\boldsymbol{\tau} = (\tau_1, \tau_2, \dots, \tau_N) \tag{2.17}$$

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_N) \tag{2.18}$$

$$\Omega_p = \mathrm{E}[g_I^2(t)] + \mathrm{E}[g_Q^2(t)] = \sum_{n=1}^{N} C_n^2 \tag{2.19}$$

and Ω_p is the total received envelope power. Note that the power in the bandpass waveform $r(t)$ is $\mathrm{E}[r^2(t)] = \Omega_p/2$.

Similarly, the crosscorrelation $\phi_{g_I g_Q}(\tau)$ is

$$\phi_{g_I g_Q}(\tau) = \mathrm{E}_{\boldsymbol{\tau}, \boldsymbol{\theta}}[g_I(t) g_Q(t + \tau)]$$

$$= \frac{\Omega_p}{2} \mathrm{E}_{\theta}[\sin(2\pi f_m \tau \cos \theta)] \ . \tag{2.20}$$

Evaluation of the expectations in (2.16) and (2.20) requires the distribution of incident power on the receiver antenna, $p(\theta)$, and the receiver antenna gain $G(\theta)$ as a function of the AoA, θ. For macrocellular applications where the radio path lengths are long compared to the antenna heights, one simple model assumes that the plane waves propagate in a 2-D (x, y) plane and arrive at the MS from all directions with equal probability, i.e., $p(\theta) = 1/(2\pi), \theta \in [-\pi, \pi]$. This model was first suggested by Clarke [64], and is commonly referred to as Clarke's 2-D **isotropic scattering** model. With 2-D isotropic scattering and an isotropic receiver antenna with gain $G(\theta) = 1$, the expectation in (2.16) becomes

$$\phi_{g_I g_I}(\tau) = \frac{\Omega_p}{2} \int_{-\pi}^{\pi} \cos\left(2\pi f_m \tau \cos \theta\right) p(\theta) G(\theta) d\theta$$

$$= \frac{\Omega_p}{2} \frac{1}{2\pi} \int_{-\pi}^{\pi} \cos\left(2\pi f_m \tau \cos \theta\right) d\theta$$

$$= \frac{\Omega_p}{2} \frac{1}{\pi} \int_{0}^{\pi} \cos\left(2\pi f_m \tau \sin \theta\right) d\theta$$

$$= \frac{\Omega_p}{2} J_0(2\pi f_m \tau) \tag{2.21}$$

where $J_0(x)$ is the zero-order Bessel function of the first kind. The normalized autocorrelation function $\phi_{g_I g_I}(\tau)/(\Omega_p/2)$ in (2.21) is plotted against the normalized time delay $f_m \tau$ in Fig. 2.3.

Likewise, for 2-D isotropic scattering and an isotropic antenna, the crosscorrelation in (2.20) becomes

$$\phi_{g_I g_Q}(\tau) = \frac{\Omega_p}{2} \frac{1}{2\pi} \int_{-\pi}^{\pi} \sin\left(2\pi f_m \tau \cos \theta\right) d\theta \tag{2.22}$$

$$= 0 \ .$$
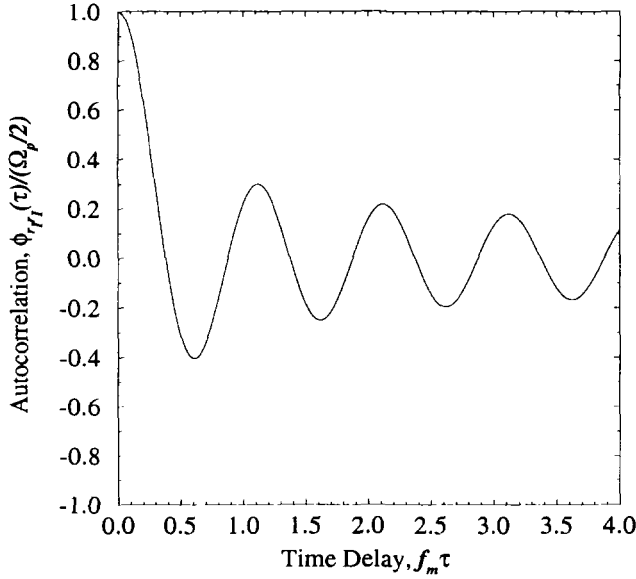
Figure 2.3. Autocorrelation of the real and imaginary components of the received complex envelope with isotropic scattering.

This means that $g_I(t)$ and $g_Q(t)$ are uncorrelated and, since they are Gaussian, independent random processes. The fact that $g_I(t)$ and $g_Q(t)$ are independent is a direct result of the symmetry of the 2-D isotropic scattering environment and the isotropic antenna gain pattern. Some scattering environments and antenna gain patterns will lead to independent $g_I(t)$ and $g_Q(t)$ processes, while others will not.

The power density spectrum (psd) of $g_I(t)$ and $g_Q(t)$ is the Fourier transform of $\phi_{g_I g_I}(\tau)$ or $\phi_{g_Q g_Q}(\tau)$. For the autocorrelation in (2.21), the corresponding psd is [147, 6.671.7]

$$
\begin{aligned}
S_{g_I g_I}(f) &= \mathcal{F}[\phi_{g_I g_I}(\tau)] \\
&= \begin{cases} \dfrac{\Omega_p}{2\pi f_m} \dfrac{1}{\sqrt{1-(f/f_m)^2}} & |f| \leq f_m \\ 0 & \text{otherwise} \end{cases}
\end{aligned} \qquad (2.23)
$$

The autocorrelation of the received complex envelope $g(t) = g_I(t) + jg_Q(t)$ is

$$
\begin{aligned}
\phi_{gg}(\tau) &= \frac{1}{2}\mathrm{E}[g^*(t)g(t+\tau)] \\
&= \phi_{g_I g_I}(\tau) + j\phi_{g_I g_Q}(\tau)
\end{aligned} \qquad (2.24)
$$

and its power spectral density is

$$
S_{gg}(f) = S_{g_I g_I}(f) + jS_{g_I g_Q}(f) \; . \qquad (2.25)
$$

Sometimes $S_{gg}(f)$ is called the **Doppler power spectrum** From (2.13) we have

$$\phi_{rr}(\tau) = \text{Re}\left[\phi_{gg}(\tau)e^{j2\pi f_c \tau}\right] . \tag{2.26}$$

By using the identity

$$\text{Re}[z] = \frac{z + z^*}{2} \tag{2.27}$$

and the property $\phi_{gg}(\tau) = \phi_{gg}^*(-\tau)$, it follows that the band-pass Doppler power spectrum is

$$S_{rr}(f) = \frac{1}{2}\left[S_{gg}(f - f_c) + S_{gg}(-f - f_c)\right] . \tag{2.28}$$

With 2-D isotropic scattering and an isotropic antenna $\phi_{g_I g_Q}(\tau) = 0$ and $S_{gg}(f) = S_{g_I g_I}(f)$ (which is real and even), so that

$$S_{rr}(f) = \frac{\Omega_p}{4\pi f_m} \frac{1}{\sqrt{1 - \left(\frac{f-f_c}{f_m}\right)^2}} , \qquad |f - f_c| \le f_m . \tag{2.29}$$

The psd in (2.29) can be derived by using a different approach that is sometimes more useful. As $N \to \infty$, the incident power on the receiver antenna as a function of the angle of incidence θ approaches a continuous distribution, denoted by $p(\theta)$. The fraction of the total incoming power that arrives between θ and $\theta + d\theta$ is $p(\theta)d\theta$. If the antenna has a gain of $G(\theta)$ at angle θ, then the corresponding received power is $G(\theta)p(\theta)d\theta$. Therefore, the psd of the received signal can be expressed as

$$S_{rr}(f)|df| = \frac{\Omega_p}{4}\{G(\theta)p(\theta) + G(-\theta)p(-\theta)\}|d\theta| . \tag{2.30}$$

From Fig. 2.2, the frequency of the incident plane wave arriving at angle θ is

$$f = f_m \cos\theta + f_c , \tag{2.31}$$

where $f_m = v/\lambda_c$ is the maximum Doppler shift and, hence,

$$|df| = f_m|-\sin\theta d\theta| = \sqrt{f_m^2 - (f - f_c)^2}|d\theta| . \tag{2.32}$$

Therefore,

$$S_{rr}(f) = \frac{\Omega_p/4}{\sqrt{f_m^2 - (f - f_c)^2}}\{G(\theta)p(\theta) + G(-\theta)p(-\theta)\} \tag{2.33}$$

where

$$\theta = \cos^{-1}\left(\frac{f - f_c}{f_m}\right) . \tag{2.34}$$
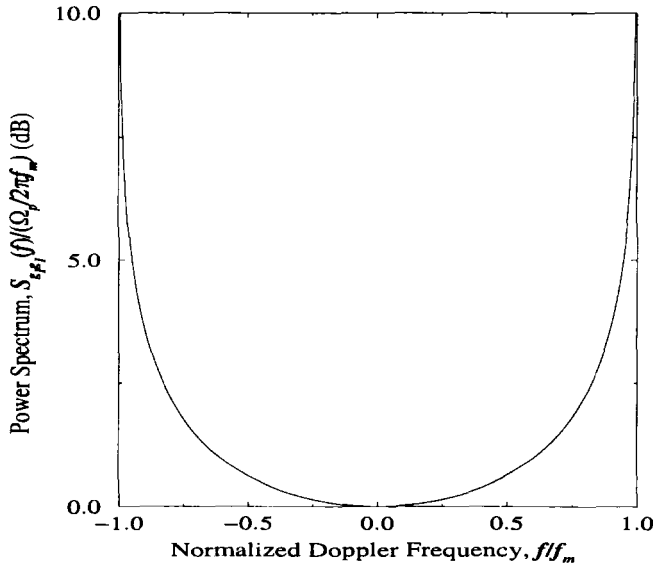
Figure 2.4. Psd of the received quadrature envelope components for a 2-D isotropic scattering channel with an isotropic antenna.

Once again, with 2-D isotropic scattering and an isotropic antenna $G(\theta)p(\theta) = 1/(2\pi)$, so that

$$S_{rr}(f) = \begin{cases} \dfrac{\Omega_p}{4\pi f_m} \dfrac{1}{\sqrt{1-\left(\frac{f-f_c}{f_m}\right)^2}} & |f - f_c| \le f_m \\ 0 & \text{otherwise} \end{cases} \qquad (2.35)$$

The same result was obtained in (2.29).

The normalized psd $S_{g_I g_I}(f)/(\Omega_p/2\pi f_m)$ in (2.23) is plotted against the normalized Doppler frequency f/f_m in Fig. 2.4. Notice that $S_{g_I g_I}(f)$ is limited to the range of frequencies $|f| \le f_m$ and $S_{g_I g_I}(f) = \infty$ at $f = \pm f_m$. In reality the Doppler psd can never go to infinity, and the reason for this behavior is that the plane waves were assumed to propagate in a 2-D plane, whereas in reality the propagation is actually three dimensional. Aulin [16] modified Clarke's 2-D model to account for 3-D propagation. The psd that Aulin obtained is very similar to Fig. 2.4, except that it remains finite at $f = \pm f_m$.

In some cases, it is appropriate to model the propagation environment as consisting of a strong specular component plus a scatter component. In this case, the AoA distribution $p(\theta)$ might have the form

$$p(\theta) = \frac{1}{K+1}\hat{p}(\theta) + \frac{K}{K+1}\delta(\theta - \theta_0) \qquad (2.36)$$

where $\hat{p}(\theta)$ is the continuous AoA distribution of the *scatter* component, θ_0 is the AoA of the specular component, and K is the ratio of the received specular to scattered power. Fig. 2.5 shows a polar plot of $p(\theta)$ for such a scattering environment, where $\hat{p}(\theta) = 1/(2\pi), \theta \in [-\pi, \pi]$. The correlation functions $\phi_{g_I g_I}(\tau)$ and $\phi_{g_I g_Q}(\tau)$ corresponding to (2.36) can be readily obtained from (2.16) and (2.20) as

$$\phi_{g_I g_I}(\tau) = \frac{1}{K+1} \frac{\Omega_p}{2} J_0(2\pi f_m \tau) + \frac{K}{K+1} \frac{\Omega_p}{2} \cos(2\pi f_m \tau \cos\theta_0) \quad (2.37)$$

$$\phi_{g_I g_Q}(\tau) = \frac{K}{K+1} \frac{\Omega_p}{2} \sin(2\pi f_m \tau \cos\theta_0) \ . \quad (2.38)$$

The AoA distribution in (2.36) yields a complex envelope having a psd of the form

$$S_{gg}(f) = \frac{1}{K+1} S_{gg}^c(f) + \frac{K}{K+1} S_{gg}^d(f) \quad (2.39)$$

where $S_{gg}^d(f)$ is the discrete portion due to the specular component and $S_{gg}^c(f)$ is the continuous portion due to the scatter component. For the case when $\hat{p}(\theta) = 1/(2\pi), \theta \in [-\pi, \pi]$ with the resulting correlation functions in (2.37) and (2.38), we have

$$S_{gg}(f) = \begin{cases} \frac{1}{K+1} \cdot \frac{\Omega_p}{2\pi f_m} \frac{1}{\sqrt{1-(f/f_m)^2}} \\ \quad + \frac{K}{K+1} \frac{\Omega_p}{2} \delta(f - f_m \cos\theta_0) & 0 \le |f| \le f_m \\ 0 & \text{otherwise} \end{cases} \quad (2.40)$$

The corresponding band-pass psd $S_{rr}(f)$ has the same form as Fig. 2.4, except for a discrete tone at frequency $f_c + f_m \cos\theta_0$.

For microcells that are deployed in dense urban areas, the plane waves may be channeled by the buildings along the streets and arrive at the receiver antenna from just one direction, as shown in Fig. 2.6. Clearly, the scattering is non-isotropic. In this case, a variety of models may be used for distribution of arriving plane waves. One plausible distribution is

$$p(\theta) = \begin{cases} \frac{\pi}{4|\theta_m|} \cos\left(\frac{\pi}{2} \cdot \frac{\theta}{\theta_m}\right) & , \ |\theta| \le |\theta_m| \le \frac{\pi}{2} \\ 0 & , \ \text{elsewhere} \end{cases} \quad (2.41)$$

The parameter θ_m determines the directivity of the incoming waves. Fig. 2.7 shows a plot of $p(\theta)$ for $\theta_m = 30^\circ, 60^\circ$, and 90°. Note that the pdf is symmetric about $\theta = 0$.

The correlation functions $\phi_{g_I g_I}(\tau)$ and $\phi_{g_I g_Q}(\tau)$ can be readily obtained by evaluating the expectations in (2.16) and (2.20), respectively, with the density in (2.41). Again, the psd of the received band-pass signal $S_{rr}(f)$ can be obtained by using (2.24), taking Fourier transforms, and substituting into (2.28).
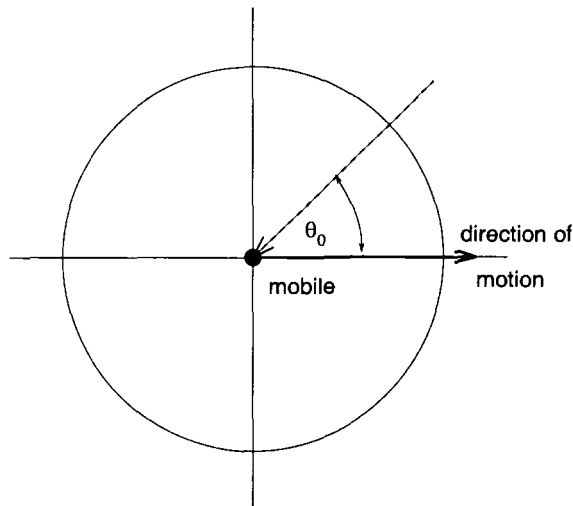
Figure 2.5. Polar plot of $p(\theta)$ with 2-D isotropic scattering plus a LoS or specular component arriving at angle θ_0.
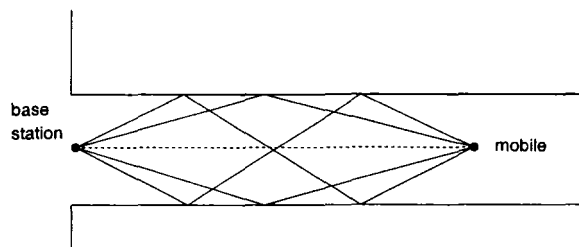
Figure 2.6. An urban microcellular propagation environment is often characterized by non-isotropic scattering.

1.2 RECEIVED ENVELOPE AND PHASE DISTRIBUTION

1.2.1 RAYLEIGH FADING

When the composite received signal consists of a large number of plane waves, the received complex envelope $g(t) = g_I(t) + jg_Q(t)$ can be treated as a wide-sense stationary complex Gaussian random process. For some types of scattering environments, e.g., 2-D isotropic scattering, $g_I(t)$ and $g_Q(t)$ are independent identically distributed zero-mean Gaussian random variables at
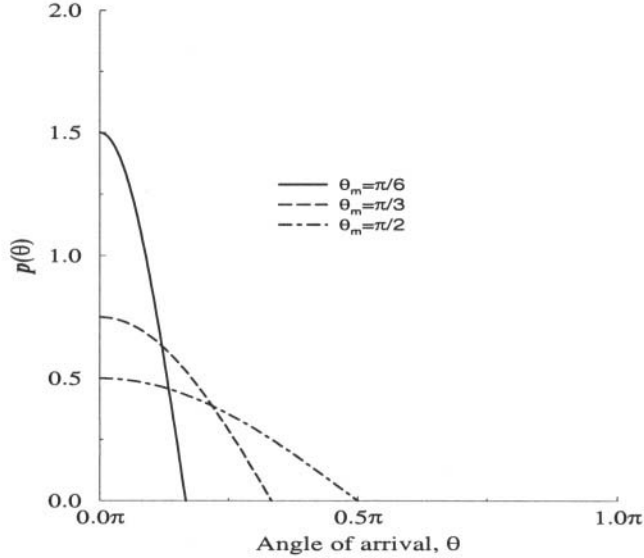
Figure 2.7. Probability density function of the arriving planes waves in 2.41. The pdf is symmetrical about $\theta = 0$.

any time t_1, with variance b_0. Under these conditions the magnitude of the received complex envelope $\alpha(t) = |g(t)|$ has a Rayleigh distribution at any time t_1, as shown in (A.26) i.e.,

$$p_\alpha(x) = \frac{x}{b_0} \exp\left\{ -\frac{x^2}{2b_0} \right\} \ . \tag{2.42}$$

The average envelope power is $E[\alpha^2] = \Omega_p = 2b_0$ so that

$$p_\alpha(x) = \frac{2x}{\Omega_p} \exp\left\{ -\frac{x^2}{\Omega_p} \right\} \qquad x \geq 0 \ . \tag{2.43}$$

This type of fading is called **Rayleigh fading.** The corresponding squared-envelope $\alpha^2(t) = |g(t)|^2$ is exponentially distributed at any time t_1 with density

$$p_{\alpha^2}(x) = \frac{1}{\Omega_p} \exp\left\{ -\frac{x}{\Omega_p} \right\} \ . \tag{2.44}$$

The squared-envelope at time t is significant because it is proportional to the *instantaneously* received signal power at time t.

1.2.2   RICEAN FADING

Some types of scattering environments have a specular or LoS component. In this case, $g_I(t)$ and $g_Q(t)$ are Gaussian random processes with non-zero

means $m_I(t)$ and $m_Q(t)$, respectively. If we again assume that these processes are uncorrelated and the random variables $g_I(t_1)$ and $g_Q(t_1)$ have the same variance b_0, then the magnitude of the received complex envelope at time t_1 has a Ricean distribution as shown in (A.32), i.e.,

$$p_\alpha(x) = \frac{x}{b_0} \exp\left\{-\frac{x^2 + s^2}{2b_0}\right\} I_o\left(\frac{xs}{b_0}\right) \qquad x \geq 0, \qquad (2.45)$$

where

$$s^2 = m_I^2(t) + m_Q^2(t) \qquad (2.46)$$

is called the non-centrality parameter. This type of fading is called **Ricean fading** and is very often observed in microcellular and mobile satellite applications.

A very simple Ricean fading model assumes that the means $m_I(t)$ and $m_Q(t)$ are constants, i.e., $m_I(t) = m_I$ and $m_Q(t) = m_Q$. Such an approach will certainly yield a Ricean distributed envelope, but will not realistically model the higher order envelope statistics for a particular scattering environment. A better approach has been suggested by Aulin [16], where $p(\theta)$ is defined in (2.36) and shown in Fig. 2.5. In this case, the means $m_I(t)$ and $m_Q(t)$ corresponding to the in phase and quadrature components of the LoS signal are given by

$$m_I(t) \;=\; s \cdot \cos(2\pi f_m \cos\theta_0 t + \phi_0) \qquad (2.47)$$
$$m_Q(t) \;=\; s \cdot \sin(2\pi f_m \sin\theta_0 t + \phi_0) \qquad (2.48)$$

where $f_m \cos\theta_0$ and ϕ_0 are the Doppler shift and random phase offset associated with the LoS or specular component, respectively.

The **Rice factor,** K, is defined as the ratio of the specular power s^2 to scattered power $2b_0$, i.e., $K = s^2/2b_0$. When $K = 0$ the channel exhibits Rayleigh fading, and when $K = \infty$ the channel does not exhibit any fading at all. The envelope distribution can be rewritten in terms of the Rice factor and the average envelope power $E[\alpha^2] = \Omega_p = s^2 + 2b_0$ by first noting that

$$s^2 = \frac{K\Omega_p}{K+1}, \qquad 2b_0 = \frac{\Omega_p}{K+1} \;. \qquad (2.49)$$

It then follows

$$p_\alpha(x) = \frac{2x(K+1)}{\Omega_p} \exp\left\{-K - \frac{(K+1)x^2}{\Omega_p}\right\} I_o\left(2x\sqrt{\frac{K(K+1)}{\Omega_p}}\right), \quad x \geq 0$$
$$(2.50)$$

Fig. 2.8 shows the Rice pdf for several values of K. The curve for $K = 0$ is the Rayleigh pdf.
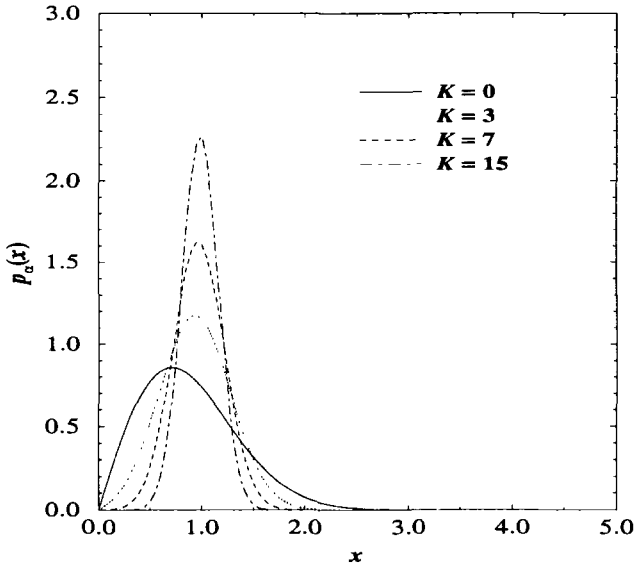
Figure 2.8.   The Rice pdf for several values of K with $\Omega_p = 1$.

The squared-envelope has the following non-central chi-square distribution with two degrees of freedom

$$p_{\alpha^2}(x) = \frac{(K+1)}{\Omega_p} \exp\left\{-K - \frac{(K+1)x}{\Omega_p}\right\} I_0\left(2\sqrt{\frac{K(K+1)x}{\Omega_p}}\right), \quad x \geq 0 \tag{2.51}$$

1.2.3   NAKAGAMI FADING

The Nakagami distribution was introduced by Nakagami in the early 1940's to characterize rapid fading in long distance HF channels [243]. The Nakagami distribution was selected to fit empirical data, and is known to provide a closer match to some experimental data than either the Rayleigh, Ricean, or log-normal distributions [47].

The Nakagami distribution describes the magnitude of the received envelope by the distribution

$$p_\alpha(x) = \frac{2m^m x^{2m-1}}{\Gamma(m)\Omega_p^m} \exp\left\{-\frac{mx^2}{\Omega_p}\right\} \quad m \geq \frac{1}{2} \tag{2.52}$$

where $\Omega_p = E[\alpha^2]$. Fig. 2.9 shows the Nakagami distribution for several values of m. Beyond its empirical justification, the Nakagami distribution is often used
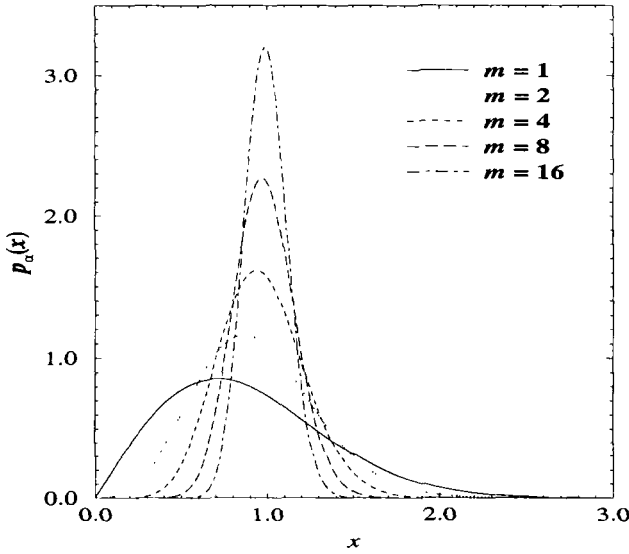
Figure 2.9. The Nakagami pdf for several values of m with $\Omega_p = 1$.

for the following reasons. First, the Nakagami distribution can model fading conditions that are either more or less severe than Rayleigh fading. When $m = 1$, the Nakagami distribution becomes the Rayleigh distribution, when $m = 1/2$ it becomes a one-sided Gaussian distribution, and when $m \to \infty$ the distribution becomes an impulse (no fading). Second, the Rice distribution can be closely approximated by using the following relation between the Rice factor K and the Nakagami shape factor m [243];

$$K = \frac{\sqrt{m^2 - m}}{m - \sqrt{m^2 - m}} \qquad m > 1 \qquad (2.53)$$

$$m = \frac{(K+1)^2}{(2K+1)} \, . \qquad (2.54)$$

Since the Rice distribution contains a Bessel function while the Nakagami distribution does not, the Nakagami distribution often leads to convenient closed form analytical expressions that are otherwise unattainable.

The squared-envelope has the Gamma density

$$p_{\alpha^2}(x) = \left(\frac{m}{\Omega_p}\right)^m \frac{x^{m-1}}{\Gamma(m)} \exp\left\{-\frac{mx}{\Omega_p}\right\} \, . \qquad (2.55)$$

By using the relationship between the K factor and the shape factor m in (2.53), the cumulative distribution function (cdf), $F_{\alpha^2}(x) = P(\alpha^2 \leq x)$ of the
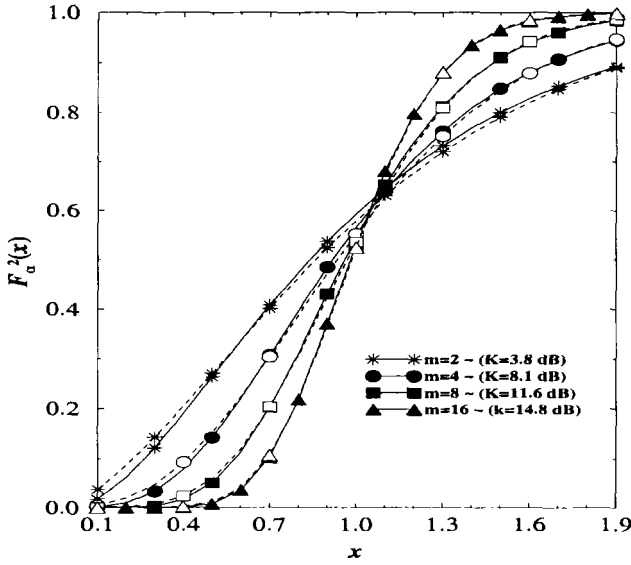
Figure 2.10. Comparison of the cdfs for the squared-envelope with Ricean and Nakagami fading.

squared-envelope with Nakagami and Ricean fading is plotted in Fig. 2.10. It is apparent from Fig. 2.10 that a Gamma distribution can approximate a non-central chi-square distribution to a reasonable degree. However, the reader is cautioned that the tails of the pdf are often the most important. Fig. 2.10 does not show how well tails of a Ricean pdf are approximated by a Nakagami pdf.

1.2.4   ENVELOPE PHASE

The phase of the received complex envelope $g(t) = g_I(t) + jg_Q(t)$ is

$$\phi(t) = \text{Tan}^{-1}\left(\frac{g_Q(t)}{g_I(t)}\right) \quad . \tag{2.56}$$

For Rayleigh fading, $g_I(t)$ and $g_Q(t)$ are independent identically distributed zero-mean Gaussian random variables at any time t_1. It follows (see Appendix A.3.2) that the phase at time t_1 is uniformly distributed over the interval $[-\pi, \pi]$, i.e.,

$$p_{\phi(t)}(x) = \frac{1}{2\pi} \quad -\pi \leq x \leq \pi \quad . \tag{2.57}$$

For Ricean fading channels, the phase is not uniformly distributed and takes on a more complicated form.

1.3 ENVELOPE CORRELATION AND SPECTRA

The autocorrelation of the envelope $\alpha(t) = |r(t)|$ of a complex Gaussian random process can be expressed in terms of the hypergeometric function $F[\cdot, \cdot; \cdot, \cdot]$ as [78]

$$
\begin{aligned}
\phi_{\alpha\alpha}(\tau) &= \mathrm{E}[\alpha(t)\,\alpha(t+\tau)] \\
&= \frac{\pi}{2}|\phi_{gg}(0)|F\left[-\frac{1}{2}, -\frac{1}{2}; 1, \frac{|\phi_{gg}(\tau)|^2}{|\phi_{gg}(0)|^2}\right]
\end{aligned}
\tag{2.58}
$$

where

$$
|\phi_{gg}(\tau)|^2 = \phi_{g_I g_I}^2(\tau) + \phi_{g_I g_Q}^2(\tau) \ .
\tag{2.59}
$$

Note that $|\phi_{gg}(0)|^2 = \phi_{g_I g_I}^2(0)$, since $\phi_{g_I g_Q}^2(0) = 0$.

The above expression is analytically cumbersome, but fortunately a useful approximation can be obtained by expanding the hypergeometric function into the following infinite series:

$$
F\left[-\frac{1}{2}, -\frac{1}{2}; 1, x\right] = 1 + \frac{1}{4}x + \frac{1}{64}x^2 + \cdots
\tag{2.60}
$$

Neglecting the terms beyond second order yields the approximation

$$
\phi_{\alpha\alpha}(\tau) \doteq \frac{\pi}{2}|\phi_{gg}(0)|\left[1 + \frac{1}{4}\frac{|\phi_{gg}(\tau)|^2}{|\phi_{gg}(0)|^2}\right] \ .
\tag{2.61}
$$

At $\tau = 0$, the approximation gives $\phi_{\alpha\alpha}(0) = 5\pi\Omega_p/16$, whereas the true value is $\phi_{\alpha\alpha}(0) = \Omega_p$. Hence, the relative error in the signal power is only 1.86%, leading us to believe that the approximation is probably very good.

The psd of the received envelope can be obtained by taking the Fourier transform of $\phi_{\alpha\alpha}(\tau)$. The psd will include a discrete spectral component at $f = 0$, due to the dc component of the received envelope. Since we are primarily interested in the continuous portion of the psd, the autocovariance function $\mu_{\alpha\alpha}(\tau)$ is of interest, where

$$
\begin{aligned}
\mu_{\alpha\alpha}(\tau) &= \mathrm{E}[\alpha(t)\alpha(t+\tau)] - \mathrm{E}[\alpha(t)]\mathrm{E}[\alpha(t+\tau)] \\
&= \frac{\pi}{2}|\phi_{gg}(0)|\left[1 + \frac{1}{4}\frac{|\phi_{gg}(\tau)|^2}{|\phi_{gg}(0)|^2}\right] - \frac{\pi}{2}|\phi_{gg}(0)| \\
&= \frac{\pi}{8|\phi_{gg}(0)|}|\phi_{gg}(\tau)|^2 \ .
\end{aligned}
\tag{2.62}
$$

With 2-D isotropic scattering $|\phi_{gg}(\tau)|^2 = \phi_{g_I g_I}^2(\tau)$ and, therefore,

$$
\mu_{\alpha\alpha}(\tau) = \frac{\pi\Omega_p}{16}J_0^2(2\pi f_m \tau) \ .
\tag{2.63}
$$
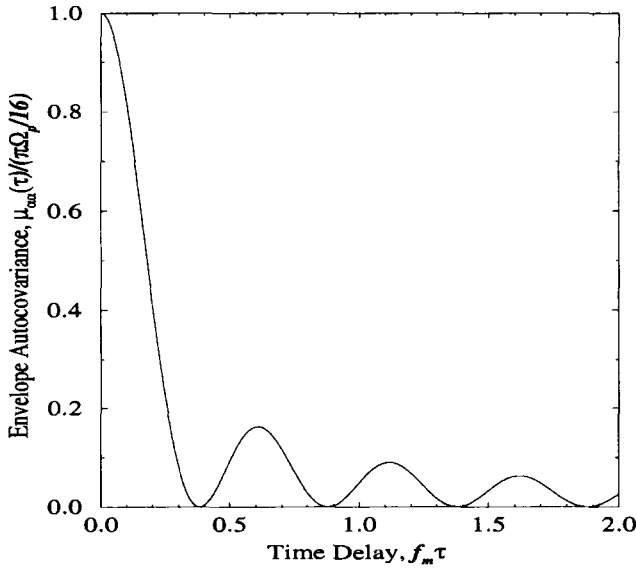
Figure 2.11.   Envelope autocovariance against the time delay $f_m \tau$ for a 2-D isotropic scattering channel.

Fig. 2.11 plots the normalized envelope auto-covariance $\mu_{\alpha\alpha}(\tau)/(\pi\Omega_p/16)$ against the normalized time delay $f_m\tau$ for the case of 2-D isotropic scattering.

The Fourier transform of $\mu_{\alpha\alpha}(\tau)$ can be calculated by using the identities $|\phi_{gg}(\tau)|^2 = \phi_{gg}(\tau)\phi_{gg}^*(\tau)$ and $\phi_{gg}(\tau) = \phi_{gg}^*(-\tau)$ to write

$$
\begin{aligned}
S_{\alpha\alpha}(f) &= \frac{\pi}{8|\phi_{gg}(0)|} S_{gg}(f) * S_{gg}(f) \\
&= \frac{\pi}{8|\phi_{gg}(0)|} \int_{-\infty}^{\infty} S_{gg}(x) S_{gg}(x-f)\, dx \\
&= \frac{\pi}{8|\phi_{gg}(0)|} \int_{-f_m}^{f_m - |f|} S_{gg}(x) S_{gg}(x + |f|)\, dx \quad 0 \le |f| \le 2f_m \;.
\end{aligned}
$$

$$(2.64)$$

Note that $S_{\alpha\alpha}(f)$ is always real, positive, and even. It is centered about $f = 0$ with a spectral width of $4f_m$, where f_m is the maximum Doppler frequency. To proceed further, we need to specify $S_{gg}(f)$. With 2-D isotropic scattering $\phi_{gg}(\tau) = \phi_{g_I g_I}(\tau)$ so that $S_{gg}(f) = S_{g_I g_I}(f)$, where $S_{g_I g_I}(f)$ is given by (2.23). The result from evaluating (2.64) is (see Problem 2.4)

$$
S_{\alpha\alpha}(f) = \frac{\Omega_p}{16\pi} \frac{1}{f_m} K\left(\sqrt{1 - \left(\frac{f}{2f_m}\right)^2} \right) \quad 0 \le |f| \le 2f_m \quad (2.65)
$$
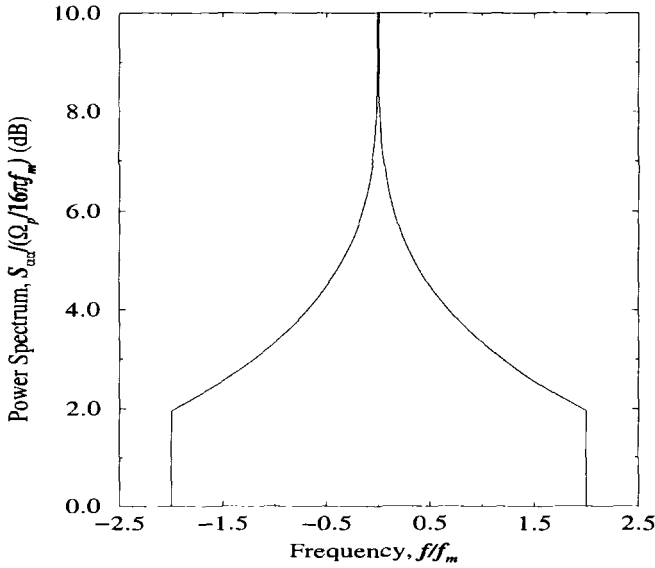
Figure 2.12. Continuous portion of the envelope psd against the normalized frequency f/f_m for a 2-D isotropic scattering channel.

where $K(\cdot)$ is the complete elliptic integral of the first kind, defined by

$$K(\gamma) = \int_0^1 \frac{dx}{\sqrt{(1 - x^2)(1 - \gamma^2 x^2)}} \quad . \tag{2.66}$$

The normalized psd $S_{\alpha\alpha}(f)/(\Omega_p/16\pi f_m)$ is plotted against the frequency f/f_m in Fig. 2.12. The psd of the complex envelope for a non-isotropic scattering channel can be obtained with some minor modifications to the above development. For example, consider the particular scattering environment shown in Fig. 2.5. In this case, the psd of $g(t)$ can be obtained from (2.24), (2.37), and (2.38) as (see Problem 2.5)

$$S_{gg}(f) = \frac{1}{K+1} \frac{\Omega_p}{2\pi f_m} \frac{1}{\sqrt{1 - (f/f_m)^2}} + \frac{K}{K+1} \frac{\Omega_p}{2} \delta(f - f_m \cos\theta_0)$$

$$|f| \leq f_m \tag{2.67}$$
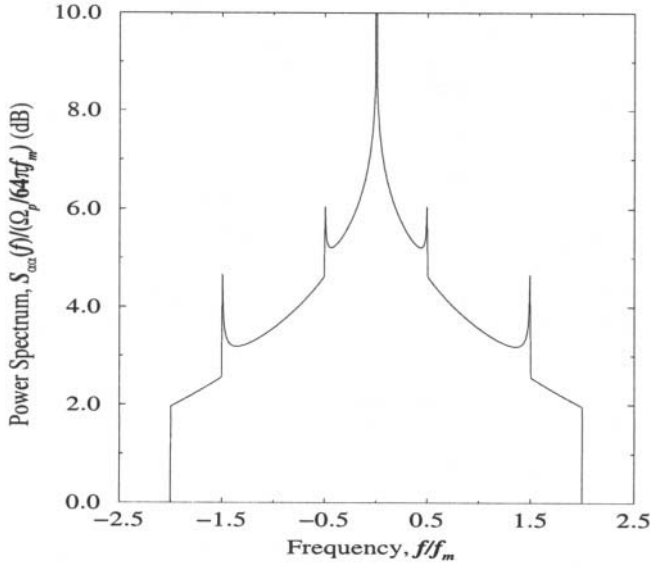
Figure 2.13. Continuous portion of the envelope psd against the normalized frequency f/f_m for the scattering environment shown in Fig. 2.5; $K = 10$ and $\theta_0 = \pi/3$.

where K is the Rice factor. Note that the psd of the received complex envelope $g(t)$ is asymmetrical. To obtain the psd of the received envelope $\alpha(t)$, we substitute (2.67) into (2.64) to obtain (see Problem 2.5)

$$S_{\alpha\alpha}(f) = \left(\frac{1}{K+1}\right)^2 \frac{\Omega_p}{16\pi f_m} \left\{ K \left(\sqrt{1 - \left(\frac{f}{2f_m}\right)^2} \right) \right. \tag{2.68}$$

$$+ \frac{K\pi}{\sqrt{1 - (f/f_m + \cos\theta_0)^2}}$$

$$\left. + \frac{K\pi}{\sqrt{1 - (f/f_m - \cos\theta_0)^2}} + K^2\pi^2 f_m \delta(f) \right\} .$$

Fig. 2.13 plots the continuous part of the normalized envelope psd $S_{\alpha\alpha}(f)/(\Omega_p/(K+1)^2 16\pi f_m)$ against the normalized frequency f/f_m.

1.3.1 SQUARED-ENVELOPE CORRELATION AND SPECTRA

The autocorrelation of the squared-envelope is

$$\phi_{\alpha^2\alpha^2}(\tau) = E[\alpha^2(t)\alpha^2(t+\tau)] \ . \tag{2.69}$$

Since $\alpha^2(t) = g_I^2(t) + g_Q^2(t)$, it follows that

$$\begin{aligned}
\phi_{\alpha^2\alpha^2}(\tau) &= E[g_I^2(t)g_I^2(t+\tau)] + E[g_Q^2(t)g_Q^2(t+\tau)] \\
&\quad + E[g_I^2(t)g_Q^2(t+\tau)] + E[g_Q^2(t)g_I^2(t+\tau)] \ . \tag{2.70}
\end{aligned}$$

First consider the case where the propagation environment is such that $g_I(t)$ and $g_Q(t)$ have zero mean. Then the squared-envelope autocorrelation is (see Problem 2.6)

$$\begin{aligned}
\phi_{\alpha^2\alpha^2}(\tau) &= 4\phi_{g_I g_I}^2(0) + 4\phi_{g_I g_I}^2(\tau) + 4\phi_{g_I g_Q}^2(\tau) \\
&= 4\phi_{g_I g_I}^2(0) + 4|\phi_{gg}(\tau)|^2 \tag{2.71}
\end{aligned}$$

Finally, the squared-envelope autocovariance is

$$\begin{aligned}
\mu_{\alpha^2\alpha^2}(\tau) &= \phi_{\alpha^2\alpha^2}(\tau) - E^2[\alpha^2(t)] \\
&= 4|\phi_{gg}(\tau)|^2 \ . \tag{2.72}
\end{aligned}$$

With isotropic scattering the above expression reduces to

$$\mu_{\alpha^2\alpha^2}(\tau) = \Omega_p^2 J_0^2(2\pi f_m\tau) \ . \tag{2.73}$$

By comparing (2.62) and (2.72), we observe that the *approximate* autocorrelation of the envelope and the *exact* autocorrelation of the squared-envelope are identical, except for a multiplicative constant. If the propagation environment is characterized by a specular or LoS component (e.g., Ricean fading), then $g_I(t)$ and $g_Q(t)$ have non-zero means and the autocovariance of the squared-envelope assumes a more complicated form. Let

$$g_I(t) = \hat{g}_I(t) + m_I(t) \tag{2.74}$$
$$g_Q(t) = \hat{g}_Q(t) + m_Q(t) \tag{2.75}$$

where $m_I(t)$ and $m_Q(t)$ are the means of $g_I(t)$ and $g_Q(t)$, respectively. From Problem 2.7,

$$\phi_{\alpha^2\alpha^2}(\tau) = 4\phi_{\hat{g}_I\hat{g}_I}^2(0) + 4\phi_{\hat{g}_I\hat{g}_I}^2(\tau) + 4\phi_{\hat{g}_I\hat{g}_Q}^2(\tau) \tag{2.76}$$

$$+ 4\left(m_I^2 + m_Q^2\right)\left(\phi_{\hat{g}_I\hat{g}_I}(0) + \phi_{\hat{g}_I\hat{g}_I}(\tau)\right) + (m_I^2 + m_Q^2)^2$$

where

$$m_I^2 = m_I(t)m_I(t+\tau) \tag{2.77}$$
$$m_Q^2 = m_Q(t)m_Q(t+\tau) \tag{2.78}$$

The squared-envelope autocovariance is

$$\mu_{\alpha^2\alpha^2}(\tau) = 4\,|\phi_{\hat{g}\hat{g}}(\tau)|^2 + 4\left(m_I^2 + m_Q^2\right)\phi_{\hat{g}_I\hat{g}_I}(\tau) \qquad (2.79)$$

Consider the scattering environment shown in Fig. 2.5. The corresponding correlation functions $\phi_{g_Ig_I}(\tau)$ and $\phi_{g_Ig_Q}(\tau)$ are given by (2.37) and (2.38), respectively, and the means $m_I(t)$ and $m_Q(t)$ are defined in (2.47) and (2.48). It can be shown that

$$\phi_{\hat{g}_I\hat{g}_I}(\tau) \;=\; \frac{1}{K+1}\frac{\Omega_p}{2}J_0(2\pi f_m\tau) \qquad (2.80)$$

$$\phi_{\hat{g}_I\hat{g}_Q}(\tau) \;=\; 0 \qquad (2.81)$$

and

$$m_I^2 + m_Q^2 = s^2\cos(2\pi f_m\cos\theta_0\tau) = \frac{K\Omega_p}{K+1}\cos(2\pi f_m\cos\theta_0\tau) \qquad (2.82)$$

where K is the Rice factor and θ_0 is the angle that the specular component makes with the MS direction of motion. Using these results in (2.79) gives

$$\mu_{\alpha^2\alpha^2}(\tau) = \left(\frac{\Omega_p}{K+1}\right)^2 J_0(2\pi f_m\tau)\left(J_0(2\pi f_m\tau) + 2K\cos(2\pi f_m\tau\cos\theta_0)\right) \qquad (2.83)$$

The corresponding normalized squared-envelope autocovariance

$$\left(\frac{K+1}{\Omega_p}\right)^2\frac{1}{1+2K}\cdot\mu_{\alpha^2\alpha^2}(\tau) \qquad (2.84)$$

is plotted in Fig. 2.14 as a function of the normalized time delay $f_m\tau$, for various values of K and θ_0.

1.4   LEVEL CROSSING RATES AND FADE DURATIONS

Two important second order statistics associated with envelope fading are the **level crossing rate** (how often the envelope crosses a specified level) and the **average fade duration** (how long the envelope remains below a specified level). These quantities are second order statistics, because they are not only affected by the scattering environment but also by the velocity of the MS. For the case of Ricean (and Rayleigh) fading, closed form expressions can be derived for these parameters.

1.4.1   ENVELOPE LEVEL CROSSING RATE

The envelope level crossing rate at a specified level R, L_R, is defined as the rate at which the envelope crosses level R in the positive (or negative) going direction. Obtaining the level crossing rate requires the joint pdf, $p(\alpha,\dot{\alpha})$, of the envelope level $\alpha = |r|$ and the envelope slope $\dot{\alpha} = |\dot{r}|$. In terms of the joint pdf $p(\alpha,\dot{\alpha})$, the expected amount of time spent in the interval $(R, R+d\alpha)$ for a given envelope slope $\dot{\alpha}$ and time duration dt is
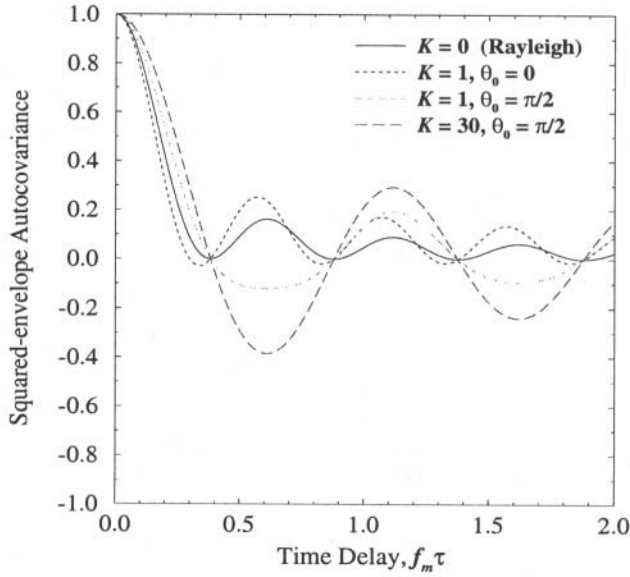
Figure 2.14. Squared-envelope auto-covariance against the normalized time delay $f_m\tau$ for the scattering environment shown in Fig. 2.5.

$$p(R,\dot\alpha)d\alpha d\dot\alpha dt \ . \tag{2.85}$$

The time required to cross the level α once for a given envelope slope $\dot\alpha$, in the interval $(R, R + d\alpha)$ is

$$d\alpha/\dot\alpha \ . \tag{2.86}$$

The ratio of these two quantities is the expected number of crossings of the envelope α within the interval $(R, R + d\alpha)$ for a given envelope slope $\dot\alpha$ and time duration dt, i.e.,

$$\dot\alpha p(R,\dot\alpha)d\dot\alpha dt \ . \tag{2.87}$$

The expected number of crossings of the envelope level R for a given envelope slope $\dot\alpha$ in a time interval of duration T is

$$\int_0^T \dot\alpha p(R,\dot\alpha)d\dot\alpha dt = \dot\alpha p(R,\dot\alpha)d\dot\alpha T \ . \tag{2.88}$$

The expected number of crossings of the envelope level R with a positive slope is

$$N_R = T\int_0^\infty \dot\alpha p(R,\dot\alpha)d\dot\alpha \ . \tag{2.89}$$

Finally, the expected number of crossings of the envelope level R per second, or the level crossing rate, is

$$L_R = \int_0^\infty \dot\alpha p(R,\dot\alpha)d\dot\alpha \ . \tag{2.90}$$

This is a general result that applies to any random process.

Rice has derived the joint pdf $p(\alpha, \dot{\alpha})$ for a sine wave plus narrow-band Gaussian noise. For this case [282]

$$p(\alpha, \dot{\alpha}) = \frac{\alpha(2\pi)^{-3/2}}{\sqrt{Bb_0}} \int_{-\pi}^{\pi} d\theta \qquad (2.91)$$

$$\times \ \exp\left\{-\frac{1}{2Bb_0}\left[B\left(\alpha^2 - 2\alpha s \cos\theta + s^2\right) + (b_0\dot{\alpha} + b_1 s \sin\theta)^2\right]\right\}$$

where s is the non-centrality parameter in the Rice distribution, and $B = b_0 b_2 - b_1^2$, where b_0, b_1, and b_2 are constants that are derived from the psd of the narrow-band noise. For the scattering environment described by (2.36) and Fig. 2.5, the sine wave corresponds to the specular component arriving at angle θ_0, while the narrow-band noise is due to the scatter component with AoA distribution $\hat{p}(\theta) = 1/(2\pi)$. Note that Rice's result in (2.91) is general enough to apply to scattering environments with other $\hat{p}(\theta)$ as well.

Suppose that the frequency of the specular or LoS component is $f_s = f_c + f_q$, where $0 \le |f_q| \le f_m$. In this case [173]

$$b_n = (2\pi)^n \int_{-f_m}^{f_m} S_{gg}^c(f)(f - f_q)^n df \qquad (2.92)$$

$$= (2\pi)^n b_0 \int_0^{2\pi} \hat{p}(\theta) \left(f_m \cos\theta - f_q\right)^n d\theta \qquad (2.93)$$

where $\hat{p}(\theta)$ is the continuous AoA distribution of the *scatter* component and $S_{gg}^c(f)$ is the corresponding continuous portion of the psd of the received complex envelope. Equivalence between (2.92) and (2.93) can be established by using (2.28) and (2.33). Note that $S_{gg}^c(f)$ is given by the Fourier transform of

$$\phi_{gg}^c(\tau) = \phi_{g_I g_I}^c(\tau) + j\phi_{g_I g_Q}^c(\tau) \qquad (2.94)$$

where

$$\phi_{g_I g_I}(\tau) = \frac{\Omega_p}{2} \int_0^{2\pi} \cos(2\pi f_m \tau \cos\theta)\hat{p}(\theta)d\theta \qquad (2.95)$$

$$\phi_{g_I g_Q}(\tau) = \frac{\Omega_p}{2} \int_0^{2\pi} \sin(2\pi f_m \tau \cos\theta)\hat{p}(\theta)d\theta \qquad (2.96)$$

In some cases, the psd $S_{gg}^c(f)$ is symmetrical about the sine wave frequency f_s. This condition occurs, for example, when $f_s = f_c$ and there is 2-D isotropic scattering. In this case, $b_n = 0$ for all odd values of n (and in particular $b_1 = 0$) so that (2.91) reduces to the convenient product form

$$p(\alpha, \dot{\alpha}) = \sqrt{\frac{1}{2\pi b_2}} \exp\left\{-\frac{\dot{\alpha}^2}{2b_2}\right\} \cdot \frac{\alpha}{b_0} \exp\left\{-\frac{(\alpha^2 + s^2)}{2b_0}\right\} I_0\left(\frac{\alpha s}{b_0}\right)$$

$$= p(\dot{\alpha}) \cdot p(\alpha) . \qquad (2.97)$$

Since $p(\alpha, \dot{\alpha}) = p(\dot{\alpha}) \cdot p(\alpha)$ in (2.97), it follows that α and $\dot{\alpha}$ are independent.

When $f_s = f_c$ and $\hat{p}(\theta) = 1/(2\pi)$, a closed form expression can be obtained for the envelope level crossing rate. Substituting (2.35) into (2.92) gives

$$b_n = \begin{cases} b_0(2\pi f_m)^n \frac{1\cdot3\cdot5\,\cdots\,(n-1)}{2\cdot4\cdot6\,\cdots\,n} & n \text{ even} \\ 0 & n \text{ odd} \end{cases} \quad . \tag{2.98}$$

Therefore, $b_1 = 0$ and $b_2 = b_0(2\pi f_m)^2/2$, where $2b_0$ is the power in the scatter component of the received band-pass signal given by (2.49). Substituting the joint density in (2.97) into (2.90) gives the envelope level crossing rate

$$L_R = \sqrt{2\pi(K+1)} f_m \rho e^{-K-(K+1)\rho^2} I_0\left(2\rho\sqrt{K(K+1)}\right) \tag{2.99}$$

where

$$\rho = \frac{R}{\sqrt{\Omega_p}} = \frac{R}{R_{\text{rms}}} \tag{2.100}$$

and $R_{\text{rms}} \triangleq \sqrt{\Omega_p}$ is the *rms* envelope level. For Rayleigh fading ($K = 0$) and isotropic scattering, the above expression simplifies to

$$L_R = \sqrt{2\pi} f_m \rho e^{-\rho^2} \quad . \tag{2.101}$$

The normalized envelope level crossing rate L_R/f_m is plotted in Fig. 2.15 as a function of ρ and K. The maximum LCR can be found by taking the derivative of (2.99) with respect to ρ and solving

$$I_0\left(2\rho\sqrt{K(K+1)}\right)\left(1 - 2(K+1)\rho^2\right)$$
$$+ 2\rho\sqrt{K(K+1)} I_1\left(2\rho\sqrt{K(K+1)}\right) = 0 \tag{2.102}$$

for ρ as a function of K. Fig. 2.16 plots the maximum envelope level crossing rate as a function of K. Finally, we note that the envelope level crossing rate around $\rho = 0$ dB is nearly independent of K. This attractive property will be exploited in Chapter 10 when we use the envelope level crossing rate to estimate the MS velocity. The simulation results in Fig. 2.15 were obtained with a fading simulator that will be described in Section 3..
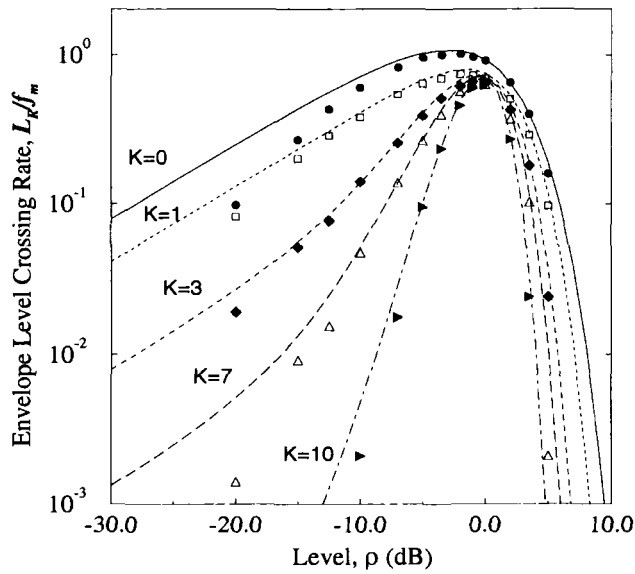
Figure 2.15. Envelope level crossing rate for the scattering environment shown in Fig. 2.5. Lines denote theoretical results, while points denote simulation results.
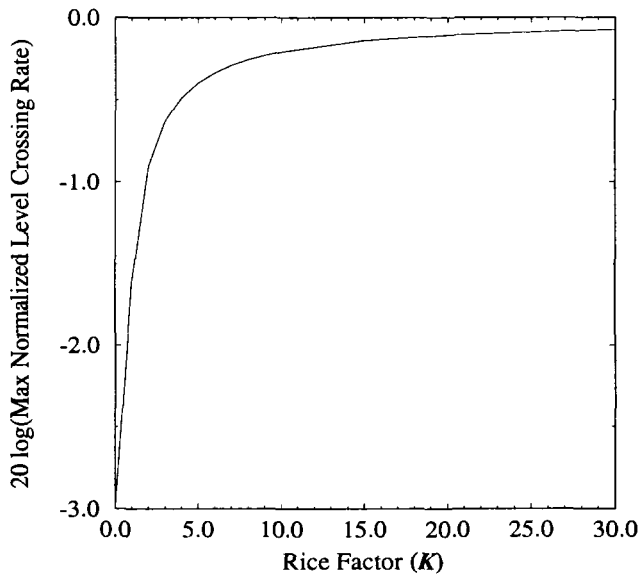
Figure 2.16. Maximum normalized envelope level crossing rate L_R/f_m for the scattering environment shown in Fig. 2.5.

1.4.2   ZERO CROSSING RATE

Recall that received complex envelope $g(t) = g_I(t) + g_Q(t)$ is a complex Gaussian random process. If the channel is characterized by a specular component then $g_I(t)$ and $g_Q(t)$ have mean values $m_I(t)$ and $m_Q(t)$, respectively. Here we are interested in the **zero crossing rate** of the zero-mean Gaussian random processes $g_I(t) - m_I(t)$ and $g_Q(t) - m_Q(t)$. Rice [282] has derived this zero crossing rate as

$$L_Z = \frac{1}{\pi}\sqrt{\frac{b_2}{b_0}} = 2f_m\sqrt{\int_0^{2\pi}\hat{p}(\theta)\cos^2\theta d\theta} \ . \tag{2.103}$$

When the scatter component is due to 2-D isotropic scattering, the zero crossing rate is

$$L_Z = \sqrt{2}f_m \ . \tag{2.104}$$

1.4.3   AVERAGE ENVELOPE FADE DURATION

Another quantity of interest is the average duration that the envelope level remains below a specified level R. Although the pdf of the envelope fade duration is unknown, the average fade duration can be calculated. Consider a very long time interval of length T and let t_i be the duration of the ith fade below the level R. The probability of the received envelope level being less than R is

$$P(\alpha \le R) = \frac{1}{T}\sum_i t_i \ . \tag{2.105}$$

The average envelope fade duration is equal to

$$\bar{t} = \frac{1}{TL_R}\sum_i t_i = \frac{P(\alpha \le R)}{L_R} \ . \tag{2.106}$$

If the envelope has the Rice distribution in (2.45), then

$$P(\alpha \le R) = \int_0^R p(\alpha)d\alpha = 1 - Q\left(\sqrt{2K}, \sqrt{2(K+1)\rho^2}\right) \tag{2.107}$$

where $Q(a, b)$ is the Marcum Q function. Therefore,

$$\bar{t} = \frac{1 - Q\left(\sqrt{2K}, \sqrt{2(K+1)\rho^2}\right)}{\sqrt{2\pi(K+1)}f_m\rho e^{-K-(K+1)\rho^2}I_0\left(2\rho\sqrt{K(K+1)}\right)} \ . \tag{2.108}$$

If the envelope is Rayleigh distributed, then

$$P(\alpha \le R) = \int_0^R p(\alpha)d\alpha = 1 - e^{-\rho^2} \tag{2.109}$$

and, therefore,

$$\bar{t} = \frac{e^{\rho^2} - 1}{\rho f_m\sqrt{2\pi}} \ . \tag{2.110}$$
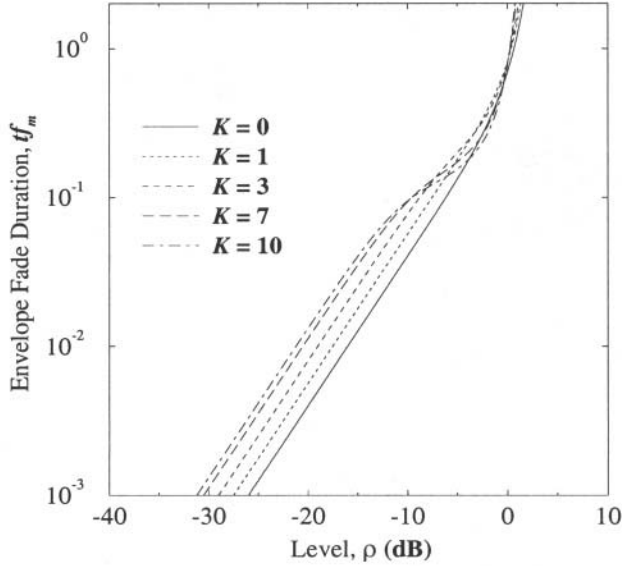
Figure 2.17. Average envelope fade duration for the scattering environment shown in Fig. 2.5.

The normalized average envelope fade duration $\bar{t}f_m$ is plotted in Fig. 2.17 as a function of ρ.

Note that the level crossing rate, zero crossing rate, and the average fade duration all depend on the velocity of the MS ($f_m = v/\lambda_c$). Very deep fades tend to occur infrequently and do not last very long. For example, at 60 mi/hr and 900 MHz, the maximum Doppler frequency is $f_m = 88$ Hz. Therefore, with isotropic scattering and Rayleigh fading ($K = 0$) there are $L_R = 81$ fades/s at $\rho = 0$ dB with an average fade duration of 7.8 ms. However, at $\rho = -20$ dB there are only 2.2 fades/s with an average fade duration of 45 μs. Observe from Fig. 2.15 that the fades are shallower when the Rice factor, K, is larger. Furthermore, we see from Fig. 2.17 that the average fade duration tends to be larger with larger Rice factors.

1.5 SPATIAL CORRELATIONS

Many mobile radio systems employ antenna diversity, where spatially separated antennas are used to provide multiple faded replicas of the same information bearing signal. A fundamental question that arises is the antenna separation needed to provide uncorrelated antenna diversity branches. This question can be answered by using our previously derived results along with the distance-time transformation $\ell = v\tau$, where v is the MS velocity. This transformation results in $f_m\tau = \ell/\lambda_c$. For the case of isotropic scattering (2.21) and (2.63) become, respectively
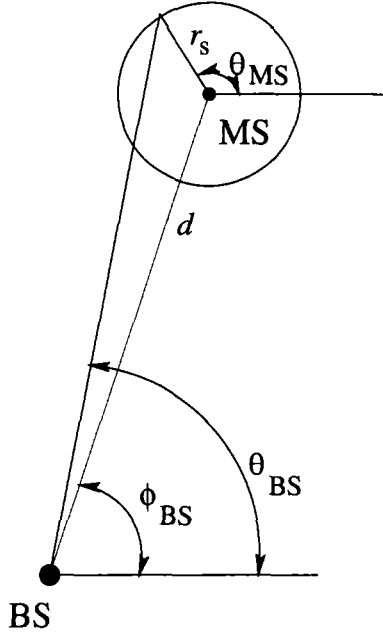
Figure 2.18.   The scattering model for propagation in macrocells. The MS is moving parallel to the x-axis with velocity $-v$, is located a distance d from the BS, and is surrounded by a scattering ring of radius r_s [173, 130].

$$p(\theta_{\rm BS}) = \begin{cases} K_A \left[(\tfrac{r_s}{d})^2 - (\phi_{\rm BS} - \theta_{\rm BS})^2 \right]^{-1/2} & ,\alpha \le \theta_{\rm BS} \le \beta \\ 0 & , \text{ otherwise} \end{cases} \qquad (2.116)$$

where

$$\begin{aligned} \phi_{\rm MS} &= {\rm Tan}^{-1}(r_s/d) \\ K_A &= \frac{1}{2{\rm Sin}^{-1}(\phi_{\rm MS}(d/r_s))} \\ \alpha &= \phi_{\rm BS} - \phi_{\rm MS} \\ \beta &= \phi_{\rm BS} + \phi_{\rm BS} \end{aligned}$$

$\phi_{\rm MS}$ is the maximum AoA at the BS for a given distance d and scattering radius r_s. Note that for $d \gg r_s$, a small angle approximation can be invoked, with the result that $\phi_{\rm MS} \approx r_s/d$ and $K_A = 1/\pi$.

To proceed further, we artificially assume that the MS is stationary and the BS is moving along the x-axis in Fig 2.18 with velocity v. To obtain the envelope and squared-envelope spatial crosscovariance at the BS, we first compute $\phi_{g_I g_I}(\tau)$ and $\phi_{g_I g_Q}(\tau)$ in (2.16) and (2.20), respectively. Then compute $\phi_{gg}(\tau)$ from (2.59) and, finally, use (2.62) and (2.72) to obtain the envelope autocovariance, $\mu_{\alpha\alpha}(\tau)$ and squared-envelope autocovariance, $\mu_{\alpha^2\alpha^2}(\tau)$, respectively.
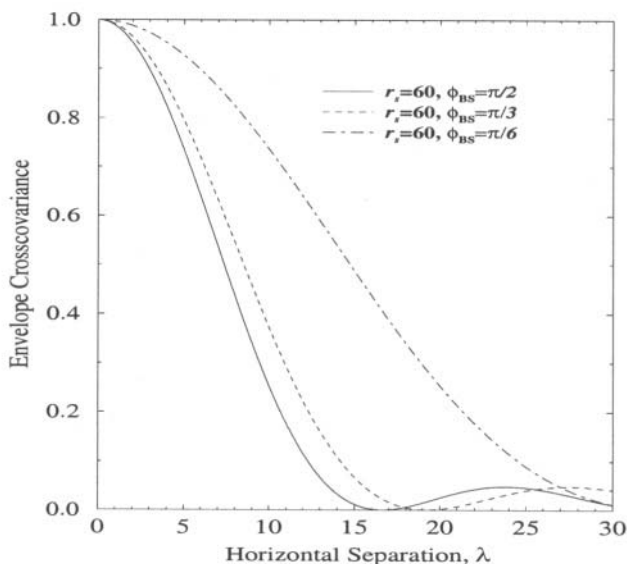
Figure 2.19. Envelope crosscovariance at the base station for $a = 60$ m and various arrival angles, ϕ_{BS}; $d = 3000$ m.

This can be accomplished by using numerical integration. Now dispense with the artificial assumption of a moving BS and fixed MS, and assume that the BS is fixed while the MS moves parallel to the x-axis with velocity $-v^2$. Finally, by using the time-distance transformation $f_m \tau = \ell / \lambda_c$, we can obtain the spatial crosscovariance functions $\mu_{\alpha\alpha}(\ell)$ and $\mu_{\alpha^2\alpha^2}(\ell)$. Fig. 2.19 plots the envelope spatial crosscovariance for $a = 60$ m and various arrival angles c ϕ_{BS}. Likewise, Fig. 2.20 plots the envelope spatial crosscovariance for $\phi_{BS} = \pi/3$ and various scattering radii. In general, we observe that a much greater spatial separation is required to achieve a given degree of envelope decorrelation at the BS as compared to the MS. Also, the correlation increases as the arrival angle and scattering radius decrease.

2. FREQUENCY-SELECTIVE MULTIPATH-FADING

To this point we have considered channel models that are appropriate for narrow-band transmission, where the inverse signal bandwidth is much greater than the time spread of the propagation path delays. For digital communication systems this means that the duration of a modulated symbol is much greater than the time spread of the propagation path delays. Under this condition all frequencies in the transmitted signal will experience the same random

---

[2]Note that the relative velocities of the MS and BS are opposite in sign.
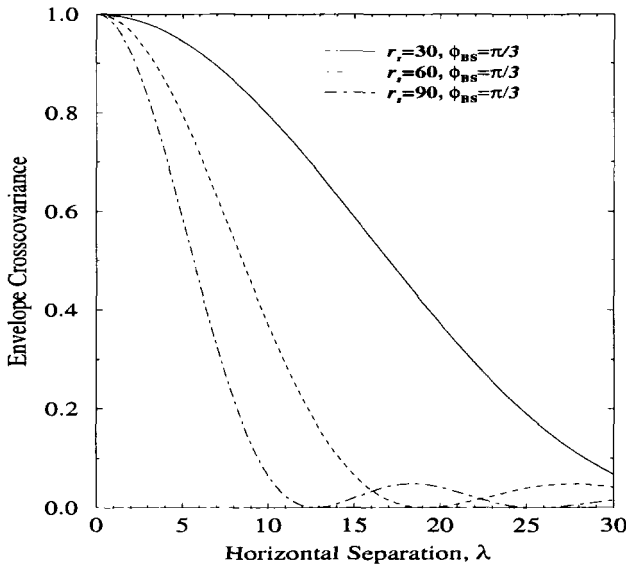
Figure 2.20.   Envelope crosscovariance at the base station for ϕ_{BS} and various scattering radii, a; $d = 3000$ m.

attenuation and phase shift due to multipath-fading. Such a channel introduces very little or no distortion into the received signal and is said to exhibit **flat fading**. If the range in the propagation path delays is large compared to the inverse signal bandwidth, then the frequency components in the transmitted signal will experience different phase shifts along the different paths. As the differential path delays become large, even closely separated frequencies in the transmitted signal can experience significantly different phase shifts. Under this condition the channel introduces amplitude and phase distortion into the message waveform. Such a channel is said to exhibit **frequency-selective fading**. The path geometry for a multipath-fading channel is shown in Fig. 2.21. Considering only single reflections, all scatterers that are associated with a particular path length are located on an ellipse with the transmitter and receiver located at the foci. Different delays correspond to different confocal ellipses. Flat fading channels have their scatterers located on ellipses corresponding to differential delays that are small compared to the duration of a modulated symbol. Frequency selective channels have strong scatterers that are located on several ellipses that correspond to differential delays that are significant compared to a symbol duration. In urban and suburban macrocellular systems, these strong scatterers usually correspond to high-rise buildings or perhaps large distant terrain features such as mountains.

    Multipath-fading channels can be modeled as time-variant linear filters,
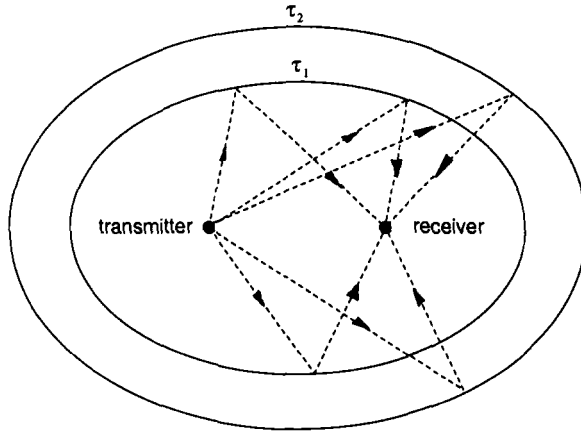
Figure 2.21. Path geometry for multipath-fading channels. Signals will arrive at the receiver antenna at the same time if they reflect off scatterers that are located on the same ellipse.

whose inputs and outputs can be described in both the time and frequency domains. This leads to four possible **transmission functions** [30]; the input delay-spread function $g(\tau, t)$, the output Doppler-spread function $H(f,v)$, the time-variant transfer function $T(f,t)$, and the delay Doppler-spread function $S(\tau, \nu)$. The complex low-pass impulse response relates the complex low-pass input and output time waveforms, $\tilde{s}(t)$ and $\tilde{r}(t)$, respectively, through the convolution

$$\tilde{r}(t) = \int_0^\infty \tilde{s}(t - \tau)g(t, \tau)d\tau \ . \tag{2.117}$$

Bello called the low-pass impulse response $g(t, \tau)$ the **input delay-spread function** [30]. In physical terms, $g(t, \tau)$ can be interpreted as the channel response at time t due to an impulse applied at time $t - \tau$. Since a physical channel cannot have an output before an input is applied $g(t, \tau) = 0$ for $\tau < 0$ and therefore the lower limit of integration in (2.117) is zero. If the convolution in (2.117) is written as a discrete sum, then

$$\tilde{r}(t) = \sum_{m=0}^{n} \tilde{s}(t - m\Delta\tau)g(t, m\Delta\tau)\Delta\tau \ . \tag{2.118}$$

This representation allows us to visualize the channel as a transversal filter with tap spacing $\Delta\tau$ and time-varying tap gains $g(t, m\Delta\tau)$ as shown in Fig. 2.22.

The second transmission function relates the input and output spectra, $\tilde{S}(f)$ and $\tilde{R}(f)$, respectively, through the integral equation

$$\tilde{R}(f) = \int_{-\infty}^\infty \tilde{S}(f - \nu)H(f - \nu, \nu)d\nu \ . \tag{2.119}$$

Bello called the function $H(f, \nu)$ the **output Doppler-spread function** [30].
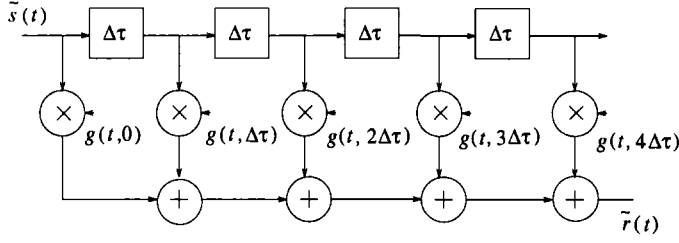
Figure 2.22. Discrete-time tapped delay line model for a multipath-fading channel, from [257].
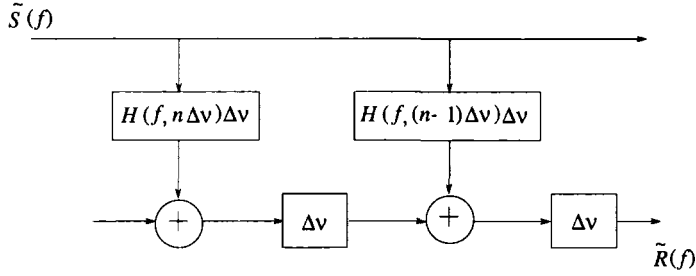
Figure 2.23. Frequency conversion model for a multipath-fading channel, from [257].

This function explicitly shows the effect of Doppler shift or spectral broadening on the output spectrum. In physical terms, the frequency-shift variable ν can be interpreted as the Doppler shift that is introduced by the channel. Once again, the integral in (2.119) can be approximated by the discrete sum

$$\tilde{R}(f) = \sum_{m=0}^{n} \tilde{S}(f - m\Delta\nu)H(f - m\Delta\nu, m\Delta\nu)\Delta\nu \ . \tag{2.120}$$

This allows the channel to be represented by a bank of filters with transfer functions $H(f, m\Delta\nu)\Delta\nu$ followed by a frequency conversion chain that produces the Doppler shifts.

The third transmission function relates the output time waveform to the input spectrum through the integral equation

$$\tilde{r}(t) = \int_{-\infty}^{\infty} \tilde{S}(f)T(f,t)e^{j2\pi ft}df \ . \tag{2.121}$$

Zadeh called the function $T(f, t)$ the **time-variant transfer function** [377].
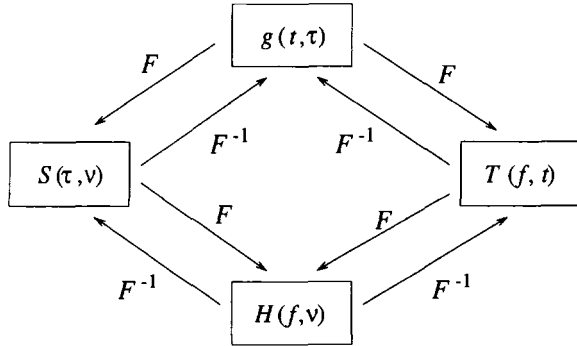
Figure 2.24. Fourier transform relations between the transmission functions, from [257].

The final description relates the input and output time waveforms through the double integral

$$\tilde{r}(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{s}(t-\tau)S(\tau,\nu)e^{-j2\pi f\tau}\,d\nu d\tau \ . \qquad (2.122)$$

The function $S(\tau,\nu)$ is called the **delay Doppler-spread function** [30], and provides a measure of the scattering amplitude of the channel in terms of the time delay τ and Doppler frequency ν. The four transmission functions are related to each other through Fourier transform pairs as shown in Fig. 2.24. In each transform pair there is always a fixed variable, so that the transform involves the other two variables.

2.1 STATISTICAL CHANNEL CORRELATION FUNCTIONS

Recall the channel impulse response $g(t,\tau) = g_I(t,\tau) + jg_Q(t,\tau)$ can be modeled as a complex Gaussian random process, where the quadrature components $g_I(t,\tau)$ and $g_Q(t,\tau)$ are correlated Gaussian random processes. Hence, all of the transmission functions defined in the last section are random processes. A thorough characterization of a channel requires knowledge of the joint pdf of all the transmission functions. Since this is rather formidable, a more reasonable approach is to obtain statistical correlation functions for the individual transmission functions. If the underlying process is Gaussian, then a complete statistical description is provided by the means and autocorrelation functions. In the following discussion, we assume zero-mean Gaussian random processes so that only the autocorrelation functions are of interest. Since there are four transmission functions, four autocorrelation functions can be defined as follows [257, 270]:
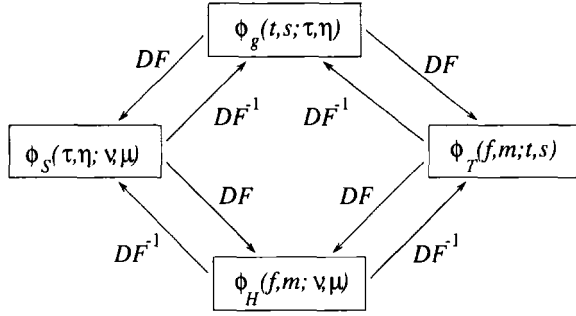
Figure 2.25. Double Fourier transform relations between the channel autocorrelation functions, from [257].

$$\phi_g(t, s; \tau, \eta) = \frac{1}{2}\mathrm{E}[g(t, \tau)g^*(s, \eta)] \qquad (2.123)$$

$$\phi_T(f, m; t, s) = \frac{1}{2}\mathrm{E}[T(f, t)T^*(m, s)] \qquad (2.124)$$

$$\phi_H(f, m; \nu, \mu) = \frac{1}{2}\mathrm{E}[H(f, \nu)H^*(m, \mu)] \qquad (2.125)$$

$$\phi_S(\tau, \eta; \nu, \mu) = \frac{1}{2}\mathrm{E}[S(\tau, \nu)S^*(\eta, \mu)] \ . \qquad (2.126)$$

These autocorrelation functions are related to each other through double Fourier transform pairs. For example,

$$\phi_S(\tau, \eta; \nu, \mu) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi_g(t, s; \tau, \eta)e^{j2\pi(\nu t - \mu s)}\,dtds \qquad (2.127)$$

$$\phi_g(t, s; \tau, \eta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi_S(\tau, \eta; \nu, \mu)e^{-j2\pi(\nu t - \mu s)}\,d\nu d\mu \qquad (2.128)$$

The complete set of such relationships is summarized in Fig. 2.25.

2.2 CLASSIFICATION OF CHANNELS

Wide sense stationary (WSS) channels have fading statistics that remain constant over short periods of time. This implies that the channel correlation functions depend on the time variables t and s only through the time difference $\Delta t = s - t$. It can be demonstrated (see Problem 2.13) that WSS channels give rise to scattering with uncorrelated Doppler shifts. This behavior suggests that the attenuations and phase shifts associated with signal components having different Doppler shifts are uncorrelated. Hence for WSS channels, the correlation functions become

$$\phi_g(t, t + \Delta t; \tau, \eta) = \phi_g(\Delta t; \tau, \eta) \tag{2.129}$$

$$\phi_T(f, m; t, t + \Delta t) = \phi_T(f, m; \Delta t) \tag{2.130}$$

$$\phi_H(f, m; \nu, \mu) = \psi_H(f, m; \nu)\delta(\nu - \mu) \tag{2.131}$$

$$\phi_S(\tau, \eta; \nu, \mu) = \psi_S(\tau, \eta; \nu)\delta(\nu - \mu) \tag{2.132}$$

where

$$\psi_H(f, m; \nu) = \int_{-\infty}^{\infty} \phi_T(f, m; \Delta t)e^{-j2\pi\nu\Delta t}d\Delta t \tag{2.133}$$

$$\psi_S(\tau, \eta; \nu) = \int_{-\infty}^{\infty} \phi_c(\Delta t; \tau, \eta)e^{-j2\pi\nu\Delta t}d\Delta t \tag{2.134}$$

are Fourier transform pairs.

Uncorrelated scattering (US) channels are characterized by an uncorrelated attenuation and phase shift with paths of different delays. Bello showed that US channels are wide sense stationary in the frequency variable so that the correlation functions depend on the frequency variables f and m only through the frequency difference $\Delta f = m - f$ [30]. Analogous to (2.131) and (2.132), the channel correlation functions can be shown (see Problem 2.14) to be singular in the time-delay variable. For US channels, the channel correlation functions become

$$\phi_g(t, s; \tau, \eta) = \psi_g(t, s; \tau)\delta(\eta - \tau) \tag{2.135}$$

$$\phi_T(f, f + \Delta f; t, s) = \phi_T(\Delta f; t, s) \tag{2.136}$$

$$\phi_H(f, f + \Delta f; \nu, \mu) = \phi_H(\Delta f; \nu, \mu) \tag{2.137}$$

$$\phi_S(\tau, \eta; \nu, \mu) = \psi_S(\tau; \nu, \mu)\delta(\eta - \tau) \tag{2.138}$$

where

$$\psi_g(t, s; \tau) = \int_{-\infty}^{\infty} \phi_T(\Delta f; t, s)e^{j2\pi\Delta f\tau}d\Delta f \tag{2.139}$$

$$\psi_S(\tau; \nu, \mu) = \int_{-\infty}^{\infty} \phi_H(\Delta f; \nu, \mu)e^{j2\pi\Delta f\tau}d\Delta f \ . \tag{2.140}$$

Wide sense stationary uncorrelated scattering (WSSUS) channels are a very special type of multipath-fading channel. These channel display uncorrelated scattering in both the time-delay and Doppler shift. Fortunately, many radio channels can be accurately modeled as WSSUS channels. For WSSUS channels, the correlation functions have singular behavior in both the time delay and Doppler shift variables, and reduce to the following simple forms:
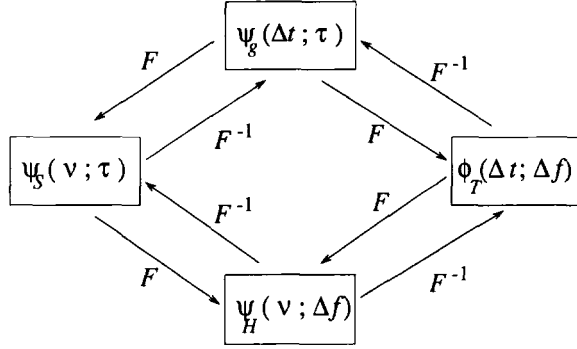
Figure 2.26. Fourier transform relations between the channel correlation functions for WSSUS channels, from [257].

$$\phi_g(t, t + \Delta t; \tau, \eta) = \psi_g(\Delta t; \tau)\delta(\eta - \tau) \qquad (2.141)$$

$$\phi_T(f, f + \Delta f; t, t + \Delta t) = \phi_T(\Delta f; \Delta t) \qquad (2.142)$$

$$\phi_H(f, f + \Delta f; \nu, \mu) = \psi_H(\Delta f; \nu)\delta(\nu - \mu) \qquad (2.143)$$

$$\phi_S(\tau, \eta; \nu, \mu) = \psi_S(\tau, \nu)\delta(\eta - \tau)\delta(\nu - \mu) . \qquad (2.144)$$

These correlation functions are related through the Fourier transform pairs shown in Fig. 2.26.

The function $\phi_g(0; \tau) \equiv \phi_g(\tau)$ is called the **multipath intensity profile** or **power delay profile** and gives the average power at the channel output as a function of the time delay τ. It can be viewed as the scattering function averaged over all Doppler shifts. A typical power delay profile is shown in Fig. 2.27. One quantity of interest is the **average delay**, defined as

$$\mu_\tau = \frac{\int_0^\infty \tau \phi_g(\tau)d\tau}{\int_0^\infty \phi_g(\tau)d\tau} . \qquad (2.145)$$

Note that the normalization $\int_0^\infty \phi_g(\tau)d\tau$ is applied because $\phi_g(\tau)$ is not a pdf. Another quantity of interest is the *rms* **delay spread**, defined as

$$\sigma_\tau = \sqrt{\frac{\int_0^\infty (\tau - \mu_\tau)^2 \phi_g(\tau)d\tau}{\int_0^\infty \phi_g(\tau)d\tau}} . \qquad (2.146)$$

There are other quantities that can also be used to describe the power delay profile. One is the width, W_x, of the middle portion of the power delay profile that contains $x\%$ of the total power in the profile. Referring to Fig. 2.27

$$W_x = \tau_3 - \tau_1 \qquad (2.147)$$

where τ_1 and τ_3 are chosen so that

$$\int_0^{\tau_1} \phi_g(\tau)d\tau = \int_{\tau_3}^\infty \phi_g(\tau)d\tau \qquad (2.148)$$
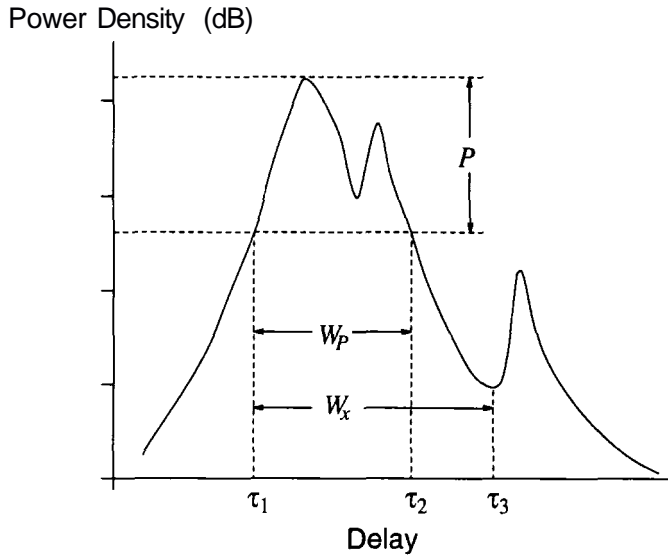
Figure 2.27. A typical power delay profile.

and

$$\int_{\tau_1}^{\tau_3} \phi_g(\tau)d\tau = x \int_0^{\infty} \phi_g(\tau)d\tau \ . \tag{2.149}$$

Another quantity is the difference in delays where the delay profile rises to a value P dB below its maximum value and where the delay profile drops to a value P dB below its maximum value for the last time. This quantity is denoted by W_P and is also illustrated in Fig. 2.27, where $W_P = \tau_2 - \tau_1$. Power delay profiles play a key role in determining whether or not an adaptive equalizer is required at the receiver. If the excess delay spread exceeds 10% to 20% of the symbol duration, then an adaptive equalizer may be required. In general, the average delay and delay spread of the channel will diminish with decreasing cell size, the reason being that the radio path lengths are shorter. While the delay spread in a typical macrocellular application may be on the order of 1 to 10 μs, the delay spreads in a typical microcellular applications are much less. Delay spreads within buildings range can anywhere from 30 to 60 ns in buildings with interior walls and little metal, to 300 ns in buildings with open plans and a significant amount of metal. The function $\phi_T(\Delta t; \Delta f)$ is called the spaced-frequency spaced-time correlation function. The function $\phi_T(0; \Delta f) \equiv \phi_T(\Delta f)$ measures the frequency correlation of the channel. The **coherence bandwidth**, B_c, of the channel is defined as the smallest value of Δf for which $\phi_T(\Delta f)$ equals some suitable correlation coefficient such as 0.5. As a result of the Fourier transform relation between $\phi_g(\tau)$ and $\phi_T(\Delta f)$, the

reciprocal of either the average delay or the delay spread is a measure of the coherence bandwidth of the channel. i.e.,

$$B_c \propto \frac{1}{\mu_\tau} \quad \text{or} \quad B_c \propto \frac{1}{\sigma_\tau} \ . \tag{2.150}$$

The function $\phi_H(\nu; 0) \equiv \phi_H(\nu)$ is called the Doppler psd and gives the average power at the channel output as a function of the Doppler frequency ν. The range of values over which $\phi_H(\nu)$ is significant is called the Doppler spread and is denoted by B_d. Since $\phi_H(\nu)$ and $\phi_T(\Delta t)$ are a Fourier transform pair, it follows that the inverse of the Doppler spread gives a measure of the **coherence time, T_c,** of the channel, i.e.,

$$T_c \approx \frac{1}{B_d} \ . \tag{2.151}$$

The coherence time of the channel is important for evaluating the performance of coding and interleaving techniques that try to exploit the inherent time diversity of the channel. Note that the Doppler spread and, hence, the coherence time depend directly on the velocity of a moving MS. Therefore, any scheme that exploits the time diversity of the channel must be evaluated over the complete range of expected MS velocities. The function $\psi_S(\tau, \nu)$ is called the **scattering function** and gives the average power output of the channel as a function of the time delay τ and the Doppler shift ν. The scattering function is widely used as a compact characterization of multipath-fading channels.

2.3   CHANNEL OUTPUT AUTOCORRELATION

The autocorrelation of the channel output can be expressed in terms of the transmission functions. For example, from (2.117) we have

$$
\begin{aligned}
\phi_{\tilde{r}\tilde{r}}(t, s) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{s}(t - \tau)\tilde{s}^*(s - \eta)\frac{1}{2}\mathrm{E}[g(t, \tau)g^*(s, \eta)]d\tau d\eta \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{s}(t - \tau)\tilde{s}^*(s - \eta)\phi_g(t, s; \tau, \eta)d\tau d\eta \ .
\end{aligned}
\tag{2.152}
$$

For WSSUS channels, the above expression reduces to

$$
\begin{aligned}
\phi_{\tilde{r}\tilde{r}}(t, t + \Delta t) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{s}(t - \tau)\tilde{s}^*(t + \Delta t - \eta)\phi_g(\Delta t; \tau)\delta(\eta - \tau)d\tau d\eta \\
&= \int_{-\infty}^{\infty} \tilde{s}(t - \tau)\tilde{s}^*(t + \Delta t - \tau)\phi_g(\Delta t; \tau)d\tau \ . \tag{2.153}
\end{aligned}
$$

The channel output autocorrelation can also be expressed in terms of the scattering function by substituting the double inverse Fourier transform in (2.128) into (2.151). For WSSUS channels, we can use (2.144) to write
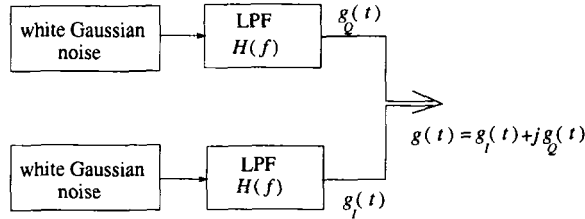
Figure 2.28. Fading simulator that uses low-pass filtered white Gaussian noise.

$$\phi_{\tilde{r}\tilde{r}}(t, t + \Delta t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{s}(t - \tau)\tilde{s}^*(t + \Delta t - \tau)\phi_S(\tau; \nu)e^{j2\pi\nu\Delta t}d\tau d\nu \ .$$

(2.154)

3. LABORATORY SIMULATION OF MULTIPATH-FADING CHANNELS

3.1 FILTERED GAUSSIAN NOISE

A straightforward method of constructing a fading simulator is to filter two independent white Gaussian noise sources with low-pass filters, as shown in Fig. 2.28. The psd of $g_I(t)$ and $g_Q(t)$ are determined by the squared amplitude response of the low-pass filters. If the noise sources have power spectral densities of $\Omega_p/2$ watts/Hz and the low-pass filters have transfer function $H(f)$, then

$$\begin{aligned} S_{g_I g_I}(f) &= S_{g_Q g_Q}(f) = \frac{\Omega_p}{2}|H(f)|^2 \\ S_{g_I g_Q}(f) &= 0 \end{aligned}$$

(2.155)

The two different noise sources must have the same psd to produce a Rayleigh faded envelope. The main limitation with this approach is that only rational forms of the Doppler spectrum can be produced, whereas the Doppler spectrum is typically non-rational as shown in Fig. 2.4. To approximate the non-rational Doppler spectrum in Fig. 2.4, a high-order pole-zero filter is required. Unfortunately, a high-order filter has a long impulse response, and this will significantly increase the run times for software simulation.

For discrete-time simulation the low-pass filter $h(t)$ is implemented as a digital filter. The simplest solution uses a first-order low-pass digital filter, which basically models the fading process as a Markov process. To describe this approach further, let $g_{I,k} \equiv g_I(kT)$ and $g_{Q,k} \equiv g_Q(kT)$ represent the real and imaginary parts of the complex envelope at epoch k, where T is the simulation step size. Then $g_{I,k}$ and $g_{Q,k}$ are Gaussian random variables with the state equation

$$(g_{I,k+1}, g_{Q,k+1}) = \zeta(g_{I,k}, g_{Q,k}) + (1 - \zeta)(w_{1,k}, w_{2,k}) \qquad (2.156)$$

where $w_{1,k}$ and $w_{2,k}$ are independent zero-mean Gaussian random variables, each with time correlation $\mathrm{E}[w_{i,k}w_{i,\ell}] = \sigma^2 \delta_{k\ell}$, $i = 1, 2$. Since $g_k = g_{I,k} + jg_{Q,k}$ has zero-mean, the envelope $\alpha_k = |g_k|^2$ is Rayleigh distributed and the phase $\phi_k = \mathrm{Tan}^{-1}(g_{Q,k}/g_{I,k})$ is uniformly distributed on the interval $[-\pi, \pi]$. It can be shown that the discrete correlation functions of $g_{I,k}$ and $g_{Q,k}$ are

$$\begin{aligned}
\phi_{g_Q g_Q}(n) = \phi_{g_I g_I}(n) &= \mathrm{E}[g_{I,k} g_{I,k+n}] \\
&= \frac{1 - \zeta}{1 + \zeta}\sigma^2 \zeta^{|n|} \qquad (2.157)
\end{aligned}$$

$$\phi_{g_I g_Q}(n) = \phi_{g_Q g_I}(n) = 0 \qquad (2.158)$$

With 2-D isotropic scattering the desired autocorrelation is, from (2.21),

$$\phi_{g_I g_I}(n) = \frac{\Omega_p}{2} J_o(2\pi f_m n T) \ . \qquad (2.159)$$

Clearly the above approach gives a different Doppler spectrum. Completion of model requires that σ^2 and ζ be specified. Taking the discrete-time Fourier transform of (2.157) gives the psd

$$S_{g_I g_I}(f) = \frac{(1 - \zeta)^2 \sigma^2}{1 + \zeta^2 - 2\zeta \cos 2\pi f T} \ . \qquad (2.160)$$

One possibility is to arbitrarily set the 3 dB point of $S_{g_I g_I}(f)$ to $f_m/4$. Solving the resulting quadratic for ζ gives

$$\zeta = 2 - \cos(2\pi f_m T) - \sqrt{(2 - \cos 2\pi f_m T)^2 - 1} \ . \qquad (2.161)$$

To normalized the mean square envelope to Ω_p, the value of σ^2 is chosen as

$$\sigma^2 = \frac{1 + \zeta}{(1 - \zeta)} \frac{\Omega_p}{2} \ . \qquad (2.162)$$

Fig. 2.29 plots an example of the received envelope. The slow roll-off of the first-order low-pass filter leaves some high frequency components in the Doppler spectrum, which are apparent in the faded envelope. Some improvement can be obtained by using a higher order filter, but as explained earlier, this will increase the complexity of the simulator. One advantage of using low-pass filtered white Gaussian noise is the ease by which multiple uncorrelated fading waveforms can be generated. We just need to use uncorrelated noise sources.

3.2   SUM OF SINUSOIDS METHOD

Another very effective channel simulator has been suggested by Jakes that is based on the sums of sinusoids. The description of this method begins with (2.8) and (2.6) and assumes equal strength multipath components ($C_n = 1$). The received complex envelope has the form
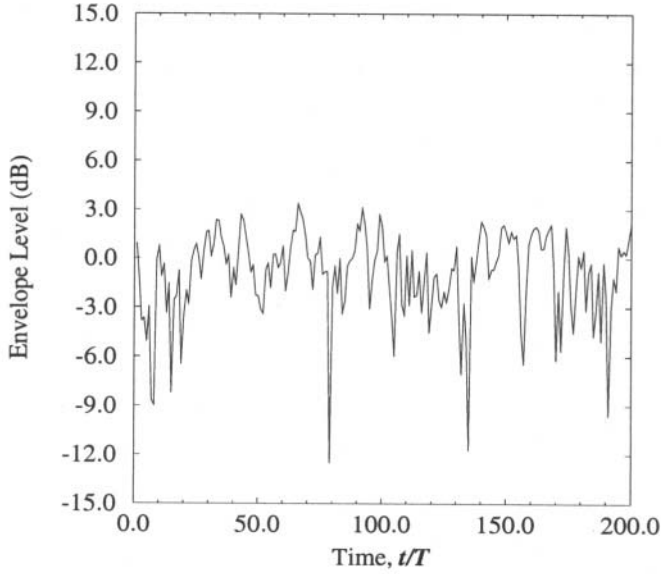
Figure 2.29. Faded envelope generated by filtering white Gaussian noise with a first-order low-pass filter; $f_m T = 0.1$.

$$g(t) = \sum_{n=1}^{N} e^{j(2\pi f_m t \cos \theta_n + \hat{\phi}_n)} , \qquad (2.163)$$

where N is the number of sinusoids and $\hat{\phi}$ is a random phase given by

$$\hat{\phi}_n = -2\pi(f_c + f_m)\tau_n . \qquad (2.164)$$

Jakes approximates a 2-D isotropic scattering environment by choosing the N components to be uniformly distributed in angle, i.e.,

$$\theta_n = \frac{2\pi n}{N} , \quad n = 1, 2, \ldots, N . \qquad (2.165)$$

By choosing $N/2$ to be an odd integer, the sum in (2.163) can be rearranged into the form

$$g(t) = \sum_{n=1}^{N/2-1} \left[e^{-j(2\pi f_m t \cos \theta_n + \hat{\phi}_{-n})} + e^{j(2\pi f_m t \cos \theta_n + \hat{\phi}_n)} \right]$$

$$+ e^{-j(2\pi f_m t + \hat{\phi}_{-N})} + e^{j(2\pi f_m t + \hat{\phi}_N)} \qquad (2.166)$$

where we have relabeled the phase indices. Note that the Doppler shifts progress from $-2\pi f_m \cos(2\pi/N)$ to $+2\pi f_m \cos(2\pi/N)$ as n progresses from 1 to $N/2 - 1$ in the first sum, while in the second sum they progress from $+2\pi f_m \cos(2\pi/N)$ to $-2\pi f_m \cos(2\pi/N)$. Therefore, the frequencies in these

terms overlap. To simplify further Jakes uses nonoverlapping frequencies to write $g(t)$ as

$$g(t) = \sqrt{2} \sum_{n=1}^{M} \left[e^{-j(\hat{\phi}_{-n} + 2\pi f_m t \cos \theta_n)} + e^{j(\hat{\phi}_n + 2\pi f_m t \cos \theta_n)} \right]$$

$$+ e^{-j(\hat{\phi}_{-N} + 2\pi f_m t)} + e^{j(\hat{\phi}_N + 2\pi f_m t)} \qquad (2.167)$$

where

$$M = \frac{1}{2} \left(\frac{N}{2} - 1 \right) \qquad (2.168)$$

and the factor $\sqrt{2}$ is included so that the total power remains unchanged. Note that (2.166) and (2.167) are not equal. In (2.166) all phases are independent. However, (2.167) implies that $-\hat{\phi}_i = \hat{\phi}_{N/2-1-i}$ and, therefore, correlation is introduced into the phases. This correlation leads to non-stationary behavior as described by Pop and Beaulieu [263]. If we further impose the constraint that $\hat{\phi}_n = -\hat{\phi}_{-n}$ in (2.167), then (2.167) can be rewritten in the form

$$g(t) = g_I(t) + jg_Q(t)$$

$$= \sqrt{2} \left\{ \left[2 \sum_{n=1}^{M} \cos \beta_n \cos 2\pi f_n t + \sqrt{2} \cos \alpha \cos 2\pi f_m t \right] \quad (2.169) \right.$$

$$\left. + j \left[2 \sum_{n=1}^{M} \sin \beta_n \cos 2\pi f_n t + \sqrt{2} \sin \alpha \cos 2\pi f_m t \right] \right\}$$

where

$$\alpha = \hat{\phi}_N = -\hat{\phi}_{-N} \qquad (2.170)$$

$$\beta_n = \hat{\phi}_n = -\hat{\phi}_{-n} \qquad (2.171)$$

From the above development, the fading simulator shown in Fig. 2.30 can be constructed. There are M low-frequency oscillators with frequencies $f_n = f_m \cos(2\pi n/N)$, $n = 1, 2, , \ldots, M$, where $M = \frac{1}{2} \left(\frac{N}{2} - 1 \right)$, and with one oscillator with frequency f_m The amplitudes of the oscillators are all unity except for the oscillator at frequency f_m which has amplitude $1/\sqrt{2}$. Note that the structure in Fig. 2.30 implements (2.169), except for a scaling factor of $\sqrt{2}$.

It is desirable that the phase of $g(t) = g_I(t) + jg_Q(t)$ be uniformly distributed. This can be accomplished by choosing the phases α and β_n so that $< g_I^2(t) > = < g_Q^2(t) >$ and $< g_I(t)g_Q(t) > = 0$, where $< \cdot >$ is a time average operator. From Fig. 2.30

offset oscillators

$g(t) = g_I(t) + j\, g_Q(t)$
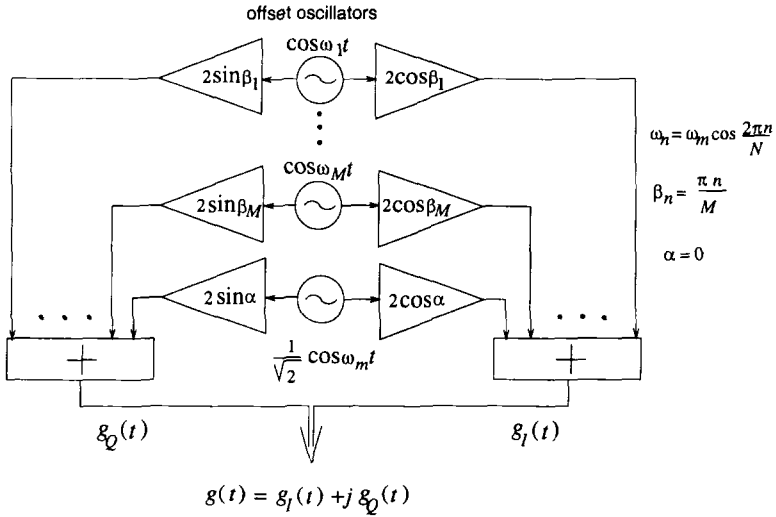
Figure 2.30. Jakes' fading simulator that generates a faded envelope by summing a number of low frequency oscillators. Choosing the phases $\alpha = 0$ and $\beta_n = \pi n/M$ will yield $< g_Q^2(t) > = M$, $< g_I^2(t) > = M + 1$, and $< g_I(t)g_Q(t) > = 0$.

$$< g_I^2(t) > \;=\; 2 \sum_{n=1}^{M} \cos^2 \beta_n + \cos^2 \alpha$$

$$=\; M + \cos^2 \alpha + \sum_{n=1}^{M} \cos 2\beta_n \qquad (2.172)$$

$$< g_Q^2(t) > \;=\; 2 \sum_{n=1}^{M} \sin^2 \beta_n + \sin^2 \alpha$$

$$=\; M + \sin^2 \alpha - \sum_{n=1}^{M} \cos 2\beta_n \qquad (2.173)$$

$$< g_I(t)g_Q(t) >\; =\; 2 \sum_{n=1}^{M} \sin \beta_n \cos \beta_n + \sin \alpha \cos \alpha \;. \qquad (2.174)$$

Choosing $\alpha = 0$ and $\beta_n = \pi n/M$, gives $< g_Q^2(t) > = M$, $< g_I^2(t) > = M + 1$, and $< g_I(t)g_Q(t) > = 0$. The mean square values $< g_I^2(t) >$ and $< g_Q^2(t) >$ can be scaled to any desired value. A typical Rayleigh faded envelope, obtained by using $N = 34$ or $(M = 8)$ is shown in Fig. 2.31. The normalized autocorrelation function

$$\phi_{gg}(\tau) = \frac{E[g^*(t)g(t+\tau)]}{E[|g(t)|^2]} \qquad (2.175)$$
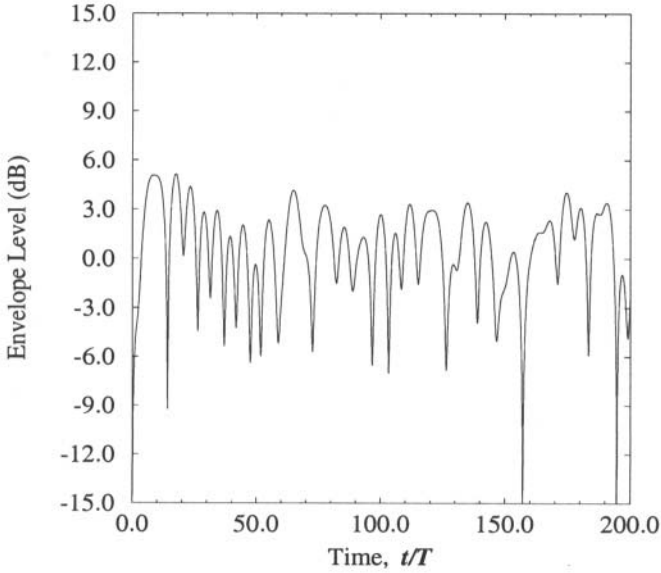
Figure 2.31. Faded envelope generated by using Jakes' fading simulator with $M = 8$ oscillators; $f_m T = 0.1$.

is plotted against the normalized time delay $f_m \tau$ in Fig. 2.32. Observe that the autocorrelation tends to deviate from the desired values at large lags. This can be improved upon by increasing the number of oscillators that are used in the simulator. For example, Fig. 2.33 shows the normalized autocorrelation function when the number of oscillators is doubled from 8 to 16. One of the advantages of using Jakes' method is that the autocorrelation and, hence, the psd of the inphase and quadrature components of the received signal can be generated so as to closely approximate a 2-D isotropic scattering environment.

3.3   MULTIPLE FADED ENVELOPES

In many cases it is desirable to generate multiple envelopes with uncorrelated fading. Jakes' extended his method to generate up to M fading envelopes by using the same M low frequency oscillators. This is accomplished by giving the nth oscillator the additional phase shift $\theta_{nk} = \gamma_{nk} + \beta_n$, $1 \leq k \leq M$, yielding the kth faded envelope

$$g_k(t) = \sum_{n=1}^{M} [\cos(\beta_n) + j \sin(\beta_n)] \cos(2\pi f_n t + \theta_{nk}) \tag{2.176}$$

The appropriate values of γ_{nk} and β_n are determined by imposing the additional

86
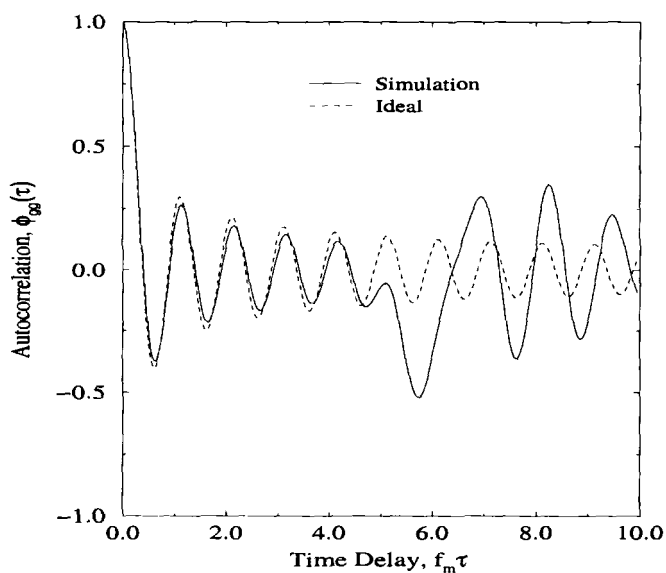
Figure 2.32. Autocorrelation of inphase and quadrature components obtained by using Jakes' fading simulator with $M = 8$ oscillators.
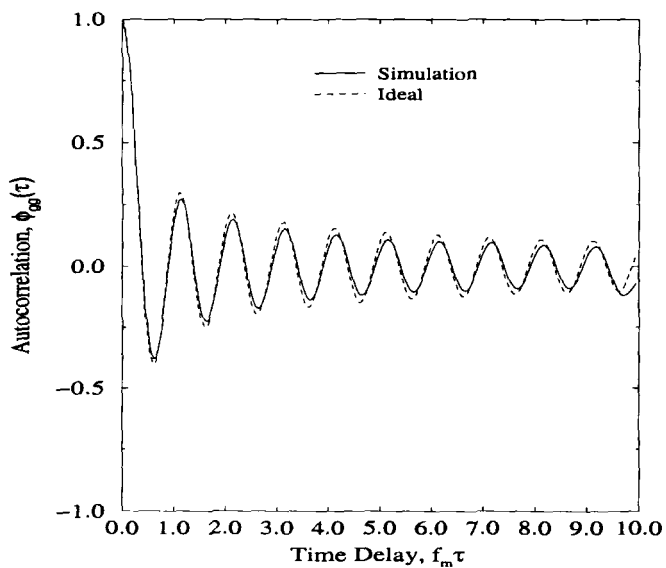
Figure 2.33. Autocorrelation of inphase and quadrature components obtained by using Jakes' fading simulator with 16 oscillators.
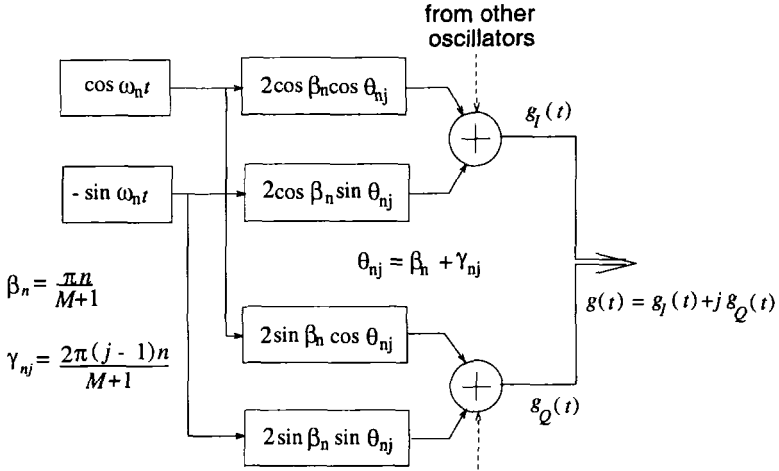
Figure 2.34.　Jakes' method for obtaining multiple Rayleigh fading envelopes.

constraint that the multiple faded envelopes be uncorrelated (or as nearly un-correlated as possible). By using two quadrature low- frequency oscillators per offset, rather than a single oscillator, the use of phase shifters can be eliminated. This leads to the fading generator shown in Fig. 2.34.

Consider the following choice for β_n and γ_{nk} with the objective yielding uncorrelated waveforms

$$\beta_n \;=\; \frac{\pi n}{M+1} \qquad n = 1, \, 2, \, \ldots, \, M \qquad\qquad (2.177)$$

$$\gamma_{nk} \;=\; \frac{2\pi(k-1)n}{M+1} \qquad n = 1, \, 2, \, \ldots, \, M \; . \qquad\qquad (2.178)$$

By using these values, the crosscorrelations between the different faded en-velopes can be computed. Fig. 2.35 plots the typical normalized crosscorrela-tion

$$\phi_{g_i g_j}(\tau) = \frac{\mathrm{E}[g_i^*(t)g_j(t+\tau)]}{\sqrt{\mathrm{E}[|g_i(t)|^2]\mathrm{E}[|g_j(t)|^2]}} \qquad\qquad (2.179)$$

against the normalized time delay $f_m\tau$. Although it is possible to make $\phi_{g_i g_j}(\tau)$ at $\tau = 0$, we observe that the envelope crosscorrelations can be quite large for $\tau \neq 0$. This property is not desirable.

Dent *et. al.* [81] suggested a modification to Jakes' approach that uses orthogonal Walsh-Hadamard codewords to decorrelate the faded envelopes. The Walsh-Hadamard codewords are obtained from a Hadamard matrix $\mathbf{H}_M$. The matrix $\mathbf{H}_M$ is generated by using the recursion

$$\mathbf{H}_M = \begin{bmatrix} \mathbf{H}_{M/2} & \mathbf{H}_{M/2} \\ \mathbf{H}_{M/2} & -\mathbf{H}_{M/2} \end{bmatrix} \qquad\qquad (2.180)$$
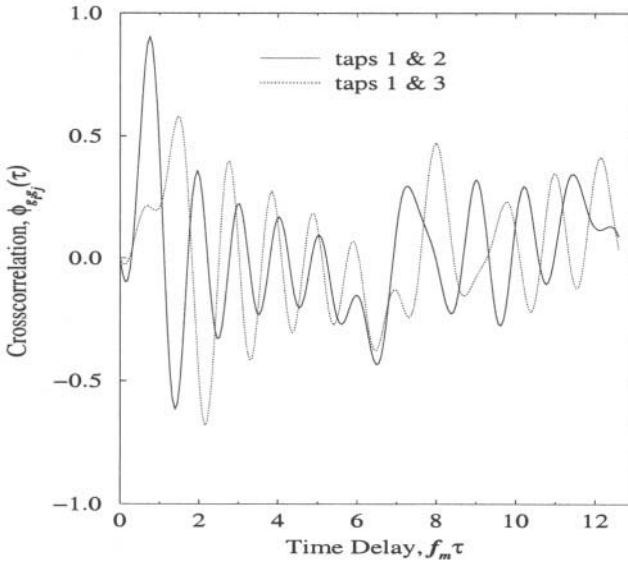
Figure 2.35. Crosscorrelation between the faded envelopes that are obtained by using Jakes' fading simulator with 8 oscillators.

where $\mathbf{H}_1 = [1]$. Let the kth row of $\mathbf{H}_M$ be the codeword $A_k(n)$. To generate the kth faded envelope, the outputs of the M low frequency oscillators are weighted by the co-ordinates of the codeword $A_k(n)$ and combined to yield

$$g_k(t) = \sum_{n=1}^{M} A_k(n) \left[\cos(\beta_n) + j \sin(\beta_n)\right] \cos(2\pi f_n t + \theta_{nk}) \qquad (2.181)$$

where $M = 2^k$ for some integer k, $\theta_{nk} = \gamma_{nk} + \beta_n$, with β_n and γ_{nk} defined in (2.177) and (2.178), respectively. This method decreases the envelope crosscorrelations, due to the orthogonality of the Walsh-Hadamard codewords. It can be shown that the autocorrelations of the faded envelopes are the same as those shown in Figs. 2.32 and 2.33. The envelope crosscorrelations are shown in Figs. 2.36 and 2.37. The crosscorrelations are zero at zero lag. At non-zero time lags the crosscorrelations are not exactly zero, but are small enough to be effectively zero. By increasing the number of oscillators, the crosscorrelations remain close to zero over a larger range of time lags.

Sometimes it is desirable to generate multiple faded envelopes with a specified crosscorrelation to study, for example, the effects of branch correlation in receivers that employ antenna diversity. One straight forward approach uses a linear combination of uncorrelated faded envelopes. Suppose that the two
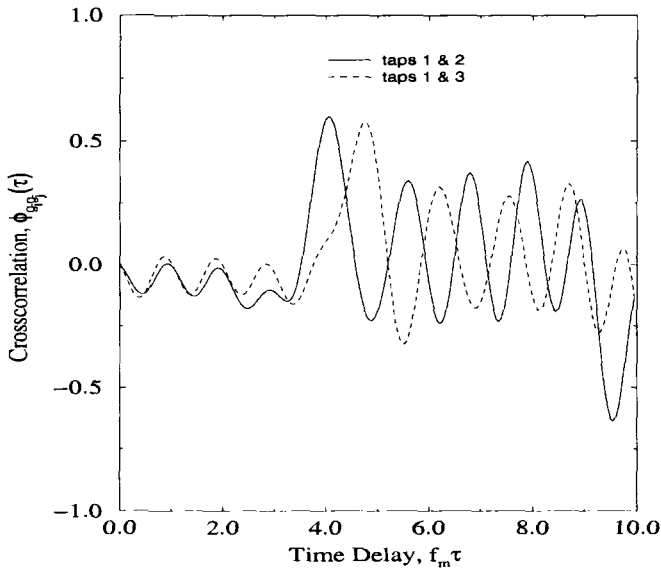
Figure 2.36. Crosscorrelation between the faded envelopes that are obtained by using the modified Jakes' fading simulator with $M = 8$ oscillators.
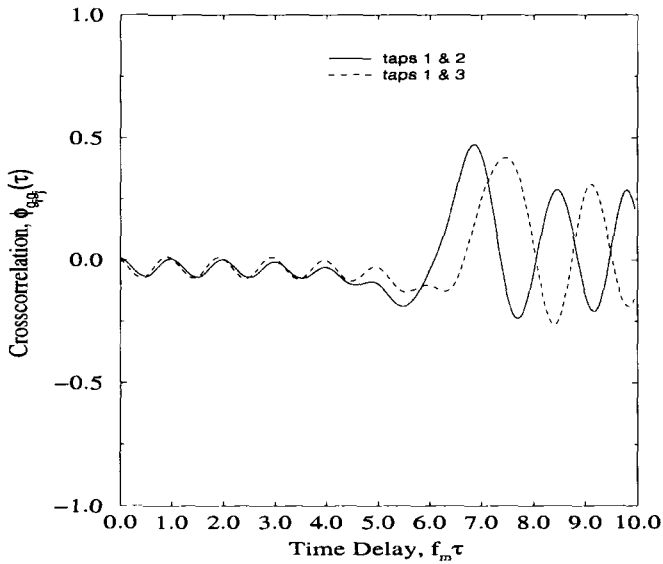
Figure 2.37. Crosscorrelation between the faded envelopes that are obtained by using the modified Jakes' fading simulator with $M = 16$ oscillators.

complex envelopes $g_i(t)$ and $g_j(t)$ are uncorrelated, and a third complex envelope $g_k(t) = ag_i(t) + (1-a)g_Q(t)$ is formed by taking a linear combination of the first two. Then the normalized crosscorrelation of $g_k(t)$ and $g_i(t)$ is

$$\phi_{g_i g_k}(\tau) = \left[1 + \frac{(1-a)^2 E[|g_j(t)|^2]}{a^2 E^2[|g_i(t)|^2]}\right]^{-1/2} \cdot \phi_{g_i g_i}(\tau) \ . \tag{2.182}$$

Notice that $\phi_{g_i g_k}(\tau)$ varies from 0 to $\phi_{g_i g_i}(\tau)$ as a varies from 0 to 1.

3.4   SIMULATION OF WIDE-BAND MULTIPATH-FADING CHANNELS

τ-spaced model:.   The τ-spaced model, models the channel by a tapped delay line with number of taps at different delays. Each tap is the result of a large number of multipath components and, therefore, the taps will experience multipath fading. Letting $\tilde{s}(t)$ be the complex envelope of the transmitted signal, the complex envelope of the received signal is

$$\tilde{r}(t) = \sum_{i=1}^{\ell} g_i(t)\tilde{s}(t - \tau_i) \tag{2.183}$$

where ℓ is the number of taps, and the $g_i(t)$ and τ_i are the complex gains and path delays associated with the taps. Although the τ_i are random, they are usually fixed in the model. It follows that the τ-spaced channel has impulse response

$$g(t, \tau) = \sum_{i=1}^{\ell} g_i(t)\delta(\tau - \tau_i) \tag{2.184}$$

and can be described by the tap gain vector

$$\mathbf{g}(t) = (g_1(t), \ g_2(t), \ \ldots, \ g_\ell(t)) \tag{2.185}$$

and the tap delay vector

$$\tau = (\tau_1, \tau_2, \ \ldots, \tau_\ell) \ . \tag{2.186}$$

Sometimes it is convenient if the tap delays are multiples of some small number τ, leading to the τ-spaced tapped delay line channel model shown in Fig. 2.38. Many of the tap coefficients in the tapped delay line are zero, reflecting the fact that no energy is received at these delays. The time varying channel tap coefficients $\{g_k(t)\}$ can be generated by using the approaches described in Section 3..

If we assume a WSSUS channel and we assume that each tap experiences isotropic scattering, then each tap should experience uncorrelated fading with the autocorrelation
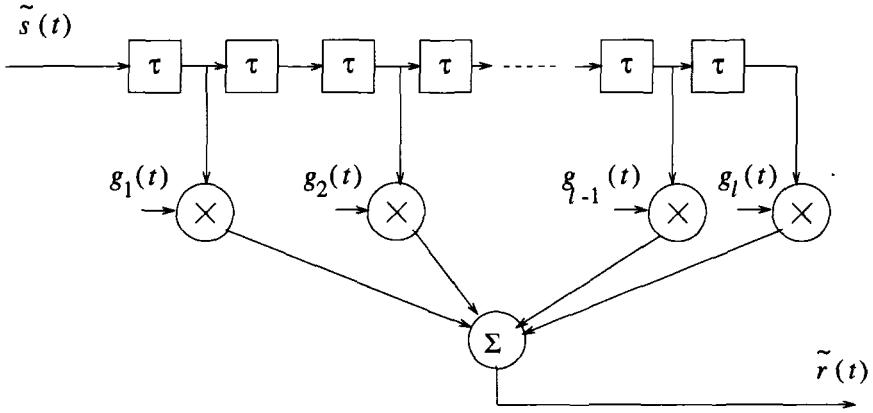
Figure 2.38. Wide-band multipath-fading channel model with discrete multipath components.

$$\phi_{g_k g_k}(\tau) = \frac{\Omega_k}{2} J_0(2\pi f_m \tau), \qquad 1 \le k \le \ell \qquad (2.187)$$

where Ω_k is the envelope power or associated with the kth tap and $J_0(\cdot)$ is the zero-order Bessel function of the first kind, and f_m is the maximum Doppler frequency. Since the taps are uncorrelated, the total envelope power is

$$\Omega_p = \sum_{k=1}^{\ell} \Omega_k \ . \qquad (2.188)$$

It follows that the tap gain vector g has the covariance matrix [173, 167]

$$\Phi_{\mathbf{g}}(\tau) = \frac{1}{2} E[\mathbf{g}^H(t)\mathbf{g}(t+\tau)] = \frac{1}{2} J_0(2\pi f_m \tau)\mathbf{\Omega} \qquad (2.189)$$

where H denotes Hermitian transposition and

$$\mathbf{\Omega} \triangleq \text{diag}[\Omega_0, \ \Omega_1, \dots, \ \Omega_\ell] \qquad (2.190)$$

COST207 models:. The COST207 models were developed and standardized for the GSM system. Four different Doppler spectra, $S_{gg}(f)$ have been specified in the COST207 models [67]. First define

$$G(f) = A \exp\left\{-\frac{(f-f_1)^2}{2f_2^2}\right\} \qquad (2.191)$$

The following types are defined;

a) CLASS is the classical Doppler spectrum, used for path delays not in excess of 500 ns ($\tau_i \le 500$ ns);

$$(\text{CLASS}) \quad S_{gg}(f) = \frac{A}{\sqrt{1 - (f/f_m)^2}} \quad |f| \leq f_m \quad (2.192)$$

b) GAUS1 is the sum of two Gaussian functions, used for path delays from 500 ns to 2 μs; (5 00 ns $\leq \tau_i \leq 2\mu$s)

$$(\text{GAUS1}) \quad S_{gg}(f) = G(A, -0.8f_m, 0.05f_m) + G(A_1, 0.4f_m, 0.1f_m)$$
$$(2.193)$$

where A_1 is 10 dB below A.

c) GAUS2 is the sum of two Gaussian functions, used for path delays exceeding 2 μs; ($\tau_i > 500$ ns)

$$(\text{GAUS2}) \quad S_{gg}(f) = G(B, 0.7f_m, 0.1f_m) + G(B_1, -0.4f_m, 0.15f_m)$$
$$(2.194)$$

where B_1 is 15 dB below B.

d) RICE is a combination of the classical Doppler spectrum and one specular path that is sometimes used for the shortest path;

$$(\text{RICE}) \quad S_{gg}(f) = \frac{0.41}{2\pi f_m \sqrt{1 - (f/f_m)^2}} + 0.91\delta(f - 0.7f_m)$$
$$|f| \leq f_m \quad (2.195)$$

A number of specific models have been defined in the COST207 study [67]. Typical urban (TU) (non-hilly) and bad urban (BU) (hilly) power delay profiles are shown in Table 2.1 and Fig. 2.39. Sometimes it is desirable to reduce the number of paths to reduce the computational requirements of computer simulations. Table 2.2 and Fig. 2.40 show the 6-ray reduced typical urban and reduced bad urban channel, as defined by COST207 [67]. Also provided are models for rural (non-hilly) areas (RA) in Table 2.3, typical hilly terrain (HT) in Table 2.4, and reduced hilly terrain (HT) in Table 2.5.

T-**spaced model:.** A typical digital communication system consists of the combination of a transmitter filter, modulator, waveform channel, demodulator, and receiver filter. Data symbols are fed into the transmit filter every T seconds, while T-spaced samples are taken at the output of the receiver filter, where T is the baud duration. The overall system from the input to the transmitter filter to the output of the sampler can be modeled by an equivalent overall T-spaced finite impulse response (FIR) channel.

	Typical Urban (TU)			Bad Urban (BU)	
delay μs	Fractional Power	Doppler Category	delay μs	Fractional Power	Doppler Category
0.0	0.092	CLASS	0.0	0.033	CLASS
0.1	0.115	CLASS	0.1	0.089	CLASS
0.3	0.231	CLASS	0.3	0.141	CLASS
0.5	0.127	CLASS	0.7	0.194	GAUS1
0.8	0.115	GAUS1	1.6	0.114	GAUS1
1.1	0.074	GAUS1	2.2	0.052	GAUS2
1.3	0.046	GAUS1	3.1	0.035	GAUS2
1.7	0.074	GAUS1	5.0	0.140	GAUS2
2.3	0.051	GAUS2	6.0	0.136	GAUS2
3.1	0.032	GAUS2	7.2	0.041	GAUS2
3.2	0.018	GAUS2	8.1	0.019	GAUS2
5.0	0.025	GAUS2	10.0	0.006	GAUS2

Table 2.1.   Typical urban (TU) ($\sigma_\tau = 1.0$ μs) and bad urban (BU) ($\sigma_\tau = 2.5$ μs) power delay profiles, from [67].
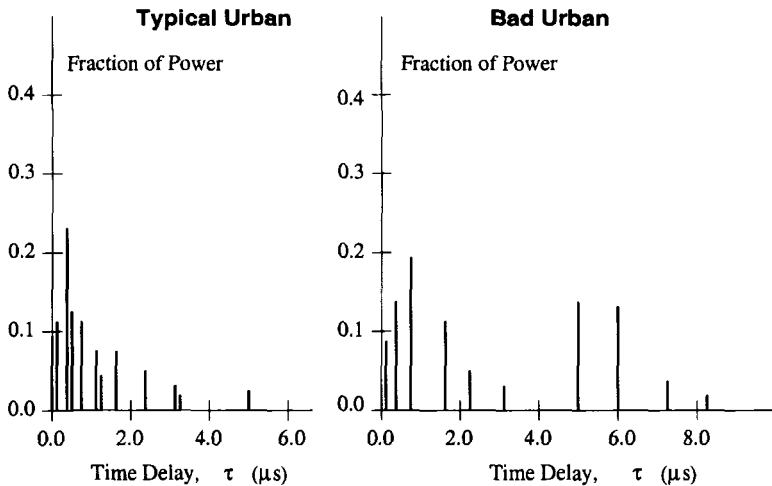
Figure 2.39.   Typical urban (TU) and bad urban (BU) power delay profiles, from [67].

The T-spaced channel model is similar to the τ-spaced channel model, except that the channel taps are T-spaced. Usually, the taps in the T-spaced model are all non-zero and correlated. The tap correlations often lead to difficulties when *analyzing* the performance of digital communication systems that are operating on these channels. These analytical difficulties are often overcome by assuming that the T-spaced taps are uncorrelated [79,99,137,201,321,187]. For computer simulations, however, such simplifications are not necessary and in fact undesirable. However, to reduce run times in discrete-time simulations,

	Typical Urban (TU)			Bad Urban (BU)	
delay μs	Fractional Power	Doppler Category	delay μs	Fractional Power	Doppler Category
0.0	0.189	CLASS	0.0	0.164	CLASS
0.2	0.379	CLASS	0.3	0.293	CLASS
0.5	0.239	CLASS	1.0	0.147	GAUS1
1.6	0.095	GAUS1	1.6	0.094	GAUS1
2.3	0.061	GAUS2	5.0	0.185	GAUS2
5.0	0.037	GAUS2	6.6	0.117	GAUS2

Table 2.2. Reduced typical urban (TU) ($\sigma_\tau = 1.0$ μs) and bad urban (BU) ($\sigma_\tau = 2.5$ μs) power delay profiles, from [67].
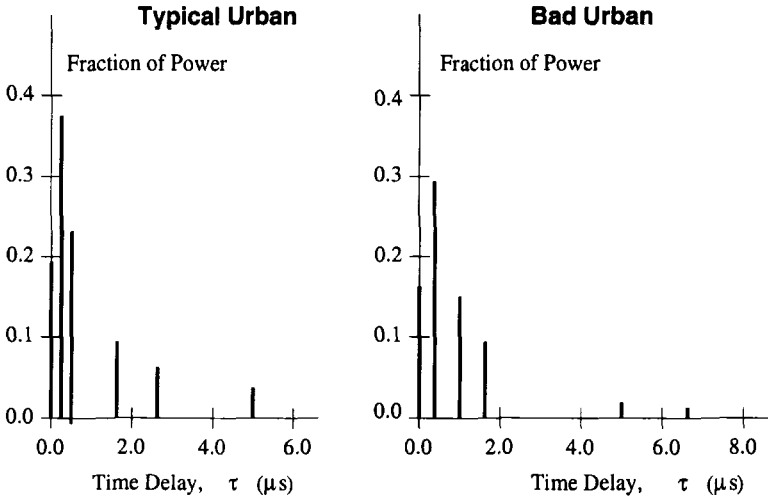
Figure 2.40. Reduced typical urban (TU) and bad urban (BU) power delay profiles, from [67].

it is sometimes desirable to set the simulation step size to the baud duration T. We now describe a method for generating the T-spaced tap coefficients with the proper crosscorrelations when a linear modulation scheme is used and the underlying channel model is τ-spaced. Consider the arrangement shown in Fig 2.41. As discussed in Chapter 4, a typical digital communication system consists of a transmit filter $h_a(t)$, channel $g(t, \tau)$, and receiver matched filter $h_a^*(-t)$. The overall pulse is $p(t) = h_a(t) * h_a^*(-t)$, and is chosen, for example, to be a raised cosine pulse. To obtain the T-space channel tap co-efficient, we pass the pulse $p(t)$ through the τ-spaced channel and extract T-spaced samples. The T-spaced samples are a linear combination of the taps in the τ-spaced model. Suppose that a vector of M, T-spaced, tap coefficients is generated in this manner

delay μs	Fractional Power	Doppler Category
0.0	0.602	RICE
0.1	0.241	CLASS
0.2	0.096	CLASS
0.3	0.036	CLASS
0.4	0.018	CLASS
0.5	0.006	CLASS

Table 2.3. Typical rural (non-hilly) area (RA) ($\sigma_\tau = 0.1$ μs), from [67].

delay μs	Fractional Power	Doppler Category
0.0	0.026	CLASS
0.1	0.042	CLASS
0.2	0.066	CLASS
0.3	0.105	CLASS
0.4	0.263	GAUS1
0.5	0.263	GAUS1
0.0	0.105	GAUS1
0.1	0.042	GAUS2
0.2	0.034	GAUS2
0.3	0.026	GAUS2
0.4	0.016	GAUS2
0.5	0.011	GAUS2

Table 2.4. Typical hilly terrain (HT) ($\sigma_\tau = 5.0$ μs), from [67].

$$\mathbf{g}_T(t) = (g_{1T}(t),\ g_{2T}(t),\ \ldots,\ g_{MT}(t)) \qquad (2.196)$$

Then $\mathbf{g}_T^T(t) = \mathbf{A}\mathbf{g}^T(t)$, where $\mathbf{g}(t)$ is defined in (2.185), and $\mathbf{A}$ is an $M \times \ell$ real matrix. The parameter M is a design parameter equal to the number of T-spaced taps that we wish to process in the receiver. As described in the example below, the entries of the matrix $\mathbf{A}$ are determined by the overall pulse response of the transmitter and receiver filters, the relative power and delays of the rays in the τ-spaced model, and the T-spaced sampler timing phase. Note that the matrix $\mathbf{A}$ only needs to be generated once each time the relative delays of the rays in the τ-spaced channel and/or the sampler timing phase change. The covariance matrix of the T-spaced tap gain vector $\mathbf{g}_T$ is

delay μs	Fractional Power	Doppler Category
0.0	0.413	CLASS
0.1	0.293	CLASS
0.3	0.145	CLASS
0.5	0.074	CLASS
15.0	0.066	GAUS2
17.2	0.008	GAUS2

Table 2.5. Reduced hilly terrain (HT) ($\sigma_\tau = 5.0 \ \mu s$), from [67].
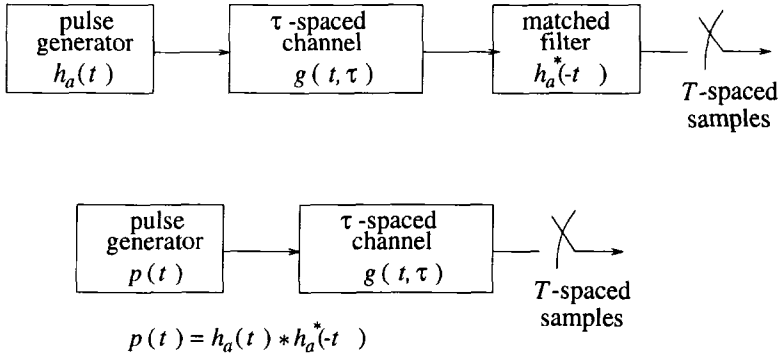
Figure 2.41. Method for generating correlated tap coefficients in a T-spaced channel model.

$$\Phi_{\mathbf{g}_T}(\tau) = \frac{1}{2}\mathrm{E}\left[\mathbf{g}_T^H(t+\tau)\mathbf{g}_T(t)\right]$$

$$= \frac{1}{2}\mathrm{E}\left[\mathbf{A}\mathbf{g}^H(t+\tau)\mathbf{g}(t)\mathbf{A}^T\right]$$

$$= \mathbf{A}\Phi_{\mathbf{g}}(\tau)\mathbf{A}^T \ .$$

For a WSSUS channel with 2-D isotropic scattering

$$\Phi_{\mathbf{g}_T}(\tau) = \frac{1}{2}\mathbf{A}\Omega\mathbf{A}^T J_0(2\pi f_m \tau) \ . \tag{2.197}$$

Example 2.1____Suppose that the combination of the transmitter and receiver filter is a raised cosine pulse having a roll-off factor of 0.35^3

---

[3] See Chapter 4 for a discussion of raised cosine pulse shaping.
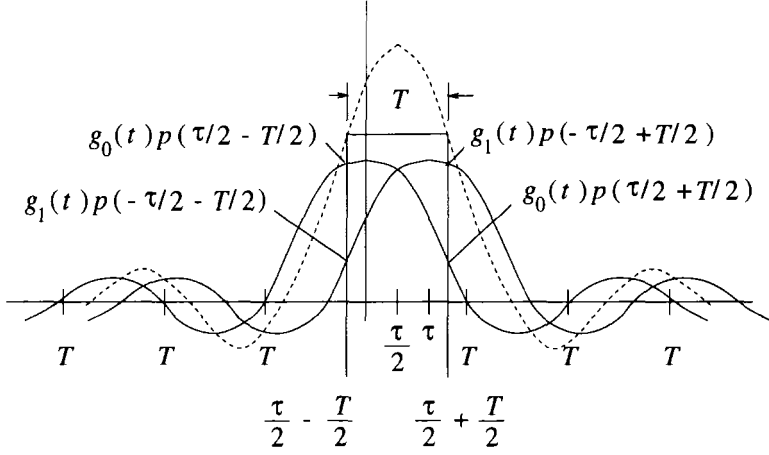
Figure 2.42. Generation T-spaced taps from a τ-spaced model.

$$p(t) = \text{Sa}(\pi t/T) \cdot \frac{\cos(\pi\beta t/T)}{1 - 4\beta^2 t^2/T^2} \qquad (2.198)$$

where $\beta = 0.35$. The τ-spaced waveform channel is characterized by two equal strength taps ($\Omega_0 = \Omega_1$) with a differential delay of $\tau = |\tau_1(t) - \tau_0(t)|$. In this example, we wish to generate the two main taps in the T-spaced channel model, $g_{0T}(t)$ and $g_{1T}(t)$, under the condition that $\tau = T/4$. Let

$$\begin{aligned} \mathbf{g}(t) &= (g_0(t), g_1(t)) \\ \mathbf{g}_T(t) &= (g_{0T}(t), g_{1T}(t)) \end{aligned}$$

and

$$\mathbf{g}_T(t) = \mathbf{g}(t)\mathbf{A}^T \ .$$

The entries of matrix $\mathbf{A}$ depend on the timing phase of the T-spaced samples taken at the output of the pulse generator. In a practical system, the sampler timing phase is determined by the synchronization process in the receiver. Suppose that the sampler timing phase is chosen so that the T-spaced taps $g_{0T}(t)$ and $g_{1T}(t)$ have equal variance. Consider Fig. 2.42. Since $\Omega_0 = \Omega_1$ for the τ-spaced channel in this example, the entries of matrix $\mathbf{A}$ can be obtained by writing

$$\begin{aligned} g_{0T}(t) &= g_0(t)p(\tau/2 - T/2) + g_1(t)p(-\tau/2 - T/2) \\ g_{1T}(t) &= g_0(t)p(\tau/2 + T/2) + g_1(t)p(-tau/2 + T/2) \end{aligned}$$

Hence,

$$\mathbf{A} = \begin{bmatrix} p(\tau/2 - T/2) & p(-\tau/2 - T/2) \\ p(\tau/2 + T/2) & p(-\tau/2 + T/2) \end{bmatrix}$$

For $\tau = T/4$ and $\beta = 0.35$

$$A = \begin{bmatrix} p(-3T/8) & p(-5T/8) \\ p(5T/8) & p(3T/8) \end{bmatrix} = \begin{bmatrix} 0.7717 & 0.4498 \\ 0.4498 & 0.7717 \end{bmatrix}$$

4. SHADOWING

Let $\Omega_v = E[\alpha(t)]$ denote the mean envelope level, where the expectation is taken over the pdf of the received envelope, e.g., the Rayleigh and Rice distribution in (2.43) and (2.50), respectively. Sometimes Ω_v is called the **local mean** because it represents the mean envelope level where the averaging is performed over a distance of a few wavelengths that represents a locality. The local mean Ω_v itself is a random variable due to shadow variations that are caused by large terrain features between the BS and MS, such as buildings and hills in macrocells and smaller objects such as vehicles in microcells. The same argument applies to the mean squared envelope level $\Omega_p = E[\alpha^2(t)]$. Empirical studies have shown that Ω_v and Ω_p have the log-normal distributions

$$p_{\Omega_v}(x) = \frac{2}{x\sigma_\Omega\xi\sqrt{2\pi}} \exp\left\{ -\frac{\left(10\log_{10}x^2 - \mu_{\Omega_v\,(\text{dBm})}\right)^2}{2\sigma_\Omega^2} \right\} \quad (2.199)$$

$$p_{\Omega_p}(x) = \frac{1}{x\sigma_\Omega\xi\sqrt{2\pi}} \exp\left\{ -\frac{\left(10\log_{10}x - \mu_{\Omega_p\,(\text{dBm})}\right)^2}{2\sigma_\Omega^2} \right\} \quad (2.200)$$

where

$$\mu_{\Omega_v\,(\text{dBm})} = 30 + 10E[\log_{10}\Omega_v^2] \quad (2.201)$$

$$\mu_{\Omega_p\,(\text{dBm})} = 30 + 10E[\log_{10}\Omega_p] \quad (2.202)$$

and $\xi = \ln 10/10$. The mean values $\mu_{\Omega_v\,(\text{dBm})}$ and $\mu_{\Omega_p\,(\text{dBm})}$ are sometimes called the **area mean** because the averaging is performed over an area that is large enough to average over the shadows. The area mean is determined by the propagation path loss between the BS and MS. By using a transformation of random variables, $\Omega_v\,_{(\text{dBm})} = 30 + 10\log_{10}\Omega_v^2$ and $\Omega_p\,_{(\text{dBm})} = 30 + 10\log_{10}\Omega_p$ can be shown to have the Gaussian densities

$$p_{\Omega_v\,(\text{dBm})}(x) = \frac{1}{\sqrt{2\pi}\sigma_\Omega} \exp\left\{ -\frac{(x - \mu_{\Omega_v\,(\text{dBm})})^2}{2\sigma_\Omega^2} \right\} \quad (2.203)$$

$$p_{\Omega_p\,(\text{dBm})}(x) = \frac{1}{\sqrt{2\pi}\sigma_\Omega} \exp\left\{ -\frac{(x - \mu_{\Omega_p\,(\text{dBm})})^2}{2\sigma_\Omega^2} \right\}. \quad (2.204)$$

Note that the logarithm of the log-normal random variable in yields a normal random variable.

Some confusion may arise in the description of log-normal shadow distribution, because some authors [235, 122, 123] treat the mean envelope Ω_v as being log-normally distributed with standard deviation σ_Ω, while other authors [203, 225, 268, 310] treat the mean square-envelope Ω_p as being log-normally distributed with the same value of σ_Ω. Clearly, these two quantities are not the same. It is shown in Appendix 2A that the standard deviation σ_Ω is the same in each case. However, with Ricean fading the means differ by

$$\mu_{\Omega_p \text{ (dBm)}} = \mu_{\Omega_v \text{ (dBm)}} + 10\log_{10} C(K) \qquad (2.205)$$

where

$$C(K) = \frac{4e^{2K}(K+1)}{\pi \, {}_1F_1^2(3/2, 1; K)} \qquad (2.206)$$

and ${}_1F_1(\cdot, \cdot; \cdot)$ denotes the confluent hypergeometric function. The shadow standard deviation σ_Ω ranges from 5 to 12 dB with 8 dB being a typical value for macrocellular applications. The shadow standard deviation increases slightly with frequency (0.8 dB higher at 1800 MHz than at 900 MHz), but has been observed to be nearly independent radio path length, even for distances that are very close to the transmitter [225]. The shadow standard deviation that is observed in microcells varies between 4 and 13dB [278, 33, 216, 142, 144, 226]. Mogensen [226] has reported $\sigma_\Omega = 6.5$ to 8.2 dB at 900 MHz in urban areas, while Mockford *et. al.* [225] report a value of 4.5 dB for urban areas. Berg [33] and Goldsmith and Greenstein [144] report that σ_Ω is around 4 dB for a spatial averaging window of 20 wavelengths and BS antenna heights of about 10 (m). Several studies suggest that σ_Ω decreases with an increase in the degree of urbanization or density of scatters. For example, the results presented by Mockford *et. al.* [225] suggest that σ_Ω is 1.3 to 1.8 dB higher in a suburban environment than in an urban environment.

4.1 LABORATORY SIMULATION OF SHADOWING

One of the challenges when constructing a shadow simulator is to account for the spatial correlation of the shadows. Several studies have investigated the spatial correlation of shadows [162, 151, 216, 172, 152]. One simple model has been suggested by Gudmundson [152], where log-normal shadowing is modeled as a Gaussian white noise process that is filtered by a first-order low-pass filter. With this model

$$\Omega_{k+1 \text{ (dBm)}} = \zeta\Omega_{k \text{ (dBm)}} + (1 - \zeta)v_k \qquad (2.207)$$

where $\Omega_{k \text{ (dBm)}}$ is the mean envelope or mean squared-envelope, expressed in decibels, that is experienced at location k, ζ is a parameter that controls the spatial correlation of the shadows, and v_k is a zero-mean Gaussian random variable with $\phi_{vv}(n) = \tilde{\sigma}^2\delta(n)$. From equation (2.157), it immediately follows that the spatial autocorrelation function of $\Omega_{k \text{ (dBm)}}$ is

$$\phi_{\Omega_{(\mathrm{dBm})}\Omega_{(\mathrm{dBm})}}(n) = \frac{1-\zeta}{1+\zeta}\tilde{\sigma}^2\zeta^{|n|} \ . \tag{2.208}$$

Since the variance of log-normal shadowing is

$$\sigma_\Omega^2 = \phi_{\Omega_{(\mathrm{dBm})}\Omega_{(\mathrm{dBm})}}(0) = \frac{1-\zeta}{1+\zeta}\tilde{\sigma}^2 \tag{2.209}$$

we can express the autocorrelation of Ω_k as

$$\phi_{\Omega_{(\mathrm{dBm})}\Omega_{(\mathrm{dBm})}}(n) = \sigma_\Omega^2\zeta^{|n|} \ . \tag{2.210}$$

This approach generates shadows that decorrelate exponentially with distance. It is interesting to note that Mandayam *et. al.* [214] have shown through an extreme value analysis that log-normal shadows cannot decorrelate exponentially with distance. Nevertheless, in the absence of a better solution, Gudmundson's model in (2.207) is still useful and effective. To use the simulator in (2.207) we must relate the decorrelation parameter ζ to the simulation index k. Suppose that we wish to model the shadows that are experienced by a MS that is traveling with velocity v. The envelope (or squared envelope) is sampled every T seconds. In kT seconds the MS moves a distance vkT. Let ζ_D be the shadow correlation between two points separated by a spatial distance of D m. Then the time autocorrelation of the shadowing is

$$\phi_{\Omega_{(\mathrm{dBm})}\Omega_{(\mathrm{dBm})}}(k) \equiv \phi_{\Omega_{(\mathrm{dBm})}\Omega_{(\mathrm{dBm})}}(kT) = \sigma_\Omega^2\zeta_D^{(vT/D)|k|} \ , \qquad k \geq 0 \ . \tag{2.211}$$

Comparing (2.210) and (2.211) we see that $\zeta = \zeta_D^{(vT/D)}$. For typical suburban propagation at 900 MHz, it has been experimentally verified by Gudmundson [150] that $\sigma_\Omega \approx 7.5$ dB with a spatial correlation of approximately 0.82 at a distance of 100 m. For typical microcellular propagation at 1700 MHz, Gudmundson has also reported $\sigma_\Omega = 4.3$ dB with a spatial correlation of 0.3 at a distance of 10 m.

4.2   COMPOSITE SHADOWING-FADING DISTRIBUTIONS

Sometimes it is desirable to know the composite distribution due to shadowing and multipath fading. This may be particularly true for the case of slow moving or stationary MSs, where the receiver is unable to average over the effects of fading and a composite distribution is necessary for evaluating link performance and other quantities. Two different approaches have been suggested in the literature for obtaining the composite distribution. The first approach is to express the envelope (or squared-envelope) as a conditional density on Ω_v (or Ω_p), and then integrate over the density of Ω_v (or Ω_p) to obtain the composite distribution. Assuming that we are interested in the composite envelope

$$p_{\alpha_c}(x) = \int_0^\infty p_{\alpha|\,\Omega_v}(x|w)p_{\Omega_v}(w)dw \ . \tag{2.212}$$

For the case of Rayleigh fading

$$\Omega_v = \mathrm{E}[\alpha(t)] = \sqrt{\frac{\pi}{2}}\sigma \tag{2.213}$$

and, hence,

$$p_{\alpha|\,\Omega_v}(x|w) = \frac{\pi x}{2w^2}\exp\left\{-\frac{\pi x^2}{4w^2}\right\} \ . \tag{2.214}$$

The composite envelope distribution with Rayleigh fading and log-normal shadowing is

$$p_{\alpha_c}(x) = \int_0^\infty \frac{\pi x}{2w^2}\exp\left\{-\frac{\pi x^2}{4w^2}\right\} \tag{2.215}$$

$$\times \frac{2}{w\sigma_\Omega\xi\sqrt{2\pi}}\exp\left\{-\frac{\left(10\log_{10}w^2 - \mu_{\Omega_v\ (\mathrm{dBm})}\right)^2}{2\sigma_\Omega^2}\right\}dw \ .$$

where $\xi = (\ln 10)/10$. Sometimes this distribution is called a Susuki distribution, after the original work by Susuki [311]. The second approach, originally suggested by Lee and Yen [194], is to express the composite received signal as the product of the short term multipath fading and the long term shadow fading. Hence, at any time t, the envelope of the composite signal has the form

$$\hat{\alpha}_c(t) = \alpha(t) \cdot \Omega_v(t) \tag{2.216}$$

and the squared-envelope of the composite signal has the form

$$\hat{\alpha}_c^2(t) = \alpha^2(t) \cdot \Omega_p(t) \ . \tag{2.217}$$

Under the assumption that the fading and shadowing are independent random processes, we now demonstrate that both approaches lead to identical results. The density function of envelope in (2.216) can be obtained by using a bivariate transformation and then integrating to obtain the marginal density. This leads to the density

$$p_{\hat{\alpha}_c}(x) = \int_0^\infty \frac{1}{w}p_\alpha\left(\frac{x}{w}\right)p_{\Omega_v}(w)dw \ . \tag{2.218}$$

Again, consider the case of log-normal shadowing and Rayleigh fading. Using (2.43) and (2.199) gives

$$p_{\hat{\alpha}_c}(x) = \int_0^\infty \frac{x}{(w\sigma)^2}\exp\left\{-\frac{x^2}{2(w\sigma)^2}\right\} \tag{2.219}$$

$$\times \frac{2}{w\sigma_\Omega\xi\sqrt{2\pi}}\exp\left\{-\frac{\left(10\log_{10}w^2 - \mu_{\Omega_v\ (\mathrm{dBm})}\right)^2}{2\sigma_\Omega^2}\right\}dw \ .$$

Observe that (2.215) and (2.219) are related by

$$p_{\alpha_c}(x) = \sqrt{\frac{\pi}{2}}\sigma p_{\hat{\alpha}_c}\left(\sqrt{\frac{\pi}{2}}\sigma x\right) \tag{2.220}$$

It follows that the random variables $\hat{\alpha}_c$ and α_c are simply related through the linear transformation

$$\alpha_c = \sqrt{\frac{2}{\pi}}\frac{1}{\sigma}\hat{\alpha}_c \ . \tag{2.221}$$

Note, however, that $\sqrt{\pi/2}\sigma$ is just the mean of the Rayleigh distribution. Therefore, if we normalize $a(t)$ to have unit mean, then α_c and $\hat{\alpha}_c$ have the exact same distribution. Voila!

4.2.1 COMPOSITE GAMMA-LOG-NORMAL DISTRIBUTION

It is sometimes very useful to model the radio propagation environment as a shadowed Nakagami fading channel, because the Nakagami distribution is mathematically convenient and can closely approximate a Ricean distribution which in turn is often used to model a specular multipath fading channel. The composite distribution of the squared-envelope due to Nakagami fading and log-normal shadowing has the Gamma-log-normal density function

$$\begin{aligned}
p_{\alpha_c^2}(x) &= \int_0^\infty \left(\frac{m}{w}\right)^m \frac{x^{m-1}}{\Gamma(m)} \exp\left\{-\frac{mx}{w}\right\} \\
&\quad \times \frac{1}{\sqrt{2\pi}\xi\sigma_\Omega w} \exp\left\{-\frac{(10\log_{10}w - \mu_{\Omega_p \text{ (dBm)}})^2}{2\sigma_\Omega^2}\right\} dw
\end{aligned} \tag{2.222}$$

where $\xi = (\ln 10)/10$. As shown in Appendix 2B, the composite Gamma-log-normal distribution in (2.222) can be approximated by a log-normal distribution with mean and standard deviation

$$\begin{aligned}
\mu_{\text{(dBm)}} &= \xi^{-1}[\psi(m) - \ln(m)] + \mu_{\Omega_p \text{ (dBm)}} \\
\sigma^2 &= \xi^{-2}\zeta(2, m) + \sigma_\Omega^2
\end{aligned} \tag{2.223}$$

where $\psi(\cdot)$ is the Euler psi function and $\zeta(\cdot, \cdot)$ is Riemann's zeta function as defined in Appendix 2B. When $m = 1$ the approximation is valid for $\sigma_\Omega > 6$ dB, and for $m > 2$ the approximation is valid for all ranges of σ_Ω of interest [165]. The effect of Nakagami fading in (2.222) is to decrease the mean and increase the variance. However, this affect decreases as the shape factor m increases (corresponding to less severe fading). For example, with $m = 1$ (Rayleigh fading) we have $\mu_{\text{(new)}} = \mu_{\Omega_p \text{ (dBm)}} - 2.50675$ and $\sigma_{\text{(new)}}^2 = \sigma^2 + 31.0215$, while with $m = 8$ we have $\mu_{\text{(new)}} = \mu_{\Omega_p \text{ (dBm)}} - 0.277$ and $\sigma_{\text{(new)}}^2 = \sigma^2 + 2.50972$. We conclude that the effects of Nakagami fading become more pronounced when the shape factor m and the shadow standard deviation σ are small.

5.   PATH LOSS MODELS

It is well known that the received signal power decays with the square of the path length in free space. That is, the received envelope power is [257]

$$\mu_{\Omega_p} = \Omega_t G_T G_R \left(\frac{\lambda_c}{4\pi d}\right)^2 \tag{2.224}$$

where Ω_t is the transmitted power, G_T and G_R are the transmitter and receiver antenna gains, and d is the radio path length. The signals in land mobile radio applications, however, do not experience free space propagation. A more appropriate theoretical model assumes propagation over a flat reflecting surface (the earth) as shown in Fig. 2.43. In this case, the received envelope power is [257]

$$\mu_{\Omega_p} = 4\Omega_t \left(\frac{\lambda_c}{4\pi d}\right)^2 G_T G_R \sin^2\left(\frac{2\pi h_b h_m}{\lambda_c d}\right) \tag{2.225}$$

where h_b and h_m are the heights of the BS and MS antennas, respectively. Under the condition that $d \gg h_b h_m$, (2.225) reduces to

$$\mu_{\Omega_p} = \Omega_t G_T G_R \left(\frac{h_b h_m}{d^2}\right)^2 \tag{2.226}$$

where we have invoked the approximation $\sin x \approx x$ for small x. Observe that when $d \gg h_b h_m$, the propagation over a flat reflecting surface differs from free space propagation in two ways. First, the path loss is not frequency dependent and, second, the envelope power decays with the fourth power rather than the square of the distance. Fig. 2.44 plots the path loss

$$
\begin{aligned}
L_{p \text{ (dB)}} &= 10\log_{10}\left\{\frac{\Omega_t G_T G_R}{\mu_{\Omega_p}}\right\} \\
&= -10\log_{10}\left\{4\left(\frac{\lambda_c}{4\pi d}\right)^2 \sin^2\left(\frac{2\pi h_b h_m}{\lambda_c d}\right)\right\} \text{ dB} \tag{2.227}
\end{aligned}
$$

against the distance d. Notice that the path loss and, hence, the received envelope power has alternate minima and maxima when the path length is small. This property has been noted in experiments by Milstein *et. al.* [223]. The last local maxima in the path loss occurs when

$$\frac{2\pi h_b h_m}{\lambda_c d} = \frac{\pi}{2}.$$

5.1   PATH LOSS IN MACROCELLS

Several highly useful empirical models for macrocellular systems have been obtained by curve fitting experimental data. Two of the more useful models for 900 MHz cellular systems are Hata's model [253] based on Okumura's
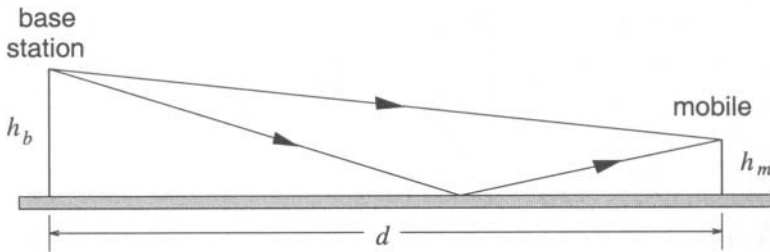
Figure 2.43. Radio propagation over a flat reflecting surface.
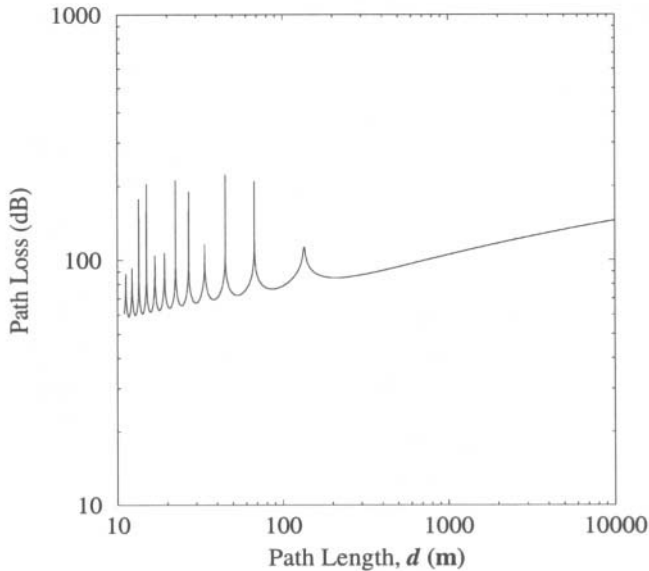
Figure 2.44. Propagation path loss with distance over a flat reflecting surface; $h_b = 7.5$ m, $h_m = 1.5$ m, $f_c = 1800$ MHz.

prediction method [162], and Lee's model [190].

5.1.1 OKUMURA-HATA AND CCIR MODELS

Hata's empirical model [162] is probably the simplest to use, and can distinguish between man-made structures. The empirical data for this model was collected by Okumura [253] in the city of Tokyo. Be cautioned, however, that the path losses for Japanese suburban areas do not match North American suburban areas very well. The latter are more like the quasi-open areas in Japan. Okumura and Hata's model is expressed in terms of the carrier frequency $150 \leq f_c \leq 1000$ (MHz), BS antenna height $30 \leq h_b \leq 200$ (m), the MS antenna height $1 \leq h_m \leq 10$ (m), and the distance $1 \leq d \leq 20$ (km)

between the BS and MS. The model is known to be accurate to within 1 dB for distances ranging from 1 to 20 km. With Okumura and Hata's model, the path loss between two isotropic BS and MS antennas is

$$L_{p \text{ (dB)}} = \begin{cases} A + B \log_{10}(d) & \text{for urban area} \\ A + B \log_{10}(d) - C & \text{for suburban area} \\ A + B \log_{10}(d) - D & \text{for open area} \end{cases} \quad (2.228)$$

where

$$\begin{aligned} A &= 69.55 + 26.16 \log_{10}(f_c) - 13.82 \log_{10}(h_b) - a(h_m) \quad (2.229) \\ B &= 44.9 - 6.55 \log_{10}(h_b) \\ C &= 5.4 + 2 \left[\log_{10}(f_c/28)\right]^2 \\ D &= 40.94 + 4.78 \left[\log_{10}(f_c)\right]^2 - 18.33 \log_{10}(f_c) \end{aligned}$$

and

$$a(h_m) = \begin{cases} (1.1 \log_{10}(f_c) - 0.7)\, h_m - (1.56 \log_{10}(f_c) - 0.8) \\ \qquad\qquad\qquad\qquad \text{for medium or small city} \\ \\ \begin{cases} 8.28 \left[\log_{10}(1.54 h_m)\right]^2 - 1.1 & \text{for } f_c \le 200 \text{ MHz} \\ 3.2 \left[\log_{10}(11.75 h_m)\right]^2 - 4.97 & \text{for } f_c \ge 400 \text{ MHz} \end{cases} \\ \qquad\qquad\qquad\qquad \text{for large city} \end{cases}$$

$$(2.230)$$

An empirical model was published by the CCIR that gives the path loss as

$$L_{p \text{ (dB)}} = A + B \log_{10}(d) - E \quad (2.231)$$

where A and B are defined in (2.229) with $a(h_m)$ being the medium or small city value in (2.230). The parameter E accounts for the degree of urbanization and is given by

$$E = 30 - 25 \log_{10}(\% \text{ of area covered by buildings}) \quad (2.232)$$

where $E = 0$ when the area is covered by approximately 16% buildings. Typical values from the Okumura-Hata "large city" model are plotted in Fig. 2.45, for a BS height of 70 m, a MS antenna height of 1.5 m, and a carrier frequency of 900 MHz. Several studies have shown that, due to a lesser degree of urbanization, the North American urban areas have path losses similar to the Japanese suburban areas.

5.1.2   LEE'S AREA-TO-AREA MODEL

Lee's area-to-area model [190] is used to predict a path loss over flat terrain. If the actual terrain is not flat, e.g., hilly, there will be large prediction errors. Two parameters are required for Lee's area-to-area model; the power at a 1 mile
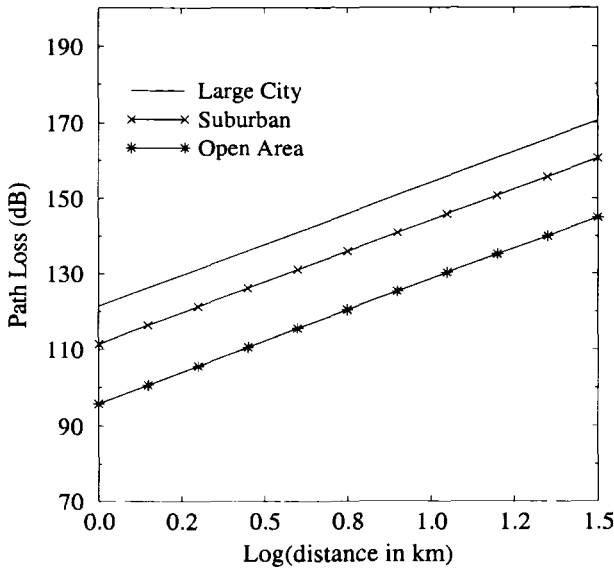
Figure 2.45. Path loss obtained by using Okumura and Hata's method; $h_b = 70$ m, $h_m = 1.5$ m, $f_c = 900$ Mhz.

(1.6 km) point of interception, $\mu_{\Omega_p}(d_o)$, and the path-loss exponent, β. The received signal power can be expressed as

$$\mu_{\Omega_p} = \mu_{\Omega_p}(d_o) \left(\frac{d_o}{d}\right)^{\beta} \left(\frac{f_c}{f}\right)^{n} \alpha_0 \tag{2.233}$$

where d is in kilometers and $d_o = 1.6$ km. The parameter α_0 is a correction factor used to account for different BS and MS antenna heights, transmit powers, and antenna gains. The following set of *nominal* conditions are assumed in Lee's area-to-area model:

- frequency $f_c = 900$ MHz

- BS antenna height = 30.48 m

- BS transmit power = 10 watts

- BS antenna gain = 6 dB above dipole gain

- MS antenna height = 3 m

- MS antenna gain = 0 dB above dipole gain

If the actual conditions are different from those listed above, then we compute the following parameters:

Terrain	$\mu_{\Omega_p}(d_o)$ (dBm)	β
Free Space	-45	2
Open Area	-49	4.35
North American Suburban	-61.7	3.84
North American Urban (Philadelphia)	-70	3.68
North American Urban (Newark)	-64	4.31
Japanese Urban (Tokyo)	-84	3.05

Table 2.6.   Parameters for Lee's area-to-area model in various propagation environments, from [190].

$$\alpha_1 = \left(\frac{\text{new BS antenna height (m)}}{30.48 \text{ m}}\right)^2$$

$$\alpha_2 = \left(\frac{\text{new MS antenna height (m)}}{3 \text{ m}}\right)^\kappa$$

$$\alpha_3 = \left(\frac{\text{new transmitter power}}{10 \text{ watts}}\right)^2$$

$$\alpha_4 = \frac{\text{new BS antenna gain with respect to } \lambda_c/2 \text{ dipole}}{4}$$

$$\alpha_5 = \text{ different antenna-gain correction factor at the MS} \quad (2.234)$$

From these parameters, the correction factor α_0 is

$$\alpha_0 = \alpha_1 \cdot \alpha_2 \cdot \alpha_3 \cdot \alpha_4 \cdot \alpha_5 . \qquad (2.235)$$

The parameters β and $\mu_{\Omega_p}(d_o)$ have been found from empirical measurements, and are listed in Table 2.6.

Experimental data suggest that n in (2.233) ranges between 2 and 3 with the exact value depending upon the carrier frequency and the geographic area. For $f_c < 450$ MHz in a suburban or open area, $n = 2$ is recommended. In an urban area with $f_c > 450$MHz, $n = 3$ is recommended. The value of κ in (2.234) can also be determined from empirical data

$$\kappa = \begin{cases} 2 & \text{for a new MS antenna height} > 10 \text{ m} \\ 3 & \text{for a new MS antenna height} < 3 \text{ m} \end{cases} . \qquad (2.236)$$

The path loss $L_{p \text{ (dB)}}$ is the difference between the transmitted and received field strengths, $L_{p \text{ (dB)}} = \mu_{\Omega_p \text{ (dBm)}}(d_o) - \mu_{\Omega_p \text{ (dBm)}}$. To compare with the Okumura-Hata model we must assume an isotropic BS antenna with 0 dB gain, so that $\alpha_4 = -6$ dB. Then by using the nominal BS transmitter power of 40 dBm (10 watts) along with the parameters in Tab. 2.6 for $\mu_{\Omega_p \text{ (dBm)}}(d_o)$ and β, the following path losses can be obtained
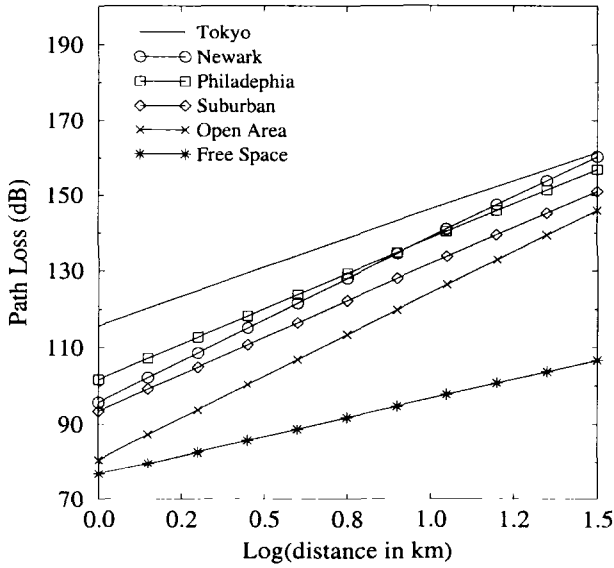
Figure 2.46. Path loss obtained by using Lee's method; $h_b = 70$ m, $h_m = 1.5$ m, $f_c = 900$ Mhz.

$$
L_{p \text{ (dB)}} = \begin{cases}
96.92 + 20.0 \log_{10} d + 10n \log_{10}(f/900) - \alpha_0 & \text{Free Space} \\
86.12 + 43.5 \log_{10} d + 10n \log_{10}(f/900) - \alpha_0 & \text{Open Area} \\
99.86 + 38.4 \log_{10} d + 10n \log_{10}(f/900) - \alpha_0 & \text{Suburban} \\
108.49 + 36.8 \log_{10} d + 10n \log_{10}(f/900) - \alpha_0 & \text{Philadelphia} \\
101.20 + 43.1 \log_{10} d + 10n \log_{10}(f/900) - \alpha_0 & \text{Newark} \\
123.77 + 30.5 \log_{10} d + 10n \log_{10}(f/900) - \alpha_0 & \text{Tokyo}
\end{cases}
$$

$$(2.237)$$

These typical values from Lee's area-to-area model are plotted in Fig. 2.46, for the same parameters used with Okumura-Hata model in Fig. 2.45.

5.2   PATH LOSS IN OUTDOOR MICROCELLS

Most of the future PCS microcellular systems are expected to operate in 1800-2000 MHz frequency bands. Some studies have suggested that the path losses experienced at 1845 MHz are about 10 dB larger than those experienced at 955 MHz when all other parameters are kept constant [68]. The COST231 study [69] has resulted in two models for urban microcellular propagation, the COST231-Hata model and the COST231-Walfish-Ikegami model.

5.2.1   COST231-HATA MODEL

The COST231-Hata model is based on the proposal by Mogensen [226]

et. al. to extend Okumura and Hata's model for use in the 1500-2000 MHz frequency range, where it is known that Okumura and Hata's model under estimates the path loss. The COST231-Hata model is expressed in terms of the carrier frequency $1500 \leq f_c \leq 2000$ **(MHz)**, BS antenna height $30 \leq h_b \leq 200$ **(m)**, MS antenna height $1 \leq h_m \leq 10$ **(m)**, and distance $1 \leq d \leq 20$ **(km)**. In particular, the path loss with the COST231-Hata model is

$$L_{p \text{ (dB)}} = A + B\log_{10}(d) + C \qquad (2.238)$$

where

$$A = 46.3 + 33.9\log_{10}(f_c) - 13.82\log_{10}(h_b) - a(h_m)$$
$$B = 44.9 - 6.55\log_{10}(h_b)$$
$$C = \begin{cases} 0 & \text{medium city and suburban areas} \\ & \text{with moderate tree density} \\ 3 & \text{for metropolitan centers} \end{cases}$$

Although both the Okumura and Hata and the COST231-Hata models are limited to BS antenna heights greater than 30 m, they can be used for lower BS antenna heights provided that the surrounding buildings are well below the BS antennas. They should not be used to predict path loss in urban canyons. The COST231-Hata model is good down to a path length of 1 km. It should not be used for smaller ranges, where path loss becomes highly dependent upon the local topography.

5.2.2   COST231-WALFISH-IKEGAMI MODEL

The COST231-Walfish-Ikegami model distinguishes between LoS and NLoS propagation.   The model is accurate for carrier frequencies in the range $800 \leq f_c \leq 2000$ **(MHz)**, and path distances in the range $0.02 \leq d \leq 5$ **(km)**.

LoS propagation:.   For LoS propagation in a street canyon, the path loss is

$$L_{p \text{ (dB)}} = 42.6 + 26\log_{10}(d) + 20\log_{10}(f_c), \quad d \geq 20 \text{ m} \qquad (2.239)$$

where the first constant is chosen so that L_p is equal to the free-space path loss at a distance of 20 m. The model parameters are the distance d (km) and carrier frequency f_c **(MHz)**.

NLoS propagation:.   As defined in Fig. 2.47, the path loss for non line-of-sight (NLoS) propagation is expressed in terms of the following parameters:

$$
\begin{aligned}
h_b &= \text{BS antenna height, } 4 \leq h_b \leq 50 \text{ (m)} \\
h_m &= \text{MS antenna height, } 1 \leq h_m \leq 3 \text{ (m)} \\
h_{\text{Roof}} &= \text{roof heights of buildings (m)} \\
\Delta h_b &= h_b - h_{\text{Roof}} = \text{height of BS relative to rooftops (m)}
\end{aligned}
$$
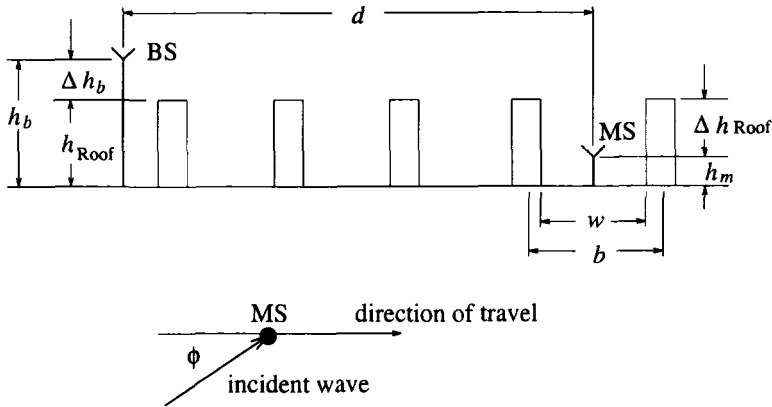
Figure 2.47. Definition of parameters used in the COST231-Walfish-Ikegami model.

$$\Delta h_m = h_{\text{Roof}} - h_m = \text{height of MS relative to rooftops (m)}$$
$$w = \text{width of streets (m)}$$
$$b = \text{building separation (m)}$$
$$\phi = \text{road orientation with respect to the direct radio path, degrees}$$

If no data on the structure of the buildings and roads are available, the following default values are recommended, $b = 20\ldots50$ (m), $w = b/2$, $\phi = 90°$, and $h_{\text{Roof}} = 3 \times$ number of floors + roof (m), where **roof** $= 3$ (m) pitched and 0 (m) flat.

The NLoS path loss is composed of three terms, viz.,

$$L_{p\,(\text{dB})} = \begin{cases} L_o + L_{\text{rts}} + L_{\text{msd}} & \text{for } L_{\text{rts}} + L_{\text{msd}} \geq 0 \\ L_o & \text{for } L_{\text{rts}} + L_{\text{msd}} < 0 \end{cases} \qquad (2.240)$$

where

$$L_o = \text{free-space loss} = 32.4 + 20\log_{10}(d) + 20\log_{10}(f_c)$$
$$L_{\text{rts}} = \text{roof-to-street diffraction and scatter loss}$$
$$L_{\text{msd}} = \text{multi-screen diffraction loss}$$

The roof-top-to-street diffraction and scatter loss is

$$L_{\text{rts}} = -16.9 - 10\log_{10}(w) + 10\log_{10}(f_c) + 20\log_{10}\Delta h_m + L_{\text{ori}} \quad (2.241)$$

where

$$L_{\text{ori}} = \begin{cases} -10 + 0.354(\phi), & 0 \leq \phi \leq 35° \\ 2.5 + 0.075(\phi - 35°), & 35° \leq \phi \leq 55° \\ 4.0 - 0.114(\phi - 55°), & 55° \leq \phi \leq 90° \end{cases} \qquad (2.242)$$

is an orientation loss.

The multi-screen diffraction loss is

$$L_{\mathrm{msd}} = L_{\mathrm{bsh}} + k_a + k_d \log_{10}(d) + k_f \log_{10}(f_c) - 9\log_{10}(b) \qquad (2.243)$$

where

$$L_{\mathrm{bsh}} = \begin{cases} -18\log_{10}(1 + \Delta h_b) & h_b > h_{\mathrm{Roof}} \\ 0 & h_b \le h_{\mathrm{Roof}} \end{cases} \qquad (2.244)$$

is the shadowing gain (negative loss) for cases when the BS antenna is above the rooftops. The parameters k_a and k_d depend on the path length, d, and base station elevation with respect to the rooftops Δh_b. The term k_a accounts for the increase in path loss when the BS antennas are situated below the roof tops of adjacent buildings, and is given by

$$k_a = \begin{cases} 54, & h_b > h_{\mathrm{Roof}} \\ 54 - 0.8\Delta h_b, & d \ge 0.5 \text{ km and } h_b \le h_{\mathrm{Roof}} \\ 54 - 0.8\Delta h_b d/0.5, & d < 0.5 \text{ km and } h_b \le h_{\mathrm{Roof}} \end{cases} \qquad (2.245)$$

The terms k_d and k_f control the dependency of the multi-screen diffraction loss on the distance and frequency, respectively, and are given by

$$k_d = \begin{cases} 18, & h_b > h_{\mathrm{Roof}} \\ 18 - 15\Delta h_b/h_{\mathrm{Roof}}, & h_b \le h_{\mathrm{Roof}} \end{cases} \qquad (2.246)$$

$$k_f = -4 + \begin{cases} 0.7(f_c/925 - 1), & \text{medium city and suburban} \\ 1.5(f_c/925 - 1), & \text{metropolitan area} \end{cases} \qquad (2.247)$$

The COST231-Walfish-Ikegami model works best for $h_b \gg h_{\mathrm{Roof}}$. Large prediction errors can be expected for $h_b \approx h_{\mathrm{Roof}}$. The model is poor for $h_b \ll h_{\mathrm{Roof}}$ because the terms in (2.245) do not consider wave guiding in street canyons and diffraction at street corners.

5.2.3   STREET MICROCELLS

For ranges less than 500 m and antenna heights less than 20 m, some empirical measurements have shown that the received signal strength for LoS propagation along city streets can be described by the two-slope model [161, 149, 175, 360, 268, 345]

$$\mu_{\Omega_p} = \frac{k\Omega_t}{d^a(1 + d/g)^b} \qquad (2.248)$$

where Ω_t is the transmitted power, k is a constant and d (m) is the distance. Close into the BS, free space propagation will prevail so that $a = 2$. The parameter g is called the break point and ranges from 150 to 300 m [161, 149, 175, 360]. At larger distances, an inverse-fourth to -eighth power law is experienced so that b ranges from 2 to 6. This is probably caused by increased shadowing at the greater distances [161]. The model parameters that were obtained by Harley [161] are listed in Table 2.7. Xia [366] has demonstrated that the break-point occurs where the Fresnel zone between the two antennas

Base Antenna Height (m)	a	b	Break point g (m)
5	2.30	-0.28	148.6
9	1.48	0.54	151.8
15	0.40	2.10	143.9
19	-0.96	4.72	158.3

Table 2.7. Two-slope path loss parameters obtained by Harley, from [161].
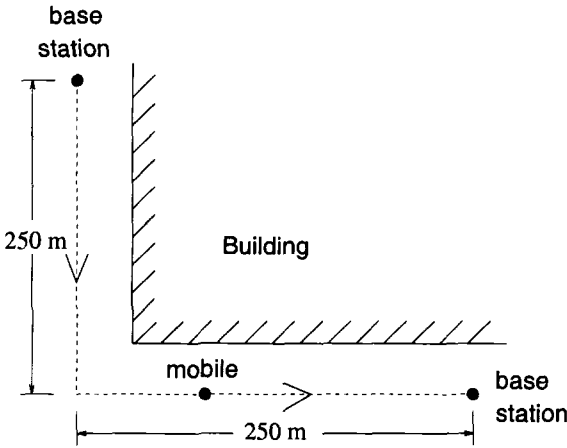
Figure 2.48. The corner effect in a street microcell environment.

just touches the ground assuming a flat surface. This distance is

$$g = \frac{1}{\lambda_c}\sqrt{(\Sigma^2 - \Delta^2)^2 - 2\,(\Sigma^2 + \Delta^2)\left(\frac{\lambda_c}{2}\right)^2 + \left(\frac{\lambda_c}{2}\right)^4} \qquad (2.249)$$

where $\Sigma = h_b + h_m$ and $\Delta = h_b - h_m$. For high frequencies this distance can be approximated as $g = 4h_b h_m/\lambda_c$. Notice that the break-point is dependent on frequency, with the break-point at 1.9 GHz being about twice that for 900 MHz.

Street microcells may also exhibit NLoS propagation when a MS rounds a street corner as shown in Fig. 2.48. In this case, the average received signal strength can drop by 25-30 dB over distances as small as 10 m for low antenna heights in an area with multi-story buildings [51, 324, 207, 238, 286], and by 25-30 dB over distances of 45-50 m for low antenna heights in a region with only one- or two-story buildings [286]. This phenomenon is called the **corner effect.**

Grimlund and Gudmundson [149] have proposed an empirical street corner path loss model. Their model assumes LoS propagation until the MS reaches a street corner. The NLoS propagation after rounding a street corner is modeled
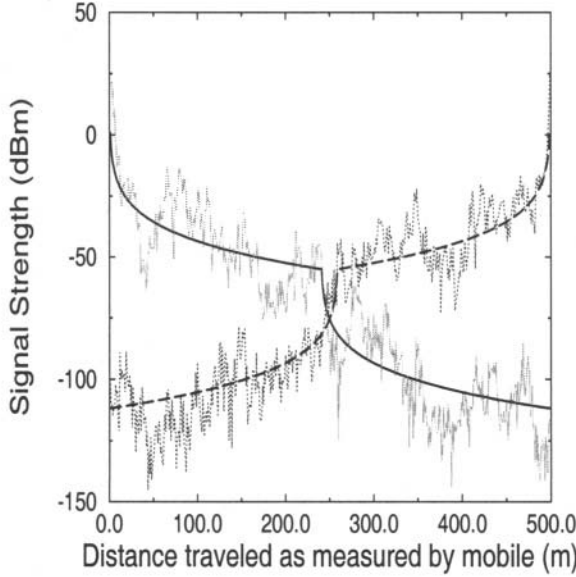
Figure 2.49.   Average and instantaneous received signal strength for the street microcell environment in Fig. 2.48. For the instantaneous received signal strength, $\sigma_\Omega = 6$ dB and $\phi_{\Omega_p \text{ (dBm)} \Omega_p \text{ (dBm)}}(d) = 0.1\sigma_\Omega^2$ at $d = 30$ m.

by assuming LoS propagation from an imaginary transmitter that is located at the street corner having a transmit power equal to the received power at the street corner from the serving BS. That is, the received signal strength (in dBm) is given by

$$
\mu_{\Omega_p} = \begin{cases} \frac{k\Omega_t}{d^a(1+d/g)^b} & d \le d_c \\ \frac{k\Omega_t}{d_c^a(1+d_c/g)^b} \cdot \frac{1}{(d-d_c)^a(1+(d-d_c)/g)^b} & d > d_c \end{cases} \tag{2.250}
$$

where d_c (m) is the distance between the serving BS and the corner. For the scenario depicted in Fig. 2.48, the received signal strength with this model is shown in Fig. 2.49. The heavy curves show the average received signal strength from the two BSs as the MS traverses the dashed path shown in Fig. 2.48. These curves were obtained by using $a = 2$, $b = 2$, $g = 150$ m, and $d_c = 250$ m in (2.250), and assuming that $\mu_{\Omega_p} = 1$ dBm at $d = 1$ m. The dotted curves superimposed on the heavy lines in Fig. 2.49 show the received signal strength with the combined effects of path loss, log-normal shadowing, and multipath-fading. The latter two were obtained by using the simulators described in Sections (4.1) and (3.2).

Building	Frequency (MHz)	β	σ_Ω (dB)
Retail stores	914	2.2	8.7
Grocery stores	914	1.8	5.2
Office, hard partition	1500	3.0	7.0
Office, soft partition	900	2.4	9.6
Office, soft partition	1900	2.6	14.1

Table 2.8. Path loss exponents and shadow standard deviations for several different types of buildings, from [11].

5.3 PATH LOSS IN INDOOR MICROCELLS

Indoor microcellular systems are becoming very important for providing wireless voice and data communications within the home and work-place. The characterization of in-building radio propagation is necessary for the effective deployment of these systems. In general, the path loss and shadowing characteristics vary greatly from one building to the next. Typical path loss exponents and shadow standard deviations are provided in Table 2.8 for several different types of buildings.

For multistory buildings, the RF attenuation between floors is important for frequency reuse on different floors of the same building. Measurements have indicated that the greatest floor loss occurs when the transmitter and receiver are separated by a single floor. Typically, the floor loss is 15 to 20 dB for one floor and an additional 6 to 10 dB per floor up to a separation of 4 floors. For 5 or more floors of separation, the overall floor loss will increase only a few dB for each additional floor. This effect is thought to be caused by signals diffracting up the sides of the building and signals scattering off the neighboring buildings. Also important for the deployment of indoor wireless systems is the building penetration loss. This loss depends on the frequency and height of the building. Turkmani *et. al.* [323] have shown that the building penetration losses decrease with increasing frequency, in particular they are 16.4, 11.6, and 7.6 dB at 441 MHz, 896.5 MHz, and 1400 MHz, respectively. In general the building penetration loss for signals propagating into a building tends to decrease with height, the reason being that a LoS path is more likely to exist at increased height. The building penetration loss decreases by about 2 dB per floor from ground level up to about 9 to 15 floors and then increases again [346]. Windows also have a significant effect on penetration loss. Plate glass provides an attenuation of about 6 dB, while lead lined glass provides an attenuation anywhere from 3 to 30 dB.

APPENDIX 2.A: Derivation of Equation (2.205)

This Appendix derives an expression for the second moment of a Ricean random variable in terms of its first moment. A Ricean random variable X has probability density function

$$p_X(x) = \frac{x}{\sigma^2} \exp\left\{-\frac{x^2 + s^2}{2\sigma^2}\right\} I_0\left(\frac{xs}{\sigma^2}\right) \quad x > 0 \qquad (2\text{-}2.\text{A}.1)$$

and moments [270]

$$E[X^n] = (2\sigma^2)^{\frac{n}{2}} \exp\left\{-\frac{s^2}{2\sigma^2}\right\} \Gamma\left((2+n)/2\right) \, {}_1F_1\left(\frac{n+2}{2}, 1; \frac{s^2}{2\sigma^2}\right) \tag{2-2.A.2}$$

where $\Gamma(\,\cdot\,)$ is the gamma function, and ${}_1F_1(a,\ b;\ x)$ is the confluent hypergeometric function. The first moment of X is

$$E[X] \equiv \Omega_v = (2\sigma^2)^{\frac{1}{2}} e^{-K} \frac{\sqrt{\pi}}{2} \, {}_1F_1(3/2, 1; K) \,, \tag{2-2.A.3}$$

where $K = s^2/2\sigma^2$ is the Rice factor. The second moment of X is

$$\begin{aligned} E[X^2] \equiv \Omega_p &= 2\sigma^2 e^{-K} \, {}_1F_1(2, 1; K) \\ &= 2\sigma^2 (K+1) \ . \end{aligned} \tag{2-2.A.4}$$

Substituting $2\sigma^2$ from (2-2.A.3) into (2-2.A.4) gives

$$\Omega_p = \frac{4 e^{2K} \, (K+1)}{\pi \, {}_1F_1^2(3/2, 1; K)} \, \Omega_v^2 = C(K) \, \Omega_v^2 \ . \tag{2-2.A.5}$$

Note that $C(0) = 4/\pi$, $C(\infty) = 1$, and $4/\pi \le C(K) \le 1$ for $0 \le K \le \infty$.

APPENDIX 2.B: Derivation of Equation (2.222)

From (2.222), the composite distribution for the squared envelope α_c^2 is

$$\begin{aligned} p_{\alpha_c^2}(x) &= \int_0^\infty \left(\frac{m}{w}\right)^m \frac{x^{m-1}}{\Gamma(m)} \exp\left\{-\frac{mx}{w}\right\} \\ &\quad \times \frac{1}{\sqrt{2\pi}\xi\sigma_\Omega w} \exp\left\{-\frac{(10\log_{10}w - \mu_{\Omega_{p\ \text{(dB)}}})^2}{2\sigma_\Omega^2}\right\} dw \ . \end{aligned} \tag{2-2.B.1}$$

where $\xi = (\ln 10)/10$. The mean of the approximate log-normal distribution is

$$\mu_{(\text{dBm})} = \text{E}[10\log_{10}(\alpha_c^2)]$$

$$= \int_0^\infty \int_0^\infty 10\log_{10}(x) \left(\frac{m}{w}\right)^m \frac{x^{m-1}}{\Gamma(m)} \exp\left\{-\frac{mx}{w}\right\}$$

$$\times \frac{1}{\sqrt{2\pi}\xi\sigma_\Omega w} \exp\left\{-\frac{(10\log_{10}w - \mu_{\Omega_p \text{ (dB)}})^2}{2\sigma_\Omega^2}\right\} dwdx$$

$$= \frac{10m^m}{\sqrt{2\pi}\xi\sigma_\Omega\Gamma(m)} \int_0^\infty \frac{1}{w^{m+1}} \exp\left\{-\frac{(10\log_{10}w - \mu_{\Omega_p \text{ (dB)}})^2}{2\sigma_\Omega^2}\right\}$$

$$\times \int_0^\infty \log_{10}(x) \, x^{m-1} e^{-\frac{mx}{w}} \, dxdw \quad . \tag{2-2.B.2}$$

Assuming that m is an integer, the inner integral becomes [147, 4.352.2]

$$\int_0^\infty \log_{10}(x) x^{m-1} e^{-\frac{mx}{w}} \, dx = \frac{\Gamma(m)w^m}{m^m \ln 10} [\psi(m) - \ln(m/w)] \quad . \tag{2-2.B.3}$$

Then by using the change of variables $x = 10\log_{10}(w)$ we obtain

$$\mu_{(\text{dBm})} = \xi^{-1}[\psi(m) - \ln(m)] + \mu_{\Omega_p \text{ (dB)}} \tag{2-2.B.4}$$

where $\psi(\,\cdot\,)$ is the Euler psi function, and

$$\psi(m) = -C + \sum_{k=1}^{m-1} \frac{1}{k} \tag{2-2.B.5}$$

and $C \simeq 0.5772$ is Euler's constant. Likewise, the second moment of the approximate log-normal distribution is

$$\text{E}[(10\log_{10}(\alpha_c^2))^2]$$

$$= \int_0^\infty \int_0^\infty [10\log_{10}(x)]^2 \left(\frac{m}{w}\right)^m \frac{x^{m-1}}{\Gamma(m)} \exp\left\{-\frac{mx}{w}\right\} \frac{1}{\sqrt{2\pi}\xi\sigma_\Omega w}$$

$$\times \exp\left\{-\frac{(10\log_{10}w - \mu_{\Omega_{p(\text{dB})}})^2}{2\sigma_\Omega^2}\right\} dwdx$$

$$= \frac{m^m}{\sqrt{2\pi}\xi\Gamma(m)} \int_0^\infty \frac{1}{w^{m+1}} \exp\left\{-\frac{(10\log_{10}w - \mu_{\Omega_p \text{ (dB)}})^2}{2\sigma_\Omega^2}\right\}$$

$$\times \int_0^\infty [10\log_{10}(x)]^2 x^{m-1} e^{-\frac{mx}{w}} \, dxdw \quad . \tag{2-2.B.6}$$

Assuming again that m is an integer, the inner integral is [147, 4.358.2]

$$\int_0^\infty [10\log_{10}(x)]^2 x^{m-1} e^{-\frac{mx}{w}} \, dx = \frac{(m-1)!w^m}{m^m \ln 10} \tag{2-2.B.7}$$

$$\times \left([\psi(m) - \ln(m/w)]^2 + \zeta(2,m)\right)$$

leading to

$$
\begin{aligned}
E[(10\log_{10}(\alpha_c^2))^2] &= \zeta^2\left([\psi(m)-\ln(m)]^2\mu_{\Omega_p\ (dB)}^2 + \zeta(2,m)\right)\\
&+ 2\zeta[\psi(m)-\ln(m)]\mu_{\Omega_p\ (dB)} + \sigma^2 + \mu_{\Omega_p\ (dB)}^2 \ .
\end{aligned}
$$

$$\text{(2-2.B.8)}$$

where

$$
\zeta(2,m) = \sum_{k=0}^{\infty}\frac{1}{(m+k)^2}
\qquad\qquad \text{(2-2.B.9)}
$$

is Reimann's zeta function. Finally, the variance of the approximate log-normal distribution is

$$
\begin{aligned}
\sigma^2 &= E[(10\log_{10}(\alpha_c^2))^2] - E^2[10\log_{10}(\alpha_c^2)]\\
&= \xi^{-2}\zeta(2,m) + \sigma_{\Omega}^2 \ .
\end{aligned}
$$

$$\text{(2-2.B.10)}$$

Problems

2.1. Suppose that $r(t)$ is a stationary band-pass random process

$$
r(t) = g_I(t)\cos 2\pi f_c t - g_Q(t)\sin 2\pi f_c t
$$

Show that the autocorrelation of $r(t)$ is

$$
\begin{aligned}
E[r(t)r(t+\tau)] &= \phi_{g_I g_I}(\tau)\cos 2\pi f_c \tau\\
&- \phi_{g_Q g_I}(\tau)\sin 2\pi f_c \tau
\end{aligned}
$$

2.2. Suppose that a vertical monopole antenna is used and the pdf of arriving plane waves, $p(\theta)$ is given by (2.41). Find the band-pass Doppler power spectrum $S_{rr}(f)$.

2.3. Determine and plot the (normalized) power spectral densities $S_{rr}(f)$ for the following cases. Assume 2-D isotropic scattering;

   a) A vertical loop antenna in the plane perpendicular to vehicle motion, $G(\theta)=\frac{3}{2}\sin^2\theta$.

   b) A vertical loop antenna in the plane of vehicle motion, $G(\theta)=\frac{3}{2}\cos^2\theta$.

   c) A directional antenna of beamwidth β directed perpendicular to vehicle motion with (see Fig 2.B.l(a))

$$
G(\theta)=\begin{cases} G_0, & |\frac{\pi}{2}-\theta| < \beta/2\\ 0, & \text{otherwise}\end{cases}
$$

   d) A directional antenna of beamwidth β directed along vehicle motion with (see Fig. 2.B.l(b))

$$
G(\theta)=\begin{cases} G_0, & |\theta| < \beta/2\\ 0, & \text{otherwise}\end{cases}
$$
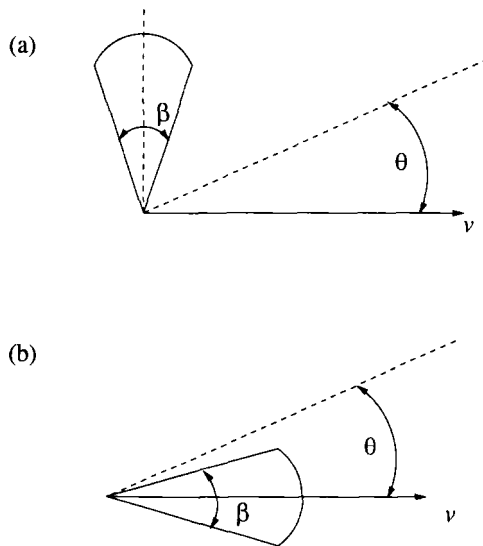
Figure 2.B.1. Scenario for Problem 2.3 parts (c) and (d).

2.4. Consider a 2-D isotropic scattering channel. Show that the psd of the received envelope $\alpha(t) = |g(t)|$ is given by (2.65).

2.5. Consider the non-isotropic scattering environment shown in Fig. 2.5.

a) Show that the psd of $g(t)$ is given by (2.67).

b) Show that the psd of the received envelope $\alpha(t) = |g(t)|$ is given by (2.68).

2.6. Consider a wide-sense stationary zero-mean complex Gaussian random process $g(t)$ having the autocorrelation function $\phi_{gg}(\tau) = \phi_{g_I g_I}(\tau) + j\phi_{g_I g_Q}(\tau)$. Show that the autocorrelation and autocovariance functions of the squared-envelope $\alpha^2(t) = |g(t)|^2$ are given by (2.71) and (2.72), respectively.

2.7. Consider a wide-sense stationary non zero-mean complex Gaussian random process $g(t) = g_I(t) + jg_Q(t)$, where

$$g_I(t) = \hat{g}_I(t) + m_I(t)$$
$$g_Q(t) = \hat{g}_Q(t) + m_Q(t)$$

and $m_I(t)$ and $m_Q(t)$ are the means of $g_I(t)$ and $g_Q(t)$, respectively. Show that the autocorrelation and autocovariance functions of the squared-envelope $\alpha^2(t) = |g(t)|^2$ are given by (2.76) and (2.79), respectively.

2.8 Establish the equivalence between (2.92) and (2.93).
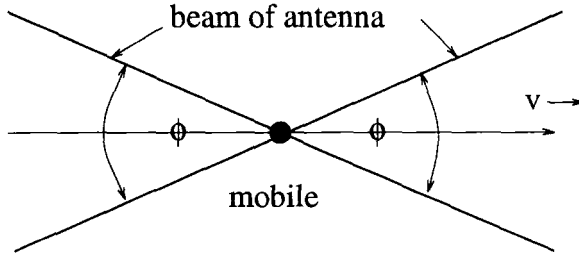
Figure 2.B.2.   Mobile with directional antenna for Problem 2.10.

2.9. Consider a situation where the received envelope is Rayleigh faded ($K = 0$), but the Doppler power spectrum $S_{gg}^c(f)$ is not symmetrical about $f = 0$, i.e., a form of non-isotropic scattering. Show that the envelope level crossing rate is given by

$$L_R = \sqrt{\frac{b_2}{b_0} - \frac{b_1^2}{b_0^2}} \cdot \frac{\rho}{\sqrt{\pi}} e^{-\rho^2}$$

where

$$\rho = \frac{R}{\sqrt{\Omega_p}} = \frac{R}{\sqrt{2b_0}} \quad .$$

and the b_i are defined in (2.92) with $f_q = 0$.

2.10. Consider the situation in the Fig. 2.B.2, where the MS employs a directional antenna with a beam width of $\phi°$. Assume a 2-D isotropic scattering environment.

  a) In receiving a radio transmission at 850 MHz, a Doppler frequency of 20 to 60 Hz is observed. What is the beam width of the MS antenna, and how fast is the MS traveling?

  b) Suppose that the MS antenna has a beam width of 13°. What is the level-crossing rate with respect to the *rms* envelope level, assuming that the MS is traveling at a speed of 30 km/h?

2.11. A vehicle experiences 2-D isotropic scattering and receives a Rayleigh faded 900 MHz signal while traveling at a constant velocity for 10 s. The average duration of fades 10 dB below the rms envelope level is 1 ms. How far does the vehicle travel during the 10 s interval? How many fades is the envelope expected to undergo that are 10 dB below the rms envelope level during the 10 s interval? Assume that the local mean remains constant during travel.

2.12. A vehicle receives a Ricean faded signal where the specular component is at the frequency f_c and scatter component is due to 2-D isotropic scattering.

a) Compute the average duration of fades that 10 dB below the rms envelope level for $K = 0, 7, 20$, and a maximum Doppler frequency of $f_m = 20$ Hz.

b) Suppose that data is transmitted using binary modulation at a rate of 1 Mbps, and an envelope level that is 10 dB below the rms envelope level represents a threshold between "error-free" and "error-prone" conditions. During error-prone conditions, bits are in error half the time. Assuming that the data is transmitted in 10,000 bit packets, how many bits errors (on the average) will each transmitted packet contain?

2.13. Show that for wide sense stationary (WSS) channels

$$\phi_H(f, m; \nu, \mu) = \psi_H(f, m; \nu)\delta(\nu - \mu)$$
$$\phi_S(\tau, \eta; \nu, \mu) = \psi_S(\tau, \eta; \nu)\delta(\nu - \mu) .$$

That is the channel correlation functions $\phi_H(f, m; \nu, \mu)$ and $\phi_S(\tau, \eta; \nu, \mu)$ have a singular behavior with respect to the Doppler shift variable. What is the physical interpretation of this property?

2.14. Show that for uncorrelated scattering (US) channels

$$\phi_g(t, s; \tau, \eta) = \psi_g(t, s; \tau)\delta(\eta - \tau)$$
$$\phi_S(\tau, \eta; \nu, \mu) = \psi_S(\tau; \nu, \mu)\delta(\eta - \tau) .$$

That is the channel correlation functions $\phi_g(t, s; \tau, \eta)$ and $.\phi_S(\tau, \eta; \nu, \mu)$ have a singular behavior with respect to the time delay variable. What is the physical interpretation of this property?

2.15. Consider the COST-207 typical urban (TU) and bad urban (BU) power delay profiles shown in Fig. 2.39 of the text with delays and fractional powers given in Tab. 2.1.

a) Calculate the average delay, μ_τ.
b) Calculate the *rms* delay spread, σ_τ.
c) Calculate the approximate values of W_{50} and W_{90}
d) If the channel is to be used with a modulation that requires an equalizer whenever the symbol duration $T < 10\sigma_\tau$, determine the maximum symbol rate that can be supported without requiring an equalizer.

2.16. The scattering function $\psi_S(\tau, \nu)$ for a multipath fading channel is non-zero for the range of values $0 \leq \tau \leq 1$ μs and $-40 \leq \lambda \leq 40$ Hz. Furthermore, $\psi_S(\tau, \nu)$ is uniform in the two variables τ and ν.

a) Find numerical values for the following parameters;
 1. the average delay, μ_τ, and *rms* delay spread, σ_τ
 2. the Doppler spread, B_d

    3. the approximate coherence time, T_c

    4. the approximate coherence bandwidth, B_c

  b) Given the answers in part a), what does it mean when the channel is

    1. frequency-nonselective

    2. slowly fading

    3. frequency-selective

2.17. Suppose that the fading gain is modeled as a Markov process with state equation given by (2.156).

  a) What are the probability density functions of the envelope magnitude

$$\alpha_k = |g_k|$$

  and envelope phase

$$\theta_k = \mathrm{Tan}^{-1}\frac{g_{Q,k}}{g_{I,k}}$$

  at any epoch k.

  b) Derive the discrete autocorrelation function

$$\phi_{g_I g_I}(n) = \mathrm{E}[g_{I,k}g_{I,k+n}] = \mathrm{E}[g_{Q,k}g_{Q,k+n}]$$

  and discrete crosscorrelation function

$$\phi_{g_I g_Q}(n) = \mathrm{E}[g_{I,k}g_{Q,k+n}]$$

2.18. Consider Jakes' fading simulator shown in Fig. 2.30.

  a) With the choice that $\alpha = 0$ and $\beta_n = \pi n/(M+1)$ show that

$$
\begin{aligned}
<g_I(t)g_Q(t)> &= 0 \\
<g_Q^2(t)> &= M+1 \\
<g_I^2(t)> &= M
\end{aligned}
$$

  b) Rederive the time averages in part a) for the choice $\alpha = 0$ and $\beta_n = \pi n/M$.

2.19. (computer exercise) You are to write a software fading simulator that uses Jakes' method and plot typical sample functions of the faded envelope. By scaling $g_I(t)$ and $g_Q(t)$ appropriately, generate a Rayleigh faded envelope having the mean-squared envelope $\Omega_p = 1$. Plot a sample function of your faded envelope assuming a maximum Doppler frequency of $f_m T = 0.1$, where T is the simulation step size.

2.20. (computer exercise) In this problem you are to generate Ricean faded by making appropriate modifications to Jakes' Rayleigh fading simulator in Figure 2.30. Assume that the means $m_I(t)$ and $m_Q(t)$ of $g_I(t)$ and $g_Q(t)$, respectively, are generated according to Aulin's model in (2.47) and (2.48). For $f_m T = 0.1$, $\Omega_p = 1$ and $K = 0, 4, 7,$ and 16, plot the following

122

a) The envelope $\alpha(t) = \sqrt{g_I^2(t) + g_Q^2(t)}$.

b) The wrapped phase $\phi(t) = \text{Tan}^{-1}(g_Q(t)/g_I(t))$, mod 2π.

2.21. (computer exercise) This problem uses the fading simulator you developed in Problem 2.20. We now want to compute an estimate of the mean-squared envelope $\Omega_p = \text{E}[\alpha^2(t)]$ from samples of $g_I(kT)$ and $g_Q(kT)$, where T is the sample spacing in seconds. The estimate is computed by forming the empirical average

$$\hat{\Omega}_p = \frac{1}{N} \sum_{i=1}^{N} \left(g_I^2(iT) + g_Q^2(iT)\right)$$

where NT is the window averaging length in seconds. Assuming a constant velocity, the distance the MS moves (in units of wavelengths) in a time of NT seconds is

$$\frac{d}{\lambda_c} = Nf_mT \ .$$

a) For $K = 0, 4, 7$, and 16, generate 1000 estimates of the of Ω_p by using non-overlapping averaging windows of size

$$N = 50, 100, 150, 200, 250, 300 \ .$$

Construct a graph that plots, for each K, the sample variance of the Ω_p estimate on the ordinate and the window size on the abscissa.

b) Can you draw any qualitative conclusions from part a)?

Note: Samples of the local mean Ω_p are often used in handoff algorithms.

2.22. Consider a scattering environment where it is known that no plane waves arrive from either directly ahead or directly behind the direction of motion. We are interested in constructing a fading simulator similar to Jakes' method to account for this fact.

a) How might you modify Jakes' method to account for the above situation, assuming that you only need to generate one faded signal?

b) Assume that the received complex envelope has the form

$$g(t) = \sum_{n=1}^{N} e^{j(2\pi f_m t \cos \theta_n) + \hat{\phi}_n}$$

where f_m is the maximum Doppler frequency, $\hat{\phi}_n$ is the random phase of the nth component, and $\theta_n = 2\pi(n - .5)/N, n = 1, \ldots, N$ is the angle of arrival for the nth component. Following the method used for deriving Jakes' fading simulator and assuming that TV/4 is even, show that $g(t)$ can be written in the form

$$g(t) = K \sum_{n=1}^{M} [\cos(\beta_n) + j \sin(\beta_n)] \cos(2\pi f_m t \cos \theta_n t + \gamma_n)$$

where $M = N/4$.

1. What are the values of β_n and $-\gamma_n$ in terms of the phases $\hat{\phi}_n$?
2. Determine K so that $E[|g(t)|^2] = 1$.
3. Assuming that $\beta_n = \pi n/M$, what is the crosscorrelation between the real and imaginary parts of $g(t)$? Is this a desirable result for the simulator?

2.23. As shown in Fig. 2.34, Jakes' approach can be used to generate M faded envelopes $g_k(t)$, $k = 1, \ldots, M$, according to

$$g_k(t) = K \sum_{n=1}^{M} [\cos(\beta_n) + j \sin(\beta_n)] \cos(2\pi f_n t + \theta_{nk})$$

where K is a normalization constant, $f_n = f_m \cos \theta_n$, $\theta_n = 2\pi n/N$, $\beta_n = \pi n/(M+1)$ and $\theta_{nk} = \beta_n + 2\pi(k-1)n/(M+1)$.

a) What are some of the problems with this technique?

b) It is claimed that this method can generate faded envelopes $g_j(t)$ and $g_k(t)$ that are almost uncorrelated for arbitrary j and k provided that

$$\theta_{nj} - \theta_{nk} = i\pi + \pi/2$$

for some integer i; otherwise, the correlation between certain pairs of faded envelopes may be significant. Justify whether this claim true or false.

2.24. (computer exercise) It is claimed that the modified Jakes's method in (2.181) can yield uncorrelated faded envelopes, because the rows of the Hadamard matrix are orthogonal.

a) By using analytical methods determine whether or not this method can yield faded envelopes having a crosscorrelation of zero at a lag of zero, i.e., $\frac{1}{2}E[g_k^*(t)g_j(t)] = 0$, for $k \neq j$.

b) Write a software fading simulator to implement the modified Jakes' method with $M = 8$ oscillators.

1. Plot the normalized autocorrelation of $g_i(t)$, defined in (2.175).
2. Choosing two of the faded envelopes at random, plot the normalized crosscorrelation between $g_k(t)$ and $g_j(t)$, defined in (2.179).

2.25. (computer exercise) In this problem we want to generate variations in the local mean Ω_p due to shadowing. The shadows are generated according to the state equation in (2.207).

a) Suppose that the simulation step size is $T = 0.1$ s and the MS velocity is $v = 30$ km/h. We want a shadow decorrelation of 0.1 at a distance

b) Using the value of ξ obtained in part a) and a shadow standard deviation of $\sigma_\Omega = 8$ dB, plot a graph of $\Omega_{p\ (dB)}$ against the *distance* traveled. Scale your plot so the distance traveled goes from 0 to 100 m.

2.26. The measured path loss at a distance of 10 km in the city of Tokyo is 160 dB. The test parameters used in the experiment were the following:

- BS antenna height $h_b = 30$ m
- MS antenna height $h_m = 3$ m
- carrier frequency $f_c = 1,000$ MHz
- isotropic BS and MS antennas.

Compare the measured path loss with the predicted path loss from Okumura and Hata's model and Lee's model.

Note: If any model parameters are undefined, then use the default values.

2.27. Consider Fig. 2.B.3 and the following data

- The symbol transmission rate is 24300 symbols/s with 2 bits/symbol
- The channel bandwidth is 30 kHz
- The propagation environment is characterized by an rms delay spread of $\sigma_\tau = 1$ ns

A MS is moving from base station A (BSA) to base station B (BSB). Base station C (BSC) is a co-channel base station with BSA.

Explain how you would construct a computer simulator to model the received signal power at the MS from (BSA) and (BSC), as the MS moves from BSA to BSB. Clearly state your assumptions and explain the relationship between the propagation characteristics in your model.
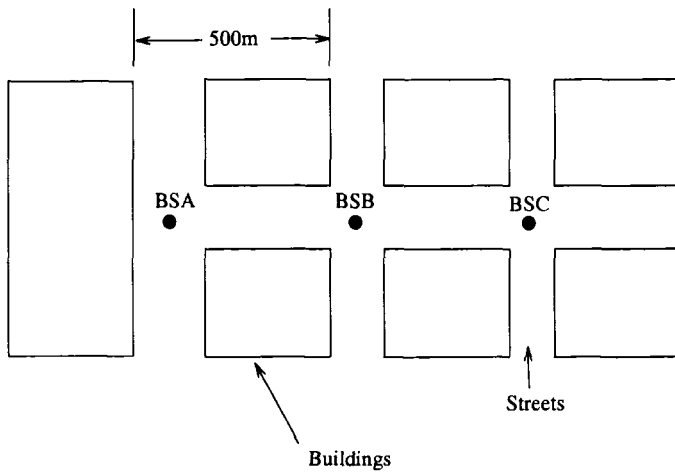
Figure 2.B.3.   Base station and street layout for Problem 2.27.

This page intentionally left blank

Chapter 3

CO-CHANNEL INTERFERENCE

For cellular radio systems the radio link performance is usually limited by interference rather than noise and, therefore, the probability of ourage due to co-channel interference outage, O_I, is of primary concern. Since this chapter deals with co-channel interference there is no need to distinguish between thermal noise and co-channel interference outages. For the remainder of the chapter, the probability of outage refers to the probability of co-channel interference outage. The definition of the outage probability depends on the assumptions made about the radio receiver and propagation environment. At higher velocities, the radio receiver can usually average over the fast envelope variations by using coding and interleaving techniques. In this case, the transmission quality will be acceptable provided that the *average* received carrier-to-interference ratio, A, exceeds a receiver threshold Λ_{th}. The receiver threshold Λ_{th} is determined by the performance of the radio link in the presence of envelope fading. Once Λ_{th} has been determined, the variations in Λ due to path loss and shadowing will determine the outage probability. At lower velocities, the radio receiver cannot average over the fast envelope variations due to the delay constraints imposed by voice traffic. In this case, the transmission quality will be acceptable provided that the *instantaneous* received carrier-to-interference ratio, λ, exceeds another receiver threshold λ_{th}[1]. Once λ_{th} has been specified, variations in λ due to path loss, shadowing, *and envelope fading,* will determine the outage probability.

The effect of co-channel interference on the radio link performance depends on the ability of the radio receiver to reject co-channel interference. Some of the more advanced receivers incorporate sophisticated signal processing methods for the rejection or cancellation of co-channel interference, e.g., equalization

---

[1]Note that Λ_{th} and λ_{th} are not the same.

and interference cancellation techniques. In this case, the radio receiver is more tolerant to co-channel interference and the receiver thresholds Λ_{th} and λ_{th} are reduced. This will reduce the outage probability.

Evaluating the outage probability for the log-normally shadowed signals that are typically found in cellular frequency reuse systems requires the probability distribution of the interference power that is accumulated from the sum several log-normal signals. Although there is no known exact expression for the probability distribution for the sum of log-normally random variables, several approximations have been derived by various authors. All of their approaches approximate the sum of log-normal random variables by another log-normal random variable. A method that matches the first two moments of the approximation has been developed by Fenton [118]. Sometimes Wilkinson is credited with this method, as in [295]. Here we called it the Fenton-Wilkinson method. Schwartz and Yeh developed another log-normal approximation that uses the exact first two moments for the sum of two log-normal random variables [295]. The Schwartz-and-Yeh method generally provides a more accurate approximation than the Fenton-Wilkinson method but it is more difficult to use. Prasad has corrected some errors in Schwartz and Yeh's paper in [264]. Another log-normal approximation is the cumulants matching approach suggested by Schleher [293]. With this approach, different log-normal approximations are applied over different ranges of the composite distribution. A good comparison of the methods of Fenton-Wilkinson, Schwartz-and-Yeh, Parley, and Schleher has been undertaken by Beaulieu, Abu-Dayya, and McLane [28].

The above log-normal approximations have been extensively applied to the calculation of the probability of outage in cellular systems. For example, Fenton's approach has been applied by Nagata and Akaiwa [240], Cox [74], Muammar and Gupta [235], and Daikoku and Ohdate [75]. Likewise, the Schwartz-and-Yeh approach has been applied by Yeh and Schwartz [372], Prasad and Arnbak [264], and Prasad, Kegel, and Arnbak [266].

Current literature also provides a thorough treatment of the probability of outage when the signals are affected by fading only, including the work of Yao and Sheikh [369], Muammar [234], and Prasad and Kegel [265]. Section 3. shows that the probability of outage is sensitive to the Rice factor of the desired signal, but it is insensitive to the number of interferers provided that the total interfering power remains constant. Calculations of the probability of outage for signals with composite log-normal shadowing and fading have considered the cases of Rayleigh fading by Linnartz [203], Nakagami fading by Ho and Stüber [165], and Ricean fading by Austin and Stüber [24]. Sections 4. and 5. show that shadowing has a more significant effect on the probability of outage than fading. Furthermore, the probability of outage is dominated by fading of the desired signal rather than fading of the interfering signals, e.g., with Ricean fading, the probability of outage is sensitive to the Rice factor of the

desired signal but is insensitive to the Rice factor of interfering signals. Finally, all of the above references assume a channel characterized by frequency non-selective (flat) fading. If the channel exhibits frequency selective fading, then the same general methodology can be used but the instantaneous carrier-to-interference ratio, λ, must be appropriately defined. The proper definition depends on the type of receiver that is employed, e.g., a maximum likelihood sequence estimation (MLSE) receiver for TDMA systems.

Most of the literature dealing with the probability of outage assumes that the interfering co-channel signals add noncoherently. The probability of outage has also been evaluated by Prasad and Kegel [267,265] for the case of coherent addition of Rayleigh faded co-channel interferers and a Ricean faded desired signal. The coherent co-channel interferers are assumed to arrive at the receiver antenna with the same carrier phase. However, as discussed by Prasad and Kegel [267] and Linnartz [203], it is more realistic to assume noncoherent addition of co-channel interferers in mobile radio systems because of the scattering environment. Coherently addition of co-channel interferers generally leads to pessimistic predictions of the probability of outage.

The remainder of this chapter begins in Section 1. where approximations are derived for the sum of multiple log-normally shadowed interferers. The various approximations are compared in terms of their accuracy. Section 2. derives the probability of outage with multiple log-normal interferers. Section 3. considers the outage probability for multiple Rayleigh or Ricean faded interferers without shadowing. Sections 4. and 5. do the same for multiple log-normally shadowed Nakagami faded interferers, and for multiple log-normally shadowed Ricean faded interferers, respectively.

1.   MULTIPLE LOG-NORMAL INTERFERERS

Consider the sum of N_I log-normal random variables

$$I = \sum_{k=1}^{N_I} \Omega_k = \sum_{k=1}^{N_I} 10^{\Omega_{k\,(\mathrm{dBm})}/10} \tag{3.1}$$

where the $\Omega_{k\,(\mathrm{dBm})}$ are Gaussian random variables with means $\mu_{\Omega_{k\,(\mathrm{dBm})}}$ and variances $\sigma_{\Omega_k}^2$, and the $\Omega_k = 10^{\Omega_{k(\mathrm{dBm})}/10}$ are the log-normal random variables. Unfortunately, there is no known closed form expression for the probability density function (pdf) of the sum of multiple $(N_I \geq 2)$ log-normal random variables. However, there is a general consensus that the sum of independent log-normal random variables can be approximated by another log-normal random variable with appropriately chosen parameters. That is,

$$I = \sum_{k=1}^{N_I} 10^{\Omega_{k(\mathrm{dBm})}/10} \approx 10^{Z_{(\mathrm{dBm})}/10} = \hat{I} \tag{3.2}$$

where $Z_{(dBm)}$ is a Gaussian random variable with mean $\mu_{Z~(dBm)}$ and variance σ_Z^2. The problem is to determine $\mu_{Z~(dBm)}$ and σ_Z^2 in terms of the $\mu_{\Omega_k~(dBm)}$ and $\sigma_{\Omega_k}^2$, $k = 1, \ldots, N_I$. Several methods have been suggested in the literature to solve this problem including those by Fenton [118], Schwartz and Yen [295], and Farley [295]. Each of these methods provides varying degrees of accuracy over specified ranges of the shadow standard deviation σ_Ω, the sum I, and the number of interferes N_I.

1.1 FENTON-WILKINSON METHOD

With the Fenton-Wilkinson method, the mean $\mu_{Z~(dBm)}$ and variance σ_Z^2 of $Z_{(dBm)}$ are obtained by matching the first two moments of the sum I with the first two moments of the approximation $\hat{I}$. To derive the appropriate moments, it is convenient to use natural logarithms. We write

$$\Omega_k = 10^{\Omega_k~(dBm)/10} = e^{\xi \Omega_k~(dBm)} = e^{\hat{\Omega}_k} \tag{3.3}$$

where $\xi = (\ln 10)/10 = 0.23026$ and $\hat{\Omega}_k = \xi \Omega_k~(dBm)$. Note that $\mu_{\hat{\Omega}_k} = \xi \mu_{\Omega_k~(dBm)}$ and $\sigma_{\hat{\Omega}_k}^2 = \xi^2 \sigma_{\Omega_k}^2$. The nth moment of the log-normal random variable Ω_k can be obtained from the moment generating function of the Gaussian random variable $\hat{\Omega}_k$ as

$$E[\Omega_k^n] = E[e^{n\hat{\Omega}_k}] = e^{n\mu_{\hat{\Omega}_k} + (1/2)n^2 \sigma_{\hat{\Omega}_k}^2} . \tag{3.4}$$

To find the appropriate moments for the log-normal approximation we can use (3.4) and equate the first two moments on both sides of the equation

$$I = \sum_{k=1}^{N_I} e^{\hat{\Omega}_k} \approx e^{\hat{Z}} = \hat{I} \tag{3.5}$$

where $\hat{Z} = \xi Z_{(dBm)}$. For example, suppose that $\hat{\Omega}_k, k = 1, \ldots, N_I$ have means $\mu_{\hat{\Omega}_k}, k = 1, \ldots, N_I$ and identical variances $\sigma_{\hat{\Omega}}^2$. Identical variances are often assumed because the standard deviation of log-normal shadowing is largely independent of the radio path length [188, 190]. Equating the means on both sides of (3.5)

$$\mu_I = E[I] = \sum_{k=1}^{N_I} E[e^{\hat{\Omega}_k}] = E[e^{\hat{Z}}] = E[\hat{I}] = \mu_{\hat{I}} \tag{3.6}$$

gives the result

$$\left(\sum_{k=1}^{N_I} e^{\mu_{\hat{\Omega}_k}} \right) e^{(1/2)\sigma_{\hat{\Omega}}^2} = e^{\mu_{\hat{Z}} + (1/2)\sigma_{\hat{Z}}^2} . \tag{3.7}$$

Likewise, we can equate the variances on both sides of (3.5) under the assumption that the $\hat{\Omega}_k, k = 1, \ldots N_I$ are independent

$$\sigma_I^2 = \mathrm{E}[I^2] - \mu_I^2 = \mathrm{E}[\hat{I}^2] = \sigma_{\hat{I}}^2 \tag{3.8}$$

giving the result

$$\left(\sum_{k=1}^{N_I} e^{2\mu_{\hat{\Omega}_k}} \right) e^{\sigma_{\hat{\Omega}}^2}(e^{\sigma_{\hat{\Omega}}^2} - 1) = e^{2\mu_{\hat{Z}}} e^{\sigma_{\hat{Z}}^2}(e^{\sigma_{\hat{Z}}^2} - 1) \ . \tag{3.9}$$

By squaring each side of (3.7) and dividing each side of resulting equation by the respective sides of (3.9) we can solve for $\sigma_{\hat{Z}}^2$ in terms of the known values of $\mu_{\hat{\Omega}_k}, k = 1, \ldots, N_I$ and $\sigma_{\hat{\Omega}}^2$. Afterwards, $\mu_{\hat{Z}}$ can be obtained from (3.7). This procedure yields the following solution:

$$\mu_{\hat{Z}} = \frac{\sigma_{\hat{\Omega}}^2 - \sigma_{\hat{Z}}^2}{2} + \ln \left(\sum_{k=1}^{N_I} e^{\mu_{\hat{\Omega}_k}} \right) \tag{3.10}$$

$$\sigma_{\hat{Z}}^2 = \ln \left((e^{\sigma_{\hat{\Omega}}^2} - 1) \frac{\sum_{k=1}^{N_I} e^{2\mu_{\hat{\Omega}_k}}}{\left(\sum_{k=1}^{N_I} e^{\mu_{\hat{\Omega}_k}} \right)^2} + 1 \right) \ . \tag{3.11}$$

Finally, $\mu_{Z\,(\mathrm{dBm})} = \xi^{-1} \mu_{\hat{Z}}$ and $\sigma_Z^2 = \xi^{-2} \sigma_{\hat{Z}}^2$.

The accuracy of this log-normal approximation can be measured in terms of how accurately the first two moments of $I_{(\mathrm{dB})} = 10 \log_{10} I$ are estimated, and how well the cumulative distribution function (cdf) of $I_{(\mathrm{dB})}$ is described by a Gaussian cdf. It has been reported in [295] that the Fenton-Wilkinson method breaks down for $\sigma_\Omega > 4$ **dB**. Unfortunately, for cellular radio applications the standard deviation of log-normal shadowing typically ranges from 6 to 12 dB. However, as pointed out in [28], the Fenton-Wilkinson method only breaks down if one considers the application of the Fenton-Wilkinson method for the prediction of the first two moments of $I_{(\mathrm{dB})}$. Moreover, in problems relating to the co-channel interference outage in cellular radio systems, we are usually interested in the tails of both the complementary distribution function (cdfc) $F_I^c(x) = \mathrm{P}(I \geq x)$ and the cdf $F_I(x) = 1 - F_I^c(x) = \mathrm{P}(I < x)$. In this case, we are interested in the accuracy of the approximation

$$F_I(x) \approx \mathrm{P}(e^{\hat{Z}} \geq x) = Q\left(\frac{\ln x - \mu_{\hat{Z}}}{\sigma_{\hat{Z}}} \right) \tag{3.12}$$

for large and small values of x. It will be shown later that the Fenton-Wilkinson method can approximate the tails of the cdf and cdfc functions with good accuracy.

1.2   SCHWARTZ-AND YEH-METHOD

The Schwartz-and-Yeh method [295] uses exact expressions for the first two moments of the sum of two log-normal random variables. Nesting and recursion techniques are then used to find exact values for the first two moments for the sum of N_I log-normal random variables. For example, suppose that $I = \Omega_1 + \Omega_2 + \Omega_3$. The exact first two moments of $\ln(\Omega_1 + \Omega_2)$ are computed. We then define $Z_2 = \ln(\Omega_1 + \Omega_2)$ as a new Gaussian random variable, let $I = e^{Z_2} + \Omega_3$, and again compute the exact first two moments of $\ln I$. Since the procedure is recursive we only need to detail the Schwartz-and-Yeh method for the case when $N_I = 2$, i.e.,

$$I = e^{\hat{\Omega}_1} + e^{\hat{\Omega}_2} \approx e^{\hat{Z}} = \hat{I} \tag{3.13}$$

or

$$\hat{Z} \approx \ln\left(e^{\hat{\Omega}_1} + e^{\hat{\Omega}_2}\right) \tag{3.14}$$

where the Gaussian random variables $\hat{\Omega}_1$ and $\hat{\Omega}_2$ have means $\mu_{\hat{\Omega}_1}$ and $\mu_{\hat{\Omega}_2}$, and variances $\sigma^2_{\hat{\Omega}_1}$ and $\sigma^2_{\hat{\Omega}_2}$, respectively.

Define the Gaussian random variable $\hat{\Omega}_d \triangleq \hat{\Omega}_2 - \hat{\Omega}_1$ so that

$$\mu_{\hat{\Omega}_d} = \mu_{\hat{\Omega}_2} - \mu_{\hat{\Omega}_1} \tag{3.15}$$

$$\sigma^2_{\hat{\Omega}_d} = \sigma^2_{\hat{\Omega}_1} + \sigma^2_{\hat{\Omega}_2} . \tag{3.16}$$

Taking the expectation of both sides of (3.14) and assuming that the approximation holds with equality gives

$$\begin{aligned} \mu_{\hat{Z}} &= \mathrm{E}\left[\ln\left(e^{\hat{\Omega}_2} + e^{\hat{\Omega}_1}\right)\right] \\ &= \mathrm{E}\left[\ln\left(e^{\hat{\Omega}_1}\left(1 + e^{\hat{\Omega}_2 - \hat{\Omega}_1}\right)\right)\right] \\ &= \mathrm{E}\left[\hat{\Omega}_1\right] + \mathrm{E}\left[\ln\left(1 + e^{\hat{\Omega}_d}\right)\right] . \end{aligned} \tag{3.17}$$

The second term in (3.17) is

$$\mathrm{E}\left[\ln\left(1 + e^{\hat{\Omega}_d}\right)\right] = \int_{-\infty}^{\infty} [\ln(1 + e^x)] p_{\hat{\Omega}_d}(x) dx . \tag{3.18}$$

We now use the power series expansion

$$\ln(1 + x) = \sum_{k=1}^{\infty} c_k x^k, \qquad c_k = \frac{(-1)^{k+1}}{k} \tag{3.19}$$

where $|x| < 1$. To ensure convergence of the power series and the resulting series of integrals, the integration in (3.18) is broken into ranges as follows:

$$\int_{-\infty}^{\infty} [\ln(1+e^x)]\, p_{\hat{\Omega}_d}(x)dx = \int_{-\infty}^{0} [\ln(1+e^x)]\, p_{\hat{\Omega}_d}(x)dx$$

$$+ \int_{0}^{\infty} [\ln(1+e^{-x}) + x]\, p_{\hat{\Omega}_d}(x)dx \quad (3.20)$$

The second integral is obtained by using the identity

$$\ln(1+e^x) = \ln[(e^{-x}+1)e^x]$$
$$= \ln(1+e^{-x}) + \ln(e^x)$$
$$= \ln(1+e^{-x}) + x \ . \quad (3.21)$$

After a very long derivation that is detailed in [295],

$$\mu_{\hat{Z}} = \mu_{\hat{\Omega}_1} + G_1 \quad (3.22)$$

where

$$G_1 = \mu_{\hat{\Omega}_d}\, \Phi\left(\frac{\mu_{\hat{\Omega}_d}}{\sigma_{\hat{\Omega}_d}}\right) + \frac{\sigma_{\hat{\Omega}_d}}{\sqrt{2\pi}} e^{-\mu_{\hat{\Omega}_d}^2/2\sigma_{\hat{\Omega}_d}^2}$$

$$+ \sum_{k=1}^{\infty} c_k e^{k^2\sigma_{\hat{\Omega}_d}^2/2}\left[e^{k\mu_{\hat{\Omega}_d}}\, \Phi\left(\frac{-\mu_{\hat{\Omega}_d} - k\sigma_{\hat{\Omega}_d}^2}{\sigma_{\hat{\Omega}_d}}\right) + T_1 \right] \quad (3.23)$$

with

$$T_1 = e^{-k\mu_{\hat{\Omega}_d}}\, \Phi\left(\frac{\mu_{\hat{\Omega}_d} - k\sigma_{\hat{\Omega}_d}^2}{\sigma_{\hat{\Omega}_d}}\right) . \quad (3.24)$$

The variance can be computed in a similar fashion, resulting in the expression [295]

$$\sigma_{\hat{Z}}^2 = \sigma_{\hat{\Omega}_1}^2 - G_1^2 - 2\sigma_{\hat{\Omega}_1}^2 G_3 + G_2 \quad (3.25)$$

where

$$G_2 = \sum_{k=1}^{\infty} b_k T_2 + \left[1 - \Phi\left(-\frac{\mu_{\hat{\Omega}_d}}{\sigma_{\hat{\Omega}_d}}\right)\right](\mu_{\hat{\Omega}_d}^2 + \sigma_{\hat{\Omega}_d}^2)$$

$$+ \frac{\mu_{\hat{\Omega}_d}\sigma_{\hat{\Omega}_d}}{\sqrt{2\pi}} e^{-\mu_{\hat{\Omega}_d}^2/(2\sigma_{\hat{\Omega}_d}^2)} \quad (3.26)$$

$$+ \sum_{k=1}^{\infty} b_k e^{-(k+1)\mu_{\hat{\Omega}_d} + (k+1)^2\sigma_{\hat{\Omega}_d}^2/2}\, \Phi\left(\frac{\mu_{\hat{\Omega}_d} - \sigma_{\hat{\Omega}_d}^2(k+1)}{\sigma_{\hat{\Omega}_d}}\right)$$

$$- 2\sum_{k=1}^{\infty} c_k e^{-\mu_{\hat{\Omega}_d}k + k^2\sigma_{\hat{\Omega}_d}^2/2}\left[\mu_{\hat{\Omega}_k}\, \Phi\left(-\frac{\mu_{\hat{\Omega}_k}}{\sigma_{\hat{\Omega}_d}}\right) - \frac{\sigma_{\hat{\Omega}_d}}{\sqrt{2\pi}} e^{-\mu_{\hat{\Omega}_k}^2/(2\sigma_{\hat{\Omega}_d}^2)}\right]$$

$$G_3 = \sum_{k=0}^{\infty} (-1)^k e^{k^2 \sigma_{\hat{\Omega}_d}^2/2} T_1 + \sum_{k=0}^{\infty} (-1)^k T_2 \tag{3.27}$$

with

$$T_2 = e^{\mu_{\hat{\Omega}_d}(k+1)+(k+1)^2\sigma_{\hat{\Omega}_d}^2/2} \Phi \left(\frac{-\mu_{\hat{\Omega}_d} - (k+1)\sigma_{\hat{\Omega}_d}^2}{\sigma_{\hat{\Omega}_d}} \right) \tag{3.28}$$

and

$$b_k = \frac{2(-1)^{k+1}}{k+1} \sum_{j=1}^{k} j^{-1} \tag{3.29}$$

$$\mu_{\hat{\Omega}_k} = -\mu_{\hat{\Omega}_d} + k\sigma_{\hat{\Omega}_d}^2 . \tag{3.30}$$

It has been reported in [295] that approximately 40 terms are required in the infinite summations to achieve four significant digits of accuracy in the moments. On the next step of the recursion it is important that we let $\sigma_{\hat{\Omega}_1}^2 = \sigma_{\hat{Z}}^2$ and $\mu_{\hat{\Omega}_1} = \mu_{\hat{Z}}$; otherwise, the procedure fails to converge.

1.3   FARLEY'S METHOD

Consider N_I normal random variables $\hat{\Omega}_k$ each with mean $\mu_{\hat{\Omega}}$ and variance $\sigma_{\hat{\Omega}}^2$. Farley approximated the cdfc of the sum

$$I = \sum_{k=1}^{N_I} = e^{\hat{\Omega}_k} \tag{3.31}$$

as

$$P(I > x) \approx 1 - \left[1 - Q \left(\frac{\ln x - \mu_{\hat{\Omega}}}{\sigma_{\hat{\Omega}}} \right) \right]^{N_I} . \tag{3.32}$$

As shown in [28], Farley's approximation is actually a strict lower bound on the cdfc. To obtain this result let

$$F_I^c(x) = P\left(\Omega_1 + \Omega_2 + \cdots + \Omega_{N_I} > x \right) \tag{3.33}$$

and define the two events

$$A = \{\text{at least one } \Omega_i > x\}$$
$$B = A^c, \text{ the complement of event } A . \tag{3.34}$$

Events A and B are mutually exclusive and partition the sample space. Therefore,

$$P(I > x) = P(I > x, A) + P(I > x, B)$$
$$= P(A) + P(I > x, B) . \tag{3.35}$$

The second term in (3.35) is positive for continuous pdfs such as the log-normal pdf. For example, the event

$$C = \{x/N_I < \Omega_i < x, \ \forall \ i\} \tag{3.36}$$

is a subset of the event B. Under the assumption that the Ω_i are independent and identically distributed, the probability of event C is

$$P(C) = \left[Q \left(\frac{\ln(x/N_I) - \mu_{\hat{\Omega}}}{\sigma_{\hat{\Omega}}} \right) - Q \left(\frac{\ln x - \mu_{\hat{\Omega}}}{\sigma_{\hat{\Omega}}} \right) \right]^{N_I} > 0 \ . \tag{3.37}$$

Therefore, $P(I > x) > P(A)$. Since the Ω_i are independent and identically distributed

$$
\begin{aligned}
P(A) &= 1 - \prod_{i=1}^{N_I} P(\Omega_i \leq x) \\
&= 1 - \left[1 - Q \left(\frac{\ln x - \mu_{\hat{\Omega}}}{\sigma_{\hat{\Omega}}} \right) \right]^{N_I} .
\end{aligned}
\tag{3.38}
$$

Finally, we have the lower bound on the cdfc

$$P(I > x) > 1 - \left[1 - Q \left(\frac{\ln x - \mu_{\hat{\Omega}}}{\sigma_{\hat{\Omega}}} \right) \right]^{N_I} \tag{3.39}$$

or, equivalently, the upper bound on the cdf

$$P(I \leq x) > \left[1 - Q \left(\frac{\ln x - \mu_{\hat{\Omega}}}{\sigma_{\hat{\Omega}}} \right) \right]^{N_I} . \tag{3.40}$$

1.4   NUMERICAL COMPARISONS

Fig. 3.1 compares the cdf for $N_I = 2$ and $N_I = 6$ log-normal random variables with the various log-normal approximations. Likewise, Figs. 3.2 - 3.4 provide comparisons of the various log-normal approximations for the cdfc. Exact results are also shown that have been obtained by computer simulation. Observe that the cdfc is approximated quite well for all the methods, but the best approximation depends on the number of interferers, shadow standard deviation, and range of distribution. The cdf is approximated less accurately, especially for $N_I = 6$ log-normal random variables.

2.   PROBABILITY OF OUTAGE

Consider the situation shown in Fig. 1.15, where a mobile station (MS) is at distance d_0 from the desired base station (BS) and at distances $d_k, k =$
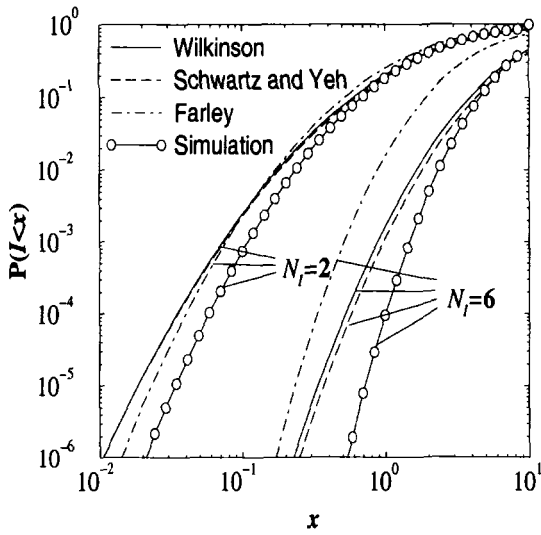
Figure 3.1. Comparison of the cdf for the sum of two and six log-normal random variables with various approximations; $\mu_{\Omega_k \ (dB)} = 0, \sigma_\Omega = 6$ dB.
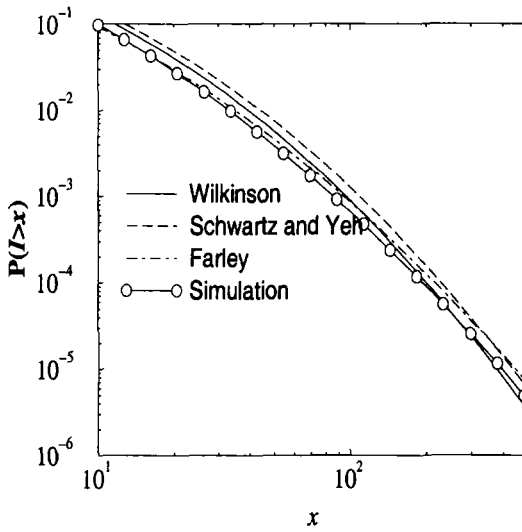
Figure 3.2. Comparison of the cdfc for the sum of two log-normal random variables with various approximations; $\mu_{\Omega_k \ (dB)} = 0, \sigma_\Omega = 6$ dB.
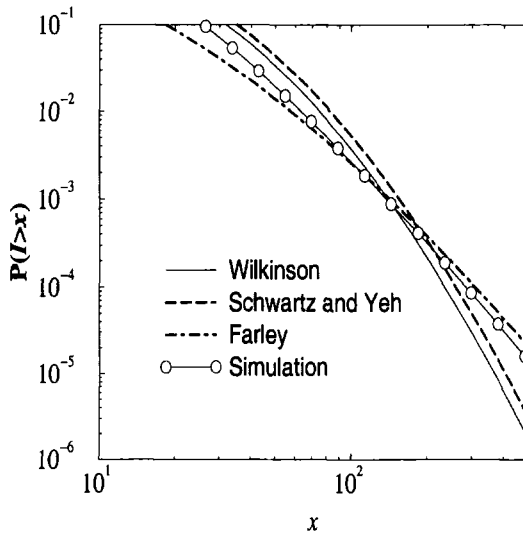
Figure 3.3. Comparison of the cdfc for the sum of six log-normal random variables with various approximations; $\mu_{\Omega_k\ (dB)} = 0, \sigma_\Omega = 6$ dB.
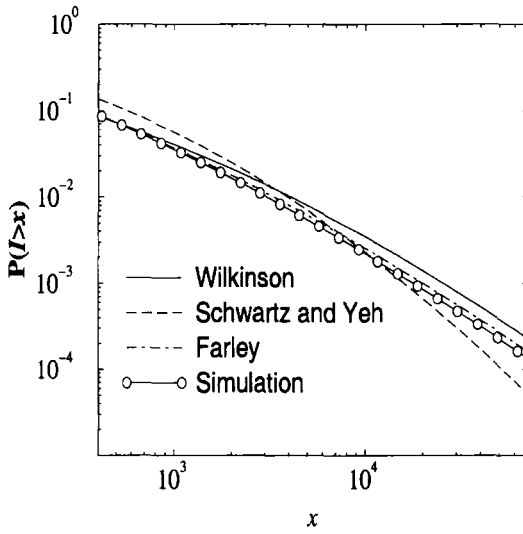
Figure 3.4. Comparison of the cdfc for the sum of six log-normal random variables with various approximations; $\mu_{\Omega_k\ (dB)} = 0, \sigma_\Omega = 12$ dB.

$1, 2, \cdots, N_I$ from the first tier of N_I co-channel BSs. For convenience, define the vector $\mathbf{d} = (d_0, d_1, \cdots, d_{N_I})$ as the set of distances for a particular MS location. The average received carrier-to-interference ratio as a function of the vector $\mathbf{d}$ is

$$\Lambda_{(\mathrm{dB})}(\mathbf{d}) = \Omega_{(\mathrm{dBm})}(d_0) - 10 \log_{10} \sum_{k=1}^{N_I} 10^{\Omega_{(\mathrm{dBm})}(d_k)/10} \ . \qquad (3.41)$$

For the case of a single interferer ($N_I = 1$) the sum on the right side of (3.41) only has one term. Therefore, $\Lambda_{(\mathrm{dB})}(\mathbf{d})$ is Gaussian with mean $\mu_{\Omega_{(\mathrm{dBm})}}(d_0) - \mu_{\Omega_{(\mathrm{dBm})}}(d_1)$ and variance $2\sigma_{\Omega}^2$. For the case of multiple interferers, the second term is approximated as a Gaussian random variable. We first obtain the mean $\mu_{\hat{Z}}$ and the variance $\sigma_{\hat{Z}}^2$ for the log-normal approximation using the techniques discussed in Section 1.. The mean and variance of $Z_{(\mathrm{dBm})} = \hat{Z}/\xi$ are, respectively,

$$\mu_{Z \ (\mathrm{dBm})} = \xi^{-1} \mu_{\hat{Z}} \qquad (3.42)$$

$$\sigma_Z^2 = \xi^{-2} \sigma_{\hat{Z}}^2 \ . \qquad (3.43)$$

Then

$$\Lambda_{(\mathrm{dB})}(\mathbf{d}) = \Omega_{(\mathrm{dBm})}(d_0) - Z_{(\mathrm{dBm})}(d_1, d_2, \ldots, d_{N_I}) \qquad (3.44)$$

where we have again shown the dependency of the co-channel interference on the set of distances. Note that $\Lambda_{(\mathrm{dB})}(\mathbf{d})$ has mean and variance

$$\mu_{\Lambda_{(\mathrm{dB})}(\mathbf{d})} = \mu_{\Omega_{(\mathrm{dBm})}}(d_0) - \mu_{Z \ (\mathrm{dBm})} \qquad (3.45)$$

$$\sigma_{\Lambda(\mathbf{d})}^2 = \sigma_{\Omega}^2 + \sigma_Z^2 \ . \qquad (3.46)$$

If there were only one possible choice of serving BS, then the probability of outage at a particular location is

$$O_I(\mathbf{d}) = Q\left(\frac{\mu_{\Omega_{(\mathrm{dBm})}}(d_0) - \mu_{Z \ (\mathrm{dBm})} - \Lambda_{\mathrm{th(dB)}}}{\sqrt{\sigma_{\Omega}^2 + \sigma_Z^2}} \right) \ . \qquad (3.47)$$

When handoffs are allowed the analysis is more complicated. In this case, the probability of outage will depend on the handoff algorithm that is employed. In the simplest case, we can consider soft handoffs where the BS that provides the most robust link is always used. In this case, an outage occurs only when no BS can provide a link having a carrier-to-interference ratio that exceeds Λ_{th}. In this case, the probability of outage at a particular location is

$$O_I = \prod_{k=0}^{M} Q\left(\frac{\mu_{\Omega_k \ (\mathrm{dBm})}(d_0) - \mu_{Z_k \ (\mathrm{dBm})} - \Lambda_{\mathrm{th(dB)}}}{\sqrt{\sigma_{\Omega}^2 + (\sigma_{Z_k})^2}} \right) \ . \qquad (3.48)$$

where M is the number of handoff candidates. The outage can then be calculated by averaging the probability of outage over the random location of a MS within a reference cell.

3.   MULTIPLE RICEAN/RAYLEIGH INTERFERERS

In microcellular environments, the received signal often consists of a direct line of sight (LOS) component, or perhaps a specular component, accompanied by a diffuse component. In this case, the envelope of the received signal experiences Ricean fading. In the same environment, the co-channel signals are often assumed to be Rayleigh faded, because a direct LOS between the co-channel cells is not likely to exist and the propagation path lengths are much longer. In this section, we calculate the probability of outage for the case of fading only. The combined effect of shadowing and fading is deferred until the next section. Let the instantaneous power in the desired signal and the N_I interfering signals be denoted by s_0 and s_k, $k = 1, \cdots, N_I$, respectively. Note that $s_i = \alpha_i^2$, where α_i^2 is the squared-envelope. For a specified receiver threshold λ_{th}, the probability of outage is

$$O_I = \mathrm{P}\left(\lambda < \lambda_{\text{th}}\right) \equiv \mathrm{P}\left(s_0 < \lambda_{\text{th}} \sum_{k=1}^{N_I} s_k\right) \tag{3.49}$$

where $\lambda = s_0 / \sum_{k=1}^{N_I} s_k$. The instantaneous received signal power, s_0, has the non-central chi-square (Ricean fading) distribution in (2.44), while the instantaneous power of each interferer, s_k, has the exponential distribution (Rayleigh fading) in (2.39).

For the case of a single interferer, the probability of outage reduces to the simple closed form [369]

$$O_I = \frac{\lambda_{\text{th}}}{\lambda_{\text{th}} + A_1} \exp\left\{-\frac{KA_1}{\lambda_{\text{th}} + A_1}\right\} \tag{3.50}$$

where K is the Rice factor of the desired signal, $A_1 = \Omega_0/(K+1)\Omega_1$, and $\Omega_k = \mathrm{E}[s_k]$. If the desired signal is Rayleigh faded, then the probability of outage can be obtained by setting $K = 0$ in (3.50). For the case of multiple interferers, each with mean power Ω_k, the probability of outage has the closed form [369]

$$O_I = 1 - \sum_{k=1}^{N_I}\left[1 - \frac{\lambda_{\text{th}}}{\lambda_{\text{th}} + A_k} \exp\left\{-\frac{KA_k}{\lambda_{\text{th}} + A_k}\right\}\right] \prod_{\substack{j=1 \\ j \neq k}}^{N_I} \frac{A_j}{A_j - A_k} \tag{3.51}$$

where $A_k = \Omega_0/(K+1)\Omega_k$. This expression is only valid if $\Omega_i \neq \Omega_j$ when $i \neq j$, i.e., the different interferers have different mean power. If some of

the interferers have the same mean power, then an appropriate expression for the probability of outage can be derived in straight forward manner. If all the interferers have the same mean power, then the total interference power $s_M = \sum_{k=1}^{N_I} s_k$ has the Gamma pdf

$$p_{s_M}(x) = \frac{x^{N_I-1}}{\Omega_1^{N_I}(N_I-1)!} \exp\left\{-\frac{x}{\Omega_1}\right\} . \tag{3.52}$$

The probability of outage can be derived as [369]

$$
\begin{aligned}
O_I &= \frac{\lambda_{\text{th}}}{\lambda_{\text{th}}+A_1} \exp\left\{-\frac{KA_1}{\lambda_{\text{th}}+A_1}\right\} \\
&\times \sum_{k=0}^{N_I-1} \left(\frac{A_1}{\lambda_{\text{th}}+A_1}\right)^k \sum_{m=0}^{k} \binom{k}{m} \frac{1}{m!} \left(\frac{K\lambda_{\text{th}}}{\lambda_{\text{th}}+A_1}\right)^m .
\end{aligned}
\tag{3.53}
$$

Again, if the desired signal is Rayleigh faded, then the probability of outage with multiple Rayleigh faded interferers can be obtained by setting $K = 0$ in either (3.51) or (3.53), which ever is appropriate. In Fig. 3.5, the probability of outage is plotted as a function of the carrier-to-interference ratio

$$\Lambda = \frac{\Omega_0}{N_I\Omega_1} \tag{3.54}$$

for various Rice factors and a single interferer. Observe that the Rice factor of the desired signal has a significant effect on the probability of outage. Fig. 3.6 plots the probability of outage for $K = 0$ and 7 and varying numbers of interferers. Observe that the number of interferers does not affect the probability of outage as much as the Rice factor, provided that the total interfering power remains constant.

4. MULTIPLE LOG-NORMAL NAKAGAMI INTERFERERS

The probability of outage has been evaluated in the literature for a single Nakagami interferer [364] and multiple Nakagami interferers [5, 370], in the absence of shadowing. Here we analytically formulate the probability of outage with multiple log-normal Nakagami interferers. For the case when the interfering signals have the same shadowing and fading statistics, we derive an exact mathematical expression for the probability of outage. Let the instantaneous power in the desired signal and the N_I interfering signals be denoted by s_0 and s_k, $k = 1, \cdots, N_I$, respectively. Again, for a specified receiver threshold λ_{th}, the probability of outage is

$$O_I = \mathrm{P}\left(\lambda < \lambda_{\text{th}}\right) \equiv \mathrm{P}\left(s_0 < \lambda_{\text{th}} \sum_{k=1}^{N_I} s_k\right) \tag{3.55}$$
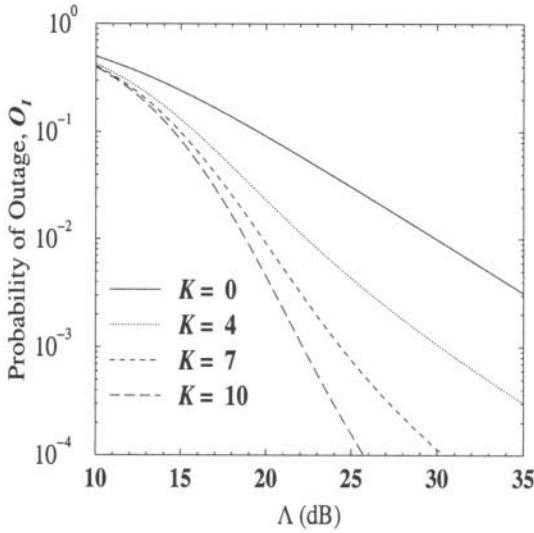
Figure 3.5. Probability of outage with a single interferer. The desired signal is Ricean faded with various Rice factors, while the interfering signal is Rayleigh faded; $\lambda_{\text{th}} = 10.0$ dB.

where $\lambda = s_0 / \sum_{k=1}^{N_I} s_k$. Since the kth signal is affected by log-normal shadowing and Nakagami fading, s_k has the composite pdf

$$p_{s_k}(x) = \int_0^\infty \left(\frac{m_k}{\Omega_k} \right)^{m_k} \frac{x^{m_k - 1}}{\Gamma(m_k)} \exp\left\{ -\frac{m_k x}{\Omega_k} \right\} \tag{3.56}$$

$$\times \frac{1}{\sqrt{2\pi} \xi \sigma_{\Omega_k} \Omega_k} \exp\left\{ -\frac{(10\log_{10}\Omega_k - \mu_{\Omega_k \ (\text{dB})})^2}{2\sigma_{\Omega_k}^2} \right\} d\Omega_k \ .$$

Let $W = \sum_{k=1}^{N_I} s_k$ be the total power from the N_I interfering signals, $X = s_0/W$, and $Y = W$. Then the joint pdf of X and Y is $p_{XY}(xy) = y p_{s_0, W}(xy, y)$ and

$$p_X(x) = \int_0^\infty y p_{s_0}(xy) p_W(y) dy \ . \tag{3.57}$$

It follows that the probability of outage is

$$O_I = \mathrm{P}\left(\frac{s_0}{W} < \lambda_{\text{th}} \right)$$

$$= 1 - \int_{\lambda_{\text{th}}}^\infty dx \int_0^\infty y p_{s_0}(xy) p_W(y) dy \tag{3.58}$$
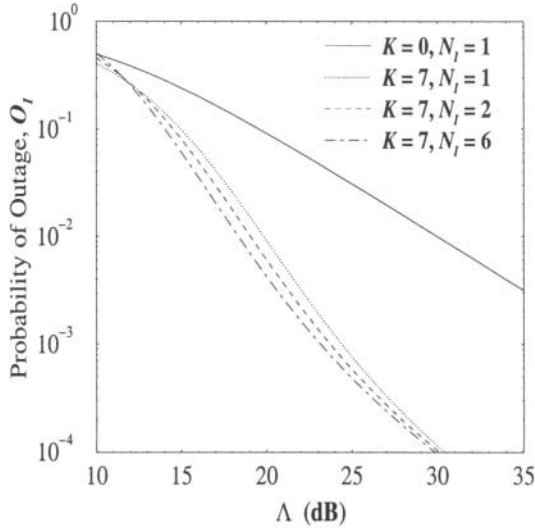
Figure 3.6. Probability of outage with multiple interferers. The desired signal is Ricean faded with various Rice factors, while the interfering signals are Rayleigh faded and of equal power; $\lambda_{\text{th}} = 10.0$ dB.

Substituting the Nakagami pdf for s_0 and integrating with respect to x gives the conditional probability [147]

$$P\left(\frac{s_0}{W} < \lambda_{\text{th}} \,|\, \Omega_0\right) = 1 - \sum_{h=0}^{m_0-1} \left(\frac{m_0\lambda_{\text{th}}}{\Omega_0}\right)^h \frac{1}{h!}$$
$$\times \int_0^\infty \exp\left\{-\frac{m_0\lambda_{\text{th}}y}{\Omega_0}\right\} y^h p_W(y)dy \quad (3.59)$$

4.1   STATISTICALLY IDENTICAL INTERFERERS

Here we assume statistically identical co-channel interferers so that $\sigma_{\Omega_k} = \sigma_\Omega$ and $m_k = m_I$, $i = 1, \ldots, N_I$. Following Linnartz [203], the integral in (3.59) can be obtained by using Laplace transform techniques. The Laplace transform of the pdf $p_W(y)$ is

$$\mathbf{L}_W(s) = \int_0^\infty e^{-sy} p_W(y)dy \quad (3.60)$$

The integral in (3.59) is then equal to the hth derivative of $\mathbf{L}_W(s)$ with respect to s evaluated at the point $s = (m_0\lambda_{\text{th}})/\Omega_0$. That is,

$$\int_0^\infty e^{-sy} y^h p_W(y) dy = (-1)^h \mathbf{L}_W^{(h)}(s) \tag{3.61}$$

$$= (-1)^h \frac{d^h}{ds^h} \left\{ \prod_{k=1}^{N_I} \int_0^\infty e^{-sy_k} p_{s_k}(y_k) dy_k \right\}$$

where the last line follows under the assumption of statistically independent interferes. By using the composite distribution in (3.56) with $m_k = m_I$

$$\int_0^\infty e^{-sy_k} p_{s_k}(y_k) dy_k = \frac{1}{\sqrt{\pi}} \int_{-\infty}^\infty \frac{m_I^{m_I} e^{-x^2}}{(10^{(\mu\Omega_k \text{ (dB)} + \sqrt{2}\sigma_\Omega x)/10} s + m_I)^{m_I}} dx \quad . \tag{3.62}$$

Averaging over the log-normal shadowing distribution of the desired signal gives the final result

$$O_I = 1 - \int_0^\infty \left\{ \sum_{h=0}^{m_0-1} \left(-\frac{m_0 \lambda_{th}}{\Omega_0} \right)^h \frac{1}{h!} \right. \tag{3.63}$$

$$\times \left. \frac{d^h}{ds^h} \left\{ \prod_{k=1}^{N_I} \int_{-\infty}^\infty \frac{m_I^{m_I} e^{-x^2} dx}{\sqrt{\pi}(10^{(\mu\Omega_k \text{ (dB)} + \sqrt{2}\sigma_\Omega)x/10} s + m_I)^{m_I}} \right\} \right|_{s=\frac{m_0 \lambda_{th}}{\Omega_0}} \right\}$$

$$\times \frac{\xi}{\sqrt{2\pi}\sigma_\Omega \Omega_0} \exp \left\{ -\frac{(10 \log_{10} \Omega_0 - \mu\Omega_0 \text{ (dB)})^2}{2\sigma_\Omega^2} \right\} d\Omega_0 \quad .$$

Equation (3.63) is an exact expression for shadowed Nakagami fading channels. When $m_d = m_I = 1$, it reduces to the simple expression obtained by Linnartz [203] for shadowed Rayleigh fading channels. If the path loss associated with each interferer is the same, then $\Omega_k = \Omega_I$ and the product in (3.63) reduces to taking the N_Ith power. Let

$$F(s) = \int_{-\infty}^\infty \frac{e^{-x^2}}{(10^{(\mu\Omega_I \text{ (dB)} + \sqrt{2}\sigma_\Omega x)/10} s + m_I)^{m_I}} dx \tag{3.64}$$

and use the identity [147]

$$G(s) = \frac{d^h}{ds^h} [F(s)]^{N_I}$$

$$= N_I \binom{h - N_I}{N_I} \sum_{i=1}^h (-1)^i \binom{h}{i} \frac{[F(s)]^{N_I-i}}{N_I - i} \frac{d^h}{ds^h} [F(s)]^i \quad . \tag{3.65}$$

Observe that $G(s)$ is just a function of the derivatives of $F(s)$, and

$$\frac{d^w}{dw}F(s) = \frac{d^w}{ds^w}\left\{\int_{-\infty}^{\infty}\frac{e^{-x^2}}{(10^{(\mu_{\Omega_I\ (dB)}+\sqrt{2}\sigma_\Omega x)/10}s+m_I)^{m_I}}dx\right\}$$

$$= (-1)^w\frac{(m_I+w-1)!}{(m_I-1)!} \tag{3.66}$$

$$\times \int_{-\infty}^{\infty}\frac{(10^{(\mu_{\Omega_I\ (dB)}+\sqrt{2}\sigma_\Omega x)/10})^w e^{-x^2}}{(10^{(\mu_{\Omega_I\ (dB)}+\sqrt{2}\sigma_\Omega x)/10}s+m_I)^{m_I+w}}dx \ .$$

We can obtain $G(s)$ from (3.65) and (3.66), and substitute it into (3.63). Then by using a change of variables the probability of outage in (3.63) becomes

$$O_I = 1 - \sum_{h=0}^{m_0-1}\left(-m_0\lambda_{th}10^{-\mu_{\Omega_0\ (dB)}/10}\right)^h\frac{m_I^{m_I N_I}}{\sqrt{\pi}^{N_I+1}h!} \tag{3.67}$$

$$\times \int_{-\infty}^{\infty}10^{-\sqrt{2}\sigma_\Omega xh/10}e^{-x^2}G\left(m_0\lambda_{th}10^{-(\mu_{\Omega_0\ (dB)}+\sqrt{2}\sigma_\Omega xh)/10}\right)dx \ .$$

The integrals in (3.66) and (3.67) can be efficiently computed by using Hermite-Gauss quadrature integration. Applying the Hermite-Gauss quadrature formula to (3.66) gives

$$F^{(w)}(s) = (-1)^w\frac{(m_I+w-1)!}{(m_I-1)!}$$

$$\times \sum_{t=1}^{N_p}H_{x_t}\frac{10^{(\mu_{\Omega_I\ (dB)}+\sqrt{2}\sigma_\Omega x_t)w/10}}{(10^{(\mu_{\Omega_I\ (dB)}+\sqrt{2}\sigma_\Omega x_t)/10}s+m_I)^{m_I+w}} \tag{3.68}$$

where H_{x_t} are weight factors, x_t are the zeros of the Hermite polynomial $H_p(x)$, and N_p is the order of the Hermite polynomial. By using this result and the values in Table 3.1 (listed for convenience) we have

$$O_I = 1 - \sum_{h=0}^{m_0-1}\left(-m_0\lambda_{th}10^{\mu_{\Omega_0\ (dB)}/10}\right)^h\frac{m_I^{m_I N_I}}{\sqrt{\pi}^{N_I+1}h!} \tag{3.69}$$

$$\times \sum_{\ell=1}^{N_p}H_{x_\ell}10^{-\sqrt{2}\sigma_\Omega x_\ell h/10}G\left(m_0\lambda_{th}10^{-(\mu_{\Omega_0\ (dB)}+\sqrt{2}\sigma_\Omega xh)/10}\right)dx \ .$$

Fig. 3.7 shows the probability of outage as a function of the carrier-to-interference ratio

$$\Lambda = \frac{\mu_{\Omega_0}}{N_I\mu_{\Omega_I}} \ . \tag{3.70}$$

Results are plotted for $N_I = 6$ interfering signals and varying degrees of fading on the desired and interfering signals. Observe that the probability of outage is

Zeros x_i	Weight Factors H_{x_i}
± 0.27348104613815	$5.079294790166 \times 10^{-1}$
± 0.82295144914466	$2.806474585285 \times 10^{-1}$
± 1.38025853919888	$8.381004139899 \times 10^{-2}$
± 1.95178799091625	$1.288031153551 \times 10^{-2}$
± 2.54620215784748	$9.322840086242 \times 10^{-4}$
± 3.17699916197996	$2.711860092538 \times 10^{-5}$
± 3.86944790486012	$2.320980844865 \times 10^{-7}$
± 4.68873893930582	$2.654807474011 \times 10^{-10}$

Table 3.1.   Zeros and weight factors of 16 order Hermite polynomials [1].
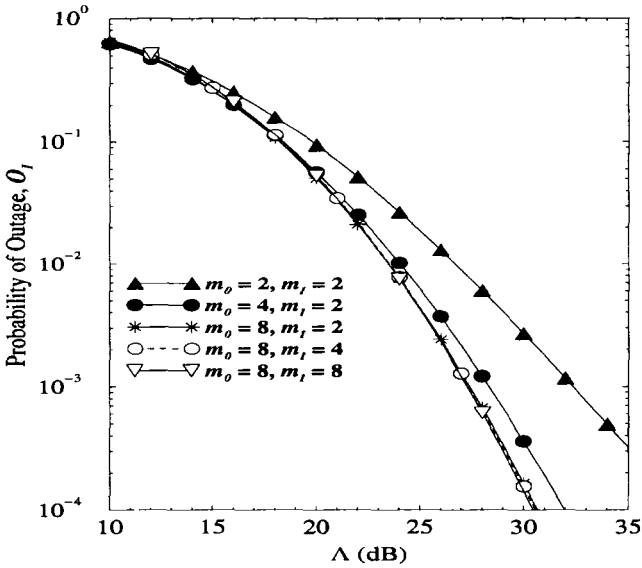
Figure 3.7.   Probability of outage when the desired and interfering signals are Nakagami faded. Results are shown for various fading distribution parameters; $\sigma_\Omega = 6$ dB, $\lambda_{\rm th} = 10.0$ dB.

insensitive to changes in the m values for interfering signals. This phenomenon demonstrates that co-channel interference is dominated by the fading of the desired signal rather than fading of the interfering signals. Fig. 3.8 shows the probability of outage for different values of the shadow standard deviation σ_Ω. We can conclude that the number of interferers and the shadow standard deviation have the most significant effect on the probability of outage.
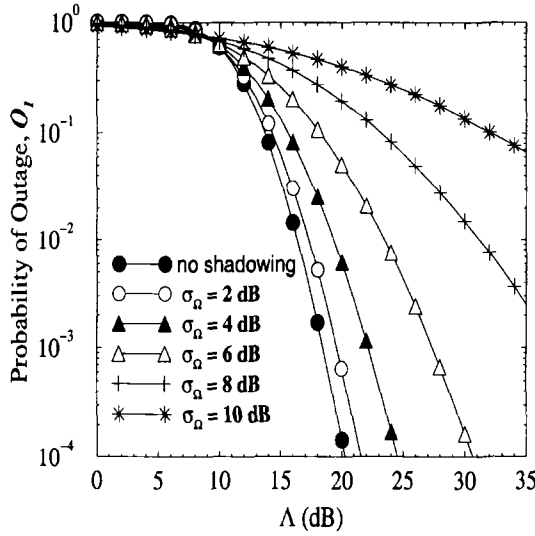
Figure 3.8. Probability of outage when the desired and interfering signals are Nakagami faded. Results are shown for various shadow standard deviations; $m_0 = 8$, $m_I = 2$, $\lambda_{\text{th}} = 10.0$ dB.

5. MULTIPLE LOG-NORMAL RICEAN/RAYLEIGH INTERFERERS

This section presents an exact method for evaluating the probability of co-channel interference for Ricean/Rayleigh faded channels with log-normal shadowing. The results can be applied for a Ricean faded desired signal and a single
Rayleigh faded interferer, or vice versa. It can also be applied for a Rayleigh faded desired signal with multiple Ricean or Rayleigh faded interfering signals. Once again, let the instantaneous power in the desired signal and the N_I interfering signals be denoted by s_0 and $s_k, k = 1, \cdots, N_I$, respectively. For a specified receiver threshold λ_{th}, the probability of outage is, again,

$$O_I = \mathrm{P}\left(\lambda < \lambda_{\text{th}}\right) \equiv \mathrm{P}\left(s_0 < \lambda_{\text{th}} \sum_{k=1}^{N_I} s_k\right) \tag{3.71}$$

where $\lambda = s_0 / \sum_{k=1}^{N_I} s_k$ and each s_k has either a composite log-normal exponential (Rayleigh fading) distribution or a composite log-normal non-central chi-square (Ricean fading) distribution. The $s_k, k = 0, \ldots, N_I$ in (3.71) can be reordered such that $(s_0 < \sum_{k=1}^{N_I} \lambda_{\text{th}} s_k) = (\tilde{s}_0 < \sum_{k=1}^{N_I} \delta_k \tilde{s}_k)$, where $\tilde{s}_0$ is exponentially distributed, the $\tilde{s}_k, k = 1, \ldots, N_I$ are either exponentially or non-central chi-square distributed. When the desired signal is Rayleigh faded

$\tilde{s}_0 = s_0$ and $\tilde{s}_k = s_k$ and $\delta_i = \lambda_{\text{th}}$. Otherwise, when the desired signal is Ricean faded and a single Rayleigh interferer is present, we observe that $P(s_0 < \lambda_{\text{th}}s_1) = P(s_1/(s_0/\lambda_{\text{th}}) > 1)$. Therefore, using $\tilde{s}_0 = s_1$, $\tilde{s}_1 = s_0$, and $\delta_1 = 1/\lambda_{\text{th}}$, the

probability of outage is $1 - P(s_1/(s_0/\lambda_{\text{th}}) < 1) = 1 - P(\tilde{s}_0/\delta_1\tilde{s}_1 < 1)$. Thus, let $X = \tilde{s}_0/W$ and $Y = W$, where $W = \sum_{k=1}^{N_I} \delta_k \tilde{s}_k$. The joint pdf of X and Y is $p_{XY}(x,y) = y\, p_{\tilde{s}_0,W}(xy,y)$ and

$$p_X(x) = \int_0^\infty y p_{\tilde{s}_0}(xy)p_W(y)dy \ . \tag{3.72}$$

Therefore,

$$O_I = P\left(\frac{\tilde{s}_0^2}{W} < 1\right) = 1 - \int_1^\infty dx \int_0^\infty y p_{\tilde{s}_0}(xy)p_W(y)dy \ . \tag{3.73}$$

Substituting the exponential pdf for $p_{\tilde{s}_0}(xy)$ and integrating with respect to x gives the conditional probability

$$P\left(\frac{\tilde{s}_0^2}{W} < 1\,|\Omega_0\right) = 1 - \int_0^\infty \exp\left\{-\frac{y}{\Omega_0}\right\}p_W(y)dy \tag{3.74}$$

where $\Omega_0 = E[\tilde{s}_0]$. Following Linnartz [203], the integral in (3.74) can be simplified by using Laplace transform techniques. Since the $\tilde{s}_k$, $k = 1, \dots, N_I$ are all independent random variables, $pw(y)$ is the convolution of the densities of the $\delta_k \tilde{s}_k$. Hence, (3.74) becomes

$$P\left(\frac{\tilde{s}_0^2}{W} < 1\,|\Omega_0\right) = 1 - \prod_{k=1}^{N_I} F_{\tilde{s}_k}(s)\Bigg|_{s=\delta_k/\Omega_0} \tag{3.75}$$

where $F_{\tilde{s}_k}(s)$ is the Laplace transform of the pdf $p_{\tilde{s}_k}(x)$, given by

$$
\begin{aligned}
F_{\tilde{s}_k}(s) &= \int_0^\infty e^{-sx} \int_0^\infty \frac{K_k+1}{\Omega_k}\exp\left\{-K_k - \frac{(K_k+1)x}{\Omega_k}\right\} \\
&\quad \times I_0\left(2\sqrt{\frac{K_k(K_k+1)x}{\Omega_k}}\right) p_{\Omega_k}(\Omega_k)d\Omega_k dx \\
&= \int_0^\infty \frac{K_k+1}{K_k+1+\delta_k s\Omega_k}\exp\left\{-\frac{\delta_k s\Omega_k K_k}{K_k+1+\delta_k v\Omega_k}\right\} \\
&\quad \times \frac{\xi}{\sqrt{2\pi}\sigma_{\Omega_k}\Omega_k}\exp\left\{-\frac{(10\log_{10}\Omega_k - \mu_{\Omega_k\,(\text{dB})})^2}{2\sigma_{\Omega_k}^2}\right\} d\Omega_k
\end{aligned} \tag{3.76}
$$

where K_k is the Rice factor of the kth signal. Averaging over the shadow distribution for the desired signal yields the final result

$$O_I = 1 - \int_0^\infty \left(\prod_{k=1}^{N_I} F_{\tilde{s}_k}(\delta_k/\Omega_k) \right) \tag{3.77}$$

$$\times \quad \frac{\xi}{\sqrt{2\pi}\sigma_{\Omega_0}\Omega_0} \exp\left\{ -\frac{(10\log_{10}\Omega_0 - \mu_{\Omega_0 \text{ (dB)}})^2}{2\sigma_{\Omega_0}^2} \right\} d\Omega_0$$

The integrals in (3.76) and (3.77) can be efficiently computed using Hermite-Gauss quadrature integration, as explained earlier. Corresponding expressions for Rayleigh fading can be obtained by setting the $K_k = 0$ in (3.76).

5.1   SINGLE INTERFERER

For a Rayleigh faded desired signal and a Ricean faded interferer, (3.77) can be used directly with $N_I = 1$ and $\delta_1 = \lambda_\text{th}$. If we assume the simple path loss model in (1.6), and define the normalized reuse distance as d_1/d_0, where d_0 and d_1 are the radio path lengths of the desired and interfering signals, respectively, then the average carrier-to-interference ratio is

$$\Lambda = \mu_{\Omega_0 \text{ (dB)}} - \mu_{\Omega_1 \text{ (dB)}} = 10\beta\log_{10}(d_1/d_0) \quad \text{dB} . \tag{3.78}$$

The probability of outage is plotted against the normalized reuse distance in Fig. 3.9, where it is shown to be insensitive to the Rice factor of the interferer. Likewise, Fig. 3.10 plots the probability of outage against the normalized reuse distance when the desired signal is Ricean faded and there is a single Rayleigh faded interferer. Observe the strong dependency of the probability of outage on the Rice factor of the desired signal.

5.2   MULTIPLE INTERFERERS

For a Rayleigh faded desired signal with multiple Ricean/Rayleigh interferers, (3.77) can be used directly leading to the same (not shown) result as the single interferer case; the probability of outage is insensitive to the Rice factors of the interferers.

For a Ricean faded desired signal with multiple Ricean/Rayleigh interferers, a different approach must be taken. An exact solution for the case of a Ricean faded desired signal with multiple Rayleigh interferers has been presented by Wang and Lea [349]. However, the case of a Rician faded desired signal with multiple Ricean interferers is still an open problem. One possibility is to approximate the Rice distribution with a Nakagami distribution as discussed in Sectionnaka, and use the results of Section 4..
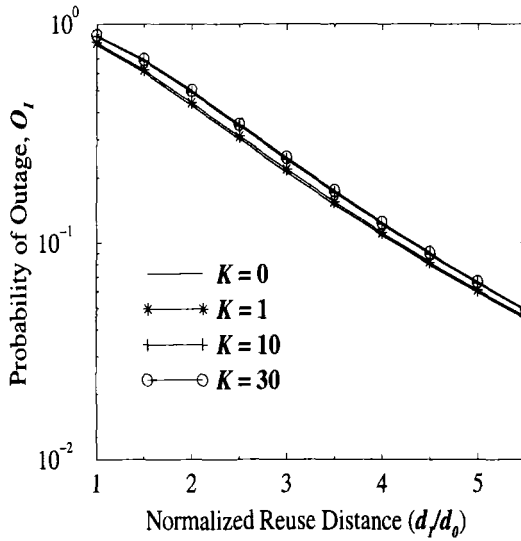
Figure 3.9. Probability of outage against the normalized reuse distance for a Rayleigh faded desired signal and one Ricean faded interferer. The Rice factors of the interfering signal varied; $\lambda_{th} = 10$ dB, $\sigma_{\Omega_0} = \sigma_{\Omega_1} = 6$ dB, $\beta = 4$.

Problems

3.1. Consider the scenario in Fig. 3.11 which depicts the worst case situation for the first tier of co-channel interference on the forward channel. Assume a cluster size of 7 cells, a cell radius of $R = 3$ km, a path loss exponent of $\beta = 3.5$, and carrier-to-interference threshold $\Lambda_{th\ (dB)} = 10$ dB. Ignore the effect of handoffs and assume that the MS must stay connected to the BS in the center cell.

  a) Using the simple path loss model in (1.6) with $\mu_{\Omega_p\ (dBm)}(d_o) = -10\,dBm$ at $d_o = 1$ km, a shadow standard deviation $\sigma_\Omega = 8$ dB, calculate the probability of outage $O_I(\mathbf{d})$ in (3.47) by using the Fenton-Wilkinson method.

  b) For $\sigma_\Omega = 4$ dB, what is required threshold Λ_{th} such that the probability of outage is less than 1%?

  c) Repeat b) for $\sigma_\Omega = 12$ dB.

3.2. Consider the Fenton-Wilkinson method for approximating the sum of log-normal random variables. Consider the sum of N log-normal random
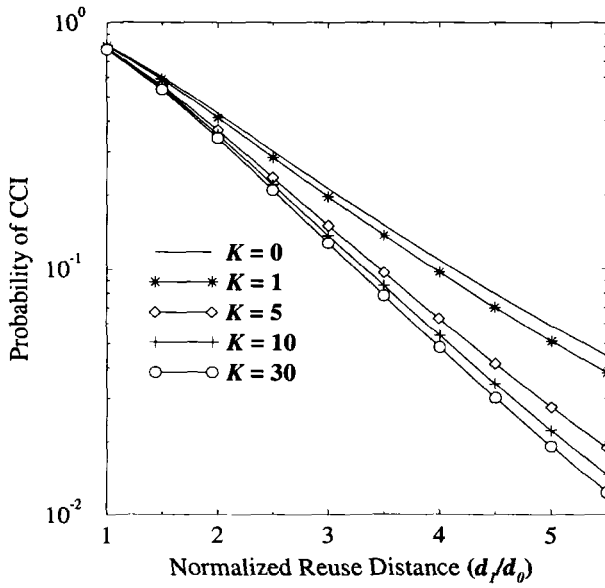
Figure 3.10. Probability of outage against the normalized reuse distance for a Ricean faded desired signal and one Rayleigh faded interferer. The Rice factors of the desired signal are varied; $\lambda_{\text{th}} = 10$ dB, $\sigma_{\Omega_0} = \sigma_{\Omega_1} = 6$ dB, $\beta = 4$.
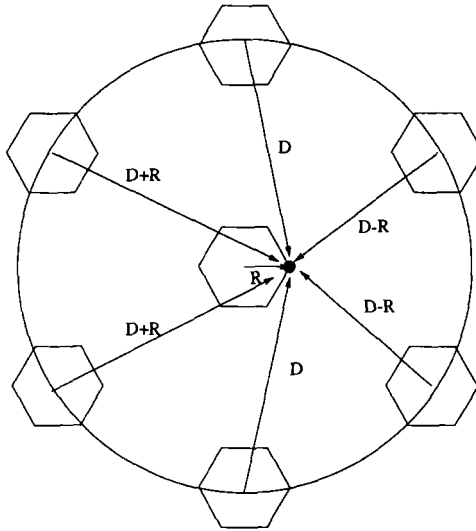
Figure 3.11. Co-channel interference on forward channel for Problem 3.1.

variables

$$I = \sum_{k=0}^{N} e^{\hat{\Omega}_k}$$
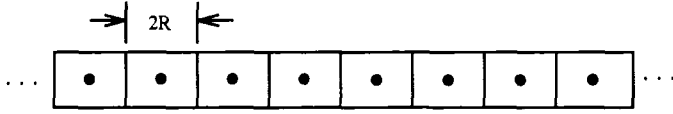
Figure 3.12.   Proposed highway microcell system for Problem 3.8.

where the $\hat{\Omega}_k$ are independent zero-mean Gaussian random variables with $\sigma_\Omega = 8$ dB. Plot the mean $\mu_{Z\ (\text{dBm})}$ and variance σ_Z^2 of the approximate Gaussian random variable $Z_{(\text{dB})}$ as a function of N for $N = 2, 3, 4, \ldots, 10$.

3.3. This problem uses computer simulation to verify the usefulness of the Schwartz-and-Yeh approximation and the Fenton-Wilkinson approximation for the sum of two log-normal random variables. Consider the sum of two log-normal random variables

$$I = \Omega_1 + \Omega_2$$

where the Gaussian random variables $\Omega_{1\ (\text{dB})}$ and $\hat{\Omega}_{2\ (\text{dB})}$ are independent and identically distributed with zero mean and variance σ_Ω^2. By using the Schwartz-and-Yeh method, plot the values of $\mu_{Z\ (\text{dB})}$ and σ_Z^2 as a function of the variance σ_Ω^2. Repeat for the Fenton-Wilkinson method. Now obtain the same results by using computer simulation and compare the analytical results. What are your conclusions?

3.4. You are asked to design a highway microcell system as shown in Fig. 3.12. Each cell has length $2R$.

   a) A BS with an omnidirectional antenna is placed at the center of each cell. Ignoring shadowing and envelope fading, determine the minimum reuse factor needed so that the worst case carrier-to-interference ratio, Λ, is at least 17 dB. State whatever assumptions you make.

   b) Now suppose that directional antennas are used to divide each cell into two sectors with boundaries perpendicular to the highway. Repeat part a).

   c) Consider again the sectored cell arrangement in part b). If shadowing is present with a standard deviation of σ_Ω dB, what is the probability of outage on a cell boundary? Assume soft handoffs between adjacent cells.

3.5. Derive equation (3.50).

3.6. Derive equation (3.51).

3.7. Derive equation (3.53).

3.8. Consider a microcellular environment where a Ricean faded signal is affected by a single Rayleigh faded interferer. Neglect the effect of path loss and shadowing. Suppose that the transmission quality is deemed acceptable if both the instantaneous carrier-to-noise ratio and the instantaneous carrier-to-interference ratio exceed the thresholds, γ_{th} and λ_{th}, respectively. Analogous to (3.53) derive an expression for the probability of outage.

Chapter 4

MODULATED SIGNALS
AND THEIR POWER SPECTRA

Modulation is the process where the message information is embedded into the radio carrier. Message information can be transmitted in the amplitude, frequency, or phase of the carrier, or a combination of these, in either analog or digital form. Most first generation cellular systems such as AMPS use analog FM, because analog technology was well understood when these systems were developed. However, the pressing need for greater spectral efficiency lead to the use of digital modulation techniques in second generation digital cellular systems.

To achieve high spectral efficiency, modulation schemes for FDMA and TDMA systems have a high **bandwidth efficiency**, measured in units of bits per second per Hertz of bandwidth (bits/s/Hz). As discussed earlier in this book, the link quality in many wireless systems is limited by co-channel interference. Hence, modulation schemes must be identified that are both bandwidth efficient and capable of tolerating high levels of co-channel interference. More specifically, digital modulation techniques are chosen for FDMA and TDMA wireless systems that satisfy the following three properties:

- *Compact Power Density Spectrum:* To minimize the effect of adjacent channel interference, the power radiated into the adjacent band should be 60 to 80 dB below that in the desired band. Hence, modulation techniques with a narrow main lobe and fast roll-off of side-lobes are needed.

- *Good Bit Error Rate Performance:* A low bit error probability must be achieved in the presence of fading, Doppler spread, intersymbol interference, adjacent and co-channel interference, and thermal noise.

- *Envelope Properties:* Portable and mobile applications typically employ non-linear (Class-C) power amplifiers to minimize battery drain. Nonlinear

amplification may degrade the bit error rate performance of modulation schemes that transmit information in the amplitude of the carrier. Also, spectral shaping is usually performed prior to up-conversion and non-linear amplification. To prevent the regrowth of spectral side-lobes during non-linear amplification, relatively constant envelope modulation schemes are preferred.

A variety of digital modulation techniques are currently being used in wireless communication systems. Two of the more widely used digital modulation techniques for cellular mobile radio are $\pi/4$-DQPSK and GMSK. The former is used in the North American IS-54 and Japanese PDC and PHS systems, while the latter is used in the European GSM, DCS 1800, DECT, and CT2 systems.

This book does not treat analog FM in detail and we refer the reader to other textbooks on the subject, such as those by Haykin [164] and Stremler [308]. Section 1. begins the chapter with a general characterization of band-pass modulated signals. Section 2. discusses Nyquist pulse shaping for ISI-free transmission. Sections 3. through 8. then provide a detailed treatment of the various linear and nonlinear digital modulations techniques that are suitable for mobile radio applications, including QAM, PSK, $\pi/4$-DQPSK, orthogonal modulation, OFDM, CPM, OMSK, and others. Since bandwidth efficiency is of great concern in mobile radio systems, Section 9. discusses the spectral characteristics of digitally modulated signals, beginning with a general framework followed by specific cases.

1. REPRESENTATION OF BAND-PASS MODULATED SIGNALS

Band-pass modulation schemes refer to modulation schemes that transmit information by using carrier modulation. The carrier modulated waveform can be expressed in the **complex envelope** form

$$s(t) = \operatorname{Re}\left[\tilde{s}(t)e^{j2\pi f_c t}\right] \tag{4.1}$$

where

$$\tilde{s}(t) = \tilde{s}_I(t) + j\tilde{s}_Q(t) \tag{4.2}$$

is the complex envelope and f_c is the carrier frequency. For any digital modulation scheme, the complex envelope can be written in the **standard form**

$$\tilde{s}(t) = A \sum_n b(t - nT, \mathbf{x}_n) \tag{4.3}$$

$$\mathbf{x}_n = (x_n,\ x_{n-1},\ \ldots,\ x_{n-K}) \tag{4.4}$$

where A is the amplitude and $\{x_n\}$ is the sequence of complex data symbols that are chosen from a finite alphabet. One data symbol is transmitted every

T seconds, so that the baud rate is $R = 1/T$ symbols/s. The function $b(t, \mathbf{x}_i)$ is an **equivalent shaping function** whose exact form depends on the type of modulation that is employed. For example, with binary phase shift keying (BPSK)

$$b(t, \mathbf{x}_n) = x_n u_T(t) \tag{4.5}$$

where

$$x_n \in \{-1, +1\} \quad = \quad \text{data symbol transmitted at epoch } n$$
$$u_T(t) = u(t) - u(t - T) \quad = \quad \text{unit amplitude rectangular pulse of length } T$$

and where $u(t)$ is the unit step function. Many other types of modulation are considered later in this chapter, where information is transmitted in the amplitude, phase, and/or frequency of the carrier. In each case, the modulated signal will be represented in the standard form in (4.3) so as to simplify the task of finding its power spectral density (psd).

By expanding (4.1), the band-pass waveform can also be expressed in the **quadrature form**

$$s(t) = \tilde{s}_I(t) \cos 2\pi f_c t - \tilde{s}_Q(t) \sin 2\pi f_c t . \tag{4.6}$$

The waveforms $\tilde{s}_I(t)$ and $\tilde{s}_Q(t)$ are known as the quadrature components $s(t)$, because they amplitude modulate the phase quadrature carrier components $\cos 2\pi f_c t$ and $\sin 2\pi f_c t$.

Finally $s(t)$ can be expressed in the **envelope-phase form**

$$s(t) = a(t) \cos(2\pi f_c t + \phi(t)) \tag{4.7}$$

where

$$a(t) \quad = \quad \sqrt{\tilde{s}_I^2(t) + \tilde{s}_Q^2(t)} \tag{4.8}$$

$$\phi(t) \quad = \quad \mathrm{Tan}^{-1} \left[\frac{\tilde{s}_Q(t)}{\tilde{s}_I(t)} \right] . \tag{4.9}$$

The three representations in (4.1), (4.6), and (4.7) are equivalent and will be used interchangeably.

1.1   VECTOR SPACE REPRESENTATIONS

For digital modulation schemes, the bandpass signal that is transmitted at each baud interval will belong to a finite set of finite energy waveforms with a few exceptions. Let $\{s_0(t), s_1(t), \ldots, s_{M-1}(t)\}$ denote the set of such waveforms, where M is the size of the set. The corresponding complex envelopes are denoted by $\{\tilde{s}_0(t), \tilde{s}_1(t), \ldots, \tilde{s}_{M-1}(t)\}$. For now we will work with the complex envelopes.

An *N*-dimensional complex vector space is defined by the set of complex orthonormal basis functions $\{\varphi_0(t), \varphi_1(t), \ldots, \varphi_{N-1}(t)\}$, where

$$\int_{-\infty}^{\infty} \varphi_i(t)\varphi_j^*(t)dt = \delta_{ij} \tag{4.10}$$

and $\delta_{ij} = 1$, $i = j$ and 0 otherwise. Each waveform $\tilde{s}_i(t)$ can be projected onto the set of basis functions to yield a signal vector

$$\tilde{\mathbf{s}}_m = (\tilde{s}_{m_0}, \tilde{s}_{m_1}, \ldots, \tilde{s}_{m_{N-1}}), \quad m = 0, \ldots, M-1 \tag{4.11}$$

where

$$\tilde{s}_{m_i} = \int_{-\infty}^{\infty} \tilde{s}_m(t)\varphi_i^*(t)dt \ . \tag{4.12}$$

If the basis functions are chosen appropriately, then the $\tilde{s}_m(t)$ can be expressed *exactly* in terms of the basis functions. That is,

$$\tilde{s}_m(t) = \sum_{i=0}^{N-1} \tilde{s}_{m_i}\varphi_i(t), \quad m = 0, \ldots, M-1 \ . \tag{4.13}$$

A systematic procedure for constructing an appropriate set of basis functions is now described.

1.2   GRAM-SCHMIDT PROCEDURE

Define the inner product between two waveforms $u(t)$ and $v(t)$ as

$$(u, v) = \int_{-\infty}^{\infty} u(t)v^*(t)dt \tag{4.14}$$

and define the norm of the waveform $u(t)$ as

$$\|u\| = \sqrt{(u, u)} \ . \tag{4.15}$$

Note that the squared-norm

$$\|u\|^2 = (u, u) = \int_{-\infty}^{\infty} |u(t)|^2 dt \tag{4.16}$$

is the energy contained in $u(t)$.

Given the finite set of finite energy signals $\{\tilde{s}_0(t), \tilde{s}_1(t), \ldots, \tilde{s}_{M-1}(t)\}$, an orthonormal set of basis functions $\{\varphi_0(t), \varphi_1(t), \ldots, \varphi_{N-1}(t)\}$ can be constructed according to the following algorithm:

1: Set $g_0(t) = \tilde{s}_0(t)$ and define

$$\varphi_0(t) = \frac{g_0(t)}{\|g_0\|} \tag{4.17}$$
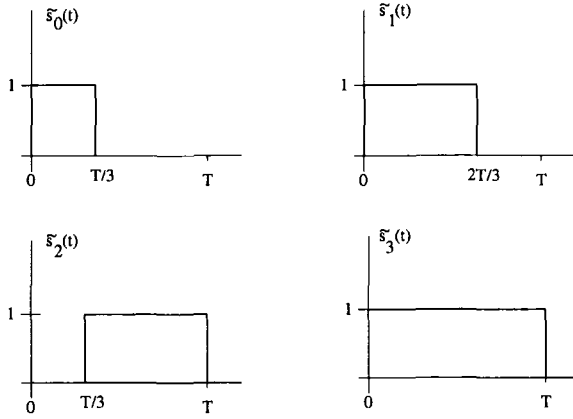
Figure 4.1.   Signal set $\{\tilde{s}_i(t)\}$ for Example 4.1.

2: Set $g_1(t) = \tilde{s}_1(t) - (\tilde{s}_0, \varphi_0)\varphi_0(t)$ and define

$$\varphi_1(t) = \frac{g_1(t)}{\|g_1\|} \tag{4.18}$$

3: Set $g_i(t) = \tilde{s}_i(t) - \sum_{j=0}^{i-1}(\tilde{s}_j, \varphi_j) = \tilde{s}_i(t) - \sum_{j=0}^{i-i} \tilde{s}_{ij}\varphi_j(t)$ and define

$$\varphi_i(t) = \frac{g_i(t)}{\|g_i\|^2} \tag{4.19}$$

Repeat Step 3 until all the $\tilde{s}_i(t)$'s have been used. If one or more of the above steps yields $g_i(t) = 0$, omit these from consideration. In the end a set of $N \leq M$ complex orthonormal basis functions $\{\varphi_0(t), \varphi_1(t), \ldots, \varphi_{N-1}(t)\}$ is obtained. The dimensionality of the complex vector space N equals M if and only if the set of waveforms $\{\tilde{s}_0(t), \tilde{s}_1(t), \ldots, \tilde{s}_{M-1}(t)\}$ is linearly independent, i.e., none of the waveforms is a linear combination of the others.

Example 4.1_____

Construct and orthonormal basis set for the set of waveforms shown in Fig. 4.1.

1: Set $g_0(t) = \tilde{s}_0(t)$. Then

$$\varphi_0(t) = \frac{g_0(t)}{\|g_0\|} = \begin{cases} \sqrt{3/T}, & 0 \leq t \leq T/3 \\ 0, & \text{else} \end{cases}$$

2: Set $g_1(t) = \tilde{s}_1(t) - (\tilde{s}_1, \varphi_0)\varphi_0(t)$, where

$$(\tilde{s}_1, \varphi_0) = \int_0^T \tilde{s}_1(t)\varphi_0^*(t)dt = \int_0^{T/3} \sqrt{3/T}dt = \sqrt{T/3}$$
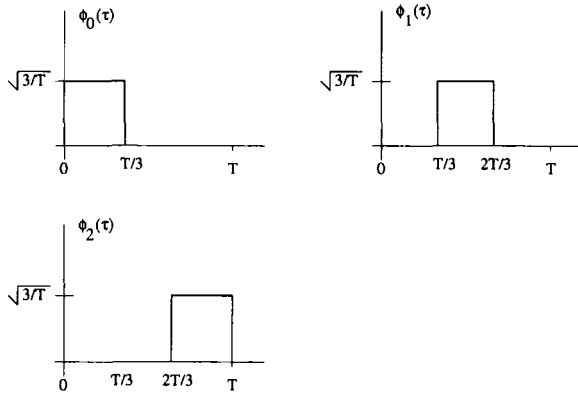
Figure 4.2. Orthonormal basis functions $\{\varphi_i\}$ for Example 4.1.

Then

$$\varphi_1(t) = \frac{g_1(t)}{\|g_1\|} = \begin{cases} \sqrt{3/T} \, , & T/3 \le t \le 2T/3 \\ 0 \, , & \text{else} \end{cases}$$

3: Set $g_2(t) = \tilde{s}_2(t) - (\tilde{s}_2, \varphi_0)\varphi_0(t) - (\tilde{s}_2, \varphi_1)\varphi_1(t)$, where

$$(\tilde{s}_2, \varphi_0) = \int_0^T \tilde{s}_2(t)\varphi_0^*(t)dt = 0$$

$$(\tilde{s}_2, \varphi_1) = \int_0^T \tilde{s}_2(t)\varphi_1^*(t)dt$$

$$= \int_{T/3}^{2T/3} \sqrt{3/T}dt = \sqrt{T/3}$$

Then

$$\varphi_2(t) = \frac{g_2(t)}{\|g_2\|} = \begin{cases} \sqrt{3/T} \, , & \sqrt{2T/3} \le t \le T \\ 0 \, , & \text{else} \end{cases}$$

4: Set $g_3(t) = \tilde{s}_3(t) - (\tilde{s}_3, \varphi_0)\varphi_0(t) - (\tilde{s}_3, \varphi_1)\varphi_1(t) - (\tilde{s}_3, \varphi_2)\varphi_2(t)$. But $g_3(t) = 0$, so ignore Step 4.

The set of basis function is shown in Fig. 4.2.

The $\tilde{s}_i(t)$ can be expressed as a linear combination of the basis functions, according to (4.13), and the corresponding signal vectors in (4.11) can be constructed. For the above example, the signal vectors are

$$\tilde{s}_0 = (\sqrt{T/3}, 0, 0)$$

$$\tilde{s}_1 = (\sqrt{T/3}, \sqrt{T/3}, 0)$$

$$\tilde{s}_2 = (0, \sqrt{T/3}, \sqrt{T/3})$$
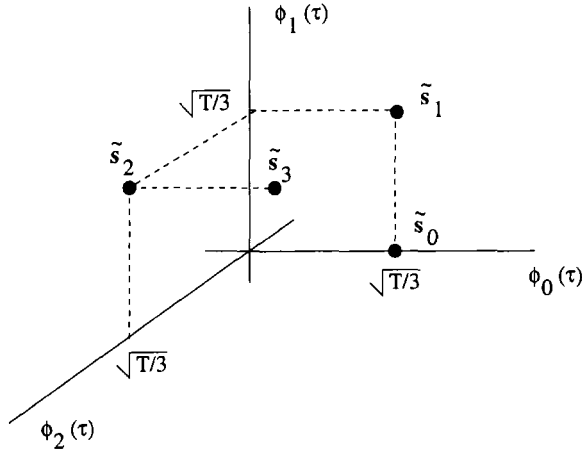
Figure 4.3.   Four signal vectors in 3-D signal space.

$$\tilde{s}_3 = (\sqrt{T/3}, \sqrt{T/3}, \sqrt{T/3})$$

These four signal vectors can be plotted in a 3-D signal space, as shown in Fig. 4.3.

1.3   SIGNAL ENERGY AND CORRELATIONS

Define the inner product between two length-N vectors $\mathbf{u}$ and $\mathbf{v}$ as

$$(\mathbf{u}, \mathbf{v}) = \mathbf{u} \cdot \mathbf{v}^* = \sum_{i=0}^{N-1} u_i v_i^* \tag{4.20}$$

and the norm of the vector $\mathbf{u}$ as

$$\|\mathbf{u}\| = \sqrt{(\mathbf{u}, \mathbf{u})} = \sum_{i=0}^{N-1} |u_i|^2 \tag{4.21}$$

Consider the set of band-pass waveforms

$$s_m(t) = \text{Re}\left[\tilde{s}_m(t)e^{j2\pi f_c t}\right], \quad m = 0, \ldots, M-1 \tag{4.22}$$

The **energy** in the waveform $s_m(t)$ is

$$E_m = (s_m, s_m) = \int_{-\infty}^{\infty} s_m^2(t)\,dt \tag{4.23}$$

Using the relation in (4.22) along with the identity $\text{Re}[z] = (z + z^*)/2$ we obtain

$$
\begin{aligned}
E_m &= \int_{-\infty}^{\infty} \left(\frac{\tilde{s}_m(t) + \tilde{s}^*(t)}{2} \right)^2 dt \\
&= \int_{-\infty}^{\infty} \frac{1}{2} \int_{-\infty}^{\infty} |\tilde{s}_m(t)|^2 dt \\
&\quad + \frac{1}{2} \int_{-\infty}^{\infty} |\tilde{s}_m(t)|^2 \cos(4\pi f_c t + 2\phi(t)) dt \\
&\approx \frac{1}{2} \int_{-\infty}^{\infty} |\tilde{s}_m(t)|^2 dt \\
&= \frac{1}{2}(\tilde{s}_m, \tilde{s}_m) \ .
\end{aligned}
\tag{4.24}
$$

where $\phi(t) = \text{Tan}^{-1}[\tilde{s}_Q(t)/\tilde{s}_I(t)]$. The above approximation is accurate when the bandwidth of the complex envelope is much less than the carrier frequency so that the double frequency term can be neglected. For digital band-pass modulated signals with baud rate $R = 1/T$, this condition is equivalent to $f_c T \gg 1$.

By using the Gram-Schmidt procedure, the $s_m(t)$ can be expressed in terms of a set of N (real) basis functions $\{\varphi_0(t), \varphi_1(t), \ldots, \varphi_{N-1}(t)\}$ where N is the dimension of the real vector space.

$$
s_m(t) = \sum_{i=0}^{N-1} s_{m_i} \varphi_i(t), \quad m = 0, \ldots, M - 1
\tag{4.25}
$$

yielding the corresponding signal vectors

$$
\mathbf{s}_m = (s_{m_0}, s_{m_1}, \ldots, s_{m_{N-1}}), \quad m = 0, \ldots, M - 1
\tag{4.26}
$$

It follows the energy in $s_m(t)$ is

$$
E_m = \int_{-\infty}^{\infty} \left(\sum_{i=0}^{N-1} s_{m_i} \varphi_i(t) \right)^2 dt = \sum_{i=0}^{N-1} s_{m_i}^2 = \|\mathbf{s}_m\|^2
\tag{4.27}
$$

where we used the orthonormal property of the basis functions in (4.10). Notice that the energy in $s_m(t)$ is equal to the squared norm of the corresponding signal vector $\mathbf{s}_m$.

Likewise, the $\tilde{s}_m(t)$ can be expressed in terms of a set of N *complex* basis functions $\{\varphi_0(t), \varphi_1(t), \ldots, \varphi_{N-1}(t)\}$ where N is the dimension of the *complex* vector space. Note that the dimensionality of the vector space and the set of basis functions for the $s_m(t)$ and the $\tilde{s}_m(t)$ are different, but related. The

energy in $s_m(t)$ is

$$E_m = \frac{1}{2} \int_{-\infty}^{\infty} \left| \sum_{i=0}^{N-1} \tilde{s}_{m_i} \varphi_i(t) \right|^2 dt = \frac{1}{2} \sum_{i=0}^{N-1} |\tilde{s}_{m_i}|^2 = \frac{1}{2} \|\tilde{s}_m\|^2 . \qquad (4.28)$$

Hence

$$E_m = \|s_m\|^2 = \frac{1}{2} \|\tilde{s}_m\|^2 . \qquad (4.29)$$

The **correlation** between the waveforms $s_m(t)$ and $s_k(t)$ is defined as

$$
\begin{aligned}
\rho_{km} &= \frac{1}{\sqrt{E_k E_m}} \int_{-\infty}^{\infty} s_m(t) s_k(t) dt \\
&= \frac{(s_m, s_k)}{\|s_m\| \|s_k\|} \\
&= \mathrm{Re} \left\{ \frac{(\tilde{s}_m, \tilde{s}_k)}{\|\tilde{s}_m\| \|\tilde{s}_k\|} \right\} .
\end{aligned} \qquad (4.30)
$$

Finally, the **squared Euclidean distance** between $s_k(t)$ and $s_m(t)$ is

$$
\begin{aligned}
d_{km}^2 &= \int_{-\infty}^{\infty} (s_m(t) - s_k(t))^2 dt \\
&= \|s_m - s_k\|^2 = \frac{1}{2} \|\tilde{s}_m - \tilde{s}_k\|^2 .
\end{aligned} \qquad (4.31)
$$

2.   NYQUIST PULSE SHAPING

Consider a modulation scheme where the complex envelope has the form

$$\tilde{s}(t) = A \sum_n x_n p(t - nT) \qquad (4.32)$$

where $p(t)$ is a shaping pulse, $\{x_n\}$ is the complex data symbol sequence, and T is the baud period. Now suppose the complex envelope is sampled every T seconds to yield the sample sequence $\{y_k\}$,

$$y_k = \tilde{s}(kT + t_o) = A \sum_n x_n p(kT + t_o - nT) \qquad (4.33)$$

where t_o is a timing offset assumed to lie in the interval $[0, T)$. First consider the case when $t_o = 0$; the effect of having $t_o \neq 0$ will be dealt with later. When $t_o = 0$

$$
\begin{aligned}
y_k &= A \sum_n x_n p_{k-n} \\
&= A x_k p_0 + A \sum_{n \neq k} x_n p_{k-n}
\end{aligned} \qquad (4.34)
$$

where $p_m = p(mT)$ is the sampled pulse. The first term in (4.34) is equal to the data symbol transmitted at the kth baud epoch, scaled by the factor p_0. The second term is the contribution of all other data symbols on the sample y_k. This term is called **intersymbol interference** (ISI).. To avoid the appearance of ISI, the sampled pulse response $\{p_k\}$ must satisfy the condition

$$p_k = \delta_{k0}p_0 \tag{4.35}$$

where δ_{jk} is the Dirac delta function defined by

$$\delta_{jk} = \begin{cases} 0, & j \neq k \\ 1, & j = k \end{cases} \tag{4.36}$$

In this case

$$y_k = x_k p_0 \quad . \tag{4.37}$$

Therefore, to avoid ISI the pulse $p(t)$ must have equally spaced zero crossings at intervals of T seconds. This requirement is known as the (first) Nyquist criterion.

We now derive an equivalent frequency domain requirement by showing that the pulse $p(t)$ satisfies the condition $p_k = \delta_{k0}p_0$ if and only if

$$P_\Sigma(f) \triangleq \frac{1}{T} \sum_{n=-\infty}^{\infty} P\left(f + \frac{n}{T}\right) = p_0 \quad . \tag{4.38}$$

The term $P_\Sigma(f)$ is called the **folded spectrum.** To avoid ISI, the folded spectrum must be a constant value or in other words "flat." Using the Fourier transform, we can write

$$
\begin{aligned}
p_k &= \int_{-\infty}^{\infty} P(f)e^{j2\pi fkT}df \\
&= \sum_{n=-\infty}^{\infty} \int_{(2n-1)/2T}^{(2n+1)/2T} P(f)e^{j2\pi kfT}df \\
&= \sum_{n=-\infty}^{\infty} \int_{-1/2T}^{1/2T} P\left(f' + \frac{n}{T}\right) e^{j2\pi k(f'+\frac{n}{T})T}df' \\
&= \int_{-1/2T}^{1/2T} \left[\sum_{n=-\infty}^{\infty} P\left(f + \frac{n}{T}\right)\right] e^{j2\pi fkT}df \\
&= T \int_{-1/2T}^{1/2T} P_\Sigma(f)e^{j2\pi fkT}df \quad .
\end{aligned}
\tag{4.39}
$$

It follows that p_k and $P_\Sigma(f)$ are a Fourier series pair, i.e.,

$$P_\Sigma(f) = \sum_{k=-\infty}^{\infty} p_k e^{-j2\pi kfT} \quad . \tag{4.40}$$

Now suppose that the condition in (4.38) is satisfied. Then $P_\Sigma(f) = p_0 T$ and from the last line of (4.39)

$$p_k = \int_{-1/2T}^{1/2T} e^{j2\pi fkT} p_0 T\, df = \frac{\sin \pi k}{\pi k} p_0 = \delta_{k0} p_0 \ . \qquad (4.41)$$

Conversely, suppose that the condition $p_k = p_0 \delta_{k0}$ is satisfied. Then from (4.40)

$$P_\Sigma(f) = p_0 \ . \qquad (4.42)$$

The requirement on the folded spectrum in (4.38) allows us to design pulses in the frequency domain that will yield zero ISI. First consider the pulse

$$P(f) = T\mathrm{rect}(fT) \qquad (4.43)$$

where

$$\mathrm{rect}(fT) = \left\{ \begin{array}{ll} 1, & |f| \le \frac{1}{2T} \\ 0, & \text{elsewhere} \end{array} \right. \qquad (4.44)$$

This pulse yields a flat folded spectrum. In the time domain

$$p(t) = \mathrm{sinc}(t/T) \ . \qquad (4.45)$$

This pulse achieves the first Nyquist criterion because it has equally spaced zero crossings at T second intervals. Furthermore, from the requirement of a flat folded spectrum, it achieves zero ISI while occupying the smallest possible bandwidth. Hence, it is called an *ideal* Nyquist pulse. Sometimes the edge frequency $f = 1/2T$ is called the Nyquist frequency.

We now examine the effect of the sampling or timing offset t_o with the aid of the ideal Nyquist pulse. With a timing offset

$$\begin{aligned} y_k &= A\sum_n x_n \mathrm{sinc}((kT + nT + t_o)/T) \\ &= Ax_k \mathrm{sinc}(t_o/T) + A\sum_{n \ne k} x_n \mathrm{sinc}((kT + nT + t_o)/T) \quad (4.46) \end{aligned}$$

In this case, the ISI term is not zero. Furthermore, with the ideal Nyquist pulse the effect of the timing offset is exasperated because the ISI term is not absolutely summable as shown in Problem 4.1. This is caused by the slow time decay of the ideal Nyquist pulse, in this case $1/t$. To make our communication systems more robust to timing errors, we would like to construct pulses that satisfy the Nyquist criterion but decay faster with time.

To construct other Nyquist pulses, we start with the ideal Nyquist pulse, $P_N(f)$, shown in Fig. 4.4(a). To the pulse $P_N(f)$, we add a "transmittence" function $P_o(f)$ as shown in Fig. 4.4(b). The critical requirement is that the
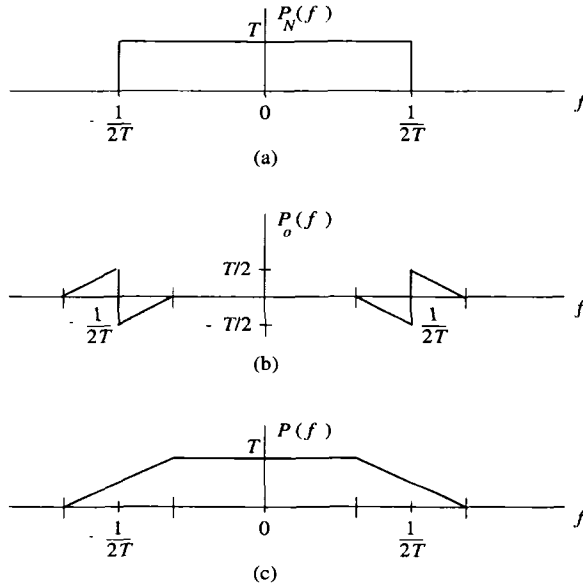
Figure 4.4. Construction of pulses satisfying the (first) Nyquist criterion.

transmittence function have *skew* symmetry about the Nyquist frequency $1/2T$. Any function with skew symmetry will do. The resulting Nyquist pulse $P(f)$ is shown in Fig. 4.4(c). Clearly, the pulse has a flat folded spectrum. The corresponding time domain pulse $p(t)$ can be obtained by taking the inverse Fourier transform of $P(f)$. Notice that the pulse $P(f)$ takes up additional bandwidth in exchange for the faster decay of the corresponding time domain pulse $p(t)$.

Raised cosine and root raised cosine pulse shaping:. The raised cosine pulse is defined by

$$P(f) = \begin{cases} T & 0 \le |f| \le (1-\beta)/2T \\ \frac{T}{2}\left[1 - \sin\frac{\pi T}{\beta}\left(f - \frac{1}{2T}\right)\right] & (1-\beta)/2T \le |f| \le (1+\beta)/2T \\ 0 & |f| \ge (1-\beta)/2T \end{cases}$$

(4.47)

The bandwidth of the raised cosine pulse is $(1+\beta)/2T$. The **roll-off factor** $\beta, 0 \le \beta \le 1$ controls the bandwidth that is in excess of the Nyquist bandwidth. Taking the inverse Fourier transform of $P(f)$ gives the time domain pulse

$$p(t) = \frac{\sin \pi t/T}{\pi t/T} \frac{\cos \beta \pi t/T}{1 - 4\beta^2 t^2/T^2}$$

(4.48)

For $\beta = 0$, $p(t)$ reduces to the sinc pulse in (4.45). Notice that the raised cosine pulse decays as a function of $1/t^3$.

As discussed in Chapter 5, the pulse shaping is usually divided between the transmitter and receiver filters. Very often the receiver filter $h_r(t)$ is "matched" to the transmitter filter $h_a(t)$, in which case $h_r(t) = h_a(-t)$. The overall pulse consisting of the cascade of the transmitter and receiver filters is $p(t) = h_a(t) * h_a(-t)$, where $*$ denotes the operation of convolution. The equivalent condition in the frequency domain is $P(f) = |H_a(f)|^2$. In this case, the transmitter filter has transfer function $H_a(f) = |P(f)|^{1/2}$. If the overall pulse $p(t)$ is a raised cosine pulse with the transfer function $P(f)$ in (4.47), then the pulse $h_a(t)$ is said to be a **root raised cosine** pulse. Taking the inverse Fourier transform of $H_a(f)$ gives the corresponding time domain root raised cosine pulse

$$h_a(t) = 4\beta \frac{\cos[(1+\beta)\pi t/T] + \sin[(1-\beta)\pi t/T](4\beta t/T)^{-1}}{\pi\sqrt{T}[1 - 16\beta^2 t^2/T^2]} \ . \qquad (4.49)$$

For $\beta = 0$, $h_a(t)$ reduces to the sinc pulse

$$h_a(t) = \sqrt{\frac{1}{T}} \, \mathrm{sinc}(t/T) \ . \qquad (4.50)$$

Raised cosine and root raised cosine pulses corresponding to $\beta = 0.5$ are shown in Fig. 4.5. Strictly speaking the pulses in (4.49) and (4.5) are non-causal. Therefore, in practice a truncated time domain pulse must be used. For example, in Fig. 4.5 the pulses are truncated to $6T$ and time shifted by $3T$ to yield causal pulses. Later we will look at the effect of the pulse truncation length. Notice that the raised cosine pulse is a Nyquist pulse with equally spaced zero crossings at the baud period T, while the root raised cosine pulse is not.

3.   QUADRATURE AMPLITUDE MODULATION (QAM)

With QAM, the complex envelope is

$$\tilde{s}(t) = A \sum_n b(t - nT, \mathbf{x}_n) \qquad (4.51)$$

where

$$b(t, \mathbf{x}_n) = x_n h_a(t) \qquad (4.52)$$

$h_a(t)$ is the **amplitude shaping pulse**, and $x_n = x_{I,n} + jx_{Q,n}$ is the complex data symbol that is transmitted at epoch n. It is apparent that both the amplitude and the phase of a QAM signal depend on the complex symbol. QAM has
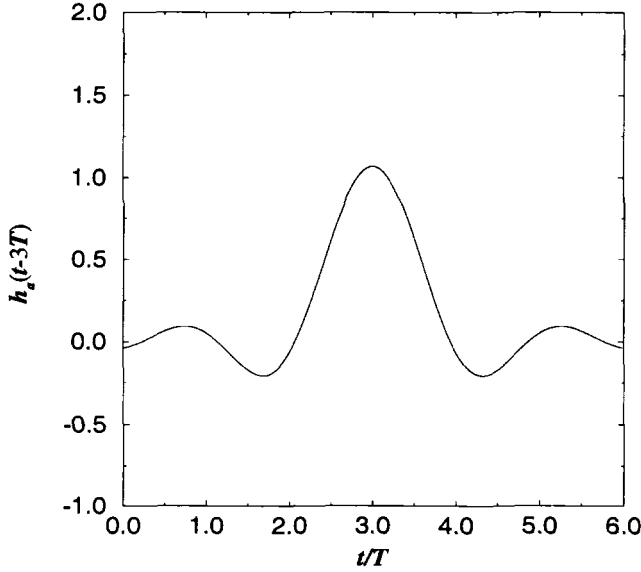
Figure 4.5. Raised cosine and root raised cosine pulses with roll-off factor $\beta = 0.5$. The pulses are truncated to length $6T$ and time shifted by $3T$ to yield causal pulses.

the advantage of high bandwidth efficiency, but amplifier nonlinearities will degrade its performance due to the non-constant envelope.

The QAM waveforms that are transmitted at each baud epoch have complex envelopes

$$\tilde{s}_m(t) = A x_m h_a(t) \quad m = 0, \ldots, M - 1 \ . \tag{4.53}$$

To represent the $\tilde{s}_m(t), m = 0, \ldots, M - 1$ in terms of a set of signal vectors, the required basis function is

$$\varphi_0(t) = \sqrt{\frac{A^2}{2E_h}} h_a(t) \tag{4.54}$$

where

$$E_h = \frac{A^2}{2} \int_{-\infty}^{\infty} h_a^2(t) dt \tag{4.55}$$

is the energy in the band-pass pulse $A h_a(t) \cos 2\pi f_c t$. Using this basis function

$$\tilde{s}_m(t) = \sqrt{2E_h} \, x_m \, \varphi_0(t) \tag{4.56}$$

and the QAM complex signal vectors are[1]

$$\tilde{s}_m = \sqrt{2E_h} \, x_m \, , \quad m = 0, \ldots, M - 1 \ . \tag{4.57}$$

---

[1] Since only one basis function is needed, we use the scalar $\tilde{s}_m$ rather than the vector $\tilde{\mathbf{s}}_m$.
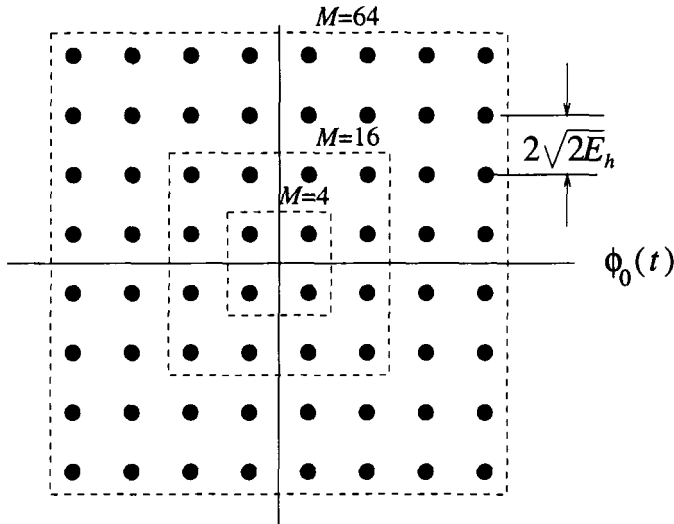
Figure 4.6.   Complex signal-space diagram for square QAM constellations.

QAM signal constellations:.   A variety of QAM signal constellations may be constructed. Square QAM constellations can be constructed when M is a power of 4 by choosing $x_{I,m}, x_{Q,m} \in \{\pm 1, \pm 3, \ldots, \pm(N-1)\}$ and $N = \sqrt{M}$. The complex signal-space diagram for the square 4-, 16-, and 64-QAM constellations is shown in Fig. 4.6. Notice that the minimum Euclidean distance between any two signal vectors is $2\sqrt{2E_h}$.

When M is not a power of 4, the signal constellation is not square. Usually, the constellation is given the shape of a cross to minimize the average energy in the constellation for a given minimum Euclidean distance between signal vectors. Examples of the QAM "cross constellations" are shown in Fig. 4.7.

Other types of QAM constellations are possible as well. Fig. 4.8 shows two different 8-QAM constellations.

PAM signal constellations:.   Pulse amplitude modulation (PAM) can be viewed as a special case of QAM, where information is transmitted only in the cosine component of the carrier. With our formulation, this can be accomplished by using real data symbols $x_m = x_{I,m}$, belonging to the set $\{\pm 1, \pm 3, \ldots, \pm(m-1)\}$. The PAM complex signal vectors are

$$\tilde{s}_m = \sqrt{2E_h}(2m - 1 - M), \quad m = 1, \ldots, M \ . \tag{4.58}$$

Typical 4- and 8-PAM signal constellations are shown in Fig. 4.9.
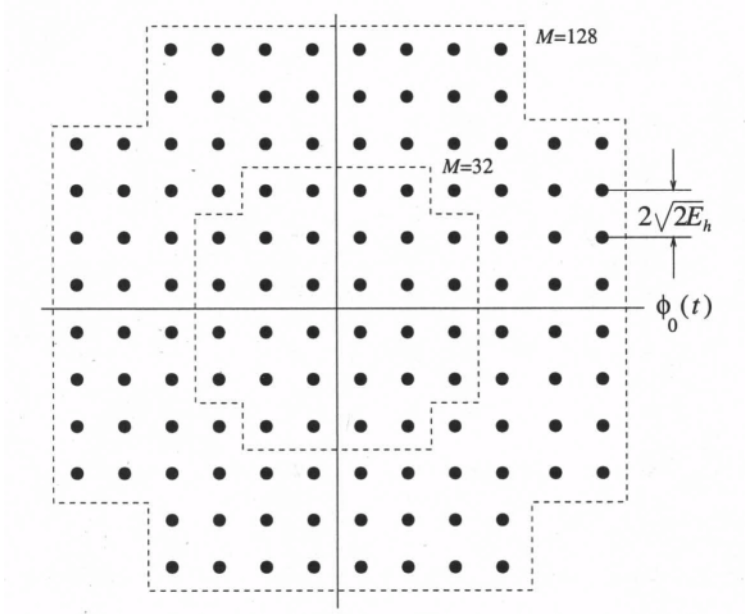
Figure 4.7. Complex signal-space diagram for cross QAM constellations.
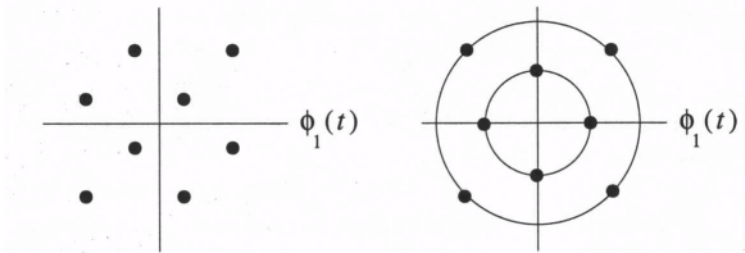
Figure 4.8. Complex signal-space diagram for 8-QAM constellations.

4. PHASE SHIFT KEYING (PSK)

The complex envelope of a PSK signal has the form

$$\tilde{s}(t) = A \sum_n b(t - nT, \mathbf{x}_n) \tag{4.59}$$

where

$$b(t, \mathbf{x}_n) = h_a(t)e^{j\theta_n} \tag{4.60}$$

$h_a(t)$ is the amplitude shaping pulse. The carrier phase takes on values

$$\theta_n = \frac{2\pi}{M}x_n + \theta_o \tag{4.61}$$
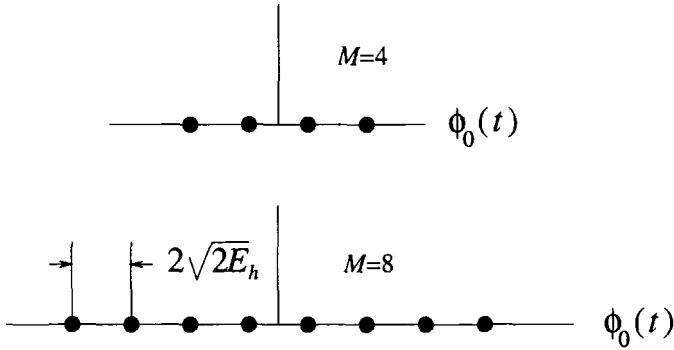
Figure 4.9.   Complex signal-space diagram for 4- and 8-PAM constellations.

where θ_o is an arbitrary constant phase, and the data symbols are defined as $x_n = n, n \in \{0, 1, \ldots, M-1\}$, with M being the alphabet size.

The PSK waveforms that are transmitted at each baud epoch have complex envelopes

$$\tilde{s}_m(t) = Ah_a(t)e^{j\theta_m} , \quad m = 0, \ldots, M-1 \tag{4.62}$$

Using the basis function in (4.54)

$$\tilde{s}_m(t) = \sqrt{2E_h}e^{j\theta_m}\varphi_0(t) , \quad m = 0, \ldots, M-1 . \tag{4.63}$$

The PSK complex signal vectors are

$$\tilde{s}_m = \sqrt{2E_h}e^{j\theta_m} , \quad m = 0, \ldots, M-1 \tag{4.64}$$

The complex signal-space diagram for 8-PSK (with $\theta_0 = 0$) is shown in Fig. 4.10. Notice that all PSK waveforms have the same energy E_h.

4.1   OFFSET QPSK (OQSPK)

QPSK or 4-PSK is equivalent to 4-QAM, where $x_n = x_{I,n} + jx_{Q,n}$ and $x_{I,n}, x_{Q,n} \in \{-1/\sqrt{2}, +1/\sqrt{2}\}$. The QPSK signal can have either $\pm 90°$ or $180°$ phase shifts from one baud interval to the next. With offset QPSK (OQPSK), the complex envelope is

$$\tilde{s}(t) = A \sum_n b(t - nT, \mathbf{x}_n) \tag{4.65}$$

where

$$b(t, \mathbf{x}_n) = x_{I,n}h_a(t) + jx_{Q,n}h_a(t - T_b) \tag{4.66}$$

and $T_b = T/2$ is the bit interval. With OQPSK signals the possibility of $180°$ phase shifts is eliminated. In fact, the phase can change by only $\pm 90°$ every
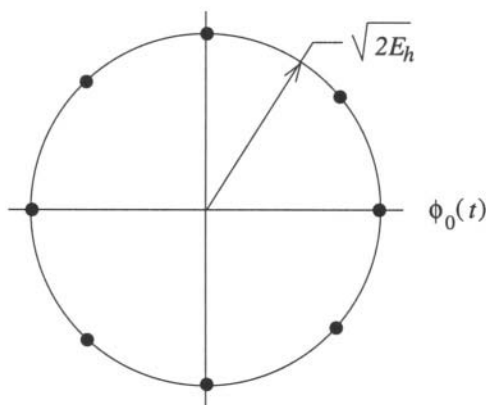
Figure 4.10. Complex signal-space diagram for 8-PSK signals; $\theta_o = 0$.

QPSK   OQPSK
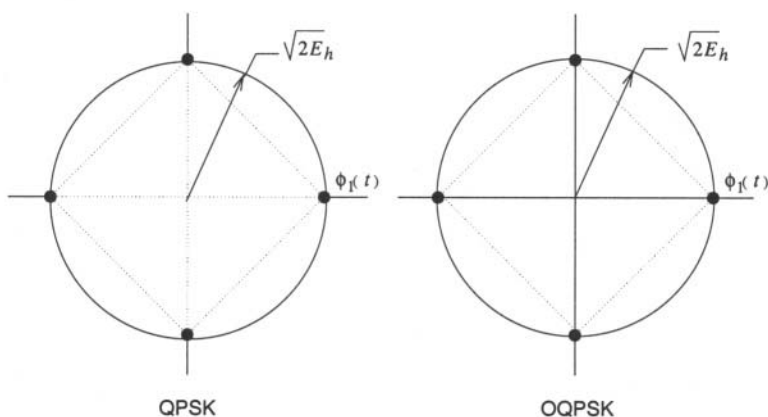
Figure 4.11. Complex signal-space diagram QPSK and OQPSK signals.

T_b seconds. With OQPSK, the amplitude shaping pulse $h_a(t)$ is often chosen to be the root raised cosine pulse in (4.49).

The signal-space diagrams for QPSK and OQPSK are shown in Fig. 4.11, where E_h is the symbol energy. The dotted lines in Fig. 4.11 show the allowable phase transitions. The exact phase trajectories depend on the amplitude shaping function. Note that the phase trajectories do not pass through the origin. This property reduces the peak-to-average ratio of the complex envelope, making the OQPSK signal less sensitive to amplifier nonlinearities than the QPSK signal. It also reduces the dynamic range that is required of the power amplifier.

4.2 $\pi/4$-**DQPSK**

QPSK transmits 2 bits/baud by transmitting sinusoidal pulses having one of 4 absolute carrier phases. $\pi/4$-DQPSK also transmits 2 bits/baud, but information is encoded into the differential carrier phase, and sinusoidal pulses having one of 8 absolute carrier phases are transmitted at each baud epoch.

Let θ_n be the absolute carrier phase for the nth data symbol, and let $\Delta\theta_n = \theta_n - \theta_{n-1}$ be the differential carrier phase. With $\pi/4$-DQPSK, the differential phase is related to the quaternary data sequence $\{x_n\}$, $x_n \in \{\pm1, \pm3\}$ through the mapping

$$\Delta\theta_n = x_n \frac{\pi}{4} \; . \tag{4.67}$$

Notice that the phase differences are $\pm\pi/4$ and $\pm3\pi/4$. The complex envelope of the $\pi/4$-DQPSK signal is

$$\tilde{s}(t) = A \sum_n b(t - nT, \mathbf{x}_n) \tag{4.68}$$

where

$$
\begin{aligned}
b(t, \mathbf{x}_n) &= h_a(t) \exp\left\{ j \left(\theta_{n-1} + x_n \frac{\pi}{4} + \theta_o \right) \right\} \\
&= h_a(t) \exp\left\{ j \frac{\pi}{4} \left(\sum_{k=-\infty}^{n-1} x_k + x_n \right) + j\theta_o \right\} \; . \tag{4.69}
\end{aligned}
$$

The summation in the exponent represents the accumulated carrier phase, while the last term is the phase change due to the nth information symbol. Assuming that $\theta_o = 0$, the absolute carrier phase during the even and odd baud intervals belongs to the sets $\{0, \pi/2, \pi, 3\pi/2\}$ and $\{\pi/4, 3\pi/4, 5\pi/4, 7\pi/4\}$, respectively, or vice versa. With $\pi/4$-DQPSK the amplitude shaping pulse $h_a(t)$ is often chosen to be the root raised cosine pulse in (4.49).

The signal-space diagrams for QPSK and $\pi/4$-DQPSK are shown in Fig. 4.12, where E_h is the symbol energy. The dotted lines in Fig. 4.12 show the allowable phase transitions. The phaser diagram for $\pi/4$-DQPSK with root raised cosine amplitude pulse shaping is shown in Fig. 4.13. Note that the phase trajectories do not pass through the origin. Like OQPSK, this property reduces the peak-to-average ratio of the complex envelope, making the $\pi/4$-DQPSK signal less sensitive to amplifier nonlinearities. Finally, we observe that unlink QPSK the carrier phase of $\pi/4$-DQPSK changes by $\pm\pi/4$ or $\pm3\pi/4$ radians during *every* baud interval. This property makes symbol synchronization is easier with $\pi/4$-DQPSK as compared to QPSK.
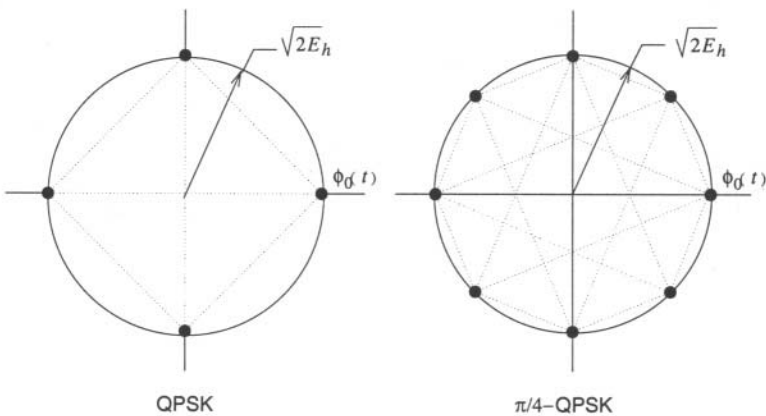
Figure 4.12. Complex signal-space diagram QPSK and $\pi/4$-DQPSK signals.
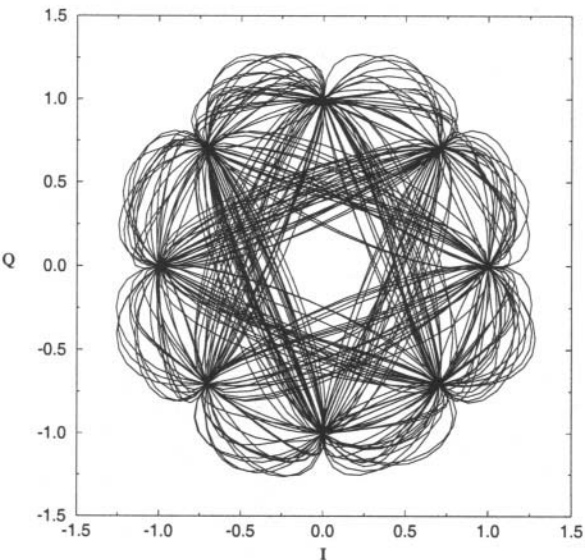
Figure 4.13. Phaser diagram for $\pi/4$-DQPSK with square root raised cosine amplitude pulse shaping; $\beta = 0.5$.

5. ORTHOGONAL MODULATION AND VARIANTS

Orthogonal modulation schemes transmit information by using a set of waveforms, $\{\tilde{s}_m(t)\}_{m=0}^{M-1}$ that are orthogonal in time. Many different types of orthogonal waveforms can be constructed and here we consider a few methods.

Orthogonal FSK modulation:.   Orthogonal M-ary frequency shift keying (MFSK) modulation uses a set of M waveforms that have different frequencies. The MFSK complex envelope is

$$\tilde{s}(t) = A \sum_n b(t - nT, \mathbf{x}_n) \tag{4.70}$$

where

$$b(t, \mathbf{x}_n) = e^{j x_n \pi \Delta_f t} u_T(t) \tag{4.71}$$

and $x_n \in \{\pm 1, \pm 3, \ldots, \pm M - 1\}$. The MFSK waveforms that are transmitted at each baud epoch have complex envelopes

$$\tilde{s}_m(t) = A e^{j x_m \pi \Delta_f t} u_T(t) , \quad m = 0, \ldots, M - 1 . \tag{4.72}$$

By choosing the frequency separation $\Delta_f = 1/2T$, the $\tilde{s}_m(t), m = 0, \ldots, M - 1$ are orthogonal (see Problem 4.1). Since the $\tilde{s}_m(t)$ are orthogonal, the MFSK signal vectors have dimension $N = M$. The appropriate set of basis functions is

$$\varphi_i(t) = \sqrt{\frac{A^2}{2 E_h}} u_T(t) , \quad i = 0, \ldots, M - 1 \tag{4.73}$$

where

$$E_h = \frac{A^2 T}{2} \tag{4.74}$$

is the energy in the band-pass pulse $A u_T(t) \cos 2\pi f_c t$. The MFSK complex signal vectors are

$$\tilde{\mathbf{s}}_m = \sqrt{2 E_h} \mathbf{e}_m , \quad m = 0, \ldots, M - 1 \tag{4.75}$$

where $\mathbf{e}_m = (e_0, e_1, \ldots, e_{M-1})$ is a vector of length M with a "1" in the mth co-ordinate and zeros elsewhere.

Orthogonal modulation with binary orthogonal codes:.   Another type of orthogonal modulation starts with the rows of a Hadamard matrix $\mathbf{H}_M$. The Hadamard matrix is generated recursively according to

$$\mathbf{H}_M = \begin{bmatrix} \mathbf{H}_{M/2} & \mathbf{H}_{M/2} \\ \mathbf{H}_{M/2} & -\mathbf{H}_{M/2} \end{bmatrix} .$$

where $\mathbf{H}_1 = [1]$. For example,

$$
\mathbf{H}_8 = \begin{bmatrix}
+1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\
+1 & -1 & +1 & -1 & +1 & -1 & +1 & -1 \\
+1 & +1 & -1 & -1 & +1 & +1 & -1 & -1 \\
+1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 \\
+1 & +1 & +1 & +1 & -1 & -1 & -1 & -1 \\
+1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 \\
+1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 \\
+1 & -1 & -1 & +1 & -1 & +1 & +1 & -1
\end{bmatrix} . \tag{4.76}
$$

Notice that the rows of the Hadamard matrix are mutually orthogonal. A set of equal energy M orthogonal waveforms can be constructed according to

$$
\tilde{s}_m(t) = A \sum_{k=1}^{M} h_{m_k} h_c(t - kT_c) , \quad m = 0, \ldots, M-1 \tag{4.77}
$$

where h_{m_k} is the kth co-ordinate in the mth row of the Hadamard matrix, $T = MT_c$ is the symbol duration, and $h_c(t)$ is a shaping pulse either having duration T_c or satisfying Nyquist's first criterion with equally spaced zero crossings at intervals of T_c seconds. The energy in the waveform $\tilde{s}_m(t)$ is

$$
E_h = \frac{MA^2}{2} \int_{-\infty}^{\infty} h_c^2(t) dt . \tag{4.78}
$$

To construct signal vectors, the appropriate choice of basis function is

$$
\varphi_i(t) = \frac{A}{E_h} \sum_{k=1}^{M} h_{i_k} h_c(t - kT_c) , \quad i = 0, \ldots, M-1 \tag{4.79}
$$

and once again

$$
\tilde{\mathbf{s}}_m = \sqrt{2E_h}\, \mathbf{e}_m , \quad m = 0, \ldots, M-1 . \tag{4.80}
$$

Biorthogonal signals:. A set of M biorthogonal signals can be easily constructed from a set of $M/2$ orthogonal signals. The M-ary biorthogonal waveforms have complex signal vectors

$$
\tilde{\mathbf{s}}_i = \begin{cases} \sqrt{2E_h}\, \mathbf{e}_i , & i = 0, \ldots, M/2-1 \\ -\tilde{\mathbf{s}}_{i-M/2} , & i = M/2, \ldots, M-1 \end{cases} \tag{4.81}
$$

where the vectors $\mathbf{e}_i$ have length $M/2$. By using an appropriate set of basis functions, for example in (4.73) or (4.79), the complex envelopes of the signal waveforms can be easily constructed.

Orthogonal multipulse modulation:.   With binary orthogonal codes only $k = \log_2 M$ bits are transmitted at each baud epoch. A more bandwidth efficient scheme can be obtained by using the rows of the Hadamard matrix $\mathbf{H}_N$ to define N orthogonal amplitude shaping pulses

$$h_i(t) = A \sum_{k=0}^{N-1} h_{ik} h_c(t - kT_c), \quad i = 0, \dots, N-1 \tag{4.82}$$

With orthogonal multipulse modulation, a block of N serial data symbols, each of duration T_c, is first converted into a block of N parallel data symbols. The block of N information symbols is transmitted in parallel by using the N orthogonal amplitude shaping pulses in (4.82). The transmitted complex envelope is

$$\tilde{s}(t) = \sum_n b(t - nT, \mathbf{x}_n) \tag{4.83}$$

where

$$b(t, \mathbf{x}_n) = \sum_{k=0}^{N-1} x_{n_k} h_k(t) \tag{4.84}$$

$T = NT_c$, and $\mathbf{x}_n = (x_{n_0}, x_{n_1}, \dots, x_{n_{N-1}})$ is the block of N data symbols transmitted at epoch n.

6.   ORTHOGONAL FREQUENCY DIVISION MULTIPLEXING (OFDM)

Orthogonal frequency division multiplexing (OFDM) is a modulation technique that has been suggested for use in cellular radio [54, 39], digital audio broadcasting [119], digital video broadcasting, and wireless LAN systems such as IEEE 802.11, HIPERLAN, and MMAC [333]. OFDM is a block modulation scheme where data symbols are transmitted in parallel by employing a (large) number of orthogonal sub-carriers. A block of N serial data symbols, each of duration T_s, is converted into a block of N parallel data symbols, each with duration $T = NT_s$. The N parallel data symbols modulate N sub-carriers that are spaced $1/T$ Hz apart.

The complex envelope of an OFDM signal is given by

$$\tilde{s}(t) = A \sum_n b(t - nT, \mathbf{x}_n) \tag{4.85}$$

where

$$b(t, \mathbf{x}_n) = h_a(t) \sum_{k=0}^{N-1} x_{n_k} \exp\left\{ j \frac{2\pi \left(k - \frac{N-1}{2}\right) t}{T} \right\} \tag{4.86}$$

where n is the block index, N is the block length, $\mathbf{x}_n = \{x_{n_0}, x_{n_1}, \ldots, x_{n_{N-1}}\}$ is the data symbol block at epoch n, and The frequency offset

$$\exp\left\{-j\frac{\pi(N-1)t}{T}\right\}$$

just ensures that band-pass signal is centered about the carrier frequency.

The data symbols x_{n_k} are often chosen from a QAM or PSK constellation, although any 2-D signal constellation can be used. If a rectangular shaping pulse $h_a(t) = Au_T(t)$ is chosen, then the $1/T$ Hz frequency separation of the sub-carriers ensures that they are orthogonal regardless of the random phases that are imparted due to data modulation (see Problem 4.4). As we will see later, other choices for $h_a(t)$ may result in a more compact psd, but the error rate performance will degrade due to the loss of sub-channel orthogonality.

The OFDM system typically operates over a non-ideal channel with transfer function $T(f)$, such that the amplitude response $|T(f)|$ is not constant across the channel bandwidth W. The power spectral density of the additive Gaussian noise $S_{nn}(f)$ may not be constant either. Shannon [299] proved that the capacity of a non-ideal additive Gaussian noise channel is achieved when the transmitted power $\Omega_t(f)$ is adjusted across the bandwidth W according to

$$\Omega_t(f) = \left\{ \begin{array}{ll} K - S_{nn}(f)/|T(f)|^2, & f \in W \\ 0, & f \notin W \end{array} \right\} \tag{4.87}$$

where K is a constant chosen to satisfy the constraint

$$\int_W \Omega_t(f) df \leq \Omega_{\text{av}} \tag{4.88}$$

and Ω_{av} is the average available power to the transmitter. One method to achieve capacity is to divide the bandwidth W into N sub-bands of width W/Δ_f, where $\Delta_f = 1/T$ is chosen small enough so that $|T(f)|^2/S_{nn}(f)$ is approximately constant within each sub-band. The signals in each sub-band may then be transmitted with the optimum power allocation $\Omega_t(f)$, while being individually coded to achieve capacity.

If the number of sub-carriers N is chosen so that $\Delta_f = 1/T$ is essentially constant across each sub-band, then no equalization is necessary because the ISI is negligible. Viewing the problem another way, if the block length N is chosen so that $T = NT_s \gg LT_s$, where LT_s is the length of the channel impulse response, then the effect of the intersymbol interference (ISI) is greatly reduced. To eliminate the ISI altogether at the expense of a small decrease in capacity, a **guard interval** of length $GT_s \geq LT_s$ can be inserted between successively modulated OFDM blocks.

It is clear from (4.86) that the data symbols $x_{k,n}$ for fixed n modulate the nth sub-carrier. From (4.87), the transmitter power should be high when

$|T(f)|^2/S_{nn}(f)$ is large and small when $T(f)/S_{nn}(f)$ is small. In a practical system with a target bit error rate, this implies the use of a larger signal constellation in sub-bands where $|T(f)|^2/S_{nn}(f)$ is larger. The technique of using different sized signal constellations on the different OFDM sub-carriers is sometimes called **discrete multitone modulation** (DMT).

6.1   MULTIRESOLUTION MODULATION

Multiresolution modulation (MRM) refers to a class of modulation techniques where multiple classes of bit streams are transmitted simultaneously that differ in their rates and error probabilities. MRM is easy to implement in OFDM schemes by using multiplexed, interleaved, and embedded signal constellations. Multiplexed MRM divides the sub-carriers into contiguous blocks, and a different size signal constellation and transmit power is used in each block. A larger signal constellation will transmit more bits/baud but also require a higher signal-to-noise ratio to achieve a given error probability. Interleaved MRM interleaves the different classes of bit streams onto the sub-carriers in a cyclic fashion, i.e., if there are K different classes of bit streams, then the sub-carriers $\ell, \ell + K, \ell + 2K, \ldots$, are assigned to the ℓth bit stream. Each class of bit stream can be transmitted by using a different sized signal constellation and power level.

Embedded MRM is more subtle and relies upon the use of asymmetric signal constellations. Fig. 4.14 shows an example of a 16-QAM MRM signal constellation, that can be used to transmit two different classes of bit streams, called low priority (LP) and high priority (HP). In Fig. 4.14, two HP bits are used to select the quadrant of the transmitted signal point, while two LP bits are used to select the signal point within the selected quadrant. In order to control the relative error probability between the two priorities a parameter $\lambda = d^l/d^h$ is used, where d^l is the distance between LP symbols and d^h is the distance between centroids in the HP symbols. In general, λ should be less than 0.5, since the MRM constellation becomes symmetric 16-QAM at $\lambda = 0.5$. As λ becomes smaller, more power is allocated to the HP bits and, hence, they are received with a smaller error probability.

6.2   FFT-BASED OFDM SYSTEM

A key advantage of using OFDM is that the modulation and demodulation can be achieved in the discrete-domain by using a discrete Fourier transform (DFT). The fast Fourier transform (FFT) algorithm efficiently implements the DFT.

Consider block $n = 0$ in (4.86) and ignore the frequency offset

$$\exp\left\{-j\frac{\pi(N-1)t}{T}\right\}$$
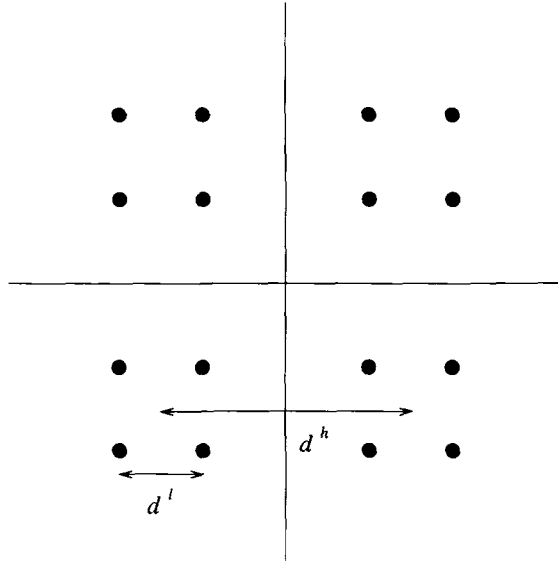
Figure 4.14. 16-QAM embedded MRM signal constellation, defining two priority classes.

Further assume that $h_a(t) = u_T(t)$. Then the complex envelope has the form[2]

$$\tilde{s}(t) = A \sum_{k=0}^{N-1} x_k \exp\left\{\frac{j2\pi kt}{NT_s}\right\} u_T(t) \ .$$ (4.89)

Suppose the complex envelope is sampled at epochs $t = nT_s$ to yield the sequence

$$X_n = \tilde{s}(nT_s) = A \sum_{k=0}^{N-1} x_k \exp\left\{\frac{j2\pi kn}{N}\right\} , \quad n = 0, 1, \ldots, N-1 \ .$$ (4.90)

Notice that the vector $\mathbf{X} = \{X_n\}_{n=0}^{N-1}$ is just the inverse DFT (IDFT) of the vector $A\mathbf{x} = A\{x_k\}_{k=0}^{N-1}$. After taking the IDFT, the sample sequence $\{X_n\}_{n=0}^{N-1}$ can be passed through a D/A converter and carrier modulated.

When the OFDM modulator is implemented as using an IFFT algorithm, the amplitude shaping pulse $h_a(t)$ is no longer the ideal rectangular pulse $u_T(t)$. Rather, the pulse is generated by using a discrete-time approximation to the rectangular pulse. To obtain this pulse, we pass the rectangular impulse train

$$\delta_T(t) = \sum_{k=0}^{N-1} \delta(t - kT_s)$$ (4.91)

---

[2]For the remainder of our discussion we remove the block index 0.
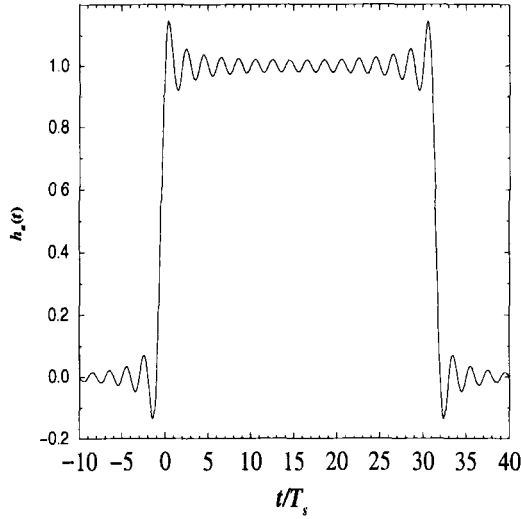
Figure 4.15.   Time domain OFDM amplitude shaping pulse.

through an ideal low-pass filter with impulse response

$$h(t) = \text{sinc}(t/T_s) \ . \tag{4.92}$$

The gives the amplitude shaping pulse

$$h_a(t) = \sum_{k=0}^{N-1} \text{sinc}(t/T_s - k) \ . \tag{4.93}$$

which is plotted in Fig. 4.15. Notice that the pulse is non-causal. As discussed in Section 9.6 this has some interesting implications for the transmitted power spectrum.

Another key advantage of OFDM is the ease by which the effects of ISI can be mitigated. A guard interval consisting of a cyclic prefix or suffix of length G can be appended to the sequence $\mathbf{X}$. Assuming a cyclic suffix, the transmitted sequence with guard interval is

$$
\begin{aligned}
X_n^g &= X_{(n)_N} & (4.94)\\
&= A \sum_{k=0}^{N-1} x_k \exp\left\{\frac{j2\pi kn}{N}\right\} , \quad n = 0, 1, \ldots, N + G - 1 & (4.95)
\end{aligned}
$$

where G is the length of the guard interval in samples, and $(n)_N$ is the residue of n modulo N. To avoid a reduction in data rate, the baud duration with guard
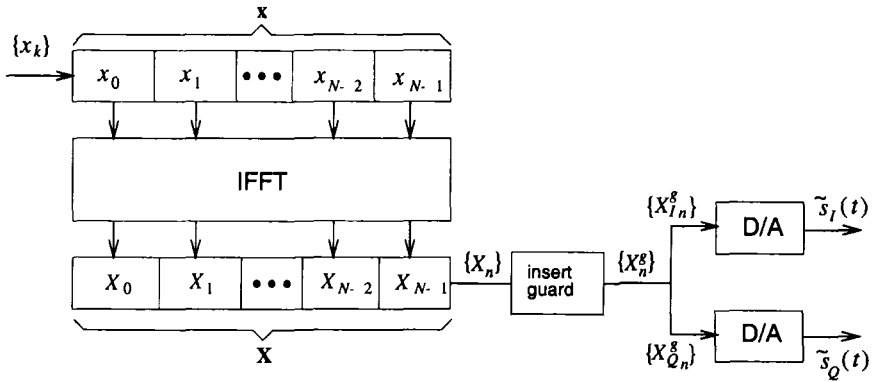
Figure 4.16. Block diagram of OFDM transmitter.

interval is $T_s^g = T_s/(1 + \frac{G}{N})$. The overall OFDM baseband modulator simply consists of an IFFT circuit to implement the IDFT followed by a D/A converter, as shown in Fig. 4.16.

The combination of the D/A converter, waveform channel $g(t)$, anti aliasing filter, and A/D converter yields an equivalent discrete-time channel with sampled impulse response $\{g_m\}_{m=0}^{L}$, where LT_s is the length of the channel impulse response. The discrete-time convolution of the transmitted sequence $\{X_n^g\}_{n=0}^{N+G-1}$ with the discrete-time channel produces the received sequence $\{R_n^g\}$, where

$$R_n^g = \sum_{m=0}^{L} g_m X_{n-m}^g \ .$$ (4.96)

Note that for our present discussion we have neglected the effects of noise. The length of the guard interval, G, is assumed to equal or exceed the channel length, L.

When a block is received, the first $G \geq L$ samples are assumed to be corrupted by ISI from the previous block. The ISI is removed by replacing these samples with the cyclic suffix according to

$$
\begin{aligned}
R_n &= R_{G+(n-G)_N}^g \\
&= \sum_{m=0}^{L} g_m X_{(n-m)_N} \ , \quad 0 \leq n \leq N-1 \ .
\end{aligned}
$$ (4.97)

This operation is illustrated in Fig. 4.17.
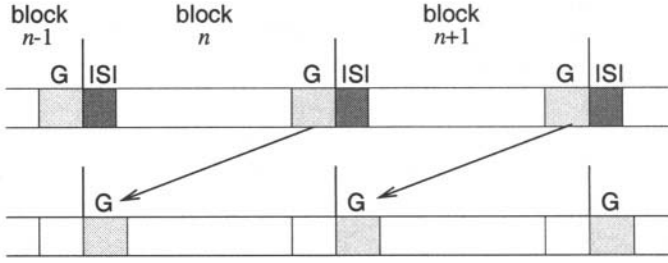
Figure 4.17.   Removal of ISI by using the cyclic suffix.
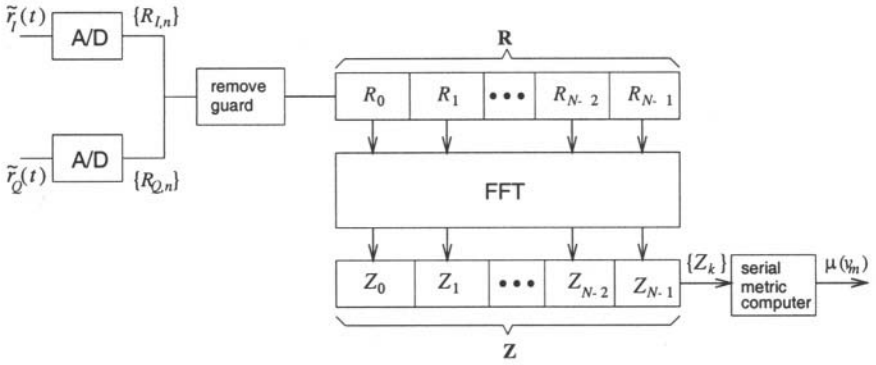
Figure 4.18.   Block diagram of OFDM receiver.

As shown in Fig. 4.18, the OFDM demodulator then performs an FFT on the vector $\mathbf{R} = \{R_n\}_{n=0}^{N-1}$. The demodulated sequence is

$$
\begin{aligned}
Z_i &= \frac{1}{N} \sum_{n=0}^{N-1} R_n e^{-j\frac{2\pi n i}{N}} \\
&= \eta_i A x_i, \quad 0 \le i \le N-1
\end{aligned}
\tag{4.98}
$$

where

$$
\eta_i = \sum_{m=0}^{L} g_m e^{-j\frac{2\pi m i}{N}}
\tag{4.99}
$$

Notice that Z_i is equal to $A x_i$ multiplied by the equivalent complex channel gain η_i. Hence, the ISI due to the channel has been completely removed. When noise is present, then the Z_i must be used to make data symbol decisions. This is the purpose of the serial metric computer in Fig. 4.18. The metric computer will be discussed further in Chapter 5.

7.   CONTINUOUS PHASE MODULATION (CPM)

Continuous phase modulation (CPM) refers to a broad class of frequency modulation techniques where the carrier phase varies in a continuous manner. A comprehensive treatment of CPM is provided in Anderson *et. al.* [12]. CPM schemes are attractive because they have constant envelope and excellent spectral characteristics, i.e., a narrow main lobe and fast roll-off of sidelobes. The complex envelope of a general CPM waveform has the form

$$\tilde{s}(t) = A \exp\left\{j\left(\phi(t) + \theta_o\right)\right\} \tag{4.100}$$

where A is the amplitude, θ_o is initial carrier phase at $t = 0$, and

$$\phi(t) = 2\pi \int_0^t \sum_{k=0}^{\infty} h_k x_k h_f(\tau - kT) d\tau \tag{4.101}$$

The term $\phi(t)$ is called the **excess phase.** In (4.101), the symbols are defined as follows:

- $\{x_k\}$ is the data symbol sequence and T is the baud period. The data symbols are chosen from the alphabet $\{\pm 1, \pm 3, \cdots, \pm(M-1)\}$, where M is the modulation alphabet size.

- $\{h_k\}$ is the sequence of modulation indices. When $h_k = h$ the modulation index is fixed for all symbols. With multi-h CPM, the sequence $\{h_k\}$ is chosen in a cyclic fashion from set $\{\hat{h}_1, \hat{h}_2, \ldots, \hat{h}_H\}$ of H modulation indices. That is, $h_{i+H} = h_i$.

- $h_f(t)$ is the **frequency shaping function**, that is zero for $t < 0$ and $t > LT$, and normalized to have an area equal to 1/2. A **full response** CPM has $L = 1$, while **partial response** CPM has $L > 1$. Some possible frequency shaping pulses are shown in Table 4.1. A more compact power density spectrum is obtained by using frequency shaping functions having continuous higher-order derivatives, such as the raised cosine pulse in Table 4.1.

An infinite variety of CPM signals can be generated by choosing different frequency shaping pulses, modulation indices, and modulation alphabet sizes.

It is useful to define the **phase shaping function**, $\beta(t)$, as

$$\beta(t) = \begin{cases} 0 & , \ t < 0 \\ \int_0^t h_f(\tau) d\tau & , \ 0 \leq t \leq LT \\ 1/2 & , \ t \geq LT \end{cases} \tag{4.102}$$

pulse type	$h_f(t)$
L-rectangular (LREC)	$\frac{1}{2LT} u_{LT}(t)$
L-raised cosine (LRC)	$\frac{1}{2LT}\left[1 - \cos\left(\frac{2\pi t}{T}\right)\right] u_{LT}(t)$
L-half sinusoid (LHS)	$\frac{\pi}{4LT}\sin(\pi t/T) u_{LT}(t)$
L-triangular (LTR)	$\frac{1}{LT}\left(1 - \frac{|t - LT/2|}{LT/2}\right)$

Table 4.1.   CPM frequency shaping functions.

7.1   FULL RESPONSE CPM

Consider a full response CPM signal with a single modulation index, $h_i = h$. Within the time interval $[nT, (n + 1)T]$, the excess phase $\phi(t)$ is

$$\phi(t) = 2\pi h \int_0^{nT} \sum_{k=0}^{n-1} x_k h_f(\tau - kT)d\tau + 2\pi h \int_{nT}^t x_n h_f(\tau - nT)d\tau$$

$$= \pi h \sum_{k=0}^{n-1} x_k + 2\pi h x_n \beta(t - nT) \tag{4.103}$$

The first term in (4.103) represents the accumulated excess phase up to time nT, while the second term represents the excess phase increment for t within the interval $[nT, (n + 1)T]$. Note that the phase is continuous so long as the frequency shaping function $h_f(t)$ does not contain impulses, which accounts for all practical cases.

Since (4.103) represents the phase within the interval $[nT, (n + 1)T]$, the corresponding complex envelopes for all such intervals can be concatenated together to write the full response CPM complex envelope as

$$\tilde{s}(t) = A \sum_n b(t - nT, \mathbf{x}_n) \tag{4.104}$$

where

$$b(t, \mathbf{x}_n) = \exp\left\{j\left(\pi h \sum_{k=0}^{n-1} x_k + 2\pi h x_n \beta(t)\right)\right\} u_T(t) \tag{4.105}$$

where we have assumed an initial phase $\theta_o = 0$.
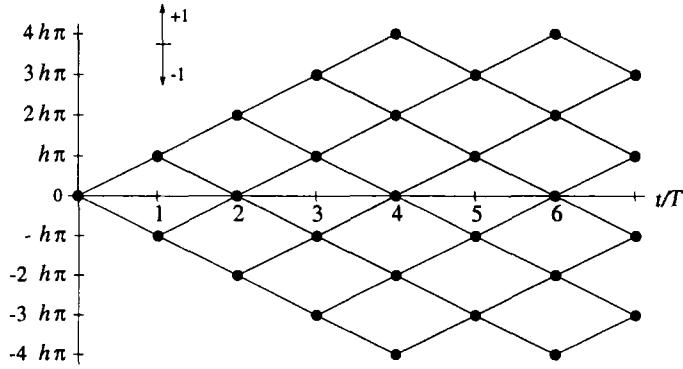
Figure 4.19. Phase tree of binary CPFSK with an arbitrary modulation index. CPFSK is characterized by linear phase trajectories.

Continuous phase frequency shift keying (CPFSK) is a special type of full response CPM obtained by using the rectangular frequency shaping function LREC with $L = 1$. For CPFSK

$$\beta(t) = \begin{cases} 0 & , t < 0 \\ t/2T & , 0 \le t \le T \\ 1/2 & , t \ge T \end{cases} . \tag{4.106}$$

CPM signals can be visualized by sketching the evolution of the excess phase $\phi(t)$ for all possible data sequences. This plot is called a **phase tree**, and a typical phase tree is shown in Fig. 4.19 for binary CPFSK. Since the CPFSK frequency shaping function is rectangular, the phase trajectories are linear as suggested by (4.106). In each baud interval, the phase increases by πh if the data symbol is +1 and decreases by πh if the data symbol is -1.

7.1.1   MINIMUM SHIFT KEYING (MSK)

Minimum shift keying (MSK) is a special case of binary CPFSK, with modulation index $h = 1/2$. The MSK bandpass signal is $s(t) = A\cos\phi_c(t)$, where $\phi_c(t) = 2\pi f_c t + \phi(t) + \theta_o$. Assuming that $\theta_o = 0$, the phase $\phi_c(t)$ within time interval $[nT, (n+1)T]$ can be obtained from (4.103) as

$$\phi_c(t) = 2\pi f_c t + \frac{\pi}{2}\sum_{k=0}^{n-1} x_k + \frac{\pi}{2}x_n \frac{t - nT}{T}$$

$$= \left(2\pi f_c + \frac{\pi x_n}{2T}\right)t + \frac{\pi}{2}\sum_{k=0}^{n-1} x_k - \frac{\pi n}{2}x_n . \tag{4.107}$$

The MSK signal can be described in terms of the phase tree in shown in Fig. 4.19 with $h = 1/2$. At the end of each symbol interval the excess phase
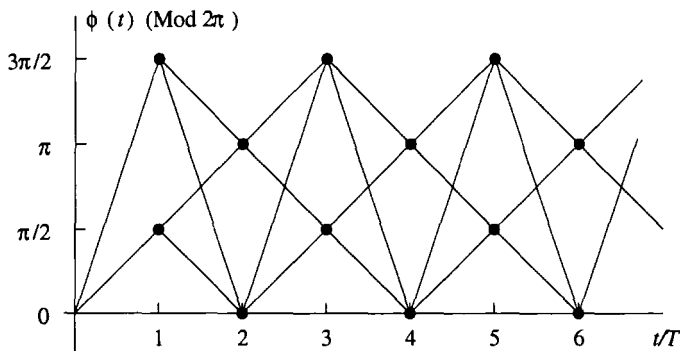
Figure 4.20.   Phase-trellis diagram for MSK.

$\phi(t)$ takes on values that are integer multiples of $\pi/2$. Since excess phases that differ by integer multiples of 2π are indistinguishable, the values taken by $\phi(t)$ at the end of each symbol interval reduced modulo 2π belong to the finite set $\{0, \pi/2, \pi, 3\pi/2\}$. In this fashion, the phase tree collapses into the phase trellis shown in Fig. 4.20.

Consider the MSK band-pass waveform in the interval $[nT, (n+1)T]$, given by

$$s(t) = A \cos\left(2\pi\left(f_c + \frac{x_n}{4T}\right)t + \frac{\pi}{2}\sum_{k=0}^{n-1}x_k - \frac{\pi n}{2}x_n\right) . \qquad (4.108)$$

Observe that the MSK signal has one of two possible frequencies

$$f_L = f_c - \frac{1}{4T} \qquad \text{and} \qquad f_U = f_c + \frac{1}{4T} . \qquad (4.109)$$

The difference between these frequencies is $\Delta f = f_U - f_L = 1/(2T)$. This is the minimum frequency separation to ensure orthogonality between two co-phased sinusoids of duration T (see Problem 4.4) and, hence, the name *minimum* shift keying.

Another interesting form for the MSK signal can be obtained by starting with (4.104) and (4.105). After a lengthy derivation which we omit here, we can write the MSK complex envelope as

$$\tilde{s}(t) = A \sum_n \left(x_{I,n}h_a(t - 2nT) + jx_{Q,n}h_a(t - 2nT - T)\right) \qquad (4.110)$$

where

$$x_{I,n} = -x_{Q,n-1}x_{2n-1} \qquad (4.111)$$

$$x_{Q,n} = x_{I,n}x_{2n} \qquad (4.112)$$

and

$$h_a(t) = \cos\left(\frac{\pi t}{2T}\right) u_{2T}(t + T) \tag{4.113}$$

$$h_a(t - T) = \sin\left(\frac{\pi t}{2T}\right) u_{2T}(t) . \tag{4.114}$$

Note that $\{x_{I,n}\}$ and $\{x_{Q,n}\}$ are independent binary symbol sequences that take on elements from the set $\{-1, +1\}$, and the half sinusoid amplitude shaping pulse $h_a(t)$ has duration $2T$. Notice that the information symbols $x_{I,n}$ and $x_{Q,n}$ are transmitted on the quadrature branches with an offset of T seconds. It follows that MSK is equivalent to OQPSK with a half-sinusoid amplitude shaping pulse. This property can be exploited in practice to generate and detect MSK signals.

8. PARTIAL RESPONSE CPM

Partial response CPM signals have a frequency shaping pulse $h_f(t)$ with duration LT where $L > 1$. Partial response CPM signals have better spectral characteristics than full response CPM signals, i.e., a narrower main lobe and faster roll-off of side lobes.

The partial response frequency shaping function can be written as

$$h_f(t) = \sum_{k=0}^{L-1} h_f(t) u_T(t - kT)$$

$$= \sum_{k=0}^{L-1} h_{f,k}(t - kT) \tag{4.115}$$

where

$$h_{f,k}(t) = h_f(t + kT) u_T(t) . \tag{4.116}$$

Likewise, for the phase shaping function

$$\beta(t) = \sum_{k=0}^{L-1} \beta_k(t - kT) \tag{4.117}$$

where

$$\beta_k(t) = \beta(t + kT) u_T(t) . \tag{4.118}$$

Note that

$$\beta_k(t) = \begin{cases} 0 & , t < 0 \\ \int_0^t h_{f,k}(\tau) d\tau & , 0 \le t \le LT \\ \beta_k(T) & , t \ge T \end{cases} \tag{4.119}$$

and

$$\sum_{k=0}^{L-1} \beta_k(T) = \frac{1}{2} \ . \tag{4.120}$$

An equivalent frequency shaping function of duration T can be derived by noting that the baseband modulating signal has the form

$$
\begin{aligned}
x(t) &= \sum_n x_n h_f(t - nT) \\
&= \sum_n \sum_{k=0}^{L-1} x_n h_{f,k}(t - (n+k)T) \\
&= \sum_m \sum_{k=0}^{L-1} x_{m-k} h_{f,k}(t - mT) \ . \tag{4.121}
\end{aligned}
$$

It follows that

$$x(t) = \sum_m h_f(t - mT, \mathbf{x}_m) \tag{4.122}$$

where

$$h_f(t, \mathbf{x}_m) = \sum_{k=0}^{L-1} x_{m-k} h_{f,k}(t) \tag{4.123}$$

and

$$\mathbf{x}_m = (x_m, \ x_{m-1}, \ \ldots, \ x_{m-L+1}) \ . \tag{4.124}$$

Likewise, an equivalent phase shaping function of duration T can be defined as

$$\beta(t, \mathbf{x}_m) = \sum_{k=0}^{L-1} x_{m-k} \beta_k(t) \ . \tag{4.125}$$

Therefore, the partial response shaping functions $h_f(t)$ and $\beta(t)$ can been replaced by equivalent shaping functions $h_f(t, \mathbf{x}_m)$ and $\beta(t, \mathbf{x}_m)$ of duration T whose value depends on the current data symbol and the $L-1$ past data symbols.

Example 4.2 LREC Frequency Shaping Function_____

For an LREC frequency shaping function

$$h_f(t) = \frac{1}{2LT} u_{LT}(t) \ .$$

Hence,

$$h_f(t, \mathbf{x}_n) = x_n h_{f,0}(t) + x_{n-1} h_{f,1}(t) + \cdots + x_{n-L+1} h_{f,L-1}(t)$$

where

$$h_{f,0}(t) = h_{f,1}(t) = \cdots = h_{f,L-1}(t) = \frac{1}{2LT} u_T(t) \ .$$

Therefore,

$$h_f(t, \mathbf{x}_n) = (x_n + x_{n-1} + \cdots + x_{n-L+1}) \frac{1}{2LT} u_T(t) \ .$$

Example 4.3 LRC Frequency Shaping Function

For an LRC frequency shaping function

$$h_f(t) = \frac{1}{2LT} \left[1 - \cos \left(\frac{2\pi t}{LT} \right) \right] u_{LT}(t) \ .$$

Hence,

$$h_f(t, \mathbf{x}_n) = x_n h_{f,0}(t) + x_{n-1} h_{f,1}(t) + \cdots + x_{n-L} h_{f,L-1}(t)$$

where

$$h_{f,k}(t) = \left[1 - \cos \left(\frac{2\pi(t + kT)}{LT} \right) \right] u_T(t) \ .$$

It follows from the above development that the complex envelope of a partial response CPM signal can be written in the standard form

$$v(t) = A \sum_n b(t - nT, \mathbf{x}_n) \tag{4.126}$$

where

$$b(t, \mathbf{x}_n) = \exp \left\{ j2\pi h \left(\sum_{i=0}^{n-1} \beta(T, \mathbf{x}_i) + \beta(t, \mathbf{x}_n) \right) \right\} u_T(t) \tag{4.127}$$

and we have assumed an initial phase $\theta_o = 0$. The excess phase over the interval $[nT, (n+1)T]$ is

$$\phi(t) \quad = \quad 2\pi h \int_0^t \sum_{k=0}^n x_k h_f(\tau - kT) d\tau \tag{4.128}$$

$$= \quad \pi h \sum_{k=0}^{n-L} x_k + 2\pi h \sum_{k=n-L+1}^n x_k \beta(t - kT) \tag{4.129}$$

$$= \quad \theta_n + 2\pi h \sum_{k=n-L+1}^{n-1} x_k \beta(t - kT) + 2\pi h x_n \beta(t - nT) \tag{4.130}$$

where

$$\theta_n = \pi h \sum_{k=0}^{n-L} x_k \qquad \text{modulo } 2\pi \tag{4.131}$$

is the accumulated **phase state.** During the interval $[nT, (n+1)T]$, the excess phase depends on the data symbol x_n, the $L - 1$ previous data symbols,

$\{x_{n-1}, x_{n-2}, \ldots, x_{n-L+1}\}$, and the accumulated phase state θ_n. The state of the CPM signal at time $t = nT$, is defined by the L-tuple

$$S_n = (\theta_n, x_{n-1}, x_{n-2}, \ldots, x_{n-L+1}) \qquad (4.132)$$

Since the vector $(x_{n-1}, x_{n-2}, \ldots, x_{n-L+1})$ can take on M^{L-1} values, the number of states equals M^{L-1} times the number of values that θ_n can take on.

The modulation index is often restricted to be a rational number, $h = m/p$, where m and p are integers that have no common factors. This constraint ensures that the number of phase states is finite which required for the implementation of some types of CPM receivers. If m is even, then

$$\theta_n \in \left\{ 0, \frac{\pi m}{p}, \frac{2\pi m}{p}, \cdots, \frac{(p-1)\pi m}{p} \right\} \qquad (4.133)$$

while if m is odd

$$\theta_n \in \left\{ 0, \frac{\pi m}{p}, \frac{2\pi m}{p}, \cdots, \frac{(2p-1)\pi m}{p} \right\} . \qquad (4.134)$$

Hence, there are p phase states for even m, while there are $2p$ phase states for odd m. In conclusion, the number of CPM states is

$$S = \begin{cases} pM^{L-1}, & m \text{ even} \\ 2pM^{L-1}, & m \text{ odd} \end{cases} \qquad (4.135)$$

For example, if $h = 1/4$, $M = 4$, and $L = 2$, then

$$\theta_n \in \left\{ 0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}, \pi, \frac{5\pi}{4}, \frac{3\pi}{2}, \frac{7\pi}{4} \right\} \qquad (4.136)$$

and the number of states is 32.

CPM signals cannot be described in terms of a signal-space diagram, like QAM and PSK. However, the CPM signal can be described in terms of the trajectories from one phase state to another. Figs. 4.22 and 4.21 show the phase state diagrams for MSK and binary CPM with $h = 1/4$, respectively. Since binary modulation is used, the phase trajectories are only allowed to adjacent phase states as shown by the dotted lines in the figures.

8.1   GAUSSIAN MINIMUM SHIFT KEYING (GMSK)

It will be shown in Section 9. that MSK has all the desirable attributes for mobile radio systems, except for a compact psd. This can be alleviated by low-pass filtering the modulating signal

$$x(t) = \sum_{n=-\infty}^{\infty} x_n h_f(t - nT) = \frac{1}{2T} \sum_{n=-\infty}^{\infty} x_n u_T(t - nT) \qquad (4.137)$$
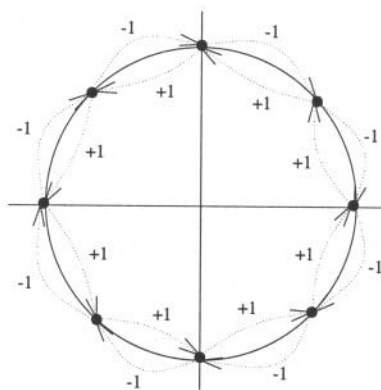
Figure 4.21. Phase state diagram for binary CPM signals with $h = 1/4$.
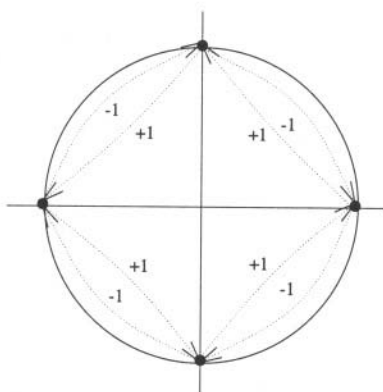
Figure 4.22. Phase state diagram for MSK signals.
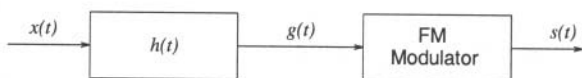
Figure 4.23. Pre-modulation filtered MSK. The modulating signal is low-pass filtered to remove the high frequency components prior to modulation.

prior to modulation, as shown in Fig. 4.23. Such filtering removes the higher frequency components in $x(t)$ and results in a more compact psd.

GMSK is a special type of partial response CPM that uses a low-pass pre-modulation filter having the transfer function

$$H(f) = \exp\left\{-\left(\frac{f}{B}\right)^2 \frac{\ln 2}{2}\right\}$$

(4.138)

where B is the 3 dB bandwidth of the filter. It is apparent that $H(f)$ is bell shaped about $f = 0$, hence the name Gaussian MSK. Transmitting the rectangular pulse

$$\frac{1}{2T}\text{rect}(t/T) = \frac{1}{2T}u_T(t + T/2)$$

through this filter yields the frequency shaping pulse

$$h_f(t) = \frac{1}{2T}\sqrt{\frac{2\pi}{\ln 2}}(BT)\int_{t/T-1/2}^{t/T+1/2}\exp\left\{-\frac{2\pi^2(BT)^2x^2}{\ln 2}\right\}dx$$

$$= \frac{1}{2T}\left[Q\left(\frac{t/T+1/2}{\sigma}\right) - Q\left(\frac{t/T-1/2}{\sigma}\right)\right] \qquad (4.139)$$

where

$$Q(\alpha) = \int_\alpha^\infty \frac{1}{\sqrt{2\pi}}e^{-x^2}dx \qquad (4.140)$$

$$\sigma^2 = \frac{\ln 2}{4\pi^2(BT)^2} . \qquad (4.141)$$

It is not difficult to show that the total pulse area is $\int_{-\infty}^\infty h_f(t)dt = 1/2$ and, therefore, the total contribution to the excess phase for each data symbol is $\pm\pi/2$.

The excess phase change over the time interval from $-T/2$ to $T/2$ is

$$\phi(T/2) - \phi(-T/2) = \pi x_0\beta_0(T) + \pi\sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} x_n\beta_n(T) \qquad (4.142)$$

where

$$\beta_n(T) = \int_{-T/2-nT}^{T/2-nT} h_f(\tau)\,d\tau . \qquad (4.143)$$

The first term in (4.142) is the desired term, and the second term is the inter-symbol interference (ISI) introduced by the premodulation filter.

Fig. 4.24 plots a GMSK frequency shaping pulse (truncated to $5T$ and time shifted by $2.5T$) for various normalized filter bandwidths BT. Because the frequency shaping pulse has a duration greater than T, ISI is introduced. As BT decreases, the ISI increases. Thus, while a smaller value of BT results in a more compact power density spectrum, the induced ISI will degrade the bit error rate performance. This leads to a tradeoff in the choice of BT. Some studies have indicated that $BT = 0.25$ provides a good tradeoff for mobile radio systems [239].
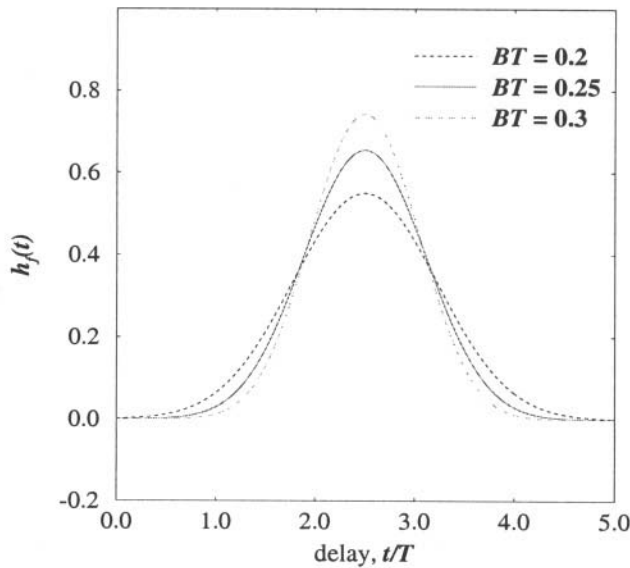
Figure 4.24. GMSK frequency shaping pulse for various normalized filter bandwidths BT.

8.2 LINEARIZED GMSK (LGMSK)

GMSK is a not linear modulation scheme. A linearized GMSK (LGMSK) signal is useful because it simplifies signal generation, receiver algorithms, and performance analysis. Here we derive a simple and accurate linear approximation to GMSK in the discrete-time domain, from which a continuous time GMSK signal can be generated with a D/A converter.

The GMSK phase shaping function is the integral of the frequency shaping function as defined in (4.102). Using integration by parts we can show that

$$\beta(t) = \frac{1}{2}\left(G\left(\frac{x}{T} + \frac{1}{2}\right) - G\left(\frac{x}{T} - \frac{1}{2}\right)\right) \tag{4.144}$$

where

$$G(x) = Q\left(\frac{x}{\sigma^2}\right) + \frac{\sigma}{\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}} \tag{4.145}$$

Since the phase shaping pulse $\beta(t)$ is strictly non-causal, it must be approximated in practice by using a truncated and time shifted version. Here we consider the time shifted pulse

$$\hat{\beta}(t) = \beta(t - T) \ . \tag{4.146}$$

The phase shaping pulse $\hat{\beta}(t)$ is shown in Fig. 4.25.
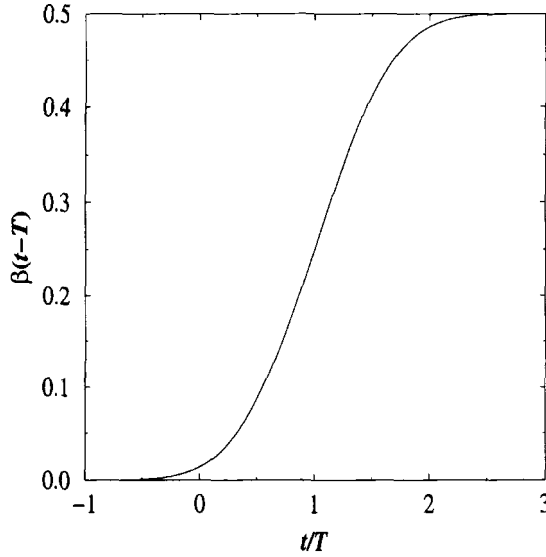
Figure 4.25.   LGMSK phase shaping pulse for $BT = 0.3$.

Suppose that the data symbol sequence $\{x_n\}$ is differentially encoded to yield the sequence

$$y_n = x_n x_{n-1} \quad x \in \{-1, +1\} \tag{4.147}$$

The sequence $\{y_n\}$ is then applied to the GMSK modulator to produce the complex envelope

$$\tilde{s}(t) = \exp\left\{ -j\pi \sum_{k=0}^{n} y_k \hat{\beta}(t - kT) \right\} . \tag{4.148}$$

Now observe from Fig. 4.26 that $\hat{\beta}(t) \approx 0, t \leq 0$ and $\hat{\beta}(t) \approx \frac{1}{2}, t \geq 2T$. We now sample the complex envelope with a sample spacing of J/T to give

$$
\begin{aligned}
\tilde{s}(nT + mT/J) &\approx \exp\left\{ j \sum_{k=0}^{n} y_k \hat{\beta}((n-k)T + mT/J) \right\} \\
&= \Pi_{k=0}^{n-2} j y_k \cdot \exp\left\{ j \sum_{k=n-1}^{n} y_k \hat{\beta}((n-k)T + mT/J) \right\} \\
&= j^{n-1} y_0 y_{n-2} \cdot \exp\left\{ j(y_{n-2} y_{n-1} \hat{\beta}(T + mT/J) \right. \\
&\qquad \left. + y_{n-1} y_n \hat{\beta}(mT/J) \right\} \\
&= j^{n-1} y_0 y_{n-2} \left[\cos(y_{n-2} y_{n-1} \hat{\beta}(T + mT/J)) \text{ (cont'd)} \right.
\end{aligned}
$$

$$+ j\sin(y_{n-2}y_{n-1}\hat{\beta}(T + mT/J))\Big]$$

$$\times \Big[\cos(y_{n-1}y_n\hat{\beta}(mT/J))$$

$$+j\sin(y_{n-1}y_n\hat{\beta}(mT/J))\Big] \tag{4.149}$$

By using the fact that $\cos(x)$ is even function and $\sin(x)$ is odd function, and $y_{n-1}y_n \in \{-1, +1\}$, the cosine and sine terms can be rewritten as

$$
\begin{aligned}
\tilde{s}(nT + mT/J) \;\approx\; & j^{n-1}y_0 y_{n-2} \Big[\cos(\hat{\beta}(T + mT/J)) \\
& +js_{n-2}s_{n-1}\sin(\hat{\beta}(T + mT/J))\Big] \\
& \times \Big[\cos(\hat{\beta}(mT/J)) + js_{n-1}s_n\sin(\hat{\beta}(mT/J))\Big]
\end{aligned} \tag{4.150}
$$

Since $y_n^2 = 1$, the sampled signal reduces to

$$
\begin{aligned}
\tilde{s}(nT + mT/J) \;\approx\; & j^{n-1}y_0 y_{n-2}\Big[\cos(\hat{\beta}(T + mT/J))\cos(\hat{\beta}(mT/J)) \\
& +jy_{n-1}y_n\cos(\hat{\beta}(T + mT/J))\sin(\hat{\beta}(mT/J)) \\
& +jy_{n-2}y_{n-1}\sin(\hat{\beta}(T + mT/J))\cos(\hat{\beta}(mT/J)) \\
& +j^2 s_{n-2}s_n\sin(\hat{\beta}(T + mT/J))\sin(\hat{\beta}(mT/J))\Big] \\
=\; & j^{n-1}\cos(\hat{\beta}(T + mT/J))\cos(\hat{\beta}(mT/J))y_{n-2} \\
& +j^n\Big[\cos(\hat{\beta}(T + mT/J))\sin(\hat{\beta}(mT/J))y_{n-2}y_{n-1}y_n \\
& + \sin(\hat{\beta}(T + mT/J))\cos(\hat{\beta}(mT/J))y_{n-1}\Big] \\
& +j^{n+1}\sin(\hat{\beta}(T + mT/J))\sin(\hat{\beta}(mT/J))y_n
\end{aligned} \tag{4.151}
$$

Finally, we can eliminate the nonlinear terms, since they are much smaller than the others. This leads to the LGMSK complex envelope

$$
\begin{aligned}
\tilde{s}(nT + mT/J) \;\approx\; & j^{n-1}\cos(\hat{\beta}(T + mT/J))\cos(\hat{\beta}(mT/J))y_{n-2} \\
& +j^n\sin(\hat{\beta}(T + mT/J))\cos(\hat{\beta}(mT/J))y_{n-1} \\
& +j^{n+1}\sin(\hat{\beta}(T + mT/J))\sin(\hat{\beta}(mT/J))y_n
\end{aligned} \tag{4.152}
$$

The sampled LGMSK shaping pulse can now be obtained from the approximate sampled complex envelope in (4.152). Assuming that $J = 8$ we have the following:

$$
h_g(mT/8) = \begin{cases}
\sin(\hat{\beta}(T + mT/8))\sin(\hat{\beta}(T + mT/8)), \\
\qquad\qquad m = 0, \ldots, 7 \\
\sin(\hat{\beta}(T + (m - 8)T/8))\cos(\hat{\beta}((m - 8)T/8)), \\
\qquad\qquad m = 8, \ldots, 15 \\
\cos(\hat{\beta}(T + (m - 16)T/8))\cos(\hat{\beta}((m - 16)T/8)), \\
\qquad\qquad m = 16, \ldots, 24
\end{cases} \tag{4.153}
$$
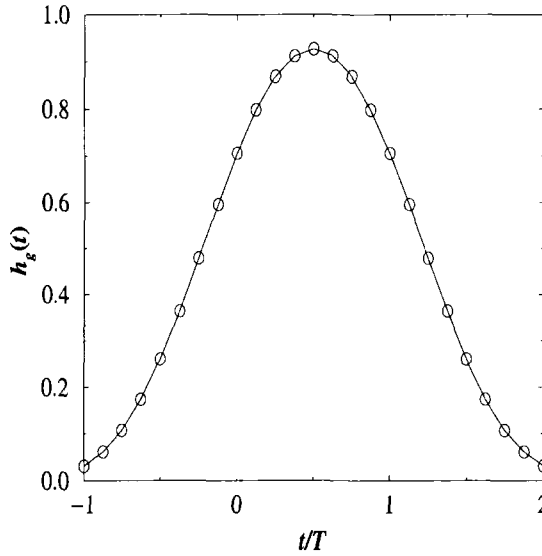
Figure 4.26. LGMSK shaping pulse for $BT = 0.3$.

Fig. 4.26 plots the interpolated LGMSK pulse, $h_g(t)$. Just as MSK is equivalent to OQPSK with a half-sinusoid shaping pulse, LGMSK is equivalent to OQPSK with the shaping pulse $h_g(t)$.

8.3   TAMED FREQUENCY MODULATION (TFM)

Tamed frequency modulation (TFM) is a special type of partial response binary CPM that was introduced by Jager and Dekker [80]. To define TFM signals, recall that the excess phase for MSK obeys the difference equation

$$\phi(nT + T) - \phi(nT) = x_n \frac{\pi}{2} \ . \tag{4.154}$$

For TFM, the excess phase trajectory is smoothed by using the partial response condition

$$\phi(nT + T) - \phi(nT) = \frac{\pi}{2} \left(\frac{x_{n-1}}{4} + \frac{x_n}{2} + \frac{x_{n+1}}{4} \right) \ . \tag{4.155}$$

The maximum excess phase change over any bit interval is equal to $\pi/2$. To complete the definition of the TFM signal, an appropriate premodulation filter must be defined. If the premodulation filter has impulse response $h_f(t)$, then the excess phase can be written as

$$\phi(t) = \sum_{k=0}^{\infty} x_k \beta(t - kT) \tag{4.156}$$

where

$$\beta(t) = 2\pi h \int_{-\infty}^{t} h_f(t)dt \ . \tag{4.157}$$

The excess phase change over the time interval $[nT, (n+1)T]$ is

$$\phi(nT + T) - \phi(nT) \quad = \quad 2\pi h \sum_{k=-\infty}^{\infty} x_k \left(\beta(nT + T - kT) - \beta(nT - kT)\right)$$

$$= \quad 2\pi h \sum_{\ell=-\infty}^{\infty} x_{n-\ell} \left(\beta(\ell T + T) - \beta(\ell T)\right) \ . \tag{4.158}$$

Expanding (4.155) in more detail gives

$$\phi(nT + T) - \phi(nT) \quad = \quad \frac{\pi}{2} \left(\ldots + x_{n-2} \cdot 0 + \frac{x_{n-1}}{4}\right.$$

$$\left. + \frac{x_n}{2} + \frac{x_{n+1}}{4} + x_{n+2} \cdot 0 + \ldots\right) . \tag{4.159}$$

Comparing (4.158) and (4.159) gives the condition

$$\beta(\ell T + T) - \beta(\ell T) = \begin{cases} \frac{1}{16h} & , \quad |\ell| = 1 \\ \frac{1}{8h} & , \quad \ell = 0 \\ 0 & , \quad \text{otherwise} \end{cases} \ . \tag{4.160}$$

From the definition of $\beta(t)$ in (4.157) the above equation leads to

$$\int_{\ell T}^{(\ell+1)T} h_f(t)dt = \begin{cases} \frac{1}{16h} & , \quad |\ell| = 1 \\ \frac{1}{8h} & , \quad \ell = 0 \\ 0 & , \quad \text{otherwise} \end{cases} \ . \tag{4.161}$$

One way of obtaining $h_f(t)$ is to use a pulse $h_N(t)$ that satisfies Nyquist's third criterion [248, 259]

$$\int_{(2\ell-1)T/2}^{(2\ell+1)T/2} h_N(t)dt = \begin{cases} 1 & , \quad \ell = 0 \\ 0 & , \quad \ell \neq 0 \end{cases} \tag{4.162}$$

and generate $h_f(t)$ by using scaling and delay operations through the filter shown in Fig. 4.27. The transfer function of this filter is

$$S(f) \quad = \quad \frac{1}{16h}e^{-j2\pi fT} + \frac{1}{8h} + \frac{1}{16h}e^{j2\pi fT}$$

$$= \quad \frac{1}{4h}\cos^2(\pi fT) \ . \tag{4.163}$$

The overall pulse $h_f(t)$ has the form

$$H_f(f) \quad = \quad H_N(f)S(f)$$

$$= \quad H_N(f)\frac{1}{4h}\cos^2(\pi fT) \ . \tag{4.164}$$
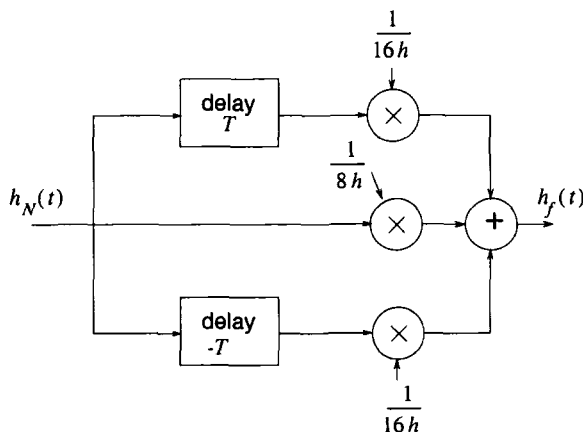
Figure 4.27.   Filter to generate a TFM frequency shaping pulse.

The filter $S(f)$ ensures that the phase constraint in (4.155) is satisfied. However, $H_N(f)$ determines the shape of the phase trajectories and, hence, can influence the TFM power density spectrum. In general, $H_N(f)$ has the form

$$H_N(f) = \frac{\pi f T}{\sin(\pi f T)} N_1(f) \qquad (4.165)$$

where $N_1(f)$ is the Fourier transform of a pulse that satisfies Nyquist's first criterion [248, 259]. One example is the raised cosine pulse $|H_a(f)|^2$ defined in (4.47). Consider, for example, the ideal Nyquist pulse (raised cosine pulse with $\beta = 0$)

$$N_1(f) = \begin{cases} 1 & , \quad 0 \le |f| \le 1/2T \\ 0 & , \quad \text{otherwise} \end{cases} . \qquad (4.166)$$

Using (4.164)–(4.166) gives

$$H_f(f) = \frac{1}{4h} \frac{\pi f T}{\sin(\pi f T)} \cos^2(\pi f T) . \qquad (4.167)$$

The corresponding frequency shaping pulse $\pi h h_f(t)$ is plotted in Fig. 4.28. Note the close similarity to the GMSK pulse in Fig. 4.24.

Generalized tamed frequency modulation (GTFM) is an extension of TFM where the phase difference has the form

$$\phi_i(nT + T) - \phi_i(nT) = \frac{\pi}{2} (ax_{n-1} + bx_n + ax_{n+1}) . \qquad (4.168)$$

The constants a and b satisfy the condition $2a + b = 1$ so that the maximum change in $\phi_i(t)$ during one symbol period is restricted to $\pm \pi/2$. A large class of signals can be constructed by varying the value of b and by varying the pulse response $\beta(t)$. TFM is a special case of GTFM where $b = 0.5$.
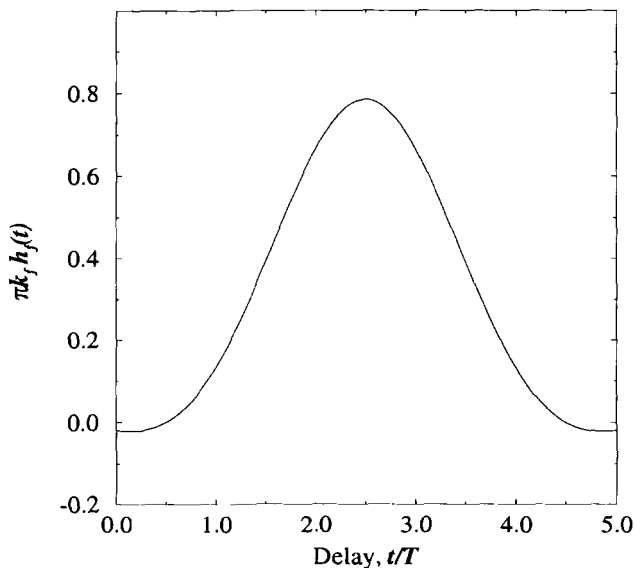
Figure 4.28. TFM frequency shaping pulse.

9. POWER SPECTRAL DENSITIES OF DIGITALLY MODULATED SIGNALS

A digitally modulated band-pass signal can be written in the generic form

$$
\begin{aligned}
s(t) &= \mathrm{Re}\left\{\tilde{s}(t)e^{j(2\pi f_c t+\theta_o)}\right\} \\
&= \frac{1}{2}\left\{\tilde{s}(t)e^{j(2\pi f_c t+\theta_o)} + \tilde{s}^*(t)e^{-j(2\pi f_c t+\theta_o)}\right\} \quad\quad (4.169)
\end{aligned}
$$

where θ_o is a random phase uniformly distributed over $(-\pi,\pi)$. Modulated signals are not wide sense stationary, but belong to a class of random processes that are cyclostationary. The autocorrelation function of $s(t)$ is

$$
\begin{aligned}
\phi_{ss}(\tau) &= \mathrm{E}\left[s(t+\tau)s(t)\right] \\
&= \frac{1}{4}\mathrm{E}\left[\left(\tilde{s}(t+\tau)e^{j(2\pi f_c t+2\pi f_c \tau+\theta_o)} + \tilde{s}^*(t+\tau)e^{-j(2\pi f_c t+2\pi f_c \tau+\theta_o)}\right)\right. \\
&\quad \left.\times \left(\tilde{s}(t)e^{j(2\pi f_c t+\theta_o)} + \tilde{s}^*(t)e^{-j(2\pi f_c t+\theta_o)}\right)\right] \\
&= \frac{1}{4}\mathrm{E}\left[\tilde{s}(t+\tau)\tilde{s}(t)e^{j(4\pi f_c t+2\pi f_c \tau+2\theta_o)} + \tilde{s}^*(t+\tau)\tilde{s}(t)e^{-j2\pi f_c \tau}\right. \\
&\quad \left.+ \tilde{s}(t+\tau)\tilde{s}^*(t)e^{j2\pi f_c \tau} + \tilde{s}^*(t+\tau)\tilde{s}^*(t)e^{-j(4\pi f_c t+2\pi f_c \tau+2\theta_o)}\right] \quad.
\end{aligned}
$$

$$(4.170)$$

To proceed further, we note that

$$\mathrm{E}_{\theta_o}\left[e^{\pm j(4\pi f_c t + 2\pi f_c \tau + 2\theta_o)}\right] = 0 \qquad (4.171)$$

where $\mathrm{E}_{\theta_o}[\,\cdot\,]$ denotes the ensemble average over the random carrier phase. Using this result,

$$\phi_{ss}(\tau) = \frac{1}{2}\phi_{\tilde{s}\tilde{s}}(\tau)e^{j2\pi f_c \tau} + \frac{1}{2}\phi_{\tilde{s}\tilde{s}}^*(\tau)e^{-j2\pi f_c \tau} \quad . \qquad (4.172)$$

Finally, the power density spectrum is the Fourier transform of $\phi_{ss}(\tau)$, i.e.,

$$S_{ss}(f) = \frac{1}{2}\left[S_{\tilde{s}\tilde{s}}(f - f_c) + S_{\tilde{s}\tilde{s}}^*(-f - f_c)\right] \qquad (4.173)$$

where $S_{\tilde{s}\tilde{s}}(f)$ is the power density spectrum of the complex envelope $\tilde{s}(t)$. Observe that $S_{\tilde{s}\tilde{s}}(f)$ is real, even though $\tilde{s}(t)$ and $\phi_{\tilde{s}\tilde{s}}(\tau)$ are complex; this property follows from the fact that $\phi_{\tilde{s}\tilde{s}}(\tau) = \phi_{\tilde{s}\tilde{s}}^*(-\tau)$ as shown in Appendix A. Therefore,

$$S_{ss}(f) = \frac{1}{2}\left[S_{\tilde{s}\tilde{s}}(f - f_c) + S_{\tilde{s}\tilde{s}}(-f - f_c)\right] \quad . \qquad (4.174)$$

From the above expression, it is apparent that the psd of the band-pass waveform $s(t)$ is completely determined by the psd of its complex envelope $\tilde{s}(t)$.

9.1   PSD OF A COMPLEX ENVELOPE

We have seen that the complex envelope of any digitally modulated signal can be expressed in the standard form

$$\tilde{s}(t) = A\sum_n b(t - nT, \mathbf{x}_n) \quad . \qquad (4.175)$$

The autocorrelation of $\tilde{s}(t)$ is

$$\phi_{\tilde{s}\tilde{s}}(t + \tau, t) = \frac{1}{2}\mathrm{E}\left[\tilde{s}(t + \tau)\tilde{s}^*(t)\right] \qquad (4.176)$$

$$= \frac{A^2}{2}\sum_i \sum_k \mathrm{E}\left[b(t + \tau - iT, \mathbf{x}_i)b^*(t - kT, \mathbf{x}_k)\right] \quad .$$

Observe that $\tilde{s}(t)$ is a cyclostationary random process, meaning that the autocorrelation function $\phi_{\tilde{s}\tilde{s}}(t + \tau, t)$ is periodic in t with period T. To see this property, first note that

$$\phi_{\tilde{s}\tilde{s}}(t + T + \tau, t + T)$$

$$= \frac{A^2}{2}\sum_i \sum_k \mathrm{E}\left[b(t + T + \tau - iT, \mathbf{x}_i)b^*(t + T - kT, \mathbf{x}_k)\right]$$

$$= \frac{A^2}{2}\sum_{i'} \sum_{k'} \mathrm{E}\left[b(t + \tau - i'T, \mathbf{x}_{i'+1})b^*(t - k'T, \mathbf{x}_{k'+1})\right] \quad .$$

$$(4.177)$$

Under the assumption that the information sequence is a stationary random process we can write

$$\phi_{\tilde{s}\tilde{s}}(t + T + \tau, t + T) = \frac{A^2}{2} \sum_{i'} \sum_{k'} \mathrm{E}\left[b(t + \tau - i'T, \mathbf{x}_{i'})b^*(t - k'T, \mathbf{x}_{k'})\right]$$

$$= \phi_{\tilde{s}\tilde{s}}(t + \tau, t) \ . \tag{4.178}$$

Therefore $\tilde{s}(t)$ is cyclostationary.

Since $\tilde{s}(t)$ is cyclostationary, the autocorrelation $\phi_{\tilde{s}\tilde{s}}(\tau)$ can be obtained by taking the time average of $\phi_{\tilde{s}\tilde{s}}(t + \tau, t)$, given by

$$\begin{aligned}
\phi_{\tilde{s}\tilde{s}}(\tau) &= <\phi_{\tilde{s}\tilde{s}}(t + \tau, t)> \\
&= \frac{A^2}{2} \sum_{i} \sum_{k} \frac{1}{T} \int_0^T \mathrm{E}\left[b(t + \tau - iT, \mathbf{x}_i)b^*(t - kT, \mathbf{x}_k)\right] dt \\
&= \frac{A^2}{2T} \sum_{i} \sum_{k} \int_{-kT}^{-kT+T} \mathrm{E}\left[b(z + \tau - (i - k)T, \mathbf{x}_i)b^*(z, \mathbf{x}_k)\right] dz \\
&= \frac{A^2}{2T} \sum_{m} \sum_{k} \int_{-kT}^{-kT+T} \mathrm{E}\left[b(z + \tau - mT, \mathbf{x}_{m+k})b^*(z, \mathbf{x}_k)\right] dz \\
&= \frac{A^2}{2T} \sum_{m} \sum_{k} \int_{-kT}^{-kT+T} \mathrm{E}\left[b(z + \tau - mT, \mathbf{x}_m)b^*(z, \mathbf{x}_0)\right] dz \\
&= \frac{A^2}{2T} \sum_{m} \int_{-\infty}^{\infty} \mathrm{E}\left[b(z + \tau - mT, \mathbf{x}_m)b^*(z, \mathbf{x}_0)\right] dz \ . \tag{4.179}
\end{aligned}$$

where $\langle \cdot \rangle$ denotes time averaging and the second last equality used the stationary property of the data sequence $\{x_k\}$. The psd of $\tilde{s}(t)$ is obtained by taking the Fourier transform of $\phi_{\tilde{s}\tilde{s}}(\tau)$[3],

$$\begin{aligned}
S_{\tilde{s}\tilde{s}}(f) &= \mathrm{E}\left[\frac{A^2}{2T} \sum_{m} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} b(z + \tau - mT, \mathbf{x}_m)b^*(z, \mathbf{x}_0)dze^{-j2\pi f\tau} d\tau\right] \\
&= \mathrm{E}\left[\frac{A^2}{2T} \sum_{m} \int_{-\infty}^{\infty} b(z + \tau - mT, \mathbf{x}_m)e^{-j2\pi f(z+\tau-mT)} d\tau \right. \\
&\qquad \times \left. \int_{-\infty}^{\infty} b^*(z, \mathbf{x}_0)e^{j2\pi fz}dze^{-j2\pi fmT}\right] \\
&= \mathrm{E}\left[\frac{A^2}{2T} \sum_{m} \int_{-\infty}^{\infty} b(\tau', \mathbf{x}_m)e^{-j2\pi f\tau'} d\tau' \right. \\
&\qquad \times \left. \int_{-\infty}^{\infty} b^*(z, \mathbf{x}_0)e^{j2\pi fz}dze^{-j2\pi fmT}\right] \text{(cont'd)}
\end{aligned}$$

---

[3]Note that expectation and integration are linear operations and their order can be exchanged.

$$= \frac{A^2}{2T} \sum_m E\left[B(f,\mathbf{x}_m)B^*(f,\mathbf{x}_0)\right] e^{-j2\pi fmT} \qquad (4.180)$$

where $B(f,\mathbf{x}_m)$ is the Fourier transform of $b(t,\mathbf{x}_m)$. To express the power density spectrum in a more convenient form, let

$$S_{b,m}(f) = \frac{1}{2}E\left[B(f,\mathbf{x}_m)B^*(f,\mathbf{x}_0)\right] \ . \qquad (4.181)$$

Then

$$S_{\tilde{s}\tilde{s}}(f) = \frac{A^2}{T} \sum_m S_{b,m}(f) e^{-j2\pi fmT} \ . \qquad (4.182)$$

Note that the psd in (4.182) depends on the correlation properties of the information sequence $\mathbf{x}_m$ and the form of the equivalent pulse shaping function $b(t,\mathbf{x}_m)$. Now suppose that the data characteristics are such that $\mathbf{x}_m$ and $\mathbf{x}_0$ are independent for $|m| \geq K$. Then

$$S_{b,m}(f) = S_{b,K}(f), \quad |m| \geq K \qquad (4.183)$$

where

$$
\begin{aligned}
S_{b,K}(f) &= \frac{1}{2}E\left[B(f,\mathbf{x}_m)\right] E\left[B^*(f,\mathbf{x}_0)\right] && |m| > K \\
&= \frac{1}{2}E\left[B(f,\mathbf{x}_0)\right] E\left[B^*(f,\mathbf{x}_0)\right] && |m| > K \\
&= \frac{1}{2}\left|E\left[B(f,\mathbf{x}_0)\right]\right|^2 , \quad |m| \geq K \ . && (4.184)
\end{aligned}
$$

It follows that

$$S_{\tilde{s}\tilde{s}}(f) = S_{\tilde{s}\tilde{s}}^c(f) + S_{\tilde{s}\tilde{s}}^d(f) \qquad (4.185)$$

where

$$
\begin{aligned}
S_{\tilde{s}\tilde{s}}^c(f) &= \frac{A^2}{T} \sum_{|m|<K} \left(S_{b,m}(f) - S_{b,K}(f)\right) e^{-j2\pi fmT} \\
S_{\tilde{s}\tilde{s}}^d(f) &= \frac{A^2}{T} S_{b,K}(f) \sum_m e^{-j2\pi fmT} \ . && (4.186)
\end{aligned}
$$

The terms $S_{\tilde{s}\tilde{s}}^c(f)$ and $S_{\tilde{s}\tilde{s}}^d(f)$ represent the continuous and discrete portions of the psd. The fact that $S_{\tilde{s}\tilde{s}}^d(f)$ represents the discrete portion, can be seen more clearly by using the identity

$$T \sum_m e^{-j2\pi fmT} = \sum_n \delta\left(f - \frac{n}{T}\right) \qquad (4.187)$$

to write

$$S^d_{\tilde{s}\tilde{s}}(f) = \left(\frac{A}{T}\right)^2 S_{b,K}(f) \sum_n \delta\left(f - \frac{n}{T}\right) . \qquad (4.188)$$

Finally, by using the property $S_{b,-m}(f) = S^*_{b,m}(f)$, the continuous portion of the psd can be written as

$$
\begin{aligned}
S^c_{\tilde{s}\tilde{s}}(f) &= \frac{A^2}{T}\left(S_{b,0}(f) - S_{b,K}(f)\right) \\
&+ \frac{A^2}{T}\sum_{m=1}^{K}\left\{\left(S_{b,m}(f) - S_{b,K}(f)\right)e^{-j2\pi fmT}\right.\\
&\qquad\qquad \left. + \left(S^*_{b,m}(f) - S_{b,K}(f)\right)e^{j2\pi fmT}\right\} \\
&= \frac{A^2}{T}\left(S_{b,0}(f) - S_{b,K}(f)\right) \\
&+ \frac{A^2}{T}2\mathrm{Re}\left\{\sum_{m=1}^{K}\left(S_{b,m}(f) - S_{b,K}(f)\right)e^{-j2\pi fmT}\right\} \qquad (4.189)
\end{aligned}
$$

Note that the ensemble average and Fourier transform are interchangeable linear operators. Therefore, if the complex envelope $\tilde{s}\tilde{s}(t)$ has zero mean, i.e., $E[b(t, \mathbf{x}_0)] = 0$, then $E[B(f, \mathbf{x}_0)] = 0$. Under this condition

$$S_{b,K}(f) = \frac{1}{2}|E[B(f, \mathbf{x}_0)]|^2 = 0 . \qquad (4.190)$$

Hence, if $b(t, \mathbf{x}_0)$ has zero mean, then $S_{\tilde{s}\tilde{s}}(f)$ contains no discrete components and $S_{\tilde{s}\tilde{s}}(f) = S^c_{\tilde{s}\tilde{s}}(f)$. Conversely, if $b(t, \mathbf{x}_0)$ has nonzero mean, then $S_{\tilde{s}\tilde{s}}(f)$ will contain discrete components.

Alternative Method. An alternative method of computing the psd is as follows. From the first line in (4.180)

$$
\begin{aligned}
S_{\tilde{s}\tilde{s}}(f) &= E\left[\frac{A^2}{2T}\sum_m\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}b(z+\tau-mT,\mathbf{x}_m)b^*(z,\mathbf{x}_0)dze^{-j2\pi f\tau}d\tau\right] \\
&= \frac{A^2}{2T}\sum_m\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}E\left[b(\tau',\mathbf{x}_m)b^*(z,\mathbf{x}_0)\right] \\
&\qquad \times e^{-j2\pi f(\tau'-z)}dzd\tau'e^{-j2\pi fmT} \qquad (4.191)
\end{aligned}
$$

Therefore, $S_{b,m}(f)$ is given by the double Fourier transform

$$S_{b,m}(f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\phi_{b,m}(\tau',z)e^{-j2\pi f(\tau'-z)}dzd\tau' . \qquad (4.192)$$

where

$$\phi_{b,m}(\tau',z) = \frac{1}{2}E\left[b(\tau',\mathbf{x}_m)b^*(z,\mathbf{x}_0)\right] . \qquad (4.193)$$

Uncorrelated Source Symbols.   Suppose that $b(t, \mathbf{x}_m)$ depends on one symbol only

$$b(t, \mathbf{x}_m) = b(t, x_m) \tag{4.194}$$

and that the data symbols x_m are independent. Then

$$S_{b,0}(f) = \frac{1}{2} \mathrm{E}\left[|B(f, x_0)|^2\right] \tag{4.195}$$

$$S_{b,m}(f) = \frac{1}{2} |\mathrm{E}\left[B(f, x_0)\right]|^2 \qquad |m| \geq 1 . \tag{4.196}$$

Hence, $S_{\tilde{s}\tilde{s}}(f)$ is given by (4.185), where

$$S_{\tilde{s}\tilde{s}}^d(f) = \frac{A^2}{T^2} S_{b,1}(f) \sum_n \delta\left(f - \frac{n}{T}\right) \tag{4.197}$$

$$S_{\tilde{s}\tilde{s}}^c(f) = \frac{A^2}{T} \left(S_{b,0}(f) - S_{b,1}(f)\right) . \tag{4.198}$$

Once again, if $b(t, x_m)$ has zero mean, then $S_{b,1}(f) = 0$ (no discrete spectral components) and

$$S_{\tilde{s}\tilde{s}}(f) = \frac{A^2}{T} S_{b,0}(f) . \tag{4.199}$$

Linear Full Response Modulation.   Consider linear full response modulation schemes where $b(t, \mathbf{x}_n) = x_n h_a(t)$ and $B(f, \mathbf{x}_n) = x_n H_a(f)$. From (4.181)

$$S_{b,m}(f) = \phi_{xx}(m) |H_a(f)|^2 \tag{4.200}$$

where

$$\phi_{xx}(m) = \frac{1}{2} \mathrm{E}[x_{k+m} x_k^*] . \tag{4.201}$$

Hence, from (4.182) the psd of the complex envelope is

$$S_{\tilde{s}\tilde{s}}(f) = \frac{A^2}{T} |H_a(f)|^2 S_{xx}(f) \tag{4.202}$$

where

$$S_{xx}(f) = \sum_m \phi_{xx}(m) e^{-j2\pi f m T} . \tag{4.203}$$

Note that the psd is the product of two components; one depends on the amplitude shaping function and the other depends on the correlation of the data sequence. With uncorrelated data symbols

$$S_{b,0}(f) = \frac{1}{2} \mathrm{E}[|x|^2] |H_a(f)|^2 \tag{4.204}$$

$$S_{b,m}(f) = \frac{1}{2} |\mu_x|^2 |H_a(f)|^2 , \quad |m| \geq 1 . \tag{4.205}$$

where $\mu_x = \mathrm{E}[x_m]$. The psd $S_{\tilde{s}\tilde{s}}(f)$ is then given by (4.185), (4.197), and (4.198). If $\mu_x = 0$, then $S_{b,1}(f) = 0$ and

$$S_{\tilde{s}\tilde{s}}(f) = \frac{A^2}{T}\sigma_x^2|H_a(f)|^2 \tag{4.206}$$

where $\sigma_x^2 = \frac{1}{2}\mathrm{E}[|x_k|^2]$.

Linear Partial Response Modulation. Consider linear partial response modulation schemes where $h_a(t)$ has duration LT. Following the development in Section 8. the equivalent shaping function has the form

$$
\begin{aligned}
b(t, \mathbf{x}_m) &= h_a(t, \mathbf{x}_m) \\
&= \sum_{k=0}^{L-1} x_{m-k} h_{a,k}(t) \tag{4.207}
\end{aligned}
$$

where

$$h_{a,k}(t) = h_a(t + kT)U_T(t) \tag{4.208}$$

Taking the Fourier transform gives

$$B(f, \mathbf{x}_m) = \sum_{k=0}^{L-1} x_{m-k} H_{a,k}(f) . \tag{4.209}$$

From (4.181),

$$
\begin{aligned}
S_{b,m}(f) &= \frac{1}{2}\mathrm{E}\left[\sum_{k=0}^{L-1} x_{m-k} H_{a,k}(f) \sum_{\ell=0}^{L-1} x_{-\ell}^* H_{a,\ell}^*(f)\right] \\
&= \sum_{k=0}^{L-1}\sum_{\ell=0}^{L-1} \frac{1}{2}\phi_{xx}(m - k + \ell)H_{a,k}(f)H_{a,\ell}^*(f) . \tag{4.210}
\end{aligned}
$$

For the special case of uncorrelated zero-mean data symbols, $\phi_{xx}(m-k+\ell) = \sigma_x^2\delta(m - k + \ell)$. Hence,

$$S_{b,m}(f) = \sigma_x^2 \sum_{\ell=0}^{L-1} H_{a,m+\ell}(f)H_{a,\ell}^*(f) \tag{4.211}$$

where

$$\sigma_x^2 = \frac{1}{2}\mathrm{E}[|x_0|^2] .$$

Example 4.4 Duobinary Signaling ____ For duobinary signaling, $L = 2$ and $h_{a,0}(t) = h_{a,1}(t) = \mathrm{Sa}(\pi t/T)$ and $H_{a,0}(f) = H_{a,1}(f) = T\mathrm{rect}(fT)$, where

$$
\mathrm{rect}(fT) = \begin{cases} T, & |f| \leq 1/(2T) \\ 0, & \text{elsewhere} \end{cases}
$$

With uncorrelated zero-mean data symbols

$$S_{b,m}(f) = \frac{1}{2}\mathrm{E}\left[\left(x_0^* H_{a,0}^*(f) + x_{-1}^* H_{a,1}^*(f)\right)\left(x_m H_{a,0}(f) + x_{m-1} H_{a,1}(f)\right)\right]$$

$$= \begin{cases} 2\sigma_x^2 T^2 \mathrm{rect}(fT) & , \quad m = 0 \\ \sigma_x^2 T^2 \mathrm{rect}(fT) & , \quad |m| = 1 \\ 0 & , \quad \text{otherwise} \end{cases}$$

and from (4.182)

$$S_{\tilde{s}\tilde{s}}(f) = 4A^2 T\sigma_x^2 \cos^2(\pi fT)\mathrm{rect}(fT) \ . \tag{4.212}$$

Example 4.5 Modified Duobinary **Signaling**_____

For modified duobinary signaling, $L = 3$ and $h_{a,0}(t) = h_{a,2}(t) = \mathrm{Sa}(\pi t/T)$ and $h_{a,1}(t) = 0$. With uncorrelated zero-mean data symbols,

$$S_{b,m}(f) = \begin{cases} 2\sigma_x^2 T^2 \mathrm{rect}(fT) & , \quad m = 0 \\ -\sigma_x^2 T^2 \mathrm{rect}(fT) & , \quad |m| = 2 \\ 0 & , \quad \text{otherwise} \end{cases}$$

and from (4.182)

$$S_{\tilde{s}\tilde{s}}(f) = 4A^2 T\sigma_x^2 \sin^2(2\pi fT)\mathrm{rect}(fT) \ .$$

9.2   PSD OF QAM

The psd of QAM with uncorrelated zero-mean data symbols is given by (4.206). If $h_a(t) = u_T(t)$, then

$$S_{\tilde{s}\tilde{s}}(f) = A^2 T\sigma_x^2 \left(\frac{\sin \pi fT}{\pi fT}\right)^2 \ . \tag{4.213}$$

To fairly compare bandwidth efficiencies with different M, the frequency variable should be normalized by the bit interval T_b. For M-ary signaling $T = T_b \log_2 M$. Hence,

$$S_{\tilde{s}\tilde{s}}(f) = A^2 T\sigma_x^2 \left(\frac{\sin \pi fT_b \log_2 M}{\pi fT_b \log_2 M}\right)^2 \ . \tag{4.214}$$

With root raised cosine pulse shaping, $|H_a(f)|^2 = P(f)$ has the form defined in (4.47) with $h_a(t)$ in (4.49). The root raised cosine pulse is non-causal. When the pulse is implemented as a digital FIR filter, it must be truncated to a finite length τ. This truncation produces the new pulse $\tilde{h}_a(t) = h_a(t)\,\mathrm{rect}(t/\tau)$. The Fourier transform of the truncated pulse $\tilde{h}_a(t)$ is $\tilde{H}_a(f) =$
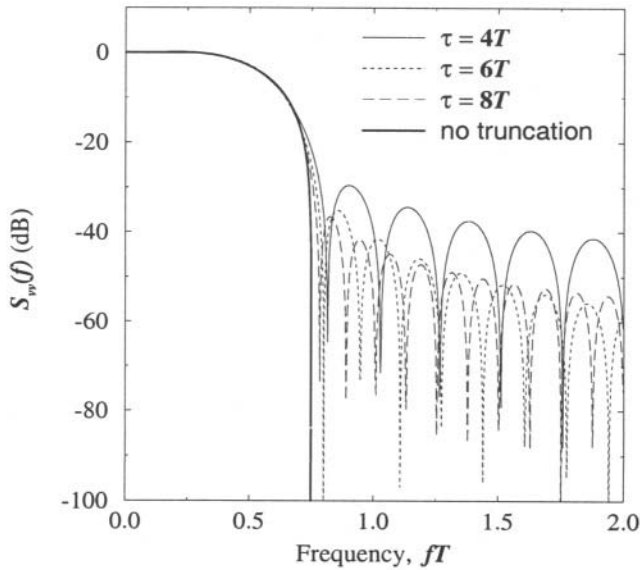
Figure 4.29. Psd of QAM with a truncated square root raised cosine pulse with various truncation lengths; $\beta = 0.5$. Pulse truncation leads to side lobe regeneration.

$H_a(f) * \tau \text{Sa}(f\tau)$, where $*$ denotes the operation of convolution taken over the frequency variable f. The psd of QAM with the pulse $\tilde{h}_a(t)$ can again be obtained from (4.206) by simply replacing $H_a(f)$ with $\tilde{H}_a(f)$. As shown in Fig. 4.29, pulse truncation can lead to significant side lobe regeneration. Again, to fairly compare bandwidth efficiencies with different M, the frequency variable should be normalized by the bit interval T_b. The has the effect of dividing the elements on the horizontal axis in Fig. 4.29 by a factor of $\log_2 M$.

9.3   PSD OF PSK

For PSK signals with the uncorrelated data symbols and the equivalent shaping function in (4.60), the psd is given by (4.206) with $\sigma_x^2 = 1/2$. Hence, PSK signals have the same psd as QAM signals. The psd with rectangular and root raised cosine pulse shaping is given by (4.213) and (4.214), respectively. Again, to fairly compare bandwidth efficiencies with different M, the frequency variable must be normalized by the bit interval T_b.

9.4   PSD OF OQPSK

For OQPSK, the equivalent shaping function is

$$b(t, \mathbf{x}_n) = b(t, x_n) = x_{I,n} h_a(t) + j x_{Q,n} h_a(t - T/2) \qquad (4.215)$$

where $x_{I,n}, x_{Q,n} \in \{-1/\sqrt{2}, +1/\sqrt{2}\}$. It follows that

$$B(f, \mathbf{x}_n) = \left(x_{I,n} + j x_{Q,n} e^{-j2\pi fT/2} \right) H_a(f) \tag{4.216}$$

With uncorrelated data symbols,

$$
\begin{aligned}
S_{b,0}(f) &= \frac{1}{2} \mathrm{E}\left[|B(f, x_0)|^2 \right] \\
&= \frac{1}{2} |H_a(f)|^2
\end{aligned}
\tag{4.217}
$$

Therefore,

$$S_{\tilde{s}\tilde{s}}(f) = \frac{A^2}{2T} |H_a(f)|^2 \tag{4.218}$$

Hence, OQPSK has the same power density spectrum as QPSK.

9.5   PSD OF $\pi/4$-DQPSK

To find the power density spectrum of $\pi/4$-DQPSK we first compute the autocorrelation

$$\phi_{b,m}(\tau', z) = \frac{1}{2} \mathrm{E}\left[b(\tau', \mathbf{x}_m) b^*(z, \mathbf{x}_0) \right] \quad . \tag{4.219}$$

For $m > 0$,

$$
\begin{aligned}
\phi_{b,m}(\tau', z) &= \frac{1}{2} \mathrm{E}\left[h_a(\tau') \exp\left\{ j\frac{\pi}{4} \sum_{n=1}^{m} x_n \right\} h_a(z) \right] \\
&= \frac{1}{2} \mathrm{E}\left[\exp\left\{ j\frac{\pi}{4} \sum_{n=1}^{m} x_m \right\} \right] h_a(\tau') h_a(z) \\
&= 0 \quad .
\end{aligned}
\tag{4.220}
$$

For $m = 0$,

$$
\begin{aligned}
\phi_{b,m}(\tau', z) &= \frac{1}{2} \mathrm{E}\left[h_a(\tau') \exp\left\{ j\frac{\pi}{4} x_0 \right\} h_a(z) \exp\left\{ -j\frac{\pi}{4} x_0 \right\} \right] \\
&= \frac{1}{2} h_a(\tau') h_a(z) \quad .
\end{aligned}
\tag{4.221}
$$

Taking the double Fourier transform gives

$$
\begin{aligned}
S_{b,0}(f) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi_{b,m}(\tau', z) e^{-j2\pi f(\tau' - z)} dz\, d\tau' \\
&= \frac{1}{2} |H_a(f)|^2 \quad .
\end{aligned}
\tag{4.222}
$$

Finally, the psd is

$$S_{\tilde{s}\tilde{s}}(f) = \frac{A^2}{2T} |H_a(f)|^2 \quad . \tag{4.223}$$

Just like OQPSK, $\pi/4$-DQPSK has the same power density spectrum as QPSK.

9.6   PSD OF OFDM

The psd of an OFDM signal can be obtained by treating OFDM as independent modulation on N sub-carriers that are separated in frequency by $1/T$. Ignoring the guard interval the data symbol period on each sub-carrier is $T = NT_s$, where T_s is the serial source data symbol period. Suppose the data symbols that are modulating each sub-carrier have zero mean and variance $\sigma_x^2 = \frac{1}{2}E[|x_{k,n}|^2]$. To keep the $1/T$-spaced sub-carriers orthogonal in time, the amplitude shaping function $h_a(t)$ must be the rectangular pulse $h_a(t) = u_T(t)$. However, if a loss of sub-channel orthogonality can be tolerated, then other types of amplitude shaping pulses can be used, such as the root raised cosine pulse. Assuming that the amplitude shaping pulse is $h_a(t)$, the psd of the OFDM complex envelope is

$$S_{\tilde{s}\tilde{s}}(f) = A^2 T \sigma_x^2 \sum_{k=0}^{N-1} \left| H_a\left(f - \frac{1}{T}\left(k - \frac{N-1}{2} \right) \right) \right|^2 \qquad (4.224)$$

Consider the rectangular amplitude shaping pulse $h_a(t) = u_T(t)$ with Fourier transform $H_a(f) = \mathrm{Sa}(\pi f T)$. The corresponding OFDM psd is shown in Figs. 4.30 and 4.31 for block sizes of $N = 4$ and $N = 32$, respectively. As the block size N is increased, the psd becomes flat in the $N/T = 1/T_s$ bandwidth containing containing the sub-carriers, while the side lobes decrease. In fact, in the limit as TV becomes very large, the sidelobes diminish to zero and the complex envelope of the OFDM signal occupies the band

$$|f| \stackrel{<}{\approx} \frac{N}{(2T)} = \frac{1}{2T_s} \quad .$$

This is the minimum possible bandwidth required for transmitting data symbols at a rate of $1/T_s$ symbols per second without intersymbol interference (ISI). It can be achieved by transmitting the serial source data symbols using single carrier modulation and the amplitude shaping pulse $h_a(t) = \mathrm{Sa}(\pi t/T_s)$. However, as mentioned earlier, the transmission of data symbols with such a high baud rate will suffer from channel induced ISI and require equalization at the receiver.

For smaller values of N, improvement in the psd can be obtained by using a root raised cosine pulse shaping on each of the sub-carriers. Fig. 4.32 shows the effect of using such pulse shaping, which can be compared directly with Fig. 4.30. However, we repeat that the use of root raised cosine pulse shaping will destroy the sub-carrier orthogonality. As shown in Chapter 5 the cost is a floor in the bit error rate performance.

Finally, it is interesting to examine the OFDM power spectrum, when the discrete-time IFFT modulator is used. After D/A conversion, the complex time
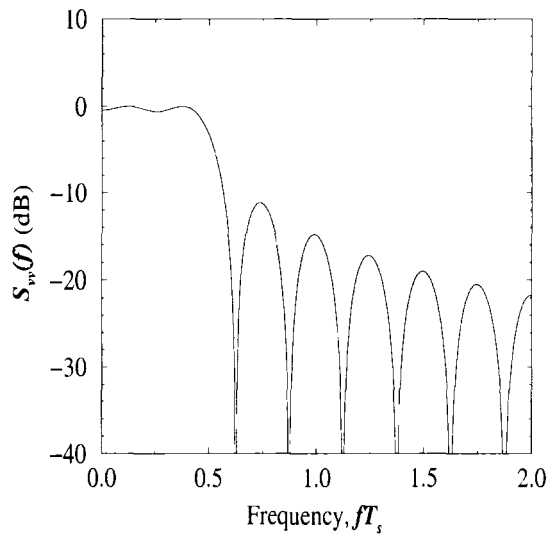
Figure 4.30.   Psd of OFDM with $N = 4$.
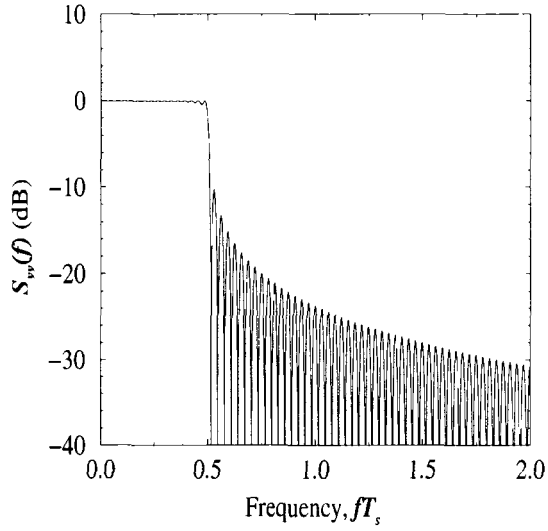
Figure 4.31.   Psd of OFDM with $N = 32$.
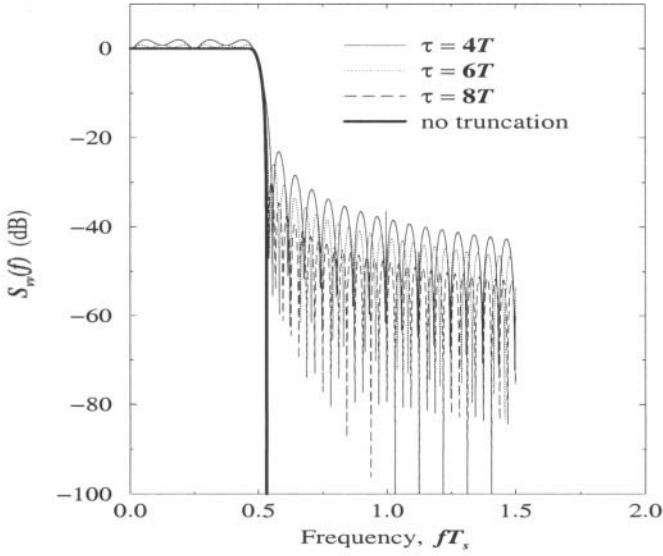
Figure 4.32. Psd of OFDM with $N = 4$ and truncated square root raised cosine pulse shaping; $\beta = 0.25$.

domain waveform that is transmitted on the *kth* sub-carrier is

$$\tilde{s}_k(t) = Ax_k h_{ak}(t) \tag{4.225}$$

where

$$h_{ak}(t) = \sum_{n=0}^{N-1} e^{j2\pi f_k nT_s} \mathrm{sinc}(t/T_s - n) \tag{4.226}$$

and $f_k = k/NT_s$. It follows that the psd for the *kth* sub-carrier is

$$\tilde{S}_{kk}(f) = \frac{A^2}{T}\sigma_x^2 |H_{ak}(f)|^2 \quad . \tag{4.227}$$

Taking the Fourier transform of $h_{ak}(t)$ gives

$$H_{ak}(f) = T_s \mathrm{rect}\left(\frac{f}{T_s}\right)\frac{\sin(\pi(f - f_k)NT_s)}{\sin(\pi(f - f_k)T_s)}e^{-j\pi(f-f_k)(N-1)T_s}$$

$$= T_s \mathrm{rect}\left(\frac{f}{T_s}\right)\frac{\sin(\pi(NfT_s - k))}{\sin(\pi(fT_s - k/N))}e^{-j\pi\left(f - \frac{k}{NT_s}\right)(N-1)T_s} \tag{4.228}$$

The power spectrum for the *kth* subcarrier is

$$S_{kk}(f) = A^2 T\sigma_x^2 \mathrm{rect}\left(\frac{f}{T_s}\right)\frac{1}{N^2}\left(\frac{\sin(\pi(NfT_s - k))}{\sin(\pi(fT_s - k/N))}\right)^2 \tag{4.229}$$
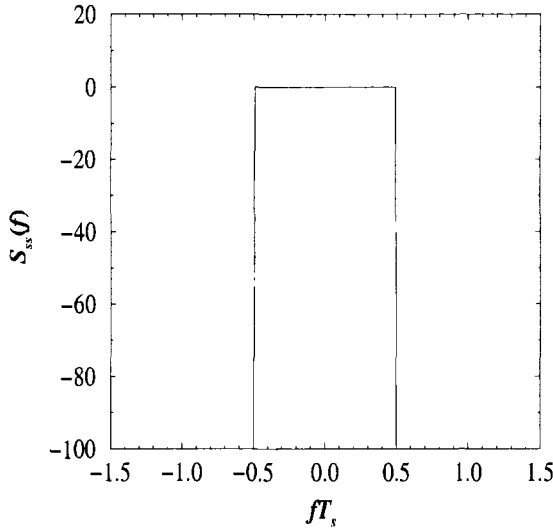
Figure 4.33.   Psd of OFDM with implemented with IFFT algorithm.

Since the data symbols on the subcarriers are independent, we can just add their psds together to yield the overall psd

$$S_{\tilde{s}\tilde{s}}(f) = A^2 T \sigma_x^2 \text{rect}\left(\frac{f}{T_s}\right) \frac{1}{N^2} \sum_{k=0}^{N-1} \left(\frac{\sin(\pi(NfT_s - k))}{\sin(\pi(fT_s - k/N))}\right)^2 \qquad (4.230)$$

which is plotted in Fig. 4.33. Note that the psd has the ideal rectangular form rect(fT).

Finally, the above results were obtained by using an ideal reconstruction filter in the D/A converter. This leads to a non-causal amplitude shaping pulse $h_a(t)$ shown in Fig. 4.15. Any practical implementation will truncate this pulse in the time domain. This in turn will lead to spectral sidelobes outside the band $|f| \leq 1/2T_s$.

9.7   PSD OF FULL RESPONSE CPM

Recall that the equivalent shaping function for CPM is given by (4.105). To compute the psd, we define the auxiliary function

$$r(t, x_k) \triangleq e^{jx_k \beta(t)} u_T(t) \qquad (4.231)$$

and calculate its mean and autocorrelation function. If M-ary signaling is used with the values of x_k defined by

$$x_k \in \{2m - 1 - M : m = 1, 2, \cdots, M\} . \qquad (4.232)$$

then

$$
\begin{aligned}
m_r(t) &= \mathrm{E}[r(t, x_k)] \\
&= \frac{1}{M} \sum_{i=1}^{M} e^{j(2i-1-M)\beta(t)} u_T(t) \\
&= \mathrm{sinf}(\beta(t)) u_T(t) \ .
\end{aligned}
\tag{4.233}
$$

where $\mathrm{sinf}(x)$ is defined by

$$
\mathrm{sinf}(x) \triangleq \frac{\sin Mx}{M \sin x} \ .
\tag{4.234}
$$

Also

$$
\phi_{r,m}(t, t') = \frac{1}{2} \mathrm{E} \left[r(t, x_m) r^*(t', x_0) \right] \ .
\tag{4.235}
$$

Evaluating the above expression for $m = 0$ gives the following result which will be used later

$$
\begin{aligned}
\phi_{r,0}(t, t') &= \frac{1}{2} \mathrm{E} \left[r(t, x_0) r^*(t', x_0) \right] \\
&= \frac{1}{2} \mathrm{E} \left[e^{j x_0 \beta(t)} e^{-j x_0 \beta(t')} \right] u_T(t) u_T(t') \\
&= \frac{1}{2} \mathrm{E} \left[e^{j(x_0 \beta(t) - x_0 \beta(t'))} \right] u_T(t) u_T(t') \\
&= \frac{1}{2} \mathrm{sinf} \left(\beta(t) - \beta(t') \right) u_T(t) u_T(t') \ .
\end{aligned}
\tag{4.236}
$$

To evaluate the psd, it is necessary to compute the autocorrelation of $b(t, \mathbf{x}_m)$. This can be done as follows

$$
\begin{aligned}
\phi_{b,m}(t, t') &= \frac{1}{2} \mathrm{E} \left[b(t, \mathbf{x}_m) b^*(t', \mathbf{x}_0) \right] \\
&= \frac{1}{2} \mathrm{E} \left[\exp \left\{ j\beta(T) \sum_{k=0}^{m-1} x_k \right\} r(t, x_m) r^*(t', x_0) \right] \\
&= \frac{1}{2} \mathrm{E} \left[\left(\prod_{k=0}^{m-1} r(T, x_k) \right) r(t, x_m) r^*(t', x_0) \right] \\
&= \frac{1}{2} \mathrm{E} \left[\left(\prod_{k=1}^{m-1} r(T, x_k) \right) r(t, x_m) r(T, x_0) r^*(t', x_0) \right] \ .
\end{aligned}
\tag{4.237}
$$

Now suppose that the data sequence is uncorrelated. Then for $m > 0$

$$\phi_{b,m}(t,t') \;=\; \frac{1}{2}\,[m_r(T)]^{m-1}\,m_r(t)\phi_{r,0}(T,t')$$

$$=\; \frac{1}{2}\,[\mathrm{sinf}\beta(T)]^{m-1}\,[\mathrm{sinf}\beta(t)]\,[\mathrm{sinf}\,(\beta(T)-\beta(t'))]\,u_T(t)u_T(t')$$

$$\tag{4.238}$$

where we have used (4.236). Likewise, for $m=0$

$$\phi_{b,0}(t,t') \;=\; \frac{1}{2}\mathrm{E}\,[b(t,\mathbf{x}_0)b^*(t',\mathbf{x}_0)]$$

$$=\; \frac{1}{2}\mathrm{E}\left[e^{j(x_0\beta(t)-jx_0\beta(t'))}\right]u_T(t)u_T(t')$$

$$=\; \frac{1}{2}\mathrm{sinf}\left(\beta(t)-\beta(t')\right)u_T(t)u_T(t')$$

$$=\; \phi_{r,0}(t,t')\;.\tag{4.239}$$

Finally, the psd is obtained by using (4.238) and (4.239) along with (4.182) and (4.192).

Alternative Method.   There is an alternate method for obtaining the psd that provides more insight. Similar to the way that (4.189) was derived, we use (4.182) along with the property $S_{b,-m}(f)=S^*_{b,m}(f)$ to obtain

$$S_{\tilde{s}\tilde{s}}(f)=\frac{A^2}{T}\left(S_{b,0}(f)+2\mathrm{Re}\left\{\sum_{m=1}^{\infty}S_{b,m}(f)e^{-j2\pi fmT}\right\}\right)\;.\tag{4.240}$$

Taking the double Fourier transforms of (4.238) and (4.239) gives

$$S_{b,m}(f)=\begin{cases}S_{r,0}(f) & m=0\\ m_r^{m-1}(T)M_r(f)\hat{M}^*_r(f) & m>0\end{cases}\tag{4.241}$$

where

$$m_r^{m-1}(T) \;=\; \mathrm{sinf}^{m-1}\beta(T)$$

$$M_r(f) \;=\; \mathcal{F}[m_r(t)]=\mathcal{F}[\mathrm{sinf}\beta(t)u_T(t)]$$

$$\hat{M}^*_r(f) \;=\; \frac{1}{2}\mathrm{E}\,[r(T,x_0)R^*(f,x_0)]=\frac{1}{2}\mathrm{E}\left[e^{j2\pi\overline{k_f}Tx_0}R^*(f,x_0)\right]$$

$\mathcal{F}[\,\cdot\,]$ denotes the Fourier transform and

$$R^*(f,x_0)=\mathcal{F}[r^*(t,x_0)]=\mathcal{F}\left[e^{-jx_0\beta(t)}u_T(t)\right]\;.\tag{4.242}$$

Then,

$$S_{\tilde{s}\tilde{s}}(f) = \frac{A^2}{T} \left(S_{r,0}(f) + 2\text{Re}\left\{ M_r(f)\hat{M}_r^*(f) \right. \right.$$
$$\left. \times \sum_{m=1}^{\infty} m_r^{m-1}(T)e^{-j2\pi f mT} \right\} \right)$$
$$= \frac{A^2}{T} \left(S_{r,0}(f) + 2\text{Re}\left\{ M_r(f)\hat{M}_r^*(f) \right. \right.$$
$$\left. \times \sum_{n=0}^{\infty} \left[m_r(T)e^{-j2\pi fT} \right]^n e^{-j2\pi fT} \right\} \right) .$$

(4.243)

Observe that

$$|r(t, x_k)| = \left| e^{jx_k\beta(t)} u_T(t) \right| = 1$$

(4.244)

so that

$$\left| m_r(T)e^{-j2\pi fT} \right| = |m_r(T)| \le 1 .$$

(4.245)

The implication of equation (4.245) is that two separate cases must be considered when evaluating the psd.

Case 1: $|m_r(T)| < 1.$ In this case the sum in (4.243) converges so that

$$S_{\tilde{s}\tilde{s}}(f) = \frac{A^2}{T} \left(S_{r,0}(f) + 2\text{Re}\left\{ \frac{M_r(f)\hat{M}_r^*(f)}{\exp\{j2\pi fT\} - m_r(T)} \right\} \right) .$$

(4.246)

and the psd has no discrete components.

Case 2: $|m_r(T)| = 1.$ This case is possible only if

$$|m_r(T)| = \left| \text{E}\left[e^{jx_k\beta(T)} \right] \right| = 1 .$$

(4.247)

For this condition to be true we must have

$$e^{jx_k\beta(T)} = e^{jc}, \quad \forall\, k$$

(4.248)

where c is a constant. However, $\beta(T) = 2\pi\overline{k_f}T$ so that $x_k\beta(T) = 2\pi\overline{k_f}T$ mod(2π) for all x_k. Then $r(T, x_0) = \exp\left\{ j2\pi\overline{k_f}T \right\}$ is a constant so that

$$m_r(T) = \text{E}[r(T, x_0)] = e^{j2\pi\overline{k_f}T}$$

(4.249)

and

$$\hat{M}_r^*(f) = M_r^*(f)e^{j2\pi\overline{k_f}T} .$$

(4.250)

Hence, the psd is

$$S_{\tilde{s}\tilde{s}}(f) = \frac{A^2}{T}\left(S_{r,0}(f) + |M_r(f)|^2 2\mathrm{Re}\left\{\sum_{m=1}^{\infty} e^{j2\pi(f-\overline{k_f})mT}\right\}\right)$$

$$= \frac{A^2}{T}\left(S_{r,0}(f) - |M_r(f)|^2 + |M_r(f)|^2\right.$$
$$\times \left.\sum_{m=-\infty}^{\infty} e^{-j2\pi(f-\overline{k_f})mT}\right)$$

$$= \frac{A^2}{T}\left(S_{r,0}(f) - |M_r(f)|^2 + \frac{1}{T}|M_r(f)|^2 \sum_{n=-\infty}^{\infty} \delta\left(f - \overline{k_f} - \frac{n}{T}\right)\right)$$

$$= \frac{A^2}{T}\left(S_{r,0}(f) - |M_r(f)|^2\right)$$
$$+ \left(\frac{A}{T}\right)^2 \sum_{m=-\infty}^{\infty} \left|M_r(\overline{k_f} + n/T)\right|^2 \delta\left(f - \overline{k_f} - \frac{n}{T}\right) \ .$$

$$(4.251)$$

Clearly, the second term in the above expression is a discrete spectral compo-
nent. Finally, if x_k assumes the values defined in (4.232), then

$$x_k 2\pi \overline{k_f} T = 2\pi \overline{k_f} T \qquad \mathrm{mod}\,(2\pi) \ . \tag{4.252}$$

However, $h = \beta(T)/\pi = 2\overline{k_f}T$. Therefore,

$$x_k h \pi = h\pi \qquad \mathrm{mod}\,(2\pi) \ . \tag{4.253}$$

Hence, h must be an integer for there to be a discrete spectral component.

9.7.1 PSD OF CPFSK

Suppose that h is a non-integer so that the psd has no discrete components.
Then

$$
\begin{aligned}
R(f, x_0) &= \int_0^T e^{j2\pi k_f t x_0} \cdot e^{-j2\pi f t} dt \\
&= T e^{-j\pi(f - x_0 k_f)T} \mathrm{Sa}\left(\pi(f - k_f x_0)T\right) \qquad (4.254)
\end{aligned}
$$

where

$$
\begin{aligned}
M_r(f) &= \mathrm{E}\left[R(f, x_0)\right] \\
&= \frac{T}{M} \sum_{m=1}^{M} e^{-j\pi\left(f-(2m-1-M)k_f\right)T} \\
&\quad \times \mathrm{Sa}\left(\pi\left(f-(2m-1-M)k_f\right)T\right) \quad\quad (4.255)
\end{aligned}
$$

$$
\begin{aligned}
S_{r,0}(f) &= \frac{1}{2}\mathrm{E}\left[|R(f, x_0)|^2\right] \\
&= \frac{T^2}{M} \sum_{m=1}^{M} {}^2\left(\pi\left(f-(2m-1-M)k_f\right)T\right) . \quad (4.256)
\end{aligned}
$$

Also,

$$
\begin{aligned}
\hat{M}_r^*(f) &= \frac{T}{2M} \sum_{m=1}^{M} e^{j\pi\left(f+(2m-1-M)k_f\right)T} \\
&\quad \times \mathrm{Sa}\left(\pi\left(f-(2m-1-M)k_f\right)T\right) \quad\quad (4.257)
\end{aligned}
$$

These expressions are used in (4.246) to obtain the psd. If h is an integer, then the psd will have a discrete component and

$$
M_r(f) = \frac{T}{M} e^{-j2\pi fT} \sin(\pi fT) \sum_{m=1}^{M} e^{-j\pi\left(f-(2m-1-M)k_f\right)T} . \quad (4.258)
$$

This expression can be used in (4.250) to obtain the psd.

If a binary modulation ($M = 2$) is used, the above expressions simplify even more. For the case when h is not an integer

$$
\Phi_{r,0}(f) = \frac{T^2}{4}\left\{\mathrm{Sa}^2\left(\pi(f-k_f)T\right) + \mathrm{Sa}^2\left(\pi(f+k_f)T\right)\right\} \quad (4.259)
$$

$$
\begin{aligned}
M_r(f) &= \left\{\frac{T}{2} e^{-j\pi(f+k_f)T}\mathrm{Sa}\left(\pi(f+k_f)T\right)\right. \\
&\quad \left. e^{-j\pi(f-k_f)T}\mathrm{Sa}\left(\pi(f-k_f)T\right)\right\} \quad\quad (4.260)
\end{aligned}
$$

$$
\begin{aligned}
\hat{M}_r^*(f) &= \frac{T}{4}\left\{e^{j\pi(f-k_f)T}\mathrm{Sa}\left(\pi(f+k_f)T\right)\right. \\
&\quad \left. + e^{j\pi(f+k_f)T}\mathrm{Sa}\left(\pi(f-k_f)T\right)\right\} \quad\quad (4.261)
\end{aligned}
$$

$$
m_r(T) = \mathrm{sinf}(2\pi k_f T) = \mathrm{sinf}(h\pi) . \quad (4.262)
$$

When h is an integer

$$
\Phi_{r,0}(f) = \frac{T^2 \sin^2 \pi fT}{2}\left\{\left(\frac{1}{\pi(f-k_f)T}\right)^2 + \left(\frac{1}{\pi(f+k_f)T}\right)^2\right\} \quad (4.263)
$$
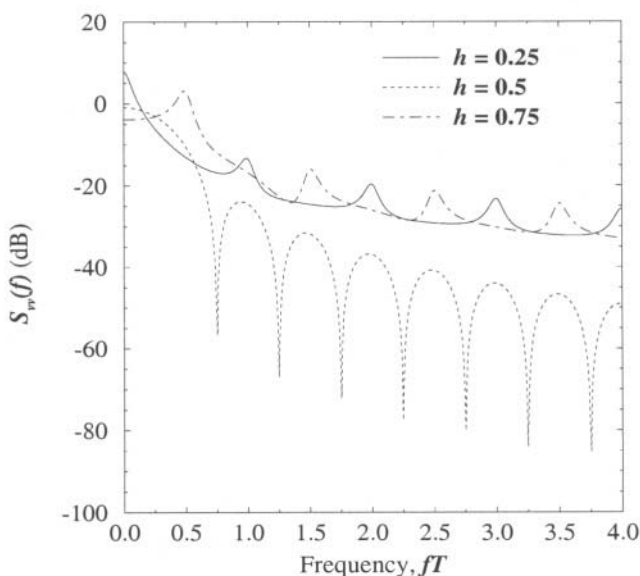
Figure 4.34a. Psd of binary CPFSK for various modulation indices.

$$M_r(f) = \frac{Te^{-j\pi fT}\sin\pi fT}{2}\left\{\frac{1}{\pi(f-k_f)T} + \frac{1}{\pi(f+k_f)T}\right\} \quad (4.264)$$

Figs. 4.34a and 4.34b plot the psd $4S_{\tilde{s}\tilde{s}}(f)/A^2 T$ against the frequency fT. The psd of MSK corresponds to $h = 0.5$. Note that modulation indices other than $h = 0.5$ result in a narrower main lobe than MSK, but larger sidelobes. Fig. 4.34b demonstrates the appearance of discrete components as $h \to 1$.

9.7.2   PSD OF MSK

An alternative method for computing the psd of MSK starts by recognizing that MSK is equivalent to OQASK with a half-sinusoid amplitude shaping pulse. It follows from (4.110) that the MSK baseband signal has the form

$$\tilde{s}(t) = A\sum_n b(t - 2nT, x_n) \quad (4.265)$$

where

$$b(t, x_n) = x_n^I h_a(t) + jx_n^Q h_a(t - T) \quad (4.266)$$

and

$$h_a(t) = \cos\left(\frac{\pi t}{2T}\right). \quad (4.267)$$
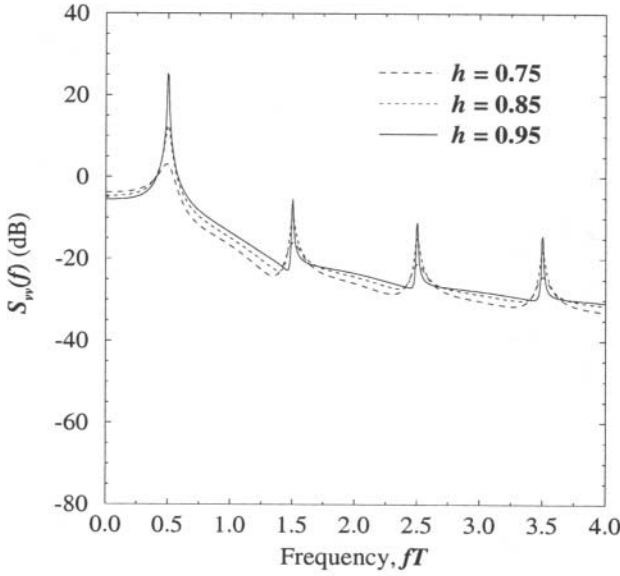
Figure 4.34b. Psd of binary CPFSK for various modulation indices.

The Fourier transform of (4.266) is

$$B(f, x_n) = \left(x_n^I + j x_n^Q \exp\{-j2\pi fT\} \right) H_a(f) \ . \tag{4.268}$$

Since the data sequence is uncorrelated, the psd can be computed from (4.195) and (4.196). Since the data sequence has zero mean, $S_{b,1}(f) = 0$. Also,

$$
\begin{aligned}
S_{b,0}(f) &= \frac{1}{2} \mathrm{E} \left[(x_n^I)^2 + (x_n^Q)^2 \right] |H_a(f)|^2 \\
&= \mathrm{E}[(x_n^I)^2] |H_a(f)|^2 \\
&= |H_a(f)|^2 \ . \tag{4.269}
\end{aligned}
$$

From (4.199)

$$S_{\tilde{s}\tilde{s}}(f) = \frac{32 A^2 T}{\pi^2} \left[\frac{\cos 2\pi fT}{1 - 16 f^2 T^2} \right]^2 \ . \tag{4.270}$$

Once again, the psd of MSK is plotted in Fig. 4.34a.

9.8 PSD OF GMSK AND TFM

GSMK and TFM are special cases of partial response CPM. In general, the psd of partial response CPM is difficult to obtain except for a rectangular shaping function. One solution has been suggested by Garrison [131], where the modulating pulses are approximated by using a large number of rectangular sub-pulses with properly chosen amplitudes.
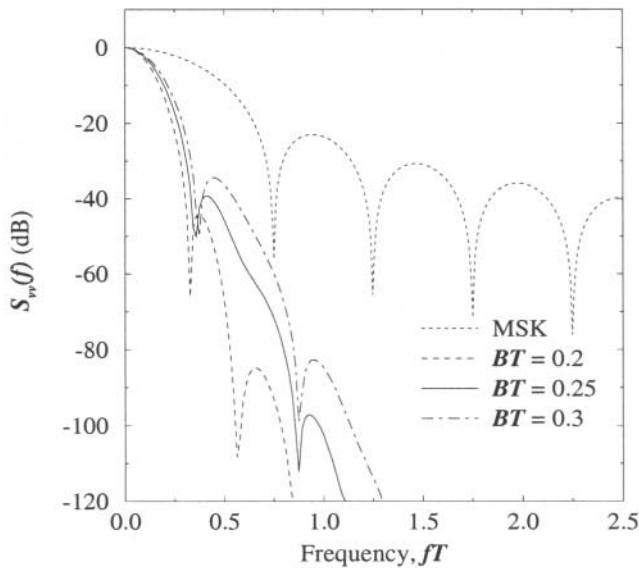
Figure 4.35.   Psd of GMSK with various normalized filter bandwidths BT.

Fig. 4.35 plots the psd of GMSK with various normalized filter bandwidths BT. Note that a smaller BT results in a more compact psd. Likewise, Fig. 4.36 plots the psd of TFM and GMSK with $BT = 0.25$. Observe that the psd of TFM compares well with that of GMSK. This is not surprising since their corresponding frequency shaping pulses are quite similar, comparing Figs. 4.24 and 4.28.

Finally, it is interesting to compare the spectral characteristics of GMSK and $\pi/4$-DQPSK since both methods are extensively used in mobile communication systems. To make a fair comparison, we must remember that GMSK transmits 1 bit/baud while $\pi/4$DQPSK transmits 2 bits/baud. If $\pi/4$-DQPSK uses root raised cosine pulse shaping, then the spectral occupancy normalized to a bit duration is is obtained by dividing the elements on the horizontal axis of Fig. 4.5 by a factor of 2. For example at $f = 1/(2T_b)$ (corresponding to $fT = 1.0$) the side lobes are about 44 dB down from the main lobe ($f = 0$) when $\tau = 6T$. From Fig. 4.35, with $f = 1/(2T)$, almost the same roll off is obtained. However, for larger values of f, the GMSK pulse is seen to decay faster.
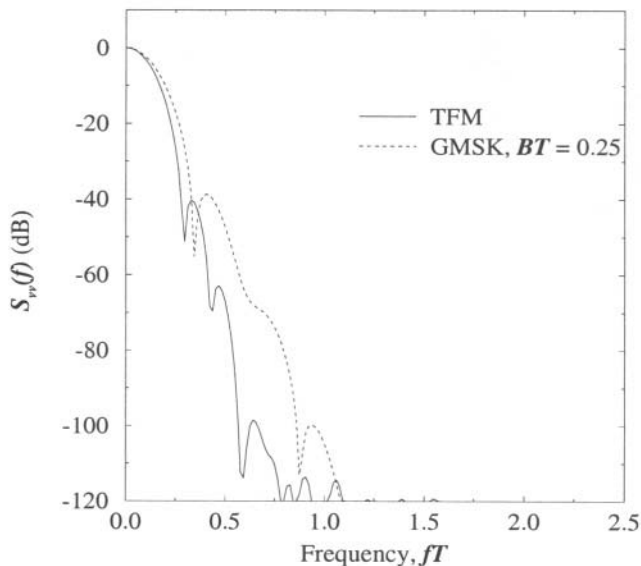
Figure 4.36. Psd of TFM and GMSK with $BT = 0.25$.

Problems

4.1. Assume that a received signal is given by

$$y(t) = \sum_{n=-\infty}^{\infty} x_n p(t - nT)$$

where $x_k = \pm 1$, and $p(t)$ is the ideal Nyquist pulse

$$p(t) = \text{sinc}(t/T)$$
$$P(f) = T\text{rect}(fT) .$$

There are two problems associated with this pulse shape. One is the problem of realizing a pulse having the rectangular spectral characteristic $P(f)$ given above. The other problem arises from the fact that the tails in $p(t)$ decay as $1/t$. Consequently, a sampling timing error results in an infinite series of ISI components. Such a series is not absolutely summable and, hence, the sum of the resulting interference does not converge.

Assume that $p(t) = 0$ for $|t| > NT$, where N is a positive integer. In spite of the restriction that the channel is band-limited, this assumption holds in all practical communication systems.

a) Due to a slight timing error, the received signal is sampled at $t = kT + t_0$, where $0 < t_0 < T$. Calculate the response for $t = kT + t_0$. Separate the response into two components, the desired term and an ISI term.

b) Assume that the polarities of x_i are such that every term in the ISI is positive, i.e., worst case ISI. Under this assumption show that the ISI term is

$$\text{ISI} = \frac{2}{\pi} \sin(\pi t_0/T) \sum_{n=1}^{N} \frac{n}{n^2 - t_0^2/T^2} \, .$$

and, therefore, $\text{ISI} \to \infty$ as $N \to \infty$.

4.2. Show that 16-QAM can be represented as a superposition of two four-phase constant envelope signals where each component is amplified separately before summing, i.e.,

$$s(t) = [A_n \cos 2\pi f_c t + B_n \sin 2\pi f_c t] + [C_n \cos 2\pi f_c t + D_n \sin 2\pi f_c t]$$

where $\{A_n\}$, $\{B_n\}$, $\{C_n\}$, and $\{D_n\}$ are statistically independent binary sequences with elements from the set $\{-1, +1\}$. Thus, show that the resulting signal is equivalent to

$$s(t) = I_n \cos 2\pi f_c t + Q_n \sin 2\pi f_c t$$

and determine I_n and Q_n in terms of A_n, B_n, C_n, and D_n.

4.3. An important parameter for digital modulation schemes is the *peak-to-mean envelope power ratio* (PMEPR), defined by

$$\text{PMEPR} = \frac{|\tilde{s}(t)|^2_{\text{peak}}}{\langle|\tilde{s}(t)|^2\rangle}$$

where $|\tilde{s}(t)|^2_{\text{peak}}$ is the largest value of $|\tilde{s}(t)|^2$ and $\langle|\tilde{s}(t)|^2\rangle$ is its time average. When non-linear power amplifiers are used it is desirable to keep the PMEPR as small as possible.

a) Plot PMEPR for $\pi/4$-DQPSK with root raised cosine pulse shaping, as a function of the roll-off factor β.

b) Repeat part a) for QPSK. What conclusions can you draw?

4.4. Two new modulation schemes have very recently been proposed for the UWC-136HS third generation system, called Q-O-QAM and B-O-QAM. Q-O-QAM transmits 2 bits/symbol, while B-O-QAM transmits 1 bit/symbol. The mapping of Q-O-QAM data bits (a_{2k}, a_{2k+1}) to symbols b_k is as follows: The symbols b_k are used to generate the symbols x_k which are given by

$$x_k = b_k e^{jk\frac{\pi}{2}}$$

For B-O-QAM the mapping of data bits a_k to symbols b_k is as follows: The

(a_{2k}, a_{2k+1})	b_k
0,0	+3
0,1	+1
1,0	-3
1,1	-1

a_k	b_k
0	+3
1	-3

symbols a_k are also used to generate the symbols x_k which are given by

$$x_k = b_k e^{jk\frac{\pi}{2}}$$

a) Plot the signal space diagram for Q-O-QAM and B-O-QAM and show the allowable transitions between the signal points in the signal constellation. Why would these modulation schemes be useful for radio transmitters that use non-linear power amplifiers.

b) Assuming an AWGN channel and coherent detection, write down an expression for the probability of *symbol* error for Q-O-QAM and B-O-QAM in terms of the bit energy to noise ratio γ_b.

4.5. Consider two sinusoids waveforms

$$\begin{aligned} s_1(t) &= A\cos 2\pi f_c t \\ s_2(t) &= A\cos 2\pi (f_c + \Delta_f)t \end{aligned}$$

a Determine the minimum value of Δ_f such that the inner product $(s_1, s_2) = 0$ over the interval $0 \le t \le T$. Assume that $f_c T \gg 1$.

b Repeat part a) for the two sinusoids

$$\begin{aligned} s_1(t) &= A\cos(2\pi f_c t + \phi_1) \\ s_2(t) &= A\cos(2\pi (f_c + \Delta_f)t + \phi_2) \end{aligned}$$

where ϕ_1 and ϕ_2 are arbitrary phases.

4.6. Suppose that an OFDM system is implemented with a guard interval that is a cyclic extension of the IFFT co-efficients as shown in (4.95).

a) Show that the output of the OFDM demodulator is given by (4.98).

b) Now suppose that the guard interval simply consists of a blank interval where nothing is transmitted. Assuming that $G \ge L > 0$ can the

data block $\mathbf{x}_0$ be recovered by taking an FFT of the received block $\mathbf{R}_0 = \{R_{0,k}\}_{k=0}^{N-1}$?

4.7. Consider a CPM signal that is generated by using a triangular frequency shaping pulse shown below

$h_f(t)$

1

0 $\dfrac{T}{2}$ T t
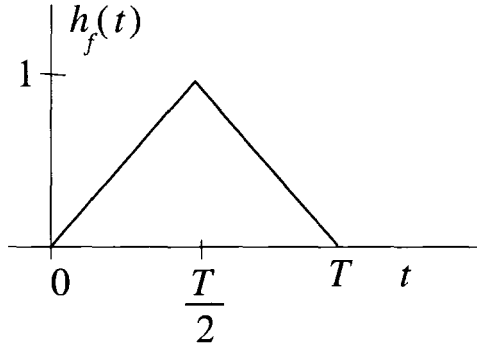
Figure 4.37.   Frequency Shaping Pulse

a) Find the peak frequency deviation k_f so that $\beta(T) = \pi/2$.

b) Sketch the phase tree and phase trellis for the binary source symbol sequence
$$\mathbf{x} = (+1, +1, +1, -1, -1, +1, -1, -1)$$

4.8. A CPM signal is generated from a baseband signal with a half-sinusoid shaping function.

a) If $h = 1/2$ find the peak frequency deviation from the carrier frequency.

b) Sketch the phase tree and phase trellis if the data symbol sequence is

$$\mathbf{x} = \{+3, -1, +1, +3, -3, +1, -1\} \ .$$

4.9. Sketch the phase-tree, the phase trellis, and phase state diagram for partial response CPM with $h = 1/2$ and

$$h_f(t) = u_{2T}(t) \ .$$

4.10. Consider a partial response CPM signal

a) Generate a shaping function of duration $3T$ by convolving a rectangular shaping function of duration T with a rectangular shaping function of duration $2T$.

b) Define and sketch the three segments of the shaping function, $h_{f,k}(t)$, $k = 0, 1, 2$.

c) Sketch the baseband signal if the symbol sequence is

$$\mathbf{x} = \{+1, -1, +1, -1, -1\} \ .$$

4.11. What are the phase states and states for the following CPM signals:

a) Full response binary CPFSK with either $h = 2/3$ or $h = 3/4$.

b) Partial response $L = 3$ binary CPFSK with either $h = 2/3$ or $h = 3/4$.

4.12. Consider a multi-h CPM waveform with the h sequence $\{h_1, h_2\} = \{\frac{1}{4}, \frac{1}{2}\}$ and the frequency shaping pulse $h_f(t) = U_T(t)$. In the ith signaling interval the excess phase changes by $\pm\pi/4$ radians if $h_1 = \frac{1}{4}$ is used and by $\pm\pi/2$ radians if $h_2 = \frac{1}{2}$ is used.

a) Plot the phase-trellis assuming the initial phase $\theta_o = 0$.

b) Indicate the phase trajectory for the symbol sequence

$$\mathbf{x} = \{1, -1, -1, 1, 1, 1, -1\} \ .$$

4.13. Design a Gaussian pulse-shaping filter with $BT = 0.5$ for a symbol rate of 19.2 kbps. Write expressions for and plot, i) the impulse response and frequency response of the filter, and ii) the frequency shaping pulse $h_f(t)$. Repeat for the case of $BT = 0.2$ and $BT=0.75$.

4.14. Consider TFM with the frequency shaping pulse

$$H_f(f) = \frac{\pi}{4h} \frac{\pi f T}{\sin(\pi f T)} \cos^2(\pi f T) \ .$$

Suppose that this pulse is obtained by exciting a filter $\tilde{h}(t)$ with a gate function rect(t/T). Find and sketch the impulse response of the filter $\tilde{h}(t)$.

4.15. Prove the identity

$$T \sum_m e^{-j2\pi f m T} = \sum_n \delta \left(f - \frac{n}{T} \right)$$

4.16. Consider the case of uncorrelated data symbols.

a) Show that if the symbols are equiprobable, then

$$E\left[|B(f, x_0)|^2 \right] - |E\left[B(f, x_0) \right]|^2$$

$$= \frac{1}{2M^2} \sum_{i=1}^{M} \sum_{k=1}^{M} |B(f, x_i) - B(f, x_k)|^2 \ .$$

b) Compute the value of part a) for $M = 2$.

4.17. Consider the complex low-pass binary modulated signal

$$\tilde{s}(t) = A \sum_n x_n h_a(t - nT)$$

where $x_n \in \{-1, +1\}$. The data sequence $\{x_n\}$ is correlated such that

$$\phi_{xx}(n) = \frac{1}{2}\mathrm{E}[x_k x^*_{k+n}] = \rho^{|n|} .$$

Compute the power density spectrum of $\tilde{s}(t)$.

4.18. Suppose that a binary data sequence $\mathbf{x}, x_i \in \{-1, +1\}$ is correlated such that $P(x_n = x_{n+1}) = 3/4$, i.e., adjacent data bits are the same with probability 3/4 and different with probability 1/4.

a) Compute the autocorrelation function $\phi_{xx}(m)$ for this data sequence.

b) Compute the power spectrum $S_{xx}(f)$.

4.19. Suppose that an uncorrelated binary data sequence is transmitted by using binary PAM with a root-Gaussian amplitude shaping pulse

$$H_a(f) = \left[A\tau e^{-\pi(f\tau)^2} \right]^{1/2}$$

a) What is the transmitted power density spectrum?

b) Find the value of τ so that the power density spectrum is 20 dB below its peak value at frequency $1/T$, where T is the baud duration.

c) What is the corresponding time domain pulse $h_a(t)$?

4.20. Consider a system that uses a set of $M = 16$ bi-orthogonal signals that are derived from the Hadamard matrix $\mathbf{H}_8$ in (4.76). The set of 16 signals is constructed according to

$$\tilde{s}_i(t) = A \sum_{k=0}^{7} h_{ik} h_c(t - kT_c), \qquad k = 1, \ldots, 8$$

$$= -\tilde{s}_i(t), \qquad k = 9, \ldots, 16$$

where $T = 7T_c$ is the baud period. Note that 4 bits are transmitted per baud.

Assume an uncorrelated data sequence and assume that all 16 waveforms are used with equal probability.

a If $h_c(t) = u_{T_c}(t)$, find the psd of the transmitted complex envelope $S_{\tilde{s}\tilde{s}}(f)$.

b Plot the power spectrum $S_{\tilde{s}\tilde{s}}(f)$ against the normalized frequency fT_b, where $T_b = T/4$ is the bit duration.

This page intentionally left blank

Chapter 5

DIGITAL SIGNALING
 ON FLAT FADING CHANNELS

The performance of a digital modulation scheme is degraded by many transmission impairments including fading, delay spread, Doppler spread, co-channel and adjacent channel interference, and noise. Fading causes a very low instantaneous received signal-to-noise ratio (SNR) or carrier-to-noise ratio (CNR) when the channel exhibits a deep fade, delay spread causes intersymbol interference (ISI) between the transmitted symbols, and a large Doppler spread is indicative of rapid channel variation and necessitates a receiver with a fast convergent algorithm. Co-channel interference, adjacent channel interference, and noise, are all additive distortions that degrade the bit error rate performance by reducing the CNR or SNR.

This chapter derives the bit error rate performance of digital signaling on frequency non-selective (flat) fading channel with AWGN. Flat fading channel models are appropriate for narrow-band land mobile radio systems or mobile satellite systems. Flat fading channels affect all frequency components of a narrow-band signal in exactly the same way and, therefore, do not introduce amplitude or phase distortion into the received signal. Frequency selective channels distort the transmitted signal and will be the subject of Chapter 6. Flat fading channel will be shown to significantly degrade the bit error rate performance unless appropriate countermeasures are taken. Diversity and coding techniques are well known methods for combating fading. The basic idea of diversity systems is to provide the receiver with multiple replicas of the same information bearing signal, where the replicas are affected by uncorrelated fading. Coding techniques introduce a form of time diversity into the transmitted signal which can be exploited to mitigate the effects of fading.

The remainder of this chapter is organized as follows. Section 1. introduces a vector representation for digital signaling on flat fading channels with additive white Gaussian noise (AWGN). Section 3. provides a generalized analysis of

the error rate performance of digital signaling on flat fading channels. Section 2. derives the structure of the optimum coherent receiver for the detection of known signals in AWGN. The error probability performance of various coherently detected digital signaling schemes is considered including PSK in Section 4., QAM in Section 5., orthogonal signals in Section 6., and OFDM in Section 7.. Section 9. considers differential detection of PSK and $\pi/4$-QPSK. Section 10. considers non-coherent detection and, finally, Section 11. considers coherent and non-coherent detection of CPM signals.

1. VECTOR SPACE REPRESENTATION OF RECEIVED SIGNALS

Suppose that one of M complex low-pass waveforms $\{\tilde{s}_k(t)\}_{k=0}^{M-1}$, say $\tilde{s}_i(t)$, is transmitted on a flat fading channel with additive white Gaussian noise (AWGN). For such a channel, the received complex envelope is

$$\tilde{r}(t) = g(t)\tilde{s}_i(t) + \tilde{n}(t) \tag{5.1}$$

where $g(t) = \alpha(t)e^{j\phi(t)}$ is the complex fading gain introduced by the channel, and $\tilde{n}(t)$ is zero-mean complex AWGN with a power spectral density (psd) of N_o watts/Hz. At any time t the complex fading gain $g(t)$ is a complex Gaussian random variable. The receiver must determine which message waveform $\tilde{s}_k(t)$ was transmitted from the observation of received signal $\tilde{r}(t)$.

In our present development, the pulses $\tilde{s}_m(t)$ are assumed to have duration T. However, our results will also apply to the case of root Nyquist pulses with duration $\tau \neq T$; for example, the root raised cosine pulse. The only difference is the length of the required observation interval.

If the channel changes very slowly with respect to the data symbol duration, i.e., $f_m T \ll 1$, then $g(t)$ will effectively remain constant over the observation interval1. Under this condition, the explicit time dependency of $g(t)$ can be removed so the received signal becomes

$$\tilde{r}(t) \quad = \quad g\tilde{s}_i(t) + \tilde{n}(t) \tag{5.2}$$

where $g = \alpha e^{j\phi}$ is the fading gain. If the Gaussian fading process has zero (non-zero) mean then the magnitude α is Rayleigh (Ricean) distributed and the phase ϕ is uniformly (non-uniformly) distributed over $[-\pi, \pi]$ as described in Chapter 2.1.2.

To facilitate the derivation of the optimum receiver and its analysis, it is useful to introduce a vector space representation for the received signals. If the channel is affected by AWGN, then the required basis functions are those obtained by using the Gram-Schmidt orthonormalization procedure outlined in

---

[1] For land mobile radio applications $f_m T \ll 1$ is a reasonable assumption.

Chapter 4.1.2. Using these basis functions, the received signal can be expressed as

$$\tilde{r}(t) = \sum_{n=0}^{N-1} \tilde{r}_n \varphi_n(t) + \tilde{z}(t) \tag{5.3}$$

where

$$\tilde{r}_n = \int_0^T \tilde{r}(t) \varphi_n^*(t) dt \tag{5.4}$$

$$= g \int_0^T \tilde{s}_i(t) \varphi_n^*(t) dt + \int_0^T \tilde{n}(t) \varphi_n^*(t) dt \tag{5.5}$$

$$= g\tilde{s}_{i_n} + \tilde{n}_n \tag{5.6}$$

and

$$\tilde{z}(t) = \tilde{n}(t) - \sum_{n=0}^{N-1} \tilde{n}_n \varphi_n(t) \ . \tag{5.7}$$

The above process yields the vector

$$\tilde{\mathbf{r}} = g\tilde{\mathbf{s}}_i + \tilde{\mathbf{n}} \tag{5.8}$$

where

$$\begin{aligned}
\tilde{\mathbf{r}} &= (\tilde{r}_0, \tilde{r}_1, \dots, \tilde{r}_{N-1}) \\
\tilde{\mathbf{s}}_i &= (\tilde{s}_{i_0}, \tilde{s}_{i_1}, \dots, \tilde{s}_{i_{N-1}}) \\
\tilde{\mathbf{n}} &= (\tilde{n}_0, \tilde{n}_1, \dots, \tilde{n}_{N-1}) \ .
\end{aligned}$$

For an AWGN channel, the $\tilde{n}_k$, $k = 0, \dots, N-1$ are Gaussian random variables that can be completely described by their means and covariances. The means are

$$\mathrm{E}[\tilde{n}_k] = \int_0^T \mathrm{E}[\tilde{n}(t)] \varphi_k(t) dt = 0 \tag{5.9}$$

and covariances are

$$\begin{aligned}
\mu_{\tilde{n}_j \tilde{n}_k} = \frac{1}{2} \mathrm{E}[\tilde{n}_j \tilde{n}_k^*] &= \int_0^T \int_0^T \frac{1}{2} \mathrm{E}[\tilde{n}(t)\tilde{n}^*(s)] \varphi_j(t) \varphi_k^*(s) dt ds \\
&= N_o \int_0^T \int_0^T \delta(t-s) \varphi_j(t) \varphi_k^*(s) dt ds \\
&= N_o \int_0^T \varphi_j(t) \varphi_k^*(t) dt \\
&= N_o \delta_{jk}
\end{aligned}$$

It follows that the $\tilde{n}_k$ are independent complex Gaussian random variables with zero mean and variance N_o. Hence, the vector $\tilde{\mathbf{n}}$ has the multivariate Gaussian pdf (A.40)

$$
\begin{aligned}
p(\tilde{\mathbf{n}}) &= \prod_{i=0}^{N-1} \frac{1}{2\pi N_o} \exp\left\{-\frac{1}{2N_o}|\tilde{n}_i|^2\right\} \\
&= \frac{1}{(2\pi N_o)^N} \exp\left\{-\frac{1}{2N_o}\|\tilde{\mathbf{n}}\|^2\right\} \; .
\end{aligned} \tag{5.10}
$$

Such noise is said to be *circularly symmetric,* because the joint pdf $p(\tilde{\mathbf{n}})$ appears as a hyperspherical cloud that is centered at the origin in the N-D vector space.

The waveform $\tilde{z}(t)$ is a "remander process" due to the fact that $\tilde{n}(t)$ lies outside the vector space that is spanned by the basis functions $\{\varphi_n(t)\}_{n=0}^{N-1}$. However,

$$
\begin{aligned}
\mathrm{E}[\tilde{z}(t)r_j^*] &= \mathrm{E}[\tilde{z}(t)]g\tilde{s}_{m_j}^* + \mathrm{E}[\tilde{z}(t)\tilde{n}_j^*] \\
&= \mathrm{E}[\tilde{z}(t)\tilde{n}_j^*] \\
&= \mathrm{E}\left[\left(\tilde{n}(t) - \sum_{n=0}^{N-1} \tilde{n}_n\varphi_n(t)\right)\tilde{n}_j^*\right] \\
&= \int_0^T \mathrm{E}[\tilde{n}(t)\tilde{n}^*(\tau)]\varphi_j(\tau)d\tau - \sum_{n=0}^{N-1}\mathrm{E}[\tilde{n}_n\tilde{n}_j^*]\varphi_n(t) \\
&= N_o\varphi_j(t) - N_o\varphi_j(t) = 0
\end{aligned}
$$

Since $\mathrm{E}[\tilde{z}(t)\tilde{r}_j^*] = 0$, $j = 0, \ldots, N-1$, it follows that $\tilde{z}(t)$ is uncorrelated with the received vector $\tilde{\mathbf{r}}$. This property implies that the remainder process $\tilde{z}(t)$ is irrelevant when making the decision as to which signal waveform was transmitted, a result known as Wozencraft's irrelevance theorem [365]. In other words, the received vector $\tilde{\mathbf{r}}$ is the only data useful for the decision process and, hence, represents "sufficient statistics" for the problem at hand.

2. DETECTION OF KNOWN SIGNALS IN ADDITIVE WHITE GAUSSIAN NOISE

The maximum *a posteriori* probability (MAP) receiver generates the vector $\tilde{\mathbf{r}}$ and from its observation decides in favor of the message vector $\tilde{\mathbf{s}}_i$ that maximizes the *a posteriori* probability $\mathrm{P}(g\tilde{\mathbf{s}}_i|\tilde{\mathbf{r}})$. If $\tilde{\mathbf{r}}$ is received and the decision is made in favor of the signal vector $\tilde{\mathbf{s}}_m$, then the conditional probability of decision error is

$$
\mathrm{P}_{e|\tilde{\mathbf{r}}} = 1 - \mathrm{P}(g\tilde{\mathbf{s}}_m|\tilde{\mathbf{r}}) \; . \tag{5.11}
$$

The unconditional probability of decision error is

$$
\mathrm{P}_e = \int_{-\infty}^{\infty} (1 - \mathrm{P}(g\tilde{\mathbf{s}}_m|\tilde{\mathbf{r}}))\, p(\tilde{\mathbf{r}})d\tilde{\mathbf{r}} \; . \tag{5.12}
$$

The MAP receiver clearly minimizes the probability of decision error, since the integrand is minimized for all possible received vectors $\tilde{\mathbf{r}}$.

By using Bayes' theorem, the *a posteriori* probability $P(g\tilde{\mathbf{s}}_m|\tilde{\mathbf{r}})$ can be expressed in the form

$$P(g\tilde{\mathbf{s}}_m|\tilde{\mathbf{r}}) = \frac{p(\tilde{\mathbf{r}}|g\tilde{\mathbf{s}}_m)P_m}{p(\tilde{\mathbf{r}})} \ , \quad m = 0, \ \cdots, \ M-1 \tag{5.13}$$

where $p(\tilde{\mathbf{r}}|g\tilde{\mathbf{s}}_m)$ is the joint conditional probability density function (pdf) of the received vector $\tilde{\mathbf{r}}$ given the transmitted message vector $\tilde{\mathbf{s}}_m$, and P_m is the prior probability of transmitting $\tilde{\mathbf{s}}_m$. Since the pdf of the received vector $p(\tilde{\mathbf{r}})$ is independent of the transmitted message vector, the MAP receiver chooses the vector $\tilde{\mathbf{s}}_m$ to maximize $p(\tilde{\mathbf{r}}|g\tilde{\mathbf{s}}_m)P_m$. In other words, the MAP decision rule is

$$\text{choose } \tilde{\mathbf{s}}_m \text{ if } \quad p(\tilde{\mathbf{r}}|g\tilde{\mathbf{s}}_m)P_m \geq p(\tilde{\mathbf{r}}|g\tilde{\mathbf{s}}_{\hat{m}})P_{\hat{m}} \ \ \forall \, \hat{m} \neq m \ . \tag{5.14}$$

Note that the MAP receiver requires knowledge of the complex channel gain *g,* implying that the receiver must employ an adaptive channel estimator.

A receiver that chooses the vector $\tilde{\mathbf{s}}_m$ to maximize $p(\tilde{\mathbf{r}}|g\tilde{\mathbf{s}}_m)$ regardless of the prior messages probabilities is called a **maximum likelihood** (ML) receiver. The ML decision rule is

$$\text{choose } \tilde{\mathbf{s}}_m \text{ if } \quad p(\tilde{\mathbf{r}}|g\tilde{\mathbf{s}}_m) \geq p(\tilde{\mathbf{r}}|g\tilde{\mathbf{s}}_{\hat{m}}) \ \ \forall \, \hat{m} \neq m \ . \tag{5.15}$$

If the prior message probabilities are equal, i.e., $P_m = 1/M$, then selection of the signal vector that maximizes $p(\tilde{\mathbf{r}}|g\tilde{\mathbf{s}}_m)$ also maximizes $p(g\tilde{\mathbf{s}}_m|\tilde{\mathbf{r}})$. Under this condition the ML receiver also minimizes the probability of decision error. The prior message probabilities will be equal when the source coding is good. In practice, an ML receiver is quite often implemented regardless of the prior message probabilities, because they may unknown. Note also, that the ML receiver requires knowledge of the channel gain *g.*

To proceed further we need the joint conditional pdf $p(\tilde{\mathbf{r}}|g\tilde{\mathbf{s}}_m)$. Since $\tilde{\mathbf{r}} = g\tilde{\mathbf{s}}_m + \tilde{\mathbf{n}}$ and $\tilde{\mathbf{n}}$ has the joint pdf in (5.10), we have

$$p(\tilde{\mathbf{r}}|g\tilde{\mathbf{s}}_m) = \frac{1}{(2\pi N_o)^N} \exp\left\{ -\frac{1}{2N_o}\|\tilde{\mathbf{r}} - g\tilde{\mathbf{s}}_m\|^2 \right\} \ . \tag{5.16}$$

By using (5.16) in (5.15), it is apparent that the signal vector $\tilde{\mathbf{s}}_m$ that maximizes $p(\tilde{\mathbf{r}}|g\tilde{\mathbf{s}}_m)$ also maximized the metric (or distance measure)

$$\mu_1(\tilde{\mathbf{s}}_m) = -\|\tilde{\mathbf{r}} - g\tilde{\mathbf{s}}_m\|^2 \tag{5.17}$$

In other words, the ML receiver decides in favor of the scaled message vector $g\tilde{\mathbf{s}}_m$ that is closest in squared Euclidean distance to the received vector $\tilde{\mathbf{r}}$. Such a receiver is said to make **minimum distance** decisions.

An alternative form of the ML receiver can be derived by expanding (5.17) as

$$\mu_1(\tilde{\mathbf{s}}_m) = -\|\tilde{\mathbf{r}}\|^2 + 2\mathrm{Re}\left\{(\tilde{\mathbf{r}}, g\tilde{\mathbf{s}}_m)\right\} - |g|^2 \|\tilde{\mathbf{s}}_m\|^2 \qquad (5.18)$$

Then notice that $\|\tilde{\mathbf{r}}\|^2$ is independent of $\tilde{\mathbf{s}}_m$ and $\|\tilde{\mathbf{s}}_m\|^2 = 2E_m$. Hence, the ML just needs to maximize the metric

$$\mu_2(\tilde{\mathbf{s}}_m) = \mathrm{Re}\left\{(\tilde{\mathbf{r}}, g\tilde{\mathbf{s}}_m)\right\} - |g|^2 E_m \ . \qquad (5.19)$$

Using the definition of the inner product, the above metric can be written in the alternate form

$$\mu_2(\tilde{\mathbf{s}}_m) = \mathrm{Re}\left\{\int_0^T \tilde{r}(t) g^* \tilde{s}_m^*(t) dt\right\} - |g|^2 E_m$$

$$\equiv \mathrm{Re}\left\{\int_0^T \tilde{r}(t) e^{-j\phi} \tilde{s}_m^*(t) dt\right\} - \alpha E_m \ . \qquad (5.20)$$

The last line in (5.20) follows because the $\mu_2(\tilde{\mathbf{s}}_m)$ can be divided by α without altering the decision process.

From the above development, the form of the ML receiver is clear. The receiver must first perform **quadrature demodulation** as shown in Fig. 5.1 to extract the complex envelope $\tilde{r}(t) = \tilde{r}_I(t) + j\tilde{r}_Q(t)$. The received bandpass waveform is

$$r(t) = \tilde{r}_I(t) \cos 2\pi f_c t - \tilde{r}_Q(t) \sin 2\pi f_c t \ . \qquad (5.21)$$

Then

$$[r(t) \cdot 2 \cos 2\pi f_c t]_{\mathrm{LP}} = \tilde{r}_I(t) \qquad (5.22)$$
$$[-r(t) \cdot 2 \sin 2\pi f_c t]_{\mathrm{LP}} = \tilde{r}_Q(t) \qquad (5.23)$$

where $[\ \cdot\ ]_{\mathrm{LP}}$ is just a low pass filter to reject the double frequency term after demodulation. After quadrature demodulation, there are several receiver structures that are functionally equivalent, but differ in their method of implementation and complexity. As shown in Fig. 5.2, one possibility is to generate the observation vector $\tilde{\mathbf{r}}$ by correlating the received complex envelope with the basis functions. This receiver structure is called a **correlation detector.** An functionally equivalent structure is shown in Fig. 5.3, where the complex envelope is filtered with a bank of filters having impulse responses $\varphi_i^*(T - t)$ and sampling the outputs at time T. The filter $<\varphi_i^*(T - t)$ is the *matched filter* to $\varphi_i(t)$. This receiver structure is called a **matched filter detector** The matched filter can be shown to be the filter that maximizes the signal-to-noise ratio at the sampling instant when the input consists of a signal corrupted by AWGN. This result is available in numerous textbooks and we do not present it here (but see Problem 5.2). Finally, the metric computer in Fig. 5.4 processes the
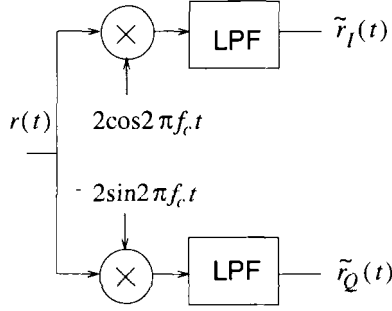
Figure 5.1.   Quadrature demodulator.

observation vector $\tilde{\mathbf{r}}$ to produce the metrics $\mu(\tilde{\mathbf{b}}_m)$, $m = 0, \ldots, M - 1$. The decision is made in favor of the data symbol corresponding to the largest metric.

To show equivalence of the correlation and matched filter detectors, let $h_i(t) = \varphi_i^*(T - t)$ denote the filter matched to $\varphi_i(t)$. Then the output of the matched filter is the convolution

$$y(t) = \int_0^t \tilde{r}(\tau)h_i(t - \tau)d\tau$$

$$= \int_0^t \tilde{r}(\tau)\varphi_i^*(T - t + \tau)d\tau \ . \tag{5.24}$$

Sampling the filter output at time T gives

$$y(T) = \int_0^T \tilde{r}(\tau)\varphi_i^*(\tau)d\tau \tag{5.25}$$

This is exactly the same as the correlation in (5.4). We note that other variations of the ML receiver can be constructed in a similar fashion by direct implementation of (5.20) by using either a bank of M correlators or a bank of M matched filters.

Some simplifications can be made for certain types of signal sets. If the message waveforms have equal energy such as PSK signals, then $E_m = E$ for all m. Hence, the bias term αE_m in (5.20) can be neglected, leading to

$$\mu_3(\tilde{\mathbf{s}}_m) = \text{Re}\left\{(\tilde{\mathbf{r}}, e^{j\phi}\tilde{\mathbf{s}}_m)\right\} \tag{5.26}$$

$$= \text{Re}\left\{\int_0^T \tilde{r}(t)e^{-j\phi}\tilde{s}_m^*(t)dt\right\} \ . \tag{5.27}$$

In this case, the receiver does not need to know the complete complex channel gain $g = \alpha e^{j\phi}$, but only the random carrier phase ϕ. The random carrier phase can be obtained in practice by using a phase locked loop.
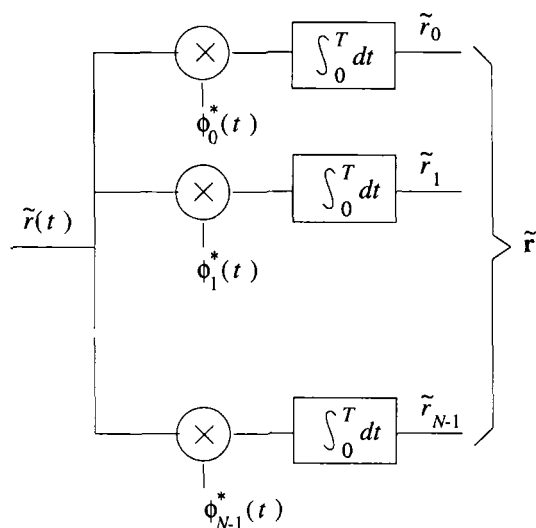
Figure 5.2. Correlator detector.
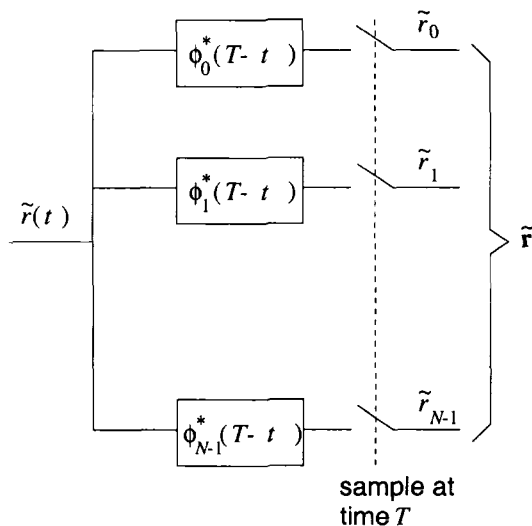
Figure 5.3. Matched filter detector.

3. PROBABILITY OF ERROR

Consider a signal constellation defined by the set M signal vectors $\{\tilde{\mathbf{s}}_m\}_{m=0}^{M-1}$. Throughout this section, we assume equally likely messages so that $P_m = 1/M$. By observing the vector $\tilde{\mathbf{r}}$, the ML receiver chooses the message vector $\tilde{\mathbf{s}}_m$ that minimizes the squared Euclidean distance $\|\tilde{\mathbf{r}} - g\tilde{\mathbf{s}}_m\|^2$. To compute the probability of ML decision error for an arbitrary set of signal vectors, we
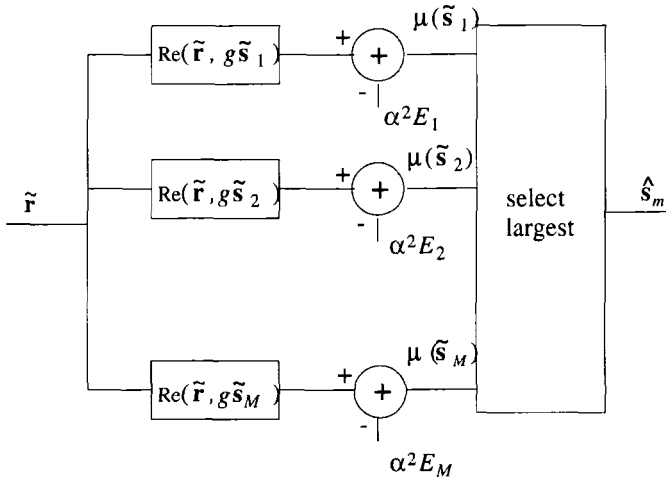
Figure 5.4. Metric computer.

first a define convex decision regions R_m around each of the signal points $g\tilde{\mathbf{s}}_m$ in the N-D signal space. Fig. 5.5 shows an example of the decision regions. Formally, the decision regions are defined by

$$R_m = \left\{ \tilde{\mathbf{r}} : \|\tilde{\mathbf{r}} - g\tilde{\mathbf{s}}_m\|^2 \leq \|\tilde{\mathbf{r}} - g\tilde{\mathbf{s}}_{\hat{m}}\|^2 , \ \forall \hat{m} \neq m \right\} . \tag{5.28}$$

Observe that all $\tilde{\mathbf{r}} \in R_m$ are closer to $g\tilde{\mathbf{s}}_m$ than to any other signal point $g\tilde{\mathbf{s}}_k, k \neq m$. The ML decision rule becomes

$$\text{choose } \tilde{\mathbf{s}}_m \text{ if } \tilde{\mathbf{r}} \in R_m \tag{5.29}$$

The decision boundaries are hyperplanes that are defined by the locus of signal points that are equidistant from two neighboring signal vectors.

The conditional error probability associated with $\tilde{\mathbf{s}}_m$ is

$$
\begin{aligned}
P(e|\tilde{\mathbf{s}}_m) &= P(\tilde{\mathbf{r}} \notin R_m) & (5.30) \\
&= 1 - P(\tilde{\mathbf{r}} \in R_m) & (5.31) \\
&= 1 - P(c|\tilde{\mathbf{s}}_m) & (5.32)
\end{aligned}
$$

where $P(c|\tilde{\mathbf{s}}_m)$ is the conditional probability of correct reception. By using the joint conditional pdf in (5.16) we can write

$$P(e|\tilde{\mathbf{s}}_m) = 1 - \int_{R_m} p(\tilde{\mathbf{r}}|g\tilde{\mathbf{s}}_m)d\tilde{\mathbf{r}} \tag{5.33}$$

Finally, the average probability of decision error is

$$P(e) = \frac{1}{M} \sum_{m=0}^{M-1} P(e|\tilde{\mathbf{s}}_m) . \tag{5.34}$$
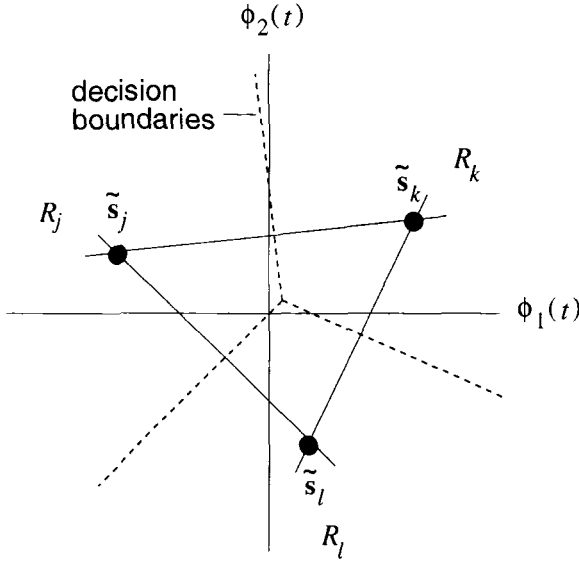
Figure 5.5. Decision regions in an 2-D signal-space.

Sometimes it may be difficult to compute the exact probability of decision error, due to the difficulty in defining the decision regions R_m and performing the integration in (5.33). In this case, various upper and lower bounds, and approximations on the probability of error are useful. First we introduce the concept of the pairwise error probability.

3.1 PAIRWISE ERROR PROBABILITY

Now consider two of the M signal vectors $\tilde{\mathbf{s}}_j$ and $\tilde{\mathbf{s}}_k$. We wish to determine the probability of decision error at the receiver, as if these two signal points are the only ones that exist. This error probability is called the **pairwise error probability** because it can be defined for each pair of signal vectors in the signal constellation. The two signal vectors $\tilde{\mathbf{s}}_j$ and $\tilde{\mathbf{s}}_k$ are separated *at the receiver* by the squared Euclidean distance $\|g\tilde{\mathbf{s}}_j - g\tilde{\mathbf{s}}_k\|^2 = \alpha^2 \|\tilde{\mathbf{s}}_j - \tilde{\mathbf{s}}_k\|^2$. Due to the circularly symmetric property of the AWGN noise, the pdf of the noise vector $\tilde{\mathbf{n}}$ is invariant to its rotation about the origin in the vector space. Hence, the noise component along the vector $g\tilde{\mathbf{s}}_j - g\tilde{\mathbf{s}}_k$ that joins the two signal vectors has zero mean and variance N_o.

A decision boundary can be established at the midpoint between the two signal vectors as shown in Fig. 5.6. Suppose that vector $\tilde{\mathbf{s}}_j$ is sent, and let $P(e|\tilde{\mathbf{s}}_j)$ denote the probability of ML decision error. The error probability is just the probability that the noise along the vector $g\tilde{\mathbf{s}}_j - g\tilde{\mathbf{s}}_k$ forces the received vector $\tilde{\mathbf{r}} = \tilde{\mathbf{s}}_j + \tilde{\mathbf{n}}$ to cross the decision boundary. This probability is just equal
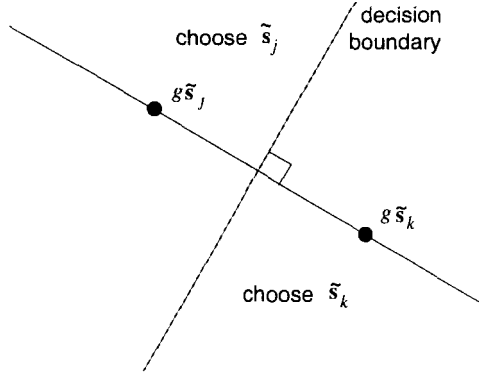
Figure 5.6.   Two received signal points in an *N*-D signal-space.

to

$$P(e|\tilde{\mathbf{s}}_j) = Q\left(\sqrt{\frac{\alpha^2 \tilde{d}_{jk}^2}{4N_o}}\right) . \qquad (5.35)$$

where $\tilde{d}_{jk}^2 = \|\tilde{\mathbf{s}}_j - \tilde{\mathbf{s}}_k\|^2$ is the squared Euclidean distance between $\tilde{\mathbf{s}}_j$ and $\tilde{\mathbf{s}}_k$. Finally, we note that $P(e|\tilde{\mathbf{s}}_j) = P(e|\tilde{\mathbf{s}}_k)$. Hence, the pairwise error probability between the message vectors $\tilde{\mathbf{s}}_j$ and $\tilde{\mathbf{s}}_k$ is

$$P(\tilde{\mathbf{s}}_j, \tilde{\mathbf{s}}_k) = P(e|\tilde{\mathbf{s}}_j) = P(e|\tilde{\mathbf{s}}_k) = Q\left(\sqrt{\frac{\alpha^2 \tilde{d}_{jk}^2}{4N_o}}\right) . \qquad (5.36)$$

3.2   UPPER BOUNDS ON ERROR PROBABILITY

Suppose that $\tilde{\mathbf{s}}_k$ is transmitted and let E_j denote the event that the receiver chooses $\tilde{\mathbf{s}}_j$ instead. The probability of the event E_j is the pairwise error probability $P(\tilde{\mathbf{s}}_j, \tilde{\mathbf{s}}_k)$. The conditional probability of decision error is

$$P(e|\tilde{\mathbf{s}}_k) = P\left(\bigcup_{j \neq k} E_j\right) \qquad (5.37)$$

By using the **union bound**

$$P\left(\bigcup_{j \neq k} E_j\right) \leq \sum_{j \neq k} P(E_j) \qquad (5.38)$$

we have

$$P(e|\tilde{\mathbf{s}}_k) \leq \sum_{j \neq k} P(\tilde{\mathbf{s}}_j, \tilde{\mathbf{s}}_k) \qquad (5.39)$$

Combining the above result with (5.36) gives

$$P(e|\tilde{\mathbf{s}}_k) \leq \sum_{j \neq k} Q\left(\sqrt{\frac{\alpha^2 \tilde{d}_{jk}^2}{4N_o}}\right) \tag{5.40}$$

and using (5.34) to further average over the signal set gives

$$P(e) \leq \frac{1}{M} \sum_{k=0}^{M-1} \sum_{j \neq k} Q\left(\sqrt{\frac{\alpha^2 \tilde{d}_{jk}^2}{4N_o}}\right) . \tag{5.41}$$

A further upper bound can be obtained by first computing the **minimum squared Euclidean distance** between any two signal points

$$\tilde{d}_{\min}^2 = \min_{n,m} \|\tilde{\mathbf{s}}_n - \tilde{\mathbf{s}}_m\|^2 . \tag{5.42}$$

Thus the pairwise error probability between $\tilde{\mathbf{s}}_j$ and $\tilde{\mathbf{s}}_k$ is bounded by

$$P(\tilde{\mathbf{s}}_j, \tilde{\mathbf{s}}_k) \leq Q\left(\sqrt{\frac{\alpha^2 \tilde{d}_{\min}^2}{4N_o}}\right) . \tag{5.43}$$

Hence, we can write

$$P(e) \leq (M-1)Q\left(\sqrt{\frac{\alpha^2 \tilde{d}_{\min}^2}{4N_o}}\right) . \tag{5.44}$$

Finally, some other upper bounds can be obtained using the upper bound on the Q-function (see Problem 5.1)

$$Q(x) \leq \frac{1}{2} e^{-x^2/2} \qquad x \geq 0 . \tag{5.45}$$

Combining with the union bound in (5.41) gives

$$P(e) \leq \frac{1}{2M} \sum_{k=0}^{M-1} \sum_{j \neq k} \exp\left\{\frac{\alpha^2 \tilde{d}_{jk}^2}{8N_o}\right\} \tag{5.46}$$

and combining with the upper bound in (5.44) gives the simplest but loosest upper bound

$$P(e) \leq \frac{(M-1)}{2} \exp\left\{\frac{\tilde{d}_{\min}^2}{8N_o}\right\} . \tag{5.47}$$

3.3   LOWER BOUND ON ERROR PROBABILITY

A useful lower bound on the probability of decision error can be obtained by bounding the error probability

$$
P(e|\mathbf{b}_k) \geq \begin{cases} Q\left(\sqrt{\frac{\alpha^2 \tilde{d}_{\min}^2}{4 N_o}}\right) & , \text{ if } \tilde{\mathbf{s}}_k \text{ at least one neighbor at distance } \tilde{d}_{\min} \\ 0 & , \text{ otherwise} \end{cases}
$$

(5.48)

Then

$$
P(e) = \frac{1}{M} \sum_{m=0}^{M-1} P(e|\tilde{\mathbf{s}}_m) \tag{5.49}
$$

$$
\geq \frac{w_{\min}}{M} Q\left(\sqrt{\frac{\alpha^2 \tilde{d}_{\min}^2}{4 N_o}}\right) \tag{5.50}
$$

where $w_{\min}$ is the number of signal vectors having at least one minimum distance neighbor. Certainly $w_{\min} \geq 2$, so that

$$
P(e) \geq \frac{2}{M} Q\left(\sqrt{\frac{\alpha^2 \tilde{d}_{\min}^2}{4 N_o}}\right) . \tag{5.51}
$$

3.4   BIT VERSUS SYMBOL ERROR PROBABILITIES

Thus far our figure of merit has been the probability of decision error or *symbol* error probability, P_M. However, we are very often interested in the *bit* error probability, P_b. In general, this error probability depends on the particular mapping between data bits and data symbols. Since each data symbol corresponds to $\log_2 M$ data bits the bit error probability is bounded as follows:

$$
\frac{P_M}{\log_2 M} \leq P_b \leq P_M . \tag{5.52}
$$

Gray coding:.   For signal constellations such as PSK and QAM, it is possible to map the binary data bits onto the M-ary symbols in such a way that the nearest neighboring symbols (in Euclidean distance) differ in only one bit position. Such a mapping is called a **Gray code.** When the signal-to-noise ratio is high, we find that symbol errors are made onto the nearest neighboring symbols with high probability. In these cases, symbol errors correspond to single bit errors. Hence,

$$
P_b \approx \frac{P_M}{k} . \tag{5.53}
$$

where $k = \log_2 M$.

$$
\begin{array}{ccc}
0\,0\,0\ldots\ldots 0\,0 & \longleftarrow & \tilde{s}_0 \\
0\,0\,0\ldots\ldots 0\,1 & \longleftarrow & \tilde{s}_1 \\
0\,0\,0\ldots\ldots 1\,0 & \longleftarrow & \tilde{s}_2 \\
0\,0\,0\ldots\ldots 1\,1 & \longleftarrow & \tilde{s}_3 \\
\cdots \qquad \cdots \\
1\,1\,1\ldots\ldots 1\,0 & \longleftarrow & \tilde{s}_{M-2} \\
1\,1\,1\ldots\ldots 1\,1 & \longleftarrow & \tilde{s}_{M-1}
\end{array}
$$

$$\text{choose} \atop i\text{th row}$$
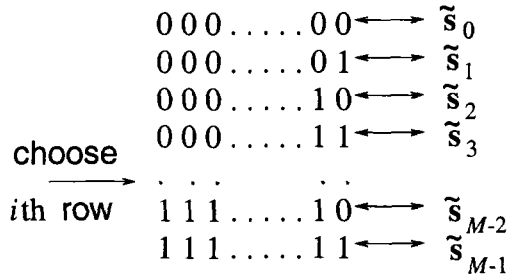
Figure 5.7. Mapping of binary k-tuples onto M-ary symbols.

Equally likely symbol errors:. Consider the case where all symbol errors are equally likely. To compute the probability of bit error, first note that the $M = 2^k$ symbols have a one-to-one mapping onto all possible 2^k binary k-tuples as shown in Fig. 5.7. Now suppose that all zeros k-tuple, or first row, corresponds to the correct symbol. However, the receiver makes an error and chooses ith row (symbol), $i \neq 0$. Since there are 2^{k-1} zeros and ones in each column and a zero corresponds to a correct bit, the probability of a particular bit position being in error is

$$P_b = \frac{2^{k-1}}{2^k - 1} P_M = \frac{M}{2(M-1)} P_M \ . \tag{5.54}$$

4. ERROR PROBABILITY OF PSK

Error probability of binary PSK (BPSK):. From (4.64) and (4.61) with $\theta_o = 0$, the BPSK signal vectors are[2]

$$\tilde{s}_0 = -\tilde{s}_1 = \sqrt{2E_h} \ . \tag{5.55}$$

Since there are only two signal vectors, the error probability is given by (5.36. For BPSK signals, $\tilde{d}_{01} = 2\sqrt{2E_h}$. Also BPSK transmits 1 bit/symbol so the symbol energy is $E_h = E_b$, where E_b is the bit energy. Therefore, the probability of bit error is

$$P_b(\gamma_b) = Q\left(\sqrt{2\gamma_b}\right) \tag{5.56}$$

where γ_b is defined as the *bit energy-to-noise ratio*

$$\gamma_b = \frac{\alpha^2 E_b}{N_o} \ . \tag{5.57}$$

---

[2]When the signal vectors lie in a 1-D complex vector space, we simplify notation by using the scalars $\tilde{s}_i$, $\tilde{n}$, $\tilde{r}$ rather than the vectors $\tilde{\mathbf{s}}_i$, $\tilde{\mathbf{n}}$, and $\tilde{\mathbf{r}}$.

Error probability of quaternary PSK (QPSK):.   From (4.64) and (4.61) with $\theta_o = \pi/4$, the QPSK (or 4-PSK) signal vectors are

$$\tilde{s}_0 = -\tilde{s}_2 = \sqrt{E_h}(1+j) \tag{5.58}$$
$$\tilde{s}_1 = -\tilde{s}_3 = \sqrt{E_h}(-1+j) \tag{5.59}$$

The QPSK signal constellation is shown in Fig. 5.8. The decision boundaries correspond to the real and imaginary axis of the complex vector space. The noise components $\tilde{n}_I$ and $\tilde{n}_Q$ are independent zero-mean Gaussian random variables with variance N_o. With minimum distance decisions, the probability of symbol error is

$$
\begin{aligned}
P_M &= P(e|\tilde{s}_0) \\
&= 1 - P(c) \\
&= 1 - P\left\{\tilde{n}_I > -\alpha\tilde{d}/2, \tilde{n}_Q > -\alpha\tilde{d}/2\right\} \\
&= 1 - P\left\{\tilde{n}_I > -\alpha\tilde{d}/2\right\} P\left\{\tilde{n}_Q > -\alpha\tilde{d}/2\right\} \\
&= 1 - \left[1 - Q\left(\sqrt{\frac{\alpha^2\tilde{d}^2}{4N_o}}\right)\right]^2
\end{aligned}
$$

where, again, α is the channel attenuation. Since $\tilde{d}^2 = 4E_h$, we have

$$P_M = 1 - [1 - Q(\sqrt{\gamma_s})]^2 \tag{5.60}$$

where γ_s is defined as the *symbol energy-to-noise ratio*

$$\gamma_s = \frac{\alpha^2 E_h}{N_o} \ . \tag{5.61}$$

Suppose the data bits are mapped onto the data symbols using a Gray code as shown in Fig. 5.8. Letting P_b denote the probability of bit error, it follows that

$$P(c) = (1 - P_b)^2$$

and

$$P_M = 1 - (1 - P_b)^2 \ .$$

Comparing the above equation with (5.60).

$$P_b = Q(\sqrt{\gamma_s})$$

QPSK transmits 2 bits/symbol so the symbol energy is $E_h = 2E_b$, where E_b is the bit energy. Since $\gamma_s = 2\gamma_b$

$$P_b(\gamma_b) = Q\left(\sqrt{2\gamma_b}\right) \ .$$
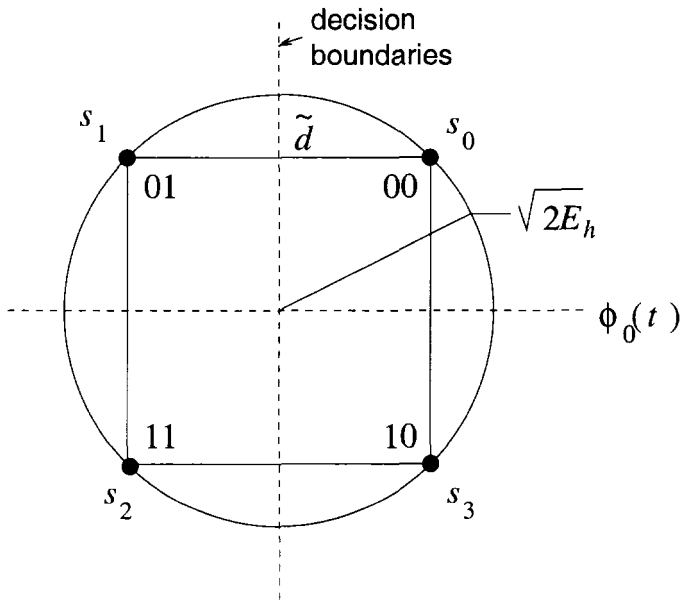
Figure 5.8. Complex signal-space diagram for QPSK.

Notice that the bit error probabilities of BPSK and QPSK are identical. Finally, since OQPSK is identical to QPSK with the exception of the inphase and quadrature branches being offset by $T_b = T/2$ seconds, the error performance of OQPSK is identical to QPSK.

Error probability of M-PSK:. To derive the error probability of M-PSK consider, for example, the 8-PSK signal constellation and associated decision regions shown in Fig. 5.9. Once again data bits are mapped onto data symbols by using a Gray code. Consider (4.46) and suppose that the message vector $\tilde{s}_0 = \sqrt{2E_h}$ is transmitted. The received signal vector is

$$\tilde{r} = \alpha e^{j\phi} \tilde{s}_0 + \tilde{n} \; . \tag{5.62}$$

Since the error probability is invariant to the angle rotation ϕ, we can arbitrarily set $\phi = 0$ so that

$$
\begin{aligned}
\tilde{r} &= \alpha \tilde{s}_0 + \tilde{n} \\
&= \alpha \sqrt{2E_h} + \tilde{n}
\end{aligned}
\tag{5.63}
$$

It follows that $\tilde{r} = \tilde{r}_I + j\tilde{r}_Q$ is a complex Gaussian random variable with pdf

$$p_{\tilde{r}}(\tilde{r}) = \frac{1}{\pi N_o} \exp \left\{ -\frac{1}{N_o} \left| \tilde{r} - \alpha \sqrt{2E_h} \right|^2 \right\} \; . \tag{5.64}$$
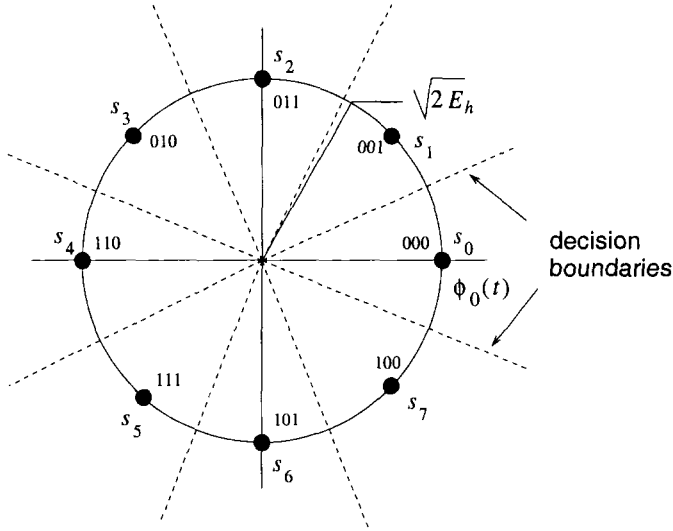
Figure 5.9. Complex signal-space diagram for 8-PSK along and the associated decision regions.

Since $\tilde{s}_0$ was transmitted, the probability of correct symbol reception with minimum distance decisions is the probability that the received angle $\Theta = \text{Tan}^{-1}[\tilde{r}_Q/\tilde{r}_I]$ lies in the interval $[-\pi/8, \pi/8]$.

To find the pdf of the angle Θ, we first define the new random variables

$$R = \sqrt{\tilde{r}_I^2 + \tilde{r}_Q^2}, \qquad \Theta = \text{Tan}^{-1}[\tilde{r}_Q/\tilde{r}_I] \qquad (5.65)$$

such that

$$\tilde{r}_I = R\cos\Theta, \qquad \tilde{r}_Q = R\sin\Theta. \qquad (5.66)$$

Then by using a bivariate transformation of random variables (Appendix A), the joint pdf of R and Θ can be obtained as

$$p_{R,\Theta}(r,\theta) = \frac{r}{\pi N_o} \exp\left\{-\frac{1}{N_o}\left(r^2 - 2\alpha\sqrt{2E_h}\,r\cos\theta + 2\alpha^2 E_h^2\right)\right\}. \qquad (5.67)$$

Since we are interested only in the phase Θ, we obtain the marginal pdf of Θ

$$
\begin{aligned}
p_\Theta(\theta) &= \int_0^\infty p_{R,\Theta}(r,\theta)\,dr \qquad &(5.68)\\
&= \frac{1}{\pi} e^{-2\gamma_s \sin^2\theta} \int_0^\infty x \exp\left\{\left(x - \sqrt{2\gamma_s}\cos\theta\right)^2\right\} dx \qquad &(5.69)
\end{aligned}
$$

where $\gamma_s = \alpha^2 E_h/N_o$ is the received symbol energy-to-noise ratio. The probability of M-ary symbol error, P_M is just the probability that Θ lies outside

the region $[-\pi/M, \pi/M]$. Thus

$$P_M(\gamma_s) = 1 - \int_{-\pi/M}^{\pi/M} p(\theta)d\theta \ . \tag{5.70}$$

A closed form expression for this integral does not exist, except for the cases $M = 2, 4$ which were considered earlier.

Error probability with Rayleigh fading:. When the channel experiences fading, the error probability must be averaged over the fading statistics. If the channel is Rayleigh faded, then α is a Rayleigh random variable. By using a transformation of random variables, $\bullet\gamma_b$ and γ_s have the exponential pdfs

$$p_{\gamma_b}(x) = \frac{1}{\bar{\gamma}_b}e^{-x/\bar{\gamma}_b} \qquad\qquad p_{\gamma_s}(x) = \frac{1}{\bar{\gamma}_x}e^{-x/\bar{\gamma}_x} \qquad\qquad x \geq 0 \tag{5.71}$$

where $\bar{\gamma}_b$ and $\bar{\gamma}_s$ are the average received bit and symbol energy-to-noise ratios, respectively, and $\bar{\gamma}_s = \log_2 M \bar{\gamma}_b$.

For BPSK and QPSK the average probability of bit error is

$$\begin{aligned} P_b &= \int_0^\infty Q(\sqrt{2x})p_{\gamma_b}(x)dx \\ &= \frac{1}{2}\left[1 - \sqrt{\frac{\bar{\gamma}_b}{1+\bar{\gamma}_b}}\right] \\ &\approx \frac{1}{4\bar{\gamma}_b} \qquad \text{for } \bar{\gamma}_b \gg 1. \end{aligned} \tag{5.72}$$

The BPSK and QPSK bit error probability is plotted in Fig. 5.10 for an AWGN channel and a flat Rayleigh fading channel with AWGN. Observe that Rayleigh fading converts an exponential dependency of the bit error probability on the bit energy-to-noise ratio into an inverse linear one. This behavior is typical for any uncoded modulation scheme in Rayleigh fading, and results in a huge loss in performance unless appropriate countermeasures are taken. For M-PSK, the average symbol error probability is

$$P_M = \int_0^\infty P_M(x)p_{\gamma_s}(x)dx \tag{5.73}$$

where $P_M(x)$ is given by (5.70). However, no closed form expression exists. With Gray coding the bit error probability is approximately $P_b = P_M/\log_2 M$.

Differential PSK (DPSK):. The received carrier phase for PSK signals is

$$\theta_k = \frac{2\pi}{M}x_k + \theta_o + \phi \tag{5.74}$$
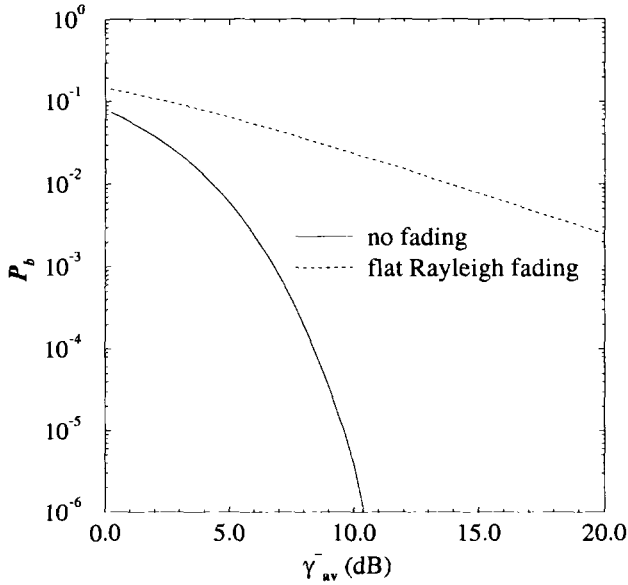
Figure 5.10.   Bit error probability for BPSK and QPSK for a slow flat Rayleigh fading channel with AWGN.

where θ_o is an arbitrary constant phase and ϕ is the random phase due to the channel. The receiver corrects for the phase ϕ by multiplying the received complex envelope by $e^{-j\phi}$ as shown in (5.27). However, in practice this operation is not quite that simple, because the symmetries in the signal constellation create **phase ambiguity.** In particular, we note that any channel induced phase of the form $\phi + 2k\pi/M$, k an integer, will lead to exactly the same set of received carrier phases. While the receiver can use a phased locked loop to recover the received carrier phase, there will remain a phase ambiguity which is a multiple of $2\pi/M$. This phase ambiguity must be resolved if the information is to be recovered correctly.

Differential encoding is one of the most popular methods for resolving phase ambiguity, where information is transmitted in the carrier phase differences between successive baud intervals rather than the absolute carrier phases. Differential encoding of PSK signals is done as follows. The information sequence $\{x_k\}, x_k \in \{0, 1, \dots M - 1\}$ is differentially encoded into a new sequence $\{d_k\}$ according to

$$d_k = x_k \oplus x_{k-1} \tag{5.75}$$

where $\oplus$ denotes modulo-M addition. Then the sequence $\{d_k\}$ is transmitted in the absolute carrier phase according to

$$\theta_k = \frac{2\pi}{M} d_k + \theta_o \ . \tag{5.76}$$

After carrier recovery the received carrier phase is

$$\tilde{\theta}_k = \frac{2\pi}{M} d_k + \theta_o + \frac{2\pi\ell}{M} \tag{5.77}$$

where the additional term $2\pi\ell/M$, ℓ an integer, represents the phase ambiguity. The receiver computes the differential phase

$$\begin{aligned}
\tilde{\theta}_k - \tilde{\theta}_{k-1} \bmod 2\pi &= \frac{2\pi}{M}(d_k - d_{k-1}) \bmod 2\pi \\
&= \frac{2\pi}{M}(d_k \ominus d_{k-1}) \\
&= \frac{2\pi}{M} x_k \tag{5.78}
\end{aligned}$$

where $\ominus$ denotes modulo-M subtraction. Hence, the data sequence $\{x_k\}$ is recovered regardless of the phase ambiguity.

In the presence of AWGN noise, the receiver must form estimates $\hat{\theta}_k$ of the received carrier phases $\tilde{\theta}_k$. However, the noise will cause errors in these estimates and occasionally $\hat{\theta}_k \neq \tilde{\theta}_k$. We note that an incorrect phase estimate $\hat{\theta}_k$ causes the decisions on both x_k and x_{k-1} to be in error, assuming that the phase estimates $\hat{\theta}_{k-1}$ and $\hat{\theta}_{k+1}$ are both correct. Hence, at high signal-to-noise ratios where errors occur infrequently, the bit error probability of DPSK is roughly two times that of PSK.

5. ERROR PROBABILITY OF M-QAM

Error probability of M-PAM:. Consider the Gray coded 8-PAM system signal constellation shown in Fig. 5.11. For the $M - 2$ inner points on the signal constellation, the probability of symbol error is

$$P_i = 2Q\left(\frac{2\alpha^2 E_h}{N_o}\right) \tag{5.79}$$

where the 2 appears in front of the Q function because errors can be made by crossing either of the two decision boundaries. Likewise, for the 2 outer points on the signal constellation the probability of symbol error is

$$P_o = Q\left(\sqrt{\frac{2\alpha^2 E_h}{N_o}}\right) \tag{5.80}$$
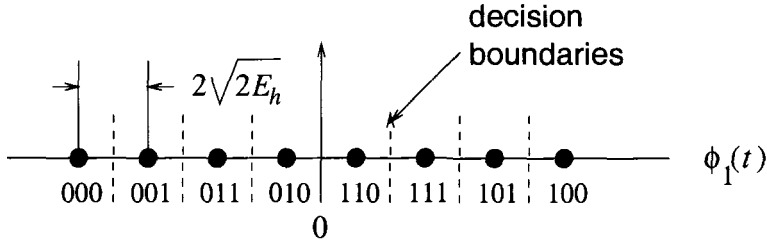
Figure 5.11.   Complex signal-space diagram for 8-PAM.

Hence, the probability of symbol error is

$$
\begin{aligned}
P_M &= \frac{M-2}{M}P_i + \frac{2}{M}P_o \\
&= 2\left(1 - \frac{1}{M}\right)Q\left(\sqrt{\frac{2\alpha^2 E_h}{N_o}}\right).
\end{aligned}
\tag{5.81}
$$

Next we have to relate E_h to the average symbol energy. Since

$$
\tilde{s}_m = \sqrt{2E_h}(2m - 1 - M), \quad m = 1,\ldots,M
\tag{5.82}
$$

the energy in $\tilde{s}_m$ is

$$
E_m = \frac{1}{2}\tilde{s}_m^2 = E_h(2m - 1 - M)^2
\tag{5.83}
$$

The average energy is

$$
\begin{aligned}
E_{\text{av}} &= E_h \frac{1}{M} \sum_{m=1}^{M} (2m - 1 - M)^2 \\
&= E_h \frac{1}{M}\left(4\sum_{m=1}^{M} m^2 - 4(M+1)\sum_{m=1}^{M} m + M(M+1)^2\right)
\end{aligned}
\tag{5.84}
$$

Using the identities

$$
\sum_{k=1}^{n} k = \frac{n(n+1)}{2} \qquad \sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}
\tag{5.85}
$$

and simplifying gives the result

$$
E_{\text{av}} = E_h \frac{M^2 - 1}{3}
\tag{5.86}
$$

Hence from (5.81)

$$P_M(\gamma_s) = 2\left(1 - \frac{1}{M}\right) Q\left(\sqrt{\frac{6}{M^2 - 1}\gamma_s}\right) \tag{5.87}$$

where

$$\gamma_s = \frac{\alpha^2 E_{av}}{N_o} \tag{5.88}$$

is the average symbol energy-to-noise ratio.

Error probability of M-QAM:.   Consider an M-QAM system having a square constellation size $M = 4^m$ for some integer m.   Such an M-QAM system can be viewed as two $\sqrt{M}$-**PAM** systems in quadrature, each allocated one-half the power of the M-QAM system.   For example, the Gray coded 16-QAM system in Fig. 5.12 can be treated as two independent Gray coded 4-PAM systems in quadrature, each operating with half the power of the 16-QAM system.   From (5.87), the symbol errorprobability for each $\sqrt{M}$-**PAM** system is

$$P_{\sqrt{M}} = 2\left(1 - \frac{1}{\sqrt{M}}\right) Q\left(\sqrt{\frac{6}{M - 1}\frac{\gamma_s}{2}}\right) \tag{5.89}$$

where γ_s is the average symbol energy-to-noise ratio of the M-QAM system. Finally, the probability of correct symbol reception in the M-QAM system is

$$P_c(e) = (1 - P_{\sqrt{M}})^2 \tag{5.90}$$

and the probability of symbol error is

$$P_M(\gamma_s) = 1 - (1 - P_{\sqrt{M}})^2 \ . \tag{5.91}$$

 For other types of M-QAM constellations, such as those in Figs. 4.7 and 4.8, the error probability can be obtained by defining convex decision regions and using the approach suggested in Section 3..

Error probability with Rayleigh fading:.   If the channel is Rayleigh faded, then γ_s has the exponential pdf in (5.71). It follows that the average symbol error probability is

$$P_M = \int_0^\infty P_M(x) p_{\gamma_s}(x) dx \ . \tag{5.92}$$

Fig. 5.13 plots the (approximate) bit errorprobability $P_b \approx P_M/\log_2 M$ against the average received *bit* energy-to-noise ratio, $\bar{\gamma}_b = \bar{\gamma}_s/\log_2 M$, for several values of M. Notice that the $\bar{\gamma}_b$ required to achieve a given bit error probability increases with the alphabet size M. However, the bandwidth efficiency also increases with M, since there are $\log_2 M$ bits/symbol.
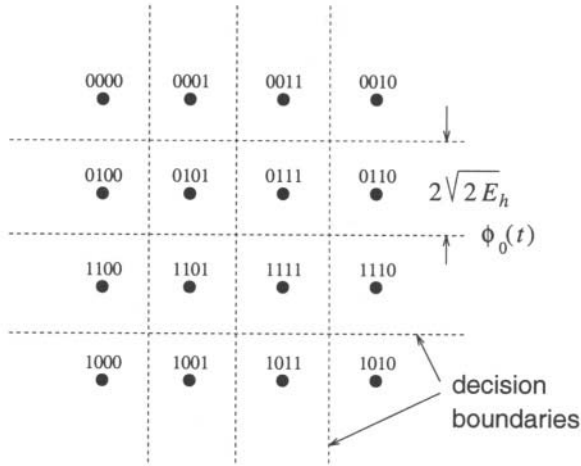
Figure 5.12.   Complex signal-space diagram for 16-QAM constellation.
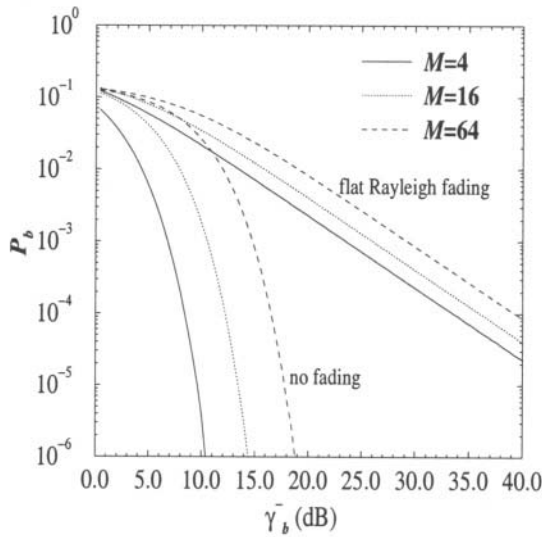
Figure 5.13.   Bit error probability for M-QAM on an AWGN channel and a Rayleigh fading channel with AWGN.

6.   ERROR PROBABILITY OF ORTHOGONAL SIGNALS

Orthogonal signals:.   Consider the M-ary orthogonal signal set

$$\tilde{\mathbf{s}}_i = \sqrt{2E_h}\mathbf{e}_m, m = 0, \ldots, M - 1 \ ,$$

where $\mathbf{e}_m$ is a length-M vector with a "1" in the mth coordinate. If the signal $\tilde{\mathbf{s}}_0$ is transmitted, then the received signal vector is

$$\tilde{\mathbf{r}} = (g\sqrt{2E_h} + \tilde{n}_0, \tilde{n}_1, \ldots, \tilde{n}_{M-1}) \tag{5.93}$$

where the n_i are independent zero mean complex Gaussian random variables with variance N_o. The ML receiver computes the decision variables

$$\mu_3(\tilde{\mathbf{s}}_m) = \mathrm{Re}(\tilde{\mathbf{r}}, g\mathbf{s}_m) \tag{5.94}$$

and chooses the signal with the largest $\mu(\mathbf{s}_m)$. We have

$$\begin{aligned}
\mu(\mathbf{s}_0) &= 2\alpha^2 E_h + \tilde{n}_{I,0}\alpha\sqrt{2E_h} & (5.95)\\
\mu(\mathbf{s}_m) &= \tilde{n}_{I,m}\alpha\sqrt{2E_h}, \quad m = 2, \ldots, M-1 & (5.96)
\end{aligned}$$

where we have ignored the phase rotation on the noise samples. The $\mu(\mathbf{s}_m)$ are independent Gaussian random variables with variance $2\alpha^2 E N_o$; the mean of $\mu(\mathbf{s}_0)$ is $2\alpha_h^E$ while the $\mu(\mathbf{s}_m), m \neq 0$, have zero mean. The probability of correct decision conditioned on $\mu(\mathbf{s}_1) = x$ is the probability that all the $\mu(\mathbf{s}_m), m \neq 0$ are less than x. This is just

$$P(c|\mu(\mathbf{s}_1) = x) = \left[\Phi\left(\frac{x}{\sqrt{2\alpha^2 E_h N_o}}\right)\right]^{M-1} \tag{5.97}$$

Hence,

$$\begin{aligned}
P(c) &= \int_{-\infty}^{\infty} \left[\Phi\left(\frac{x}{\sqrt{2\alpha^2 E_h N_o}}\right)\right]^{M-1}\\
&\quad \times \frac{1}{\sqrt{4\pi\alpha^2 E_h N_o}}\exp\left\{-\frac{(x - 2\alpha^2 E_h)^2}{4\alpha^2 E_h N_o}\right\} dx
\end{aligned} \tag{5.98}$$

Now let $y = (x - 2\alpha^2 E_h)/\sqrt{2\alpha^2 E_h N_o}$. Then

$$P(c) = \int_{-\infty}^{\infty} \left[\Phi\left(y + \sqrt{2\gamma_s}\right)\right]^{M-1} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dx \tag{5.99}$$

where

$$\gamma_s = \frac{\alpha^2 E_h}{N_o} . \tag{5.100}$$

Finally, the probability of symbol error is

$$P_M = 1 - P(c) . \tag{5.101}$$

An alternate expression for the error probability can be derived by first condi-
tioning on the event that one of the $M - 1$ decision variables $\mu(\mathbf{s}_m), m \neq 0$ is
the largest. This gives

$$P_M = (M - 1) \int_{-\infty}^{\infty} \Phi\left(\frac{x - 2\alpha^2 E_h}{\sqrt{2\alpha^2 E_h N_o}}\right) \left[\Phi\left(\frac{x}{\sqrt{2\alpha^2 E_h N_o}}\right)\right]^{M-2}$$

$$\times \frac{1}{\sqrt{4\pi\alpha^2 E_h N_o}} \exp\left\{-\frac{x^2}{4\alpha^2 E_h N_o}\right\} dx \quad . \tag{5.102}$$

Now let $y = x/\sqrt{2\alpha^2 E_h N_o}$. Then

$$P_M = (M - 1) \int_{-\infty}^{\infty} \Phi\left(y - \sqrt{2\gamma_s}\right) [\Phi(y)]^{M-2} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{y^2}{2}\right\} dx \quad . \tag{5.103}$$

For orthogonal signals $\gamma_s = \gamma_b \log_2 M$, and the bit error probability is given
by (5.54). If the channel is Rayleigh faded, then γ_b has the exponential pdf in
(5.71), and the average bit error probability is

$$P_b = \int_0^{\infty} P_b(x) p_{\gamma_b}(x) dx \quad . \tag{5.104}$$

Biorthogonal signals:.   Consider the biorthogonal signal set

$$\tilde{\mathbf{s}}_i = \begin{cases} \sqrt{2E_h}\mathbf{e}_i , & i = 0, \ldots, M/2 - 1 \\ -\tilde{\mathbf{s}}_{i-M/2} , & i = M/2, \ldots, M - 1 \end{cases} \tag{5.105}$$

Now suppose that $\mathbf{s}_0$ is transmitted. The receiver computes the $M/2$ decision
variables
$$\mu(\mathbf{s}_m) = \text{Re}(\tilde{\mathbf{r}}, g\mathbf{s}_m) , \quad m = 0, \ldots, M/2 - 1 \tag{5.106}$$

and chooses the one having the largest magnitude. The sign of $\mu(\mathbf{s}_m)$ is
used to decide whether $\mathbf{s}_m$ or $\mathbf{s}_{M/2+m} = -\mathbf{s}_m$ was sent. As before, the
$\mu(\mathbf{s}_m)$ are independent Gaussian random variables with variance $2\alpha^2 E N_o$;
the mean of $\mu(\mathbf{s}_0)$ is $2\alpha_h^E$ while the $\mu(\mathbf{s}_m), m = 1, \ldots, M/2 - 1$, have zero
mean. The probability of correct decision is the probability that $\mu(\mathbf{s}_0) > 0$ and
$|\mu(\mathbf{s}_m)| < \mu(\mathbf{s}_0), m = 1, \ldots, M/2 - 1$. Condition on $\mu(\mathbf{s}_1) = x, x > 0$, we
have

$$p\left(|\mu(\mathbf{s}_m)| < x\right) = \Phi\left(\frac{x}{\sqrt{2\alpha^2 E_h N_o}}\right) - \Phi\left(-\frac{x}{\sqrt{2\alpha^2 E_h N_o}}\right) \quad . \tag{5.107}$$

Hence,

$$P(c) = \int_0^\infty \left[\Phi\left(\frac{x}{\sqrt{2\alpha^2 E_h N_o}} \right) - \Phi\left(-\frac{x}{\sqrt{2\alpha^2 E_h N_o}} \right) \right]^{M/2-1}$$
$$\times \frac{1}{\sqrt{4\pi\alpha^2 E_h N_o}} \exp\left\{ -\frac{(x - 2\alpha^2 E_h)^2}{4\alpha^2 E_h N_o} \right\} dx \qquad (5.108)$$

Now let $y = (x - 2\alpha^2 E_h)/\sqrt{2\alpha^2 E_h N_o}$. Then

$$P(c) = \int_{-\sqrt{2\gamma_s}}^\infty \left[\Phi\left(y + \sqrt{2\gamma_s} \right) - \Phi\left(-y - \sqrt{2\gamma_s} \right) \right]^{M/2-1} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \, dy \ .$$
$$(5.109)$$

Finally, $P_M = 1 - P(c)$. For biorthogonal signals $\gamma_s = \gamma_b \log_2 M$, but the bit error probability is not given by (5.54).

7.   ERROR PROBABILITY OF OFDM

For an AWGN channel, the error probability of OFDM can be calculated by taking advantage of the property that the OFDM sub-carriers are orthogonal. The optimum receiver for OFDM on an AWGN channel consists of a bank of correlator detectors, one for each sub-carrier. Since the sub-carriers are orthogonal, there is no cross talk between them, and the symbol error probability for each of the subcarriers can be obtained independently of the others.

A key advantage of OFDM is that the receiver can be implemented by using fast Fourier transform (FFT) algorithm, as discussed in Chapter 4.6. In the following discussion we assume that the guard interval is long enough to isolate the OFDM blocks. Hence, we suppress the block index. Following the development in Chapter 4.6, suppose that the discrete-time sequence $\mathbf{X}^g = \{X_n^g\}_{n=0}^{N+G-1}$ is transmitted over a flat fading channel with complex gain g. The received sequence is $\mathbf{R}^g = \{R_n^g\}_{n=0}^{N+G-1}$, where

$$R_n^g = gX_n^g + \tilde{n}_n \qquad (5.110)$$

$g = \alpha^{j\phi}$ is the channel gain, and the $\tilde{n}_n$ are the noise samples. Suppose that the $\tilde{n}_n$ are obtained by passing the received noise waveform through an ideal anti aliasing filter having a bandwidth $1/T_s$ followed by a sampler. In this case, the $\tilde{n}_n$ are independent zero complex Gaussian random variables with variance $\sigma^2 = \frac{1}{2}E[|\tilde{n}_n|^2] = N_o/T_s$.

The receiver first removes the guard interval according to

$$R_n = R_{G+(n-G)_N}^g, \quad 0 \le n \le N - 1 \qquad (5.111)$$

where $(n)_N$ is the residue of n modulo N. Demodulation is then performed by computing the FFT on the block R to yield the vector $\mathbf{Z} = \{Z_i\}_{i=0}^{N-1}$ of N
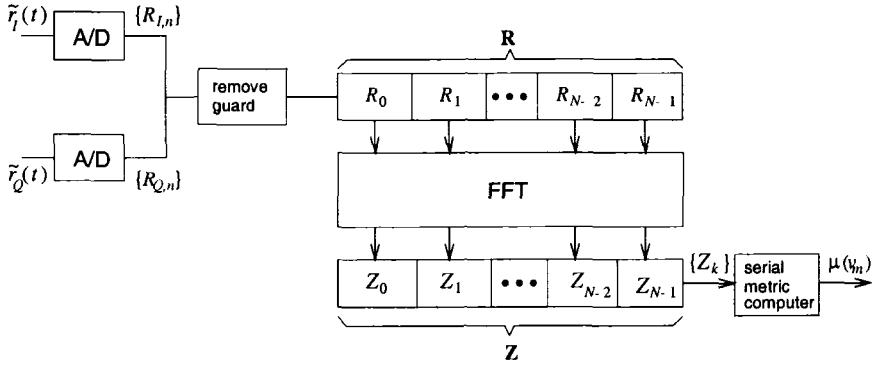
Figure 5.14. Block diagram of OFDM receiver implemented by using a DFT or an FFT.

decision variables

$$Z_i = \frac{1}{N}\sum_{n=0}^{N-1} R_n e^{-\frac{j2\pi in}{N}}$$

$$= gAx_i + \nu_i , \quad i = 0, \ldots, N-1 . \tag{5.112}$$

where $A = \sqrt{2E_h/T}$ and the noise terms are given by

$$\nu_i = \frac{1}{N}\sum_{n=0}^{N-1} \tilde{n}_n e^{-\frac{j2\pi in}{N}} . \tag{5.113}$$

A block diagram of an OFDM receiver is shown in Fig. 5.14.

It can be shown that the ν_i are zero mean complex Gaussian random variables with covariance

$$\mu_{n,m} = \frac{1}{2}\mathrm{E}[\nu_n\nu_m^*] = \frac{N_o}{NT_s}\delta_{nm} . \tag{5.114}$$

Therefore, the Z_i are independent Gaussian random variables having mean $g\sqrt{2E_h/T}x_i$ and variance N_o/NT_s. To be consistent with our earlier results for PSK and QAM signals, we can multiply the Z_i for convenience by the factor $\sqrt{NT_s} = \sqrt{T}$ so they have variance N_o. Such scaling gives

$$\tilde{Z}_i = g\sqrt{2E_h}x_i + \tilde{\nu}_i \tag{5.115}$$

where $\tilde{\nu}_i$ has variance N_o. Notice that $\sqrt{2E_h}x_i = \tilde{s}_i$ is equal to the complex signal vector that is transmitted on the ith sub-carrier. For *each* of the $\tilde{Z}_i$, the receiver decides in favor of the *signal vector* $\tilde{s}_{i,m}$ that minimizes the squared Euclidean distance

$$\mu(\tilde{s}_{i,m}) = \|\tilde{Z}_i - g\tilde{s}_{i,m}\|^2 . \tag{5.116}$$

Thus for each OFDM block N symbol decisions must be made, one for each of the N sub-carriers. This can be done in either a serial fashion as in Fig. 5.14, or a parallel fashion.

Finally, it is clear that the probability of symbol error is identical to that achieved with independent modulation on each of the sub-carriers. This is expected, because the sub-carriers are orthogonal in time.

Interchannel Interference (ICI):. Perhaps a more interesting issue is the effect of Doppler on the OFDM receiver performance. Although our analysis will be undertaken for slow flat fading channels, a similar analysis will apply provided that the guard interval is longer than the length of the channel impulse response. We will show that variations in the complex channel gain $\{g_k\}_{k=0}^{N-1}$ over the duration of an OFDM block causes **interchannel interference** (ICI) due to a loss of subchannel orthogonality. The ICI has an effect similar to AWGN.

To isolate the Doppler effects, AWGN is ignored. The received discrete-time sequence after removal of the guard interval is

$$R_n = g_{G+(n-G)_N} X_n \ . \tag{5.117}$$

The vector **Z** at the output of the FFT demodulator circuit is

$$Z_i = \sqrt{2E_h/T} \sum_{n=0}^{N-1} x_n H(n-i) \tag{5.118}$$

where

$$H(n-i) = \frac{1}{N} \sum_{k=0}^{N-1} g_{G+(k-G)_N} e^{\frac{j2\pi}{N}(n-i)k} \ , \quad 0 \le i \le N-1 \ . \tag{5.119}$$

To highlight the effect of channel time variations, (5.118) can be rewritten as

$$Z_i = \sqrt{2E_h/T} H(0) x_i + c_i \tag{5.120}$$

where

$$c_i = \sqrt{2E_h/T} \sum_{\substack{n=1 \\ n \neq i}}^{N-1} x_n H(n-i) \ . \tag{5.121}$$

Note that $H(0)$ is a multiplicative noise term, while c_i is an additive noise term due to ICI. If the channel is time-invariant, then $g_k = g$ and $Z_i = g\sqrt{2E_h/T}x_i$ as before. But for a time-varying channel Z_i is a function of all the data symbols within a block and, hence, interchannel interference (ICI) is introduced.

For N sufficiently large in (5.121), the central limit theorem can be invoked and the ICI treated as a Gaussian random variable. Since x_n and $H(n-i)$ are independent random variables and $\mathrm{E}[x_n] = 0$, it follows that $\mathrm{E}[c_l] = 0$. The variance of c_i is computed by evaluating the autocorrelation function at

lag zero. Since $2E_h \cdot \frac{1}{2}\mathrm{E}[x_n x_m^*] = E_{\mathrm{av}}\delta_{nm}$, where E_{av} is the average symbol energy, the autocorrelation of c_l is

$$\phi_{cc}(r) = \frac{1}{2}\mathrm{E}[c_i c_{i+r}^*] = \frac{E_{\mathrm{av}}}{T}\sum_{n\neq i,i+r}\mathrm{E}[H(n-i)H^*(n-i-r)] \ . \quad (5.122)$$

If we further assume the normalization $\mathrm{E}[|g_k|^2] = 1$ and 2-D isotropic scattering with an isotropic receiver antenna (see Chapter 2), then the autocorrelation becomes

$$\phi_{cc}(r) = \frac{E_{\mathrm{av}}}{T}\delta_r - \frac{E_{\mathrm{av}}}{TN^2}\sum_{k=0}^{N-1}\sum_{k'=0}^{N-1}J_0(2\pi f_m T_s(k-k'))$$
$$\cdot \left[\exp\left(j\frac{2\pi k'r}{N}\right) + (1-\delta_r)\exp\left(j\frac{2\pi kr}{N}\right)\right] \quad (5.123)$$

where f_m is the maximum Doppler frequency. Note that the autocorrelation is not influenced by the positioning of the guard interval, due to the symmetry of the summations.

For symbol-by-symbol detection, it is sufficient to examine the variance of the ICI term

$$\phi_{cc}(0) = \frac{E_{\mathrm{av}}}{T} - \frac{E_{\mathrm{av}}}{TN^2}\left\{N + 2\sum_{i=1}^{N-1}(N-i)J_0(2\pi f_m T_s i)\right\}, \quad (5.124)$$

where the fact that $J_0(\cdot)$ is an even function has been used. Note that variance of the ICI terms is only a function of E_{av}, N, T_s, and f_m, but is otherwise independent of the signal constellation. Fig. 5.15 plots the signal-to-interference ratio

$$\mathrm{SIR} = \frac{E_{\mathrm{av}}/T}{\phi_{cc}(0)} \quad (5.125)$$

as a function of $f_m T_s$ for several values of N.

Suppose that the data symbols $x_{k,n}$ are chosen from a 16-QAM alphabet. From Section 5., the symbol error probability for 16-QAM is

$$P_M = 3Q\left(\sqrt{\frac{1}{5}\gamma_s}\right)\left[1 - \frac{3}{4}Q\left(\sqrt{\frac{1}{5}\gamma_s}\right)\right] \quad (5.126)$$

where γ_s is the average received symbol energy-to-noise ratio. With Rayleigh fading, the symbol error probability is obtained by averaging (5.126) over the pdf in (5.71). Assuming validity of the Gaussian approximation for the ICI, the error floor due to ICI can be obtained by substituting the SIR in (5.125) for $\bar{\gamma}_s$. The results are shown in Fig. 5.16. Simulation results are also shown in Fig. 5.16 corroborating the Gaussian approximation for the ICI. Fig. 5.17
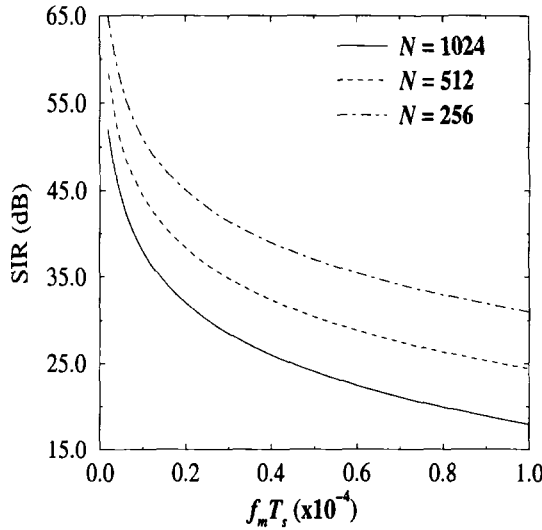
Figure 5.15. Signal-to-interference ratio due to ICI.

shows the bit error rate performance of OFDM with $N = 512$ subcarriers, a 16-QAM signal constellation, and a 20 Mbps bit rate for various Doppler frequencies. At low $\bar{\gamma}_b$, additive noise dominates the performance so that the extra noise due to ICI has little effect. However, at large $\bar{\gamma}_b$ ICI dominates the performance and causes an error floor. Further measures are needed to improve the performance.

8. ERROR PROBABILITY OF MSK

MSK signals can be recovered using a variety of techniques. One method suggested by De Buda uses the fact that MSK is equivalent to OQPSK with a half-sinusoid shaping function. From (4.110), the MSK complex envelope is

$$\tilde{s}(t) = A \sum_n \left(x_{I,n} h_a(t - 2nT) + j x_{Q,n} h_a(t - 2nT - T) \right) \qquad (5.127)$$

where

$$h_a(t) = \cos\left(\frac{\pi t}{2T}\right) u_{2T}(t + T) . \qquad (5.128)$$

The received complex envelope is

$$\tilde{r}(t) = g\tilde{s}(t) + \tilde{n}(t) \qquad (5.129)$$

where $g = \alpha e^{j\phi}$. The receiver first removes the effect of the phase rotation by
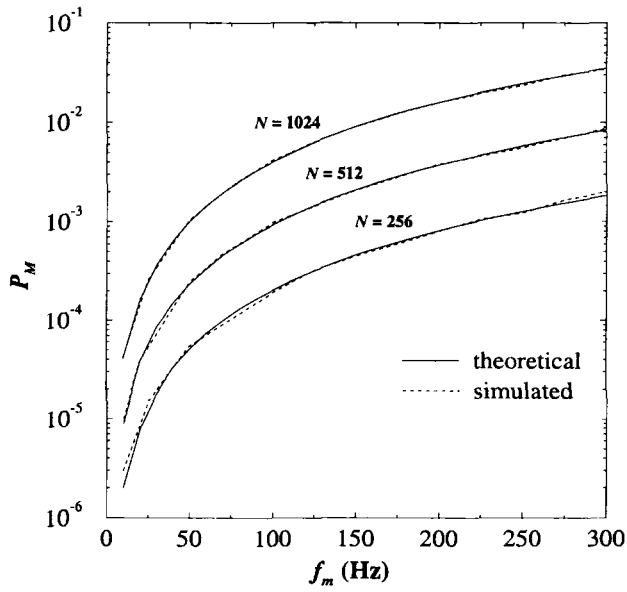
Figure 5.16. Error floor due to ICI with 16-QAM.
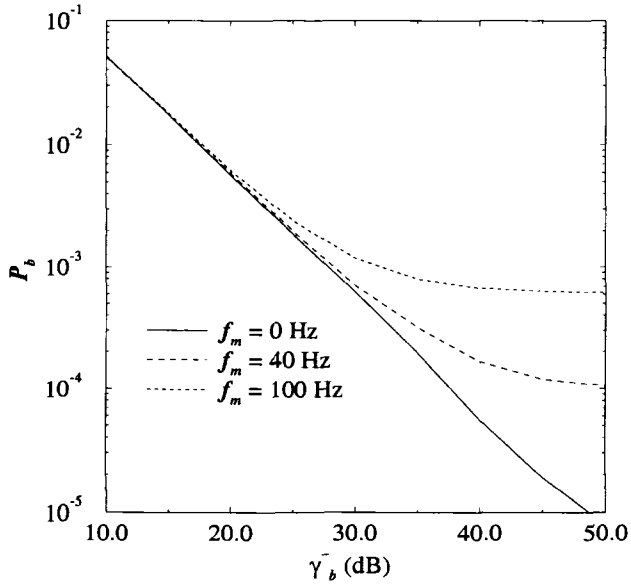
Figure 5.17. Bit error probability for 16-QAM OFDM on a Rayleigh fading channel with various Doppler frequencies.
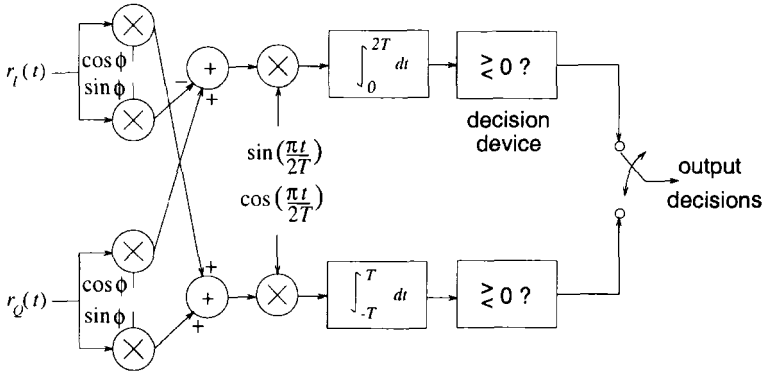
Figure 5.18. Coherent detector for MSK signals.

$$e^{-j\phi}\tilde{r}(t) = \tilde{r}_I(t)\cos\phi + \tilde{r}_Q(t)\sin\phi + j\left[\tilde{r}_Q(t)\cos\phi - \tilde{r}_I(t)\sin\phi\right]$$
$$= \alpha\tilde{s}_I(t) + \tilde{n}_I(t) + j\left[\alpha\tilde{s}_Q(t) + \tilde{n}_Q(t)\right] \qquad (5.130)$$

where we have ignored the effect of the phase rotation on the noise $\tilde{n}(t)$ because it is circularly symmetric. Detection then proceeds by processing the real and imaginary parts of $e^{-j\phi}\tilde{r}(t)$ as separate binary PAM streams. The resulting MSK detector is shown in Fig. 5.18. Note that the source symbols on the in-phase and quadrature carrier components must be detected over intervals of length $2T$, the duration of the amplitude shaping pulse $h_a(t)$, and bit decisions are made every T seconds. Since bit error performance of OQPSK and QPSK (and BPSK) are identical, and MSK can be viewed as a form of OQPSK, it follows that MSK has the same bit error performance as QPSK or BPSK.

9. DIFFERENTIAL DETECTION

Differentially encoded PSK (DPSK) can also be detected by using differentially coherent detection, where the receiver estimates the change in the phase of the received carrier between two successive signaling intervals. Since the differential carrier phase between baud intervals is precisely what contains the data, the basic mechanism for differential detection is obvious. If the carrier phase changes slowly with respect to the baud period, then the phase difference between waveforms received in two successive signaling intervals will be independent of the absolute carrier phase. However for fading channels, the carrier phase can change over two successive baud intervals. This leads to an error floor that increases with the Doppler frequency.

Binary DPSK:. Consider binary DPSK. Let θ_n denote the absolute carrier phase for the nth symbol, and $\Delta\theta_n = \theta_n - \theta_{n-1}$ denote the differential carrier

phase, where

$$\Delta\theta_n = \begin{cases} 0 & , & x_n = +1 \\ \pi & , & x_n = -1 \end{cases} \tag{5.131}$$

The complex envelope of the transmitted signal is

$$\tilde{s}(t) = A \sum_n h_a(t - nT)e^{j\theta_n} \tag{5.132}$$

and the complex envelope of the received signal is

$$\tilde{r}(t) = \alpha e^{j\phi} A \sum_n h_a(t - nT)e^{j\theta_n} + \tilde{n}(t) \tag{5.133}$$

where $g = \alpha e^{j\phi}$ is the complex channel gain.

A block diagram of a differentially coherent receiver for DPSK is shown in Fig. 5.19. During the time interval $[nT, (n+1)T]$, the values of X_n, X_{nd}, Y_n and Y_{nd} in Fig. 5.19 are

$$\begin{aligned} X_n &= 2\alpha E_h \cos(\theta_n + \phi) + \tilde{n}_I \\ X_{nd} &= 2\alpha E_h \cos(\theta_{n-1} + \phi) + \tilde{n}_{I,d} \\ Y_n &= 2\alpha E_h \sin(\theta_n + \phi) + \tilde{n}_Q \\ Y_{nd} &= 2\alpha E_h \sin(\theta_{n-1} + \phi) + \tilde{n}_{Q,d} \end{aligned} \tag{5.134}$$

where

$$E_h = \frac{A^2}{2} \int_0^T h_a^2(t)dt \tag{5.135}$$

is the bit energy, and the noise terms are

$$\begin{aligned} \tilde{n}_I &= A \int_{nT}^{(n+1)T} \tilde{n}_I(t)h_a(t)dt \\ \tilde{n}_{I,d} &= A \int_{(n-1)T}^{nT} \tilde{n}_I(t)h_a(t)dt \\ \tilde{n}_Q &= A \int_{nT}^{(n+1)T} \tilde{n}_Q(t)h_a(t)dt \\ \tilde{n}_{Q,d} &= A \int_{(n-1)T}^{nT} \tilde{n}_Q(t)h_a(t)dt \end{aligned} \tag{5.136}$$

Note that $\tilde{n}_I$, $\tilde{n}_{I,d}$, $\tilde{n}_Q$, and $\tilde{n}_{Q,d}$ are independent Gaussian random variables with variance $2E_h N_o$.

In the absence of noise, it is easy to verify that $U_n = 4\alpha^2 E_h^2 x_n$. To determine the pdf of the decision variable U_n, it is convenient to express U_n as

$$U_n = \text{Re}\{Z_n Z_{nd}^*\} = \frac{1}{2}(Z_n Z_{nd}^* + Z_n^* Z_{nd}) \tag{5.137}$$
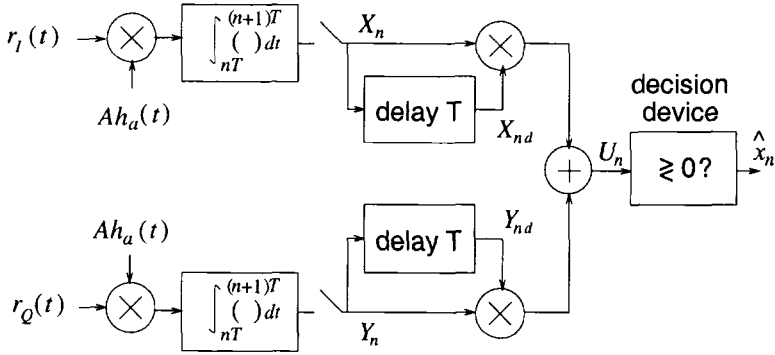
Figure 5.19. Differentially coherent receiver for DPSK signals.

where

$$Z_n = X_n + jY_n \qquad (5.138)$$

$$Z_{nd} = X_{nd} + jY_{nd} . \qquad (5.139)$$

It can be shown by using characteristic functions that $U_n = W_n - Y_n$, where W and Y are non-central and central chi-square random variables with densities [309]

$$f_{W_n}(w) = \frac{1}{2E_h N_o} e^{-\frac{(wx_n + 4\alpha^2 E_h^2)}{2E_h N_o}} I_0 \left(\frac{2\sqrt{wx_n \alpha^2 E_h^2}}{E_h N_o} \right), \quad wx_n \geq 0$$

$$(5.140)$$

$$f_{Y_n}(y) = \frac{1}{2E_h N_o} e^{-\frac{yx_n}{2E_h N_o}}, \quad yx_n \geq 0 . \qquad (5.141)$$

Defining $V_n = W_n$, the pdf of U_n is

$$f_{U_n}(u) = \int_{R_{U_n V_n}} f_{W_n}(v) f_{Y_n}(v - u) dv$$

$$= \begin{cases} \frac{1}{4E_h N_o} \exp \left\{ \frac{x_n u - 2\alpha^2 E_h^2}{2E_h N_o} \right\}, & -\infty < x_n u < 0 \\ \\ \frac{1}{4E_h N_o} \exp \left\{ \frac{x_n u - 2\alpha^2 E_h^2}{2E_h N_o} \right\} Q \left(\sqrt{\frac{2\alpha^2 E_h}{N_o}}, \sqrt{\frac{2x_n u}{E_h N_o}} \right), & \\ & 0 < x_n u < \infty \end{cases}$$

$$(5.142)$$

where $Q(a,b)$ is the Marcum Q function, defined by

$$Q(a,b) = 1 - \int_0^b ze^{-\frac{z^2 + a^2}{2}} I_0(za) dz \qquad (5.143)$$

From (5.142), the bit error probability of DPSK with differential detection is

$$P_b = \int_0^\infty \frac{1}{4E_h N_o} \exp\left\{ -\frac{u + 2\alpha^2 E_h^2}{2E_h N_o} \right\} du = \frac{1}{2} e^{-\gamma_b} \qquad (5.144)$$

where $\gamma_b = \alpha^2 E_h/N_o$ is the received bit energy-to-noise ratio. For a slow Rayleigh fading channel, α is Rayleigh distributed so the received bit energy-to-noise ratio has the exponential pdf in (5.71). It follows that the bit error probability with slow Rayleigh fading is

$$P_b = \frac{1}{2(1 + \bar\gamma_b)} \ . \qquad (5.145)$$

9.1   DIFFERENTIAL DETECTION OF $\pi/4$-DQPSK

The above results can be extended to differential detection of $\pi/4$-DQPSK. Once again the complex envelopes of the transmitted and received signals are given by (5.132) and (5.133), respectively. However, with $\pi/4$-DQPSK, $\Delta\theta_n = \pi x_n/4$ where $x_n \in \{\pm 1, \pm 3\}$. A block diagram of the differentially coherent receiver for $\pi/4$-DQPSK is shown in Fig. 5.20. The values of X_n, X_{nd}, Y_n and Y_{nd} are again given by (5.134). The detector outputs are

$$U_n = \text{Re}\{Z_n Z_{nd}^*\} = \frac{1}{2}(Z_n Z_{nd}^* + Z_n^* Z_{nd}) \qquad (5.146)$$

$$V_n = \text{Im}\{Z_n Z_{nd}^*\} = \frac{1}{j2}(Z_n Z_{nd}^* - Z_n^* Z_{nd}) \qquad (5.147)$$

where $.Z_n$ and Z_{nd} are defined in (5.138) and (5.139), respectively. In the absence of noise, it can be verified that the detector outputs are

$$\begin{aligned}
U_n &= -a, \ V_n = -a, \ x_n = -3 \\
U_n &= a, \ V_n = -a, \ x_n = -1 \\
U_n &= a, \ V_n = a, \ x_n = +1 \\
U_n &= -a, \ V_n = a, \ x_n = +3
\end{aligned} \qquad (5.148)$$

where $a = 2\sqrt{2}\alpha^2 E_h^2$. The bit error probability for $\pi/4$-DQPSK with Gray coding is quite complicated to derive, but can be expressed in terms of well known functions [270]

$$P_b = Q(a, b) - \frac{1}{2} I_0(ab) \exp\left\{ -\frac{1}{2}(a^2 + b^2) \right\} \qquad (5.149)$$

where $I_0(x)$ is the zero-order modified Bessel function of the first kind defined by

$$I_0(x) = \frac{1}{2\pi} \int_0^{2\pi} e^{-x \cos\theta} d\theta \qquad (5.150)$$
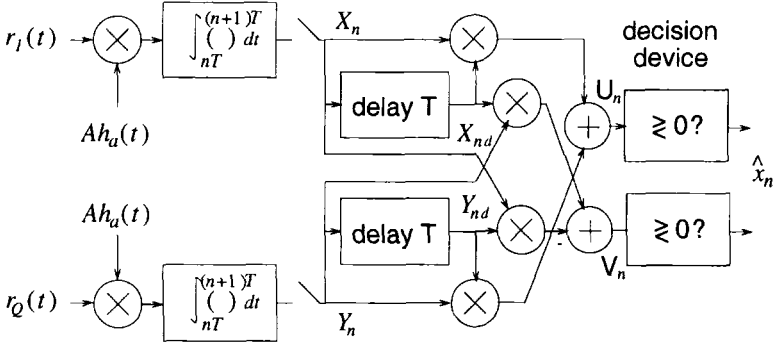
Figure 5.20. Differentially coherent receiver for $\pi/4$-QPSK signals.

$$a = \sqrt{2\gamma_b \left(1 - \frac{1}{\sqrt{2}}\right)} \qquad (5.151)$$

$$b = \sqrt{2\gamma_b \left(1 + \frac{1}{\sqrt{2}}\right)} \qquad (5.152)$$

and γ_b is the bit energy-to-noise ratio. Once again, if the channel is faded, then the bit error probability can be obtained by averaging over the fade distribution.

10.   NON-COHERENT DETECTION

If information is transmitted in the amplitude or frequency of a waveform, then a **non-coherent** receiver can be used. Non-coherent receivers make no attempt to determine the carrier phase. Non-coherent receivers are easier to implement than coherent receivers. They typically allow the receiver to be implemented with less expensive components that may also consume less power. Non-coherent receivers trade implementation complexity for transmitter power and bandwidth. Typical applications for non-coherent receivers include one-way paging, where the receiver must be inexpensive and operate at a low power budget. With one-way paging transmit functions are not required in the terminal equipment.

Suppose that one of M complex low-pass waveforms, $\tilde{s}_m(t), m = 0, \ldots, M-1$ is transmitted, say $\tilde{s}_i(t)$, on a flat fading channel with AWGN. The received complex envelope is

$$\tilde{r}(t) = g\tilde{s}_i(t) + \tilde{n}(t) \qquad (5.153)$$

where $g = \alpha e^{j\phi}$ is the channel gain that includes the random phase ϕ, and $\tilde{n}(t)$ is the AWGN.

By projecting $\tilde{r}(t)$ onto the basis functions obtained through the Gram-Schimdt orthonormalization procedure on the signal set $\{\tilde{s}_m(t)\}_{m=0}^{M-1}$, we obtain

the received vector

$$\tilde{\mathbf{r}} = g\tilde{\mathbf{s}}_i + \tilde{\mathbf{n}} \ . \tag{5.154}$$

The joint pdf of $\tilde{\mathbf{n}}$ is given in (5.10). The maximum likelihood non-coherent detector does not use the random phase ϕ in the decision process, and chooses the message vector $\tilde{\mathbf{s}}_m$ to maximize the joint conditional pdf $p(\tilde{\mathbf{r}}|\alpha\tilde{\mathbf{s}}_m)$:

$$\text{choose } \tilde{\mathbf{s}}_m \text{ if } \quad p(\tilde{\mathbf{r}}|\alpha\tilde{\mathbf{s}}_m) \geq p(\tilde{\mathbf{r}}|\alpha\tilde{\mathbf{s}}_{\hat{m}}) \ \forall \hat{m} \neq m \ . \tag{5.155}$$

Letting $p(\phi)$ denote the pdf of $<\phi$, we have

$$p(\tilde{\mathbf{r}}|\alpha\tilde{\mathbf{s}}_m) = \mathrm{E}_\phi[p(\tilde{\mathbf{r}}|\alpha\tilde{\mathbf{s}}_m, \phi)] = \int_0^{2\pi} p(\tilde{\mathbf{r}}|\alpha\tilde{\mathbf{s}}_m, \phi)p(\phi)d\phi \ . \tag{5.156}$$

Using the joint pdf of $\tilde{\mathbf{n}}$ in (5.10)

$$
\begin{aligned}
p(\tilde{\mathbf{r}}|\alpha\tilde{\mathbf{s}}_m, \phi) &= \frac{1}{(2\pi N_o)^N} \exp\left\{ -\frac{1}{2N_o}\|\tilde{\mathbf{r}} - g\tilde{\mathbf{s}}_m\|^2 \right\} \\
&= \frac{1}{(2\pi N_o)^N} \exp\left\{ -\frac{\|\tilde{\mathbf{r}}\|^2 + 2\alpha^2 E_m}{2N_o} \right\} \\
&\quad \times \exp\left\{ \frac{1}{N_o}\mathrm{Re}(\tilde{\mathbf{r}}, g\tilde{\mathbf{s}}_m) \right\} \ .
\end{aligned}
\tag{5.157}
$$

Now let

$$(\tilde{\mathbf{r}}, g\tilde{\mathbf{s}}_m) = g^*(\tilde{\mathbf{r}}, \tilde{\mathbf{s}}_m) = g^* X_m^{j\theta_m} = \alpha X_m e^{j(\theta_m - \phi)} \ . \tag{5.158}$$

Hence

$$
\begin{aligned}
p(\tilde{\mathbf{r}}|\alpha\tilde{\mathbf{s}}_m, \phi) &= \frac{1}{(2\pi N_o)^N} \exp\left\{ -\frac{\|\tilde{\mathbf{r}} + 2\alpha^2 E_m\|^2}{2N_o} \right\} \\
&\quad \times \exp\left\{ \frac{\alpha X_m}{N_o}\cos(\theta_m - \phi) \right\} \ .
\end{aligned}
\tag{5.159}
$$

Assuming a uniformly distributed random phase ϕ gives

$$
\begin{aligned}
p(\tilde{\mathbf{r}}|\alpha\tilde{\mathbf{s}}_m) &= \frac{1}{(2\pi N_o)^N} \exp\left\{ -\frac{2\alpha^2 E_m + \|\tilde{\mathbf{r}}\|^2}{2N_o} \right\} \\
&\quad \times \frac{1}{2\pi}\int_0^{2\pi} \exp\left\{ \frac{\alpha X_m}{N_o}\cos(\theta_m - \phi) \right\} d\phi \\
&= \frac{1}{(2\pi N_o)^N} \exp\left\{ -\frac{2\alpha^2 E_m + \|\tilde{\mathbf{r}}\|^2}{2N_o} \right\} I_0\left(\frac{\alpha X_m}{N_o} \right) \tag{5.160}
\end{aligned}
$$

Since the term $\|\tilde{\mathbf{r}}\|^2$ is independent of the 111111111 $\tilde{\mathbf{s}}_m$,111hypothesis   the signal maximizes $p(\tilde{\mathbf{r}}|\alpha\tilde{\mathbf{s}}_m)$ also maximizes the metric

$$\mu_1(\mathbf{s}_m) = \frac{\exp\left\{-\frac{\alpha^2 E_m}{N_o}\right\}}{(2\pi N_o)^N} I_0\left(\frac{\alpha X_m}{N_o}\right) . \tag{5.161}$$

If all message waveforms have equal energy, then considerable simplification results. The ML receiver can choose $\tilde{\mathbf{s}}_m$ to maximize

$$\mu_2(\tilde{\mathbf{s}}_m) = I_0\left(\frac{\alpha X_m}{N_o}\right) . \tag{5.162}$$

However $I_0(x)$ increases monotonically with x. Therefore, the ML receiver can simply choose $\tilde{\mathbf{s}}_m$ to maximize

$$\mu_3(\tilde{\mathbf{s}}_m) = X_m \tag{5.163}$$

From the above development, the structure of the ML non-coherent receiver is clear. The receiver first uses the quadrature demodulator in Fig. 5.1 to extract the real and imaginary components of the complex envelope $\tilde{\mathbf{r}}_I(t)$ and $\tilde{\mathbf{r}}_Q(t)$. Then compute

$$\begin{aligned} X_m &= |(\tilde{\mathbf{r}}, \tilde{\mathbf{s}}_m)| \\ &= \left| \int_0^T \tilde{r}(t)\tilde{s}_m^*(t)dt \right| \end{aligned} \tag{5.164}$$

and maximize over the choice of $\tilde{\mathbf{s}}_m$. Continuing further, note that

$$\begin{aligned} X_m &= \left[(\tilde{\mathbf{r}}_I, \tilde{\mathbf{s}}_{I,m})^2 + (\tilde{\mathbf{r}}_Q, \tilde{\mathbf{s}}_{Q,m})^2 \right]^{1/2} \\ &= \left[\left(\int_0^T \tilde{r}_I(t)\tilde{s}_{I,m}(t)dt \right)^2 + \left(\int_0^T \tilde{r}_Q(t)\tilde{s}_{Q,m}(t)dt \right)^2 \right]^{1/2} \end{aligned} \tag{5.165}$$

This leads to the detector structure shown in Fig. 5.21, commonly known as a **square-law detector.** Note that the square-law detector generates $.X_m^2$. However, the choice of $\tilde{\mathbf{s}}_m$ that maximizes X_m^2 also maximizes X_m. If the $\tilde{\mathbf{s}}_m$ do not have equal energy then the metric in (5.161) must be used. This add considerable complexity to the ML receiver because the channel gain a must be determined and the Bessel function $I_0(x)$ must be calculated.

Error probability of M-ary orthogonal signals.   Consider the case of M-ary orthogonal signals as discussed in Chapter 4.5. Assume without loss of generality that $\tilde{\mathbf{s}}_0$ is sent. Then

$$\begin{aligned} \tilde{r}_0 &= g\sqrt{2E} + \tilde{n}_0 \\ \tilde{r}_i &= \tilde{n}_i , \quad i = 1, \dots, M-1 \end{aligned} \tag{5.166}$$
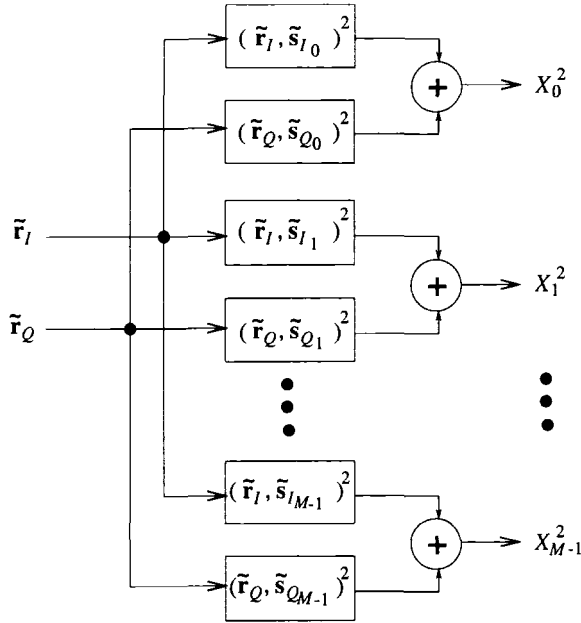
Figure 5.21. Non-coherent square-law detector.

Since the *M*-ary orthogonal signals have equal energy, we can use the metric in (5.163). Then

$$X_m = |\sqrt{2E}r_m| = \sqrt{2E}\left(r_{I,m}^2 + r_{Q,m}^2\right)^{1/2} . \tag{5.167}$$

The receiver will make a correct decision if

$$X_0 > X_i , \ \forall i \neq 0 . \tag{5.168}$$

From Appendix A, X_0 has the Rice distribution

$$p_{X_0}(x_0) = \frac{x_0}{2EN_o} \exp\left\{-\frac{x_0^2 + 4\alpha^2 E^2}{4EN_o}\right\} I_0\left(\frac{\alpha x_0}{N_o}\right) \tag{5.169}$$

while the $X_i, i \neq 0$ are independent Rayleigh random variables with pdf

$$p_{X_i}(x_i) = \frac{x_i}{2EN_o} \exp\left\{-\frac{x_i^2}{4EN_o}\right\} . \tag{5.170}$$

The probability of correct symbol reception is

$$
\begin{aligned}
P(c) &= P\left(X_1 < X_0, X_2 < X_0, \ldots, X_{M-1} < X_0\right) \\
&= \int_0^\infty \left(\int_0^{x_0} \frac{x_i}{2EN_o} \exp\left\{-\frac{x_i^2}{4EN_o}\right\} dx_i\right)^{M-1} p_{X_0}(x_0) dx_0 \\
&= \int_0^\infty \left(1 - \exp\left\{-\frac{x_0^2}{4EN_o}\right\}\right)^{M-1} p(x_0) dx_0 \quad\quad (5.171)
\end{aligned}
$$

Using the binomial expansion

$$
(1 - x)^n = \sum_{k=0}^{n} \binom{n}{k} (-1)^k x^k
$$

gives

$$
P(c) = \sum_{k=0}^{M-1} (-1)^k \binom{M-1}{k} \int_0^\infty \exp\left\{-\frac{kx_0^2}{4EN_o}\right\} p_{X_0}(x_0) dx_0 \ . \quad (5.172)
$$

However, the above integral is

$$
\begin{aligned}
\mathcal{I} &= \int_0^\infty \exp\left\{-\frac{kx_0^2}{4EN_o}\right\} p_{X_0}(x_0) dx_0 \\
&= \int_0^\infty \exp\left\{-\frac{kx_0^2}{4EN_o}\right\} \frac{x_0}{2EN_o} e^{-\frac{x_0^2 + 4\alpha^2 E^2}{4EN_o}} I_0\left(\frac{\alpha x_0}{N_o}\right) dx_0 \\
&= \int_0^\infty \frac{x_0}{2EN_o} \exp\left\{-\frac{(k+1)x_0^2 + 4\alpha^2 E^2}{4EN_o}\right\} I_0\left(\frac{\alpha x_0}{N_o}\right) dx_0
\end{aligned}
$$

The trick is to manipulate the integrand to look like a Ricean pdf. To do so, let

$$
N_o' = \frac{2EN_o}{k+1} \qquad\qquad E' = \frac{E}{(k+1)}
$$

Then

$$
I = \frac{1}{k+1} \exp\left\{-\frac{2k\alpha^2 E'^2}{N_o'}\right\} = \frac{1}{k+1} \exp\left\{-\frac{k\gamma_s}{(k+1)}\right\}
$$

where $\gamma_s = \alpha^2 E/N_o$ is the symbol energy-to-noise ratio. Hence,

$$
P(c) = \sum_{k=0}^{M-1} \frac{(-1)^k \binom{M-1}{k}}{k+1} \exp\left\{-\frac{k\gamma_s}{(k+1)}\right\} \quad\quad (5.173)
$$

and the probability of symbol error is

$$
P_M = 1 - P(c) = \sum_{k=1}^{M-1} \frac{(-1)^{k+1} \binom{M-1}{k}}{k+1} \exp\left\{-\frac{k\gamma_s}{(k+1)}\right\} \quad\quad (5.174)
$$

Since orthogonal modulation is used, the probability of bit error is

$$P_b = \frac{M}{2(M-1)} P_M \ .$$

11. DETECTION OF CPM SIGNALS

CPM receivers can be categorized into three different types of dectection schemes; coherent detection, differential detection, and non-coherent detection. Furthermore, in each category there are two approaches; symbol-by-symbol detectors and sequence estimators. For mobile radio channels, sequence estimation based approaches are not favored for at least two reasons. First is the hardware limitation due to the large number of CPM states, since the complexity of sequence estimators grows ex ponentially with the number of system states. Second is the possible utilization of coding and interleaving to combat fading. Coding issues will be discussed in detail in Chapter 8 The use of interleaving requires separation of the CPM demodulator and the decoder. When convolutional or trellis coding is used with CPM, a soft output symbol-by-symbol detector is typically used to detect the CPM signals, while a sequence estimator is still used for decoding. Therefore, this section only treats symbol-by-symbol CPM detectors. While there exist a large variety of coherent and non-coherent symbol-by-symbol CPM detectors, we present two structures. Both receivers use multiple-symbol observation intervals to detect partial response CPM signals, and both generate soft outputs making them well suited to systems that employ convolutional, trellis, or Turbo coding.

The CPM complex envelope during the time interval $[nT, (n+1)T]$ is

$$
\begin{aligned}
\tilde{s}(t, \boldsymbol{x}) &= \sqrt{2E_s/T} \exp\{j\phi(t, \boldsymbol{x})\} \\
&= \sqrt{2E_s/T} \exp j\{\theta_n + 2\pi h \sum_{i=n-L+1}^{n} x_i q(t - iT)\} \quad (5.175)
\end{aligned}
$$

where E_s is the symbol energy, $T = (\log_2 M)T_b$ is the symbol duration, h is the modulation index, and x is the data symbol sequence chosen from the M-ary alphabet $\{\pm 1, \pm 3, \cdots, \pm(M-1)\}$. The phase shaping function is $\beta(t) = \int_0^t h_f(\tau)d\tau$, where $h_f(t)$ is a partial response frequency shaping pulse of duration LT. The accumulated phase θ_n is equal to $\left[\pi h \sum_{i=-\infty}^{n-L} x_i\right]$ mod 2π. The CPM state at time $t = nT$ is defined by the L-tuple

$$S_n = (\theta_n, x_{n-1}, x_{n-2}, \cdots, x_{n-L+1}) \ . \quad (5.176)$$

In the sequel, the CPM signal will also be denoted by $\tilde{s}(S_n, x_n)$ to emphasize the finite state nature of the signal. For a slow flat fading channel, the received signal is

$$\tilde{r}(t) = g\tilde{s}(t, \boldsymbol{x}) + \tilde{n}(t) \quad (5.177)$$

where $\tilde{n}(t)$ is a zero-mean complex AWGN with psd N_o watts/Hz.

11.1 COHERENT CPM DEMODULATOR

A coherent CPM demodulator was proposed by Osborn and Luntz [254], and Schonhoff[294]. The metrics for symbol x_n are obtained by observing $\tilde{r}(t)$ over N_p+1 successive symbol intervals and generating metrics for all M^{N_p+1} possible symbol vectors $\boldsymbol{x}_n = \{x_n, \mathbf{b}_n\}$, where $\mathbf{b}_n = \{x_{n+1}, \cdots, x_{n+N_p}\}$. The ML metric for x_n is proportional to the conditional density $p(\tilde{r}(t)|S_n, x_n, \mathbf{b}_n, g)$ and is given by[3]

$$\lambda(S_n, x_n, \mathbf{b}_n) = - \sum_{i=n}^{n+N_p} \int_{iT}^{(i+1)T} |\tilde{r}(t) - g\tilde{s}(S_i, x_i)|^2 \, dt \ . \qquad (5.178)$$

The metrics for x_n can be obtained by averaging (5.178) over the M^{N_p} possible values of $\mathbf{b}_n$ and averaging over all possible initial states S_n. This leads to the metric

$$\lambda(x_n) = \sum_{S_n} \sum_{\mathbf{b}_n} \lambda(S_n, x_n, \mathbf{b}_n) P(\mathbf{b}_n) P(S_n) = \sum_{S_n} \sum_{\mathbf{b}_n} \lambda(S_n, x_n, \mathbf{b}_n) \ ,$$
$$(5.179)$$

where $P(\mathbf{b}_n)$ and $P(S_n)$ are the probabilities of $\mathbf{b}_n$ and S_n, respectively, and the last equality follows because all the $\mathbf{b}_n$ are equal, and all the S_n are equal, for equally likely data symbols. By using (5.179) a set of M metrics is calculated for the M possible x_n. The receiver decides in favor of the symbol having the largest metric.

A simplified receiver that will yield almost the same performance uses the suboptimum metric [254, 294]

$$\lambda(x_n) = \max_{S_n} \max_{\mathbf{b}_n} \left\{ - \sum_{i=n}^{n+N_p} \int_{iT}^{(i+1)T} |\tilde{r}(t) - g\tilde{s}(S_i, x_i)|^2 \, dt \right\} \ , \qquad (5.180)$$

which is also exactly the same as the metric proposed by Kerr and McLane for full response CPFSK [181].

11.2 NON-COHERENT CPM DEMODULATOR

The receiver adopts the same multiple-symbol detection strategy used for the coherent receiver described in the previous section. After observing $\tilde{r}(t)$ over an N-symbol interval $[(n - n_1)T, (n + n_2)T]$, where $N = n_1 + n_2 + 1$, the non-coherent CPM demodulator generates the following set of M^{N+L-2}

---

[3]Slowly varying channels are assumed where g remains constant over the observation interval.

conditional symbol metrics for x_n;

$$\lambda(x_n, \mathbf{b}_n) = \left| \sum_{i=n-n_1}^{n+n_2} \int_{iT}^{(i+1)T} \tilde{r}(t)\tilde{s}^*(S_i, x_i)\, dt \right|^2 \tag{5.181}$$

where $\mathbf{b}_n = \{x_{n-n_1-L+1}, \cdots, x_{n-1}, x_{n+1}, \cdots, x_{n+n_2}\}$ is the adjacent symbol vector. Note that the phase term θ_{n-n_1} in $, S_{n-n_1}$ does not affect the value of (5.181) and can therefore be assumed zero. A simple symbol metric can be formed by choosing the largest among all possible $.\lambda(x_n, \mathbf{b}_n)$, viz,

$$\lambda(x_n) = \max_{\mathbf{b}_n} \left| \sum_{i=n-n_1}^{n+n_2} \int_{iT}^{(i+1)T} \tilde{r}(t)\tilde{s}^*(S_i, x_i)\, dt \right|^2 . \tag{5.182}$$

The set of M symbol metrics so obtained is then used to make decisions on the transmitted symbols.

For $N = 1$ $(n_1 = n_2 = 0)$, the symbol metric in (5.182) is the same one used by the single-symbol receiver in [2] and, as a result, the single-symbol receiver can be treated as a special case of the receiver presented here. In order to calculate the metrics recursively, we can follow a similar approach to [305] and rewrite $\lambda(x_n, \mathbf{b}_n)$ as

$$\lambda(x_n, \mathbf{b}_n) = \left| \sum_{i=n-n_1}^{n+n_2} \Gamma_i F_i \right|^2 \tag{5.183}$$

where

$$\Gamma_i = \int_{iT}^{(i+1)T} \tilde{r}(t)\tilde{s}^*(x_{i-L+1}, \cdots, x_i)\, dt$$

$$F_i = e^{-j\pi h x_{i-L}} F_{i-1}; \quad F_{n-n_1} = 1 . \tag{5.184}$$

The metric generator structure is shown in Fig. 5.22. Generally, the metric calculator requires M^L matched filters and generates M^{N+L-1} metrics. However, unlike the coherent receiver the complexity is independent of the modulation index h. Actually, since the term θ_k is not explicitly exploited, h is not even required to be a rational number. Finally, it is observed that no channel information is needed and, therefore, the receiver complexity is greatly reduced.
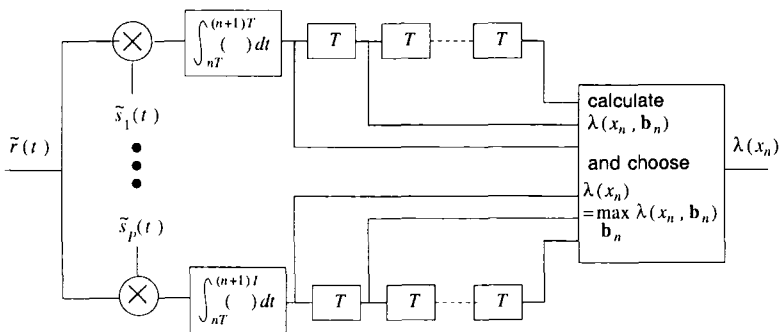
Figure 5.22. The symbol metric calculator. Note that the signal $\tilde{s}^*(t)$ is labeled to account for $P = M^L$ possible matched filters.

Problems

5.1. Derive the upper bound

$$Q(x) \le \frac{1}{2}e^{-x^2/2} \qquad x \ge 0 .$$

Hint: Note that $4Q^2(x)$ is the probability that a pair of independent zero-mean, unit variance, Gaussian random variables u, v lies within the shaded region of Fig. 5.23(a). This probability is exceeded by the probability that u, v lies within the shaded region of Fig. 5.23(b).

(a)   (b)
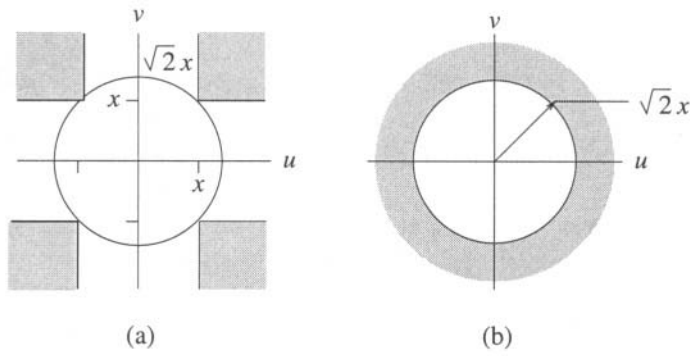
Figure 5.23.   Figure for Problem 5.1.

5.2. Consider the receiver model shown in Fig. 5.24, consisting of a linear time-invariant filter $\tilde{h}(t)$ followed by a sampler. The input to the filter consists of a pulse $\tilde{p}(t)$ of duration T corrupted by AWGN

$$\tilde{r}(t) = \tilde{p}(t) + \tilde{n}(t) , \quad 0 \le t \le T$$
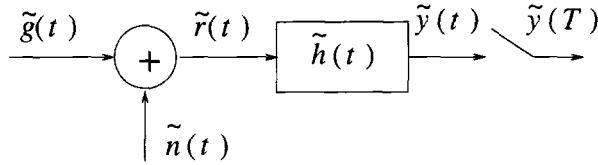
Figure 5.24.   Figure for Problem 5.2.

The output of the filter is

$$y(t) = \tilde{p}_o(t) + \tilde{z}(t)$$

where $\tilde{p}_o(t) = \tilde{p}(t) * \tilde{h}(t)$ and $\tilde{z}(t) = \tilde{n}(t) * \tilde{h}(t)$. The signal-to-noise ratio at the output of the sampler is defined as

$$\text{SNR} = \frac{|\tilde{p}_o(T)|^2}{\text{E}[|\tilde{z}(t)|^2]}$$

Find the filter $\tilde{h}(t)$ (and corresponding transfer function $\tilde{H}(f)$) that will maximize the SNR.

5.3. Derive the expression for the symbol error probability of $\sqrt{M}$-PAM in (5.87).

5.4. Show that the symbol error probability for coherent M-ary PSK is bounded by $p \leq P_M \leq 2p$, where

$$p = Q\left(\sqrt{2\gamma} \sin \frac{\pi}{M}\right)$$

and γ is the symbol energy-to-noise ratio.

5.5. Suppose that BPSK signaling is used with coherent detection. The channel is affected by flat Rayleigh fading and log-normal shadowing with a shadow standard deviation of σ_Ω dB. Plot the average probability of bit error against the average received bit energy-to-noise ratio $\bar{\Omega}_v$ under the assumption that the MS is stationary, i.e., use the Susuki distribution in (2.180). Plot several curves with different values of σ_Ω.

5.6. (computer exercise) This problem requires that you first complete the computer exercise in Problem 2.20, wherein you will construct a Ricean fading simulator. The objective of this question is to evaluate the performance of BPSK signaling on a Ricean fading channel through computer simulation.

Suppose that we send one of two possible signal vectors $\tilde{s}_0 = -\tilde{s}_1 = \sqrt{2E_n}$ where $E_h = E_b$ is the transmitted bit energy. Assuming ideal coherent detection, the received signal vector is

$$\tilde{r} = \alpha\tilde{s}_i + \tilde{n}$$

where α is a Ricean distributed random variable and $\tilde{n}$ is a zero-mean complex Gaussian random variable with variance N_o. For a given α, the probability of bit error is

$$P_b(\gamma_b) = Q\left(\sqrt{2\gamma_b}\right)$$

where $\gamma_b = \alpha^2 E_b/N_o$. The probability of bit error with Ricean fading is

$$P_b = \int_0^\infty Q\left(\sqrt{2\gamma_b}\right) p(\gamma_b)d\gamma_b$$

Evaluate the bit error probability by using computer simulation, where α is generated by the Ricean fading simulator that you developed in Problem 2.20. Assume that the value of α stays constant for a bit duration, i.e., update your fading simulator every T seconds, where T is the bit duration. Assume $f_m T = 0.1$.

Plot the simulated bit error probability, P_b, against the average received bit energy-to-noise ratio $\bar{\gamma}_b = \mathrm{E}[\alpha^2]E_b/N_o$. Show your results for $0.5 < P_b < 10^{-3}$ and for Rice factors $K = 0, 4, 7$ and 16.

Note: To adjust $\bar{\gamma}_b$ you will need to adjust the value of Ω_p in your faded envelope generator.

5.7. Consider an AWGN channel where the channel gain, α, has the following probability density function

$$p(\alpha) = 0.2\delta(\alpha) + 0.5\delta(\alpha - 1) + 0.3\delta(\alpha - 2)$$

a) Determine the average probability of bit error for binary DPSK signaling over a channel with gain α in terms over the *average* received bit energy-to-noise ratio $\bar{\gamma}_b$. What value does the probability of bit error approach as $\bar{\gamma}_b$ gets large?

b) Now suppose that two-branch antenna diversity is used with predetection selective combining. Assume that the diversity branches are perfectly uncorrelated. Determine the average probability of error in terms of the average bit energy-to-noise ratio per diversity branch $\bar{\gamma}_c$. What value does the probability of error approach as $\bar{\gamma}_c$ gets large?

c) Plot the probability of error for parts a) and b) on the same graph.

5.8. Suppose that the average bit energy-to-noise ratio, $\bar{\gamma}_b$, in a cell is uniformly distributed between 12 and 16 dB. Calculate the average probability of bit error in the cell assuming that there is also Rayleigh fading, and binary DPSK signaling is used.

5.9. Consider the differentially coherent receiver shown in Fig. 5.19. Show that the pdf of U is given by (5.142).

5.10. Consider binary, orthogonal signaling using non-coherent FSK modulation and demodulation. The bit error rate for non-coherent FSK on an AWGN channel is

$$P_b = \frac{1}{2}e^{-\gamma_b/2}$$

where $\gamma_b = \alpha^2 E_b/N_o$ is the received bit energy-to-noise ratio. Derive the bit-error-rate for

a) a flat Rayleigh fading channel

b) a flat Ricean fading channel

5.11. Consider binary CPFSK modulation with modulation index $h \leq 0.5$. Compute the minimum squared Euclidean distance between any pair of band-pass waveforms as given by

$$D_{\min}^2 = \lim_{n\to\infty} \min_{i,j} \int_0^{nT} \left[s(t;\mathbf{x}^{(i)}) - s(t;\mathbf{x}^{(j)})\right]^2 dt$$

where $s(t;\mathbf{x}^{(i)})$ and $s(t;\mathbf{x}^{(j)})$ are the two band-pass signals whose phase trajectories diverge at time $t = 0$ and remerge sometime later. What is the pairwise error probability between two such signals?

5.12. The squared Euclidean distance between a pair of CPM band-pass waveforms, $s(t;\mathbf{x}^{(i)})$ and $s(t;\mathbf{x}^{(j)})$, is

$$D^2 = \int_0^\infty \left[s(t;\mathbf{x}^{(i)}) - s(t;\mathbf{x}^{(j)})\right]^2 dt$$

Show that

$$D^2 = 2(\log_2 M)E_b\frac{1}{T}\int_0^\infty [1 - \cos\Delta_\phi(t)]\, dt$$

where M is the symbol alphabet size, E_b is the energy per bit, and $\Delta_\phi(t)$ is the phase difference between the two signals.

5.13. Construct a differential detector for MSK signaling. Obtain an expression for the probability of bit error for differentially detected MSK on an AWGN channel.

5.14. Suppose that GMSK signaling is used. Unfortunately, a GMSK pulse is noncausal and, therefore, a truncated version of the pulse is usually employed in a practical system, i.e., the time domain pulse is

$$g(t) = \left[A \sqrt{\frac{2\pi}{\ln 2}} B \int_{t-T/2}^{t+T/2} \exp\left\{ -\frac{2\pi^2 B^2 x^2}{\ln 2} \right\} dx \right] u_{2L_T T}(t - L_T T) \ .$$

Compute the maximum value of the ISI term in (4.173) as a function of the normalized filter bandwidth BT when $L_T = 3$.

Chapter 6

ANTENNA DIVERSITY

Rayleigh fading has been shown to convert an exponential dependency of the bit error probability on the signal-to-noise ratio into an inverse linear one, thereby resulting in a very large signal-to-noise ratio penalty. Diversity is one very effective remedy that exploits the principle of providing the receiver with multiple faded replicas of the same information bearing signal. To understand the mechanism, let p denote the probability that the instantaneous signal-to-noise ratio is below a critical threshold on each diversity branch. Then with independently faded branches, p^L is the probability that the instantaneous signal-to-noise ratio is below the same critical threshold on all L diversity branches.

The methods by which diversity can be achieved generally fall into seven categories: i) space, ii) angle, iii) polarization, iv) field v) frequency, vi) multipath, and vii) time. Space diversity is achieved by using multiple transmit or receiver antennas. The spatial separation between the multiple antennas is chosen so that the diversity branches experience uncorrelated fading. Chapter 2 showed that a spatial separation of about a half-wavelength will suffice with 2-D isotropic scattering and an isotropic antenna. Angle (or direction) diversity requires a number of directional antennas. Each antenna selects plane waves arriving from a narrow angle of arrival spread, so that uncorrelated branches are achieved. Polarization diversity exploits the property that a scattering environment tends to depolarize a signal. Receiver antennas having different polarizations can be used to obtain uncorrelated branches. Field diversity exploits the fact that the electric and magnetic field components at any point are uncorrelated. Frequency diversity uses multiple channels that are separated by at least the coherence bandwidth of the channel. In many cases, this can be several hundred kilohertz. However, frequency diversity is not a bandwidth efficient solution for TDMA and FDMA systems. Frequency hop spread spec-

trum CDMA systems can exploit frequency diversity through the principle of fast frequency hopping, where each symbol is transmitted sequentially on multiple hops (or carriers) that experience uncorrelated fading. Multipath diversity is obtained by resolving multipath components at different delays by using direct sequence spread spectrum signaling along with a RAKE receiver. Spread spectrum concepts will be discussed in detail in Chapter 8. Time diversity is obtained by transmitting the same information at multiple time periods that are separated by at least the coherence time of the channel. Error correction coding techniques can be viewed as an efficient method of time diversity. Unfortunately, the coherence time of the channel depends on the Doppler spread, and a small Doppler spread implies a large coherence time. Under this condition, it may not be possible to obtain time diversity without introducing unacceptable delay. Finally, the above techniques can be combined together. For example space and time diversity can be combined together by using space-time coding techniques.

This chapter concentrates on antenna diversity techniques. Section 1. discusses receiver antenna diversity techniques where there is a single transmit antenna and multiple receiver antennas. Section 7. considers transmit diversity schemes where there are multiple transmit antennas and a single or multiple receiver antennas.

1. DIVERSITY COMBINING

There are many methods for combining the signals that are received on the disparate diversity branches, and several ways of categorizing them. Diversity combining that takes place at RF is called **predetection combining,** while diversity combining that takes place at baseband is called **postdetection combining.** In many cases there is no difference in performance, at least in an ideal sense. Here we concentrate on implementations that use postdetection combining.

Consider the receiver diversity system shown in Fig. 6.1. The signals that are received by the different antenna branches are demodulated to baseband with a quadrature demodulator in Fig. 5.1, processed with correlator or matched filter detector, and then applied to a diversity combiner as shown in Fig. 6.1. Here we consider maximal ratio, equal gain, selective, and switched combining.

If the signal $\tilde{s}_m(t)$ is transmitted, the received complex envelopes on the different diversity branches are

$$\tilde{r}_k(t) = g_k \tilde{s}_m(t) + \tilde{n}_k(t), \qquad k = 1, \ldots, L \qquad (6.1)$$

where $g_k = \alpha_k e^{-j\phi_k}$ is the fading gain associated with the kth branch. The AWGN processes $\tilde{n}_k(t)$ are independent from branch to branch. The corresponding received signal vectors are

$$\tilde{\mathbf{r}}_k = g_k \tilde{\mathbf{s}}_m + \tilde{\mathbf{n}}_k, \qquad k = 1, \ldots, L \qquad (6.2)$$
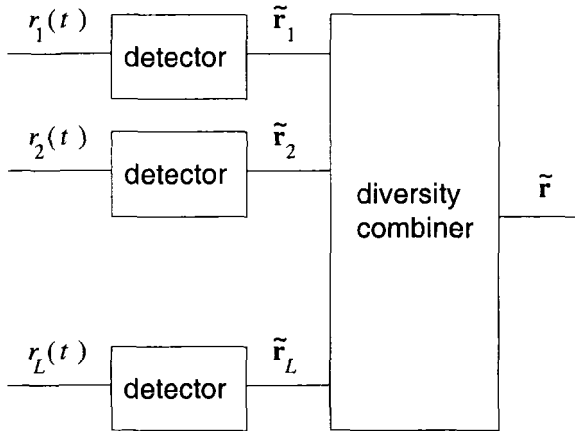
Figure 6.1.   Postdetection diversity receiver.

where
$$\tilde{r}_{k_i} = g_k \tilde{s}_{m_i} + \tilde{n}_{k_i} . \tag{6.3}$$

The fading gains of the various diversity branches typically have some degree of correlation, and the degree of correlation depends on the type of diversity being used and the propagation environment. To simplify analysis, the the diversity branches are usually assumed to be uncorrelated. However, branch correlation will reduce the achievable diversity gain and, therefore, the uncorrelated branch assumption gives optimistic results. Nevertheless, we will evaluate the performance of the various diversity combining techniques under the assumption of uncorrelated branches.

The fade distribution will affect the diversity gain. In general, the *relative* advantage of diversity is greater for Rayleigh fading than Ricean fading, because as the Rice factor K increases there is less difference between the instantaneous received signal-to-noise ratios on the various diversity branches. However, the performance will always be better with Ricean fading than with Rayleigh fading, for a given average received signal-to-noise ratio and diversity order. For our purpose, we will consider the performance with Rayleigh fading.

2.   SELECTIVE COMBINING

With selective combining (SC), the branch yielding the highest signal-to-noise ratio is always selected. In this case, the diversity combiner in Fig. 6.1 performs the operation
$$\tilde{r} = \max_{|g_k|} \tilde{r}_k . \tag{6.4}$$

For radio systems that use continuous transmission, SC is impractical because it requires continuous monitoring of all diversity branches to obtain the time-

varying complex gains g_k. If such monitoring is performed, then it is probably better to use maximal ratio combining, as discussed in the next section, since the implementation is not that much more complicated and the performance is better. However, in systems that use TDMA, a form of SC can sometimes be implemented where the diversity branch is selected prior to the transmission of a TDMA burst. The selected branch is then used for the duration of the entire burst. Obviously, such an approach is only useful if the channel does not change significantly over a TDMA burst. In this section, however, we evaluate selection diversity under the assumption of continuous branch selection.

With Rayleigh fading, the instantaneous received symbol energy-to-noise ratio on the kth diversity branch has the exponential pdf

$$p_{\gamma_k}(x) = \frac{1}{\bar{\gamma}_c} e^{-x/\bar{\gamma}_c} \tag{6.5}$$

where $\bar{\gamma}_c$ is the average received branch symbol energy-to-noise ratio. With ideal SC, the branch with the largest symbol energy-to-noise ratio is always selected so the instantaneous symbol energy-to-noise ratio at the output of the selective combiner is

$$\gamma_s^{\rm s} = \max\{\gamma_1, \gamma_2, \cdots, \gamma_L\} \tag{6.6}$$

where L is the number of branches. If the branches are independently faded, then order statistics gives the cumulative distribution function (cdf)

$$F_{\gamma_s^{\rm s}}(x) = \Pr[\gamma_1 \le x, \gamma_2 \le x, \cdots, \gamma_L \le x] = \left[1 - e^{-x/\bar{\gamma}_c}\right]^L . \tag{6.7}$$

Differentiating the above expression gives the pdf

$$p_{\gamma_s^{\rm s}}(x) = \frac{L}{\bar{\gamma}_c} \left[1 - e^{-x/\bar{\gamma}_c}\right]^{L-1} e^{-x/\bar{\gamma}_c} . \tag{6.8}$$

The average symbol energy-to-noise ratio with SC is

$$
\begin{aligned}
\bar{\gamma}_s^{\rm s} &= \int_0^\infty x p_{\gamma_s^{\rm s}}(x) dx \\
&= \int_0^\infty \frac{Lx}{\bar{\gamma}_c} \left[1 - e^{-x/\bar{\gamma}_c}\right]^{L-1} e^{-x/\bar{\gamma}_c} dx \\
&= \bar{\gamma}_c \sum_{k=1}^L \frac{1}{k} .
\end{aligned}
\tag{6.9}
$$

Fig. 6.2 plots the cdf $F_{\gamma_s^{\rm s}}(x)$ against the normalized symbol energy-to-noise ratio $x/\bar{\gamma}_c$. Note that the largest diversity gain is obtained in going from $L = 1$ to $L = 2$, and diminishing returns are obtained with increasing L. This is typical for all diversity techniques.
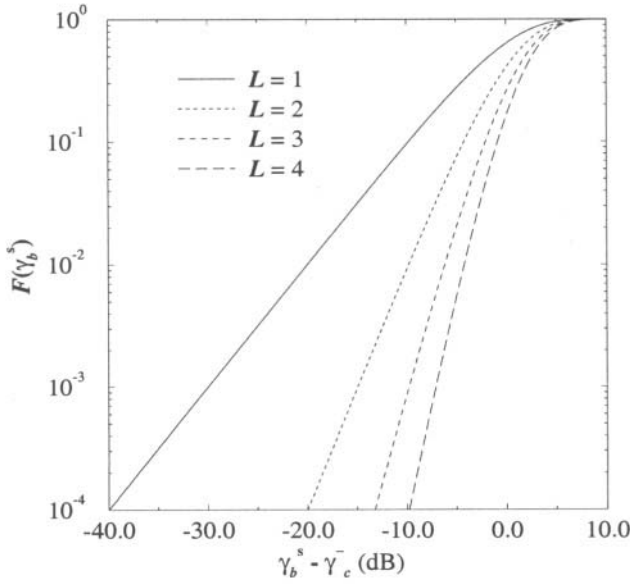
Figure 6.2. Cdf of γ_b^s for selective combining; $\bar{\gamma}_c$ is the average branch bit energy-to-noise ratio.

The bit error probability with slow fading is obtained by averaging over the pdf of γ_s^s. For example, consider binary DPSK with differential detection having the bit error probability

$$P_b(\gamma_s^s) = \frac{1}{2}e^{-\gamma_s^s} .$$ (6.10)

Hence, with SC

$$
\begin{aligned}
P_b &= \int_0^\infty P_b(x)p_{\gamma_s^s}(x)dx \\
&= \int_0^\infty \frac{L}{2\bar{\gamma}_c}e^{-(1+1/\bar{\gamma}_c)x}\left(1 - e^{-x/\bar{\gamma}_c}\right)^{L-1} dx \\
&= \frac{L}{2\bar{\gamma}_c}\sum_{n=0}^{L-1}\binom{L-1}{n}(-1)^n \int_0^\infty e^{-(1+(n+1)/\bar{\gamma}_c)x}dx \\
&= \frac{L}{2}\sum_{n=0}^{L-1}\frac{\binom{L-1}{n}(-1)^n}{1+n+\bar{\gamma}_c}
\end{aligned}
$$ (6.11)

where we have used the binomial expansion

$$(1-x)^{L-1} = \sum_{n=0}^{L-1}\binom{L-1}{n}(-1)^n x^n .$$ (6.12)
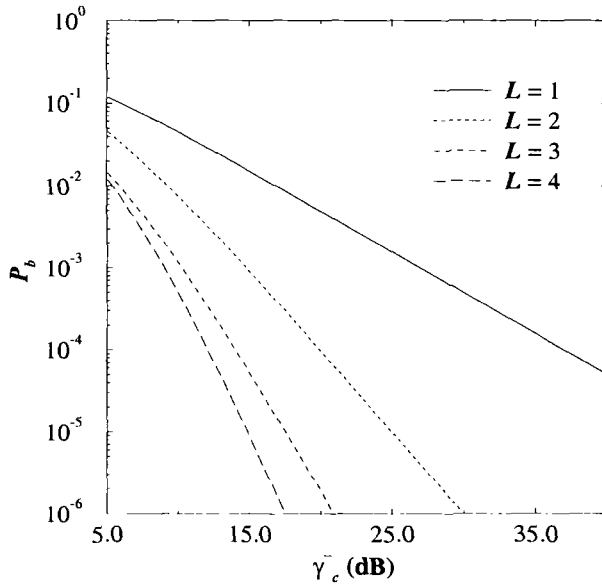
Figure 6.3. Bit error probability for selective combining and binary DPSK with differential detection.

The bit error probability is plotted in Fig. 6.3, where $\bar{\gamma}_c$ is equal to the branch *bit* energy-to-noise ratio since binary modulation is being used. SC is seen to give a very large improvement in error performance. When $\bar{\gamma}_c/gg1$, (6.11) shows that the bit error probability is proportional to $1/\bar{\gamma}_c^L$. Again, the largest diversity gain is achieved with 2-branch diversity and diminishing returns are realized with increasing L.

3. MAXIMAL RATIO COMBINING

With maximal ratio combining (MRC), the diversity branches are weighted by their respective complex fading gains and combined. MRC realizes an ML receiver as we now show. Referring to (6.2), the vector

$$\tilde{\mathbf{r}} \overset{\triangle}{=} (\tilde{\mathbf{r}}_1, \ \tilde{\mathbf{r}}_2, \ \cdots, \ \tilde{\mathbf{r}}_L) \tag{6.13}$$

has the multivariate Gaussian distribution

$$p(\tilde{\mathbf{r}}|\mathbf{g}, \tilde{\mathbf{s}}_m) = \prod_{k=1}^{L} \prod_{i=1}^{N} \frac{1}{2\pi N_o} \exp\left\{ -\frac{1}{2N_o}|\tilde{r}_{k,i} - g_k\tilde{s}_{m,i}|^2 \right\}$$

$$= \frac{1}{(2\pi N_o)^{LN}} \exp\left\{ -\frac{1}{2N_o} \sum_{k=1}^{L} \|\tilde{\mathbf{r}}_k - g_k\tilde{\mathbf{s}}_m\|^2 \right\} \tag{6.14}$$

where $\mathbf{g} = (g_1, g_2, \ldots, g_L)$ is the channel vector. From this expression, the ML receiver chooses the message vector $\tilde{\mathbf{s}}_m$ that maximizes the metric

$$\mu(\tilde{\mathbf{s}}_m) = -\sum_{k=1}^{L} \|\tilde{\mathbf{r}}_k - g_k \tilde{\mathbf{s}}_m\|^2$$

$$= -\sum_{k=1}^{L} \left\{ \|\tilde{\mathbf{r}}_k\|^2 - 2\mathrm{Re}\,(g_k^* \tilde{\mathbf{r}}_k, \tilde{\mathbf{s}}_m) + |g_k|^2 \|\tilde{\mathbf{s}}_m\|^2 \right\} . \quad (6.15)$$

Since $\sum_{k=1}^{L} \|\tilde{\mathbf{r}}_k\|^2$ is independent of the hypothesis as to which $\tilde{\mathbf{s}}_m$ was sent and $\|\tilde{\mathbf{s}}_m\|^2 = 2E_m$, the receiver just needs to maximize the metric

$$\mu_2(\tilde{\mathbf{s}}_m) = \sum_{k=1}^{L} \mathrm{Re}\,(g_k^* \tilde{\mathbf{r}}_k, \tilde{\mathbf{s}}_m) - E_m \sum_{k=1}^{L} |g_k|^2$$

$$= \sum_{k=1}^{L} \mathrm{Re}\left\{ g_k^* \int_0^T \tilde{r}_k(t) \tilde{s}_m^*(t)dt \right\} - \sum_{k=1}^{L} |g_k|^2 E_m \quad (6.16)$$

If signals have equal energy then the last term can be neglected, since it is the same for all message vectors. This results in

$$\mu_3(\tilde{\mathbf{s}}_m) = \sum_{k=1}^{L} \mathrm{Re}\,(g_k^* \tilde{\mathbf{r}}_k, \tilde{\mathbf{s}}_m)$$

$$= \sum_{k=1}^{L} \mathrm{Re}\left\{ g_k^* \int_0^T \tilde{r}_k(t) \tilde{s}_m^*(t)dt \right\} \quad (6.17)$$

An alternative form of the ML receiver can also be obtained by rewriting the metric in (6.16) as

$$\mu_4(\tilde{\mathbf{s}}_m) = \mathrm{Re}\left(\sum_{k=1}^{L} g_k^* \tilde{\mathbf{r}}_k, \tilde{\mathbf{s}}_m \right) - E_m \sum_{k=1}^{L} |g_k|^2$$

$$= \int_0^T \mathrm{Re}\left\{ \left(\sum_{k=1}^{L} g_k^* \tilde{r}_k(t) \right) \tilde{s}_m^*(t) \right\} dt \sum_{k=1}^{L} |g_k|^2 E_m \quad (6.18)$$

From the above development, the ML receiver can be constructed. The diversity combiner in Fig. 6.1 just generates the sum

$$\tilde{\mathbf{r}} = \sum_{k=1}^{L} g_k^* \tilde{\mathbf{r}}_k \quad (6.19)$$

which is then applied to the metric computer shown in Fig. 6.4.
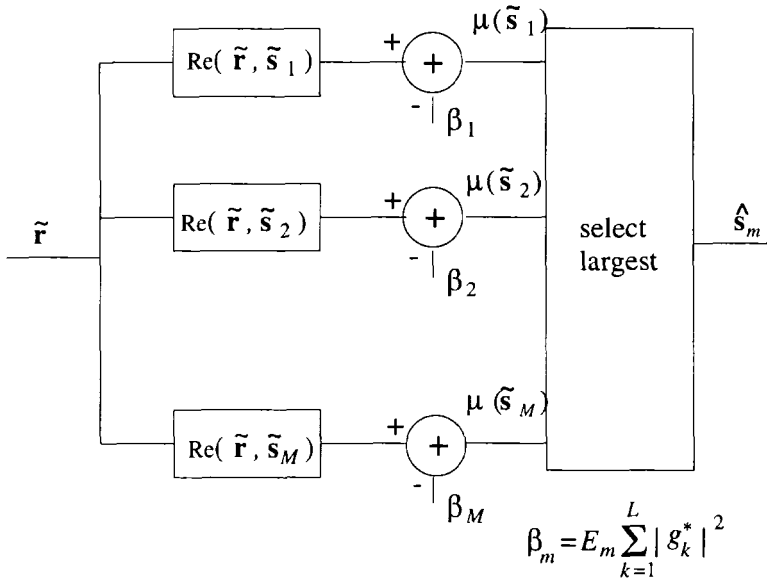
Figure 6.4. Metric computer for maximal ratio combining.

After weighting, co-phasing and combining, the envelope of the composite signal component is

$$\alpha_M = \sum_{k=1}^{L} \alpha_k^2 \ . \tag{6.20}$$

The weighted sum of the branch noise powers is

$$\sigma_{\tilde{n},\text{tot}}^2 = N_o \sum_{k=1}^{L} \alpha_k^2 \ . \tag{6.21}$$

Hence, the symbol energy-to-noise ratio is

$$\gamma_s^{\text{mr}} = \frac{\alpha_M^2 E_{\text{av}}}{\sigma_{\tilde{n},\text{tot}}^2} = \sum_{k=1}^{L} \frac{\alpha_k^2 E_{\text{av}}}{N_o} = \sum_{k=1}^{L} \gamma_k \tag{6.22}$$

where $\gamma_k = \alpha_k^2 E_{\text{av}}/N_o$, and E_{av} is the average symbol energy in the signal constellation. Hence, γ_s^{mr} is the sum of the symbol energy-to-noise ratios of the diversity branches.

If the branches are balanced (which is a reasonable assumption with antenna diversity) and uncorrelated, then γ_s^{mr} has a chi-square distribution with $2L$ degrees of freedom. That is,

$$p_{\gamma_s^{\text{mr}}}(x) = \frac{1}{(L-1)!(\bar{\gamma}_c)^L} x^{L-1} e^{-x/\bar{\gamma}_c} \tag{6.23}$$
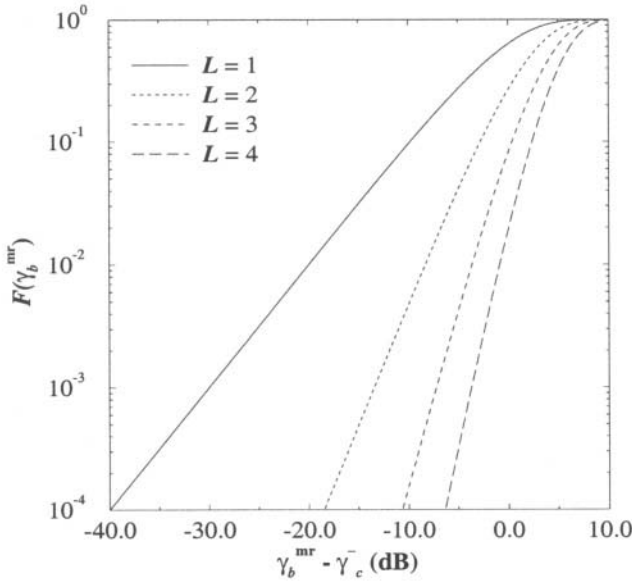
Figure 6.5. Cdf of γ_s^{mr} for maximal ratio combining; $\bar{\gamma}_c$ is the average branch symbol energy-to-noise ratio.

where

$$\bar{\gamma}_c = \mathrm{E}[\gamma_k] \qquad k = 1, \ldots, L . \tag{6.24}$$

The cdf of γ_s^{mr} is

$$F_{\gamma_s^{\mathrm{mr}}}(x) = 1 - e^{-x/\bar{\gamma}_c} \sum_{k=0}^{L-1} \frac{1}{k!} \left(\frac{x}{\bar{\gamma}_c} \right)^k . \tag{6.25}$$

It follows from (6.22) that the average symbol energy-to-noise ratio with MRC is

$$\bar{\gamma}_s^{\mathrm{mr}} = \sum_{k=1}^{L} \bar{\gamma}_k = \sum_{k=1}^{L} \bar{\gamma}_c = L\bar{\gamma}_c . \tag{6.26}$$

Fig. 6.2 plots the cdf $F_{\gamma_b^{\mathrm{mr}}}(x)$. Plots of the cumulative distribution function allow easy comparison of the various combining schemes, without the need to consider particular modulation schemes. For example, with SC the cdf in Fig. 6.2 shows that $F_{\gamma_s^s} = 10^{-4}$ at $\gamma_s^s - \bar{\gamma}_c = 20$ dB. However, for MRC Fig. 6.5 shows that $F_{\gamma_s^{\mathrm{mr}}}(x) = 10^{-4}$ at $\gamma_s^{\mathrm{mr}} - \bar{\gamma}_c = 18$ dB. The implication is that MRC is 2 dB more effective than SC.

Since MRC is a coherent detection technique we must limit our attention to coherent signaling techniques, e.g., BPSK and M-QAM. For example, if BPSK
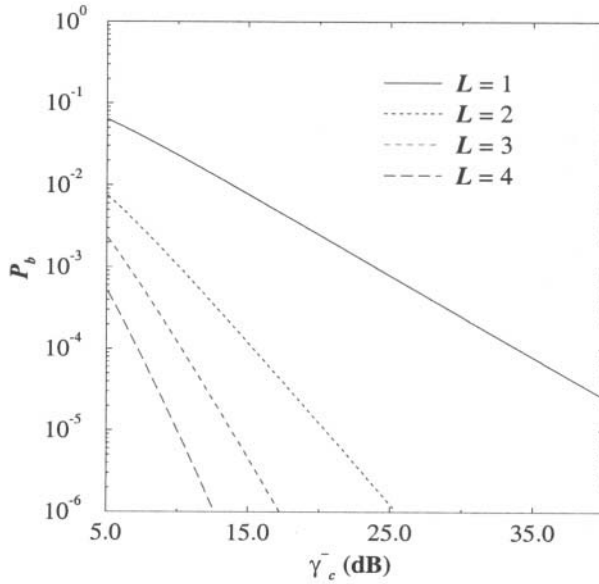
Figure 6.6. Bit error probability for predetection maximal ratio combining and coherent BPSK signaling.

is used the bit error probability is:

$$
\begin{aligned}
P_b &= \int_0^\infty P_b(x) p_{\gamma_s^{\mathrm{mr}}}(x)\,dx \\
&= \int_0^\infty Q\left(\sqrt{2x}\right) \frac{1}{(L-1)!(\bar{\gamma}_c)^L} x^{L-1} e^{-x/\bar{\gamma}_c}\,dx \\
&= \left(\frac{1-\mu}{2}\right)^L \sum_{k=0}^{L-1} \binom{L-1+k}{k} \left(\frac{1+\mu}{2}\right)^k
\end{aligned}
$$
(6.27)

where

$$
\mu = \sqrt{\frac{\bar{\gamma}_c}{1+\bar{\gamma}_c}}
$$
(6.28)

The last step follows after some algebra. The expression in (6.27) is plotted in Fig. 6.6. Once again, diversity significantly improves the performance.

4. EQUAL GAIN COMBINING

Equal gain combining (EGC) is similar to MRC because the diversity branches are co-phased, but different from MRC because the diversity branches are not weighted. In practice, such a scheme is useful for modulation techniques having equal energy symbols, e.g., M-PSK. With signals of unequal

energy, the complete channel vector $\mathbf{g} = (g_1, g_2, \ldots, g_L)$ is required anyway and MRC might as well be used. With EGC, the receiver maximizes the metric

$$
\begin{aligned}
\mu(\tilde{\mathbf{s}}_m) &= \sum_{k=1}^{L} \text{Re}\left(\tilde{\mathbf{r}}_k e^{-j\phi_k}, \tilde{\mathbf{s}}_m\right) \\
&= \sum_{k=1}^{L} \text{Re}\left\{ e^{-j\phi_k} \int_0^T \tilde{r}_k(t)\tilde{s}_m^*(t)dt \right\} .
\end{aligned}
\tag{6.29}
$$

This metric can be rewritten in the alternate form

$$
\begin{aligned}
\mu(\tilde{\mathbf{s}}_m) &= \text{Re}\left(\sum_{k=1}^{L} e^{-j\phi_k}\tilde{\mathbf{r}}_k, \tilde{\mathbf{s}}_m^*\right) \\
&= \int_0^T \text{Re}\left\{ \left(\sum_{k=1}^{L} e^{-j\phi_k}\tilde{r}_k(t)\right) \tilde{s}_m^*(t) \right\} dt .
\end{aligned}
\tag{6.30}
$$

The combiner in Fig 6.1 just generates the sum

$$
\tilde{\mathbf{r}} = \sum_{k=1}^{L} e^{-\phi_k}\tilde{\mathbf{r}}_k
\tag{6.31}
$$

The vector $\tilde{\mathbf{r}}$ is then applied to the metric computer shown in Fig. 6.4 with $\beta_m = 0, m = 1, \ldots, L$. The reason for setting $\beta_m = 0$ comes from the assumption of equal energy signals.

After co-phasing and combining, the envelope of the composite signal is

$$
\alpha_E = \sum_{k=1}^{L} \alpha_k
\tag{6.32}
$$

and the sum of the branch noise powers is LN_o. The resulting symbol energy-to-noise ratio is

$$
\gamma_s^{\text{eg}} = \frac{\alpha_E^2 E_{\text{av}}}{LN_o} .
\tag{6.33}
$$

The cdf and pdf for γ_s^{eg} does not exist in closed form for $L > 2$. However, for $L = 2$ and $\bar{\gamma}_1 = \bar{\gamma}_2 = \bar{\gamma}_c$, the cdf is equal to

$$
F_{\gamma_s^{\text{eg}}}(x) = 1 - e^{-2x/\bar{\gamma}_c} - \sqrt{\pi \frac{x}{\bar{\gamma}_c}}e^{-x/\bar{\gamma}_c}\left(1 - 2Q\left(\sqrt{2\frac{x}{\bar{\gamma}_c}}\right)\right) .
\tag{6.34}
$$

Differentiating the above expression yields the pdf

$$
\begin{aligned}
p_{\gamma_s^{\text{eg}}}(x) &= \frac{1}{\bar{\gamma}_c}e^{-2x/\bar{\gamma}_c} + \sqrt{\pi}e^{-x/\bar{\gamma}_c}\left(\frac{1}{2\sqrt{x\bar{\gamma}_c}} - \frac{1}{\bar{\gamma}_c}\sqrt{\frac{x}{\bar{\gamma}_c}}\right) \\
&\quad \times \left(1 + 2Q\left(\sqrt{2\frac{x}{\bar{\gamma}_c}}\right)\right) .
\end{aligned}
\tag{6.35}
$$

The average symbol energy-to-noise ratio with EGC is

$$\bar{\gamma}_s^{eg} = \frac{E_{av}}{LN_o} E\left[\left(\sum_{k=1}^{L} \alpha_k\right)^2\right]$$

$$= \frac{E_{av}}{LN_o} \sum_{j=1}^{L} \sum_{\ell=1}^{L} E\left[\alpha_k \alpha_\ell\right] . \qquad (6.36)$$

With Rayleigh fading, $E[\alpha_k^2] = 2\sigma^2$ and $E[\alpha_k] = \sqrt{\pi/2}\ \sigma$. Furthermore, if the branches experience uncorrelated fading, then $E[\alpha_k \alpha_\ell] = E[\alpha_k]E[\alpha_\ell]$ for $k \neq \ell$. Hence,

$$\bar{\gamma}_s^{eg} = \frac{E_{av}}{LN_o}\left(2L\sigma^2 + L(L-1)\frac{\pi\sigma}{2}\right)$$

$$= \frac{2\sigma^2 E_{av}}{N_o}\left(1 + (L-1)\frac{\pi}{4}\right)$$

$$= \bar{\gamma}_c\left(1 + (L-1)\frac{\pi}{4}\right) . \qquad (6.37)$$

The error probability with 2-branch combining can be obtained by using the pdf in (6.35). For example, with coherent BPSK signaling the bit error probability is (see Problem 6.8)

$$P_b = \int_0^\infty P_b(x) p_{\gamma_s^{eg}}(x) dx$$

$$= \frac{1}{2}\left(1 - \sqrt{1 - \mu^2}\right) \qquad (6.38)$$

where

$$\mu = \frac{1}{1 + \bar{\gamma}_c} . \qquad (6.39)$$

5.   SWITCHED COMBINING

A switched combiner scans through the diversity branches until it finds one that has a signal-to-noise ratio exceeding a specified threshold. This diversity branch is selected and used until the signal-to-noise ratio again drops below the threshold. When this happens another diversity branch is chosen which has a signal-to-noise ratio exceeding the threshold. The big advantage of switched combining is that only one detector is needed. There are several variations of switched diversity. Here, we analyze two-branch **switch and stay** combining (SSC). With SSC, the receiver switches to, and stays with, the alternate branch when the signal-to-noise ratio drops below a specified threshold. It does this regardless of whether or not the signal-to-noise ratio with the alternate branch is above or below the threshold.

Let the symbol energy-to-noise ratios associated with the two branches be denoted by γ_1 and γ_2, and let the switching threshold be denoted by T. By using (6.5), the probability that γ_i is less than T is

$$
\begin{aligned}
q &= \Pr[\gamma_i < T] \\
&= 1 - e^{-T/\bar{\gamma}_c} \ .
\end{aligned}
\tag{6.40}
$$

Likewise, the probability that γ_i is less than S is

$$
p = 1 - e^{-S/\bar{\gamma}_c}
\tag{6.41}
$$

Let γ_s^{sw} denote the symbol energy-to-noise ratio at the output of the switched combiner. Then

$$
\Pr[\gamma_s^{\text{sw}} \leq S] = \Pr\left[\{\gamma_s^{\text{sw}} \leq S | \gamma_s^{\text{sw}} = \gamma_1\} \bigcup \{\gamma_s^{\text{sw}} \leq S | \gamma_s^{\text{sw}} = \gamma_2\}\right]
\tag{6.42}
$$

Since γ_1 is statistically identical to γ_2, we can assume that branch 1 is currently in use. It follows that

$$
\Pr[\gamma_s^{\text{sw}} \leq S] = \left\{ \begin{array}{ll} \Pr\left[\{\gamma_1 \leq T\} \bigcap \{\gamma_2 \leq S\}\right], & S < T \\ \Pr\left[\{T \leq \gamma_1 \leq S\} \bigcup \{\gamma_1 \leq T \bigcap \gamma_2 \leq S\}\right], & S \geq T \end{array} \right.
\tag{6.43}
$$

The region $S < T$ corresponds to the case where γ_1 has dropped below the threshold T and a switch to branch 2 is initiated, but $\gamma_2 < T$ so that the switch does not result in a γ_b^{sw} greater than T. On the other hand, the region $S \geq T$ corresponds to the case when either γ_1 is between T and S or when γ_1 has dropped below the threshold T so that a switch to branch 2 occurs, and $T \leq \gamma_2 \leq S$. Since γ_1 and γ_2 are independent, the above probabilities are

$$
\Pr\left[\{\gamma_1 \leq T\} \bigcap \{\gamma_2 \leq S\}\right] = qp
\tag{6.44}
$$

$$
\Pr\left[\{T \leq \gamma_1 \leq S\} \bigcup \{\gamma_1 \leq T \bigcap \gamma_2 \leq S\}\right] = p - q + qp
\tag{6.45}
$$

Therefore,

$$
\Pr[\gamma_b^{\text{sw}} \leq S] = \left\{ \begin{array}{ll} qp & S < T \\ p - q + qp & S \geq T \end{array} \right. .
\tag{6.46}
$$

Fig. 6.7 plots the cdf $F_{\gamma_s^{\text{sw}}}(x)$ for several values of the normalized threshold $R = 10 \log_{10}(T/\bar{\gamma}_c)$ (dB). Observe that SSC always performs worse than SC except at the switching threshold, where the performance is the same. Since SSC offers the most improvement just above the threshold level, the threshold level should be chosen as γ_{th}, the minimum acceptable instantaneous symbol energy-to-noise ratio that the radio system can tolerate and still provide an acceptable error probability performance. Finally, the optimum threshold,
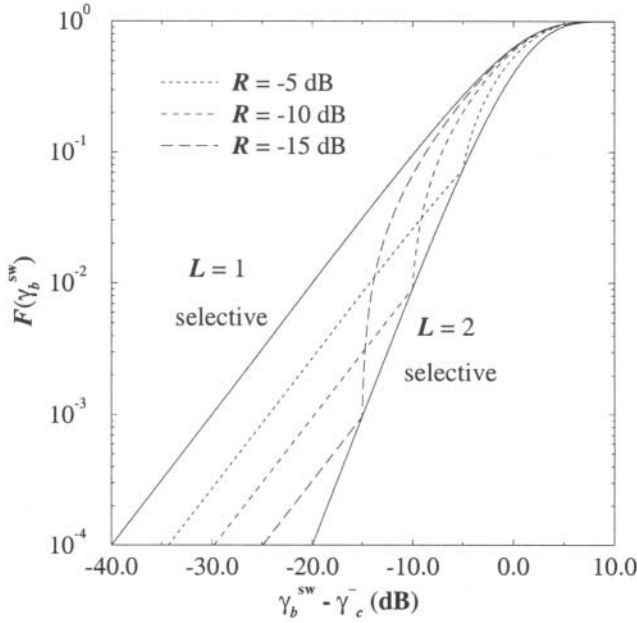
Figure 6.7. Cdf of γ_{sw} for 2-branch switched diversity for several values of the normalized threshold $R = 10\log_{10}(T/\bar{\gamma}_c)$; $\bar{\gamma}_c$ is the average branch symbol energy-to-noise ratio.

$T = R\bar{\gamma}_c$, depends on $\bar{\gamma}_c$. Since $\bar{\gamma}_c$ varies due to path loss and shadowing, the threshold must be adaptive.

The probability of bit error can be also be computed for SSC. The pdf for γ_s^{sw} is

$$p_{\gamma_s^{sw}}(x) = \begin{cases} q\frac{1}{\bar{\gamma}_c}e^{-x/\bar{\gamma}_c} & , \quad x < T \\ (1+q)\frac{1}{\bar{\gamma}_c}e^{-x/\bar{\gamma}_c} & , \quad x \geq T \end{cases} \qquad (6.47)$$

If binary DPSK is used, then the probability of error is

$$\begin{aligned} P_b &= \int_0^\infty P_b(x)p_{\gamma_s^{sw}}(x)dx \\ &= \frac{1}{2(1+\bar{\gamma}_c)}\left(q + (1-q)e^{-T}\right) \ . \end{aligned} \qquad (6.48)$$

where $\bar{\gamma}_c$ is the average branch bit energy-to-noise ratio. The above expression is plotted in Fig. 6.8 for several values of T. The performance with $T = 0$ is the same as using no diversity at all, because no switching occurs. The performance changes little for $T > 6$. As T increases, the probability of switching q also increases, as shown in Fig. 6.9. For some system, it may desirable to keep q as small as possible to minimize the number of switches.
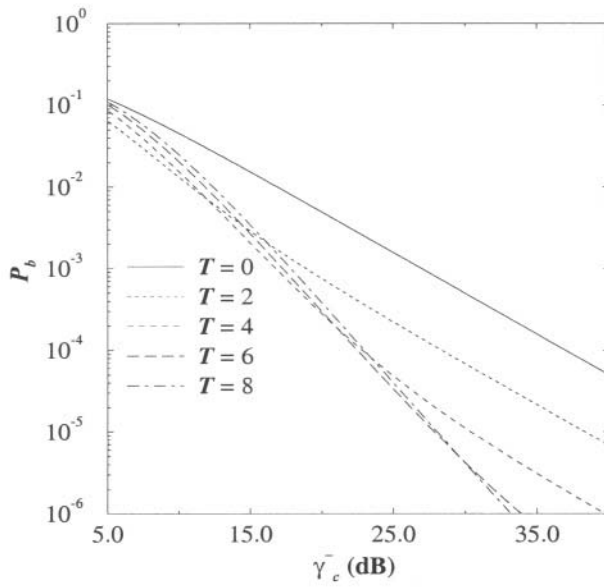
Figure 6.8. Bit error probability for 2-branch switched combining and differentially detected DPSK.
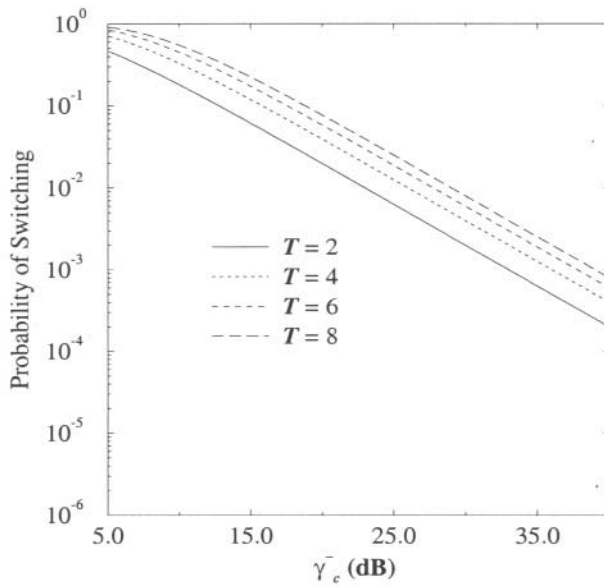
Figure 6.9. Probability of switching for two-branch switched combining.
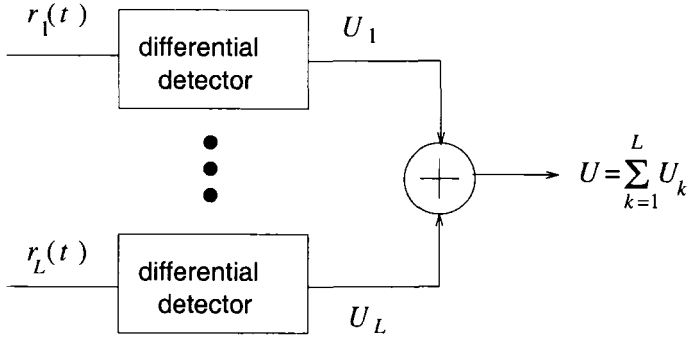
Figure 6.10. Differential detection with postdetection equal gain combining.

6. DIFFERENTIAL DETECTION WITH EQUAL GAIN COMBINING

Equal gain combining has a simple implementation and very good performance when used in conjunction with differential detection. Differential detection circumvents the need to co-phase and weight the diversity branches. The overall receiver structure is shown in Fig. 6.10. The structure of the individual differential detectors depends on the type of modulation that is being used. For DPSK, the detector is shown in Fig. 5.19, while for $\pi/4$-QPSK the detector is shown in Fig. 5.20. In the latter case, the U and V branches are combined separately.

For DPSK the decision variable at the output of the combiner at epoch n is, from (5.137),

$$U_n = \sum_{k=1}^{L} U_{n,k} = \frac{1}{2} \sum_{k=1}^{L} (Z_{n,k} Z_{nd,k}^* + Z_{n,k}^* Z_{nk,d}) \ . \tag{6.49}$$

Once again, by using characteristic functions it can be shown that $U_n = W_n - Y_n$, where W_n and Y_n are non-central and central chi-square random variables with $2L$ degrees of freedom, respectively, and having the densities

$$f_{W_n}(w) = \frac{1}{2 E_h N_o} \left(\frac{w}{s^2} \right)^{\frac{L-1}{2}} \exp\left\{ -\frac{(s^2 + w)}{2 E_h N_o} \right\} I_{L-1}\left(\sqrt{w} \frac{s}{E_h N_o} \right) \tag{6.50}$$

$$f_{Y_n}(y) = \left(\frac{1}{2 E_h N_o} \right)^{L} \frac{1}{(L-1)!} y^{L-1} \exp\left\{ -\frac{y}{2 E_h N_o} \right\} \ . \tag{6.51}$$

where

$$s^2 = 4 E_h \sum_{k=1}^{L} \alpha_k^2 \tag{6.52}$$

is the non-centrality parameter, and $I_n(x)$ is the nth-order modified Bessel function of the first kind, defined by

$$I_n(x) = \frac{1}{2\pi} \int_0^{2\pi} e^{x\cos\theta} \cos(n\theta)d\theta \ . \tag{6.53}$$

After some algebraic detail, the probability of error can be expressed in the closed form [270]

$$P_b(\gamma_t) = \frac{1}{2^{2L-1}} e^{-\gamma_t} \sum_{k=0}^{L-1} b_k \gamma_t^k \tag{6.54}$$

where

$$b_k = \frac{1}{k!} \sum_{n=0}^{L-1-k} \binom{2L-1}{n} \tag{6.55}$$

and

$$\gamma_t = \sum_{k=1}^{L} \gamma_k \ . \tag{6.56}$$

Since γ_t has the central chi-square distribution in (6.23), averaging $P_b(\gamma_t)$ gives the result

$$P_b = \frac{1}{2^{2L-1}(L-1)!(1+\bar{\gamma}_c)^L} \sum_{k=1}^{L-1} b_k(L-1+k)! \left(\frac{\bar{\gamma}_c}{1+\bar{\gamma}_c}\right)^k \ . \tag{6.57}$$

This can be manipulated in the same form as (6.27) with

$$\mu = \frac{\bar{\gamma}_c}{1+\bar{\gamma}_c} \ . \tag{6.58}$$

The various diversity combining techniques are compared in Fig. 6.11 for differentially detected binary DPSK signals. It is apparent that SSC results in the worst performance, followed by SC. Differential detection followed by EGC give the best performance. Once again, we stress that it does not make sense to use MRC with differential detection since MRC is a coherent combining technique. Therefore, a curve for MRC is not included in Fig. 6.11.

7. TRANSMITTER DIVERSITY

Transmitter diversity uses multiple transmit antennas to provide the receiver with multiple uncorrelated replicas of the same signal. The obvious advantage is that the complexity of having multiple antenna is placed on the transmitter which may be shared among many receivers. For example, the forward (base-to-mobile) link in many wireless systems. The portable receivers can use just a single antenna and still benefit from a diversity gain.
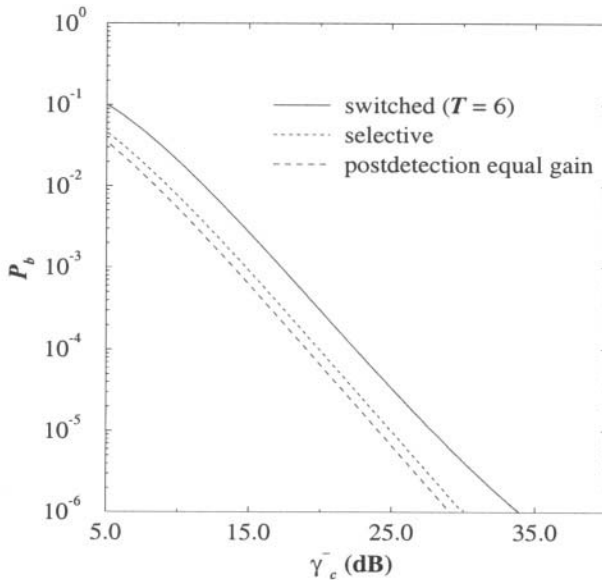
Figure 6.11. Comparison of 2-branch diversity combining techniques for differentially detected binary DPSK.

Transmitter diversity can take on many forms, distinguished by the method of using the multiple transmit antennas. Transmit diversity is straight forward for systems that use time division duplexing (TDD), where different time slots on the same carrier are used in the forward and reverse links, because of the reciprocity of the channel impulse response. At the base station the signals received on all antennas can be processed during every received burst. During the next forward burst, the antenna that provided the highest received symbol energy-to-noise ratio is selected and used. This is a form of selective transmit diversity (STD). Obviously, this scheme requires that the channel change slowly.

For frequency division duplexed (FDD) systems, transmit diversity is more complicated to implement, because the forward and reverse links are not reciprocal. Time division transmit diversity (TDTD) can be used for FDD by switching the transmitted signal between two or more transmit antennas. Alternate bursts are transmitted through two or more separate antennas, a technique known as time switched transmit diversity (TSTD). Another method is delay transmitter diversity, where copies of the same symbol are transmitted through multiple antennas at different times. This has the effect of creating artificial delay spread so the resulting channel looks like a fading ISI channel. An equalizer can then be used to recover the signal and provide a diversity gain.

More elaborate forms of transmit diversity use space-time or space-frequency encoding of the transmitted information. These scheme require three functions: (1) the encoding and transmission of the information sequence at the transmitter, (2) the combining scheme at the receiver, (3) the decision rule for making decisions. Alamouti [10] has introduced a simple repetition transmit diversity scheme with maximum likelihood combining at the receiver. By using two transmit antennas and one receiver antenna, the scheme provides the same diversity order as maximal ratio receiver combining with one transmit antenna and two receiver antennas. This scheme requires no feedback from the receiver to the transmitter, and requires no bandwidth expansion. However, to estimate the channel, the scheme requires separate pilot sequence insertion and extraction for each of the transmit diversity antennas.

The scheme proposed by Alamouti can be considered a form of space-time coding. More sophisticated forms of transmit diversity use specially designed space-time error correcting codes [298]. The data is encoded by a space-time encoder that chooses the transmitted signal constellation points so that the coding and diversity gains are maximized. The encoded data is split into n streams that are simultaneously transmitted by using n antennas. Bandwidth efficient space-time trellis codes have designed for PSK and QAM constellations [298]. These techniques have been applied to IS-136 with good results [241, 242]. Furthermore, space-time codes may be designed with multilevel structures, and multistage decoding can be useful when the number of transmit antennas is large. This enables us to significantly reduce the decoding complexity.

7.1   SPACE-TIME TRANSMIT DIVERSITY

Here we describe the scheme suggested by Alamouti [10] as an example of transmit diversity. The scheme uses two transmit antennas and one receiver antenna, referred to as 2×1 diversity. In any given baud period, two data symbols are transmitted simultaneously from the two transmit antennas. Suppose the symbols transmitted from Antennas 1 and 2 are denoted by $\tilde{\mathbf{s}}_{(1)}$ and $\tilde{\mathbf{s}}_{(2)}$, respectively. During the next baud period, the symbols transmitted from Antennas 1 and 2 are $-\tilde{\mathbf{s}}_{(2)}^{*}$ and $\tilde{\mathbf{s}}_{(1)}^{*}$, respectively. The channel gains for the two antennas are denoted by $g_1(t)$ and $g_2(t)$. If the channel stays constant over two baud intervals then we can write

$$g_k(t) = g_k(t+T) = g_k = \alpha_k e^{j\phi_k} \tag{6.59}$$

where T is the baud period. The received complex vectors are

$$
\begin{aligned}
\tilde{\mathbf{r}}_{(1)} &= g_1 \tilde{\mathbf{s}}_{(1)} + g_2 \tilde{\mathbf{s}}_{(2)} + \tilde{\mathbf{n}}_{(1)} \\
\tilde{\mathbf{r}}_{(2)} &= -g_1 \tilde{\mathbf{s}}_{(2)}^{*} + g_2 \tilde{\mathbf{s}}_{(1)}^{*} + \tilde{\mathbf{n}}_{(2)}
\end{aligned} \tag{6.60}
$$

Figure 6.12. Space-time diversity receiver for 2 × 1 diversity.

where $\tilde{\mathbf{r}}_{(1)}$ and $\tilde{\mathbf{r}}_{(2)}$ represent the received vectors at time t and $t + T$, respectively, and $\tilde{\mathbf{n}}_{(1)}$ and $\tilde{\mathbf{n}}_{(2)}$ are the corresponding noise vectors.

The diversity combiner for this scheme is shown in Fig. 6.12. The combiner constructs the following two signal vectors

$$
\begin{aligned}
\tilde{\mathbf{v}}_{(1)} &= g_1^* \tilde{\mathbf{r}}_{(1)} + g_2 \tilde{\mathbf{r}}_{(2)}^* \\
\tilde{\mathbf{v}}_{(2)} &= g_2^* \tilde{\mathbf{r}}_{(1)} - g_1 \tilde{\mathbf{r}}_{(2)}^*
\end{aligned}
\tag{6.61}
$$

Afterwards, the receiver applies the vectors $\tilde{\mathbf{v}}_{(1)}$ and $\tilde{\mathbf{v}}_{(2)}$ in a sequential fashion to the metric computer in Fig 6.4, to make decisions by maximizing the metric

$$
\begin{aligned}
\mu(\tilde{\mathbf{s}}_{(1),m}) &= \mathrm{Re}\left(\tilde{\mathbf{v}}_{(1)}, \tilde{\mathbf{s}}_{(1),m}\right) - E_m(|g_1|^2 + |g_2|^2) \\
\mu(\tilde{\mathbf{s}}_{(2),m}) &= \mathrm{Re}\left(\tilde{\mathbf{v}}_{(2)}, \tilde{\mathbf{s}}_{(2),m}\right) - E_m(|g_1|^2 + |g_2|^2)
\end{aligned}
\tag{6.62}
$$

Using (6.59) and (6.60) in (6.61) gives

$$
\begin{aligned}
\tilde{\mathbf{v}}_{(1)} &= (\alpha_1^2 + \alpha_2^2)\tilde{\mathbf{s}}_{(1)} + g_1^* \tilde{\mathbf{n}}_{(0)} + g_2 \tilde{\mathbf{n}}_{(1)}^* \\
\tilde{\mathbf{v}}_{(2)} &= (\alpha_1^2 + \alpha_2^2)\tilde{\mathbf{s}}_{(2)} - g_1 \tilde{\mathbf{n}}_{(2)}^* + g_2^* \tilde{\mathbf{n}}_{(1)}
\end{aligned}
\tag{6.63}
$$

This is to be compared with the output of the MRC metric computers in Fig. 6.4. With $L = 2$,

$$\begin{aligned}
\tilde{\mathbf{r}} &= g_1^* \tilde{\mathbf{r}}_1 + g_2^* \tilde{\mathbf{r}}_2 \\
&= (\alpha_1^2 + \alpha_2^2)\tilde{\mathbf{s}}_m + g_1^* \tilde{\mathbf{n}}_1 + g_2^* \tilde{\mathbf{n}}_2
\end{aligned} \tag{6.64}$$

Comparison of (6.63) and (6.64) shows that the combined signals in each case are the same. The only difference is the phase rotations of the noise vectors which will not change the error probability due to their circular symmetry.

$2 \times L$ **diversity:.** We now consider the case of 2 transmit antennas and L receiver antennas, and show that the performance is equivalent to a receiver diversity order of $2L$. The results are illustrated for the case of 2×2 diversity, and the extension to $L \times L$ diversity will be obvious. To describe the scheme, we need to introduce the following notation

$$g_{j,i} = \text{channel gain between receiver antenna } j \text{ and transmit antenna } i.$$
$$\tilde{\mathbf{r}}_{(1),j} = \text{received signal at antenna } j \text{ at time } t.$$
$$\tilde{\mathbf{r}}_{(2),j} = \text{received signal at antenna } j \text{ at time } t + T.$$

The encoding scheme remains the same as before: symbols $\tilde{\mathbf{s}}_{(1)}$ and $\tilde{\mathbf{s}}_{(2)}$ are transmitted from from Antennas 1 and 2 at time t, and symbols $-\tilde{\mathbf{s}}_{(2)}^*$ and $\tilde{\mathbf{s}}_{(1)}^*$ are transmitted from Antennas 1 and 2 at time $t + T$. The received signals are

$$\begin{aligned}
\tilde{\mathbf{r}}_{(1),1} &= g_{1,1}\tilde{\mathbf{s}}_{(1)} + g_{2,1}\tilde{\mathbf{s}}_{(2)} + \tilde{\mathbf{n}}_{(1),1} \\
\tilde{\mathbf{r}}_{(1),2} &= -g_{1,1}\tilde{\mathbf{s}}_{(2)}^* + g_{2,1}\tilde{\mathbf{s}}_{(1)}^* + \tilde{\mathbf{n}}_{(1),2} \\
\tilde{\mathbf{r}}_{(2),1} &= g_{1,2}\tilde{\mathbf{s}}_{(1)} + g_{2,2}\tilde{\mathbf{s}}_{(2)} + \tilde{\mathbf{n}}_{(2),1} \\
\tilde{\mathbf{r}}_{(2),2} &= -g_{1,2}\tilde{\mathbf{s}}_{(2)}^* + g_{2,2}\tilde{\mathbf{s}}_{(1)}^* + \tilde{\mathbf{n}}_{(2),2}
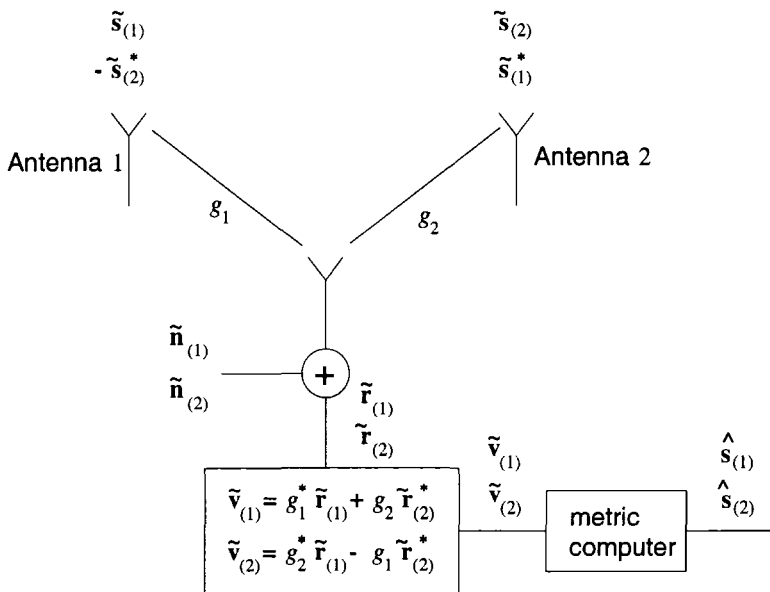\end{aligned}$$

The combiner shown in Fig. 6.13 constructs the following two signal vectors

$$\tilde{\mathbf{v}}_{(1)} = g_{1,1}^* \tilde{\mathbf{r}}_{(1),1} + g_{2,1}\tilde{\mathbf{r}}_{(1),2}^* + g_{1,2}^* \tilde{\mathbf{r}}_{(2),1} + g_{2,2}\tilde{\mathbf{r}}_{(1),1}^* \tag{6.65}$$

$$\tilde{\mathbf{v}}_{(2)} = g_{2,1}^* \tilde{\mathbf{r}}_{(1),1} - g_{1,1}\tilde{\mathbf{r}}_{(1),2}^* + g_{2,2}^* \tilde{\mathbf{r}}_{(2),1} - g_{2,1}\tilde{\mathbf{r}}_{(1),1}^* \tag{6.66}$$

Again, the receiver applies the vectors $\tilde{\mathbf{v}}_{(1)}$ and $\tilde{\mathbf{v}}_{(2)}$ to the metric computer in Fig 6.4 and decisions are made by maximizing the metric in (6.62).

To compare with MRC, we substitute appropriate equations to obtain

$$\begin{aligned}
\tilde{\mathbf{v}}_{(1)} &= (\alpha_{1,1}^2 + \alpha_{1,2}^2 + \alpha_{2,1}^2 + \alpha_{2,2}^2)\tilde{\mathbf{s}}_{(1)} \\
&\quad + g_{1,1}^* \tilde{\mathbf{n}}_{(1),1} + g_{2,1}\tilde{\mathbf{n}}_{(1),2}^* + g_{1,2}^* \tilde{\mathbf{n}}_{(2),1} + g_{2,2}\tilde{\mathbf{n}}_{(1),1}^* \tag{6.67}
\end{aligned}$$

$$\begin{aligned}
\tilde{\mathbf{v}}_{(2)} &= (\alpha_{1,1}^2 + \alpha_{1,2}^2 + \alpha_{2,1}^2 + \alpha_{2,2}^2)\tilde{\mathbf{s}}_{(2)} \\
&\quad + g_{2,1}^* \tilde{\mathbf{n}}_{(1),1} - g_{1,1}\tilde{\mathbf{n}}_{(1),2}^* + g_{2,2}^* \tilde{\mathbf{n}}_{(2),1} - g_{2,1}\tilde{\mathbf{n}}_{(1),1}^* \tag{6.68}
\end{aligned}$$

This is to be compare with the output of the MRC in Fig. 6.4. With $L = 4$,

$$\begin{aligned}
\tilde{\mathbf{r}} &= g_1^* \tilde{\mathbf{r}}_1 + g_2^* \tilde{\mathbf{r}}_2 + g_3^* \tilde{\mathbf{r}}_3 + g_4^* \tilde{\mathbf{r}}_4 \\
&= (\alpha_1^2 + \alpha_2^2 + \alpha_3^2 + \alpha_4^2)\tilde{\mathbf{s}}_m + g_1^* \tilde{\mathbf{n}}_1 + g_2^* \tilde{\mathbf{n}}_2 + g_3^* \tilde{\mathbf{n}}_3 + g_4^* \tilde{\mathbf{n}}_4 \tag{6.69}
\end{aligned}$$
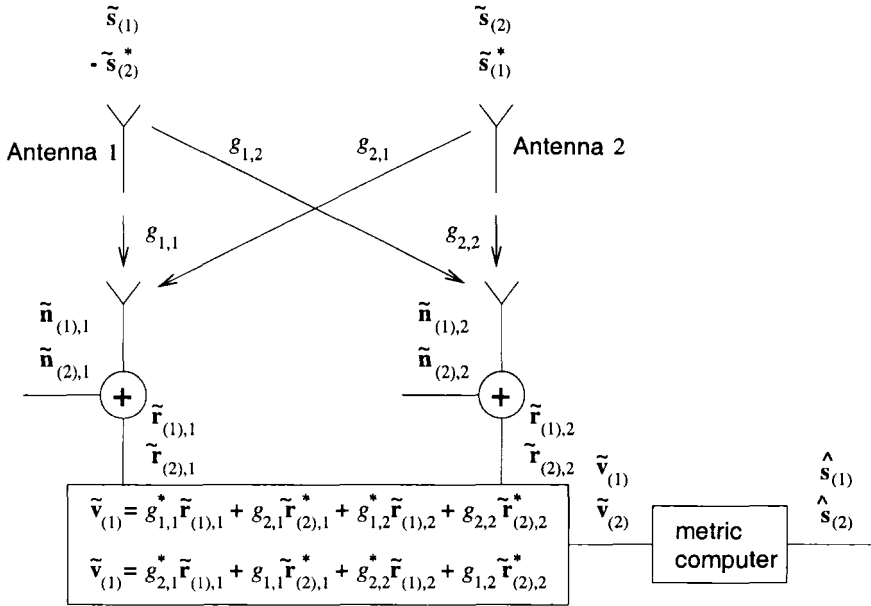
Figure 6.13. Space-time diversity receiver for 2 × 2 diversity.

Again, we see the 2×2 transmit diversity scheme is equivalent to a 4×1 diversity scheme with MRC. The extension to show that an $2 \times L$ transmit diversity scheme is equivalent to a $2L$ diversity scheme with MRC is obvious.

Implementation Issues:. There are several key implementation issues with the above transmit diversity scheme, including the following:

- Since there are 2 transmit antennas, the power per antenna must be halved to maintain a constant transmit power. This results in a 3 dB loss in performance compared to a single transmit antenna.

- With 2 transmit antennas, twice as many pilot symbols are needed compared to the case of one transmit antenna. The pilots must alternate between the antennas. Alternatively, orthogonal pilot sequences can be used.

- In order to achieve sufficient fading decorrelation on the diversity branches, the transmit antennas must be spaced at the same distance that would required if the same two antennas were to provide receiver diversity. We have seen in Chapter 2.1.5.1 that the required separation is several tens of wavelengths.

Problems

6.1. Consider a Rayleigh random variable, X, with a pdf given by (2.42) in the text.

a) Let $\{X_1, X_2, \ldots, X_N\}$ be a set of independent Rayleigh random variables each with an *rms* value of $1/\sqrt{N}$. Derive the pdf of

$$Y = \max\left(X_1^2, X_2^2, \ldots, X_N^2\right) \ .$$

This result is useful for the study of *selective combining* diversity systems.

b) Again, using the set $\{X_1, X_2, \ldots, X_N\}$, derive the pdf of

$$Z = X_1^2 + X_2^2 + \cdots + X_N^2 \ .$$

This result is useful for the study of *maximal ratio* combining diversity systems.

6.2. Suppose that two-branch selective combining is used. However, the branches are mismatched such that $\bar{\gamma}_1 \neq \bar{\gamma}_2$ where the $\bar{\gamma}_i, i = 1, \ 2$ are the average received symbol energy-to-noise ratios for the two branches. Plot the cdf of γ_s^s against the average normalized symbol energy-to-noise ratio $10\log_{10}(\gamma_s^s/\bar{\gamma}_t)$, where $\bar{\gamma}_t = (\bar{\gamma}_1 + \bar{\gamma}_2)/2$. Show several curves while varying the ratio $\xi = \bar{\gamma}_1/\bar{\gamma}_2$.

6.3. Consider using selective combining with coherent BPSK. For BPSK, the probability of bit error is $P_b(\gamma_b^s) = Q(\sqrt{2\gamma_b^s})$ The instantaneous bit energy-to-noise ratio is given by (6.8).

a) Derive an expression for the average bit error probability

$$P_b = \int_0^\infty P_b(x) p_{\gamma_s^s}(x) dx$$

b) Repeat part a) for two-branch switched diversity combining where the pdf of γ_s^{sw} is given by (6.47).

c) Plot and compare the results in parts a) and b) for two-branch diversity.

6.4. Suppose that binary DPSK signaling ($x_k \in \{-1, +1\}$) is used on a flat Rayleigh fading channel with 3-branch diversity. The diversity branches are assumed to experience uncorrelated fading. The signal that is received over each diversity branch is corrupted with AWGN having a one-sided psd of N_o watts/Hz. The noise processes that are associated with the diversity branches are mutually uncorrelated.

a Suppose that a separate differential detector is used on each diversity branch, yielding three independent estimates of each transmitted bit, i.e., for x_k the receiver generates the three independent estimates $(\hat{x}_k^1, \hat{x}_k^2, \hat{x}_k^3)$. Majority logic combining is then used to combine the three estimates together to yield the final decision $\hat{x}_k$, i.e.,

$$\hat{x}_k = \begin{cases} -1 & \text{if two or more } \hat{x}_k^i = -1 \\ +1 & \text{if two or more } \hat{x}_k^i = +1 \end{cases}$$

Find an expression for the probability of bit error, P_b. Evaluate P_b for $\bar{\gamma}_c = 20\,\text{dB}$, where $\bar{\gamma}_c$ is the average received branch bit energy-to-noise ratio.

b) Evaluate the probability of bit error for $\bar{\gamma}_c = 20\,\text{dB}$ if the receiver uses 3-branch diversity with postdetection equal gain combining. Compare with the result in part a).

c) Generalize the expression for the probability of bit error in part a) to L-branch diversity.

6.5. Derive (6.27) for BPSK and maximal ratio combining.

6.6. Derive (6.54) for DPSK with differential detection followed by equal gain combining.

6.7. The bit error probability MSK signaling on a Rayleigh fading channel with additive white Gaussian noise is

$$P_b = \frac{1}{2}\left[1 - \sqrt{\frac{\bar{\gamma}_b}{1 + \bar{\gamma}_b}}\right]$$

a) Derive a Chernoff upper bound (see Appendix A) on the probability of bit error for the same channel and compare the Chernoff bound with the exact error probability.

b) Repeat part a) if the receiver employs L-branch diversity. Assume uncorrelated diversity branches with $\bar{\gamma}_1 = \bar{\gamma}_2 = \cdots \bar{\gamma}_L = \bar{\gamma}_c$.

6.8. Suppose that BPSK modulation is used with two-branch diversity and coherent equal gain combining. Assume uncorrelated diversity branches with $\bar{\gamma}_1 = \bar{\gamma}_2 = \bar{\gamma}_c$. Show that the probability of bit error for a Rayleigh fading channel is given by (6.38).

6.9. Consider a system that uses L-branch selection diversity. The instantaneous received signal power on each diversity branch, $s_{0,i}$, $i = 1, \ldots, L$, has the non-central chi-square (Ricean fading) distribution in (2.51). The instantaneous received signal power from each interferer on each diversity branch, $s_{k,i}$, $i = 1, \ldots, L$ has the exponential (Rayleigh fading)

distribution in (2.44). Note that all the $s_{0,i}$ and $s_{k,i}$ are all independent. Let $\lambda_i = s_{0,i}/\sum_{k=1}^{N_I} s_{k,i}$, $i = 1, \ldots, L$ be the instantaneous carrier-to-interference ratio for each diversity branch and $\lambda_s = \max_i \lambda_i$. Derive an expression for the probability of co-channel interference outage

$$O_I = \mathrm{P}(\lambda_s < \lambda_{\mathrm{th}}) \ .$$

Plot O_I against λ_{th} for various L.

6.10. Consider a coherent MSK system that operates over a slow flat Rayleigh fading channel in the presence of a single flat Rayleigh faded co-channel interferer. The received carrier-to-interference ratio is $\lambda = s_0/s_1$, where s_0 and s_1 are independent exponentially distributed random variables with density in (2.44). The average signal-to-interference ratio is defined as

$$\bar{\lambda} \triangleq \frac{\bar{s}_0}{\bar{s}_1}$$

   a) Derive an expression for the probability density function of λ. What is the mean value of λ?

   b) To improve the bit error probability performance, L-branch antenna diversity is employed at the receiver. Assume that the λ_i experienced on each of the diversity branches are independent. If selective diversity combining is used, what is the probability density function of the signal-to-interference ratio at the output of the selective combiner?

6.11. Suppose that two-branch antenna diversity is used with selective combining. However, the branches have correlated fading so that the maximum diversity gain is not achieved. Let γ_1 and γ_2 be the joint pdf for the instantaneous bit energy-to-noise ratio for each diversity branch, and let $\bar{\gamma}_c = \mathrm{E}[\gamma_i]$. It is known that joint pdf of γ_1 and γ_2 is

$$p_{\gamma_1,\gamma_2}(x_1, x_2) = \frac{1}{\bar{\gamma}_c(1 - \rho^2)} I_0\left(\frac{2\rho\sqrt{x_1 x_2}}{\bar{\gamma}_c(1 - \rho^2)}\right) \exp\left\{-\frac{x_1 + x_2}{\bar{\gamma}_c(1 - \rho^2)}\right\}$$

where ρ is magnitude of the covariance of the two complex, jointly Gaussian random processes that are associated with each diversity branch. Derive an expression for the cdf of the output of the selective combiner

$$\gamma_s = \max\{\gamma_1, \gamma_2\} \ .$$

Plot the cdf for various ρ. What conclusions can you make?

This page intentionally left blank.

Chapter 7

EQUALIZATION
AND INTERFERENCE CANCELLATION

Land mobile radio channels are modeled as fading dispersive channels, because of the multipath propagation and the randomly changing medium characteristics. Many types of impairments are observed on these channels such as multipath spread (or delay spread), fading, Doppler spread, nonlinear distortion, frequency offset, phase jitter, impulse noise, thermal noise, and co-channel and adjacent channel interference arising from spectrum sharing. This chapter concentrates on the effects of delay spread, fading, Doppler spread, thermal noise, and co-channel interference. Delay spread causes interference between adjacent symbols, known as intersymbol interference (ISI), a large Doppler spread indicates rapid channel variations and necessitates a fast convergent algorithm when an adaptive receiver is employed, and fading results in a very low received signal-to-noise ratio or signal-to-interference ratio when the channel exhibits a deep fade.

An adaptive equalizer is an arrangement of adjustable filters at the receiver whose purpose is to mitigate the combined effect of ISI and noise [210, 274]. Two broad categories of equalizers have been documented extensively in the literature; symbol-by-symbol equalizers and sequence estimators. Symbol-by-symbol equalizers include a decision device to make symbol-by-symbol decisions on the received symbol sequence, while sequence estimators make decisions on sequences of received symbols. Many structures and adaptive algorithms have been proposed for each type of equalizer for different channel characteristics. Sequence estimators are generally more complex than symbol-by-symbol equalizers, but can potentially offer better performance.

This chapter begins with a brief survey of adaptive equalization techniques. This is followed by a discussion of ISI channel modeling in Section 2.. The optimum receiver for digital signaling on an ISI channel is presented in Section 3.. Section 4. provides a treatment of symbol-by-symbol equalizers and Section 5.

provides a treatment of sequence estimators. Section 6. provides an analysis of the bit error rate performance of maximum likelihood sequence estimation (MLSE) on static ISI channels and multipath fading ISI channels. Finally, Section 7. analyzes the performance of fractionally-spaced MLSE receivers on ISI channels.

1. OVERVIEW

1.1 SYMBOL-BY-SYMBOL EQUALIZERS

Lucky [208, 209] was the first to develop an adaptive (linear) equalizer for digital communication systems in the mid-1960s. This equalizer was based on the peak distortion criterion, where the equalizer forces the ISI to zero, and it is called a **zero-forcing** (ZF) equalizer. Soon after, Proakis and Miller [271], Lucky *et. al.* [210], and Gersho [134] developed the linear **LMS equalizer**, based on the least mean square (LMS) criterion. The LMS equalizer is more robust than the ZF equalizer, because the latter ignores the effects of noise. Thaper [319] examined the performance of trellis coded modulation for high speed data transmission on voiceband telephone channels, and proposed a simple receiver structure that used an adaptive linear equalizer. He reported that the performance was close to ideal, but his work did not include the more severely distorted multipath fading ISI channels.

Linear equalizers have the drawback of enhancing channel noise while trying to eliminate ISI, a characteristic known as **noise enhancement**. As a result, satisfactory performance is unattainable with linear equalizers for channels having severe amplitude distortion. In 1967, Austin [20] proposed the nonlinear **decision feedback equalizer** (DFE) to mitigate noise enhancement. Because only the precursor ISI is eliminated by the feedforward filter of the DFE, noise enhancement is greatly reduced. To eliminate the postcursor ISI, the estimated symbols are fed back through the feedback filter of the DFE. However, this introduces error propagation which can seriously degrade the performance of the DFE and complicate analysis of its performance. Belfiore and Park [29] proposed an equivalent DFE, called a **predictive DFE**, by using a linear predictor as the feedback filter. This structure is useful when a DFE is combined with a sequence estimator for equalization and decoding of trellis-coded modulation on an ISI channel [107].

Early adaptive equalizers were implemented by using a transversal filter with a tap-spacing equal to the signal interval, *T*, known as symbol-spaced equalizers. The performance of a symbol-spaced equalizer is very sensitive to the sampling instant and can be very poor with an improperly chosen sampling time [42, 329, 140]. Even with perfect timing and matched filtering, the symbol-spaced equalizer cannot realize the optimal linear receiver because of the finite tapped delay line structure. Brady [42], Monson [230], Ungerboeck

[329], and Gitlin and Weinstein [140] solved this problem by proposing a **fractionally-spaced equalizer** (FSE), where the tap-spacing is less than T. If a symbol-spaced equalizer is preceded by a matched filter, then an FSE and a symbol-spaced equalizer are equivalent. However, the exact matched filter is difficult to obtain in practical applications because its structure depends on the unknown channel characteristics and, hence, an FSE is quite attractive. It can also be argued that the FSE can achieve an arbitrary linear filter with a finite-length fractionally-spaced tapped delay line. Hence, the FSE is expected to outperform a (finite-length) symbol-spaced equalizer even with ideal matched filtering and sampling.

In the 1980's, Gersho and Lim [135], Mueller and Salz [237], and Wesolowski [354] proposed an interesting decision-aided equalizer, known as an **ISI canceller.** Theoretically, ISI cancelers can eliminate ISI completely without any noise enhancement. However, a decision-aided mechanism is employed in the equalizer so that it suffers from error propagation, similar to a DFE.

Various adaptation algorithms have been proposed to adjust the equalizer coefficients. The **LMS algorithm**, proposed by Widrow *et. al.* [357], and analyzed by Gitlin *et. al.* [139], Mazo [220, 221], Ungerboeck [327], and Widrow *et. al.* [358], is the most popular because of its simplicity and numerical stability. However, the LMS algorithm converges very slowly for channels with severe amplitude distortion. This slow convergence is intolerable for many practical applications. For example, Hsu *et. al.* [171] reported that the LMS algorithm is not suitable for an HF shortwave ionospheric channel, because the channel has severe amplitude distortion when a deep fade occurs and the channel characteristics change very rapidly.

A considerable research effort has been directed to finding a fast-convergent algorithms for adaptive equalizers. In 1974, Gordard [146] described a fast-convergent algorithm later known as the **recursive least square** (RLS) algorithm. This algorithm utilizes all available information from the beginning of processing, and converges much faster than the LMS algorithm. Unfortunately, the computational complexity is proportional to N^2, where N is the order of the equalizer, which is too high for many practical applications. To reduce the complexity, Falconer and Ljung [113], and Cioffi and Kailath [58] developed different fast RLS algorithms in 1978 and 1984, respectively. These algorithms have a complexity proportional to the equalizer order N. However, when the algorithms are implemented with finite precision arithmetic, they tend to become unstable. Examples of this numerical instability were reported by Mueller [236].

Another RLS algorithm, called the **recursive least square lattice** (RLSL) algorithm, was investigated by Morf *et. al.* [232], Satorius [291, 292], Friedlander [124], and Ling and Proakis [200]. The RLSL algorithm has a higher complexity that the fast RLS algorithms, but has better numerical stability.

However, numerical instability of the RLSL algorithm was still reported by Perl *et. al.* [260].

Some applications of symbol-by-symbol equalization techniques to multi-path fading channels were studied by Monson [230, 231], *Hsu et. al.* [171], Ling and Proakis [201], and Eleftheriou and Falconer [99]. For rapidly time-varying channels, a reinitialization procedure might be needed for fast-convergent algorithms in order to avoid numerical instability [97]. Finally, Wong and McLane [363] examined the performance of trellis-coded modulation for HF radio channels having in-band spectral nulls. They considered both linear and non-linear equalization and proposed a modified DFE (MDFE).

1.2   SEQUENCE ESTIMATION

The Viterbi algorithm was originally devised by Viterbi for maximum likelihood decoding of convolutional codes [341, 342]. Forney recognized the analogy between an ISI channel and a convolutional encoder, and applied the Viterbi algorithm for the detection of digital signals corrupted by ISI and additive white Gaussian noise [127]. Because of the efficiency of the Viterbi algorithm, the implementation of optimum **maximum likelihood sequence estimation** (MLSE) for detecting ISI-corrupted signals is feasible.

After Forney's initial work [127], the MLSE receiver was modified and extended. Magee and Proakis [215] proposed an adaptive MLSE receiver that employed an adaptive channel estimator for estimating the channel impulse response. Ungerboeck [328] developed a simpler MLSE that also accounted for the effect of carrier phase errors and sampling time errors. Acampora [6] used MLSE for combining convolutional decoding and equalization, and extended the application of MLSE to quadrature amplitude modulation (QAM) systems [7].

MLSE has a complexity that grows exponentially with the size of signal constellation and the length of channel impulse response. MLSE is impractical for systems having a large signal constellation and/or having a long channel impulse response. Considerable research has been undertaken to reduce the complexity of MLSE while retaining most of its performance. Early efforts concentrated on shaping the original channel impulse response into the one having a shorter length. Then a sequence estimator with a smaller number of states can be applied. In [275], Qureshi and Newhall employed a linear equalizer as the shaping filter. This method is quite successful if the original channel and the desired channel have a similar channel spectrum. Falconer and Magee [112], and Beare [27] adaptively optimized the linear equalizer and the desired channel response, by minimizing the mean square error between the output of the equalizer and the desired channel. This scheme has improved performance when the original channel is quite different from the desired one, but it has a higher complexity. As mentioned earlier, linear equalizers enhance

the channel noise. Lee and Hill [191] proposed using a DFE to truncate the channel impulse response so as to reduce the system complexity and mitigate noise enhancement.

Another approach for reducing the complexity of MLSE lies in simplifying the Viterbi algorithm itself. By employing suitable decision regions, Vermuelen and Hellman [337] and Foschini [120] observed that only a small number of likely paths need to be extended to obtain a near maximum likelihood performance. Wesolowski [355] employed a DFE to determine a small set of likely signal points, and then used the Viterbi algorithm to find the most likely sequence path through a reduced-state trellis. Clark *et. al.* [60] and Clark and Clayden [61] also proposed some similar detection methods.

Recently, two novel reduced-state sequence estimation techniques have been proposed. **Eyuboğlu** and Qureshi [109] proposed **reduced-state sequence estimation** (RSSE), a technique that is especially useful for systems with large signal constellations. Duel-Hallen and Heegard [90, 89] proposed **delayed decision-feedback sequence estimation** (DDFSE), a technique that is useful for channels with long impulse responses (DDFSE can be applied on channels with an infinite impulse response). Chevillat and Eleftheriou [49] independently proposed the same algorithm, but for a finite length channel. Both RSSE and DDFSE use the Viterbi algorithm to search for the most likely path, and provide a good performance/complexity trade-off. In both schemes, a feedback mechanism must be introduced to compute the branch metrics, because of the reduction in the number of system states. This feedback introduces error propagation. However, the effect of the error propagation is much smaller than with a DFE [109, 90]. **Eyuboğlu** and Qureshi [109] also observed that for channels with a finite channel impulse response, DDFSE can be conveniently modeled as a special case of RSSE. **Eyuboğlu** [107] and Chevillat and Eleftheriou [49] also suggested using RSSE for systems employing trellis-coded modulation. Sheen and Stüber have obtained error probability upper bounds and approximations for RSSE and DDFSE for uncoded systems [301] and trellis-coded systems [302].

Eyuboğlu and Forney [108] proposed a combined precoding and coded modulation technique that achieves the best coding gain of any known trellis code. With their technique, equalization is achieved by using Tomlinson-Harashima precoding [108], which requires that the channel impulse response be known at the transmitter.

For decoding convolutional codes, a sequential decoding algorithm is a good alternative to the Viterbi algorithm, especially when the encoder has a long constraint length and the system has a moderate-to-high SNR [199]. It is apparent that **sequential sequence estimation** (SSE) can be applied for detecting ISI-corrupted signals. Long and Bush [206, 205], and Xiong *et. al.* [368] reported some results on this application. In [206, 205], the Fano algorithm [114] was

employed as the detection algorithm, and a DFE was used to determine the path to be extended. If the DFE makes correct decisions most of the time, then the number of nodes visited by the Fano algorithm can be reduced. The multiple stack algorithm [48] was employed in [368] for avoiding the erasure or buffer overflow problem encountered with sequential detection algorithms. Systems with an infinite impulse response were also considered in [368].

Applications of sequence estimation techniques to multipath fading ISI channels were studied by D'aria and Zingarelli [76], D'avella *et. al.* [77], and Eleftheriou and Falconer [99]. MLSE was employed for equalizing UHF land mobile radio channels in [76, 77], and employed for equalizing HF shortwave ionospheric channels in [99]. Tight upper bound on the error probability of digital signaling on fading ISI channels with MLSE have been provided by Sheen and Stüber for uncoded systems [300] and trellis-coded systems [303]. Katz and Stüber [285] have applied SSE for the detection of trellis-coded signals on multipath fading ISI channels.

1.3   CO-CHANNEL INTERFERENCE CANCELLATION

The spectral efficiency of TDMA cellular systems, such as IS-54/136 and GSM, is limited primarily by co-channel interference (CCI). Several approaches may be used to combat CCI. Adaptive antenna arrays that use beam and null steering principles are one solution. Co-channel interference cancellation (CCIC) is another approach where signal processing techniques are used to cancel the CCI. CCIC receivers can use either a single antenna or multiple antennas. Single antenna CCIC receivers treat the channel as a multiple-input single-output (MISO) channel. The problem in this case is very similar to CDMA multiuser detection. However for narrowband systems, such as IS-54/136, the CCIC receivers usually employ multiple receiver antennas to gain additional degrees of freedom. The use of multiple receiver antennas creates a multiple-input multiple-output (MIMO) channel.

Winters [361, 362] suggested an optimum linear minimum-mean square error (MMSE) combining technique for flat fading channels with CCI. The optimum combiner jointly combats the effects of fading and CCI through digital beam forming with an *M*-element spatial diversity combiner. For the case of two antenna elements, direct matrix inversion (DMI) was suggested as a means of updating the antenna weighting coefficients. Optimum linear combining has the disadvantage that it cannot combat ISI. Co-channel interference and ISI can be jointly combated by using symbol-by-symbol equalization techniques. DFE-based approaches have been suggested by Duel-Hallen [91], Tidestav *et al.* [320] and Uesugi *et al.* [326].

By extending Forney's maximum likelihood receiver [127], Van Etten [103] proposed MLSE for joint maximum likelihood sequence estimation. Variations of the MLSE approach have been suggested by many authors, including

Giridhar *et al.* [138], Yoshino *et al.* [375], Yokota *et al.* [374], and Ranta *et al.* [277]. Bottomley and Jamal [41] have developed a scheme that combines adaptive antenna arrays and MLSE equalization. CCIC is performed in the Viterbi metric and the receiver is equivalent to Winter's optimum linear combiner under flat fading channel conditions. This work was extended by Molnar and Bottomley to a receiver that uses horizontal and vertical polarized antenna arrays [229]. The polarization diversity increases the diversity gain against fading that is lost when the Viterbi branch metric is modified for the purpose of CCIC. An interesting modification to the receiver in [41] was very recently proposed by Bottomley and Molnar [40], where CCIC is used for both channel and data estimation.

Finally, we note that CDMA multi-user detection techniques can be readily extended to narrow-band to perform CCIC in narrow-band TDMA systems. In many cases, the mathematical framework is the same or very similar. The optimum multiuser detector for asynchronous CDMA systems was developed by Verdú [336]. A variety of less complex suboptimal CDMA multi-user detectors have also been developed, including the decorrelator detector [211, 212] linear MMSE detectors [367], non-linear decision feedback detectors [92, 93], and multi-stage detectors [334, 335].

2.   MODELING OF ISI CHANNELS

Chapter 4 showed that the complex envelope of any modulated signal can be expressed in the general form

$$\tilde{s}(t) = A \sum_n b(t - nT, \mathbf{x}_n) \ . \tag{7.1}$$

This chapter restricts attention to linear modulation schemes where

$$b(t, \mathbf{x}_n) = x_n h_a(t) \tag{7.2}$$

$h_a(t)$ is the amplitude shaping pulse, and $\{x_n\}$ is a complex symbol sequence. In general, ASK and PSK waveforms are included, but most FSK waveforms are not.

Suppose that the signal in (7.2) is transmitted over a channel having a time-invariant complex low-pass impulse response $g(t)$. The received complex envelope is

$$\tilde{r}(t) = \sum_n x_n h(t - nT) + \tilde{n}(t) \tag{7.3}$$

where

$$h(t) = \int_{-\infty}^{\infty} h_a(\tau) g(t - \tau) d\tau \tag{7.4}$$

is the convolution of the transmitted pulse $h_a(t)$ and the channel impulse response $g(t)$, and $\tilde{n}(t)$ is a zero-mean complex additive white Gaussian noise
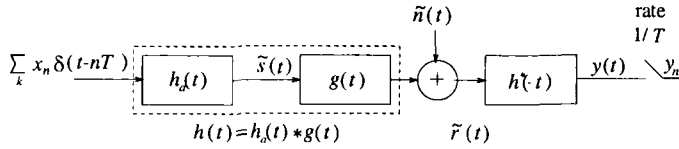
Figure 7.1. Digital signaling on an ISI channel. The receiver implements a filter that is matched to the received pulse followed by a symbol rate sampler.

(AWGN) with a power spectral density of N_o watts/Hz. Since the physical channel is causal, the lower limit of integration in (7.4) can be replaced by zero, resulting in

$$h(t) = \int_0^\infty h_a(\tau)g(t - \tau)d\tau \qquad t \geq 0 . \qquad (7.5)$$

Finally, the overall pulse $h(t)$ is assumed to have a finite duration so that $h(t) = 0$ for $t < 0$ and $t > (L + 1)T$, where L is some positive integer. We will show in Section 3. that the maximum likelihood receiver consists of an analog filter $h^*(-t)$ that is matched to the received pulse $h(t)$, followed by a symbol- or T-spaced sampler. Assuming that a matched filter has been implemented, the complex low-pass signal at the output of the matched filter is

$$y(t) = \sum_n x_n f(t - nT) + \nu(t) \qquad (7.6)$$

where

$$f(t) = \int_{-\infty}^\infty h^*(\tau)h(\tau + t)d\tau \qquad (7.7)$$

and

$$\nu(t) = \int_{-\infty}^\infty h^*(\tau)z(t - \tau)d\tau \qquad (7.8)$$

is the filtered noise. Note that the overall pulse response $f(t)$ accounts for the transmit filter, channel, and receive filter. The overall system as described above is shown in Fig. 7.1.

Sampling the matched filter output every T seconds yields the sample sequence

$$\begin{aligned} y_k = y(kT) &= \sum_n x_n f(kT - nT) + \nu(kT) \\ &\equiv \sum_n x_n f_{k-n} + \nu_k \\ &= x_k f_0 + \sum_{n \neq k} x_n f_{k-n} + \nu_k \qquad (7.9) \end{aligned}$$

where $f_n = f(nT)$ and $\nu_n = \nu(nT)$. The first term in (7.9) is the desired term, the second term is the ISI, and the last term is the noise at the output of
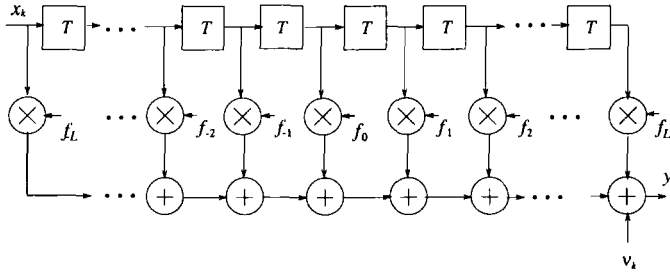
Figure 7.2.   Discrete-time model for digital signaling on an ISI channel.

the matched filter. It follows that the overall discrete-time system in Fig. 7.1 can be represented by a discrete-time transversal filter with coefficients

$$\mathbf{f} = (f_{-L},\ f_{-L+1},\ \dots,\ f_{-1},\ f_0,\ f_1,\ \dots,\ f_{L-1},\ f_L)\ . \tag{7.10}$$

This representation is depicted in Fig. 7.2.

From (7.9), the condition for ISI-free transmission is

$$f_n = \delta_{n0} f_0 \tag{7.11}$$

in which case

$$y_k = x_k f_0 + \nu_k\ . \tag{7.12}$$

Chapter 4.2 shows that the pulse $f(t)$ satisfies $f_n = \delta_{n0} f_0$ if and only if

$$F_\Sigma(f) \triangleq \frac{1}{T} \sum_{n=-\infty}^{\infty} F\left(f + \frac{n}{T}\right) = f_0\ . \tag{7.13}$$

That is, it is sufficient and necessary that the folded spectrum $F_\Sigma(f)$ be flat. For ISI-free transmission, the pulse $f(t)$ can be any function that has equally spaced zero crossings.

2.1   VECTOR REPRESENTATION OF RECEIVED SIGNALS

As discussed in Chapter 5.1, a Gram-Schmidt orthonormalization procedure can be used to express the received signal as

$$\tilde{r}(t) = \sum_{n=0}^{N-1} \tilde{r}_n \varphi_n(t) + \tilde{z}(t) \tag{7.14}$$

where the $\{\varphi_n(t)\}$ form a complete set of complex orthonormal basis functions defined over the interval $(0, T)$ and

$$\tilde{z}(t) = \tilde{n}(t) - \sum_{n=0}^{N-1} \tilde{n}_n\ . \tag{7.15}$$

It can be readily shown that

$$\tilde{r}_k = \sum_n x_n h_{k_n} + \tilde{n}_k \tag{7.16}$$

where[1]

$$h_{k_n} = \int_0^T h(t - nT)\varphi_k^*(t)dt$$

$$\tilde{n}_k = \int_0^T \tilde{n}(t)\varphi_k^*(t)dt \ . \tag{7.17}$$

Since the $\tilde{n}_n$ are complex Gaussian random variables with zero-mean and covariance $\frac{1}{2}\mathrm{E}[\tilde{n}_k^*\tilde{n}_m] = N_o\delta_{km}$, the vector $\tilde{\mathbf{r}} = (\tilde{r}_1, \tilde{r}_2, \cdots, \tilde{r}_N)$ has the multivariate Gaussian distribution

$$p(\tilde{\mathbf{r}}|\mathbf{x}, \mathbf{H}) = \prod_{k=0}^{N-1} \frac{1}{2\pi N_o} \exp\left\{ -\frac{1}{2N_o}\left| \tilde{r}_k - \sum_n x_n h_{k_n} \right|^2 \right\} \tag{7.18}$$

where

$$\mathbf{H} = (\mathbf{h}_0, \ \mathbf{h}_2, \ \ldots, \ \mathbf{h}_{N-1})^T \tag{7.19}$$

and

$$\mathbf{h}_k = (h_{k_0}, h_{k_1}, \ldots, h_{k_L}) \ . \tag{7.20}$$

3. OPTIMUM RECEIVER FOR ISI CHANNELS WITH AWGN

The maximum likelihood receiver decides in favor of the symbol sequence $\mathbf{x}$ that maximizes the likelihood function $p(\mathbf{w}|\mathbf{x}, \mathbf{H})$ or the log-likelihood function $\log p(\mathbf{w}|\mathbf{x}, \mathbf{H})$, i.e.,

$$\text{choose } \mathbf{x} \text{ if} \quad \log p(\mathbf{w}|\mathbf{x}, \mathbf{H}) > \log p(\mathbf{w}|\hat{\mathbf{x}}, \mathbf{H}) \quad \forall \hat{\mathbf{x}} \neq \mathbf{x} \ . \tag{7.21}$$

For an AWGN channel, $p(\mathbf{w}|\mathbf{x}, \mathbf{H})$ has the form in (7.18) and the decision rule in (7.21) is equivalent choosing $\mathbf{x}$ to maximize the quantity

$$
\begin{aligned}
\mu(\mathbf{x}) &= -\sum_{k=0}^{N-1} \left| \tilde{r}_k - \sum_n x_n h_{k_n} \right|^2 \\
&= -\sum_{k=0}^{N-1} |\tilde{r}_k|^2 + \sum_{k=0}^{N-1} \left(\tilde{r}_k^* \sum_n x_n h_{k_n} + \tilde{r}_k \sum_n x_n^* h_{k_n}^* \right) \\
&\quad - \sum_{k=0}^{N-1} \left(\sum_n x_n h_{k_n} \right) \left(\sum_m x_m^* h_{m_n}^* \right) \ .
\end{aligned}
\tag{7.22}
$$

---

[1] As in Chapter 5.1, we assume that $h(t)$ has duration T, although the development applies to longer pulses as well.

Since the term $\sum_{k=0}^{N-1} |\tilde{r}_k|^2$ is independent of $\mathbf{x}$, it may be omitted so that the maximum likelihood receiver chooses $\mathbf{x}$ to maximize

$$\mu(\mathbf{x}) = 2\mathrm{Re}\left\{\sum_n x_n^* \sum_{k=0}^{N-1} \tilde{r}_k h_{k_n}^*\right\} - \sum_n \sum_m x_n x_m^* \sum_{k=0}^{N-1} h_{k_n} h_{m_n}^* \qquad (7.23)$$

where $\mathrm{Re}\{z\}$ denotes the real part of z. To proceed further, note that

$$\lim_{N\to\infty} \sum_{k=0}^{N-1} \tilde{r}_k h_{k_n}^* = \int_{-\infty}^{\infty} \tilde{r}(\tau) h^*(\tau - nT) d\tau = y_n \qquad (7.24)$$

$$\lim_{N\to\infty} \sum_{k=0}^{N-1} h_{k_n} h_{m_n}^* = \int_{-\infty}^{\infty} h(\tau - nT) h^*(\tau - mT) d\tau = f_{m-n}(7.25)$$

where y_n and f_{m-n} were introduced earlier. The variables $\{y_n\}$ are obtained by passing the received low-pass signal $\tilde{r}(t)$ through the matched filter $h^*(-t)$ and sampling the output. Note that the T-spaced samples at the output of the matched filter must be obtained with the correct timing phase, and in the above development perfect symbol synchronization is implied. The $\{f_n\}$ are called the **ISI coefficients** and have the property that $f_n = f_{-n}^*$. By using (7.24) and (7.25) in (7.23) we have the final form

$$\mu(\mathbf{x}) = 2\mathrm{Re}\left\{\sum_n x_n^* y_n\right\} - \sum_n \sum_m x_n x_m^* f_{m-n} \quad . \qquad (7.26)$$

The noise samples at the matched filter output are, from (7.8),

$$\nu_n = \int_{-\infty}^{\infty} h^*(\tau) \tilde{n}(nT + \tau) d\tau \qquad (7.27)$$

and their discrete autocorrelation function is

$$\phi_{\nu\nu}(n) = \frac{1}{2}\mathrm{E}[\nu_{k+n}\nu_k^*] = N_o f_n \quad . \qquad (7.28)$$

3.1   DISCRETE-TIME WHITE NOISE CHANNEL MODEL

The correlation between the noise samples poses some complications when implementing the various equalization schemes. To overcome this difficulty, a **noise whitening filter** can be employed to process the sampled sequence $\{y_n\}$ as described below, resulting in an equivalent **discrete-time white noise channel model.** The z-transform of the vector $\mathbf{f}$ is

$$F(z) = \sum_{n=-L}^{L} f_n z^{-n} \quad . \qquad (7.29)$$

Using the property $f_n^* = f_{-n}$ we can write

$$F^*(1/z^*) = F(z) \ . \tag{7.30}$$

It follows that $F(z)$ has $2L$ roots with the factorization

$$F(z) = G(z)G^*(1/z^*) \tag{7.31}$$

where $G(z)$ and $G^*(1/z^*)$ are polynomials of degree L having conjugate reciprocal roots. There are 2^L possible choices for the roots of $G^*(1/z^*)$ and any one will suffice for a noise whitening filter. However, some equalization techniques such as RSSE and DDFSE require that the polynomial of the overall response $G(z)$ have **minimum-phase**. In this case, we can choose the unique $G(z)$ that has minimum phase, i.e., all the roots of $G(z)$ are inside the unit circle. With this choice of $G(z)$, the noise whitening filter $1/G^*(1/z^*)$ is a stable but noncausal filter. In practice, such an noncausal noise whitening filter can be implemented by using an appropriate delay. If the overall response $G(z)$ need not have minimum phase, then we can choose $G^*(1/z^*)$ to have minimum phase, i.e., all the roots of $G^*(1/z^*)$ are inside the unit circle. This choice ensures that the noise whitening filter $1/G^*(1/z^*)$ is both causal and stable.

Example 7.1 _____

Consider a simple T-spaced two-ray channel where the received pulse is

$$h(t) = h_a(t) + ah_a(t - T)$$

and the transmitted pulse $h_a(t)$ is normalized to have unit energy. The ISI coefficients are

$$
\begin{aligned}
f_n &= \int_{-\infty}^{\infty} h^*(t)h(t + nT)dt \\
&= \begin{cases} 1 + |a|^2 & n = 0 \\ a & n = 1 \\ a^* & n = -1 \end{cases}
\end{aligned}
$$

and, hence,

$$
\begin{aligned}
F(z) &= a^*z + (1 + |a|^2) + az^{-1} \\
&= (az^{-1} + 1)(a^*z + 1) \ .
\end{aligned}
$$

There are two possible choices for the noise whitening filter. Under the assumption that $|a| > 1$, suppose that the zero of $G^*(1/z^*)$ is chosen to be inside the unit circle. That is,

$$
\begin{aligned}
G(z) &= 1 + az^{-1} \\
G^*(1/z^*) &= 1 + a^*z \ .
\end{aligned}
$$

In this case, the noise whitening filter is stable and causal, and the overall system is characterized by the non-minimum phase polynomial

$$G(z) = 1 + az^{-1} \ .$$

Again, under the assumption that $|a| > 1$, suppose that the zero of $G^*(1/z^*)$ is chosen to be outside the unit circle. That is,

$$
\begin{aligned}
G(z) &= 1 + a^* z \\
G^*(1/z^*) &= 1 + az^{-1} \ .
\end{aligned}
$$

In this case, the noise whitening filter is stable and noncausal, and the overall system is characterized by the minimum phase polynomial

$$G(z) = 1 + a^* z \ .$$

For any choice of noise whitening filter, the filter output is

$$
\begin{aligned}
V(z) &= \left(X(z)F(z) + \nu(z) \right) \frac{1}{G^*(1/z^*)} \\
&= X(z)G(z) + \nu(z) \frac{1}{G^*(1/z^*)} \ .
\end{aligned}
\tag{7.32}
$$

From (7.28), the power spectral density of the noise at the input to the noise whitening filter is

$$S_{\nu\nu}(f) = N_o F(e^{j2\pi fT}), \qquad |f| \le \frac{1}{2T} \ . \tag{7.33}$$

Therefore, the power spectral density of the noise at the output of noise whitening filter is

$$
\begin{aligned}
S_{\eta\eta}(f) &= N_o \frac{F(e^{j2\pi fT})}{|G^*(e^{j2\pi fT})|^2} \\
&= N_o \frac{G(e^{j2\pi fT})G^*(e^{j2\pi fT})}{G(e^{j2\pi fT})G^*(e^{j2\pi fT})} \\
&= N_o, \qquad |f| \le \frac{1}{2T}
\end{aligned}
\tag{7.34}
$$

which is clearly white. The above development leads to the system shown in Fig. 7.3, with the equivalent discrete-time white noise channel model shown in Fig. 7.4. The discrete-time samples at the output of the noise whitening filter are

$$v_k = \sum_{n=0}^{L} g_n x_{k-n} + \eta_k \ . \tag{7.35}$$
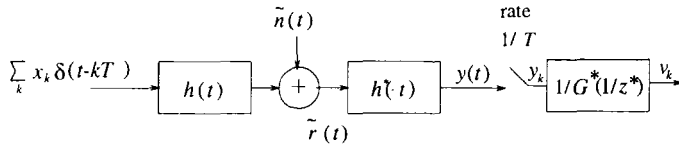
Figure 7.3. Block diagram of system that implements a filter matched to $h(t)$ followed by a discrete-time noise whitening filter.
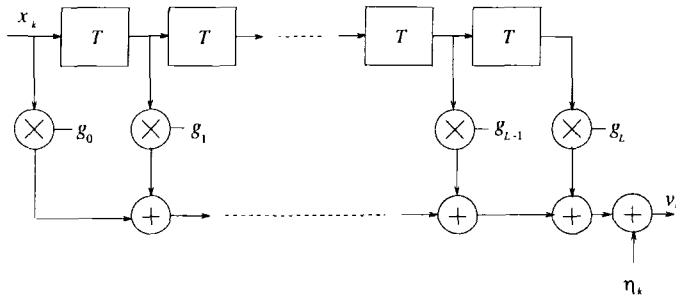
Figure 7.4. Discrete-time white noise channel model.

It follows that the effective overall channel impulse response can be described by the **channel vector**

$$\mathbf{g} = (g_0, \ g_1, \ldots, \ g_L)^T \ . \tag{7.36}$$

The symbol energy-to-noise ratio is

$$\gamma_s = \frac{E}{N_o} = \frac{\mathrm{E}[|x_k|^2] \sum_{i=0}^{L} |g_i|^2}{2N_o} \tag{7.37}$$

and the bit energy-to-noise ratio is $\gamma_b = \gamma_s / \log_2 M$ where M is the modulation alphabet size.

3.1.1   TIME VARYING CHANNELS WITH DIVERSITY

For time-varying channels with D-branch diversity, the corresponding discrete-time white noise channel model is shown in Fig. 7.5. At epoch k, the tap gains associated with diversity branch d are described by the vector

$$\mathbf{g}_d(k) = (g_{0,d}(k), \ g_{1,d}(k), \ \ldots \ g_{L,d}(k))^T \tag{7.38}$$

The $\{g_{i,d}(k)\}$ are discrete-time complex Gaussian random processes that are generally correlated with the covariance matrix

$$\Phi_{\mathbf{g}_d}(m) = \frac{1}{2}\mathrm{E}[\mathbf{g}_d(k+m)\mathbf{g}_d^H(k)] \tag{7.39}$$
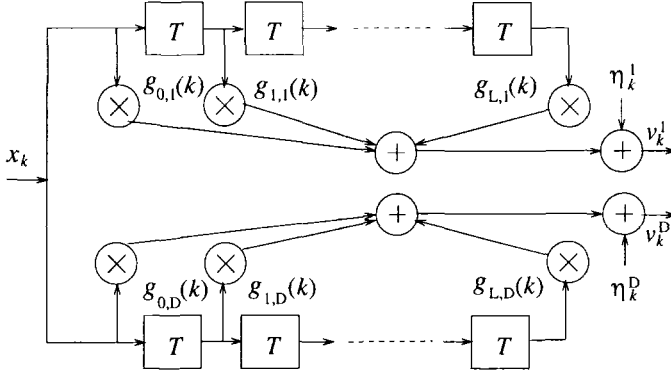
Figure 7.5. Discrete-time white noise channel model with D-branch diversity, from [300].

where H denotes Hermitian transposition. The received sample on branch d at epoch k is

$$v_{k,d} = \sum_{i=0}^{L} g_{i,d}(k)x_{k-i} + \eta_{k,d} \ . \tag{7.40}$$

The $\eta_{k,d}$ are independent complex zero-mean white Gaussian noise samples with variance $\frac{1}{2}\mathrm{E}[|\eta_{k,d}|^2] = N_o$. The average received symbol energy-to-noise ratio for branch d is

$$\bar{\gamma}_s^d = \frac{\bar{E}}{N_o} = \frac{\mathrm{E}[|x_k|^2]\sum_{i=0}^{L}\mathrm{E}[|g_{i,d}|^2]}{2N_o} \ . \tag{7.41}$$

In many cases, the branches are balanced so that $\bar{\gamma}_s^d = \bar{\gamma}_s$, $d = 1,\ldots,D$. The averaged received branch bit energy-to-noise ratio is $\bar{\gamma}_c = \bar{\gamma}_s/\log_2 M$.

3.1.2   *T/2*-SPACED RECEIVER

In practice the the matched filter outputs are often **oversampled** for the purpose of extracting timing information and to mitigate the effects of timing errors. One important example that will be considered at various points in this chapter is when the output of the receiver filter, $y(t)$, is sampled with rate $2/T$. In this case the overall channel impulse response and sampler can be represented by a $T/2$-spaced discrete-time transversal filter with coefficients

$$\mathbf{f}^{(2)} = (\ f_{-2L}^{(2)},\ f_{-2L+1}^{(2)},\ \cdots,\ f_{-1}^{(2)},\ f_0^{(2)},\ f_1^{(2)},\ \cdots,\ f_{2L-1}^{(2)},\ f_{2L}^{(2)}\ ) \ . \tag{7.42}$$

where $(\ \cdot\ )^{(2)}$ indicates rate $2/T$ sampling. If it so happens that the samples in (7.42) are obtained with the correct timing phase, i.e., $f_n^{(2)} = f(nT/2)$, then

$$\begin{aligned} \mathbf{f} &= (\ f_{-L},\ f_{-L+1},\ \cdots,\ f_{-1},\ f_0,\ f_1,\ \cdots,\ f_{L-1},\ f_L\ ) \\ &= (\ f_{-2L}^{(2)},\ f_{-2L+2}^{(2)},\ \cdots,\ f_{-2}^{(2)},\ f_0^{(2)},\ f_2^{(2)},\ \cdots,\ f_{2L-2}^{(2)},\ f_{2L}^{(2)}\ ) \end{aligned} \tag{7.43}$$
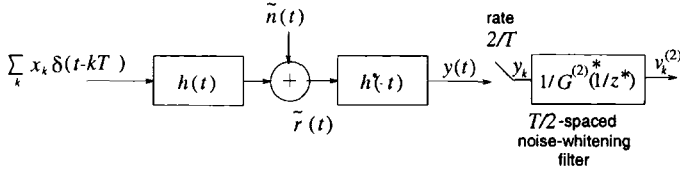
Figure 7.6. Block diagram of system that implements a filter matched to $h(t)$ followed by a $T/2$-spaced sampler and a $T/2$-spaced noise whitening filter.

where $f_n^{(2)} = \left(f_{-n}^{(2)}\right)^*$ and $f_n = f_{2n}^{(2)}$. More details on timing phase sensitivity will be provided in Section 7.4.

The $T/2$-spaced noise samples at the matched filter output have the autocorrelation

$$\phi_{\nu\nu}(n) = N_o f_n^{(2)} \ . \tag{7.44}$$

The z-transform of $\mathbf{f}^{(2)}$, denoted as $F^{(2)}(z)$, has $4L$ roots with the factorization

$$F^{(2)}(z) = G^{(2)}(z)(G^{(2)}(1/z^*))^* \tag{7.45}$$

where $G^{(2)}(z)$ and $(G^{(2)}(1/z^*))^*$ are polynomials of degree $2L$ having conjugate reciprocal roots. The $T/2$-spaced correlated noise samples can be whitened by using a filter with transfer function $1/(G^{(2)}(1/z^*))^*$. Once again, $(G^{(2)}(1/z^*))^*$ can be chosen such that all its roots are inside the unit circle, yielding a stable and causal noise whitening filter. On the other hand, we could choose the overall response $G^{(2)}(z)$ to have minimum phase, if necessary. The output of the noise whitening filter is

$$v_n^{(2)} = \sum_{k=0}^{2L} g_k^{(2)} x_{n-k}^{(2)} + \eta_n^{(2)} \tag{7.46}$$

where $\{\eta_n^{(2)}\}$ is a $T/2$-spaced white Gaussian noise sequence with variance $\frac{1}{2}\mathrm{E}[|\eta_n^{(2)}|^2] = N_o$ and the $\{g_n^{(2)}\}$ are the coefficients of a $T/2$-spaced discrete-time transversal filter having a transfer function $G^{(2)}(z)$. The sequence $\{x_n^{(2)}\}$ is the corresponding $T/2$-spaced input symbol sequence and is given by

$$x_n^{(2)} = \begin{cases} x_{n/2} \,, & n = 0, 2, 4, \ldots \\ 0 \,, & n = 1, 3, 5, \ldots \end{cases} \tag{7.47}$$

The overall system and equivalent discrete-time models are shown in Figs. 7.6 and 7.7, respectively.

Comparing (7.31) and (7.45), we have

$$\sum_{k=0}^{2L} |g_k^{(2)}|^2 = \sum_{k=0}^{L} |g_k|^2 = f_0^{(2)} = f_0 \ . \tag{7.48}$$
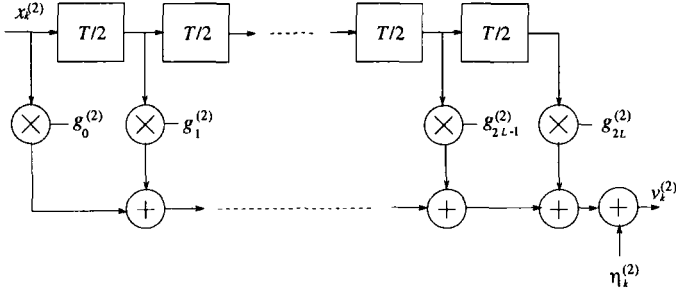
Figure 7.7.   Discrete-time white noise channel model with rate $2/T$ sampling.

Notice that the samples $v_{2n}^{(2)}$ and $v_{2n+1}^{(2)}$ correspond to the nth received baud, where

$$v_{2n}^{(2)} = \sum_{k=0}^{L} g_{2k}^{(2)} x_{n-k} + \eta_{2n}^{(2)} \tag{7.49}$$

$$v_{2n+1}^{(2)} = \sum_{i=0}^{L-1} g_{2k+1}^{(2)} x_{n-k} + \eta_{2n+1}^{(2)} . \tag{7.50}$$

Finally, by comparing (7.35) and (7.50), we note that $v_{2n}^{(2)}$ is not necessarily equal to v_n because a different noise whitening filter is used to whiten the $T/2$-spaced samples.

4.   SYMBOL-BY-SYMBOL EQUALIZERS

There are two broad categories of symbol-by-symbol equalizers, linear forward equalizers and nonlinear decision feedback equalizers.   As shown in Fig. 7.8, a linear forward equalizer consists of a transversal filter with adjustable tap coefficients.   The tap co-efficients of the equalizer are denoted by the column vector

$$\mathbf{c} = (c_0, c_1, \cdots, c_{N-1})^T \tag{7.51}$$

where N is the number of equalizer taps.   Assuming that the equalizer is preceded by a whitened matched filter that outputs the sequence $\{v_n\}$, the output of the equalizer is

$$\tilde{x}_n = \sum_{j=0}^{N-1} c_j v_{n-j} \tag{7.52}$$

where the v_n are given by (7.35). The equalizer output $\tilde{x}_k$ is quantized to the nearest (in Euclidean distance) information symbol to form the decision $\hat{x}_k$.
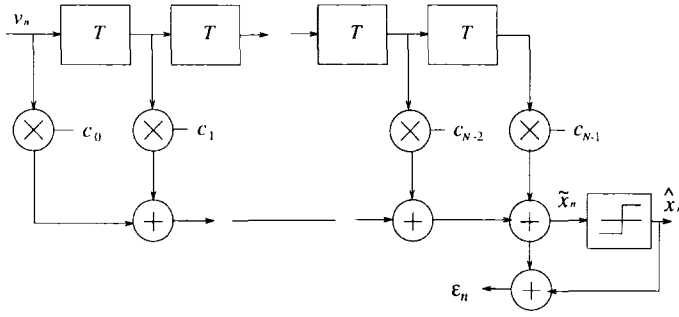
Figure 7.8. Linear transversal equalizer with adjustable T-spaced taps.

Observe that the overall discrete-time white noise channel and equalizer can be represented by a single filter having the sampled impulse response

$$q = (q_0, q_1, \ldots, q_{N+L-1})^T \tag{7.53}$$

where

$$
\begin{aligned}
q_n &= \sum_{j=0}^{N-1} c_j g_{n-j} \\
&= \mathbf{c}^T \mathbf{g}(n) \tag{7.54}
\end{aligned}
$$

with

$$\mathbf{g}(n) = (g_n, g_{n-1}, g_{n-2}, \ldots, g_{n-N+1})^T \tag{7.55}$$

and $g_i = 0, i < 0, i > L$. That is, $\mathbf{q}$ is the discrete convolution of $\mathbf{g}$ and $\mathbf{c}$.

Let the component of g of greatest magnitude be denoted by g_{d_1}. Note that any choice of noise whitening filter that does not result in an overall transfer function $G(z)$ with minimum phase may have $d_1 \neq 0$. Also, let the number of equalizer taps be equal to $N = 2d_2 + 1$ where d_2 is an integer. Perfect equalization means that

$$q = \mathbf{e}_d = (\underbrace{0, 0, \ldots, 0}_{d-1 \text{ zeroes}}, 1, 0, \ldots, 0, 0)^T \tag{7.56}$$

where $d - 1$ zeroes precede the "1" and d is an integer representing the overall delay. Unfortunately, perfect equalization is difficult to achieve and does not always yield the best performance.

4.1   LINEAR EQUALIZER

4.1.1   ZERO-FORCING (ZF)

With a zero-forcing (ZF) equalizer, the tap coefficients $\mathbf{c}$ are chosen to minimize the peak distortion of the equalized channel, denned as

$$D_{\mathrm{p}} = \frac{1}{|q_d|} \sum_{\substack{n=0 \\ n \neq d}}^{N+L-1} |q_n - \hat{q}_n| \tag{7.57}$$

where $\hat{q} = (\hat{q}_0, \ldots, \hat{q}_{N+L-1})^T$ is the *desired equalized channel* and the delay d is a positive integer optimized to have the value $d = d_1 + d_2$ [59]. Lucky showed that if the initial distortion without equalization is less than unity, i.e.,

$$D = \frac{1}{|g_{d_1}|} \sum_{\substack{n=0 \\ n \neq d_1}}^{L} |g_n| < 1 \,, \tag{7.58}$$

then D_p is minimized by those N tap values which simultaneously cause $q_j = \hat{q}_j$ for $d - d_2 \leq j \leq d + d_2$. However, if the initial distortion before equalization is greater than unity, the ZF criterion is not guaranteed to minimize the peak distortion. For the case when $\hat{q} = \mathbf{e}_d$ the equalized channel is given by

$$\mathbf{q} = (q_0, \ldots, q_{d_1-1}, 0, \ldots, 0, 1, 0, \ldots, 0, q_{d_1+N}, \ldots, q_{N+L-1})^T \,. \tag{7.59}$$

In this case the equalizer forces zeroes into the equalized channel and, hence, the name "zero-forcing equalizer."

Equalizer Tap Solution.   For a known channel impulse response, the tap gains of the ZF equalizer can be found by the direct solution of a simple set of linear equations [59]. To do so, we form the matrix

$$\mathbf{G} = [\mathbf{g}(d_1), \ldots, \mathbf{g}(d), \ldots, \mathbf{g}(N + d_1 - 1)] \tag{7.60}$$

and the vector

$$\check{q} = (\hat{q}_{d_1}, \ldots, \hat{q}_d, \ldots, \hat{q}_{N+d_1-1})^T \,. \tag{7.61}$$

Then the vector of optimal tap gains, $\mathbf{c}_{\mathrm{op}}$, satisfies

$$\mathbf{c}_{\mathrm{op}}^T \mathbf{G} = \check{q}^T \longrightarrow \mathbf{c}_{\mathrm{op}} = (\mathbf{G}^{-1})^T \check{q} \,. \tag{7.62}$$

Example 7.2.

Suppose that a system has the channel vector

$$\mathbf{g} = (0.90, -0.15, 0.20, 0.10, -0.05)^T \,,$$

where $g_i = 0, i < 0, i > 4$. The initial distortion before equalization is

$$D = \frac{1}{|g_0|} \sum_{n=1}^{4} |g_n| = 0.5555$$

and, therefore, the minimum distortion is achieved with the ZF solution. Suppose that we wish to design a 3-tap ZF equalizer. Since g_0 is the component of g having the largest magnitude, $d_1 = 0$ and the optimal equalizer delay is $d = 1$. Suppose that the desired response is $\hat{q}_0 = e_0$ so that $\tilde{q} = (1, 0, 0)$. We then construct the matrix

$$\begin{aligned} \mathbf{G} &= [\mathbf{g}(0), \mathbf{g}(1), \mathbf{g}(2)] \\ &= \begin{bmatrix} 0.90 & -0.15 & 0.20 \\ 0.00 & 0.90 & -0.15 \\ 0.00 & 0.00 & 0.90 \end{bmatrix} \end{aligned}$$

and obtain the optimal tap solution

$$\mathbf{c}_{op} = (\mathbf{G}^{-1})^T \tilde{\mathbf{q}} = (1.1111, 0.1852, -0.2160)^T .$$

The overall response of the channel and equalizer is

$$q = (1.0, 0.0, 0.0, 0.1806, -0.080, -0.031, -0.011, 0, \dots)^T .$$

Hence, the minimum distortion with this equalizer is

$$D_{\min} = \frac{1}{|q_0|} \sum_{n=1}^{6} |q_n - \hat{q}_n| = 0.30247 .$$

Adaptive Solution. In practice, the channel impulse response is unknown to the receiver and a known finite length sequence **x** is used to train the equalizer. During the training mode, the equalizer taps can be obtained by using the following steepest-descent recursive algorithm:

$$c_j^{n+1} = c_j^n + \alpha \epsilon_n x_{n-j-d_1}^* , \qquad j = 0, \dots, N-1 \qquad (7.63)$$

where

$$\begin{aligned} \epsilon_n &= x_{n-d} - \tilde{x}_n \\ &= x_{n-d} - \sum_{i=0}^{N-1} c_i v_{n-i} \end{aligned} \qquad (7.64)$$

is the error sequence, $\{c_j^n\}$ is the set of equalizer tap gains at epoch n, and α is an adaptation step-size that can be optimized to provide rapid convergence.

spectral density (psd)

$$S_{\nu\nu}(f) = N_o F(e^{j2\pi fT}), \qquad |f| \le \frac{1}{2T} . \qquad (7.70)$$

Therefore, the psd of the noise sequence $\{\zeta_n\}$ at the output of the equalizer is

$$S_{\zeta\zeta}(f) = \frac{N_o}{F(e^{j2\pi fT})}, \qquad |f| \le \frac{1}{2T} . \qquad (7.71)$$

and the noise samples have variance

$$
\begin{aligned}
\sigma_\zeta^2 &= T \int_{-1/2T}^{1/2T} S_{\zeta\zeta}(f) df \\
&= T \int_{-1/2T}^{1/2T} \frac{N_o}{F(e^{j2\pi fT})} df .
\end{aligned}
\qquad (7.72)
$$

If $\mathrm{E}[|x_k|^2] = 1$ and $\hat{q}_n = \delta_{n0}$, then the signal-to-noise ratio at the output of the infinite-tap equalizer is

$$\gamma_\infty = \frac{\mathrm{E}[|x_k|^2]}{\sigma_\zeta^2} = \frac{1}{\sigma_\zeta^2} . \qquad (7.73)$$

Finally, we can show that (see Problem 6.2)

$$F(e^{j2\pi fT}) = F_\Sigma(f), \qquad |f| \le \frac{1}{2T} \qquad (7.74)$$

where $F_\Sigma(f)$ is the folded spectrum of $F(f)$ defined in (7.13), and $F(f) = |H(f)|^2$ is the Fourier transform of the pulse $f(t) = h(t) * h^*(-t)$. It is clear from (7.72) that ZF equalizers are unsuitable for channels that have severe ISI, where the folded spectrum $F_\Sigma(f)$ has spectral nulls or very small values. Under these conditions, the equalizer tries to compensate for nulls in the folded spectrum by introducing infinite gain at these frequencies. Unfortunately, this results in severe noise enhancement at the output of the equalizer. Mobile radio channels often exhibit spectral nulls and, therefore, ZF equalizers are typically not used for mobile radio applications.

4.1.2   MINIMUM MEAN-SQUARE-ERROR (MMSE)

The minimum mean-square-error (MMSE) equalizer is more robust and superior to the ZF equalizer in its performance and convergence properties [271, 270, 274]. By defining the vector

$$\mathbf{v}_n = (v_n, v_{n-1}, \ldots, v_{n-N+1}) \qquad (7.75)$$

the output of the equalizer in (7.52) can be expressed in the form

$$\tilde{x}_n = \mathbf{c}^T \mathbf{v}_n = \mathbf{v}_n^T \mathbf{c} . \qquad (7.76)$$

A MMSE equalizer adjusts the tap coefficients to minimize the mean square error (MSE)

$$
\begin{aligned}
J &\triangleq \mathrm{E}[|x_{n-d} - \tilde{x}_n|^2] \\
&= \mathrm{E}\left[\mathbf{c}^T \mathbf{v}_n \mathbf{v}_n^H \mathbf{c}^* - 2\mathrm{Re}\{\mathbf{v}_n^H \mathbf{c}^* x_{n-d}\} + |x_{n-d}|^2\right] \quad.
\end{aligned}
\tag{7.77}
$$

Equalizer Tap Solution. If the channel impulse response is known, the optimum equalizer taps can be obtained by direct solution. Define

$$
\begin{aligned}
\mathbf{M}_v &\triangleq \mathrm{E}[\mathbf{v}_n \mathbf{v}_n^H] \\
\mathbf{v}_x^H &\triangleq \mathrm{E}[\mathbf{v}_n^H x_{n-d}]
\end{aligned}
\tag{7.78}
$$

where $\mathbf{M}_v$ is an $N \times N$ Hermitian matrix and $\mathbf{v}_x$ is a length N column vector. Using these definitions and assuming that $\mathrm{E}[|x_{n-d}|^2] = 1$, the mean-square-error is

$$
J = \mathbf{c}^T \mathbf{M}_v \mathbf{c}^* - 2\mathrm{Re}\{\mathbf{v}_x^H \mathbf{c}^*\} + 1 \quad.
\tag{7.79}
$$

The tap vector c that minimizes the mean square error can obtained by equating the gradient $\nabla_{\mathbf{c}} J$ to zero. It can be shown that (see Problem 6.15)

$$
\nabla_{\mathbf{c}} J = \left(\frac{\partial J}{\partial c_0}, \cdots, \frac{\partial J}{\partial c_{N-1}}\right) = 2\mathbf{c}^T \mathbf{M}_v - 2\mathbf{v}_x^H \quad.
\tag{7.80}
$$

Setting $\nabla_{\mathbf{c}} J = 0$ gives

$$
\mathbf{c}_{\mathrm{op}} = (\mathbf{M}_v^T)^{-1} \mathbf{v}_x^* \quad.
\tag{7.81}
$$

By using the identity $(\mathbf{A}^{-1})^T = (\mathbf{A}^T)^{-1}$ and the fact that $\mathbf{M}_v$ is Hermitian, the minimum mean-square-error (MMSE) is

$$
\begin{aligned}
J_{\min} &= \mathbf{c}_{\mathrm{op}}^T \mathbf{M}_v \mathbf{c}_{\mathrm{op}}^* - 2\mathrm{Re}\{\mathbf{v}_x^H \mathbf{c}_{\mathrm{op}}^*\} + 1 \\
&= 1 - \mathbf{v}_x^H \mathbf{M}_v^{-1} \mathbf{v}_x \quad.
\end{aligned}
\tag{7.82}
$$

Since the overall channel and equalizer can be represented as a single filter with impulse response $\mathbf{q}$ in (7.53) it follows that the MMSE can also be expressed in the form

$$
J_{\min} = \|\mathbf{q} - \hat{\mathbf{q}}\|^2 + N_o \|\mathbf{c}\|^2 \quad.
\tag{7.83}
$$

Example 7.3 ———————————————————————

Consider a system having the same channel vector as in Example 7.2. Suppose that we wish to design a 3-tap MSE equalizer. In this case

$$
\mathbf{v}_x^H = (-0.15, \ 0.90, \ 0.00)
$$

and

$$
\mathbf{M}_v = \begin{bmatrix} \beta & -0.1500 & 0.1550 \\ -0.1500 & \beta & -0.1500 \\ 0.1550 & -0.1500 & \beta \end{bmatrix}
$$

where $\beta = 0.8850 + N_o$. The inverse of $\mathbf{M}_v$ is

$$\mathbf{M}_v^{-1} = \frac{\mathrm{adj}(\mathbf{M}_v)}{\det(\mathbf{M}_v)}$$

where $\det(\mathbf{M}_v) = \beta(\beta^2 - 0.069025) + 0.006975$ and

$$\mathrm{adj}(\mathbf{M}_v) = \begin{bmatrix} \beta^2 - 0.0225 & 0.15\beta - 0.02325 & 0.0225 - 0.155\beta \\ 0.15\beta - 0.02325 & \beta^2 - 0.024025 & 0.15\beta - 0.02325 \\ 0.0225 - 0.155\beta & 0.15\beta - 0.02325 & \beta^2 - 0.0225 \end{bmatrix}$$

Hence,

$$\mathbf{c}_{\mathrm{op}} = \frac{1}{\det(\mathbf{M}_v)} \begin{pmatrix} -0.15\beta^2 + 0.135\beta - 0.1755 \\ 0.90\beta^2 - 0.0225\beta - 0.018135 \\ 0.15825\beta - 0.0243 \end{pmatrix}$$

With this tap solution,

$$J_{\min} = 1 - \frac{1}{\det(\mathbf{M}_v)} \left(0.8325\beta^2 - 0.013689 \right)$$

and as $N_o \to 0$, $J_{\min} = 0.001089424$.

Adaptive Solution. In practice, the channel impulse response is unknown. However, the equalizer taps can be obtained by using the stochastic gradient algorithm

$$c_j^{n+1} = c_j^n + \alpha \epsilon_n v_{n-j}^* \qquad j = 0, \ldots, N-1 \qquad (7.84)$$

where ϵ_n is given in (7.64). To show that (7.84) leads to the desired solution, note from (7.80) that

$$\begin{aligned} \nabla_{\mathbf{c}} J &= 2\mathrm{E}[\mathbf{c}^T \mathbf{v}_n \mathbf{v}_n^H - x_{n-d} \mathbf{v}_n^H] \\ &= 2\mathrm{E}[(\mathbf{c}^T \mathbf{v}_n - x_{n-d}) \mathbf{v}_n^H] \\ &= 2\mathrm{E}[\epsilon_n \mathbf{v}_n^H] = 0 . \end{aligned} \qquad (7.85)$$

It follows that

$$\mathrm{E}[\epsilon_n v_{n-j}^*] = 0, \qquad j = 0, \ldots, N-1 . \qquad (7.86)$$

Performance of the MSE Equalizer. The performance of an MSE equalizer having an infinite number of taps provides some useful insight. In this case

$$\begin{aligned} \mathbf{c} &= (c_{-\infty}, \ldots, c_0, \ldots, c_\infty) \\ \mathbf{v}_n &= (v_{n+\infty}, \ldots, v_n, \ldots, v_{n-\infty}) . \end{aligned}$$

Since the delay d with an infinite-tap equalizer is irrelevant we can choose $d = 0$ so that

$$E[x_n v_{n-j}^*] = \begin{cases} g_{-j}^* & , \quad -L \le j \le 0 \\ 0 & , \quad \text{otherwise} \end{cases} \tag{7.87}$$

The equation for the optimal tap gain vector $\mathbf{c}^T \mathbf{M}_v = \mathbf{v}_x^H$ can be written in the form

$$\sum_{i=-\infty}^{\infty} c_i \left(f_{j-i} + N_o \delta_{ij} \right) = g_{-j}^* \qquad -\infty < j < \infty \ . \tag{7.88}$$

Taking the z-transform of both sides of (7.88) gives

$$C(z) \left(G(z) G^*(1/z^*) + N_o \right) = G^*(1/z^*) \tag{7.89}$$

and, therefore,

$$C(z) = \frac{G^*(1/z^*)}{G(z) G^*(1/z^*) + N_o} \ . \tag{7.90}$$

The equivalent MSE equalizer that includes the noise whitening filter $1/G^*(1/z^*)$ is

$$C'(z) = \frac{1}{G(z) G^*(1/z^*) + N_o} = \frac{1}{F(z) + N_o} \ . \tag{7.91}$$

Notice that $C'(z)$ has the same form as the ZF equalizer in (7.69), except for the noise term N_o in the denominator. Clearly, the ZF and MSE criterion lead to the same solution in the absence of noise.

The most meaningful measure of performance is the bit error probability. However, for many equalization techniques, the bit error probability is a highly nonlinear function of the equalizer co-efficients. One possibility is to evaluate the MMSE of an infinite-length MMSE equalizer [270]

$$\begin{aligned} J_{\min} &= 1 - \frac{1}{T} \int_{-1/2T}^{1/2T} \frac{F_\Sigma(f)}{F_\Sigma(f) + N_o} df \\ &= T \int_{-1/2T}^{1/2T} \frac{N_o}{F_\Sigma(f) + N_o} df \ . \end{aligned} \tag{7.92}$$

Note that $0 \le J_{\min} \le 1$, and that $J_{\min} = 0$ when there is no ISI or noise and $J_{\min} = 1$ when the folded spectrum $F_\Sigma(f)$ exhibits a spectral null.

Another useful measure for the effectiveness of linear equalization techniques is the *signal-to-noise-plus-interference ratio* (SNIR) defined as

$$\text{SNIR} = \frac{|q_d|^2}{\sum_{\substack{j=0 \\ j \ne d}}^{N+L-1} |q_j|^2 + N_o \sum_{j=0}^{N-1} |c_j|^2} \ . \tag{7.93}$$
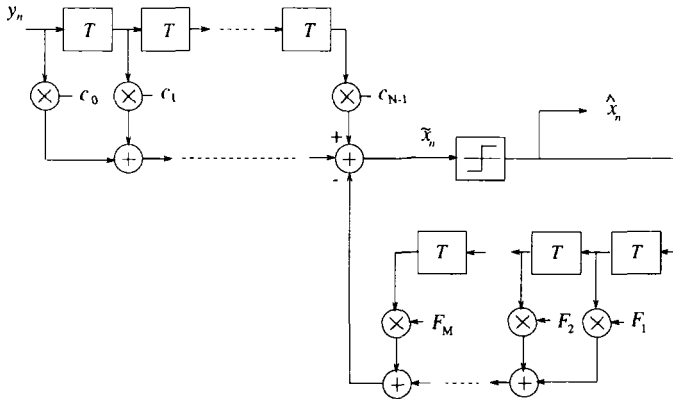
Figure 7.9. Decision feedback equalizer.

Although the MSE equalizer accounts for the effects of noise, satisfactory performance still cannot be achieved for channels with severe ISI or spectral nulls, because of the noise enhancement at the output of the equalizer [270, 107]. Another problem with a linear equalizer is the adaptation of the equalizer during data mode. This problem is especially acute for systems that use trellis-coded modulation, because the equalizer-based decisions are unreliable and inferior to those in uncoded systems due to the reduced separation between the points in the signal constellation. This problem can be partially alleviated by using periodic training, where the equalizer taps are allowed to converge in the periodic training modes. When the equalizer has converged, the updating algorithm is disabled [88]. However, this approach is only suitable for fixed channels or channels with very slow variations such as voiceband data channels.

4.2 DECISION FEEDBACK EQUALIZER (DFE)

The deleterious effects of noise enhancement that degrade the performance of linear equalizers can be mitigated by using a nonlinear decision feedback equalizer (DFE). The DFE consists of two sections; a feedforward section and a feedback section. A typical DFE is illustrated in Fig. 7.9. The DFE is nonlinear because the feedback path includes a decision device. The feedforward section has an identical structure to the linear forward equalizer discussed earlier, and its purpose is to reduce the precursor ISI. For reasons to be discussed later, the input to the feedforward filter is the sampled output of the matched filter. A separate noise whitening filter is not used in this case. Decisions made on the equalizer outputs are propagated through the feedback filter, which is used to estimate the ISI contributed by these symbols. The coefficients of the feedback filter are the sampled impulse response of the tail of the system impulse response including the forward part of the DFE.

The output of the DFE is

$$\tilde{x}_n = \sum_{i=0}^{N-1} c_{N-1-i} v_{n+i} - \sum_{i=1}^{M} F_i \hat{x}_{n-i} \tag{7.94}$$

where $\{c_i\}$ and $\{F_i\}$ are the tap coefficients of the feedforward and feedback filters, respectively, and $\{\hat{x}_i\}$ is the sequence of previously detected symbols. Recall that the overall channel and feedforward portion of the equalizer can be represented by the sampled impulse response in (7.53). By using (7.35), the DFE output can be written as

$$
\begin{aligned}
\tilde{x}_n &= \sum_{i=-L}^{N-1} q_{N-1-i} x_{n+i} - \sum_{i=1}^{M} F_i \tilde{x}_{n-i} + n_n \\
&= x_n q_{N-1} + \sum_{i=1}^{M} (x_{n-i} - \tilde{x}_{n-i}) q_{N-1+i} + \sum_{i=1}^{M} (q_{N-1+i} - F_i) \tilde{x}_{n-i} \\
&\quad + \sum_{i=1}^{N-1} q_{N-1+i} x_{n+i} + \sum_{i=M+1}^{L} q_{N-1+i} x_{n-i} + n_n
\end{aligned} \tag{7.95}
$$

If we choose

$$F_i = q_{N-1+i} = \mathbf{c}^T \mathbf{g}_{N-1+i}, \quad i = 1, 2, \dots, M \tag{7.96}$$

so that the second summation is zero and if correct decisions are made so that the first summation is zero, then

$$\tilde{x}_n = x_n q_{N-1} + \sum_{i=1}^{N-1} q_{N-1+i} x_{n+i} + \sum_{i=M+1}^{L} q_{N-1+i} x_{n-i} + n_n \ . \tag{7.97}$$

The first and second summations in (7.97) are the residual ISI associated with the feedforward and feedback filters, respectively. Note that feedback coefficients in (7.96) result in the complete removal of ISI from the previously detected symbols if $L \leq M$.

Equalizer Tap Solution. The co-efficients $\{c_i\}$ and $\{F_i\}$ can be adjusted simultaneously to minimize the mean square error, sometimes called a MMSE-DFE. Define

$$
\begin{aligned}
\mathbf{c} &= (c_0, c_1, \dots, c_{N-1})^T & (7.98) \\
\mathbf{y}_n &= (y_{n+N-1}, y_{n+N-2}, \dots, y_n)^T & (7.99) \\
\hat{\mathbf{x}}_n &= (\hat{x}_{n-1}, \hat{x}_{n-2}, \dots, \hat{x}_{n-M})^T & (7.100) \\
\mathbf{F} &= (F_1, F_2, \dots, F_M)^T & (7.101)
\end{aligned}
$$

and

$$\tilde{\mathbf{c}} = (\mathbf{c}^T, \mathbf{F}^T)^T \tag{7.102}$$

$$\tilde{\mathbf{y}}_n = (\mathbf{y}_n^T, \hat{\mathbf{x}}_n^T)^T . \tag{7.103}$$

Then the MSE can be expressed as

$$
\begin{aligned}
J &= \mathrm{E}[|x_n - \tilde{x}_n|^2] \\
&= \mathrm{E}\left[\tilde{\mathbf{c}}^T \tilde{\mathbf{y}}_n \tilde{\mathbf{y}}_n^H \tilde{\mathbf{c}}^* - 2\mathrm{Re}\{\tilde{\mathbf{y}}_n^H \tilde{\mathbf{c}}^* x_n\} + |x_n|^2\right] .
\end{aligned}
\tag{7.104}
$$

Since (7.104) and (7.77) have the same form it follows that the optimal tap solution can be obtained by defining

$$\tilde{\mathbf{M}}_y \triangleq \mathrm{E}[\tilde{\mathbf{y}}_n \tilde{\mathbf{y}}_n^H] \tag{7.105}$$

$$\tilde{\mathbf{y}}_x^H \triangleq \mathrm{E}[\tilde{\mathbf{y}}_n^H x_n] \tag{7.106}$$

$$\tilde{\mathbf{c}}_{\mathrm{op}} = (\tilde{\mathbf{M}}_y^T)^{-1}\tilde{\mathbf{y}}_x^* . \tag{7.107}$$

Adaptive Solution. The feedforward taps of the DFE can be adjusted by using

$$c_j^{n+1} = c_j^n + \alpha \epsilon_n v_{n+j}^* \qquad j = 0, \ldots, N-1 \tag{7.108}$$

while the feedback coefficients can be adjusted according to

$$F_j^{n+1} = F_j^n + \alpha \epsilon_n \hat{x}_{n-j}^* \qquad j = 1, \ldots, M . \tag{7.109}$$

To see that this leads to the desired solution, observe that $\nabla_{\mathbf{c}} J = 2\mathrm{E}[\epsilon_n \tilde{\mathbf{v}}_n^H] = \mathbf{0}$ implies that

$$\mathrm{E}[\epsilon_n v_{n+j}^*] = 0, \qquad j = 0, \ldots, N-1 \tag{7.110}$$

$$\mathrm{E}[\epsilon_n x_{n-j}^*] = 0, \qquad j = 1, \ldots, M . \tag{7.111}$$

Performance of the DFE. Since the feedback section of the DFE eliminates the postcursor residual ISI at the output of the forward filter, it is apparent that the optimum setting for the forward filter for an infinite length DFE is identical to a stable, non-causal, noise whitening filter that results in a overall channel with a minimum phase response [274]. The MMSE for the infinite length DFE is [289]

$$J_{\min} = \exp\left\{ T \int_{-1/2T}^{1/2T} \ln\left[\frac{N_o}{F(e^{j2\pi fT}) + N_o}\right] df \right\} \tag{7.112}$$

where $0 \le J_{\min} \le 1$.

4.3   COMPARISON OF SYMBOL-BY-SYMBOL EQUALIZERS

The typical steady-state performance for the various symbol-by-symbol equalizers is now illustrated. Consider 4-PSK modulation on the static ISI channels shown in Fig. 7.10, where we have plotted

$$F(e^{j2\pi fT}) = G(e^{j2\pi fT})G^*(e^{j2\pi fT}) \ .$$

Channel A is an 11-tap typical data-quality twisted copper pair telephone channel with [363]

$$
\begin{aligned}
\mathbf{g}_A \ = \ & (0.0000 + j0.0000, \ 0.0485 + j0.0194, \hspace{2cm} (7.113) \\
& 0.0573 + j0.0253, \ 0.0786 + j0.0282, \ 0.0874 + j0.0447, \\
& 0.9222 + j0.3031, \ 0.1427 + j0.0349, \ 0.0835 + j0.0157, \\
& 0.0621 + j0.0078, \ 0.0359 + j0.0049, \ 0.0214 + j0.0019) \ .
\end{aligned}
$$

Channels B and C have [270]

$$
\begin{aligned}
\mathbf{g}_B \ &= \ (0.407, \ 0.815, \ 0.407) \hspace{3cm} (7.114) \\
\mathbf{g}_C \ &= \ (0.227, \ 0.460, \ 0.688, \ 0.460, \ 0.227) \ . \hspace{1cm} (7.115)
\end{aligned}
$$

Channels B and C have severe ISI, with Channel C having the worst spectral characteristics because of the in-band spectral null.

Fig. 7.11 shows the performance of the linear ZF and MMSE equalizers for Channel A. The equalizers have 21 taps and the tap gains are obtained using the previously discussed iterative techniques. The linear ZF and MMSE equalizers have about the same performance for Channel A.

Fig. 7.12 shows the performance for Channel B. With linear equalization, the optimum tap weights are obtained from a direct solution that assumes a known channel response. Obviously, the linear ZF equalizer is not suitable for Channel B and the linear MMSE equalizer does not perform much better. The performance of a non-linear MMSE-DFE with 11-tap forward section and 10-tap feedback section is also shown. The non-linear MMSE-DFE offers much better performance than the linear ZF or MMSE equalizers for the same complexity. Likewise, Fig. 7.13 shows the performance on Channel C. Again, both the linear ZF and MMSE equalizers perform quite poorly, while the non-linear MMSE-DFE offers much better performance.

5.   SEQUENCE ESTIMATION
5.1   MLSE AND THE VITERBI ALGORITHM

Recall that the overall discrete-time white noise channel with D-branch diversity reception can be modeled by collection of D transversal filters that are
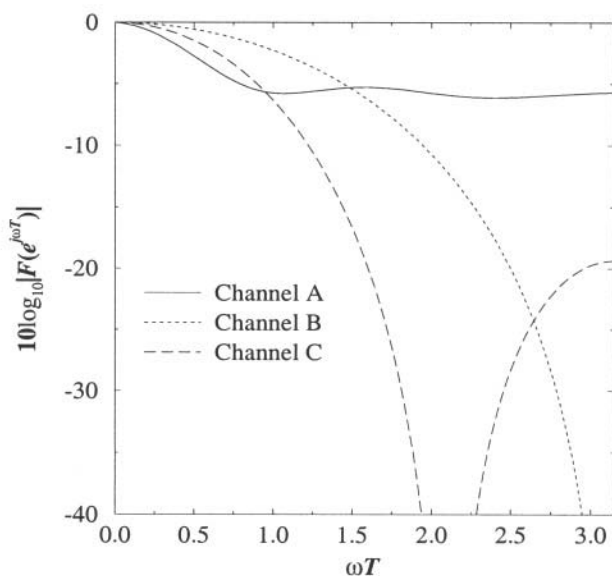
Figure 7.10. Amplitude spectrum for several static ISI channels.
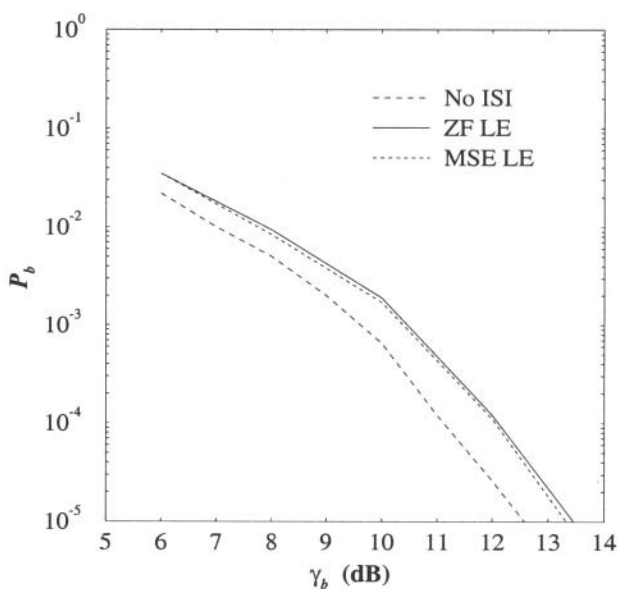
Figure 7.11. Performance of 21-tap ZF and MSE linear equalizers for 4-PSK signaling Channel A in Fig. 7.10.

T-spaced and have $(L + 1)$-taps, as shown in Fig. 7.5. From Fig. 7.5, it can be seen that the channel has a finite number of states. If the size of the signal
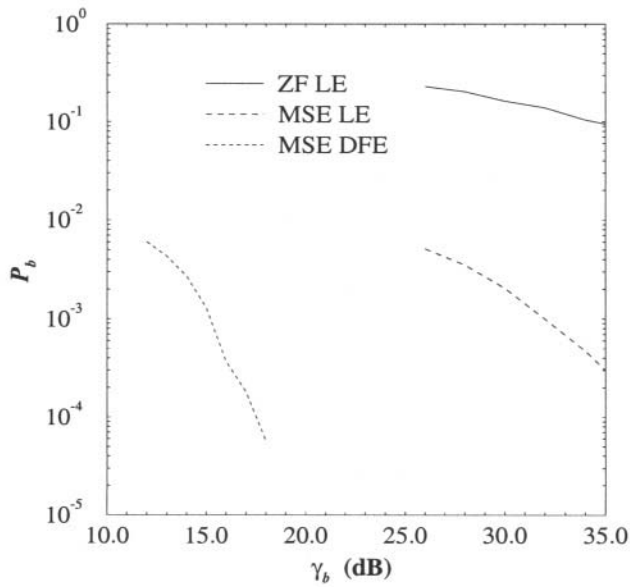
Figure 7.12. Performance of the 21-tap ZF and MSE linear equalizers and a 21-tap DFE equalizers for 4-PSK signaling on Channel B in Fig. 7.10.
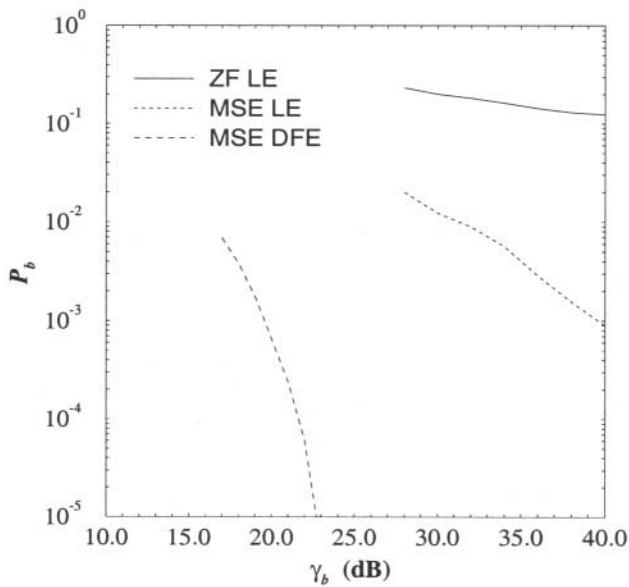
Figure 7.13. Performance of the 21-tap ZF and MSE linear equalizers and a 21-tap DFE equalizers for 4-PSK signaling on Channel C in Fig. 7.10.

constellation is 2^n, there are total of $N_S = 2^{nL}$ **states.** The state at epoch k is

$$\varrho_k = (x_{k-1}, x_{k-2}, \cdots, x_{k-L}) \; . \tag{7.116}$$

Assume that k symbols have been transmitted over the channel. Let $\mathbf{V}_n = (v_{n,1}, \; v_{n,2}, \; \cdots, \; v_{n,D})$ denote the vector of signals received on all diversity branches at epoch n. After receiving the sequence $\{\mathbf{V}_n\}_{n=1}^k$, the ML receiver decides in favor of the sequence $\{x_n\}_{n=1}^k$ that maximizes the **likelihood function**

$$p(\mathbf{V}_k, \; \cdots, \; \mathbf{V}_1 | x_k, \; \cdots, \; x_1) \tag{7.117}$$

or, equivalently, the **log-likelihood function**

$$\log p(\mathbf{V}_k, \; \cdots, \; \mathbf{V}_1 | x_k, \; \cdots, \; x_1) \; . \tag{7.118}$$

Since the noise samples $\{\eta_{n,d}\}$ in (7.35) are independent, and $\mathbf{V}_n$ depends only on the L most recent transmitted symbols, the log-likelihood function (7.118) can be rewritten as

$$\log p(\mathbf{V}_k, \cdots, \mathbf{V}_1 | x_k, \cdots, x_1) =$$

$$\log p(\mathbf{V}_k | x_k, \cdots, x_{k-L}) + \log p(\mathbf{V}_{k-1}, \cdots, \mathbf{V}_1 | x_{k-1}, \cdots, x_1 \tag{7.119}$$

where $x_{k-L} = 0$ for $k - L \leq 0$. If the second term on the right side of (7.119) has been calculated previously at epoch $k-1$ then only the first term, called the **branch metric**, has to be computed for the incoming signal vector $\mathbf{V}_k$ at epoch k.

The model in Fig. 7.5 gives the conditional pdf

$$p(\mathbf{V}_k | x_k, \; \cdots, x_{k-L}) = \frac{1}{(\pi N_o)^D} \exp \left\{ -\frac{1}{N_o} \sum_{d=1}^{D} \left| v_{k,d} - \sum_{i=0}^{L} g_{i,d} x_{k-i} \right|^2 \right\} \tag{7.120}$$

so that $\log p(\mathbf{V}_k | x_k, \; \cdots, x_{k-L})$ yields the branch metric

$$\mu_k = -\sum_{d=1}^{D} \left| v_{k,d} - \sum_{i=0}^{L} g_{i,d} x_{k-i} \right|^2 \; . \tag{7.121}$$

Note that the receiver requires knowledge of the channel vectors $\{\mathbf{g}_d\}$ to compute the branch metrics.

Based on the recursion in (7.119) and the branch metric in (7.121), the well-known **Viterbi algorithm** [342] can be used to implement the ML receiver by searching through the N_S-state trellis for the most likely transmitted sequence

x. This search process is called maximum likelihood sequence estimation (MLSE). Here, we give a very brief outline of the Viterbi algorithm followed by an example. At epoch k, assume that the algorithm has stored N_S **surviving sequences** $\check{\mathbf{x}}(\varrho_k^{(i)})$ (paths through the trellis) along with their associated **path metrics** $\Gamma(\varrho_k^{(i)})$ (distances from the received sequence) that terminate at state $\varrho_k^{(i)}$, $i = 0, \cdots, N_S - 1$. The path metric is defined as

$$\Gamma(\varrho_k^{(i)}) = \sum_{\{k\}} \mu_k \tag{7.122}$$

where $\{\mu_k\}$ is the sequence of branch metrics along the surviving path $\check{\mathbf{x}}(\varrho_k^{(i)})$. After the vector $\mathbf{V}_k$ has been received, the Viterbi algorithm executes the following steps for each state $\varrho_{k+1}^{(j)}$, for $j = 0, \cdots, N_S - 1$:

1. Compute the set of path metrics $\Gamma(\varrho_k^{(i)} \to \varrho_{k+1}^{(j)}) = \Gamma(\varrho_k^{(i)}) + \mu(\varrho_k^{(i)} \to \varrho_{k+1}^{(j)})$ for all possible paths through the trellis that terminate in state $\varrho_{k+1}^{(j)}$.

2. Find $\Gamma(\varrho_{k+1}^{(j)}) = \max_i \Gamma(\varrho_k^{(i)} \to \varrho_{k+1}^{(j)})$ where, again, the maximization is over all possible paths through the trellis that terminate in state $\varrho_{k+1}^{(j)}$.

3. Store $\Gamma(\varrho_{k+1}^{(j)})$ and its associated surviving sequence $\check{\mathbf{x}}(\varrho_{k+1}^{(j)})$. Drop all other paths.

In Step 1 above, $\mu(\varrho_k^{(i)} \to \varrho_{k+1}^{(j)})$ is the branch metric associated with the transition $\varrho_k^{(i)} \to \varrho_{k+1}^{(j)}$ and is computed according to the following variation of (7.121)

$$\mu(\varrho_k^{(i)} \to \varrho_{k+1}^{(j)}) = -\sum_{d=1}^{D} \left| v_{k,d} - g_{0,d} x_k(\varrho_k^{(i)} \to \varrho_{k+1}^{(j)}) - \sum_{m=1}^{L} g_{m,d} x_{k-m}(\varrho_k^{(i)}) \right|^2 \tag{7.123}$$

where $x_k(\varrho_k^{(i)} \to \varrho_{k+1}^{(j)})$ is a symbol that is uniquely determined by the transition $\varrho_k^{(i)} \to \varrho_{k+1}^{(j)}$, and the L most recent symbols $\{x_{k-m}(\varrho_k^{(i)})\}_{m=1}^{L}$ are uniquely specified by the previous state $\varrho_k^{(i)}$.

After all states have been processed, the time index k is incremented and the whole algorithm repeats. As implied in (7.119), the ML receiver waits until the entire sequence $\{\mathbf{V}_n\}_{n=1}^{\infty}$ has been received before making a decision. In practice, such a long delay (maybe infinite) is intolerable and, therefore, a decision about x_{k-Q} is usually made when $\mathbf{V}_k$ is received and processed. It is well known that if $Q > 5L$, the performance degradation caused by the resulting path metric truncation is negligible [342]. MLSE and the Viterbi algorithm is best explained by example and one follows.
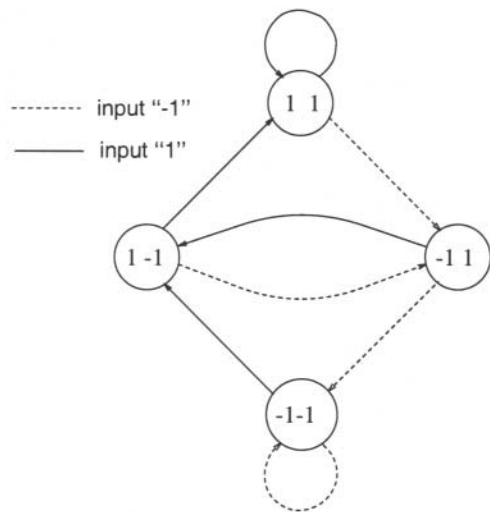
Figure 7.14. State diagram for binary signaling on a three-tap ISI channel.
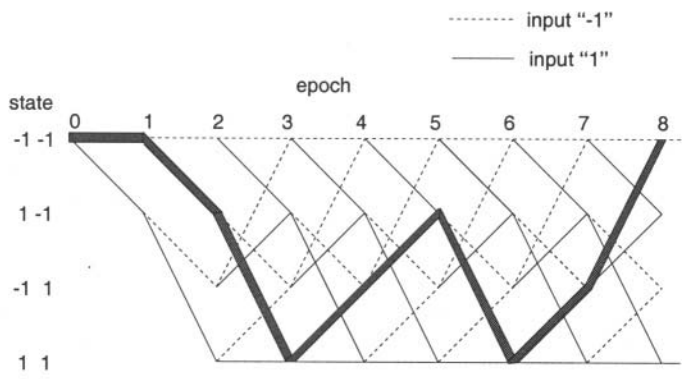
Figure 7.15. Trellis diagram for binary signaling on a three-tap ISI channel.

Example 7.4

Suppose that the binary sequence $\mathbf{x}$, $x_n \in \{-1, +1\}$, is transmitted over a three-tap static ISI channel with channel vector $\mathbf{g} = (1, 1, 1)$. In this case there are four states, and the system can be described the **state diagram** shown in Fig. 7.14. Note that there are two branches entering and leaving each state. In general there are 2^n such branches.

The system state diagram can be used to construct the **trellis diagram** shown in Fig. 7.15, where the initial zero state is assumed to be $\varrho_0^{(0)} = (-1, -1)$. State transitions with a solid line correspond to an input +1, while those with a dashed line correspond to an input –1.
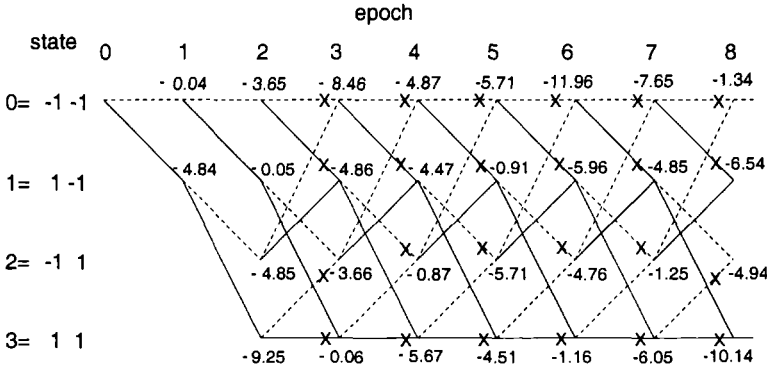
Figure 7.16. Cumulative path metrics and surviving sequences with the Viterbi algorithm.

Suppose that the data sequence $\mathbf{x} = (-1,\ 1,\ 1,\ -1,\ 1,\ 1,\ -1,\ -1,\ldots)$ is transmitted. Then the state sequence follows the shaded path in Fig. 7.15. The noiseless received sequence is $\mathbf{v} = (v_0,\ v_1,\ v_2,\ v_3,\ v_4,\ldots)$ where

$$
\begin{aligned}
v_n &= g_0 x_n + g_1 x_{n-1} + g_2 x_{n-2} \\
&= x_n + x_{n-1} + x_{n-2}
\end{aligned}
$$

Hence, for the data sequence $\mathbf{x} = (-1,\ 1,\ 1,\ -1,\ 1,\ 1,\ -1,\ -1,\ldots)$ the noiseless received sequence is $\mathbf{v} = (-3,\ -1,\ 1,\ 1,\ 1,\ 1,\ 1,\ -1,\ldots)$.

Suppose that the noisy received sequence is

$$
\begin{aligned}
\mathbf{v} &= (v_0,\ v_1,\ v_2,\ v_3,\ v_4,\ldots) \\
&= (-3.2,\ -1.1,\ 0.9,\ 0.1,\ 1.2,\ 1.5,\ 0.7,\ -1.3,\ldots)
\end{aligned}
$$

The Viterbi algorithm is initialized with $\Gamma(\varrho_0^{(i)}) = 0$ for $i = 0, \cdots, N_S - 1$. The initial state is assumed to be $\varrho_0^{(0)} = (-1,\ -1)$. Executing the Viterbi algorithm yields the result shown in Fig. 7.16, where the X's on the branches in the trellis denote dropped paths and the numbers in the trellis are the path metrics corresponding to the surviving sequences. The path metrics are equal to the square Euclidean distance between the surviving sequence $\check{\mathbf{x}}(\varrho_k^{(i)})$ and the corresponding received sequence $\mathbf{v}$.

5.1.1   ADAPTIVE MLSE RECEIVER

The Viterbi algorithm requires knowledge of the channel vectors $\mathbf{g}_d$ to compute the branch metrics in (7.121) so that an adaptive channel estimator is needed. Various channel estimators have been proposed in the literature [62, 215, 98]. Usually, a transversal digital filter with the LMS algorithm is

used for this purpose, because of its good performance, numerical stability, and simplicity in implementation [163, 215]. Another possible adaptation algorithm is the Recursive Least Squares (RLS) or the Kalman algorithm [163]. The RLS algorithm has a very fast convergence rate as compared to the LMS algorithm. However, it is very complicated to implement and it is sensitive to roundoff noise that accumulates due to recursive computations which may cause numerical instabilities in the algorithm [270]. It has also been reported that the tracking properties of the LMS algorithm for the purpose of channel estimation in a fast varying environment are quite similar to those of the RLS algorithm [98, 198, 304]. For these reasons the LMS algorithm is commonly used during the tracking mode in adaptive MLSE receivers. During the training mode, it is possible that the RLS algorithm could offer better performance than the LMS algorithm.

A straightforward method for adaptive channel estimation with an MLSE receiver is to use the final decisions at the output of the Viterbi algorithm to update the channel estimator during the tracking mode. With the LMS algorithm, the tap coefficients are updated according to

$$\hat{g}_{i,d}(k+1) = \hat{g}_{i,d}(k) + \alpha \epsilon_{k-Q,d} \hat{x}^*_{k-i-Q} , \quad \begin{aligned} i &= 0, \ldots, L \\ d &= 1, \ldots, D \end{aligned} \quad (7.124)$$

where α is the adaptation step size, and

$$\epsilon_{k-Q,d} = v_{k-Q,d} - \sum_{i=0}^{L} \hat{g}_{i,d}(k)\hat{x}_{k-i-Q} \qquad (7.125)$$

is the error associated with branch d at epoch k. A major problem with this channel estimator is that it lags behind the true channel vector by the decision delay Q that is used in the Viterbi algorithm. To see this, we can write

$$v_{k-Q,d} = \sum_{i=0}^{L} g_{i,d}(k-Q)x_{k-i-Q} + \eta_{k-Q,d} \qquad (7.126)$$

so that

$$\epsilon_{k-Q,d} = \sum_{i=0}^{L} \left(g_{i,d}(k-Q) - \hat{g}_{i,d}(k) \right) x_{k-i-Q} + \eta_{k-Q,d} . \qquad (7.127)$$

Hence, channel time variations over the decision delay Q will cause the terms $\{g_{i,d}(k-Q) - \hat{g}_{i,d}(k)\}_{k=1}^{L}$ to be non-zero, and this will degrade the tracking performance. The decision delay Q could be reduced but this will also reduce the reliability of the decisions $\hat{x}_{k-i-Q}$ that are used to update the channel estimates in (7.124). Since decision errors will also degrade the performance

of the estimator, the overall performance improvement obtained by reducing Q is often minimal.

One solution to this problem is to use **per-survivor processing** [296, 297, 276, 197], where *each state* has its own channel estimator that tracks the channel. In this case, the tap coefficients are updated according to

$$\hat{g}_{i,d}(k+1) = \hat{g}_{i,d}(k) + \alpha \epsilon_{k,d} \check{x}_{k-i}^*, \quad \begin{aligned} i &= 0, \ldots, L \\ d &= 1, \ldots, D \end{aligned} \quad (7.128)$$

where $\check{x}$ is the surviving sequence associated with *each state*. Notice that the individual channel estimators for each state use zero-delay symbols in their adaptation algorithm and, therefore, good channel tracking performance is expected. These zero-delay symbols are uniquely defined by the state transitions in the trellis diagram.

5.1.2   *T/2*-SPACED MLSE RECEIVER

Suppose that the matched filter output is sampled at rate $2/T$ and the $T/2$-spaced samples are processed with a $T/2$-spaced noise whitening filter as shown in Fig. 7.6. Once again, the channel can be modeled as a finite-state machine with the states defined in (7.116). The Viterbi decoder searches for the most likely path in the trellis based on the $T/2$-spaced received sequence. For each state transition $\varrho_k^{(j)} \to \varrho_{k+1}^{(i)}$ at epoch k, the samples $v_{2k}^{(2)}$ and $v_{2k+1}^{(2)}$ are used by the Viterbi algorithm to evaluate the branch metric[2]

$$\gamma_k(\varrho_k^{(i)} \to \varrho_{k+1}^{(j)}) = \left| v_{2k}^{(2)} - g_0^{(2)} x_k(\varrho_k^{(i)} \to \varrho_{k+1}^{(j)}) - \sum_{m=1}^{L} g_{2m}^{(2)} x_{k-m}(\varrho_k^{(i)}) \right|^2$$

$$+ \left| v_{2k+1}^{(2)} - g_1^{(2)} x_k(\varrho_k^{(i)} \to \varrho_{k+1}^{(j)}) - \sum_{m=1}^{L-1} g_{2m+1}^{(2)} x_{k-m}(\varrho_k^{(i)}) \right|^2 .$$

$$(7.129)$$

5.2   DELAYED DECISION-FEEDBACK SEQUENCE ESTIMATION

Unfortunately, the complexity of the MLSE receiver grows exponentially with the channel memory length. When the channel memory length becomes large, the MLSE receiver becomes impractical. One solution is to reduce the receiver complexity by truncating the effective channel memory to μ terms, where μ is an integer that can be varied from 0 to L. Thus, a suboptimum decoder is obtained with complexity controlled by parameter μ. This is the basic principle of delayed decision-feedback sequence estimation (DDFSE).

---

[2] For notational simplicity we assume $D = 1$.

Let the z transform of the overall discrete-time white noise channel, $G(z)$, be represented as a rational function $\beta(z)/\Psi(z)$, where $\beta(z)$ and $\Psi(z)$ are polynomials. It is assumed that $\Psi(z)$ has degree n_2 and $\Psi_0 = 1$. If L is finite then $\Psi(z) = 1$. The polynomial $G(z)$ can be written as

$$G(z) = G_\mu(z) + z^{-(\mu+1)}G^+(z) \tag{7.130}$$

where

$$G_\mu(z) \quad = \quad \sum_{i=0}^{\mu} g_i z^{-i} \tag{7.131}$$

$$G^+(z) \quad = \quad \sum_{i=0}^{L-\mu-1} g_{i+\mu+1} z^{-i} \quad . \tag{7.132}$$

From (7.130), $G^+(z)$ is a rational function that can be written as $\beta^+(z)/\Psi(z)$ where $\beta^+(z)$ is a polynomial of degree n_1 satisfying the equality

$$\beta^+(z) = [\beta(z) - G_\mu(z)\Psi(z)]z^{\mu+1} \quad . \tag{7.133}$$

Let $W(z) = G^+(z)X(z)$, where $X(z)$ is the z-transform of the input sequence. Then

$$w_k = \begin{cases} \sum_{i=0}^{n_1} \beta_i^+ x_{k-i} - \sum_{i=1}^{n_2} \Psi_i w_{k-i} & , \ L = \infty \\ \sum_{i=0}^{L-\mu-1} g_{i+\mu+1} x_{k-i} & , \ L < \infty \text{ or } n_2 = 0 \end{cases} \tag{7.134}$$

and

$$v_k = \sum_{i=0}^{\mu} g_i x_{k-i} + w_{k-\mu-1} + \eta_k \quad . \tag{7.135}$$

From (7.134) and (7.135), the system state at epoch k can be decomposed into the state

$$\varrho_k^\mu = (x_{k-1}, \ \ldots, \ x_{k-\mu}) \tag{7.136}$$

and a partial state

$$\kappa_k = \begin{cases} x_{k-\mu-1}, \ldots, \ x_{k-\mu-n_1-1}, w_{k-\mu-2}, \ \ldots, \ w_{k-\mu-n_2-2} & ; \ L = \infty \\ x_{k-\mu-1}, \ \ldots, \ x_{k-L} & , \ L < \infty \end{cases} \quad . \tag{7.137}$$

There are $N_\mu = 2^{n\mu}$ states in (7.136).

The DDFSE receiver can be viewed as a combination of the Viterbi algorithm and a decision feedback detector. For each state transition $\varrho_k^{\mu(i)} \to \varrho_{k+1}^{\mu(j)}$, the DDFSE receiver stores N_μ estimates of the partial states κ_k associated with $\varrho_k^{\mu(i)}$. The DDFSE receiver uses the branch metric

$$\mu_k(\varrho_k^{\mu(i)} \rightarrow \varrho_{k+1}^{\mu(j)}) = -\left|y_k - g_0 x_k(\varrho_k^{\mu(i)} \rightarrow \varrho_{k+1}^{\mu(j)})\right.$$

$$\left. - \sum_{l=1}^{\mu} g_l x_{k-l}(\varrho_k^{\mu(i)}) - \hat{w}_{k-\mu-1}\right|^2 . \quad (7.138)$$

The estimate $\hat{w}_{k-\mu-1}$ of $w_{k-\mu-1}$ is obtained from the estimate of the partial state using (7.134). For finite length channels, the DDFSE branch metric can be written as

$$\mu_k(\varrho_k^{\mu(i)} \rightarrow \varrho_{k+1}^{\mu(j)}) = -\left|y_k - g_0 x_k(\varrho_k^{\mu(i)} \rightarrow \varrho_{k+1}^{\mu(j)})\right.$$

$$\left. - \sum_{l=1}^{\mu} g_l x_{k-l}(\varrho_k^{\mu(i)}) - \sum_{l=\mu+1}^{L} g_l \check{x}_l(\varrho_k^{\mu(i)})\right|^2 \quad (7.139)$$

where $\check{x}_l(\varrho_k^{\mu(i)})$ is the lth component of the surviving sequence $\check{\mathbf{x}}(\varrho_k^{\mu(i)})$. Since each path uses decision-feedback based on its own history, the DDFSE receiver avoids using a single unreliable decision for feedback. Hence, error propagation with a DDFSE receiver is not a severe as with a DFE receiver. When $\mu = 0$ the DDFSE receiver is equivalent to Driscoll's decoder [88] and when $\mu = L$ the DDFSE receiver is equivalent to the MLSE receiver.

Finally, since only the μ most recent symbols are represented by the state in (7.136), it is important to have most of the signal energy contained in these terms. Hence, it is very important that the noise whitening filter be selected so that the overall channel $G(z)$ has *minimum* phase. This requirement can present some practical problems. For example, if one of the zeros is close to the unit circle, then the *non-causal* noise whitening filter has a very long impulse response and will be hard to approximate. Also, when the channel is time-varying or unknown, the receiver cannot ensure that $G(z)$ will have minimum phase. Without $G(z)$ having minimum phase, DDFSE does not work well. This point will be repeated again in Chapter 8.

Example 7.5

Consider again the system in Example 7.4, where the received sequence is

$$\mathbf{v} = (v_0, v_1, v_2, v_3, v_4, \ldots)$$
$$= (-3.2, -1.1, 0.9, 0.1, 1.2, 1.5, 0.7, -1.3, \ldots)$$

Recall that $\varrho_k^{(i)} = (x_{k-1}, x_{k-2})$ so there are 4 system states. However, we wish to apply DDFSE with the state $\varrho_k^{\mu(i)} = x_{k-1}$, $i = 0, 1$. The initial state is assumed to be $\Gamma(\varrho_0^{(0)}) = -1$. Since the channel has finite length, (7.139) gives the branch metric

$$\mu_k(\varrho_k^{\mu(i)} \rightarrow \varrho_{k+1}^{\mu(j)}) = \left|y_k - x_k(\varrho_k^{\mu(i)} \rightarrow \varrho_{k+1}^{\mu(j)}) - x_{k-1}(\varrho_k^{\mu(i)}) - \check{x}_2(\varrho_k^{\mu(i)})\right|^2 .$$
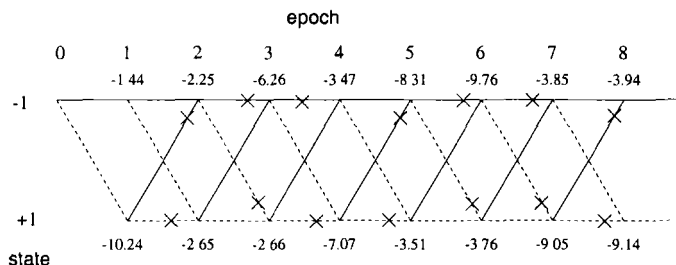
340

Figure 7.17. Cumulative path metrics and surviving sequences with DDFSE.

Applying DDFSE with the Viterbi algorithm gives the result shown in Fig. 7.17. Once again, the *X's* on the branches in the trellis denote dropped paths and the numbers in the trellis denote the path metrics.

5.3 REDUCED-STATE SEQUENCE ESTIMATION

For large signal constellations the number of states with DDFSE, $2^{n\mu}$, is substantial even for small μ. One possible remedy is to reduce the number of states by using Ungerboeck-like set partitioning principles. As described in [109], for each element x_{k-n} in ϱ_k^μ, a set partitioning $\Omega(n), 1 \leq n \leq \mu \leq L$ is defined where the signal set is partitioned into J_i subsets in a way of increasing intrasubset minimum Euclidean distance[3].

The subset in the partitioning $\Omega(i)$ to which x_{k-i} belongs is denoted by $c_i(x_{k-i})$. The subset partitioning is constrained such that $\Omega(i)$ is a finer partition of $\Omega(i+1)$, $1 \leq i \leq \mu - 1$ and $J_1 \geq J_2 \ldots \geq J_\mu$. In this case this following **subset-state** can be defined

$$t_k^\mu = (c_1(x_{k-1}), \ c_2(x_{k-2}), \ \ldots, \ c_\mu(x_{k-\mu})) . \qquad (7.140)$$

Note that the RSSE subset-state does not completely specify the μ most recent symbols $\{x_{k-i}\}_{i=1}^\mu$. Rather, the subset-state only specifies the subsets to which these symbols belong.

The constraints on the subset partitioning ensure a properly defined **subset-trellis.** Given the current subset-state t_k^μ and the subset $c_1(x_k)$ to which the current symbol x_k belongs, the next subset-state t_{k+1}^μ is uniquely determined. Since $c_i(x_{k-i})$ can only assume J_i possible values, there are $\prod_{i=1}^\mu J_i$ subset-states which could be much less than 2^{nL}. Note that if $J_1 < 2^n$, there are parallel transitions associated with each **subset-transition**. The number of

---

[3] If $J_1 = J_2 = \cdots = J_\mu = M$ and $\mu < L$, then RSSE becomes DDFSE.

the parallel transitions is equal to the number of symbols in the corresponding subset.

The Viterbi algorithm used to search the subset-trellis is the same one used for MLSE except for a different branch metric and the possibility of parallel transitions associated with the subset-transitions.[4] When there are parallel transitions, the Viterbi algorithm chooses the parallel transition with the maximum branch metric first[5] and then execute steps for the Viterbi algorithm as defined in Section 5.1.

With RSSE, the branch metric in (7.121) is not uniquely determined by the associated pair of subset-states. This is solved by introducing a decision feedback mechanism for the branch metric calculation [109, 90]. The RSSE branch metric for a particular parallel transition associated with the subset-transition $(t_k^{\mu(i)} \to t_{k+1}^{\mu(j)})$ is

$$\gamma_k(t_k^{\mu(i)} \to t_{k+1}^{\mu(j)}) = -\left| y_k - g_0 x_k(t_k^{\mu(i)} \to t_{k+1}^{\mu(j)}) - \sum_{l=1}^{L} g_l \check{x}_l(t_k^{\mu,(i)}) \right|^2 \quad (7.141)$$

where $x_k(t_k^{\mu(i)}) \to x_k(t_{k+1}^{\mu(j)})$ is the source symbol corresponding to the particular parallel transition, and $:\check{x}_l(t_k^{\mu,(i)})$ is the lth component of the source symbol sequence $\check{x}(t_k^{\mu,(i)})$ that corresponds to the surviving path leading to the subset-state $t_k^{(i)}$. Similar to DDFSE, each path uses decision-feedback based on its own history.

6.   ERROR PROBABILITY FOR MLSE ON ISI CHANNELS

Let $\mathbf{x}$ and $\hat{\mathbf{x}}$ be the transmitted and estimated symbol sequences, respectively. For every pair $\mathbf{x}$ and $\hat{\mathbf{x}}$, the error sequence $\boldsymbol{\epsilon} = \{\epsilon_i\}$ can be formed by defining $\epsilon_i = x_i - \hat{x}_i$. We arbitrarily assume that the bit error probability at epoch j_1 is of interest, so that $\epsilon_{j_1} \neq 0$ for all error sequences that are considered. For each error sequence $\boldsymbol{\epsilon}$, define the following useful error events.

$\mathcal{E}'(\varepsilon)$:   The sequence $\mathbf{x} - \boldsymbol{\epsilon}$ is the maximum likelihood sequence.
$\mathcal{E}(\boldsymbol{\epsilon})$:   The sequence $\mathbf{x} - \boldsymbol{\epsilon}$ has a larger path metric than sequence $\mathbf{x}$.

It is also convenient to define the events

$$\mathcal{E}'_G = \bigcup_{\boldsymbol{\epsilon} \in G} \mathcal{E}'(\boldsymbol{\epsilon}) \quad (7.142)$$

---

[4]With DDFSE there are no parallel transitions.
[5] If the signal constellation has some symmetries, this step can be easily done by using a slicing operation [109].

and

$$\mathcal{E}_F = \bigcup_{\boldsymbol{\epsilon} \in F} \mathcal{E}(\boldsymbol{\epsilon}) \tag{7.143}$$

where G is the set of all possible error sequences having $\epsilon_{j_1} \neq 0$ and $F \subset G$ is the set of error sequences containing no more than $L-1$ consecutive zeroes amid nonzero elements.

Let $\boldsymbol{\varrho} = \{\varrho_k\}$ and $\hat{\boldsymbol{\varrho}} = \{\hat{\varrho}_k\}$ be the system state sequences corresponding to the symbol sequences $\mathbf{x}$ and $\hat{\mathbf{x}}$, respectively. An **error event** occurs between k_1 and k_2, of length $k_2 - k_1$, if

$$\varrho_{k_1} = \hat{\varrho}_{k_1}, \ \varrho_{k_2} = \hat{\varrho}_{k_2}, \ \text{and} \ \varrho_j \neq \hat{\varrho}_j \ \text{for} \ k_1 < j < k_2 \tag{7.144}$$

where $k_1 \leq j_1 \leq k_2$, The symbol error probability at epoch j_1 is

$$
\begin{aligned}
P_s(j_1) \ &\overset{\triangle}{=} \ \mathrm{P}(x_{j_1} \neq \hat{x}_{j_1}) \\
&= \ \mathrm{P}(\mathcal{E}'_G) \\
&= \ \sum_{\boldsymbol{\epsilon} \in G} \sum_{\mathbf{x} \in \mathcal{X}(\boldsymbol{\epsilon})} \mathrm{P}(\mathcal{E}'(\boldsymbol{\epsilon})|\mathbf{x})\mathrm{P}(\mathbf{x}) \tag{7.145}
\end{aligned}
$$

where $\mathcal{X}(\boldsymbol{\epsilon})$ is the set of symbol sequences that can have $\boldsymbol{\epsilon}$ as the error sequence. For different $\boldsymbol{\epsilon}$, the set $\mathcal{X}(\boldsymbol{\epsilon})$ might be different. The third equation in (7.145) is obtained by using the property that the events $\mathcal{E}'(\boldsymbol{\epsilon})$ are disjoint for $\boldsymbol{\epsilon} \in G$. Unfortunately, (7.145) does not admit an explicit expression and, hence, upper bounding techniques are needed for the performance evaluation. A union bound on the error probability will be employed in our analysis.

To obtain a tighter union bound, we now prove that the symbol error probability at epoch j_1 is

$$P_s(j_1) = \mathrm{P}(\mathcal{E}_F) \ . \tag{7.146}$$

Consider the typical trellis diagram as shown in Fig. 7.18, where $\mathbf{x}$ denotes the transmitted symbol sequence, and $\tilde{\mathbf{x}}^{(1)}$ and $\tilde{\mathbf{x}}^{(2)}$ denote two different symbol sequences. It can be seen that the error sequence $\boldsymbol{\epsilon}^{(1)}$ associated with $\tilde{\mathbf{x}}^{(1)}$ and the error sequence $\boldsymbol{\epsilon}^{(2)}$ associated with $\tilde{\mathbf{x}}^{(2)}$ belong to sets F and $G \setminus F$, respectively. For every $\boldsymbol{\epsilon}^{(2)} \in G \setminus F$ there always exists an $\boldsymbol{\epsilon}^{(1)} \in F$. If the sequence $\mathbf{x} - \boldsymbol{\epsilon}^{(2)}$ is the ML sequence, i.e., the event $\mathcal{E}'_G$ has occurred, then the sequence $\mathbf{x} - \boldsymbol{\epsilon}^{(1)}$ has a larger path metric than the sequence $\mathbf{x}$, i.e., the event $\mathcal{E}_F$ has occurred. This means that $\mathcal{E}'_G$ implies $\mathcal{E}_F$. On the other hand, if $\boldsymbol{\epsilon}^{(1)} \in F$ and the sequence $\mathbf{x} - \boldsymbol{\epsilon}^{(1)}$ has a larger path metric than sequence $\mathbf{x}$, then there exists a sequence $\boldsymbol{\epsilon} \in G$ such that the sequence $\mathbf{x} - \boldsymbol{\epsilon}$ is the ML sequence. Therefore, $\mathcal{E}_F$ implies $\mathcal{E}'_{G^n}$ and (7.146) is proven.
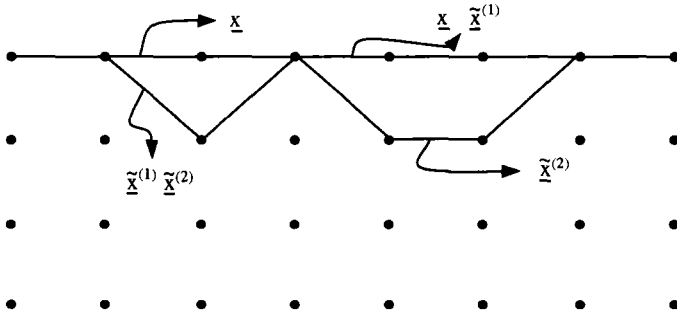
Figure 7.18.   A typical Error state trellis diagram

The union bound on (7.146) yields

$$P_s(j_1) \;\leq\; \sum_{\epsilon \in F} \mathrm{P}(\mathcal{E}(\epsilon))$$

$$= \sum_{\epsilon \in F} \sum_{\mathbf{x} \in \mathcal{X}(\epsilon)} \mathrm{P}(\mathcal{E}(\epsilon)|\mathbf{x})\mathrm{P}(\mathbf{x}) \qquad (7.147)$$

or, equivalently,

$$P_s(j_1) \;\leq\; \sum_{\epsilon \in E} w_s(\epsilon) \sum_{\mathbf{x} \in \mathcal{X}(\epsilon)} \mathrm{P}(\mathcal{E}(\epsilon)|\mathbf{x})\mathrm{P}(\mathbf{x}) \qquad (7.148)$$

where $E \in F$ is the set of error sequences that have the first non-zero element starting at time j_1, and $w_s(\epsilon)$ is the number of symbol errors associated with the error sequence ϵ. To obtain (7.148), we have used the following observations; i) there are $w_s(\epsilon)$ places for the error sequence ϵ to start such that $\epsilon_{j_1} \neq 0$, and ii) the error probability $\mathrm{P}\{\mathcal{E}(\epsilon)|\mathbf{x}\}$ is independent of the place where the error sequence ϵ starts. If the transmitted symbol sequence is long enough, then the symbol error probability $P_s(j_1)$ is independent of the time index j_1 and, therefore, the time index will be omitted hereafter. Finally, for a given transmitted symbol sequence $\mathbf{x}$, the events $\{\mathcal{E}(\epsilon)\}$ for $\epsilon \in F$ in (7.147) might overlap. The reason is that there may be multiple symbol sequences that simultaneously have a larger path metric than the path metric of the transmitted symbol sequence. When the system is operating at a low SNR, there are more overlapping events $\mathcal{E}(\epsilon)$ and, hence, the union bound (7.147) becomes looser.

From the definition of event $\mathcal{E}(\epsilon)$, the union bound (7.148) becomes

$$P_s \;\leq\; \sum_{\epsilon \in E} w_s(\epsilon) \sum_{\mathbf{x} \in \mathcal{X}(\epsilon)} \mathrm{P}(\Gamma(\mathbf{x} - \epsilon) \geq \Gamma(\mathbf{x})|\mathbf{x})\mathrm{P}(\mathbf{x}) \qquad (7.149)$$

where $\Gamma(\mathbf{x})$ is the path metric associated with the input sequence $\mathbf{x}$. To obtain the bit error probability, (7.149) can be easily modified as

$$P_b \leq \frac{1}{n} \sum_{\boldsymbol{\epsilon} \in E} w_b(\boldsymbol{\epsilon}) \sum_{\mathbf{x} \in \mathcal{X}(\boldsymbol{\epsilon})} P(\Gamma(\mathbf{x} - \boldsymbol{\epsilon}) \geq \Gamma(\mathbf{x})|\mathbf{x})P(\mathbf{x}) \qquad (7.150)$$

where n is the number bits transmitted per unit time, and $w_b(\boldsymbol{\epsilon})$ is the number of bit errors associated with the error sequence $\boldsymbol{\epsilon}$. The probability

$$P(\Gamma(\mathbf{x} - \boldsymbol{\epsilon}) \geq \Gamma(\mathbf{x})|\mathbf{x}) \qquad (7.151)$$

is called **pairwise error probability.**

We will see in the following two sections that the pairwise error probability is independent of the transmitted symbol sequence $\mathbf{x}$. Therefore, the union bounds (7.149) and (7.150) simplify to

$$P_s \leq \sum_{\boldsymbol{\epsilon} \in E} w_s(\boldsymbol{\epsilon}) P\left(\Gamma(\mathbf{x} - \boldsymbol{\epsilon}) \geq \Gamma(\mathbf{x})|\mathcal{X}(\boldsymbol{\epsilon})\right) P\left(\mathcal{X}(\boldsymbol{\epsilon})\right) \qquad (7.152)$$

and

$$P_b \leq \frac{1}{n} \sum_{\boldsymbol{\epsilon} \in E} w_b(\boldsymbol{\epsilon}) P\left(\Gamma(\mathbf{x} - \boldsymbol{\epsilon}) \geq \Gamma(\mathbf{x})|\mathcal{X}(\boldsymbol{\epsilon})\right) P\left(\mathcal{X}(\boldsymbol{\epsilon})\right) \qquad (7.153)$$

respectively. The expressions in (7.152) and (7.153) are easier to calculate than those in (7.149) and (7.150), because not all of the symbol sequences have to be considered in the calculation.

6.1 STATIC ISI CHANNELS

The pairwise error probability associated with the error event of length ℓ in (7.144) is (see Problem 6.16)

$$P(\Gamma(\mathbf{x} - \boldsymbol{\epsilon}) \geq \Gamma(\mathbf{x})|\mathbf{x}) = Q\left(\sqrt{\frac{\Delta^2}{4N_o}}\right) \qquad (7.154)$$

where

$$\Delta^2 = \sum_{k=k_1}^{k_1+\ell-1} \left| \sum_{i=0}^{L} g_i \epsilon_{k-i} \right|^2 \qquad (7.155)$$

and Δ^2 is the squared Euclidean path distance. At high signal-to-noise ratios the error event probability is approximately

$$P_e \approx N_{\min} Q\left(\sqrt{\frac{d_{\min}^2}{4N_o}}\right) \qquad (7.156)$$

where $d_{\min}^2$ is the minimum value of Δ^2 and $N_{\min}$ denotes the average number of error events at distance $d_{\min}$.

The squared Euclidean path distance in (7.155) can be rewritten as

$$\Delta^2 = \sum_{k=k_1}^{k_1+\ell-1} \Delta_k^2 \tag{7.157}$$

where

$$\Delta_k^2 = \mathbf{g}^H \mathbf{E}_k \mathbf{g} \tag{7.158}$$

is the squared branch distance and

$$\mathbf{E}_k = [(e_{mn})_k] \tag{7.159}$$

is the $(L+1) \times (L+1)$ **branch distance matrix** having elements $(e_{mn})_k = \epsilon_{k-m+1}^* \epsilon_{k-n+1}$. Define the error vector $\boldsymbol{\varepsilon}_k = (\epsilon_k^*, \epsilon_{k-1}^*, \ldots, \epsilon_{k-L}^*)^T$. It follows that $\mathbf{E}_k = \boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_k^H$ and, hence, $\mathbf{E}_k$ has rank one. Note that $\mathbf{E}_k \boldsymbol{\varepsilon}_k = (\boldsymbol{\varepsilon}_k^H \boldsymbol{\varepsilon}_k) \boldsymbol{\varepsilon}_k$ and, therefore, $\boldsymbol{\varepsilon}_k$ is an eigenvector of $\mathbf{E}_k$ and the only eigenvalue of $\mathbf{E}_k$ is $\lambda(k) = \sum_{i=0}^{L} |\epsilon_{k-i}|^2$. The path distance matrix of the length ℓ error event in (7.144) is defined as

$$\mathbf{E} \triangleq \sum_{k=k_1}^{k_1+\ell-1} \mathbf{E}_k \ . \tag{7.160}$$

Using (7.116) and (7.144), the elements of $\mathbf{E}$ are

$$e_{mn} = r_\ell(n - m) \tag{7.161}$$

where

$$r_\ell(i) = \begin{cases} \sum_{k=k_1}^{k_1+\ell-(L+1)-i} \epsilon_k \epsilon_{k+i}^* & i \geq 0 \\ r_\ell^*(-i) & i < 0 \end{cases} \ . \tag{7.162}$$

It follows that (7.157) has the Hermitian form $\Delta^2 = \mathbf{g}^H \mathbf{E} \mathbf{g}$. Since $\Delta^2 > 0$, $\mathbf{E}$ is a positive definite matrix with all eigenvalues being real and positive. The matrix $\mathbf{E}$ depends on the signal constellation and the length of the channel $L + 1$.

By using (7.37) and the normalization $\mathrm{E}[|x_k|^2] = 1$, the squared Euclidean path distance can be expressed in the form

$$\Delta^2 = 2E \frac{\mathbf{g}^H \mathbf{E} \mathbf{g}}{\mathbf{g}^H \mathbf{g}} \ . \tag{7.163}$$

The ratio of the Hermitian form $\mathbf{g}^H \mathbf{E} \mathbf{g}$ to the inner product $\mathbf{g}^H \mathbf{g}$ is called the **Rayleigh quotient** of the vector $\mathbf{g}$ and is denoted $R(\mathbf{g})$ [163]. The eigenvalues of $\mathbf{E}$ are equal to the Rayleigh quotient of the corresponding eigenvectors. The

Rayleigh quotient of $\mathbf{E}$ satisfies $\lambda_{\min} \leq R(\mathbf{g}) \leq \lambda_{\max}$. The minimum value of $R(\mathbf{g})$ occurs when $\mathbf{g} = \mathbf{v}_{\min}$ and the maximum value occurs when $\mathbf{g} = \mathbf{v}_{\max}$. The eigenvalues of $\mathbf{E}$ are bounded by [163]

$$\lambda_{\max} \leq \sum_{i=0}^{L} |r_\ell(i)| \quad \text{and} \quad \lambda_{\min} \geq r_\ell(0) - \sum_{i=1}^{L} |r_\ell(i)| \ . \tag{7.164}$$

The condition number of $\mathbf{E}$ is defined as $c(\mathbf{E}) \triangleq \lambda_{\max}/\lambda_{\min}$.

6.2   FADING ISI CHANNELS

For fading ISI channels with D-branch diversity reception and maximal ratio combining, the pairwise error probability is still given by (7.154) but the squared Euclidean path distance associated with an error event of length ℓ is [300]

$$\Delta^2 = \sum_{d=1}^{D} \Delta_d^2 \tag{7.165}$$

where

$$\Delta_d^2 = \sum_{k=k_1}^{k_1+\ell-1} \left| \sum_{i=0}^{L} g_{i,d}(k)\epsilon_{k-i} \right|^2 \ . \tag{7.166}$$

The above expression can be written in the form

$$\Delta_d^2 = \sum_{k=k_1}^{k_1+\ell-1} \mathbf{g}_d^H(k)\mathbf{E}_k\mathbf{g}_d(k) \ . \tag{7.167}$$

In general, the covariance matrix $\mathbf{\Phi}_{\mathbf{g}_d}(0)$ defined in (7.39) is not diagonal. A non-diagonal $\mathbf{\Phi}_{\mathbf{g}_d}(0)$ matrix leads to considerable analytical difficulty and loss of insight. However, if $\mathbf{\Phi}_{\mathbf{g}_d}(0)$ is diagonal, then a normalized channel vector $\mathbf{f}_d(k)$ can be defined such that $\mathbf{\Phi}_{\mathbf{f}_d}(0) = \mathbf{I}_{L+1}$. As a result, (7.167) can be rewritten as

$$\Delta_d^2 = \sum_{k=k_1}^{k_1+\ell-1} \mathbf{f}_d^H(k)\mathbf{A}_{k,d}\mathbf{f}_d(k) \tag{7.168}$$

where

$$\mathbf{A}_{k,d} = \mathbf{\Sigma}_d\mathbf{E}_k\mathbf{\Sigma}_d \tag{7.169}$$

and where

$$\mathbf{\Sigma}_d = \text{diag}[\sigma_{0,d}, \ \sigma_{1,d}, \ \ldots, \ \sigma_{L,d}] \tag{7.170}$$

with $\sigma_d = \frac{1}{2}\text{E}[|g_{i,d}|^2]$. It follows that $\mathbf{A}_{k,d} = \mathbf{u}_{k,d}\mathbf{u}_{k,d}^H$ where $\mathbf{u}_{k,d} = \mathbf{\Sigma}_d\epsilon_k$ and, hence, $\mathbf{A}_{k,d}$ is a rank one matrix and $\mathbf{u}_{k,d}$ is an eigenvector of $\mathbf{A}_{k,d}$. The only nonzero eigenvalue of $\mathbf{A}_{k,d}$ is $\lambda_d = \sum_{i=0}^{L} \sigma_{i,d}^2|\epsilon_{k-i}|^2$, where $\sigma_{i,d}^2 = \frac{1}{2}\text{E}[|g_{i,d}|^2]$.

For slowly time-variant channels it is reasonable to assume that $\mathbf{g}_d(k)$ remains constant over the length of the dominant error events, i.e., $\mathbf{g}_d(k) \equiv \mathbf{g}_d$. This assumption holds even for relatively large Doppler frequencies and error event lengths. For example, if the channel exhibits 2-D isotropic scattering and $f_m T = 0.0025$, then error events up to length 20 have $, J_0(2\pi f_m |k|T) \geq J_0(2\pi f_m 10T) = 0.9984 \approx 1$. By using the above assumption, (7.168) can be written as

$$\Delta_d^2 = \mathbf{f}_d^H \mathbf{A}_d \mathbf{f}_d \tag{7.171}$$

where

$$\mathbf{A}_d = \sum_{k=k_1}^{k_1+\ell-1} \mathbf{A}_{k,d}$$

$$= \boldsymbol{\Sigma}_d \mathbf{E} \boldsymbol{\Sigma}_d . \tag{7.172}$$

The matrix $\mathbf{A}_d$ is also positive definite with all its eigenvalues real and positive. The elements of $\mathbf{A}_d$ are given by $[(a_{mn})]_d = \sigma_{m-1,d}\,\sigma_{n-1,d}\,r_\ell(n-m)$ where $r_\ell(i)$ is given by (7.162). The trace of the matrix $\mathbf{A}_d$ is

$$\mathrm{tr}(\mathbf{A}_d) = \sum_{i=1}^{L+1} \lambda_{i,d} = r_\ell(0) \sum_{i=0}^{L} \sigma_{i,d}^2 = \bar{E} r_\ell(0) \tag{7.173}$$

where the $\lambda_{i,d}$, $i = 1, \ldots, L+1$ are the eigenvalues of $\mathbf{A}_d$. The last equality in (7.173) is obtained by using (7.41) along with the normalization $\mathrm{E}[|x_k|^2] = 1$. Since $\mathbf{A}_d$ is Hermitian, there exists a diagonalization $\mathbf{A}_d = \mathbf{U}_d \boldsymbol{\Lambda}_d \mathbf{U}_d^H$ such that $\mathbf{U}_d$ is a unitary matrix and $\boldsymbol{\Lambda}_d$ is a diagonal matrix consisting of the eigenvalues of $\mathbf{A}_d$. Let $\boldsymbol{\omega}_d = \mathbf{U}_d^H \mathbf{f}_d$ be the corresponding diagonal transformation. Hence,

$$\Delta_d^2 = \boldsymbol{\omega}_d^H \boldsymbol{\Lambda}_d \boldsymbol{\omega}_d = \sum_{i=1}^{L+1} \lambda_{i,d} |\omega_{i,d}|^2 \tag{7.174}$$

where $\frac{1}{2}\mathrm{E}[\boldsymbol{\omega}_d \boldsymbol{\omega}_d^H] = \mathbf{I}_{L+1}$ so that the $\{\omega_{i,d}\}$ are independent zero-mean unit-variance Gaussian random variables. Using (7.165) and (7.174) gives

$$\Delta^2 = \sum_{d=1}^{D} \sum_{i=1}^{L+1} \alpha_{i,d} \tag{7.175}$$

where $\alpha_{i,d} = \lambda_{i,d} |\omega_{i,d}|^2$. The $\alpha_{i,d}$ are chi-square distributed with 2 degrees of freedom and, therefore, the characteristic function of Δ^2 is

$$\Psi_{\Delta^2}(z) = \prod_{i=1}^{D(L+1)} \frac{1}{1 - \bar{\alpha}_{i,d} z} \tag{7.176}$$

where $\bar{\alpha}_{i,d} = 2\lambda_{i,d}$. Finally, the pairwise error probability is

$$P(\Gamma(\mathbf{x} - \boldsymbol{\epsilon}) \geq \Gamma(\mathbf{x})|\mathbf{x}) = \int_0^\infty Q\left(\sqrt{2x}\right) f_{\Delta^2}(x)\, dx \qquad (7.177)$$

where $f_{\Delta^2}(x)$ is the probability density function of Δ^2. Note that if some of the eigenvalues $\lambda_{i,d}$ are the same, then there will be repeated poles in the characteristic function in (7.176). This can be expected to be the case for balanced diversity branches, and will also be the case if the channel vector $\mathbf{g}_d$ has equal strength taps. Consider the case where D-branch antenna diversity is used and the channel taps are not of equal strength. In this case, $\lambda_{i,d} \equiv \lambda_i$, $d = 1, \ldots, D$ and the characteristic function in (7.176) has the form

$$\begin{aligned}
\Psi_{\Delta^2}(z) &= \prod_{i=1}^{L+1} \frac{1}{(1 - z\bar{\alpha}_i)^D} \\
&= \sum_{i=1}^{L+1} \sum_{d=1}^{D} \frac{A_{id}}{(1 - z\bar{\alpha}_i)^d} \qquad (7.178)
\end{aligned}$$

where

$$A_{id} = \frac{1}{(D-d)!(-\bar{\alpha}_i)^{D-d}} \left\{ \frac{d^{D-d}}{dz^{D-d}} (1 - z\bar{\alpha}_i)^D\, \Psi_{\Delta^2}(z) \right\}_{z = 1/\bar{\alpha}_i} \qquad (7.179)$$

and $\bar{\alpha}_i = 2\lambda_i$. The pdf of Δ^2 is

$$f_{\Delta^2}(x) = \sum_{i=1}^{L+1} \sum_{d=1}^{D} A_{id} \frac{1}{(d-1)!(\bar{\alpha}_i)^d} x^{d-1} e^{-x/\bar{\alpha}_i} . \qquad (7.180)$$

From (7.177) and (7.180), the exact pairwise error probability is

$$\begin{aligned}
P(\Gamma(\mathbf{x} - \boldsymbol{\epsilon}) \geq \Gamma(\mathbf{x})|\mathbf{x}) &= \sum_{i=1}^{L+1} \sum_{d=1}^{D} A_{id} \left(\frac{1 - \mu_i}{2}\right)^d \\
&\times \sum_{m=0}^{d-1} \binom{d-1+m}{m} \left(\frac{1 + \mu_i}{2}\right)^m \qquad (7.181)
\end{aligned}$$

where

$$\mu_i = \sqrt{\frac{\bar{\alpha}_i}{1 + \bar{\alpha}_i}} . \qquad (7.182)$$

From (7.173), the $\bar{\alpha}_{i,d}$ have the sum value constraint

$$\sum_{i=1}^{L+1} \bar{\alpha}_{i,d} = 2 \sum_{i=1}^{L+1} \lambda_{i,d} = 2\bar{E}r_\ell(0) . \qquad (7.183)$$

Define $S \subseteq R^{L+1}$ as the set of all $(L+1)$-component vectors $\{\boldsymbol{\gamma} : \sum_{i=1}^{L+1} \gamma_i = 2\bar{E}r_\ell(0)\}$. The set 5 is convex, since for any pair of vectors $\boldsymbol{\gamma}_i$ and $\boldsymbol{\gamma}_j$ the convex combination $\theta\boldsymbol{\gamma}_i + (1-\theta)\boldsymbol{\gamma}_j$ is contained in S for any $0 \le \theta \le 1$. If the pairwise error probability is treated as a mapping from S to R, then it is a convex function of $\boldsymbol{\gamma}$ and, hence, has a unique minimum. For example, Fig. 7.19 shows the pairwise error probability for a three-tap channel ($L = 2$, $D = 1$) with equal strength taps ($\gamma_1 = \gamma_2 = \gamma_3$). Note that the value of γ_3 is determined uniquely by the values of γ_1 and γ_2, and that is why a three dimensional graph is used. By using variational calculus, it is shown in Appendix 6A that the pairwise error probability is minimized when the $\bar{\alpha}_{i,d}$ are all equal, i.e., $\lambda_{i,d} = \lambda = r_\ell(0)E/(L+1)$, resulting in the minimum pairwise error probability

$$P_{\min} = \left(\frac{1-\mu}{2}\right)^{D(L+1)} \sum_{m=0}^{D(L+1)-1} \binom{D(L+1)-1+m}{m} \left(\frac{1+\mu}{2}\right)^m$$

$$\tag{7.184}$$

where

$$\mu = \sqrt{\frac{\lambda/4N_o}{1 + \lambda/4N_o}} \quad . \tag{7.185}$$

For a given error event, the pairwise error probability is minimized when $\mathbf{A}_d$ is perfectly conditioned, i.e., $c(\mathbf{A}_d) = 1$. Recall that $c(\mathbf{A}_d) = c(\boldsymbol{\Sigma}_d \mathbf{E} \boldsymbol{\Sigma}_d) \le (c(\boldsymbol{\Sigma}_d))^2 c(\mathbf{E})$, where $(c(\boldsymbol{\Sigma}_d))^2$ represents the ratio of the maximum and minimum channel tap variances $(\sigma_d^2)_{\max}/(\sigma_d^2)_{\min}$. We have seen that $\mathbf{E}$ depends only the signal constellation being used and the channel vector length $L+1$. However, $\mathbf{A}_d$ has information about the signal constellation and power distribution of the fading ISI channel. It follows that $c(\mathbf{A}_d) \le c(\mathbf{E})$ with equality if and only if the channel has equal strength taps. This means that any system has the best performance when the fading ISI channel has equal strength taps.

6.3   COMPUTING THE UNION BOUND

Many algorithms have been suggested for evaluating the union bound on the error probability. One technique is to obtain a union-Chernoff bound by finding the transfer function of the error-state diagram and imposing a Chernoff upper bound on the complementary error function appearing in (7.177). This approach has three draw backs i) the Chernoff bound is very loose when the channel exhibits a deep fade, ii) the transfer function is difficult to obtain for large-state systems, and iii) if the exact pairwise error probability is available, then the transfer function approach cannot be used. To overcome these difficulties, a method based on the error-state transition matrix can be used for calculating the upper bound [7], but it demands a very large amount of computer memory. Here we discuss another alternative that uses an error-state diagram with a
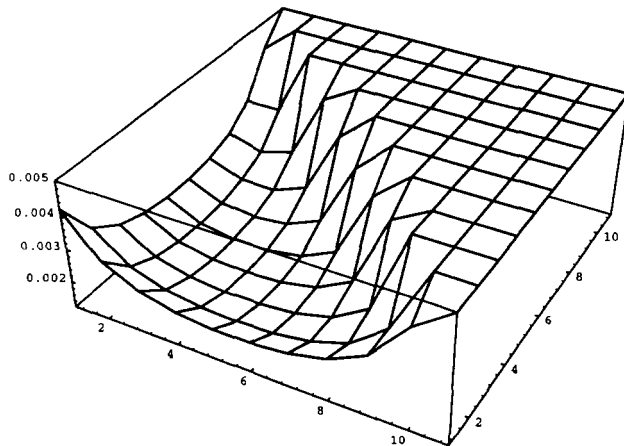
Figure 7.19.   Pairwise error probability for a three-tap fading ISI channel, from [158].

one-directional stack algorithm. Note that other types of algorithms could also be used for this calculation [284].

6.3.1   ERROR-STATE DIAGRAM

To evaluate the upper bound, the error-state diagram must be defined. Assume a system with N_V error-states, $V_0, V_1, \cdots, V_{N_V-1}$. By splitting the zero state, an $(N_V + 1)$-node error-state diagram can be constructed such that the initial and final nodes, V_0 and V_{N_V}, respectively, are zero-error states and the intermediate nodes are non-zero error states. Let t_{ij} denote the branch-weight associated with the V_i to V_j transition, defined as follows:

$$t_{ij} = P_{ij} Z_1^{u_{ij}} Z_2^{\mathbf{A}_{ij}} \tag{7.186}$$

where

- Z_1 and Z_2 are intermediate (dummy) variables.

- P_{ij} is the fraction of correct symbols x_k such that the transition from V_i to V_j is possible.

- u_{ij} is the number of bit errors associated with the transition from V_i to V_j.

- $\mathbf{A}_{ij}$ is given by $(7.169)^6$, but we emphasize that it is a function of the V_i to V_j transition.

---

[6]Here we assume equal diversity branches, i.e., $\mathbf{\Sigma}_d \equiv \mathbf{\Sigma}$, $\forall\, d$, so that $\mathbf{A}_{k,d} \equiv \mathbf{A}_k$, $\forall\, d$.

From the definition of t_{ij} in (7.186), the weight of a particular path in the error-state diagram is

$$\prod_{\{(i,j)\}} P_{ij} Z_1^{\sum_{\{(i,j)\}} u_{ij}} Z_2^{\sum_{\{(i,j)\}} \mathbf{A}_{ij}} \tag{7.187}$$

where $\{(i, j)\}$ denotes the set of state transitions associated with the path under consideration. Note that each path beginning at the initial node and ending at the final node in the error-state diagram represents an error sequence $\boldsymbol{\epsilon} \in E$, where the set E is defined in (7.148). From (7.187)

$$P\left(\mathcal{X}(\boldsymbol{\epsilon})\right) = \prod_{\{(i,j)\}} P_{ij} \tag{7.188}$$

$$w(\boldsymbol{\epsilon}) = \sum_{\{(i,j)\}} u_{ij} \tag{7.189}$$

and

$$\mathbf{A} = \sum_{\{(i,j)\}} \mathbf{A}_{ij} \ . \tag{7.190}$$

These values are required in the calculation of (7.152) or (7.153).

6.3.2   THE STACK ALGORITHM

The union bounds in (7.152) and (7.153) require the calculation of an infinite series. In practice, the mathematical rigor must be sacrificed by truncating the series at an appropriate point. The basic idea of the stack algorithm is to include the R error sequences $\boldsymbol{\epsilon} \in E$ that correspond to the R largest terms in (7.152) or (7.153). The value of R is chosen so that the rest of the terms in the union bound are insignificant. Alternatively, the union bound can be truncated by excluding all paths that have a pairwise error probability P_I less than a threshold P_T.

The stack algorithm maintains a stack with each path (entry) containing the following information; terminal node, $\prod_{\{(i,j)\}} P_{ij}$, $\sum_{\{(i,j)\}} u_{ij}$, $\sum_{\{(i,j)\}} \mathbf{A}_{ij}$, and the intermediate bit error probability P_I. Here, P_I is calculated by

$$P_I = \prod_{\{(i,j)\}} P_{ij} \cdot \sum_{\{(i,j)\}} u_{ij} \cdot \check{P}_r\{\Gamma(\mathbf{x} - \boldsymbol{\epsilon}) \geq \Gamma(\mathbf{x})|\mathbf{x}\} \tag{7.191}$$

where $\check{P}_r\{\Gamma(\mathbf{x} - \boldsymbol{\epsilon}) \geq \Gamma(\mathbf{x})|\mathbf{x}\}$ is calculated by using (7.181) along with the eigenvalues associated with the matrix $\sum_{\{(i,j)\}} \mathbf{A}_{ij}$.

The stack is ordered (from top-to-bottom) in order of decreasing intermediate bit error probability P_I. The algorithm first checks if the top path has terminated at the final node. If it has, then the algorithm outputs P_I which is one of the R terms that will be included in the calculation of (7.152) or (7.153); otherwise,

the top path is extended and the stack is reordered. Since the top path has the largest P_I, it is likely that the extensions of this path will correspond to one or more of the R dominant terms that are of interest. All paths with the same P_I can be grouped together for easier sorting of the stack. The complete algorithm is given in Fig. 7.20 and is described as follows.

1. Load the stack with the initial node, set all the parameters equal to zero, and input the threshold value P_T (described below) or R.

2. Determine if the top path terminates at the final node. If it does, then go to Step 3; otherwise go to Step 4.

3. Output P_I, and determine if the algorithm should be terminated. If yes, then terminate the algorithm; otherwise delete the top path and go to Step 2.

4. Extend the top path and calculate $\prod_{\{(i,j)\}} P_{ij}$, $\sum_{\{(i,j)\}} u_{ij}$, $\sum_{\{(i,j)\}} \mathbf{A}_{ij}$, and P_I for all of the extension paths.

5. Delete the top path.

6. Insert the new extension paths and rearrange the stack in the order of decreasing intermediate bit error probability P_I.

7. Go to Step 2.

6.4   EXAMPLES

Union bounds will now be evaluated and compared with computer simulations for two example systems by using the above procedure. In the simulations, $\sigma_{i,d}^2 = \sigma^2$, $\forall\ i$, d. The tap coefficients are generated by passing independent complex white Gaussian noise through a digital Butterworth filter with a normalized 3-dB cut off frequency equal to $fT = 8.333 \times 10^{-5}$ Hz, typical of an HF channel [201]. All analytical results are obtained by setting the threshold $P_T = 10^{-3} \cdot P_{\max}$, where $P_{\max}$ is the maximum term in the upper bound in (7.153).

Example 7.6 BPSK Modulated System ⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯

A three-tap channel with BPSK modulation is analyzed in this example, where $x_k \in \{-1, +1\}$. There are three different error symbols in this case, i.e., $\epsilon_k \in \{0, \pm 2\}$. The error-state diagram is shown in Fig. 7.21. Observe that the error-state diagram is symmetrical in that there are always two paths having the same set of parameters $\mathrm{P}\{\mathcal{X}(\varepsilon)\}$, $w(\varepsilon)$, and A. Combining all such pairs of paths together, results in the simplified error-state diagram shown in Fig. 7.22. For equal strength taps, $\mathbf{A}_{ij} = \sigma^2 \mathbf{E}_{ij}$, where $\mathbf{E}_{ij}$ is given by (7.159). The
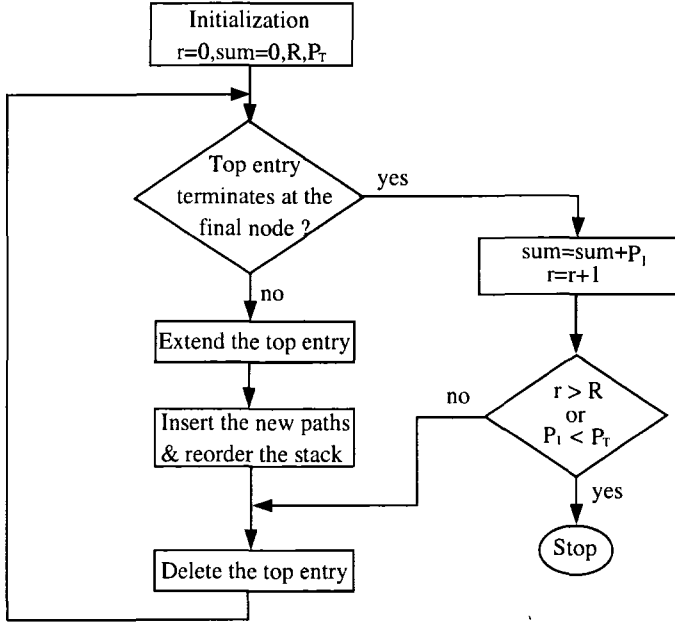
Figure 7.20.   The stack algorithm for computing the error probability union bound with MLSE, from [300].

branch weights for the error-state diagram are defined in Table 6.1. Since $\mathbf{E}_{ij}$ is Hermitian, only the lower triangular elements of the matrix $\mathbf{E}_{ij}$ are given.

Fig. 7.23 compares the union bound with simulation results. The received branch bit energy-to-noise ratio $\bar{\gamma}_c$, can be obtained from (7.41). For $D = 1$, the union bound is loose by about 2 dB for bit error probabilities less than 10^{-3}. However, for $D = 2$, the union bound is tight to within 1 dB. This is reasonable because the channel is unlikely to experience a deep fade on both diversity branches where the union bound becomes loose. In general, the bound is tighter for larger $\bar{\gamma}_c$ and D.

Example 7.7 QPSK Modulated System

This example considers QPSK on a two-tap channel model. The x_k are complex taking on the values $\exp\{j(\pi/4 \pm k\pi/2)\}$, $k = 0, 1, 2, 3$. There are nine different error symbols in this case, i.e., $\epsilon_k \in \{0, \pm\sqrt{2}, \pm j\sqrt{2}, \pm\sqrt{2} \pm j\sqrt{2}\}$. It is left as an exercise to the reader that Fig. 7.24 represents a simplified error state diagram. The branches labeled with "2" represent two error-state transitions. For example, the branch 612 represents the error-state transitions $\epsilon_{12} = \pm(\sqrt{2}, j\sqrt{2})$ and $\epsilon_{12} = \pm(\sqrt{2}, -j\sqrt{2})$. The transition-gains are shown in Table 6.2, where only the lower triangular elements of $\mathbf{E}_{ij}$ are given. Fig. 7.25
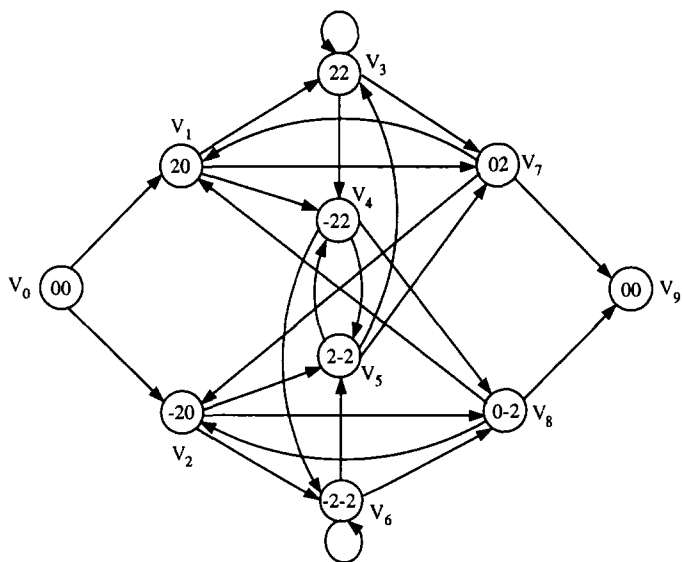
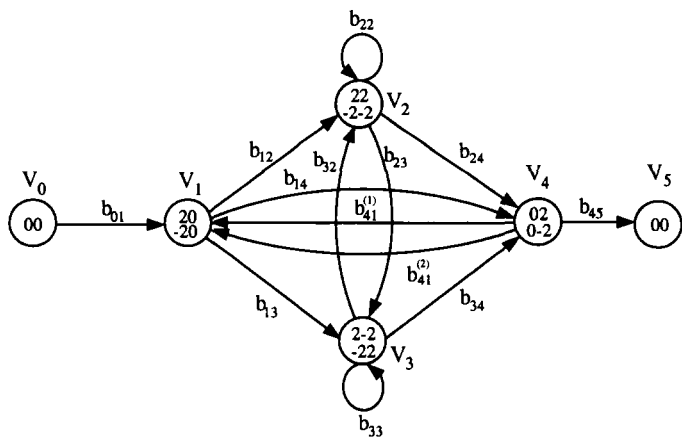*Figure 7.21.* Error-state diagram for the BPSK system, from [300].

Figure 7.22. Simplified error-state diagram for the BPSK system, from [300].

compares the union bound with simulation results. For $D = 1$, the difference is about 4 dB. However, for $D = 2$ the difference is only 1.5 dB.

branch	P_{ij}	u_{ij}	e_{11}	e_{21}	e_{22}	e_{31}	e_{32}	e_{33}
b_{01}	1.0	1.0	4.0	0.0	0.0	0.0	0.0	0.0
b_{12}	0.5	1.0	4.0	4.0	4.0	0.0	0.0	0.0
b_{13}	0.5	1.0	4.0	-4.0	4.0	0.0	0.0	0.0
b_{14}	0.5	0.0	0.0	0.0	4.0	0.0	0.0	0.0
b_{22}	0.5	1.0	4.0	4.0	4.0	4.0	4.0	4.0
b_{23}	0.5	1.0	4.0	-4.0	4.0	-4.0	4.0	4.0
b_{24}	0.5	0.0	0.0	0.0	4.0	0.0	4.0	4.0
b_{32}	0.5	1.0	4.0	4.0	4.0	-4.0	-4.0	4.0
b_{33}	0.5	1.0	4.0	-4.0	4.0	4.0	-4.0	4.0
b_{34}	0.5	0.0	0.0	0.0	4.0	0.0	-4.0	4.0
$b_{41}^{(1)}$	0.5	1.0	4.0	0.0	0.0	4.0	0.0	4.0
$b_{41}^{(2)}$	0.5	1.0	4.0	0.0	0.0	-4.0	0.0	4.0
b_{45}	0.5	0.0	0.0	0.0	0.0	0.0	0.0	4.0

Table 7.1.   Branch weights of BPSK modulated system, from [300]

7.   ERROR PROBABILITY FOR *T*/2-SPACED MLSE RECEIVER

Referring to Fig. 7.6, let $X(z)$, $V(z)$, and $V_{(2)}(z)$ be the z-transforms of the input sequence $\mathbf{x}$, the T-spaced received sequence $\mathbf{v}$ and the $T/2$-spaced received sequence $\mathbf{v}^{(2)}$, respectively. The mappings from $X(z)$ to $V(z)$ and from $X(z)$ to $V^{(2)}(z)$ are one-to-one and both the T-spaced and $T/2$-spaced MLSE receivers operate on noisy sequences that are corrupted by noise samples with variance N_o. Therefore, we only need to compare the Euclidean distances between allowed sequences of channel outputs to determine the relative performance of the T- and $T/2$-spaced receivers.

7.1   *T*-SPACED MLSE RECEIVER

From the definition of the error event in (7.144), the z-transform of the error sequence is

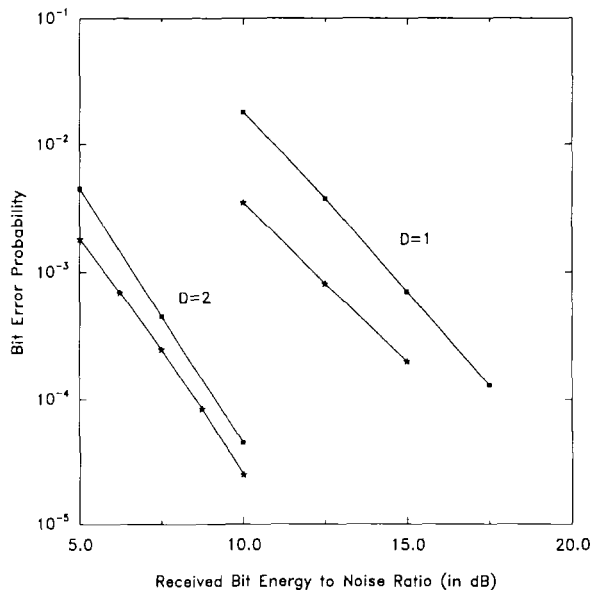$$\mathcal{E}(z) = \epsilon_{k_1} + \epsilon_{k_1+1}z^{-1} + \ldots + \epsilon_{k_2-L-1}z^{L-\ell+1} \qquad (7.192)$$

Figure 7.23. Comparison of union bounds with simulation results for BPSK on a three-equal-ray T-spaced fading ISI channel, [300].
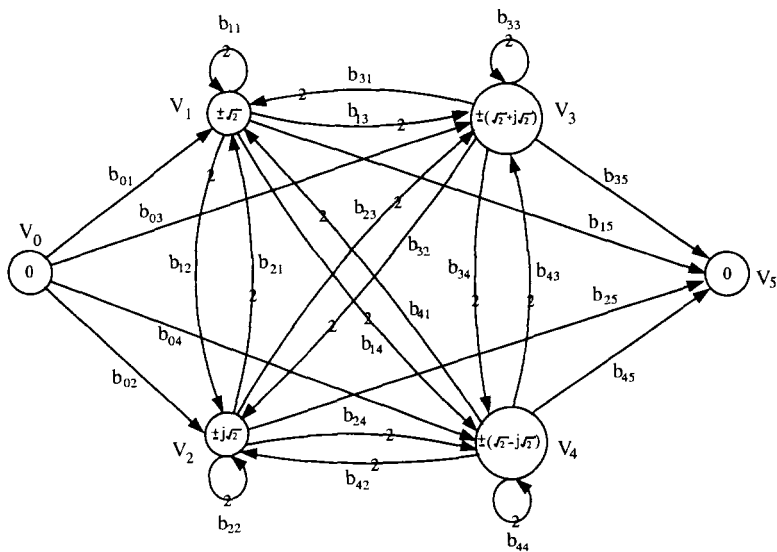

Figure 7.24. Simplified error-state diagram for QPSK system, from [300].

branch	P_{lm}	u_{lm}	e_{11}	e_{21}	e_{22}
b_{01}	1.0	1.0	$2.0 + j0.0$	$0.0 + j0.0$	$0.0 + j0.0$
b_{02}	1.0	1.0	$2.0 + j0.0$	$0.0 + j0.0$	$4.0 + j0.0$
$b_{11}^{(1)}$	0.5	1.0	$2.0 + j0.0$	$2.0 + j0.0$	$2.0 + j0.0$
$b_{11}^{(2)}$	0.5	1.0	$2.0 + j0.0$	$-2.0 + j0.0$	$2.0 + j0.0$
$b_{23}^{(1)}$	0.25	2.0	$4.0 + j0.0$	$2.0 - j2.0$	$2.0 + j0.0$
$b_{23}^{(2)}$	0.25	2.0	$4.0 + j0.0$	$-2.0 + j2.0$	$2.0 + j0.0$
$b_{34}^{(1)}$	0.25	2.0	$4.0 + j0.0$	$0.0 - j4.0$	$4.0 + j0.0$
$b_{34}^{(2)}$	0.25	2.0	$4.0 + j0.0$	$-4.0 + j0.0$	$4.0 + j0.0$
b_{35}	1.0	0.0	$0.0 + j0.0$	$0.0 + j0.0$	$4.0 + j0.0$
b_{45}	1.0	0.0	$0.0 + j0.0$	$0.0 + j0.0$	$4.0 + j0.0$

Table 7.2. Transition-gain examples of QPSK system, from [300]

where $\epsilon_k = x_k - \hat{x}_k$. The z-transform of the signal error sequence associated with the error event is

$$\mathcal{E}_v(z) = (v_{k_1} - \hat{v}_{k_1}) + (v_{k_1+1} - \hat{v}_{k_1+1})z^{-1} + \ldots + (v_{k_1-\ell-1} - \hat{v}_{k_1+\ell-1})z^{-\ell+1} \tag{7.193}$$

and we have

$$\mathcal{E}_v(z) = \mathcal{E}(z)G(z) . \tag{7.194}$$

From (7.155), the squared Euclidean distance Δ^2 of the error event in (7.144) is [127]

$$
\begin{aligned}
\Delta^2 &= \sum_{k=k_1}^{k_1+\ell-1} \left| \sum_{i=0}^{L} g_i \epsilon_{k-i} \right|^2 \\
&= [\, \mathcal{E}_v(z)\mathcal{E}_v^*(1/z^*) \,]_0 \\
&= [\, \mathcal{E}(z)F(z)\mathcal{E}^*(1/z^*) \,]_0
\end{aligned} \tag{7.195}
$$

where $[\cdot]_0$ is the coefficient of z^0.

7.2 $T/2$-SPACED MLSE RECEIVER

For the same error event described in (7.144), the corresponding z-transform of the $T/2$-spaced error sequence is

$$\mathcal{E}^{(2)}(z) = \epsilon_{k_1}^{(2)} + \epsilon_{k_1+1}^{(2)}z^{-2} + \ldots + \epsilon_{k_2-L-1}^{(2)}z^{2(L-\ell+1)} . \tag{7.196}$$
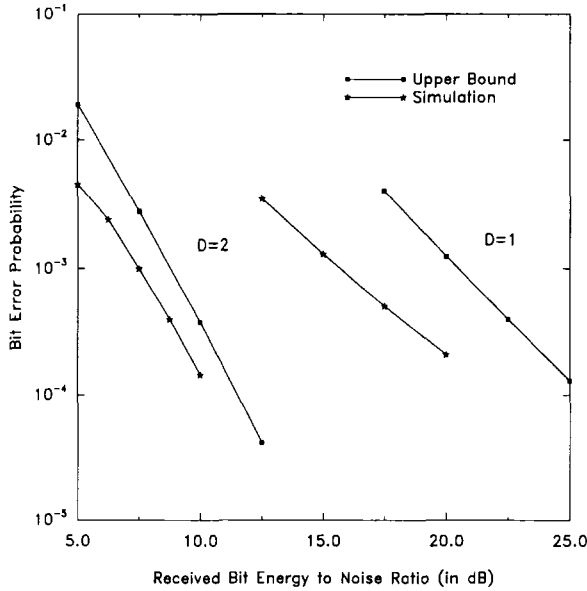
Figure 7.25. Comparison of union bound and simulation results for a QPSK on a two-equal-ray T-spaced channel, from [300].

Notice that $\epsilon_k^{(2)} = x_k^{(2)} - \hat{x}_k^{(2)}$ is zero for even k. Therefore, $\mathcal{E}^{(2)}(z) = \mathcal{E}(z^2)$. The corresponding z-transform of the $T/2$-spaced signal error sequence associated with the error event in (7.144) is

$$\mathcal{E}_v^{(2)}(z) = \mathcal{E}^{(2)}(z)G^{(2)}(z) \ . \tag{7.197}$$

From (7.155), the squared Euclidean distance of the error event in (7.144) is

$$
\begin{aligned}
\left(\Delta^{(2)}\right)^2 &= \sum_{k=2k_1}^{2(k_1+\ell-1)} \left| \sum_{i=0}^{2L} g_i^{(2)} \epsilon_{k-i}^{(2)} \right|^2 \\
&= [\, \mathcal{E}_v^{(2)}(z)\mathcal{E}_v^{(2)*}(1/z^*) \,]_0 \\
&= [\, \mathcal{E}^{(2)}(z)F^{(2)}(z)\mathcal{E}^{(2)*}(1/z^*) \,]_0 \\
&= [\, \mathcal{E}(z^2)F^{(2)}(z)\mathcal{E}^*(1/z^{*2}) \,]_0 \ . \tag{7.198}
\end{aligned}
$$

Note that polynomial $\mathcal{E}(z^2)\,\mathcal{E}^*(1/z^{*2})$ has the property that the odd powers of z have zero coefficients. Therefore, the contributions to the coefficient $[\, \mathcal{E}(z^2)F^{(2)}(z)\mathcal{E}^*(1/z^{*2}) \,]_0$ arise only from the coefficients of $F^{(2)}(z)$ associated with even powers of z. Note also from (7.42) and (7.43) that the coefficients $f_{2k}^{(2)}$ of $F^{(2)}(z)$ associated with even powers of z are equal to the

coefficients f_k of $F(z)$, i.e., $f_{2k}^{(2)} = f_k$. Therefore,

$$\left(\Delta^{(2)}\right)^2 = [\,\mathcal{E}(z^2)F^{(2)}(z)\mathcal{E}^*(1/z^{*2})\,]_0 = [\,\mathcal{E}(z)F(z)\mathcal{E}^*(1/z^*)\,]_0 = \Delta^2 \ .$$
(7.199)

Consequently, the error probability performance of the T- and $T/2$-spaced MLSE receivers are identical.

Example 7.8

Let

$$\begin{aligned}
\mathcal{E}(z) &= \epsilon_0 + \epsilon_1 z^{-1} + \epsilon_2 z^{-2} \\
F^{(2)}(z) &= f_{-2}^{(2)} z^2 + f_{-1}^{(2)} z + f_0^{(2)} + f_1^{(2)} z^{-1} + f_2^{(2)} z^{-2} \ .
\end{aligned}$$

Then

$$\begin{aligned}
F(z) &= f_{-1}z + f_0 + f_1 z^{-1} = f_{-2}^{(2)} z + f_0^{(2)} + f_2^{(2)} z^{-1} \\
\mathcal{E}^{(2)}(z) &= \epsilon_0 + \epsilon_1 z^{-2} + \epsilon_2 z^{-4} \ .
\end{aligned}$$

Therefore,

$$\begin{aligned}
[\,\mathcal{E}(z)F(z)\mathcal{E}^*(1/z^*)\,]_0 = \ & \left(|\epsilon_0|^2 + |\epsilon_1|^2 + |\epsilon_2|^2\right)^2 f_0 \\
& + \epsilon_0 \epsilon_1^* f_1 + \epsilon_1 \epsilon_0^* f_{-1}
\end{aligned}$$

and

$$\begin{aligned}
[\,\mathcal{E}^{(2)}(z)F^{(2)}(z)\mathcal{E}^{(2)*}(1/z^*)\,]_0 = \ & \left(|\epsilon_0|^2 + |\epsilon_1|^2 + |\epsilon_2|^2\right)^2 f_0^{(2)} \\
& + \epsilon_0 \epsilon_1^* f_2^{(2)} + \epsilon_1 \epsilon_0^* f_{-2}^{(2)} \ .
\end{aligned}$$

Hence, $\Delta^2 = (\Delta^{(2)})^2$.

7.3   PRACTICAL *T/2*-SPACED MLSE RECEIVER

The receivers in Figs. 7.3 and 7.6 use a filter that is matched to the received pulse $h^*(-t)$. Since this filter requires knowledge of the unknown channel impulse response, it is impractical. One solution is to implement an 'ideal' low-pass filter with a cutoff frequency of $1/T$ and sample the output at rate $2/T$. The noise samples at the output of this filter will be uncorrelated and, therefore, the $T/2$-spaced MLSE receiver can be implemented. Vachula and Hill [332] showed that this receiver is optimum; however, it has some drawbacks. First, it is not suitable for bandwidth efficient systems that are affected by adjacent channel interference such as the North American IS-54 and Japanese PDC systems, because the cutoff frequency of the low-pass filter will extend significantly into the adjacent band. Second, the ideal low-pass filter is nonrealizable
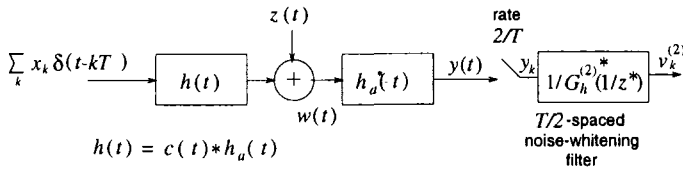
$$h(t) = c(t) * h_a(t)$$

Figure 7.26. Block diagram of system that implements a filter matched to $h_a(t)$ followed by a $T/2$-spaced noise whitening filter. The structure of the noise whitening filter depends only on the pulse $h_a(t)$.

and difficult to approximate. One solution is to use a receiver filter that is matched to the transmitted pulse $h_a(t)$ as suggested by Hamied and Stüber [159]. Chugg and Polydoros have suggested a similar approach [53]. If the received pulse $h(t)$ is time-limited, then such front-end processing is optimal only if the channel rays themselves are $T/2$ spaced. However, if the transmitted signals are strictly bandlimited with at most 100% excess bandwidth, then rate $2/T$ sampling satisfies the sampling theorem and the $T/2$-spaced samples will provide sufficient statistics as we now show.

Let $H_a^{(2)}(z)$, $C^{(2)}(z)$, and $H^{(2)}(z)$ be the z-transforms of the $T/2$-spaced discrete-time signals corresponding to $h_a(t)$, $c(t)$, and $h(t)$, respectively. The z-transform of the autocorrelation function of the noise samples at the output of the receive filter $h_a^*(-t)$ is $N_o F_h^{(2)}(z)$ where $F_h^{(2)}(z) = H_a^{(2)}(z)\left(H_a^{(2)}(1/z^*)\right)^*$. Using the factorization

$$F_h^{(2)}(z) = G_h^{(2)}(z)\left(G_h^{(2)}(1/z^*)\right)^* \tag{7.200}$$

the $T/2$-spaced noise sequence can be whitened by using a filter with transfer function $1/\left(G_h^{(2)}(1/z^*)\right)^*$. The resulting system is shown in Fig. 7.26. We now show that the receivers in Figs. 7.6 and 7.26 yield identical performance.

The z-transform of the overall $T/2$-spaced discrete-time channel that includes the noise-whitening filter is

$$
\begin{aligned}
G_{eq}^{(2)}(z) &= H_a^{(2)}(z)C^{(2)}(z)\left(H_a^{(2)}(1/z^*)\right)^* / \left(G_h^{(2)}(1/z^*)\right)^* \\
&= C^{(2)}(z)G_h^{(2)}(z) \ . \tag{7.201}
\end{aligned}
$$

On the other hand, referring to the conventional system shown in Fig. 7.6, we have

$$H^{(2)}(z) = H_a^{(2)}(z)\, C^{(2)}(z) \tag{7.202}$$

and

$$F^{(2)}(z) = H_a^{(2)}(z)\left(H_a^{(2)}(1/z^*)\right)^* C^{(2)}(z)\left(C^{(2)}(1/z^*)\right)^* \ . \tag{7.203}$$

Let

$$C^{(2)}(z)\left(C^{(2)}(1/z^*)\right)^* = G_c^{(2)}(z)\left(G_c^{(2)}(1/z^*)\right)^* \qquad (7.204)$$

be a factorization of $C^{(2)}(z)\left(C^{(2)}(1/z^*)\right)^*$ such that $\left(G_c^{(2)}(1/z^*)\right)^*$ has minimum phase. Using (7.200), (7.203) and (7.204) yields

$$F^{(2)}(z) = G_h^{(2)}(z)\left(G_h^{(2)}(1/z^*)\right)^* G_c^{(2)}(z)\left(G_c^{(2)}(1/z^*)\right)^* . \qquad (7.205)$$

The transfer function of the noise-whitening filter must be chosen as

$$1/\left(\left(G_h^{(2)}(1/z^*)\right)^*\left(G_c^{(2)}(1/z^*)\right)^*\right) . \qquad (7.206)$$

Therefore, the overall transfer function at the output of the noise whitening filter is

$$G^{(2)}(z) = G_h^{(2)}(z)G_c^{(2)}(z) . \qquad (7.207)$$

The equivalent response $G_{eq}^{(2)}(e^{j\omega})$ in (7.201) has the same amplitude as $G^{(2)}(e^{j\omega})$ but different phase. Also

$$G_{eq}^{(2)}(z)\left(G_{eq}^{(2)}(1/z^*)\right)^* = F^{(2)}(z) . \qquad (7.208)$$

Therefore, the Euclidean distance between sequences of channel outputs for the T-spaced and $T/2$-spaced systems are the same. It follows that the system shown in Fig. 7.26 has maximum likelihood performance. The main advantage of the system in Fig. 7.26 is that the noise-whitening filter does not depend on the unknown channel and has a fixed structure. The channel estimation can be performed after the noise-whitening filter and the Viterbi algorithm can be implemented using the metric in (7.129). Although the number of computations needed in the $T/2$-spaced MLSE receiver is twice that of a T-spaced receiver, the latter can not be implemented for unknown channels. Moreover, a T-spaced MLSE receiver has poor performance when it is implemented with a matched filter that is derived from an inaccurate channel estimate [252].

7.4 TIMING PHASE SENSITIVITY

The conventional MLSE receiver based on T-spaced sampling at the output of the matched filter suffers from sensitivity to the sampler timing phase [274]. We now show that a $T/2$-spaced MLSE receiver is insensitive to the sampler timing phase.

For a given a timing offset t_0, the sampled impulse response at the output of the matched filter is represented by the vector $\mathbf{f}_{t_0}^{(2)}$, where $f_{t_0,k}^{(2)} = f(kT' \pm t_0)$ and $T' = T/2$. Note that $f_{t_0,n}^{(2)} \neq \left(f_{t_0,-n}^{(2)}\right)^*$ in this case. The DTFT of $\mathbf{f}_{t_0}^{(2)}$ is

$$F_{t_0}^{(2)}(e^{j\omega}) = F^{(2)}(e^{j\omega})e^{\pm j\omega\tau_0} \qquad (7.209)$$

where $\tau_0 = t_0/T'$. If the sampler phase is known, then a discrete-time filter with response $e^{\pm j\omega\tau_0}$ after the sampler will give the symmetric signal $\mathbf{f}^{(2)}$ at its output. However, as we now show, there is no need to correct the phase.

The power spectrum of the noise at the output of the matched filter is independent of the timing offset t_0 and is given by

$$S_{\nu\nu}(f) = N_o F^{(2)}(e^{j\omega}) \ . \tag{7.210}$$

Since the DTFT of the noise-whitening filter is

$$1/(G^{(2)}(1/z^*))^*|_{z=e^{j\omega}} = 1/(G^{(2)}(e^{j\omega}))^* \tag{7.211}$$

and we have

$$F^{(2)}(e^{j\omega}) = G^{(2)}(e^{j\omega})\,(G^{(2)}(e^{j\omega}))^* = |G^{(2)}(e^{j\omega})|^2 \tag{7.212}$$

it follows from (7.210) that the noise is white at the output of the noise-whitening filter. The DTFT of the message signal at the output of the noise-whitening filter is

$$G_{t_0}^{(2)}(e^{j\omega}) = G^{(2)}(e^{j\omega})e^{\pm j\omega\tau_0} \tag{7.213}$$

and we have

$$\sum_i |g_i^{(2)}|^2 = \sum_i |g_{t_0,i}^{(2)}|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi}|G^{(2)}(e^{j\omega})|^2 d\omega \ . \tag{7.214}$$

This means that

$$G_{t_0}^{(2)}(z)(G_{t_0}^{(2)})^*(1/z^*) = F^{(2)}(z) \ . \tag{7.215}$$

Therefore, the distances between allowed sequences of channel outputs with the $T/2$-spaced MLSE receiver is insensitive to the sampler phase $e^{\pm j\omega\tau_0}$. Since the noise remains white the performance is also insensitive to the sampler phase.

8.   MIMO MLSE RECEIVERS

In this section, we derive the optimum and suboptimum MLSE receivers for co-channel demodulation of digital signals corrupted by intersymbol interference (ISI). By modeling the overall system as a discrete-time multiple-input multiple-output (MIMO) channel, the optimum MIMO MLSE receiver is derived. By following the same arguments used for single-input single-output (SISO) channels, a $T/2$-spaced MIMO MLSE receiver is shown to have the same performance as the T-spaced receiver, but with insensitivity to timing phase errors. The optimality of a practical $T/2$-spaced receiver is shown, that consists of a filter that is matched to the transmitted pulse, followed by a rate-$2/T$ sampler, a $T/2$-spaced noise whitening filter and a Viterbi algorithm. The optimum MIMO MLSE receiver requires complete knowledge of all co-channel signals. In many cases, this is impractical or even infeasible. For such cases, we discuss an interference rejection combining MLSE (IRC-MLSE) receiver.

8.1   SYSTEM AND CHANNEL MODEL

Consider a system where the signals from K co-channel signals are received by J antenna elements. This system can be modeled by a multiple-input multiple-output (MIMO) channel, where the channel inputs are the symbol sequences from the K co-channel users and the channel outputs are the combination of the signals that are received from the co-channel users at each of the J receiver antenna elements. The problem is similar to CDMA multiuser detection. However, while each user in a CDMA system uses a unique spreading sequence, the K co-channel transmitters in a TDMA system all use the same pulse shaping filter $h_a(t)$. The impulse response of the channel between the kth user and the jth antenna element is denoted by $g^{(k,j)}(t)$, where we have assumed that the channels can be modeled as time-invariant linear filters. The case of time varying channels will be considered later. While the channel introduces fading and time dispersion into the received signals, these same effects allow the co-channel signals to be distinguished at the receiver, since the received pulses $h^{(k,j)}(t) = g^{(k,j)}(t) \otimes h_a(t)$ are all distinct.

The received signal at the jth antenna element is

$$\tilde{r}^{(j)}(t) = \sum_{k=1}^{K} \sum_{\ell} x_{\ell}^{(k)} h^{(k,j)}(t - \ell T - \tau_k) + \tilde{n}^{(j)}(t) \qquad (7.216)$$

where $\tau_k, (0 \leq \tau_k \leq T)$ is the random transmission delay due to asynchronous users and $\tilde{n}^{(j)}(t)$ is additive white Gaussian noise (AWGN) assumed to be independent on the different antenna branches. For our purpose, the channel $g^{(k,j)}(t)$ is modeled by $\mathcal{L}^{(k,j)}$ arriving rays so that the impulse response has the form

$$g^{(k,j)}(t) = \sum_{n=1}^{\mathcal{L}^{(k,j)}} \alpha_n^{(k,j)} e^{j\phi_n^{(k,j)}} \delta(t - \tau_n^{(k,j)}) \qquad (7.217)$$

where $\alpha_n^{(k,j)}$, $\phi_n^{(k,j)}$, and $\tau_n^{(k,j)}$ are the amplitude, phase, and delay of the nth arriving ray from the kth transmitter at the jth antenna element. The parameters $\alpha_n^{(k,j)}$, $\phi_n^{(k,j)}$, and $\tau_n^{(k,j)}$ vary with time, but the explicit time dependency is not shown here since these parameters vary slowly compared to the baud duration.

It is safe to assume that the channel impulse responses $g^{(k,j)}(t)$, $k = 1, \ldots, K$ are uncorrelated for fixed j; however, the $g^{(k,j)}(t)$, $j = 1, \ldots J$ for fixed k are usually correlated, especially on the reverse link of a cellular system where the signals arrive at the base station with a narrow angle of arrival spread.

8.2 JOINT MAXIMUM LIKELIHOOD SEQUENCE ESTIMATION

The joint maximum likelihood sequence estimation (J-MLSE) receiver processes the total received vector

$$\tilde{\mathbf{r}}(t) = (\tilde{r}^{(1)}(t), \tilde{r}^{(2)}(t), \ldots, \tilde{r}^{(J)}(t))$$

to generate the ML estimate of the information sequence

$$\mathbf{x} = (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(K)})$$

where $\mathbf{x}^{(k)} = \{x_n^{(k)}\}$. To derive the structure of the joint ML receiver, we follow the same approach used in Section 2.1. Let $\{\varphi_n(t)\}$ denote a complete set of orthonormal basis functions defined over the interval $(0,T)$. Then

$$\tilde{r}^{(j)}(t) = \sum_{n=0}^{N-1} \tilde{r}_n^{(j)} \varphi_n(t) + \tilde{z}^{(j)}(t) \tag{7.218}$$

where

$$\tilde{r}_n^{(j)} = \sum_{k=1}^{K} \sum_{\ell} x_\ell^{(k)} h_{n_\ell}^{(k,j)} + \tilde{n}_n^{(j)} \tag{7.219}$$

where

$$h_{n_\ell}^{(k,j)} = \int_0^T h^{(k,j)}(t - \ell T - \tau_k) \varphi_n^*(t) dt \tag{7.220}$$

$$\tilde{n}_n^{(j)} = \int_0^T \tilde{n}^{(j)}(t) \varphi_n^*(t) dt \ . \tag{7.221}$$

Define the received vector

$$\tilde{\mathbf{r}} = (\tilde{\mathbf{r}}^{(1)}, \tilde{\mathbf{r}}^{(2)}, \ldots, \mathbf{r}^{(K)})$$

where $\tilde{\mathbf{r}}^{(j)} = \{\tilde{r}_n^{(j)}\}$. Since the noise components $\tilde{n}_n^{(j)}$ associated with the J antenna elements are uncorrelated zero-mean complex Gaussian random variables with variance $\frac{1}{2}E[|\tilde{n}_n^{(j)}|^2] = N_o$, the received vector $\tilde{\mathbf{r}}$ has the multivariate Gaussian density

$$p(\tilde{\mathbf{r}}|\mathbf{x}, \mathbf{H}) = \prod_{n=0}^{N-1} \frac{1}{2\pi N_o} \exp\left\{ -\frac{1}{2N_o} \sum_{j=1}^{J} \left| \tilde{r}_n^{(j)} - \sum_{k=1}^{K} \sum_{\ell} x_\ell^{(k)} h_{n_\ell}^{(k,j)} \right|^2 \right\} \ . \tag{7.222}$$

The optimum receiver chooses $\mathbf{x}$ to maximize $p(\tilde{\mathbf{w}}|\mathbf{x}, \mathbf{H})$ or, equivalently, the metric

$$\mu(\mathbf{x}) = -\sum_{n=0}^{N-1} \sum_{j=1}^{J} \left| \tilde{r}_n^{(j)} - \sum_{k=1}^{K} \sum_{\ell} x_\ell^{(k)} h_{n_\ell}^{(k,j)} \right|^2 \ . \tag{7.223}$$

Since $\sum_{n=0}^{n-1}\sum_{j=1}^{J}|\tilde{r}_n^{(j)}|^2$ is independent of $\mathbf{x}$, maximizing (7.223) is equivalent to maximizing

$$\mu(\mathbf{x}) = \sum_{j=1}^{J}\left\{2\mathrm{Re}\left\{\sum_{k=1}^{K}\sum_{\ell}x_\ell^{(k)*}\sum_{n=0}^{N-1}\tilde{r}_n^{(j)}h_{n\ell}^{(k,j)*}\right\}\right.$$
$$\left. - \sum_{k=1}^{K}\sum_{k'=1}^{K}\sum_{\ell}\sum_{\ell'}x_\ell^{(k)}x_{\ell'}^{(k')*}\sum_{n=0}^{N-1}h_{n\ell}^{(k,j)}h_{n\ell'}^{(k',j)*}\right\} . \quad (7.224)$$

To proceed further, define

$$y_\ell^{(k,j)} \triangleq \lim_{N\to\infty}\sum_{n=0}^{N-1}w_n^{(j)}h_{n\ell}^{(k,j)*}$$
$$= \int_{-\infty}^{\infty}w^{(j)}(t)h^{(k,j)*}(t-\ell T-\tau_k)dt \quad (7.225)$$

$$f_{\ell'-\ell}^{(k,k',j)} \triangleq \lim_{N\to\infty}\sum_{n=0}^{N-1}h_{n\ell}^{(k,j)}h_{n\ell'}^{(k',j)*}$$
$$= \int_{-\infty}^{\infty}h^{(k,j)}(t-\ell T-\tau_k)h^{(k',j)*}(t-\ell'T-\tau_{k'})dt \quad (7.226)$$

giving

$$\mu(\mathbf{x}) = \sum_{j=1}^{J}\left\{2\mathrm{Re}\left\{\sum_{k=1}^{K}\sum_{\ell}x_\ell^{(k)*}y_\ell^{(k,j)}\right\}\right.$$
$$\left. - \sum_{k=1}^{K}\sum_{k'=1}^{K}\sum_{\ell}\sum_{\ell'}x_\ell^{(k)}x_{\ell'}^{(k')*}f_{\ell'-\ell}^{(k,k',j)}\right\} . \quad (7.227)$$

It follows that the ML receiver processes the $\{y_\ell^{(k,j)}\}$ and $\{f_\ell^{(k,k',j)}\}$ to determine the most likely sequence. The integral in (7.225) represents the output of the kth matched filter at the jth antenna element and the $f_\ell^{(k,k',j)}$ are the ISI coefficients for the jth antenna element. It follows that the ML receiver employs a bank of K matched filters at each of the J antenna elements and combines together all JK matched filter outputs to generate the ML estimate of the transmitted sequence. From (7.225) and (7.226), we note that the ML receiver requires knowledge of the $\{h^{(k,j)}(t)\}$ and the $\{\tau_k\}$.

The noise samples at the output of the matched filter $h^{(k,j)*}(t)$ are

$$\nu_\ell^{(k,j)} = \nu^{(k,j)}(\ell T) = \int_{-\infty}^{\infty}\tilde{n}^{(j)}(t)h^{(k,j)*}(t-\ell T-\tau_k)dt \quad (7.228)$$

which from (7.226) have autocorrelation function

$$\phi_{\nu\nu}^{(k,j)}(n) = \frac{1}{2}\mathrm{E}\left[\nu_{\ell+n}^{(k,j)}\nu_\ell^{(k,j)*}\right] = N_o f_n^{(k,k,j)} . \quad (7.229)$$

This leads to the overall discrete-time model

$$y_\ell^{(k,j)} = \sum_{k'=1}^{K} \sum_{n} x_n^{(k')} f_{\ell-n}^{(k',k,j)} + \nu_\ell^{(k,j)} \ . \tag{7.230}$$

If we define

$$\mathbf{y}_\ell^{(j)} = (y_\ell^{(1,j)}, \ldots, y_\ell^{(K,j)}) \tag{7.231}$$

$$\mathbf{x}_\ell = (x_\ell^{(1)}, \ldots, x_\ell^{(K)}) \tag{7.232}$$

$$\boldsymbol{\nu}_\ell^{(j)} = (\nu_\ell^{(1,j)}, \ldots, \nu_\ell^{(K,j)}) \tag{7.233}$$

$$\mathbf{F}_m^{(j)} = \left[f_m^{(k',k,j)} \right]_{K \times K} \tag{7.234}$$

then (7.230) has the convenient matrix form

$$\mathbf{y}_\ell^{(j)} = \sum_{n} \mathbf{x}_{\ell-n} \mathbf{F}_n^{(j)} + \boldsymbol{\nu}_\ell^{(j)} \ . \tag{7.235}$$

The above development leads to the overall system model shown in Fig. 7.27.
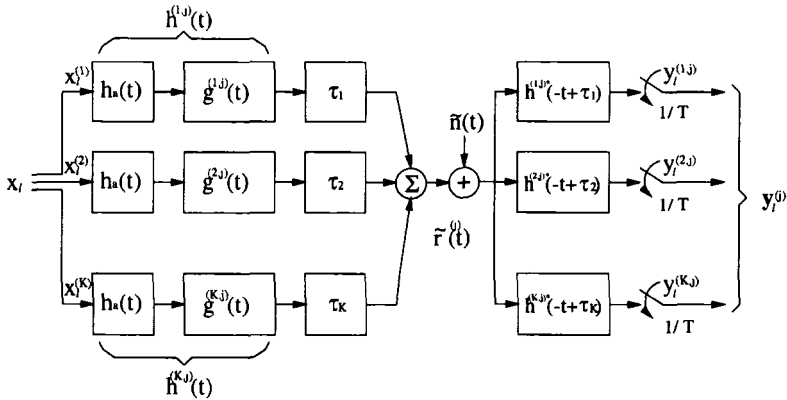

Figure 7.27. Overall MIMO system model.

8.3 DISCRETE-TIME MIMO CHANNEL MODEL

Analogous to the discrete-time model for digital signaling on a single input – single output (SISO) ISI channel, the MIMO ISI channel can also be described by a collection of *J* parallel discrete-time white noise channel models. A *matrix noise whitening filter* is used to whiten the noise samples at the outputs of the

bank of matched filters for each antenna element. Suppose that the channel impulse response $h^{(k,j)}(t)$ has length $L^{(k,j)}T$, i.e., $h^{(k,j)}(t) = 0$ for $t < 0$ and $t > L^{(k,j)}T$. Then the channel matrix filter for the *jth* antenna element can be defined as [93]

$$\mathbf{F}^{(j)}(z) = \sum_{n=-L_j}^{L_j} \mathbf{F}_n^{(j)} z^{-n} \tag{7.236}$$

where $L_j = \max_k L^{(k,j)}$. This is a straight forward extension of the conventional SISO ISI channel where $F(z) = \sum_{n=-L+1}^{L-1} f_n z^{-n}$. An asynchronous MIMO channel is described by a matrix filter and the range of summation in (7.236) must be expanded from $(-L_j + 1, L_j - 1)$ to $(-L_j, L_j)$ in order to account for the random user delays. For an ideal channel (with no ISI) the channel matrix filter in (7.236) is

$$\mathbf{F}^{(j)}(z) = \sum_{n=-N}^{N} \mathbf{F}_n^{(j)} z^{-n} \tag{7.237}$$

where NT is the length of the transmitted pulse, i.e., $h_a(t) = 0$ for $t < 0$ and $t > NT$. For a *synchronous* ideal MIMO channel

$$\mathbf{F}^{(j)}(z) = \sum_{n=-N+1}^{N-1} \mathbf{F}_n^{(j)} z^{-n} \tag{7.238}$$

By using (7.226), it can be shown that the ISI coefficients have the symmetric property

$$f_n^{(k,k',j)} = f_{-n}^{(k',k,j)*} \tag{7.239}$$

and, therefore, $\mathbf{F}_n^{(j)} = \mathbf{F}_{-n}^{(j)H}$, where H denotes Hermitian transposition. Hence, $\mathbf{F}^{(j)}(z)$ has the symmetric form

$$\mathbf{F}^{(j)}(z) = \mathbf{F}^{(j)H}(1/z^*) . \tag{7.240}$$

It follows that the channel matrix filter $\mathbf{F}^{(j)}(z)$ can be factored as

$$\mathbf{F}^{(j)}(z) = \mathbf{G}^{(j)}(z)\mathbf{G}^{(j)H}(1/z^*) . \tag{7.241}$$

Example 7.9
Consider a two user system with a single receiver antenna. Since $J = 1$, we can omit the index (j). Let $L = 1$ and $\mathbf{F}(z)$ be

$$
\begin{aligned}
\mathbf{F}(z) &= \mathbf{F}_1^H z + \mathbf{F}_0 + \mathbf{F}_1 z^{-1} \\
&= \begin{bmatrix} 1 & 0.48 + 0.48z^{-1} \\ 0.48 + 0.48z & 1 \end{bmatrix} \\
&= \begin{bmatrix} 0 & 0 \\ 0.48 & 0 \end{bmatrix} z + \begin{bmatrix} 1 & 0.48 \\ 0.48 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0.48 \\ 0 & 0 \end{bmatrix} z^{-1} .
\end{aligned}
$$

The matrix spectral factorization of $\mathbf{F}(z)$ has the form

$$
\begin{aligned}
\mathbf{F}(z) &= \mathbf{G}^H(1/z^*)\mathbf{G}(z) \\
&= [\mathbf{G}_0 + \mathbf{G}_1 z^*]^H \left[\mathbf{G}_0 + \mathbf{G}_1 z^{-1}\right] .
\end{aligned}
$$

$\mathbf{F}_0$ and $\mathbf{F}_1$ can be represented by

$$
\begin{aligned}
\mathbf{F}_0 &= \mathbf{G}_0^H \mathbf{G}_0 + \mathbf{G}_1^H \mathbf{G}_1 \\
\mathbf{F}_1 &= \mathbf{G}_0^H \mathbf{G}_1 ,
\end{aligned}
$$

where $\mathbf{G}_0$ is lower triangular and $\mathbf{F}_1$ is upper triangular with zero diagonal. In turn, $\mathbf{G}_1$ must be upper triangular with zero diagonal. This results in the spectral factorization

$$
\mathbf{F}(z) = \left[\begin{bmatrix} 0.8 & 0 \\ 0.6 & 0.8 \end{bmatrix} + \begin{bmatrix} 0 & 0.6 \\ 0 & 0 \end{bmatrix} z^* \right]^H \left[\begin{bmatrix} 0.8 & 0 \\ 0.6 & 0.8 \end{bmatrix} + \begin{bmatrix} 0 & 0.6 \\ 0 & 0 \end{bmatrix} z^{-1}\right]
$$

The matrix noise whitening filter $[\mathbf{G}^H(1/z^*)]^{-1}$ is anticausal and stable with an infinite length.

$$
\begin{aligned}
[\mathbf{G}^H(1/z^*)]^{-1} &= \begin{bmatrix} 0.8 & 0.6 \\ 0.6z & 0.8 \end{bmatrix}^{-1} \\
&= \frac{1}{0.64 - 0.36z} \begin{bmatrix} 0.8 & -0.6 \\ -0.6z & 0.8 \end{bmatrix} .
\end{aligned}
$$

In practice the filter $[\mathbf{G}^H(1/z^*)]^{-1}$ can be approximated as a finite length filter with sufficient delay. Finally, the overall discrete-time white noise matrix channel has transfer function

$$
\mathbf{G}(z) = \begin{bmatrix} 0.8 & 0 \\ 0.6 & 0.8 \end{bmatrix} + \begin{bmatrix} 0 & 0.6 \\ 0 & 0 \end{bmatrix} z^{-1}
$$

As an alternative, it is possible to choose $[\mathbf{G}(z)]^{-1}$ as the matrix noise whitening filter. In this case the matrix noise whitening filter is stable and causal. This fact makes no difference in the performance of an MIMO MLSE receiver. However, it is important to choose the matrix noise whitening filter as $[\mathbf{G}^H(1/z^*)]^{-1}$ for some reduced complexity co-channel receivers such as those that employ reduced-state sequence estimation (RSSE) [110], delayed decision feedback sequence estimation (DDFSE) [90], or symbol-by-symbol co-channel receivers that employ decision feedback MIMO equalization.

If the sequence $\tilde{\mathbf{r}}^{(j)}(z)$ is input to the matrix noise whitening filter, the output is

$$
\mathbf{v}^{(j)}(z) = \mathbf{G}^{(j)}(z)\mathbf{x}(z) + \boldsymbol{\eta}^{(j)}(z) \tag{7.242}
$$

where $\eta^{(j)}(z)$ is white Gaussian noise with power spectral density

$$S_{\boldsymbol{\eta}^{(j)}\boldsymbol{\eta}^{(j)}}(f) = N_o \mathbf{I} \ . \tag{7.243}$$

In the time domain

$$\mathbf{v}_{\ell}^{(j)} = \sum_{n=0}^{L_j} \mathbf{x}_{\ell-n} \mathbf{G}_n^{(j)} + \boldsymbol{\eta}_{\ell}^{(j)} \tag{7.244}$$

where

$$\mathbf{v}_{\ell}^{(j)} = (v_{\ell}^{(1,j)}, \ldots, v_{\ell}^{(K,j)}) \tag{7.245}$$

$$\boldsymbol{\eta}_{\ell}^{(j)} = (\eta_{\ell}^{(1,j)}, \ldots, \eta_{\ell}^{(K,j)}) \ . \tag{7.246}$$

The optimum receiver consists of a bank of K matched filters at the output of each antenna element, followed by a baud-rate sampler and a $K \times K$ matrix noise whitening filter. With J-branch diversity reception, the overall matrix channel consisting of the transmit filters, channels, matched filters, samplers, and matrix noise whitening filters, can be modeled as a parallel collection of J T-spaced matrix filters with independent white noise sequences as shown in Fig. 7.28. To determine the number of states in the overall channel model, we first define $L_k = \max_j L^{(k,j)}$ as the length of the channel memory for the kth input. Then there are $2^n \sum_{k=1}^{K} L_k$ states, where 2^n is the size of the signal constellation.
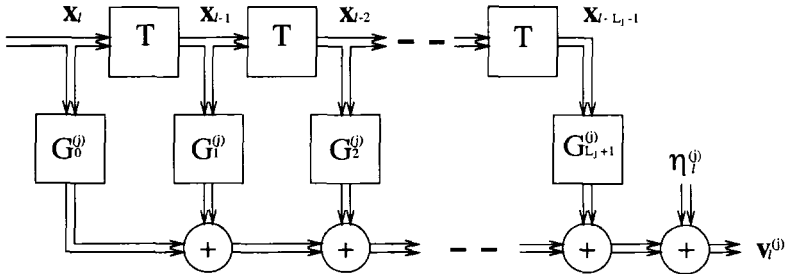

Figure 7.28. Discrete-time white noise MIMO channel model.

Similar to the receiver derived by Ungerboeck [328], it is possible to implement the ML receiver by directly operating on the sequences $\{\mathbf{y}_{\ell}^{(j)}\}$ at the outputs of the matched filters, thus eliminating the need for the matrix noise whitening filters. The metric for Ungerboeck's receiver can be obtained from (7.227).

8.4 THE VITERBI ALGORITHM

Suppose that m symbols from each of the K transmitters have been transmitted over the channel. Let $\mathbf{V}_n = (\mathbf{v}_n^{(1)}, \mathbf{v}_n^{(2)}, \ldots, \mathbf{v}_n^{(J)})$, where $\mathbf{v}_n^{(j)} = (v_n^{(1,j)}, \ldots, v_n^{(K,j)})$ denote the collection of vectors at the outputs of the matrix noise whitening filters on each of the J antenna branches at epoch n. After receiving the output sequence $\{\mathbf{V}_n\}_{n=1}^m$, the ML receiver decides in favor of the sequence of input vectors $\{\mathbf{x}_n\}_{n=1}^m$ that maximizes the log-likelihood function

$$\log p(\mathbf{V}_m, \ldots, \mathbf{V}_1 | \mathbf{x}_m, \ldots, \mathbf{x}_1)$$
$$= \log p(\mathbf{V}_m | x_m^{(1)}, \ldots, x_{m-L_1}^{(1)}; x_m^{(2)}, \ldots, x_{m-L_2}^{(2)}; \ldots; x_m^{(K)}, \ldots, x_{m-L_K}^{(K)})$$
$$+ \log p(\mathbf{V}_{m-1}, \ldots, \mathbf{V}_1 | \mathbf{x}_{m-1}, \ldots, \mathbf{x}_1) \qquad (7.247)$$

The first term on the right hand side of (7.247) is the *branch metric* used in the Viterbi algorithm. The discrete-time white noise matrix channel model leads to the conditional density function

$$\log p(\mathbf{V}_m | x_m^{(1)}, \ldots, x_{m-L_1}^{(1)}; x_m^{(2)}, \ldots, x_{m-L_2}^{(2)}; \ldots; x_m^{(K)}, \ldots, x_{m-L_K}^{(K)})$$
$$= \frac{1}{(\pi N_o)^{KJ}} \exp \left\{ -\frac{1}{N_o} \sum_{j=1}^J \| \mathbf{v}_m^{(j)} - \sum_{n=0}^L \mathbf{x}_{m-n} \mathbf{G}_n^{(j)} \|^2 \right\} \qquad (7.248)$$

where $L = \max_k L_k$. Note that some elements in the matrix $\mathbf{G}_n^{(j)}$ may be zero if $L_k \neq L$, $k = 1, \ldots, K$ in which case the branch metric computation can be simplified. The density in (7.248) leads to the branch metric

$$\mu_m = -\sum_{j=1}^J \| \mathbf{v}_m^{(j)} - \sum_{n=0}^L \mathbf{x}_{m-n} \mathbf{G}_n^{(j)} \|^2 . \qquad (7.249)$$

8.5 PAIRWISE ERROR PROBABILITY

Let $\mathbf{x}$ and $\hat{\mathbf{x}}$ be the transmitted and estimated symbol sequences, respectively, and define the error sequence $\boldsymbol{\epsilon} = \mathbf{x} - \hat{\mathbf{x}}$. The pairwise error probability is the probability that the receiver decides in favor of sequence $\hat{\mathbf{x}}$ when sequence $\mathbf{x}$ was transmitted, equal to

$$P(\Gamma(\mathbf{x} - \boldsymbol{\epsilon}) \geq \Gamma(\mathbf{x}) | \mathbf{x}) = P(\Gamma(\hat{\mathbf{x}}) > \Gamma(\mathbf{x})) \qquad (7.250)$$

where $\Gamma(\mathbf{x}) = \sum_m \mu_m$ is the path metric associated with the input sequence $\mathbf{x}$ with the branch metric μ_m defined in (7.249). From (7.249),

$$P(\Gamma(\mathbf{x}-\boldsymbol{\epsilon}) \geq \Gamma(\mathbf{x})|\mathbf{x}) = P\left(\sum_m \sum_{j=1}^{J} \| \sum_{n=0}^{L} \boldsymbol{\epsilon}_{m-n} \mathbf{G}_n^{(j)} + \boldsymbol{\eta}_m^{(j)} \|^2 \right.$$

$$< \sum_m \sum_{j=1}^{J} \| \boldsymbol{\eta}_m^{(j)} \|^2 \bigg)$$

$$= P\left(\sum_m \sum_{j=1}^{J} 2\mathrm{Re}\left\{ \sum_{n=0}^{L} \boldsymbol{\epsilon}_{m-n} \mathbf{G}_n^{(j)} \boldsymbol{\eta}_m^{(j)^H} \right\} \right.$$

$$< - \sum_m \sum_{j=1}^{J} \| \sum_{n=0}^{L} \boldsymbol{\epsilon}_{m-n} \mathbf{G}_n^{(j)} \|^2 \bigg) \quad . \quad (7.251)$$

Define

$$\Delta^2 = \sum_m \sum_{j=1}^{J} \| \sum_{n=0}^{L} \boldsymbol{\epsilon}_{n-m} \mathbf{G}_n^{(j)} \|^2 \qquad (7.252)$$

$$\chi = \sum_m \sum_{j=1}^{J} 2\mathrm{Re}\left\{ \sum_{n=0}^{L} \boldsymbol{\epsilon}_{m-n} \mathbf{G}_n^{(j)} \boldsymbol{\eta}_m^{(j)^H} \right\} \qquad (7.253)$$

It can be shown that χ is a zero-mean Gaussian random variable with variance $4N_o\Delta^2$. Therefore, the pairwise error probability becomes

$$P(\Gamma(\mathbf{x}-\boldsymbol{\epsilon}) \geq \Gamma(\mathbf{x})|\mathbf{x}) = Q\left(\sqrt{\frac{\Delta^2}{4N_o}} \right) \quad . \qquad (7.254)$$

8.6 *T/2*-SPACED MIMO MLSE RECEIVER

Suppose that the matched filter outputs $y^{(k,j)}(t)$ are sampled at the correct timing phase but with rate $2/T$. In this case, the discrete-time channel from the *kth* input to the *jth* output can be described by a $T/2$-spaced transversal filter with coefficients $\tilde{f}_\ell^{(k,k',j)} = f(\ell T/2)$ and $\tilde{f}_\ell^{(k,k',j)} = \tilde{f}_{-\ell}^{(k',k,j)^*}$, where the tilde denotes rate $2/T$ sampling. Since the timing phase is correct, we have $f_\ell^{(k,k',j)} = \tilde{f}_{2\ell}^{(k,k',j)}$. It follows that the overall discrete-time matrix channel filter, denoted by $\tilde{\mathbf{F}}^{(j)}(z)$ has the factorization

$$\tilde{\mathbf{F}}^{(j)}(z) = \tilde{\mathbf{G}}^{(j)}(z)\tilde{\mathbf{G}}^{(j)^H}(1/z^*) \quad . \qquad (7.255)$$

As with baud-rate sampling, the $T/2$-spaced correlated noise samples can be whitened by using a stable anticausal matrix noise whitening filter with the transfer function $\left[\tilde{\mathbf{G}}^{(j)^H}(1/z^*) \right]^{-1}$. The output of the matrix noise whitening filter is

$$\tilde{\mathbf{v}}^{(j)}(z) = \tilde{\mathbf{G}}^{(j)}(z)\tilde{\mathbf{x}}(z) + \tilde{\boldsymbol{\eta}}^{(j)}(z) \qquad (7.256)$$

or in the time domain

$$\tilde{\mathbf{v}}_\ell^{(j)} = \sum_{n=0}^{2L_j} \tilde{\mathbf{x}}_{\ell-n} \tilde{\mathbf{G}}_n^{(j)} + \tilde{\boldsymbol{\eta}}_\ell^{(j)} \tag{7.257}$$

where $\{\tilde{\boldsymbol{\eta}}_\ell^{(j)}\}$ is a $T/2$-spaced white noise sequence with power spectrum $S_{\tilde{\eta}\tilde{\eta}}(f) = N_o \mathbf{I}$, The sequence $\{\tilde{\mathbf{x}}_n\}$ is the corresponding $T/2$-spaced input symbol sequence and is given by

$$\tilde{\mathbf{x}}_n = \begin{cases} \mathbf{x}_{n/2} & , \quad n \text{ even} \\ 0 & , \quad n \text{ odd} \end{cases} \tag{7.258}$$

The overall system and equivalent discrete-time white noise models are shown in Figs. 7.29 and 7.30, respectively.
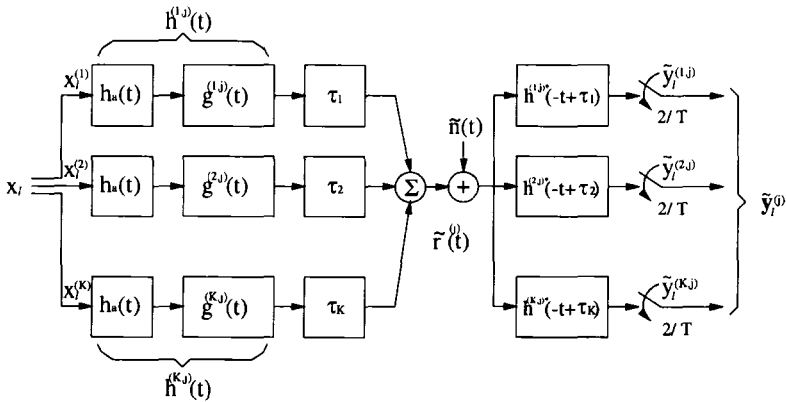
Figure 7.29. Overall MIMO system model with $T/2$-spaced sampling.

Note that the vector samples $\tilde{\mathbf{v}}_{2\ell}^{(j)}$ and $\tilde{\mathbf{v}}_{2\ell+1}^{(j)}$ correspond to the ℓth received baud, where

$$\tilde{\mathbf{v}}_{2\ell}^{(j)} = \sum_{n=0}^{L_j} \mathbf{x}_{\ell-n} \tilde{\mathbf{G}}_{2n}^{(j)} + \tilde{\boldsymbol{\eta}}_{2\ell}^{(j)} \tag{7.259}$$

$$\tilde{\mathbf{v}}_{2\ell+1}^{(j)} = \sum_{n=0}^{L_j-1} \mathbf{x}_{\ell-n} \tilde{\mathbf{G}}_{2n+1}^{(j)} + \tilde{\boldsymbol{\eta}}_{2\ell+1}^{(j)} . \tag{7.260}$$
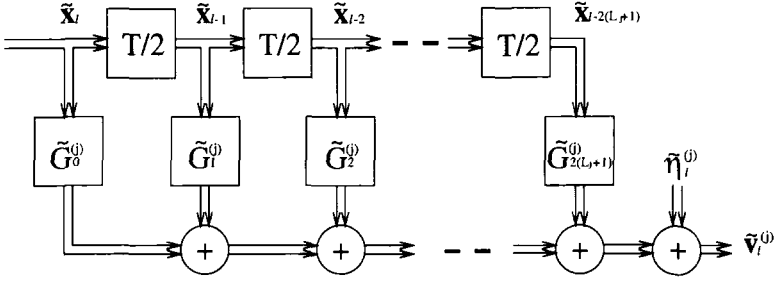
Figure 7.30.   Overall discrete-time white noise MIMO channel model with $T/2$-spaced sampling.

With $T/2$-spaced fractional sampling there are two samples per baud and the branch metric becomes

$$
\mu_m = -\sum_{j=1}^{J} \left\{ \|\tilde{\mathbf{v}}_{2m}^{(j)} - \sum_{n=0}^{L} \mathbf{x}_{m-n} \tilde{\mathbf{G}}_{2n}^{(j)} \|^2 \right.
$$

$$
\left. + \|\tilde{\mathbf{v}}_{2m+1}^{(j)} - \sum_{n=0}^{L-1} \mathbf{x}_{m-n} \tilde{\mathbf{G}}_{2n+1}^{(j)} \|^2 \right\} . \tag{7.261}
$$

Once again, if $L_k \neq L, k = 1, \ldots, K$ then some of the $\tilde{\mathbf{G}}_{2n}^{(j)}$ may be zero. Notice that $T/2$-spaced fractional sampling doubles the number of computations in forming the branch metrics as compared to T-spaced sampling.

8.6.1   ERROR PROBABILITY

We now generalize the result for SISO channels, and show that the T-spaced and $T/2$-spaced MIMO MLSE receivers have identical performance. For T-spaced sampling, define

$$
\mathbf{E}(z) = \sum_{n} \epsilon_n z^{-n} \tag{7.262}
$$

$$
\mathbf{E}_v^{(j)}(z) = \sum_{n} \mathbf{v}_n^{(j)} z^{-n} \tag{7.263}
$$

Then

$$
\mathbf{E}_v^{(j)}(z) = \mathbf{E}(z) \mathbf{G}^{(j)}(z) \tag{7.264}
$$

and

$$\Delta^2 \;=\; \sum_m \sum_{j=1}^{J} \|\sum_{n=0}^{L} \boldsymbol{\epsilon}_{n-m} \mathbf{G}_n^{(j)}\|^2 \tag{7.265}$$

$$=\; \sum_{j=1}^{J} \left[\mathbf{E}_v^{(j)}(z)\mathbf{E}_v^{(j)H}(1/z^*) \right]_0$$

$$=\; \sum_{j=1}^{J} \left[\mathbf{E}(z)\mathbf{G}^{(j)}(z)\mathbf{G}^{(j)H}(1/z^*)\mathbf{E}^*(1/z^*) \right]_0$$

$$=\; \sum_{j=1}^{J} \left[\mathbf{E}(z)\mathbf{F}^{(j)}(z)\mathbf{E}^*(1/z^*) \right]_0 \tag{7.266}$$

For $T/2$-spaced sampling, define

$$\tilde{\mathbf{E}}(z) = \sum_n \tilde{\boldsymbol{\epsilon}}_n z^{-2n} \quad . \tag{7.267}$$

Since $\tilde{\boldsymbol{\epsilon}}_n = \mathbf{x}_n - \hat{\mathbf{x}}_n$ is zero for even k, we have $\tilde{\mathbf{E}}(z) = \mathbf{E}(z^2)$. Also,

$$\tilde{\mathbf{E}}_v^{(j)}(z) = \tilde{\mathbf{E}}(z)\tilde{\mathbf{G}}^{(j)}(z) \tag{7.268}$$

Therefore

$$\tilde{\Delta}^2 \;=\; \sum_m \sum_{j=1}^{J} \|\sum_{n=0}^{L} \tilde{\boldsymbol{\epsilon}}_{n-m} \tilde{\mathbf{G}}_n^{(j)}\|^2 \tag{7.269}$$

$$=\; \sum_{j=1}^{J} \left[\tilde{\mathbf{E}}_v^{(j)}(z)\tilde{\mathbf{E}}_v^{(j)H}(1/z^*) \right]_0$$

$$=\; \sum_{j=1}^{J} \left[\tilde{\mathbf{E}}(z)\mathbf{G}^{(j)}(z)\mathbf{G}^{(j)H}(1/z^*)\tilde{\mathbf{E}}^*(1/z^*) \right]_0$$

$$=\; \sum_{j=1}^{J} \left[\tilde{\mathbf{E}}(z)\mathbf{F}^{(j)}(z)\tilde{\mathbf{E}}^*(1/z^*) \right]_0$$

$$=\; \sum_{j=1}^{J} \left[\mathbf{E}(z^2)\mathbf{F}^{(j)}(z)\mathbf{E}^*(1/z^{*^2}) \right]_0 \tag{7.270}$$

where $[\,\cdot\,]_0$ is the coefficient of z^0. Since the odd powers of $\mathbf{E}(z^2)\mathbf{E}^*(1/z^{*^2})$ are zero and $\mathbf{F}_\ell = \tilde{\mathbf{F}}_{2\ell}$ we have $\tilde{\Delta}^2 = \Delta^2$. Therefore, the T-spaced and $T/2$-spaced receivers have identical bit error probability performance.

8.6.2 TIMING PHASE SENSITIVITY

The T-spaced MIMO MLSE receiver must have knowledge of the set of delays $\{\tau_k\}_{k=1}^{K}$ to generate the branch metrics. One of the greatest advantages

of a $T/2$-spaced MIMO MLSE receiver is its insensitivity to timing phase. Section 7. showed this property for SISO channels and here we generalize the result to MIMO channels.

Suppose that the timing phase offset for the kth sampler and the jth antenna branch is $t^{(k,j)}$ seconds. The $T/2$-spaced sampled impulse response at the output of the matched filter $h^{(k,j)*}(-t)$ is $\tilde{f}_{\ell,t}^{(k,k',j)} = \tilde{f}^{(k,k',j)}(\ell T' \pm t^{(k,j)})$, where $T' = T/2$. Due to the timing phase offset, the ISI coefficients are not symmetric, i.e., $\tilde{f}_{\ell,t}^{(k,k',j)} \neq \tilde{f}_{-\ell,t}^{(k,k',j)*}$. Define the matrices

$$\tilde{\mathbf{F}}_n^{(j)} = \left[\tilde{f}_n^{(k,k',j)}\right]_{K \times K} \tag{7.271}$$

$$\tilde{\mathbf{F}}_{n,t}^{(j)} = \left[\tilde{f}_{n,t}^{(k,k',j)}\right]_{K \times K} . \tag{7.272}$$

The discrete-time Fourier transform (DTFT) of $\tilde{\mathbf{F}}_{n,t}^{(j)}$ is

$$\begin{aligned}
\tilde{\mathbf{F}}_{t}^{(j)}(e^{j\omega}) &= \sum_{n=-2L_j}^{2L_j} \tilde{\mathbf{F}}_{n,t}^{(j)}(e^{j\omega})^{-n} \\
&= e^{\pm j\omega \boldsymbol{\tau}^{(j)}} \tilde{\mathbf{F}}^{(j)}(e^{j\omega})
\end{aligned} \tag{7.273}$$

where $e^{\pm j\omega \boldsymbol{\tau}^{(j)}} = (e^{\pm j\omega \tau^{(1,j)}}, \ldots, e^{\pm j\omega \tau^{(K,j)}})$ and $\tau^{(k,j)} = t^{(k,j)}/T'$.

Since the noise is circularly symmetric, the psd of the noise at the output of the *jth* matched filter is independent of the timing offset $\mathbf{t}$ and is given by

$$S_{\boldsymbol{\nu}^{(j)}\boldsymbol{\nu}^{(j)}}(f) = N_o \tilde{\mathbf{F}}^{(j)}(e^{j\omega}) \tag{7.274}$$

The DTFT of the matrix noise whitening filter is

$$\left[\tilde{\mathbf{G}}^{(j)H}(1/z^*)\right]_{z=e^{j\omega}}^{-1} = \left[\tilde{\mathbf{G}}^{(j)H}(e^{j\omega})\right]^{-1} \tag{7.275}$$

and we have

$$\tilde{\mathbf{F}}^{(j)}(e^{j\omega}) = \tilde{\mathbf{G}}^{(j)}(e^{j\omega})\tilde{\mathbf{G}}^{(j)H}(e^{j\omega}) . \tag{7.276}$$

Hence, the noise at the output of the matrix noise whitening filter is white. Since the input data sequence is white, the DTFT of the message vector at the output of the noise whitening filter is

$$\tilde{\mathbf{G}}_{t}^{(j)}(e^{j\omega}) = e^{\pm j\omega \boldsymbol{\tau}^{(j)}} \tilde{\mathbf{G}}^{(j)}(e^{j\omega}) \tag{7.277}$$

and we have

$$\sum_n \tilde{\mathbf{G}}_n^{(j)} \tilde{\mathbf{G}}_n^{(j)H} = \sum_n \tilde{\mathbf{G}}_{n,t}^{(j)} \tilde{\mathbf{G}}_{n,t}^{(j)H} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \tilde{\mathbf{G}}^{(j)}(e^{j\omega})\tilde{\mathbf{G}}^{(j)H}(e^{j\omega})d\omega . \tag{7.278}$$

This implies that

$$\tilde{\mathbf{G}}_{\mathbf{t}}^{(j)}(z)\tilde{\mathbf{G}}_{\mathbf{t}}^{(j)^H}(1/z^*) = \tilde{\mathbf{F}}^{(j)}(z) \ . \tag{7.279}$$

It follows that the distances between allowed sequences of channel outputs with the $T/2$-spaced MLSE receiver is not sensitive to the sampler phase $e^{\pm j\omega \boldsymbol{\tau}^{(j)}}$. Since the noise remains white, the error rate performance is insensitive to the sampler phase. Finally, we note that (7.278) does not hold for the T-spaced receiver due to aliasing of the signal spectrum.

8.6.3 PRACTICAL RECEIVER

Section 7. showed that the optimal front-end processing for a SISO ISI channel can be realized by a receiver filter that is matched to the transmitted pulse $h_a(t)$ followed by a rate-$2/T$ sampler and a $T/2$-spaced noise whitening filter. Here we generalize this concept to MIMO ISI channels. For a MIMO system where all input signals have the same form, a significant complexity reduction is realized by using this receiver. No longer is a matched filter bank required at each antenna element. As shown in Fig. 7.31, the receiver simply consists of a single matched filter for each antenna element followed by a rate-$2/T$ sampler and a $T/2$-spaced noise whitening filter. Although the T-spaced samples at the output of the filter $h_a^*(-t)$ are white, the $T/2$-spaced samples are not and, therefore, the noise whitening filter is necessary. However, the structure of the noise whitening filter is completely known because it depends on the known filter $h_a^*(-t)$.
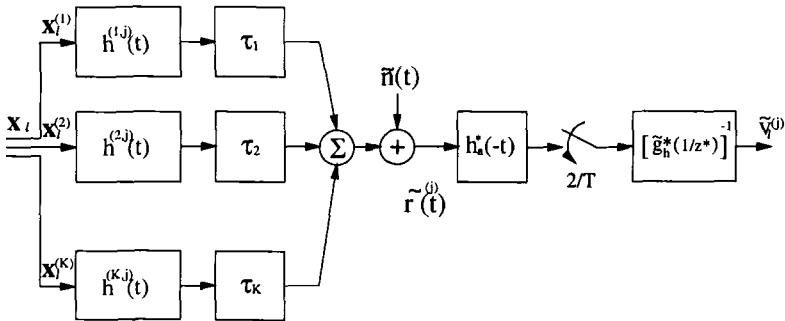

Figure 7.31. Practical MIMO system with $T/2$-spaced sampling.

We now establish that the systems shown in Figs. 7.29 and 7.31 yield identical performance. Assuming that rate-$2/T$ sampling satisfies the sampling theorem, the two systems can be completely represented by their $T/2$-spaced discrete-time signals. This is achieved, for example, by using raised cosine pulse

shaping with less than 100% excess bandwidth. Define

$$\mathbf{c}^{(j)}(t) \;=\; (c^{(1,j)}(t), \ldots, c^{(K,j)}(t))^T \tag{7.280}$$

$$\mathbf{h}^{(j)}(t) \;=\; (h^{(1,j)}(t), \ldots, h^{(K,j)}(t))^T \tag{7.281}$$

Let $\tilde{h}_a(z)$, $\tilde{\mathbf{c}}(z)$, and $\tilde{\mathbf{h}}(z)$ be the z-transforms of the $T/2$-spaced sample signals corresponding to $h_a(t)$, and $\mathbf{c}(t)$ and $\mathbf{h}(t)$, respectively. The z-transform of the autocorrelation function of the noise samples at the output of the receive filter $h_a^*(-t)$ is $N_o \tilde{F}_h(z)$ where $\tilde{F}_h(z) = \tilde{h}_a(z)\tilde{h}_a^*(1/z^*)$. Using the factorization

$$\tilde{F}_h(z) = \tilde{g}_h(z)\tilde{g}_h^*(1/z^*) \tag{7.282}$$

the $T/2$-spaced noise sequence at the output of the matched filter $h_a^*(-t)$ can be whitened by using a filter having the transfer function $[\tilde{g}_h^*(1/z^*)]^{-1}$ as shown in Fig. 7.31. Note that the noise whitening filter is not a matrix filter, but just a scalar filter.

The z-transform of the overall $T/2$-spaced discrete-time channel in Fig. 7.31 that includes the noise-whitening filter is

$$\begin{aligned}\tilde{\mathbf{g}}_{eq}(z) \;&=\; \tilde{h}_a(z)\tilde{\mathbf{c}}(z)\tilde{h}_a^*(1/z^*)/\tilde{g}_h^*(1/z^*)\\ &=\; \tilde{\mathbf{c}}(z)\tilde{g}_h(z) \;.\end{aligned} \tag{7.283}$$

Referring to the conventional system shown in Fig. 7.29, we have

$$\tilde{\mathbf{h}}(z) = \tilde{h}_a(z)\tilde{\mathbf{c}}(z) \tag{7.284}$$

and

$$\tilde{\mathbf{F}}(z) = \tilde{h}_a(z)\tilde{\mathbf{c}}(z)(\tilde{h}_a(1/z^*)\mathbf{c}(1/z^*))^H = \tilde{h}_a(z)\tilde{\mathbf{c}}(z)\tilde{\mathbf{c}}^H(1/z^*)\tilde{h}_a^*(1/z^*) \;. \tag{7.285}$$

Let

$$\tilde{\mathbf{c}}(z)\tilde{\mathbf{c}}^H(1/z^*) = \tilde{\mathbf{G}}_c(z)\mathbf{G}_c^H(1/z^*) \tag{7.286}$$

be a factorization of the matrix $\tilde{\mathbf{c}}(z)\tilde{\mathbf{c}}^H(1/z^*)$ such that $\tilde{\mathbf{G}}_c^H(1/z^*)$ has minimum phase. Combining (7.282), (7.285) and (7.286) gives

$$\tilde{\mathbf{F}}(z) = \tilde{g}_h(z)\tilde{\mathbf{G}}_c^H(z)\tilde{\mathbf{G}}_c^H(1/z^*)\tilde{g}_h^*(1/z^*) \;. \tag{7.287}$$

The transfer function of the matrix noise-whitening filter is chosen as

$$\left[\tilde{\mathbf{G}}_c^H(1/z^*)\tilde{g}_h^*(1/z^*)\right]^{-1} \;. \tag{7.288}$$

Therefore, the overall transfer function at the output of the matrix noise whitening filter is

$$\tilde{\mathbf{G}}(z) = \tilde{g}_h(z)\tilde{\mathbf{G}}_c(z) \;. \tag{7.289}$$

Finally, we note that

$$\tilde{\mathbf{g}}_{\text{eq}}(z)\tilde{\mathbf{g}}_{\text{eq}}^H(1/z^*) = \tilde{\mathbf{F}}(z) = \tilde{\mathbf{G}}(z)\tilde{\mathbf{G}}^H(1/z^*) \ . \tag{7.290}$$

Therefore, the Euclidean distance between sequences of channel outputs for the system in Fig. 7.31 is the same as those for the $T/2$-spaced MLSE receiver in Fig. 7.29. Consequently, the system shown in Fig. 7.31 achieves ML performance. The main advantage of the system in Fig. 7.31 is that the noise-whitening filter does not depend on the unknown channel and has a fixed structure. Of course, the implementation of the noise whitening filter will still require substantial complexity.

The receiver shown in Fig. 7.31 has a scalar output, while the receiver in Fig. 7.29 has a vector output and, furthermore, $\tilde{\mathbf{g}}_{\text{eq},n}^{(j)}$ is a vector while $\tilde{\mathbf{G}}_n^{(j)}$ is a matrix. As a result, the branch metric used in the Viterbi algorithm needs to be modified accordingly. From (7.261)

$$\mu_m = -\sum_{j=1}^{J}\left\{\left|\tilde{v}_{2m}^{(j)} - \sum_{n=0}^{L_j}\mathbf{x}_{m-n}\tilde{\mathbf{g}}_{\text{eq},2n}^{(j)}\right|^2\right.$$

$$\left. + \left|\tilde{v}_{2m+1}^{(j)} - \sum_{n=0}^{L_j-1}\mathbf{x}_{m-n}\tilde{\mathbf{g}}_{\text{eq},2n+1}^{(j)}\right|^2\right\} \ . \tag{7.291}$$

Although the $T/2$-spaced receiver is optimum, there are several key issues that must be resolved before it can be implemented. First, the receiver must be trained to derive an initial estimate of the chance vectors $\{\tilde{\mathbf{g}}_{\text{eq},n}^{(j)}\}$. This synchronization and training problem is particularly challenging for an asynchronous TDMA cellular system where the training sequences are not coincident. With an asynchronous system different elements of the channel matrices are trained at different times. Second, the receiver must be able to track the channel vectors $\{\tilde{\mathbf{g}}_{\text{eq}n}^{(j)}\}$ during data demodulation. Perhaps a per-survivor processing approach such as the one suggested in Section 5.1.1 could be used.

8.7   INTERFERENCE REJECTION COMBINING MLSE

In many cases, the structure of the CCI is often unknown. This is true for example with licensed cellular systems that use different common air interfaces in the same band. For example, AMPS, IS-54/136, and CDPD users all share the same band. Here we derive a MIMO MLSE receiver, called the interference rejection combining MLSE (IRC-MLSE) receiver, for such conditions.

Once again, we assume that the receiver filter on each antenna element is matched to the transmitted pulse and followed by rate $2/T$ sampling. Since

the CCI has an unknown form, a matched filter is only required for the desired signal. The overall pulse response consisting of the transmit filter, channel, and receiver filter is $f^{(j)}(t) = c^{(1,j)}(t) * p(t),$ where $p(t) = h_a(t) * h_a^*(-t)$ is the overall response of the transmit and receive filters. The vector of matched filter outputs from the J antenna elements is

$$\mathbf{y}(t) = \sum_{\ell=0}^{L} x_\ell \mathbf{f}(t - \ell T) + \mathbf{z}(t) \tag{7.292}$$

where

$$
\begin{aligned}
\mathbf{y}(t) &= (y^{(1)}(t), \dots, y^{(J)}(t))^T \\
\mathbf{f}(t) &= (f^{(1)}(t), \dots, f^{(J)}(t))^T \\
\mathbf{z}(t) &= (z^{(1)}(t), \dots, z^{(J)}(t))^T
\end{aligned}
\tag{7.293}
$$

and where LT is the length of the pulse $f^{(j)}(t)$. The vector $\mathbf{z}(t)$ is the impairment at the output of the matched filter due to the K co-channel signals plus AWGN, and has the form

$$\mathbf{z}(t) = \sum_{k=1}^{K} \mathbf{i}_k(t) + \boldsymbol{\nu}(t) \tag{7.294}$$

where

$$
\begin{aligned}
\mathbf{i}_k(t) &= (i_k^{(1)}(t), \dots, i_k^{(J)}(t))^T \\
\boldsymbol{\nu}(t) &= (\eta^{(1)}(t), \dots, \eta^{(J)}(t))^T \ .
\end{aligned}
\tag{7.295}
$$

The matched filter outputs are sampled at rate $2/T$ and passed to a noise whitening filter. The noise whitening filter is sub-optimum in the presence of CCI, since the CCI at the input to the receiver filter can be viewed as colored noise. However, the noise whitening filter ensures maximum likelihood performance in the absence of CCI. The noise whitening filter is obtained by using the same procedure leading to the overall $T/2$-spaced discrete-time channel with the transfer function defined in (7.283). It follows that the overall channel consisting of the transmit filter, channel, and receiver filter, and $T/2$-spaced sampler can be modeled as a $T/2$-spaced tapped delay line with tap coefficients

$$\tilde{\mathbf{g}}^{(j)} = (\tilde{g}_0^{(j)}, \dots, \tilde{g}_{2L_h}^{(j)})$$

where $L_h T$ is the length of the pulse $h_a(t)$. Define

$$\tilde{\mathbf{g}}_\ell = (\tilde{g}_\ell^{(1)}, \dots \tilde{g}_\ell^{(J)})^T \ . \tag{7.296}$$

Then the vectors $\tilde{\mathbf{v}}_{2k}$ and $\tilde{\mathbf{v}}_{2k+1}$ at the output of the noise whitening filter corresponding to the kth received baud are

$$\tilde{\mathbf{v}}_{2k} = \sum_{\ell=0}^{L_h} x(k-\ell)\tilde{\mathbf{g}}_{2\ell} + \tilde{\mathbf{n}}_{2k} \tag{7.297}$$

$$\tilde{\mathbf{v}}_{2k+1} = \sum_{\ell=0}^{L_h-1} x(k-\ell)\tilde{\mathbf{g}}_{2\ell+1} + \tilde{\mathbf{n}}_{2k+1} \tag{7.298}$$

To derive a feasible receiver structure, we now assume that the sampled impairment vector $\tilde{\mathbf{n}}_k$ at the output of the noise whitening filter is a vector of J correlated complex Gaussian random variables having the joint pdf

$$p(\tilde{\mathbf{n}}_k) = \frac{1}{(2\pi)^J |\mathbf{R}_k|} \exp\left\{ -\frac{1}{2}\tilde{\mathbf{n}}_k^H \mathbf{R}_k^{-1} \tilde{\mathbf{n}}_k \right\} \tag{7.299}$$

where $|\mathbf{R}_k|$ is the determinant of $\mathbf{R}_k$ and

$$\mathbf{R}_k = \frac{1}{2}\mathrm{E}\left[\tilde{\mathbf{n}}_k \tilde{\mathbf{n}}_k^H \right] . \tag{7.300}$$

Assuming an MLSE-like algorithm, the branch metric should be related to the likelihood of the impairment vector. At epoch k, the samples $\tilde{\mathbf{v}}_{2k}$ and $\tilde{\mathbf{v}}_{2k+1}$ are used by the Viterbi algorithm to evaluate the branch metric

$$\begin{aligned}
\mu_k &= [\tilde{\mathbf{v}}_{2k} - \hat{\mathbf{u}}_{2k}^1]^H \mathbf{R}_{2k}^{-1}[\tilde{\mathbf{v}}_{2k} - \hat{\mathbf{u}}_{2k}^1] \\
&\quad + [\tilde{\mathbf{v}}_{2k+1} - \hat{\mathbf{u}}_{2k+1}^2]^H \mathbf{R}_{2k+1}^{-1}[\tilde{\mathbf{v}}_{2k+1} - \hat{\mathbf{u}}_{2k+1}^2]
\end{aligned} \tag{7.301}$$

where

$$\hat{\mathbf{u}}_{2k}^1 = \sum_{\ell=0}^{2L_h} \hat{x}(k-\ell)\tilde{\mathbf{g}}_{2\ell} \qquad \hat{\mathbf{u}}_{2k+1}^2 = \sum_{\ell=0}^{2L_h-1} \hat{x}(k-\ell)\tilde{\mathbf{g}}_{2\ell+1} .$$

Notice that the metric calculation requires the correlation matrix $\mathbf{R}_k$ and its inverse, and the subchannel impulse responses

$$\tilde{\mathbf{g}}^1 = (\tilde{\mathbf{g}}_0, \tilde{\mathbf{g}}_2, \ldots, \tilde{\mathbf{g}}_{2L_h}) \tag{7.302}$$

$$\tilde{\mathbf{g}}^2 = (\tilde{\mathbf{g}}_1, \tilde{\mathbf{g}}_3, \ldots, \tilde{\mathbf{g}}_{2L_h-1}) . \tag{7.303}$$

Computing the inverse of $\mathbf{R}_k$ can be computationally intensive for large J, the number of computations required being proportional to J^3. However, when $J = 2$ (two receiver antenna elements) the inverse can be obtained by using direct matrix inversion (DMI), i.e., the inverse of the matrix $\mathbf{R}_k$ is

$$\mathbf{R}_k^{-1} = \frac{\mathrm{adj}(\mathbf{R}_k)}{|\mathbf{R}_k|} = \frac{1}{r_{k_{11}}r_{k_{22}} - r_{k_{12}}r_{k_{21}}} \begin{bmatrix} r_{k_{22}} & -r_{k_{12}} \\ -r_{k_{21}} & r_{k_{11}} \end{bmatrix} . \tag{7.304}$$

Division by the determinant $|\mathbf{R}_k|$ is unnecessary provided that $\mathbf{R}_k$ remains constant over the decision delay in the Viterbi algorithm, since the determinant just scales all the path metrics. In this case, the Viterbi algorithm can use the simplified branch metric

$$
\begin{aligned}
\mu_k \;=\; & [\tilde{\mathbf{v}}_{2k} - \hat{\mathbf{u}}^1_{2k}]^H \mathrm{adj}(\mathbf{R}_{2k})[\tilde{\mathbf{v}}_{2k} - \hat{\mathbf{u}}^1_{2k}] \\
& + [\tilde{\mathbf{v}}_{2k+1} - \hat{\mathbf{u}}^2_{2k+1}]^H \mathrm{adj}\mathbf{R}_{2k+1})[\mathbf{v}_{2k+1} - \hat{\mathbf{u}}^2_{2k+1}] \;. \quad (7.305)
\end{aligned}
$$

which only requires multiplications and additions.

Finally, a metric combining MLSE (MC-MLSE) receiver is one that zeroes the off diagonal elements of the matrix $\mathbf{R}_k$. The metric combining receiver is equivalent to maximal ratio combining when the channel is affected by additive white Gaussian noise.

8.8   EXAMPLES

The performance of the J-MLSE, IRC-MLSE and MC-MLSE receivers discussed in the previous sections is now compared and contrasted. For this purpose, and EDGE (Enhanced Data for GSM Evolution) burst format is assumed. The EGDE burst format is the same as the GSM burst format described in Fig. 1.2. However, instead of the GMSK modulation used in GSM, EDGE uses 8-PSK modulation with square-root raised cosine pulse shaping with a roll-off factor of $\beta = 0.5$. For illustrative purposes, a T-spaced two equal ray model is assumed for the desired signal. The interference impairment consists of a single flat faded EDGE interferer. In all cases, the receiver front-end consists of a receiver filter that is matched to the transmitted pulse followed by a rate $2/T$ sampler and a noise whitening filter. The J-MLSE receiver has 512 states, as defined by 2 symbols for the desired signal and 1 symbol for the co-channel interferer. The MC/IRC-MLSE receivers have 64 states, as defined by 2 symbols for the desired signal. Each simulation run consists of 3000 frames of 142 8-PSK symbols.

Fig. 7.32 shows the E_b/N_o performance of the three receivers for a fixed $C/I = 30$ dB. The J-MLSE receiver is the optimum receiver in the maximum likelihood sense and achieves the best possible performance in AWGN. The MC-MLSE receiver is also optimum for AWGN channels, but exhibits some degradation at higher E_b/N_o due to the co-channel interference that is present. The IRC-MLSE receiver give the worst E_b/N_o performance.

Fig. 7.33 shows the C/I performance of the three receivers for $E_b/N_o = 30$ dB. Observe that the MC-MLSE receiver gives the worst performance, while the J-MLSE receiver and IRC-MLSE receivers offer huge C/I performance gains. The best performance is realized with the IRC-MLSE receiver. Hence, the IRC-MLSE receiver sacrifices a small amount of E_b/N_o performance for a large gain in C/I performance.
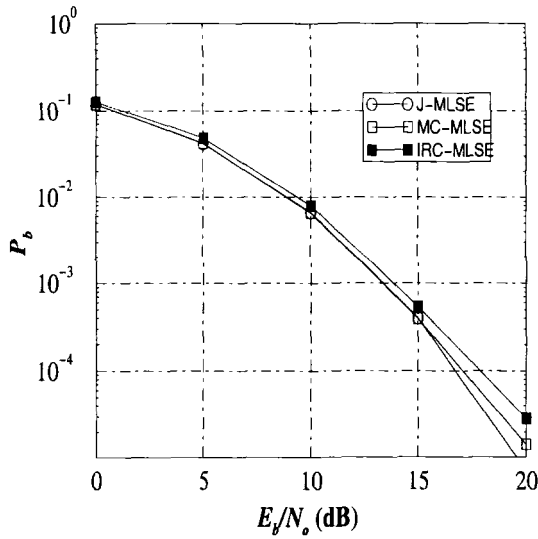
Figure 7.32. Relative E_b/N_o performance of the J-MLSE, MC-MLSE and IRC-MLSE receivers; $C/I = 30$ dB.

It is curious that the IRC-MLSE receiver outperforms the J-MLSE receiver. First the J-MLSE receiver that we have implemented, does not have a sufficient number of receiver states due to pulse truncation effects. Hence, there is some residual intersymbol interference that can be significant at low C/I. Second, the overall signal constellation produced by the combination of the desired signal and the co-channel signal may degenerate such that the constellation points overlap. In this case, errors can occur even for large E_b/N_o values.

APPENDIX 7.A: Derivation of Equation (7.184)

Assume that (7.176) has M different poles γ_1, γ_2, ..., γ_M. Then the pairwise error probability is equal to

$$P(\gamma) = \sum_{i=1}^{M} \left\{ \left(\frac{1}{2} - \frac{1}{2}\sqrt{\frac{\gamma_i}{1+\gamma_i}} \right) \prod_{j \neq i} \left(1 - \frac{\gamma_j}{\gamma_i} \right)^{-1} \right\}. \qquad (7\text{-}7.\text{A}.1)$$

Define the function $-\phi(\gamma) = \sum_{i=1}^{M} \gamma_i - C = 0$, where C is a constant. The method of Lagrange multipliers suggests that

$$\frac{\partial P}{\partial \gamma_i} + \lambda \frac{\partial \phi}{\partial \gamma_i} = 0 \quad i = 1, \ldots, M \qquad (7\text{-}7.\text{A}.2)$$
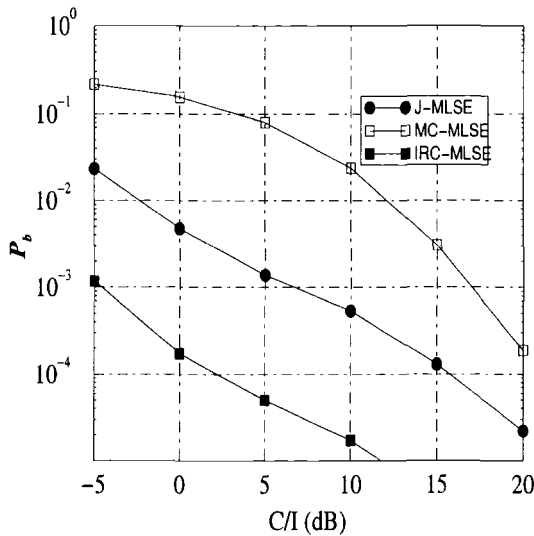
Figure 7.33. Relative C/I performance of the J-MLSE, MC-MLSE and IRC-MLSE receivers; $E_b/N_o = 30$ dB.

for any real number λ. It can be shown by induction that

$$
\begin{aligned}
\frac{\partial P}{\partial \gamma_k} &= -\left(\frac{1}{2} - \frac{1}{2}\sqrt{\frac{\gamma_k}{1+\gamma_k}}\right) \sum_{i \neq k} \left\{ \frac{\gamma_i}{\gamma_k^2} \left(1 - \frac{\gamma_i}{\gamma_k}\right)^{-2} \prod_{j \neq i,k} \left(1 - \frac{\gamma_j}{\gamma_k}\right)^{-1} \right\} \\
&\quad + \sum_{i \neq k} \left\{ \frac{1}{\gamma_i} \left(1 - \frac{\gamma_k}{\gamma_i}\right)^{-2} \left(\frac{1}{2} - \frac{1}{2}\sqrt{\frac{\gamma_i}{1+\gamma_i}}\right) \prod_{j \neq i,k} \left(1 - \frac{\gamma_j}{\gamma_i}\right)^{-1} \right\} \\
&\quad - \left(\frac{1}{4\gamma_k^{1/2}} \frac{1}{(1+\gamma_k)^{3/2}}\right) \prod_{j \neq k} \left(1 - \frac{\gamma_j}{\gamma_k}\right)^{-1}.
\end{aligned}
\tag{7-7.A.3}
$$

By solving (7-7.A.2) and observing the symmetry of $P(\boldsymbol{\gamma})$ and the derivative (7-7.A.3) with respect to the permutations of $\boldsymbol{\gamma}$, it is apparent that the minimum of $P(\boldsymbol{\gamma})$ is achieved when $\gamma_1 = \gamma_2 = \ldots = \gamma_M$.

Problems

7.1. Assume that a received signal is given by

$$
y(t) = \sum_{i=-\infty}^{\infty} x_i f(t - iT)
$$

where $x_k = \pm 1$, and $f(t)$ is a the minimum bandwidth pulse satisfying Nyquist's criterion for zero ISI, i.e.,

$$F(f) = \begin{cases} T & |f| \le 1/2T \\ 0 & |f| > 1/2T \end{cases}$$

and

$$f(t) = \frac{\sin(\pi t/T)}{\pi t/T} \ .$$

There are two problems associated with this pulse shape. One is the problem of realizing a pulse having the rectangular spectral characteristic $F(f)$ given above. The other problem arises from the fact that the tails in $f(t)$ decay as $1/t$. Consequently, a sampling timing error results in an infinite series of ISI components. Such a series is not absolutely summable and, hence, the sum of the resulting interference does not converge.

Assume that $f(t) = 0$ for $|t| > NT$, where N is a positive integer. In spite of the restriction that the channel is band-limited, this assumption holds in all practical communication systems.

a) Due to a slight timing error, the received signal is sampled at $t = kT + t_0$, where $t_0 < T$. Calculate the response for $t = kT + t_0$. Separate the response into two components, the desired term and the ISI term.

b) Assume that the polarities of x_i are such that every term in the ISI is positive, i.e., worst case ISI. Under this assumption show that the ISI term is

$$\text{ISI} = \frac{2}{\pi} \sin(\pi t_0/T) \sum_{n=1}^{N} \frac{n}{n^2 - t_0^2/T^2} \ .$$

and, therefore, $\text{ISI} \to \infty$ as $N \to \infty$.

7.2. Starting with

$$f_k = \int_{-\infty}^{\infty} h^*(\tau) h(\tau + kT) d\tau$$

show that

$$F(e^{j2\pi fT}) = F_\Sigma(f) \ .$$

7.3. Suppose that the impulse response of an overall channel consisting of the transmit filter, channel, and receive filter, is

$$F(f) = \begin{cases} 1 & , \ |f| \le f_\ell \\ \frac{f_u - |f|}{f_u - f_\ell} & , \ f_\ell \le |f| \le f_u \end{cases} \ .$$

a) Find the overall impulse response $f(t)$.

b) Is it possible to transmit data without ISI?

c) How do the magnitudes of the tails of the overall impulse response decay with large values of t?

d) Suppose that binary signaling is used with this pulse shape so that the noiseless signal at the output of the receive filter is

$$y(t) = \sum_n x_n f(t - nT)$$

where $x_n \in \{-1, +1\}$. What is the maximum possible magnitude that $y(t)$ can achieve?

7.4. Show that the ISI coefficients $\{f_n\}$ may be expressed in terms of the channel vector coefficients $\{g_n\}$ as

$$f_n = \sum_{k=0}^{L-n} g_k^* g_{k+n} \qquad n = 0, 1, 2, \ldots, L \ .$$

7.5. Suppose that BPSK is used on a static ISI channel. The complex envelope has the form

$$\tilde{s}(t) = A \sum_{k=-\infty}^{\infty} x_k h_a(t - kT)$$

where $x_k \in \{-1, +1\}$ and $h_a(t)$ is the amplitude shaping pulse. The non-return-to-zero pulse $h_a(t) = u_T(t)$ is used and the impulse response of the channel is

$$g(t) = g_0 \delta(t) - g_1 \delta(t - \tau)$$

where g_0 and g_1 are complex numbers and $0 < \tau < T$.

a) Find the received pulse $h(t)$.

b) What is the filter matched to $h(t)$?

c) What are the ISI coefficients $\{f_i\}$?

7.6. Suppose that BPSK signaling is used on a static ISI channel having impulse response

$$g(t) = \delta(t) + 0.1\delta(t - T)$$

The receiver employs a filter that is matched to the transmitted pulse $h_a(t)$, and the sampled outputs of the matched filter are

$$y_n = x_n q_0 + \sum_{k \neq n} x_k q_{n-k} + \eta_n$$

where $x_n \in \{-1, +1\}$. Decisions are made on the $\{y_n\}$ without any equalization.

a) What is the variance of noise term η_n?

b) What are the values of the $\{q_n\}$?

c) What is the probability of error in terms of the average received bit-energy-to-noise ratio?

7.7. A typical receiver for digital signaling on an ISI channel consists of a matched filter followed by an equalizer. The matched filter is designed to minimize the effect of random noise, while the equalizer is designed to minimize the effect of intersymbol interference. By using mathematical arguments, show that i) the matched filter tends to accentuate the effect of ISI, and ii) the equalizer tends to accentuate the effect of random noise.

7.8. Consider an ISI channel, where $f_n = 0$ for $|n| > 1$. Suppose that the receiver uses a filter matched to the received pulse $h(t) = h_a(t) * g(t)$, and the T-spaced samples at the output of the matched filter, $\{y_k\}$, are filtered as shown in Fig. 7.A.1. The values of g_0 and g_1 are chosen to satisfy

$$|g_0|^2 + |g_1|^2 = f_0$$
$$g_0 g_1^* = f_1$$

Find an expression for the filter output v_k in terms of g_0, g_1, x_k, x_{k-1}, and the noise component at the output of the digital filter, η_k.
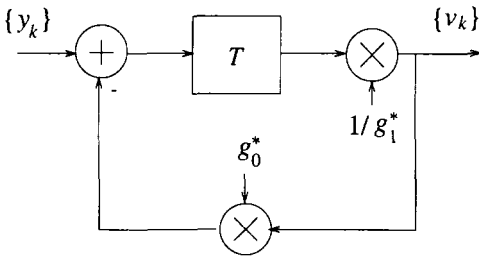
Figure 7.A.1.   Digital filter for Problem 7.8.

7.9. The z-transform of the channel vector g of a communication system is equal to

$$G(z) = 0.1 + 1.0z^{-1} - 0.1z^{-2} .$$

A binary sequence **x** is transmitted, where $x_k = \in \{-1, +1\}$. The received samples at the output of the noise whitening filter are

$$v_n = \sum_{k=0}^{2} g_k x_{n-k} + \eta_n$$

where $\{\eta_n\}$ is a white Gaussian noise sequence with variance $\sigma_\eta^2 = N_o$.

a) Evaluate the probability of error if the demodulator ignores ISI.

b) Design a 3-tap zero-forcing equalizer for this system.

c) What is the response $\{v_k\}$ for the input sequence

$$\{x_k\} = (-1)^k, \; k = 0, 1, 2, 3 \; ?$$

What is the response at the output of the equalizer?

d) Evaluate the probability of error for the equalized channel.

7.10. Suppose that a system is characterized by the received pulse

$$h(t) = \sqrt{2a}e^{-at}, \qquad 0 \le t \le \infty \; .$$

A receiver implements a filter matched to $h(t)$ and generates T-spaced samples at the output of the filter. Note that the matched filter is actually noncausal.

a) Find the ISI co-efficients f_i.

b) What is the transfer function of the noise whitening filter that yields a system having an overall minimum phase response?

c) Find the transfer function of the equivalent zero-forcing equalizer $C'(z)$.

d) Find the noise power at the output of the zero-forcing equalizer, and find the condition when the noise power becomes infinite.

7.11. Consider M-PAM on a static ISI channel, where the receiver employs a filter that is matched to the received pulse. The sampled outputs of the matched filter are

$$y_n = x_n f_0 + \sum_{k \ne n} x_k f_{n-k} + \nu_n$$

where the source symbols are from the set $\{\pm 1, \pm 3, \ldots, \pm(M-1)\}$. Decisions are made on the $\{y_n\}$ without any equalization by using a threshold detector. The ℓth ISI pattern can be written as

$$D(\ell) = \sum_{k \ne n} x_{\ell,k} f_{n-k}$$

and $D(\ell)$ is maximum when $\text{sgn}(x_{\ell,k}) = \text{sgn}(f_{n-k})$ and each of the $x_{\ell,k}$ takes on the maximum signaling level, i.e., $x_{\ell,k} = (M-1)d$ for M even. The *maximum distortion* is defined as

$$D_{\max} = \frac{1}{f_0} \sum_{n \ne 0} |f_n| \; .$$

a) Discuss and compare error performance M-ary signaling $(M > 2)$ with binary signaling $(M = 2)$, using D_{max} as a parameter.

b) Suppose that the channel has ISI coefficients

$$
\begin{aligned}
f_i &= 0.0, \quad |i| \geq 3 \\
f_2 &= f_{-2} = 0.1 \\
f_1 &= f_{-1} = -0.2 \\
f_0 &= 1.0 .
\end{aligned}
$$

Plot the probability of error against the signal-to-noise ratio and compare with the ideal channel case, i.e., $f_0 = \delta_{n0}$. Show your results for $M = 2$ and 4.

7.12. Consider a linear MSE equalizer and suppose that the tap gain vector c satisfies

$$
\mathbf{c} = \mathbf{c}_{op} + \mathbf{c}_e
$$

where $\mathbf{c}_e$ is the tap gain error vector. Show that the mean square error that is achieved with the tap gain vector c is

$$
J = J_{min} + \mathbf{c}_e^T \mathbf{M}_v \mathbf{c}_e^* .
$$

7.13. The matrix $\mathbf{M}_v$ has an eigenvalue λ_k and eigenvector $\mathbf{x}_k$ if

$$
\mathbf{x}_k \mathbf{M}_v = \lambda_k \mathbf{x}_k \qquad\qquad k = 1, \ldots, N .
$$

Prove that the eigenvectors are orthogonal, i.e.,

$$
\mathbf{x}_i \mathbf{x}_j^T = \delta_{ij}
$$

7.14. Show that the relationship between the output SNR and J_{min} for an infinite-tap mean-square error linear equalizer is

$$
\gamma_\infty = \frac{1 - J_{min}}{J_{min}}
$$

where the subscript ∞ on γ indicates that the equalizer has an infinite number of taps. Note that this relationship between γ_∞ and J_{min} holds when there is residual intersymbol interference in addition to the noise.

7.15. In this question, we will show in steps that

$$
\nabla_{\mathbf{c}} J = 2\mathbf{c}^T \mathbf{M}_v - 2\mathbf{v}_x^H .
$$

Define

$$\mathbf{M}_v = \mathbf{M}_{v_R} + j\mathbf{M}_{v_I}$$
$$\mathbf{c} = \mathbf{c}_R + j\mathbf{c}_I$$
$$\mathbf{v}_x = \mathbf{v}_{x_R} + j\mathbf{v}_{x_I}$$

a) By using the Hermitian property $\mathbf{M}_v = \mathbf{M}_v^H$ show that

$$\mathbf{M}_{v_R} = \mathbf{M}_{v_R}^T \text{ and } \mathbf{M}_{v_I} = -\mathbf{M}_{v_I}^T .$$

b) Show that

$$\nabla_{\mathbf{c}_R} \mathrm{Re}\{\mathbf{v}_x^H \mathbf{c}^*\} = \mathbf{v}_{x_R}^T$$
$$\nabla_{\mathbf{c}_I} \mathrm{Re}\{\mathbf{v}_x^H \mathbf{c}^*\} = -\mathbf{v}_{x_I}^T$$
$$\nabla_{\mathbf{c}_R} \mathbf{c}^T \mathbf{M}_v \mathbf{c}^* = 2\mathbf{c}_R^T \mathbf{M}_{v_R} - 2\mathbf{c}_I^T \mathbf{M}_{v_I}$$
$$\nabla_{\mathbf{c}_I} \mathbf{c}^T \mathbf{M}_v \mathbf{c}^* = 2\mathbf{c}_I^T \mathbf{M}_{v_R} + 2\mathbf{c}_R^T \mathbf{M}_{v_I}$$

where $\nabla_{\mathbf{x}}$ is the gradient with respect to vector x.

c) If we define the gradient of a real-valued function with respect to a complex vector c as

$$\nabla_{\mathbf{c}} = \nabla_{\mathbf{c}_R} + j\nabla_{\mathbf{c}_I}$$

show that

$$\nabla_{\mathbf{c}} \mathrm{Re}\{\mathbf{v}_x^H \mathbf{c}^*\} = \mathbf{v}_x^H$$
$$\nabla_{\mathbf{c}} \mathbf{c}^H \mathbf{M}_v \mathbf{c}^* = 2\mathbf{c}^T \mathbf{M}_v .$$

7.16. Show that the pairwise error probability for digital signaling on an ISI channel is given by (7.154).

7.17. Consider the transmission of the binary sequence x, $x_n \in \{-1, +1\}$ over the equivalent discrete-time white noise channel model shown in Fig. 7.A.2. The received sequence is

$$v_0 = .70x_0 + \eta_1$$
$$v_1 = .70x_1 - .60x_0 + \eta_2$$
$$v_2 = .70x_2 - .60x_1 + \eta_3$$
$$\vdots$$
$$v_k = .70x_k - .60x_{k-1} + \eta_k$$

a) Draw the state diagram for this system.
b) Draw the trellis diagram.
c) Suppose that the received sequence is

$$\{v_i\}_{i=0}^{6} = \{1.0, \ -1.5, \ 0.0, \ 1.5, \ 0.0, \ -1.5, \ 1.0\}$$

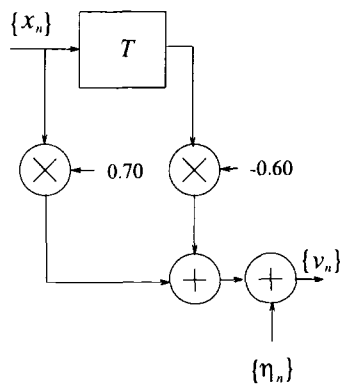Show the surviving paths and their associated path metrics after v_6 has been received.

Figure 7.A.2.   Discrete-time white noise channel model for Problem 6.17.

7.18. Suppose that BPSK signaling is used on a frequency selective fading channel. The discrete-time system consisting of the transmit filter, channel, receiver filter, and baud-rate sampler can be described by the polynomial

$$F(z) = \frac{5}{16} - \frac{1}{8}z^{-1} - \frac{1}{8}z \ .$$

The samples at the output of the receiver filter are processed by a noise whitening filter such that the overall discrete-time white noise channel model $G(z)$ has minimum phase.

a) Find $G(z)$.
b) Draw the state diagram and the trellis diagram for the discrete-time white noise channel model.
c) A block of 10 symbols $\mathbf{x} = \{x_i\}_{i=0}^{9}$ is transmitted over the channel and it is known that $x_9 = -1$. Assume that $x_i = 0, i < 0$ and the suppose that the sampled sequence at the output of the matched filter is

$$\mathbf{y} = (y_0, y_1, y_2, y_3, \ldots y_9)$$
$$= (1/2, 1/4, -3/4, 3/4, -3/4, -1/4, 3/4, -3/4, -1/4, -1/4)$$

What sequence $\mathbf{x}$ was most likely transmitted?

Chapter 8

ERROR CONTROL CODING

Channel coding and interleaving techniques have long been recognized as an effective technique for combating the deleterious effects of noise, interference, jamming, fading, and other channel impairments. The basic idea of channel coding is to introduce controlled redundancy into the transmitted signals that is exploited at the receiver to correct channel induced errors by means of forward error correction. Channel coding can also be used for error detection in schemes that employ automatic repeat request (ARQ) strategies. ARQ strategies must have a feedback channel to relay the retransmission requests from the receiver back to the transmitter when errors are detected. ARQ schemes require buffering at the transmitter and/or receiver and, therefore, are suitable for data applications but are not suitable for delay sensitive voice applications. Hybrid ARQ schemes use both error correction and error detection; the code is used to correct the most likely error patterns, and to detect the more infrequently occurring error patterns. Upon detection of errors a retransmission is requested.

There are many different types of error correcting codes, but historically they have been classified into block codes and convolutional codes. To generate a codeword of an (n, k) block code, a block of k data bits is appended by $n - k$ redundant parity bits that are algebraically related to the k data bits, thereby producing a codeword consisting of n code bits. The ratio $R_c = k/n$ is called the code rate, where $0 < R_c \leq 1$. Convolutional codes, on the other hand, are generated by the discrete-time convolution of the input data sequence with the impulse response of the encoder. The memory of the encoder is measured by the duration of the impulse response. While block encoder operates on k-bit blocks of data bits, a convolutional encoder accepts a continuous sequence of input data bits.

Both block codes and convolutional codes find potential applications mobile radio systems. Some second generation digital cellular standards (e.g., GSM, IS-54) use convolutional codes, while others (e.g., PDC) use block codes. Although hard decision block decoders are easy to implement, there exist some very simple soft decision decoding algorithms (e.g., the Viterbi algorithm) for convolutional codes. As a result convolutional codes are often preferred over block codes.

In the early application of coding to digital communications, the modulator and coder were treated a separate entities. Hence, a block code or a convolutional code was employed to obtain a coding gain at the cost of bandwidth expansion or data rate. Although this may be a feasible approach for power limited channels where bandwidth resources are plentiful, it is undesirable and sometimes not even possible for bandwidth limited applications such as cellular radio. If no sacrifices of data rate or bandwidth can be made, then schemes that separate the operations of coding and modulation require a very powerful code just to break even with an uncoded system. In 1974, Massey [219] suggested that the performance of a coded digital communication system could be improved by treating coding and modulation as a single entity. Ungerboeck, later developed the basic principles of trellis-coded modulation (TCM) [330] and identified classes of trellis codes that provide substantial coding gains on bandwidth limited additive white Gaussian noise (AWGN) channels.

TCM schemes combine the operations of coding and modulation and can be viewed as a generalization of convolutional codes. While convolutional codes attempt to maximize the minimum Hamming distance between allowed code symbol sequences, trellis-codes attempt to maximize the Euclidean distance between allowed code symbol sequences. By jointly designing the encoder and modulator Ungerboeck showed that, for an AWGN channel, coding gains of 3-6 dB could be obtained relative to an uncoded system by using trellis codes with 4-128 encoder states, without sacrificing bandwidth or data rate. This property makes TCM very attractive for cellular radio applications where high spectral efficiency is needed due to limited bandwidth resources and good power efficiency is needed to extend battery life in portable radios. TCM experienced an almost immediate and widespread application into high-speed power-efficient and bandwidth-efficient digital modems. In 1984, a variant of the Ungerboeck 8-state 2-D trellis code was adopted by CCITT for both 14.4 kb/s leased-line modems and the 9.6 kb/s switched-network modems [38]. In 1985, a TCM-based modem operating at 19.2 kb/s was introduced by Codex [331].

Ungerboeck's work [330] captured the attention of the coding community and laid the foundation for intensified research. Calderbank and Mazo introduced an analytic description of trellis codes [43]. They showed how to realize the two operations (coding and mapping) in Ungerboeck's codes by

using a single-step procedure. Calderbank and Sloane [44], and Wei [352], proposed multi-dimensional trellis codes. Spaces with larger dimensionality are attractive, because the signals are spaced at larger Euclidean distances [38]. Calderbank and Sloan [44], and Forney [128], made the observation that the signal constellation should be regarded as a finite set of points taken from an infinite lattice, and the partitioning of the constellation into subsets corresponds to the partitioning of the lattice into a sub-lattice and its cosets. They then developed a new class of codes, called coset codes, based on this principle.

Many studies have examined the performance of TCM on interleaved flat fading channels [82, 83, 94, 45]. Divsalar and Simon [83, 84] constructed trellis codes that are effective for interleaved flat Ricean and Rayleigh fading channels. Interleaving randomizes the channel with respect to the transmitted symbol sequence and has the effect of reducing the channel memory. Consequently, interleaving improves the performance of codes that have been designed for memoryless channels. Moreover, trellis codes that are designed for flat fading channels exhibit time diversity when combined with interleaving of sufficient depth. It was reported in [45] that interleaving with reasonably long interleaving depths is almost as good as ideal infinite interleaving. The design of trellis codes for interleaved flat fading channels is not guided by the minimum Euclidean distance used for AWGN channels, but rather by the minimum product squared Euclidean distance and the minimum built-in time diversity between any two allowed code symbol sequences. Wei [353] introduced an additional design parameter called the minimum decoding depth, and proposed a set of efficient codes for interleaved flat Rayleigh fading channels.

Many studies have also considered the effect of intersymbol interference (ISI) on the performance of trellis codes that have been designed for AWGN channels [319, 363, 88, 107]. The coded performance on *static* ISI channels may be significantly degraded compared to that on ISI-free channels. Receivers for trellis-coded modulation on static ISI channels typically use a linear forward equalizer followed by a soft decision Viterbi decoder. For channels with severe ISI, a more appropriate approach is to use a decision feedback equalizer (DFE) in front of the TCM decoder to avoid the problems of noise enhancement. However, the feedback section of the DFE requires that decisions be available with zero delay. Since the zero-delay decisions are unreliable, the performance improvement by using the DFE is marginal [49]. It is possible that the performance can be improved if equalization and decoding is performed in a joint manner by using maximum likelihood sequence estimation (MLSE) or some other form of sequence estimator. However, the complexity of an MLSE receiver grows exponentially with the number of encoder states and the length of the channel vector.

In 1993, Berrou *et al.,* introduced *parallel* concatenated convolutional codes (PCCCs), called Turbo coding [36]. When used in conjunction with an iter-

ative decoding scheme, PCCCs achieve near Shannon limit performance on both the AWGN channel and the interleaved flat fading channel. Simulations of a rate-1/2 Turbo code have shown a bit error probability of 10^{-5} at an $E_b/N_o = 0.5$ dB, which is only 0.5 dB from the Shannon limit! Although, the performance of Turbo codes is remarkable at low E_b/N_o, their performance at high E_b/N_o is unimpressive. There is a perceivable change in the slope of the bit error rate (BER) curves, which has been loosely termed an "error floor." In 1997, Benedetto *et* al., showed that iterative decoding of serially concatenated interleaved convolutional codes (SCCCs) can provide large coding gains without the problem of an error floor [288]. In general, SCCCs outperform PCCCs at high E_b/N_o, whereas the opposite is true for low E_b/N_o.

The remainder of the chapter is organized as follows. Section 1. gives an introduction to block codes. Sections 2. and 3. introduce convolutional codes and trellis codes. This is followed by a consideration of the design and performance analysis of trellis codes for various types of channels that are found in mobile radio applications. These include the AWGN channels in Section 4., interleaved flat fading channels in Section 5., and non-interleaved fading ISI channels in Section 6.. The evaluation of error probability upper bounds is important for performance prediction and Section 6.4 presents a technique for union bounding the error probability of TCM on a fading ISI channel; flat fading channels and static ISI channels can be treated as special cases. Finally, section 7. provides an introductory treatment of Turbo coding.

1.   BLOCK CODES

1.1   BINARY BLOCK CODES

A binary block encoder accepts a length-k input vector $\mathbf{a} = (a_1, a_2, \dots, a_k)$, where $a_i \in \{0, 1\}$, and generates a length-n codeword $\mathbf{c} = (c_1, c_2, \dots, c_n)$, where $c_i \in \{0, 1\}$, through the linear mapping $\mathbf{c} = \mathbf{aG}$, where $\mathbf{G} = [g_{ij}]_{k \times n}$ is a $k \times n$ matrix called the **generator matrix.** The matrix $\mathbf{G}$ has full row rank k, and the code $\mathcal{C}$ is generated by taking all linear combinations of the rows of the matrix $\mathbf{G}$, where field operations are performed by using modulo-2 arithmetic. The code rate is $R_c = k/n$ and there are 2^k codewords. The whole task of designing a block code is to find the generator matrices that yield codes that are both powerful and easy to decode.

For any block code with generator matrix $\mathbf{G}$, there exists an $(n-k) \times n$ **parity check matrix** $\mathbf{H} = [h_{ij}]_{(n-k) \times n}$ such that $\mathbf{GH}^T = \mathbf{0}_{k \times (n-k)}$. The matrix $\mathbf{H}$ has full row rank $n - k$ and is orthogonal to all codewords, i.e., $\mathbf{cH}^T = \mathbf{0}_{n-k}$. The matrix H is the generator matrix of a **dual code** $\mathcal{C}^T$, consisting of 2^{n-k} codewords. The parity check matrix of $\mathcal{C}^T$ is the matrix $\mathbf{G}$.

A systematic block code is one having a parity check matrix of the form $\mathbf{G} = [\mathbf{I}_{k \times k} | \mathbf{P}]$ where $\mathbf{P}$ is a $k \times (n-k)$ matrix. For a systematic block

code, the first k coordinates of each codeword are equal to the k-bit input vector a, while the last $n - k$ coordinates are the parity check bits. By using elementary row operations, the generator matrix of any linear block code can be put into systematic form. A systematic block code has the parity check matrix $\mathbf{H} = [\mathbf{I}_{(n-k)\times(n-k)}| - \mathbf{P}^T]$. For a systematic block code, $\mathbf{G}\mathbf{H}^T = [\mathbf{P}|\mathbf{I}_{k\times k}][-\mathbf{P}^T|\mathbf{I}_{(n-k)\times(n-k)}]^T = -\mathbf{P} + \mathbf{P} = \mathbf{0}_{k\times(n-k)}$.

Example 8.1———————————————————The parity check matrix of an (n, k) Hamming code consists of all non-zero binary $(n - k)$-tuples. For example, the systematic (7,4) Hamming code has the parity check matrix

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} . \tag{8.1}$$

which consists of all non-zero binary 3-tuples. The generator matrix of the (7,4) systematic Hamming code is

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix} . \tag{8.2}$$

The 16 codewords of the (7,4) Hamming code are generated by taking all linear combinations of the rows of **G** using modulo-2 arithmetic.

1.1.1   MINIMUM DISTANCE

Let $d(\mathbf{c}_1, \mathbf{c}_2)$ denote the **Hamming distance** between the codewords $\mathbf{c}_1$ and $\mathbf{c}_2$, equal to the number of coordinates in which they differ. For linear block codes, $d(\mathbf{c}_1, \mathbf{c}_2) = w(\mathbf{c}_1 + \mathbf{c}_2)$, where $w(\mathbf{c}_1 + \mathbf{c}_2)$ is the weight of $\mathbf{c}_1 + \mathbf{c}_2$, equal to the number of non-zero coordinates of $w\mathbf{c}_1 + \mathbf{c}_2$. The **free Hamming distance,** d_{free}, of a linear block code is the minimum number of coordinates in which any two codewords differ. For a linear code, the sum of any two codewords $\mathbf{c}_1 + \mathbf{c}_2$ is another codeword. Hence, the free Hamming distance is

$$d_{\text{free}} = \min_{\mathbf{c}_1, \mathbf{c}_2} d(\mathbf{c}_1, \mathbf{c}_2) \tag{8.3}$$

$$= \min_{\mathbf{c}\neq\mathbf{0}} d(\mathbf{c}, \mathbf{0}) \tag{8.4}$$

$$= \min_{\mathbf{c}\neq\mathbf{0}} w(\mathbf{c}) . \tag{8.5}$$

Therefore, d_{free} is equal to the weight of the minimum weight non-zero codeword.

To derive an upper bound on d_{free}, recall that any linear block code can be put into systematic form, $\mathbf{G} = [\mathbf{P}|\mathbf{I}_{k\times k}]$ where **P** is a $k \times (n - k)$ matrix. It

is certainly the case the number of non-zero elements in any row of $\mathbf{P}$ cannot exceed $n - k$. Hence, the number of non-zero elements in any row of $\mathbf{G}$ cannot exceed $n - k + 1$. Since all rows of $\mathbf{G}$ are valid codewords,

$$d_{\text{free}} \leq n - k + 1 \tag{8.6}$$

a result known as the **Singleton bound..** A code that has $d_{\text{free}} = n - k + 1$ is called a **maximum distance separable** (MDS) code.

An example of a simple block code that meets the Singleton bound is the binary repetition code

$$0 \longrightarrow \mathbf{c}_0 = (0, 0, \ldots, 0)_n$$
$$1 \longrightarrow \mathbf{c}_1 = (1, 1, \ldots, 1)_n$$

In this case, $d_{\text{free}} = d(\mathbf{c}_0, \mathbf{c}_1) = n - k + 1$. The repetition code is the only MDS binary code. The non-binary Reed-Solomon codes are also MDS codes.

1.1.2 SYNDROMES

Suppose that the codeword $\mathbf{c}$ is transmitted and the vector $\mathbf{y} = \mathbf{c} + \mathbf{e}$ is received, where $\mathbf{e}$ is defined as the **error vector.** The **syndrome** of the received vector $\mathbf{y}$ is defined as

$$\mathbf{s} = \mathbf{y}\mathbf{H}^T . \tag{8.7}$$

If $\mathbf{s} = \mathbf{0}$, then y is a codeword; conversely if $\mathbf{s} \neq \mathbf{0}$, then an error must have occurred. Note that if y is a codeword, then $\mathbf{s} = \mathbf{0}$. Hence, $\mathbf{s} = \mathbf{0}$ does not mean that no errors have occurred. They are just undetectable. Since for a linear code the sum of any two codewords is another codeword, it follows that the number of undetectable error patterns is equal to $2^k - 1$, the number of non-zero codewords. The syndrome only depends upon the error vector because

$$\mathbf{s} = \mathbf{y}\mathbf{H}^T = \mathbf{c}\mathbf{H}^T + \mathbf{e}\mathbf{H}^T = 0 + \mathbf{e}\mathbf{H}^T = \mathbf{e}\mathbf{H}^T . \tag{8.8}$$

In general, $\mathbf{s} = \mathbf{e}\mathbf{H}^T$ is a system of $n - k$ equations in n variables. Hence, for any given syndrome s, there are 2^k solutions for e. However, the most likely error pattern $\mathbf{e}$ is the one that has minimum Hamming weight.

1.1.3 ERROR DETECTION

A linear block code can detect all error patterns of $d_{\text{free}} - 1$ or fewer error. If $\mathbf{e} \neq \mathbf{0}$ is a codeword, then no errors are detected. There are $2^k - 1$ undetectable error patterns, but there are $2^n - 1$ possible non-zero error patterns. Hence, the number of detectable error patterns is

$$2^n - 1 - (2^k - 1) = 2^n - 2^k$$

Usually, $2^k - 1$ is a small fraction of $2^n - 2^k$. For the (7,4) Hamming code considered in Example 8.1, there are $2^4 - 1 = 15$ undetectable error patterns and $2^7 - 2^4 = 112$ detectable error patterns.

1.1.4   WEIGHT DISTRIBUTION

Consider a block code C and let A_i be the number of codewords of weight i. The set $\{A_0, A_1, \ldots, A_n\}$ is called the **weight distribution** of C. The weight distribution can be expressed as a **weight enumerator polynomial**

$$A(z) = A_0 z^0 + A_1 z^1 + \cdots + A_n z^n \ . \tag{8.9}$$

For the (7,4) Hamming code in Example 8.1,

$$A_0 = 1, A_2 = 0, A_3 = 7, A_4 = 7, A_5 = 0, A_6 = 0, A_7 = 1 \ .$$

Hence,

$$A(z) = 1 + 7z^3 + 7z^4 + z^7 \ .$$

1.1.5   PROBABILITY OF UNDETECTED ERROR

The probability of undetected error is

$$\begin{aligned} P_e(U) \ &= \ P(\mathbf{e} \text{ is a nonzero codeword}) \\ &= \ \sum_{i=1}^{n} A_i P(w(\mathbf{e}) = i) \end{aligned} \tag{8.10}$$

The error probability $P(w(\mathbf{e}) = i)$ depends on the coding channel, defined as that portion of the communication system that is seen by the coding system. The simplest coding channel is the **binary symmetric channel** (BSC), where

$$P(y_i \neq c_i) = p = 1 - P(y_i = c_i) \ . \tag{8.11}$$

For a BSC, $P(w(\mathbf{e}) = i) = p^i (1-p)^{n-i}$ and, hence,

$$P_e(U) = \sum_{i=1}^{n} A_i p^i (1-p)^{n-i} \ . \tag{8.12}$$

The (7,4) Hamming code in Example 8.1 has an undetected error probability of

$$P_e(U) = 7p^3 (1-p)^4 + 7p^4 (1-p)^3 + p^7 \ . \tag{8.13}$$

For a raw channel error rate of $p = 10^{-2}$, we have $P_e(U) = 7 \times 10^{-6}$. Hence, the undetected error rate can be very small even for a fairly simple block code.

1.1.6   ERROR CORRECTION

A linear block code can correct all error patterns of t or fewer errors, where

$$t \leq \lfloor \frac{d_{\text{free}} - 1}{2} \rfloor \tag{8.14}$$

and $\lfloor x \rfloor$ is the largest integer contained in x. A code is usually capable of correcting many error patterns of $t + 1$ or more errors. In fact, up to 2^{n-k} error patterns may be corrected, which is equal to the number of syndromes.

For a BSC, the probability of codeword error is

$$
\begin{aligned}
\mathrm{P}(E) \quad &\leq \quad 1 - \mathrm{P}(t \text{ or fewer errors}) \\
&= \quad 1 - \sum_{i=0}^{t} \binom{n}{i} p^i (1-p)^{n-i} \quad . \tag{8.15}
\end{aligned}
$$

1.1.7   STANDARD ARRAY DECODING

One conceptually simple method for decoding any linear block codes is **standard array decoding.** The standard array of an (n, k) linear block code is constructed as follows:

1. Write out all 2^k codewords in a row starting with $\mathbf{c}_0 = \mathbf{0}$.

2. From the remaining $2^n - 2^k$ n-tuples, select an error pattern $\mathbf{e}_2$ of weight 1 and place it under $\mathbf{c}_0$. Under each codeword put $\mathbf{c}_i + \mathbf{e}_2, i = 1, \ldots, 2^k - 1$.

3. Select a minimum weight error pattern $\mathbf{e}_3$ from the remaining unused n-tuples and place it under $\mathbf{c}_0 = \mathbf{0}$. Under each codeword put $\mathbf{c}_i + \mathbf{e}_3, i = 1, \ldots, 2^k - 1$.

4. Repeat Step 3 until all n-tuples have been used.

Note that every n-tuple appears once and only once in the standard array.

Example 8.2——————————————————————

Consider the $(4,2)$ code with generator matrix

$$\mathbf{G} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} .$$

The standard array is

$$
\begin{bmatrix}
\mathbf{e}_1 & 0000 & 1100 & 0101 & 1001 \\
\mathbf{e}_2 & 0001 & 1101 & 0100 & 1000 \\
\mathbf{e}_3 & 0010 & 1110 & 0111 & 1011 \\
\mathbf{e}_4 & 0011 & 1111 & 0110 & 1010
\end{bmatrix}
$$

The standard array consists of 2^{n-k} disjoint rows of 2^k elements. These rows are called **cosets** and the ith row has the elements

$$F_i = \{\mathbf{e}_i, \mathbf{e}_i + \mathbf{c}_1, \ldots, \mathbf{e}_i + \mathbf{c}_{2^k-1}\} \ .$$

The first element, $\mathbf{e}_i$, is called the **coset leader.** The standard array also consists of 2^k disjoint columns. The *jth* column has the elements

$$D_j = \{\mathbf{c}_j, \mathbf{c}_j + \mathbf{e}_2, \ldots, \mathbf{c}_j + \mathbf{e}_{2^{n-k}}\} \ .$$

To correct errors, the following procedure is used. When $\mathbf{y}$ is received, find $\mathbf{y}$ in the standard array. If $\mathbf{y}$ is in row i and column j, then the coset leader from row i, $\mathbf{e}_i$, is the most likely error pattern to have occurred and $\mathbf{y}$ is decoded into $\mathbf{y} + \mathbf{e}_i = \mathbf{c}_j$. A code is capable of correcting all error patterns that are coset leaders. If the error pattern is not a coset leader then erroneous decoding will result.

1.1.8   SYNDROME DECODING

Syndrome decoding relies on the fact that all 2^k n-**tuples** in the same coset of the standard array have the same syndrome. This is because the syndrome only depends on the coset leader as shown in (8.8). To perform syndrome decoding

1. Compute the syndrome $\mathbf{s} = \mathbf{y}\mathbf{H}^T$.

2. Locate the coset leader $\mathbf{e}_\ell$ where $\mathbf{e}_\ell \mathbf{H}^T = \mathbf{s}$.

3. Decode $\mathbf{y}$ into $\mathbf{y} + \mathbf{e}_\ell = \hat{\mathbf{c}}$.

This technique can be used for any linear block code. The calculation in Step 2 can be done by using a simple look-up table. However, for large $n - k$ it becomes impractical because 2^{n-k} syndromes and 2^{n-k} error patterns must be stored.

2.   CONVOLUTIONAL CODES
2.1   ENCODER DESCRIPTION

The encoder for a rate-$1/n$ binary convolutional code can be viewed as a **finite-state machine** (FSM) that consists of an v-stage binary shift register with connections to n modulo-2 adders, and a multiplexer that converts the adder outputs to serial codewords. The **constraint length** of a convolutional code is defined as the number of shifts through the FSM over which a single input data bit can affect the encoder output. For an encoder having a v-stage shift register, the constraint length is equal to $K = v + 1$. A very simple rate-1/2, constraint length-3, binary convolutional encoder is shown in Fig. 8.1.
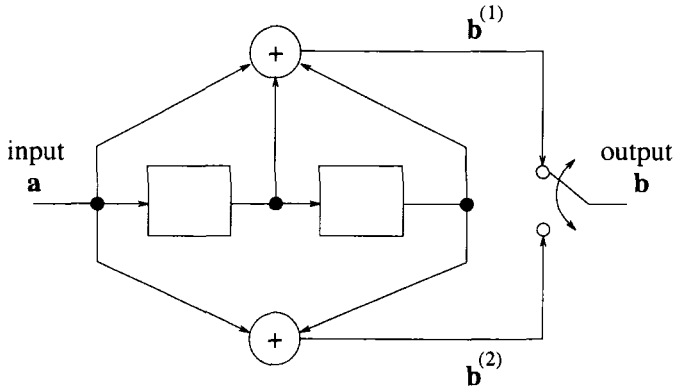
Figure 8.1. Binary convolutional encoder; $R_c = 1/2$, $K = 3$.
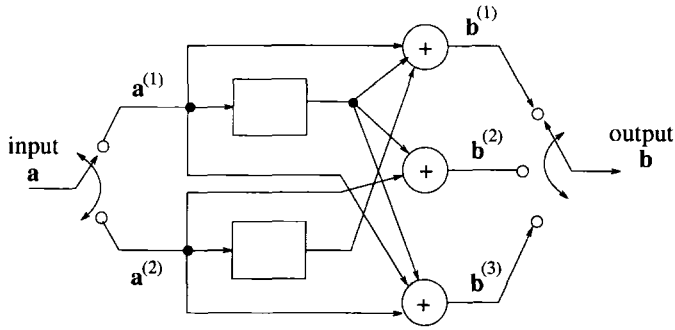
Figure 8.2. Binary convolutional encoder; $R_c = 2/3$, $K = 2$.

The above concept can be generalized to *rate-k/n* binary convolutional code by using k shift registers, n modulo-2 adders, along with input and output multiplexers. For a rate-k/n code, the k-bit information vector $\mathbf{a}_\ell = (a_\ell^{(1)}, \ldots, a_\ell^{(k)})$ is input to the encoder at epoch ℓ to generate the n-bit code vector $\mathbf{b}_\ell = (b_\ell^{(1)}, \ldots, b_\ell^{(n)})$. If K_i denotes the constraint length of the ith shift register, then the overall constraint length is defined as $K = \max_i K_i$. Fig. 8.2 shows a simple rate-2/3, constraint length-2 convolutional encoder.

A convolutional encoder can be described by the set of impulse responses, $\{\mathbf{g}_i^{(j)}\}$, where $\mathbf{g}_i^{(j)}$ is the *jth* output sequence $\mathbf{b}^{(j)}$ that results from the *ith* input sequence $\mathbf{a}^{(i)} = (1, 0, 0, 0, \ldots)$. The impulse responses can have a duration of at most K and have the form $\mathbf{g}_i^{(j)} = (g_{i,0}^{(j)}, g_{i,1}^{(j)}, \ldots, g_{i,K-1}^{(j)})$. Sometimes the $\{\mathbf{g}_i^{(j)}\}$ are called **generator sequences.** For the encoder in Fig. 8.1

$$\mathbf{g}^{(1)} = (1, 1, 1) \qquad \mathbf{g}^{(2)} = (1, 0, 1) \tag{8.16}$$

and for the encoder in Fig. 8.2

$$
\begin{aligned}
\mathbf{g}_1^{(1)} &= (1,\,1), \ \mathbf{g}_1^{(2)} = (0,\,1), \ \mathbf{g}_1^{(3)} = (1,\,1) \\
\mathbf{g}_2^{(1)} &= (0,\,1), \ \mathbf{g}_2^{(2)} = (1,\,0), \ \mathbf{g}_2^{(3)} = (1,\,0) \ .
\end{aligned}
\tag{8.17}
$$

It follows that the *j*th output, $\mathbf{b}_i^{(j)}$, corresponding to the *i*th input sequence $\mathbf{a}^{(i)}$ is the discrete convolution $\mathbf{b}_i^{(j)} = \mathbf{a}^{(i)} \circledast \mathbf{g}_i^{(j)}$, where $\circledast$ denotes modulo-2 convolution. The time domain convolutions can be conveniently replaced by polynomial multiplications in a *D*-transform domain according to

$$
\mathbf{b}_i^{(j)}(D) = \mathbf{a}^{(i)}(D)\mathbf{g}_i^{(j)}(D)
\tag{8.18}
$$

where

$$
\mathbf{a}^{(i)}(D) = \sum_{k=0}^{\infty} a_{i,k}D^k
\tag{8.19}
$$

is the *i*th input data polynomial,

$$
\mathbf{b}_i^{(j)}(D) = \sum_{k=0}^{\infty} b_{i,k}^{(j)}D^k
\tag{8.20}
$$

is the *j*th output polynomial corresponding to the *i*th input, and

$$
\mathbf{g}_i^{(j)}(D) = \sum_{k=0}^{K-1} g_{i,k}^{(j)}D^k
\tag{8.21}
$$

is the associated **generator polynomial.** It follows that the *j*th output sequence is

$$
\mathbf{b}^{(j)}(D) = \sum_{i=1}^{k} \mathbf{b}_i^{(j)}(D) = \sum_{i=1}^{k} \mathbf{a}^{(i)}(D)\mathbf{g}_i^{(j)}(D) \ .
\tag{8.22}
$$

The above expression leads to the matrix form

$$
\begin{aligned}
&\left[\mathbf{b}^{(i)}(D), \ldots, \mathbf{b}^{(n)}(D) \right] \\
&= \left[\mathbf{a}^{(1)}(D), \ldots, \mathbf{a}^{(k)}(D) \right]
\begin{bmatrix}
\mathbf{g}_1^{(1)}(D), & \cdots, & \mathbf{g}_1^{(n)}(D) \\
& \vdots & \\
\mathbf{g}_k^{(1)}(D), & \cdots, & \mathbf{g}_k^{(n)}(D)
\end{bmatrix}
\end{aligned}
\tag{8.23}
$$

where

$$
\mathbf{G}(D) =
\begin{bmatrix}
\mathbf{g}_1^{(1)}(D), & \cdots, & \mathbf{g}_1^{(n)}(D) \\
& \vdots & \\
\mathbf{g}_k^{(1)}(D), & \cdots, & \mathbf{g}_k^{(n)}(D)
\end{bmatrix}
\tag{8.24}
$$

is the generator matrix of the code. For the encoder in Fig. 8.1

$$\mathbf{G}(D) = \begin{bmatrix} 1 + D + D^2 & 1 + D^2 \end{bmatrix} \qquad (8.25)$$

while for the encoder in Fig. 8.2

$$\mathbf{G}(D) = \begin{bmatrix} 1 + D & D & 1 + D \\ D & 1 & 1 \end{bmatrix} . \qquad (8.26)$$

After multiplexing the outputs, the final codeword has the polynomial representation

$$\mathbf{b}(D) = \sum_{j=1}^{n} D^{j-1} \mathbf{b}^{(j)}(D^n) . \qquad (8.27)$$

Systematic convolutional codes are those where first k encoder output sequences, $\mathbf{b}^{(1)}, \ldots, \mathbf{b}^{(k)}$ are equal to the k encoder input sequences $\mathbf{a}^{(1)}, \ldots, \mathbf{a}^{(k)}$.

2.2 STATE AND TRELLIS DIAGRAMS, AND WEIGHT DISTRIBUTION

Since the convolutional encoder is a FSM, its operation can be described by a state-diagram and trellis diagram in a manner very similar to the treatment of ISI channels in Chapter 7. The state of the encoder is defined by the shift register contents. For a rate-k/n code, the ith shift register contains ν_i previous information bits. The state of the encoder at epoch ℓ is defined as

$$\sigma_\ell = \left(a_{\ell-1}^{(1)}, \cdots, a_{\ell-\nu_1}^{(1)} ; \cdots ; a_{\ell-1}^{(k)}, \cdots, a_{\ell-\nu_m}^{(k)} \right) . \qquad (8.28)$$

There are a total of $N_S = 2^{\nu_T}$ encoder states, where $\nu_T \triangleq \sum_{i=1}^{k} \nu_i$ is defined as the **total encoder memory.** For a rate-$1/n$ code, the encoder state at epoch ℓ is simply $\sigma_\ell = (a_{\ell-1}, \cdots, a_{\ell-\nu})$.

Figs. 8.3 and 8.4 show the state diagrams for codes in Figs. 8.1 and 8.2, respectively. The states are labeled using the convention $\sigma^{(i)}, i = 0, \ldots, \nu_T - 1$, where $\sigma^{(i)}$ represents the encoder state $(c_0, \ldots, c_{\nu_T-1})$ corresponding to the integer $i = \sum_{j=0}^{\nu_T-1} c_j 2^j$. In general, for a *rate-k/n* code there are 2^k branches entering and leaving each state. The branches in the state diagram are labeled with the convention $\mathbf{a}/\mathbf{b} = (a^{(1)}, a^{(2)}, \ldots, a^{(k)})/(b^{(1)}, b^{(2)}, \ldots, b^{(n)})$. For example, the state transition $\sigma^{(1)} \to \sigma^{(3)}$ in Fig. 8.3 has the label 1/01. This means that the input $\mathbf{a} = 1$ to the encoder in Fig. 8.1 with state $\sigma^{(1)} = (01)$ gives the output $\mathbf{b} = (01)$ and transitions the encoder to state $\sigma^{(3)} = (11)$.

Convolutional codes are linear codes, meaning that the sum of any two codewords is another codeword and the all-zeroes sequence is a codeword. It follows that the weight distribution and other distance properties of a convolutional code can be obtained from the state diagram. Consider, for example,
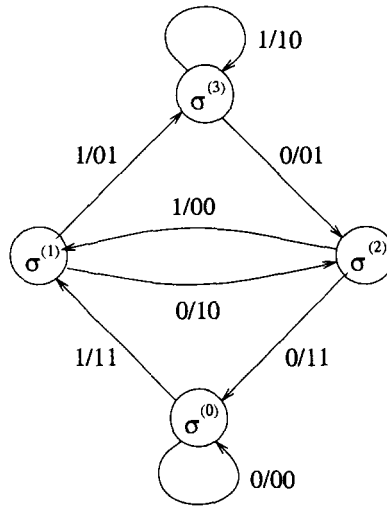
Figure 8.3.   State diagram for the binary convolutional encoder in Fig. 8.1.
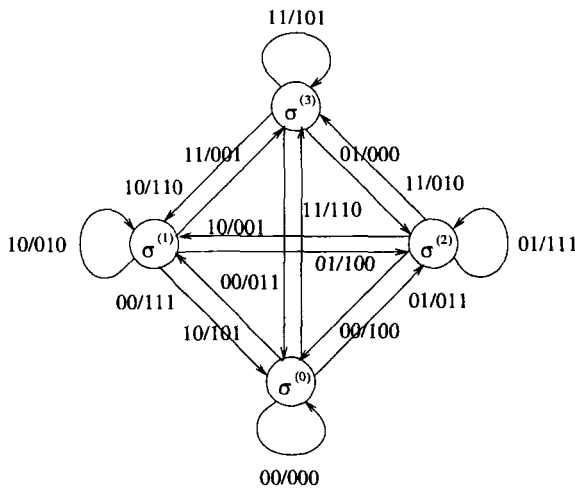

Figure 8.4.   State diagram for the binary convolutional encoder in Fig. 8.2.

the encoder in Fig. 8.1 along with its state diagram in Fig. 8.3. Since the self-loop at the zero state $\sigma^{(0)}$ corresponds to the all-zeroes codeword, we can split the zero state $\sigma^{(0)}$ into two nodes, representing the input and output of the state diagram. This leads to the **modified state diagram** shown in Fig. 8.5. The branches in the modified state diagram have labels of the form $D^i N^j L$, where i is the number of 1's in the encoder output sequence corresponding to a particular state transition, and j is the number of input 1's into the encoder for
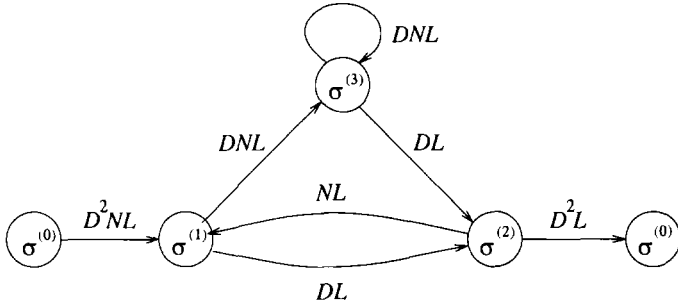
Figure 8.5.   Modified state diagram for the binary convolutional encoder in Fig. 8.1.

that transition. Every branch is labeled with the letter L, and the exponent of L is unity because each branch has length one.

The distance properties of a convolutional code can be obtained by computing the transfer function $T(D, N, L)$ of the modified state diagram. Any appropriate technique can be used to obtain the transfer function, such as Mason's formula [218]. For the example shown in Fig. 8.5, the transfer function is

$$
\begin{aligned}
T(D, N, L) &= \frac{D^5 N L^3}{1 - DNL(L+1)} \\
&= D^5 L^3 N + D^6 N^2 L^4 (L+1) + D^7 N^3 L^5 (L+1)^2 \\
&\quad + \cdots + D^{k+5} N^{k+1} L^{k+3} (L+1)^k + \cdots \qquad (8.29)
\end{aligned}
$$

The term $D^{k+5} N^{k+1} L^{k+3} (L+1)^k$ appearing in the transfer function means there are 2^k paths at Hamming distance $k + 5$ from the all-zeroes path, caused by $k + 1$ input ones. Of these 2^k paths, $\binom{k}{n}$ have length $k + n + 3$.

Sometimes the transfer function can be simplified if we are only interested in extracting certain distance properties of the convolutional code. For example, the weight distribution of the code can be obtained by setting $N = 1$ and $L = 1$ in the transfer function. For the particular transfer function in (8.29) this leads to

$$
\begin{aligned}
T(D) &= \frac{D^5}{1 - 2D} \\
&= D^5 + 2D^6 + 4D^7 + \cdots + 2^k D^{5+k} + \cdots \qquad (8.30)
\end{aligned}
$$

meaning that there are 2^k codewords at Hamming distance $5 + k$ from the all-zeroes codeword. Notice that no non-zero codeword exists with a Hamming distance less than 5 from the all-zeroes codeword. For the code in Fig. 8.1, we see that $d_{\text{free}} = 5$ from (8.30). The free Hamming distance can also be seen by inspecting the trellis diagram in Fig. 8.6. The branches in the trellis diagram

epoch
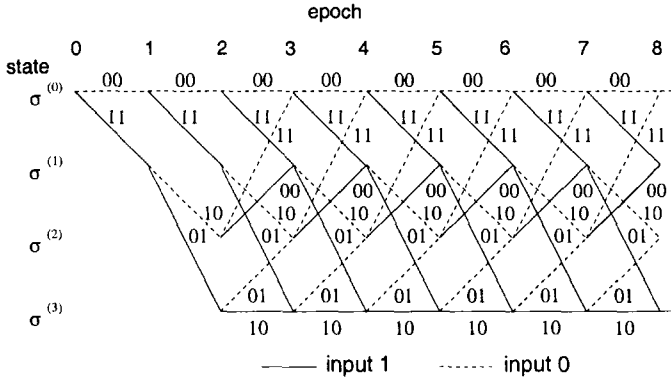
Figure 8.6.   Trellis diagram for the binary convolutional encoder in Fig. 8.1.

are labeled with the encoder output bits that correspond to the various state transitions.

Convolutional codes are designed to have the largest possible d_{free} for a given code rate and total encoder memory. Tabulation of convolutional codes that are optimal in this sense can be found in many references, e.g., Proakis [270], Lin and Costello [199], and Clark and Cain [63].

2.3   RECURSIVE SYSTEMATIC CONVOLUTIONAL (RSC) CODES

Forney [126] and Costello [70] showed that it is possible to construct a recursive systematic convolutional (RSC) encoder from every rate $R_c = 1/n$ feed-forward non-systematic convolutional encoder, such that the weight distributions of the codes are identical. Consider a rate-$1/n$ code with generator polynomials $\mathbf{g}_1(D), \ldots, \mathbf{g}_n(D)$. The output sequences are described by the polynomials

$$\mathbf{b}^{(j)}(D) = \mathbf{a}(D)\mathbf{g}^{(j)}(D), \qquad j = 1, \ldots, n \ . \tag{8.31}$$

To obtain a systematic code, we need to have $\mathbf{b}^{(1)}(D) = \mathbf{a}(D)$. To do this, suppose that both sides of (8.31) are divided by $\mathbf{g}^{(1)}(D)$, so that

$$\tilde{\mathbf{b}}^{(1)}(D) \;=\; \frac{\mathbf{b}^{(1)}(D)}{\mathbf{g}^{(1)}(D)} = \mathbf{a}(D) \tag{8.32}$$

$$\tilde{\mathbf{b}}^{(j)}(D) \;=\; \frac{\mathbf{b}^{(j)}(D)}{\mathbf{g}^{(1)}(D)} = \mathbf{a}(D)\frac{\mathbf{g}^{(j)}(D)}{\mathbf{g}^{(1)}(D)}, \qquad j = 2, \ldots, n \ . \tag{8.33}$$

Sometimes the $\mathbf{g}^{(j)}(D)$ are called the feed-forward polynomials, while $\mathbf{g}^{(1)}(D)$ is called the feed-back polynomial. Define a new input sequence $\tilde{\mathbf{a}}(D)$ as

$$\tilde{\mathbf{a}}(D) \triangleq \frac{\mathbf{a}(D)}{\mathbf{g}^{(1)}(D)} \qquad (8.34)$$

so that

$$\tilde{\mathbf{b}}^{(1)}(D) = \tilde{\mathbf{a}}(D)\mathbf{g}^{(1)}(D) \qquad (8.35)$$

$$\tilde{\mathbf{b}}^{(j)}(D) = \tilde{\mathbf{a}}(D)\mathbf{g}^{(j)}(D), \qquad j = 2, \ldots, n . \qquad (8.36)$$

Observe that the transformation between $\mathbf{a}(D)$ and $\tilde{\mathbf{a}}(D)$ in (8.34) is that of a recursive digital filter with modulo-2 operations. This transformation simply reorders the input sequences $\mathbf{a}(D)$. Since the input sequences consist of all possible binary sequences, the filtered sequences $\tilde{\mathbf{a}}(D)$ also consist of all possible binary sequences. Hence, the set of coded sequences $\tilde{\mathbf{b}}(D)$ is the same as the set of coded sequences $\mathbf{b}(D)$ and thus the non-systematic and systematic codes have the same weight distribution functions. However, the input weight distributions for the two codes are completely different as we will see.

Example 8.3

Consider, for example, the rate-1/2 encoder in Fig. 8.1 with generators

$$\mathbf{g}^{(1)}(D) = 1 + D + D^2 \qquad (8.37)$$

$$\mathbf{g}^{(2)}(D) = 1 + D^2 \qquad (8.38)$$

By following the above described procedure a RSC code is obtained with generators

$$\hat{\mathbf{g}}^{(1)}(D) = 1$$

$$\hat{\mathbf{g}}^{(2)}(D) = \frac{\mathbf{g}^{(2)}(D)}{\mathbf{g}^{(1)}(D)} = \frac{1 + D^2}{1 + D + D^2}$$

The RSC is shown in Fig. 8.7

Similar to their feed-forward counterparts, the weight distribution and other distance properties of RSC codes can be obtained by constructing their corresponding modified state diagram and computing the transfer function $T(D, N, L)$. The RSC encoder in Fig. 8.7 has transfer function

$$T(D, N, L) = \frac{D^5 N^3 L^3 - D^6 N^4 L^4 + D^6 N^2 L^4}{1 - DNL - DNL^2 - D^2 L^3 + D^2 N^2 L^3} \qquad (8.39)$$

$$= D^5 N^3 L^3 + D^6 N^2 L^4 + D^6 N^4 L^5 + \cdots \qquad (8.40)$$

By setting $N = 1$ and $L = 1$, we obtain the weight distribution of the code, $T(D)$, which is identical to the weight distribution of the corresponding feed-forward non-systematic encoder in (8.30). However, by comparing the first
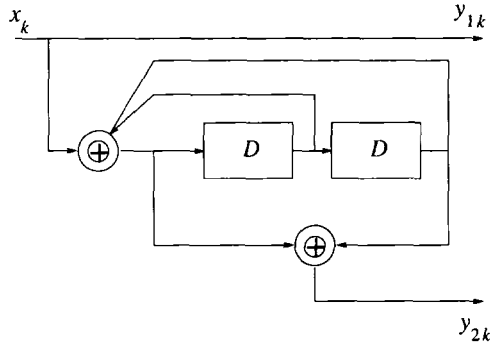
Figure 8.7. Recursive systematic convolutional (RSC) encoder derived from the feed-forward non-systematic encoder in Fig. 8.1.

few terms in their respective transfer functions in (8.29) and (8.40) it can be observed that the input weight distributions are completely different. In particular, codewords can be generated by weight-1 input sequences for the feed-forward non-systematic encoder, while the RSC requires input sequences having at least weight-2 to generate codewords. In fact, any finite weight codeword for the RSC code in Fig 8.7 is generated by an input polynomial $\mathbf{a}(D)$ that is divisible by $1 + D + D^2$. We will see later that these properties are crucial for Turbo codes.

Finally, both the feed-forward non-systematic and RSC codes are time invariant. This means that if the input sequence $a(D)$ produces codeword $b(D)$, then the input sequence $D^i a(D)$ produces the codeword $. D^i b(D)$. Note that the codewords $b(D)$ and $D^i b(D)$ have the same weight.

3. TRELLIS CODED MODULATION
3.1 ENCODER DESCRIPTION

Conventional convolutional codes realize a coding gain at the expense of data rate or bandwidth. Although, such coding schemes are attractive for power-limited applications, they are not suitable for bandwidth-limited applications. Ungerboeck showed that a coding gain can be achieved without sacrificing data rate or bandwidth by using a rate-$m/(m + r)$ convolutional encoder, and mapping the coded bits onto signal points $\{x_k\}$ through a technique called **mapping by set partitioning** [330]. This combination of coding and modulation, called trellis coded modulation (TCM), has three basic features;

1. An expanded signal constellation is used that is larger than the one necessary for uncoded modulation at the same data rate. The additional signal points allow redundancy to be inserted without sacrificing data rate or bandwidth.
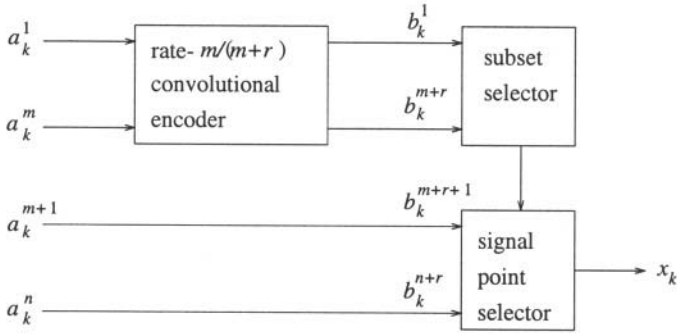
Figure 8.8. Ungerboeck trellis encoder.

2. The expanded signal constellation is partitioned such that the intra-subset minimum squared Euclidean distance is maximized at each step in the partition chain.

3. Convolutional encoding and signal mapping is used so that only certain sequences of signal points are allowed.

Fig. 8.8 shows the basic encoder structure for Ungerboeck's trellis codes. The n-bit information vector $\mathbf{a}_k = (a_k^{(1)}, \cdots, a_k^{(n)})$ is transmitted at epoch k. At each epoch, $m \leq n$ information bits are encoded into $m + r$ code bits by using a rate-$m/(m+r)$ linear convolutional encoder. The $m+r$ code bits select one of 2^{m+r} subsets of a 2^{n+r}-point signal constellation. The uncoded $n - m$ information bits select one of the 2^{n-m} signal points within the selected subset. This principle is best explained by example, and Fig. 8.9 shows a 4-state 8-PSK Ungerboeck trellis code. The equivalent uncoded system is 4-PSK which has a bit rate of 2 bits/symbol. The 4-state 8-PSK code uses a rate-1/2 convolutional code along with one uncoded bit to select signal points in an expanded 8-PSK signal constellation. Note that the overall rate is still 2 bits/symbol. Fig. 8.10 shows an 8-state 8-PSK Ungerboeck trellis code. The equivalent uncoded system is again 4-PSK with 2 bits/symbol. The 8-state 8-PSK code uses a rate-2/3 convolutional code to select one of the points in an expanded 8-PSK signal constellation so that the overall rate is again 2 bits/symbol.

3.2   MAPPING BY SET PARTITIONING

The critical step in the design of Ungerboeck's codes is the method of mapping the outputs of the convolutional encoder to points in the expanded signal constellation. Fig. 8.11 shows how the 8-PSK signal constellation is partitioned into subsets such that the intra-subset minimum squared Euclidean distance is maximized for each step in the partition chain. In the 8-PSK signal constellation there are 8 signal points equally spaced around a circle
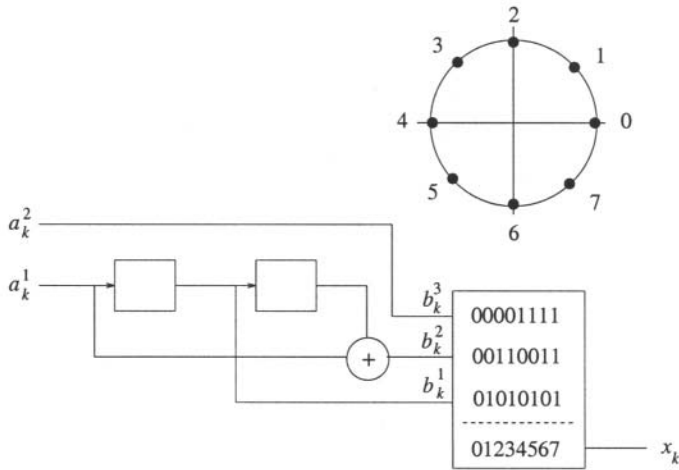
Figure 8.9.   Encoder and signal mapping for the 4-state 8-PSK Ungerboeck trellis code.
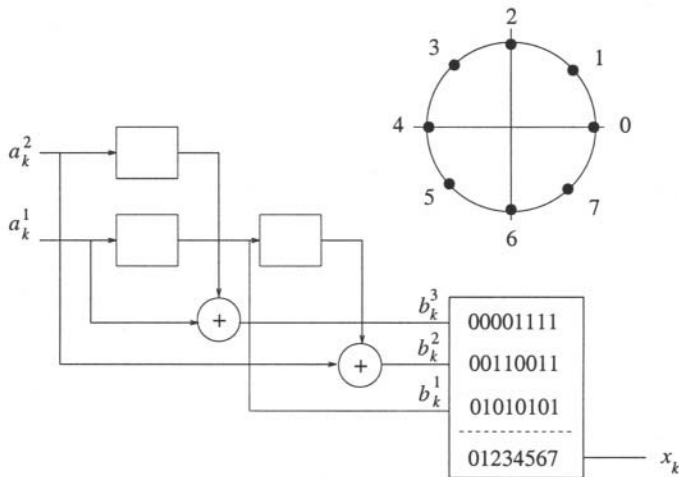
Figure 8.10.   Encoder and signal mapping for the 8-state 8-PSK Ungerboeck trellis code.

of unit radius. Notice that the minimum Euclidean distance between signal points in the 8-PSK signal constellation is $\Delta_0 = 0.765$, while the minimum Euclidean distances between signal points in the first and second level partitions are $\Delta_1 = \sqrt{2}$ and $\Delta_2 = 2$, respectively. The minimum Euclidean distance increases at each level of partitioning.

The advantages of using TCM can most easily be seen by considering the trellis diagram. For both the 4-state and 8-state 8-PSK trellis codes the equivalent uncoded system is 4-PSK. The trellis diagram for uncoded 4-PSK is shown in Fig. 8.12. The trellis only has one state and there are 4 **parallel**
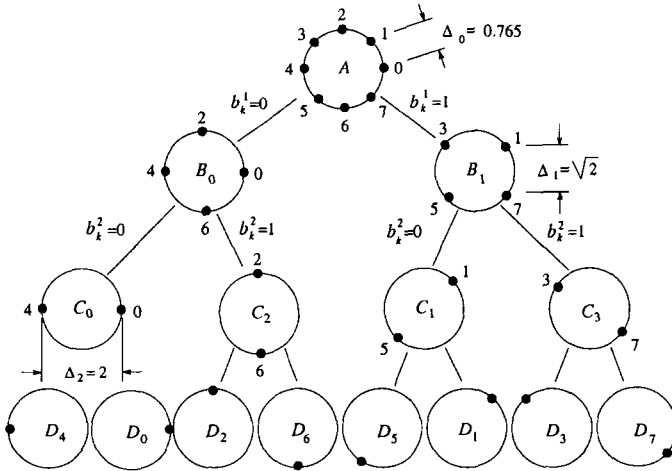
Figure 8.11. Set partitioning for an 8-PSK signal constellation.

transitions between the states. The subsets D_0, D_2, D_4, and D_6 are used as the signal points. The label D_0, D_2, D_4, D_6 means that the branches in the trellis diagram are labeled from top to bottom with signal points taken from the sets D_0, D_2, D_4, D_6. The minimum Euclidean distance between any two paths through the trellis is $d_{\min} = \sqrt{2}$.

The trellis diagram for the 4-state 8-PSK code is shown in Fig. 8.13. Each branch in the 4-state trellis is labeled with one of the four subsets C_0, C_1, C_2, and C_3. Again, the label $C_i C_j$ associated with a state means that the branches in the trellis diagram originating from that state are labeled from top to bottom with the subsets C_i and C_j. As shown in Fig. 8.11, each subset C_i contains two signal points. Thus, each branch in the trellis diagram actually contains two parallel transitions. For example branches with the label C_0 have two parallel transitions that are labeled with the signal points 0 and 4. For the 4-state 8-PSK code, it is possible that two coded sequences could differ by just a single parallel transition and, hence, their minimum Euclidean distance is $d = 2$. Also, any two signal paths that diverge from a state and remerge with the same state after more that one transition have a minimum Euclidean distance of $d = \sqrt{\Delta_1^2 + \Delta_0^2 + \Delta_1^2} = 2.141$. For example, the closest non-parallel code sequence to the all-zeroes sequence $\mathbf{x} = (0, 0, 0)$ is the sequence $\mathbf{x} = (2, 1, 2)$ at distance $d = 2.141$. Hence, the minimum Euclidean distance of the code over all parallel and non-parallel pairs of sequences for the 4-state 8-PSK code is $d_{\min} = 2$.

At high signal-to-noise ratio (SNR), the bit error rate performance on an AWGN channel is dominated by the minimum Euclidean distance error events. The pairwise error probability between two coded sequences x and $\hat{\mathbf{x}}$ separated

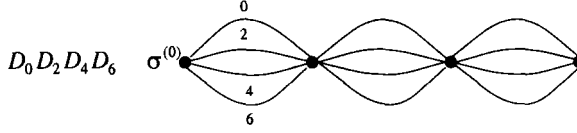$D_0 D_2 D_4 D_6$ $\sigma^{(0)}$

Figure 8.12.   Trellis diagram for uncoded 4-PSK.

$C_0 C_2$ $\sigma^{(0)}$

$C_1 C_3$ $\sigma^{(1)}$

$C_2 C_0$ $\sigma^{(2)}$
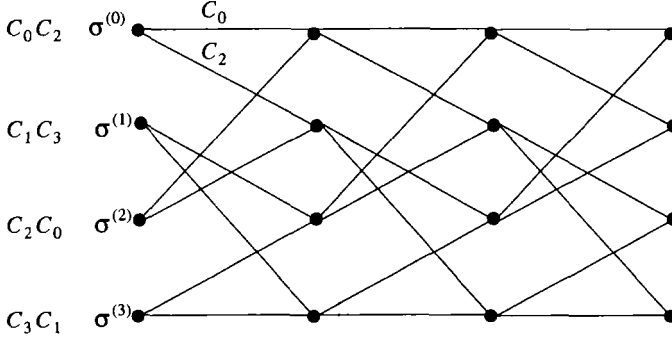
$C_3 C_1$ $\sigma^{(3)}$

Figure 8.13.   Trellis diagram for 4-state 8-PSK Ungerboeck trellis code.

by Euclidean distance $d_{\min}$ is

$$P(\mathbf{x} \rightarrow \hat{\mathbf{x}}) = Q\left(\sqrt{\frac{d_{\min}^2}{4N_o}}\right) \tag{8.41}$$

The **asymptotic coding gain** is defined by [38]

$$G_a = 10 \log_{10} \frac{(d_{\min,\text{coded}}^2 / E_{\text{av,coded}})}{(d_{\min,\text{uncoded}}^2 / E_{\text{av,uncoded}})} \text{ dB} \tag{8.42}$$

where E_{av} is the average energy per symbol in the signal constellation. For the 4-state 8-PSK code, the asymptotic coding gain is $G_a = 3$ dB.

The concept of mapping by set partitioning was developed by Ungerboeck as a method for maximizing the minimum Euclidean distance of a code and optimize the performance on an AWGN channel. Ungerboeck's construction of the optimum 4-state 8-PSK code was based on the following heuristic rules [331];

1. Parallel transitions (when they occur) are assigned signal points having the maximum Euclidean distance between them.

2. The transition starting or ending in any state is assigned the subsets (C_0, C_2) or (C_1, C_3) which have a maximum distance between them.

3. All signal points are used in the trellis diagram with equal frequency.
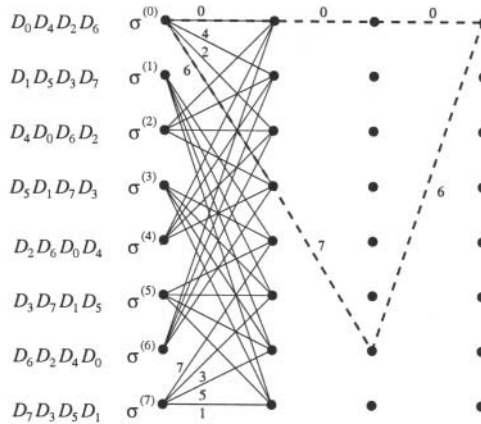
Figure 8.14. Trellis diagram for 8-state 8-PSK Ungerboeck trellis code. The dashed lines show two minimum distance paths.

It is clear that the performance of the 4-state 8-PSK code is limited by the parallel transitions. Larger asymptotic coding gains can be obtained by introducing more code states so that the parallel transitions are eliminated. For example, the above design rules can be applied to the 8-state 8-PSK code to obtain the code trellis shown in Fig. 8.14. In this case, the minimum Euclidean distance is $d_{\min} = \sqrt{\Delta_1^2 + \Delta_0^2 + \Delta_1^2} = 2.141$. This yields an asymptotic coding gain of 3.6 dB over uncoded 4-PSK.

4.   CODED PERFORMANCE ON AWGN CHANNELS

Viterbi originally exploited the trellis structure of convolutional codes and developed the Viterbi algorithm for ML decoding of convolutional codes [341]. Given the similarity between the trellis structures of ISI channels, convolutional codes, and trellis codes (e.g., compare Figs. 7.15, 8.6 and 8.13), it is not surprising that the union bounding techniques that were developed to evaluate the error probability of digital signaling on ISI channels with an MLSE receiver in Chapter 7.6 can also be applied, with some modification, to evaluate the error probability of convolutional and trellis codes with an MLSE receiver.

To develop the union bound, let $\mathbf{a} = \{a_k\}$ denote the transmitted information sequence. For any other sequence $\hat{\mathbf{a}} \neq \mathbf{a}$, define the corresponding error sequence as $\mathbf{e} = \{e_k\} = \mathbf{a} \oplus \hat{\mathbf{a}}$, where $\oplus$ denotes modulo-2 addition. Since the bit error probability at epoch j_1 is of interest, $e_{j_1} \neq 0$ for all error sequences. An error event occurs between k_1 and k_2 of length $k_2 - k_1$, if $\sigma_{k_1} = \hat{\sigma}_{k_1}$ and $\sigma_{k_2} = \hat{\sigma}_{k_2}$, but $\sigma_j \neq \hat{\sigma}_j$ for $k_1 < j < k_2$, where $k_1 \leq j_1 < k_2$, and $\sigma = \{\sigma_k\}$ and $\hat{\sigma} = \{\hat{\sigma}_k\}$ are the system state sequences associated with $\mathbf{a}$ and $\hat{\mathbf{a}}$, respectively. Let E be the set of error sequences corresponding to all

possible error events at epoch j_1. Then, the average bit error probability is bounded by

$$P_b \le \frac{1}{n} \sum_{\mathbf{e} \in \mathbf{E}} w_b(\mathbf{e}) \sum_{\mathbf{a}} P(\mathbf{a}) \, P\left(\Gamma(\mathbf{a} \oplus \mathbf{e}) \ge \Gamma(\mathbf{a}) \bigg| \mathbf{a} \right) \qquad (8.43)$$

where $\Gamma(\mathbf{a})$ is the path metric of $\mathbf{a}$, and $w_b(\mathbf{e})$ is the number of bit errors associated with e. The factor $1/n$ appears in front of the first summation, because n information bits are transmitted per epoch (or per branch in the trellis diagram). The second summation is over all possible information sequences, because each sequence $\mathbf{a}$ can have $\mathbf{e}$ as the error sequence. This is necessary for trellis codes because the signal mapping and, hence, the codes are nonlinear.

Another way of writing the bound on the bit error probability in (8.43) is

$$P_b \le \sum_{\mathbf{x} \in \mathbf{C}} \sum_{\hat{\mathbf{x}} \in \mathbf{C}} w_b(\mathbf{x}, \hat{\mathbf{x}}) P(\mathbf{x}) P(\mathbf{x} \to \hat{\mathbf{x}}) \qquad (8.44)$$

where $\mathbf{C}$ is the set of all coded symbol sequences, $w_b(\mathbf{x}, \hat{\mathbf{x}})$ is the number of bit errors that occur when the sequence $\mathbf{x}$ is transmitted and the sequence $\hat{\mathbf{x}} \ne \mathbf{x}$ is chosen by the decoder, $P(x)$ is the a priori probability of transmitting x, and $P(\mathbf{x} \to \hat{\mathbf{x}})$ is the pairwise error probability.

For convolutional codes the upper bound in (8.43) simplifies because the codes are linear, meaning that the sum of any two codewords is another codeword and that all-zeroes sequence is a codeword [199]. Because of this property, we can assume that $\mathbf{a} = \mathbf{0}$, so that the union bound becomes

$$P_b \le \frac{1}{k} \sum_{\mathbf{e} \in \mathbf{E}} w_b(\mathbf{e}) \, P\left(\Gamma(\mathbf{e}) \ge \Gamma(\mathbf{0}) \right) \qquad (8.45)$$

Note that we divide by k rather than n in front of the summation, because a convolutional code transmits k bits per epoch whereas a trellis code transmits n bits per epoch.

4.1   UNION BOUND FOR CONVOLUTIONAL CODES

Evaluation of the error probability upper bound for trellis codes is complicated by the fact that trellis codes are nonlinear and, therefore, all possible correct sequences must be considered when computing the upper bound. We will defer treatment of the coded error probability upper bound for trellis codes until Section 6.4, where we will consider the more general case of TCM on a fading ISI channel. In this section we will show how the error probability upper bound can be computed for convolutional codes with Viterbi decoding.

For convolutional codes the set $\mathbf{E}$ in (8.45) consists of all sequences that begin and end at the zero-state in the state diagram. The enumeration of these

sequences (or codewords) along with their associated Hamming distances, information weights, and lengths, was obtained earlier by computing the transfer function, $T(D, N, L)$, of the augmented state diagram. When a particular incorrect path through the trellis is selected over the all-zeroes path at a given node in the trellis, the corresponding number of bits errors, $w_b(\mathbf{e})$, is given by the exponent of N in the transfer function. Multiplying $w_b(\mathbf{e})$ by the pairwise error probability $P\left(\Gamma(\mathbf{e}) \geq \Gamma(\mathbf{0})\right)$ for that path and dividing by the number of input bits per branch, k, gives the bit error rate for that path. Summing over the set of all possible incorrect sequences $\mathbf{E}$ yields a union bound on the bit error probability.

In general, the transfer function $T(D, N)$ for a convolutional code has the form

$$T(D, N) = \sum_{d=d_{\text{free}}}^{\infty} a_d D^d N^{f(d)} \tag{8.46}$$

where $f(d)$ is the exponent of N as a function of d. For the example in (8.29), $a_d = 2^{d-5}$ and $f(d) = d - 4$. Differentiating $T(D, N)$ with respect to N and setting $N = 1$ gives

$$\frac{dT(D, N)}{dN}\bigg|_{N=1} = \sum_{d=d_{\text{free}}}^{\infty} a_d f(d) D^d \tag{8.47}$$

Once again, for the example in (8.29) this leads to

$$\frac{dT(D, N)}{dN}\bigg|_{N=1} = \sum_{d=d_{\text{free}}}^{\infty} 2^{d-5}(d-4)D^d \ . \tag{8.48}$$

The pairwise error probability in (8.45) depends on the type of modulation, detection, and decoding that is employed. The code bits are mapped onto symbols taken from a signal constellation, and transmitted over the channel. The sampled output of the receiver matched filter at epoch k is

$$y_k = x_k + \eta_k \tag{8.49}$$

where x_k is one of the M *low-pass* points in the signal constellation and η_k is a zero-mean complex-valued Gaussian random variable with variance N_o. For convolutional codes, two types of decoding can be used, **hard decision decoding** and **soft decision decoding.** Soft decision decoders do not make symbol by symbol decisions on the received symbols, rather, the decoder operates directly on the sequence of matched filter outputs $\mathbf{y}$. For an AWGN channel, the MLSE receiver searches for the symbol sequence $\hat{\mathbf{x}}$ that is closest in Euclidean distance to the received sequence y. Following the same argument used in Chapter 5.2, the MLSE receiver decides in favor of the sequence $\hat{\mathbf{x}}$ that maximizes the metric

$$\mu(\mathbf{x}) = -\|\mathbf{y} - \mathbf{x}\|^2 \ . \tag{8.50}$$

The sequence $\hat{\mathbf{x}}$ corresponds to a unique sequence $\hat{\mathbf{a}}$ that is the final estimate of the transmitted information sequence $\mathbf{a}$.

In general, the pairwise error probability for an AWGN channel that is associated with an error event of length ℓ beginning at epoch k_1 is

$$P_2(\ell) = Q\left(\sqrt{\frac{\Delta^2}{4N_o}}\right) \tag{8.51}$$

where

$$\Delta^2 = \sum_{k=k_1}^{k_1+\ell+1} \delta_k^2 \tag{8.52}$$

$$\delta_k^2 = |x_k - \hat{x}_k|^2 \tag{8.53}$$

and $\mathbf{x} = \{x_k\}$ and $\hat{\mathbf{x}} = \{\hat{x}_k\}$ are the symbol sequences corresponding to the information sequences $\mathbf{a}$ and $\hat{\mathbf{a}}$, respectively. The parameter δ_k^2 is the **squared branch Euclidean distance** associated with branch k, and Δ^2 is the **squared path Euclidean distance** associated with the error event. Clearly, the pairwise error probability depends on the particular mapping between the encoder output bits and the points in the signal constellation. Suppose for example that code bits are mapped onto a BPSK signal constellation. Then the pairwise error probability between the two codewords $\mathbf{b}$ and $\hat{\mathbf{b}}$ that differ in d positions is

$$P_2(d) = Q(\sqrt{2R_c d \gamma_b}) \tag{8.54}$$

where γ_b, is the received bit energy-to-noise ratio[1]. Therefore, the union bound on bit error probability becomes

$$P_b \le \frac{1}{k} \sum_{d=d_{\mathbf{free}}}^{\infty} a_d f(d) P_2(d) \ . \tag{8.55}$$

Note that we have explicitly shown the pairwise error probability to be a function of the Hamming distance between the codewords in (8.54). However, it is very important to realize that this property does not apply to all convolutionally encoded systems. For example, suppose that the outputs of the rate-2/3 convolutional encoder in Fig. 8.2 are mapped onto symbols from an 8-PSK signal constellation. In this case, the pairwise error probability depends not only on the Hamming distance between codewords, but also upon the particular mapping between the 8-PSK symbols and the encoder outputs.

Hard decision decoders make symbol by symbol decisions on the received sequence of matched filter outputs $\mathbf{y} = \{y_k\}$ to yield the received symbol

---

[1]The received symbol energy-to-noise ratio is $\gamma_s = R_c \gamma_b$

sequence $\tilde{\mathbf{x}}$. A **minimum distance decoder** decides in favor of the symbol sequence $\hat{\mathbf{x}}$ that is closest in Hamming distance to the received symbol sequence $\tilde{\mathbf{x}}$. Again, the pairwise error probability depends on the particular mapping between the encoder outputs and the points in the signal constellation. If BPSK signaling is used, for example, then the pairwise error probability between two codewords $\mathbf{b}$ and $\hat{\mathbf{b}}$ at Hamming distance d is

$$P_2(d) = \begin{cases} \sum_{k=(d+1)/2}^{d} \binom{d}{k} p^k (1-p)^{d-k} , & d \text{ odd} \\ \sum_{k=d/2+1}^{d} \binom{d}{k} p^k (1-p)^{d-k} + \frac{1}{2}\binom{d}{d/2} p^{d/2}(1-p)^{d/2} , & d \text{ even} \end{cases}$$

(8.56)

where

$$p = Q(\sqrt{2R_c\gamma_b})$$

(8.57)

is the probability of symbol error. Once again, the pairwise error probability for BPSK is a function of the Hamming distance between the codewords.

The union bound in (8.55) can be simplified by imposing a Chernoff bound (see Appendix A) on the pairwise error probability. First consider the case of soft decision decoding. Suppose that sequence $\mathbf{x}$ is transmitted and $\mathbf{y}$ is the received sequence. Then the pairwise error probability between sequences $\mathbf{x}$ and $\hat{\mathbf{x}}$ with an ML receiver can be Chernoff bounded by

$$\begin{aligned} \mathrm{P}(\mathbf{x} \to \hat{\mathbf{x}}) &= \mathrm{P}\left(\|\mathbf{y}-\hat{\mathbf{x}}\|^2 < \|\mathbf{y}-\mathbf{x}\|^2\right) \\ &\leq \mathrm{E}\left[\exp\left\{\lambda\left(\|\mathbf{y}-\mathbf{x}\|^2 - \|\mathbf{y}-\hat{\mathbf{x}}\|^2\right)\right\}|\mathbf{x}\right] . \end{aligned}$$

(8.58)

Substituting $\mathbf{y} = \mathbf{x} + \boldsymbol{\eta}$, taking the expectation over the Gaussian random vector $\boldsymbol{\eta}$, and simplifying gives

$$\mathrm{P}(\mathbf{x} \to \hat{\mathbf{x}}) \leq \exp\left\{-\lambda\|\mathbf{x}-\hat{\mathbf{x}}\|^2(1-\lambda 2N_o)\right\} .$$

(8.59)

The tightest upper bound is obtained with $\lambda^* = 1/(4N_o)$ yielding

$$\mathrm{P}(\mathbf{x} \to \hat{\mathbf{x}}) \leq \exp\left\{\frac{-\|\mathbf{x}-\hat{\mathbf{x}}\|^2}{8N_o}\right\} .$$

(8.60)

Finally, if the signal constellation is *normalized* so that $\mathrm{E}[|x_i|^2] = 1$, then the Chernoff bound can be written in the form

$$\mathrm{P}(\mathbf{x} \to \hat{\mathbf{x}}) \leq \exp\left\{-\frac{\gamma_s}{4}\|\mathbf{x}-\hat{\mathbf{x}}\|^2\right\}$$

(8.61)

where γ_s is the received symbol energy-to-noise ratio.

For the case of BPSK signaling on an AWGN channel, the Chernoff bound on the pairwise error probability becomes

$$P_2(d) \leq e^{-R_c d\gamma_b} .$$

(8.62)

Likewise, if BPSK signaling is used with hard decision decoding, then the pairwise error probability has the Chernoff bound

$$P_2(d) \leq [4p(1-p)]^{d/2} .$$ (8.63)

Notice that the Hamming distance d appears in the exponent of the pairwise error probability. The resulting upper bound on bit error probability is called a **union-Chernoff bound** and has the simple form

$$P_b \leq \frac{1}{k} \frac{dT(D,N)}{dN} \Big|_{N=1,D=Z}$$ (8.64)

where

$$Z = \begin{cases} \sqrt{4p(1-p)} , & \text{hard decision decoding} \\ e^{-R_c \gamma_b} , & \text{soft decision decoding} \end{cases} .$$ (8.65)

At high SNR, the performance is dominated by the error events with minimum Hamming distance. Since the minimum distance error events are not necessarily mutually exclusive, the bit error probability at high SNR is approximately

$$\begin{aligned} P_b &\approx \frac{1}{k} a_{d_{\text{free}}} f(d_{\text{free}}) P_2(d_{\text{free}}) \\ &\leq \frac{1}{k} a_{d_{\text{free}}} f(d_{\text{free}}) Z^{d_{\text{free}}} . \end{aligned}$$ (8.66)

The above procedure for upper bounding the error probability is called the **transfer function approach,** because it relies upon the transfer function of the state diagram. The transfer function approach, however, has its limitations. First, if the number of encoder states is large, then obtaining the transfer function $T(D, N)$ quickly becomes intractable. Second, if the pairwise error probability is not just a function of the Hamming distance between allowable code sequences, then the branch labeling in the augmented state diagram must be done differently and the Chernoff bound cannot be employed. These problems can be overcome by a using a different approach to compute the upper bound, such as the stack algorithm presented in Section 6.4.

5. CODED PERFORMANCE ON INTERLEAVED FLAT FADING CHANNELS

Fig. 8.15 is a block diagram of a coded communication system operating on an interleaved flat fading channel. The information sequence **a** is encoded and mapped onto a signal set to generate the symbol sequence **x** by using either convolutional coding or trellis coded modulation. The symbol sequence is then interleaved (or scrambled), and the resulting sequence $\check{\mathbf{x}}$ is filtered for spectral shaping and transmitted over the channel. The receiver employs a filter
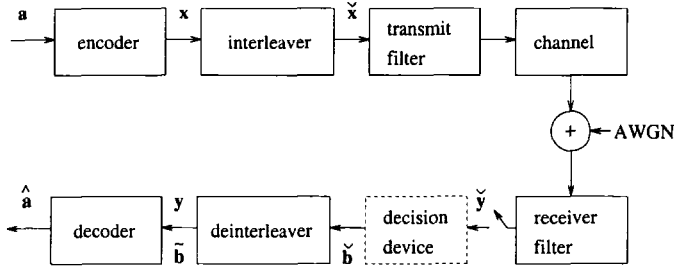
Figure 8.15. Digital communication on an interleaved flat fading channel.

that is matched to the transmitted pulse and symbol- or T-spaced samples are taken at the output of the matched filter. With hard decision decoding, these outputs are applied to a decision device and deinterleaved to yield the received code sequence $\tilde{\mathbf{b}}$. With soft decision decoding, the received samples $\check{\mathbf{y}}$ are deinterleaved to generate the sequence $\mathbf{y}$ which is then input to the Viterbi decoder.

The channel is characterized by flat Ricean fading, so that the sampled output of the matched filter is

$$\check{y}_k = \check{\alpha}_k \check{x}_k + \check{\eta}_k \tag{8.67}$$

where the pdf of $\check{\alpha}_k$ is

$$p_{\check{\alpha}}(x) = \frac{2x(1+K)}{\Omega_p} \exp\left\{-K - \frac{(K+1)x^2}{\Omega_p}\right\} I_0\left(2x\sqrt{\frac{K(K+1)}{\Omega_p}}\right) \tag{8.68}$$

where $\Omega_p = \mathrm{E}[\check{\alpha}_k^2]$. The $\check{\eta}_k$ are independent zero-mean complex Gaussian random variables with variance $.N_o$.

The interleaver serves to reduce the correlation between the fades experienced by successive source symbols that are transmitted over the channel. There are a variety of interleaver structures [63], and the interleaver that we consider here is a **block interleaver.** A block interleaver can be regarded as a buffer with J rows and M columns, where J represents the interleaving depth and M represents the interleaving span. The length of the interleaver is JM symbols. Source symbols are fed into the buffer in successive rows and transmitted out of the buffer in columns. The deinterleaver performs the reverse operation. In practise, the interleaver depth J should be chosen so that successive source symbols, which are actually transmitted J symbol durations apart, are independently faded. In a 2-D isotropic scattering environment, the fades experienced at two different locations separated by a half wavelength are approximately uncorrelated. If the signals are received by a mobile station (MS) traveling with a speed of v km/h relative to the base station (BS), then the spatial distance associated with one symbol duration T is equal to vT.

Therefore, we should have $JvT > 0.5\lambda$, where λ is the carrier wavelength. For example, a carrier frequency of 900 MHz yields

$$J > 0.6\frac{R}{v} \quad \text{(symbols)} \tag{8.69}$$

where R is the signaling rate (symbols/s), v is the vehicle speed (km/h) and J is in units of symbols. Observe that the required interleaving depth is inversely proportional to the speed and, therefore, slow moving MSs require large interleaving depths. For $R = 24$ ks/s and $v = 30$ km/h, $J = 478$ symbols.

The basic objective of any interleaver is to at least separate any $L_D + 1$ successive source symbols as far apart as possible, where L_D is the decoding depth. Hence, M should be at least equal to $L_D + 1$ yielding an interleaving delay of

$$t_d = 0.6 \frac{(L_D + 1)}{v} \quad \text{(seconds)} . \tag{8.70}$$

For example, with $L_D = 13$ symbols and a MS speed of 30 km/h, the interleaving delay should be at least 280 ms. This delay is quite large, especially for voice applications, and the problem is exasperated by lower MS speeds. One possible solution is to design codes that minimize the decoding depth, L_D. The other solution is to use better interleaving techniques so that the effective interleaving length is longer than the actual interleaving length JM.

For analytical purposes, an infinite interleaving depth is often assumed so that the deinterleaved sequence $\{\alpha_k\}$ is a sequence of independent random variables. In this case the conditional density of y has the product from

$$p(\mathbf{y}|\boldsymbol{\alpha} \cdot \mathbf{x}) = \prod_k p(y_k|\alpha_k x_k) . \tag{8.71}$$

Suppose that sequence x is transmitted and the vector $\mathbf{y} = \boldsymbol{\alpha} \cdot \mathbf{x} + \boldsymbol{\eta}$ is received. An ML receiver having perfect knowledge of a chooses the sequence $\hat{\mathbf{x}}$ that maximizes the metric

$$\mu(\hat{\mathbf{x}}) = -\|\mathbf{y} - \boldsymbol{\alpha} \cdot \hat{\mathbf{x}}\|^2 . \tag{8.72}$$

The pairwise error probability between the sequences $\mathbf{x}$ and $\hat{\mathbf{x}}$ has the Chernoff bound

$$P(\mathbf{x} \to \hat{\mathbf{x}}) \leq \exp\left\{-\frac{\|\boldsymbol{\alpha} \cdot (\mathbf{x} - \hat{\mathbf{x}})\|^2}{4N_o}\right\} . \tag{8.73}$$

Once again, if we assume the normalization $E[|x_k|^2] = 1$ then the Chernoff bound becomes

$$P(\mathbf{x} \to \hat{\mathbf{x}}) \leq \exp\left\{-\frac{E}{4N_o}\|\boldsymbol{\alpha} \cdot (\mathbf{x} - \hat{\mathbf{x}})\|^2\right\} . \tag{8.74}$$

Averaging (8.74) over the probability density function in (8.68) gives [82]

$$P(\mathbf{x} \to \hat{\mathbf{x}}) \le \prod_{i \in A} \frac{1 + K}{1 + K + \frac{\bar{\gamma}_s}{4}|x_i - \hat{x}_i|^2} \exp\left\{ -\frac{K\frac{\bar{\gamma}_s}{4}|x_i - \hat{x}_i|^2}{1 + K + \frac{\bar{\gamma}_s}{4}|x_i - \hat{x}_i|^2} \right\}$$

(8.75)

where $\bar{\gamma}_s = \mathrm{E}[\alpha^2]E/N_o$ is the average received symbol energy-to-noise ratio, and $A = \{i|x_i \ne \hat{x}_i\}$. At sufficiently high $\bar{\gamma}_s$, (8.75) simplifies to

$$P(\mathbf{x} \to \hat{\mathbf{x}}) \le \prod_{i \in A} \frac{4(1 + K)}{\bar{\gamma}_s |x_i - \hat{x}_i|^2} e^{-K} \ .$$

(8.76)

It follows that the bound in (8.43) will be dominated by the error event path having the smallest number of elements in set A. Divsalar and Simon [82, 83] called this path the **shortest error event path** and defined its length as $L_{\min}$. Based on previous arguments, the bit error probability can be approximated as

$$P_b \simeq C\left(\frac{(1 + K)e^{-K}}{\bar{\gamma}_s} \right)^{L_{\min}} \quad \bar{\gamma}_s \gg K$$

(8.77)

where C is a constant that depends on the distance structure of the code. Observe that P_b varies inversely with $(\bar{\gamma}_s)^{L_{\min}}$, yielding a diversity effect of order $L_{\min}$. Wei [353] called $L_{\min}$ the **minimum built-in time diversity** (MTD). The MTD dominates the performance of TCM on an interleaved flat fading channel, and the maximization of the MTD is the major design criterion for TCM on interleaved flat fading channels.

The pairwise error probability in (8.75) can be written in the form

$$P(\mathbf{x} \to \hat{\mathbf{x}}) \le e^{-\frac{\bar{\gamma}_s}{4}d^2}$$

(8.78)

where

$$
\begin{aligned}
d^2 &= \sum_{i \in A} \frac{|x_i - \hat{x}_i|^2 K}{1 + K + \frac{\bar{\gamma}_s}{4}|x_i - \hat{x}_i|^2} + \left(\frac{\bar{\gamma}_s}{4} \right)^{-1} \ln\left(\frac{1 + K + \frac{\bar{\gamma}_s}{4}|x_i - \hat{x}_i|^2}{1 + K} \right) \\
&= \sum_{i \in A} d_{1i}^2 + d_{2i}^2
\end{aligned}
$$

(8.79)

Two special cases are associated with (8.79), $K = \infty$ and $K = 0$. For $K = \infty$ (no fading),

$$d_{1i}^2 = |x_i - \hat{x}_i|^2, \qquad d_{2i}^2 = 0$$

(8.80)

and, therefore, d^2 becomes the sum of the squared Euclidean distances over the error event path. Maximizing d^2 under this condition is the TCM design criterion for AWGN channels.

For $K = 0$ (Rayleigh fading),

$$d_{1i}^2 = 0, \qquad d_{2i}^2 = \left(\frac{\bar{\gamma}_s}{4}\right)^{-1} \ln\left(1 + \frac{\bar{\gamma}_s}{4}|x_i - \hat{x}_i|^2\right) \ . \qquad (8.81)$$

For reasonably large SNR, d^2 is the sum of the logarithms of the squared Euclidean distances, each weighted by $\bar{\gamma}_s$. In this case, the pairwise error probability is given by

$$P(\mathbf{x} \to \hat{\mathbf{x}}) \leq \left(\prod_{i \in A} \frac{\bar{\gamma}_s}{4}|x_i - \hat{x}_i|^2\right)^{-1} \qquad (8.82)$$

which is inversely proportional to the product of the squared Euclidean distances along the error event path. The **minimum product squared Euclidean distance** (MPSD) between any two valid sequences,

$$\min_{\mathbf{x}, \hat{\mathbf{x}}} \prod_{i \in A} |x_i - \hat{x}_i|^2 \qquad (8.83)$$

is another design parameter for Rayleigh fading channels. For values of K between 0 and ∞, the equivalent squared Euclidean distance of (8.79) becomes a mixture of the two limiting cases given above.

If interleaving is not used, then the assumption that the fading is independent from symbol to symbol is no longer valid. If the fading is slow enough to be considered constant over the duration of the minimum distance error event path, then for coherent detection with a Gaussian metric the bit error probability at high SNR is, approximately,

$$P_b \simeq C_1 \mathrm{E}\left[\exp\left\{-\frac{\gamma_s}{4}d_{\min}^2\right\}\right] \qquad (8.84)$$

where C_1 is a constant, $\gamma_s = \alpha^2 E/N_o$ is the received symbol energy-to-noise ratio, $d_{\min}^2$ is the minimum Euclidean distance of the code, and the averaging is over the density in (8.68). Taking this average gives

$$P_b \simeq C_1 \frac{1 + K}{1 + K + d_{\min}^2 \frac{\bar{\gamma}_s}{4}|x_i - \hat{x}_i|^2} \exp\left\{-\frac{K d_{\min}^2 \frac{\bar{\gamma}_s}{4}|x_i - \hat{x}_i|^2}{1 + K + d_{\min}^2 \frac{\bar{\gamma}_s}{4}|x_i - \hat{x}_i|^2}\right\} \qquad (8.85)$$

which can be approximated at large $\bar{\gamma}_s$ by

$$P_b \simeq 4C_1 \frac{1 + K}{d_{\min}^2 \bar{\gamma}_s} e^{-K} \ . \qquad (8.86)$$

Observe that without interleaving, P_b is asymptotically inverse linear with $\bar{\gamma}_s$, independent of the trellis code. If follows that interleaving is required to achieve diversity with TCM on a flat fading channel.

5.1 DESIGN RULES FOR TCM ON FLAT FADING CHANNELS

According to the previous section, when TCM is used on Ricean fading channel with interleaving/deinterleaving, the design of the code for optimum performance is guided by the minimum built-in time diversity (MTD) of the code. For Rayleigh fading channels, the design of the code is also guided by the minimum product squared distance (MPSD) of the code. The minimum Euclidean distance, which is the principal design criterion for trellis coded modulation AWGN channels, plays a less significant role on Ricean fading channels as the K factor decreases, and no role for Rayleigh fading channels ($K = 0$). A third design criterion is to minimize the decoding depth of the code.

The design of trellis codes for interleaved flat fading channels is based on Ungerboeck's principle of mapping by set partitioning, but now the partitioning is done to maximize the MTD and MPSD of the code. This can be accomplished by maximizing the intra-subset MTD and MPSD, but it should be pointed out that large MTD and MPSD can be sometimes achieved even if the partitioning is done to maximize the minimum Euclidean distance as in Ungerboeck's codes for AWGN channels.

In general, the following guidelines are followed when designing trellis codes for interleaved flat fading channels;

1. All signals occur with equal frequency and with regularity and symmetry.

2. Transitions originating from the even and odd numbered states are assigned signals from the first and second subsets, respectively, of the first partitioning level.

3. Whenever possible, the transitions joining in the same state receive signals from either the first or second subset of the first partitioning level.

4. Parallel transitions receive signals from the same subset of the finest partitioning level.

5. The state transitions originating from each current state and going to even-numbered next states are assigned signals from subsets whose inter-subset MTD and MPSD are maximized. The same applies for the transition originating from each current state and going to odd-numbered next states.

The first four rules are similar to those suggested by Ungerboeck [330], but now the subsets used may be different. The fifth rule is used to reduce the **decoding depth** of the code. Wei [353] developed several codes based on minimizing the decoding depth of a code. He defined two minimum decoding depths (MDD1, MDD2) to characterize a code. MDD1+1 is defined as the

length (in symbols) of the longest valid sequence of signal points, say $\mathbf{x_1}$, which originates from the same state as another valid sequence x and merges into the same last state as x and whose Hamming distance from x is the same as the MTD of the code. Note that the performance of a code is mainly governed by the pairs of sequences which determine the MTD of the code. Each such pair of sequences differ in at most MDD1+1 successive symbols. The farther these symbols are separated, the better the performance of the code. Hence, to benefit from the MTD of the code, the interleaver should separate the symbols in each sequence of MDD1+1 input symbols as far as possible.

MDD2 is defined as the length of the longest unmerged valid sequence of signal points, say $\mathbf{x_2}$, which originates from the same state as another valid sequence, say x, and whose Hamming distance from x is not greater than the MTD of the code. In case the Hamming distance between the two sequences is equal to the MTD of the code, the squared product distance between the two sequences must be less than the MPSD of the code. Since MDD2 is greater than MDD1, the decoding depth should be at least equal to MDD2 to realize the MTD and MPSD of a code. It suffices if the decoding depth is few symbols longer than MDD2. Finally, to benefit from both the MTD and MPSD of a code, the interleaver should separate the symbol in each sequence of MDD2+1 input symbols as far as possible.

5.1.1   MULTIDIMENSIONAL TCM

Recall that the length of the shortest error event with conventional trellis codes (one symbol per trellis branch) is equal to the number of branches along that error event path. If the trellis code has parallel transitions, then MTD = 1. Unfortunately, parallel transitions are inevitable when the size of the signal constellation exceeds the number of states. In this case, the bit error probability for Rayleigh fading channels has an inverse linear dependency on the bit energy-to-noise ratio. To solve this problem, multidimensional TCM techniques can be used.

Multidimensional TCM uses signal spaces having a larger dimensionality so as to increase the minimum Euclidean distance between signal points. Another feature of multidimensional trellis codes is noticed when comparing the coding gain of these codes to 1- or 2-D codes. When the size of the signal constellation is doubled with respect to uncoded modulation, the average signal energy may also increase. For example, doubling the size of a 2-D M-QAM constellation implies a 3 dB increase in average signal energy. However, if this increase in average signal energy can be avoided, then the TCM coding gain would be greater. This 3 dB cost falls to 1.5 dB when four dimensions are used, and to 0.75 dB when eight dimensions are used [38]. Multidimensional TCM is also attractive for fading channels. 4-D TCM schemes are special because they can be implemented in radio communications without any increase in bandwidth,

by transmitting on the same carrier frequency with two spatially orthogonal electric field polarizations [38].

A $2N$-D constellation is formed by first selecting a constituent 2-D constellation and then concatenating N such constellations together in the time domain. If the size of the $2N$-D constellation is larger than needed, then some of the *less desirable* points are deleted [353]. The resulting constellation is then partitioned into a chain of increasingly large numbers of subsets. The partitioning is performed first to maximize the intra-subset MTD, and then to maximize the MPSD between any pair of $2N$-D signal points within the same subset having that MTD.

When $N > 1$, an MTD of at least two is easily achieved for each subset in the finest partition. Fig. 8.16 shows a 32-point 4-D 8-PSK constellation that is partitioned into 8-subsets. The 4-D constellation is formed by concatenating a pair of 2-D 8-PSK constellations in the time domain and deleting those points having the form (even,odd),(odd,even). The intra-subset MTD within each of the finest partitions is 2 with an intra-subset MPSD of 4. Fig. 8.17 shows a rate-2/3, 4-D, 4-state, 8-PSK trellis code with 2 bits/symbol. The bits b_1, b_2, b_3 are used to select one of the 8 subsets in Fig. 8.16 and bits b_4, b_5 are used to select one of the four 4-D elements within each subset. The MTD and MPSD of the code are the same as the intra-subset MTD and MPSD and, hence, are maximized for the partitioning in Fig. 8.16.

Fig. 8.17 also shows the trellis diagram of the code, along with examples of the longest sequences which determine the values of MDD1, and MDD2. Note that MDD1 and MDD2 are measured in units of 2-D symbols. Since there are parallel transitions of length 2 symbols, MDD1 = 1. To find MDD2, suppose that the all zeroes sequence is the reference sequence. Note that the 2-D sequence {1, 5, 0, 0, 0} associated with the 4-D sequence {(1,5), (0,0), (0,2)} has a time diversity of 2 (which is equal to the MTD) and an MPSD of $d^2(0,1) \times d^2(0,5) = 0.5857 \times 3.414 < 4$ and, hence, MDD2 = 5. Wei [353] investigated different multi-dimensional codes. He found that multi-dimensional TCM requires longer decoding depths than 2-D TCM. This longer decoding depth has proven to be very detrimental and, therefore, Wei considered only 4-D codes.

5.1.2   MULTIPLE TCM (MTCM)

Multiple TCM is implemented by using a rate-b/s encoder, where the encoder outputs are mapped onto k M-ary symbols in each transmission interval, as shown in Fig. 8.18. The s encoder output symbols are divided into k groups of $m = \log_2 M$ symbols each, in this case $s = k \log_2 M$. Another method is to divide the s binary symbols into k groups of m_i symbols where each group now corresponds to a signal constellation of different size M_i. If $m_i = \log_2 M_i$ for the *ith* group, then $s = \sum_{i=1}^{k} m_i$. Notice that $k = 1$ corresponds to conven-
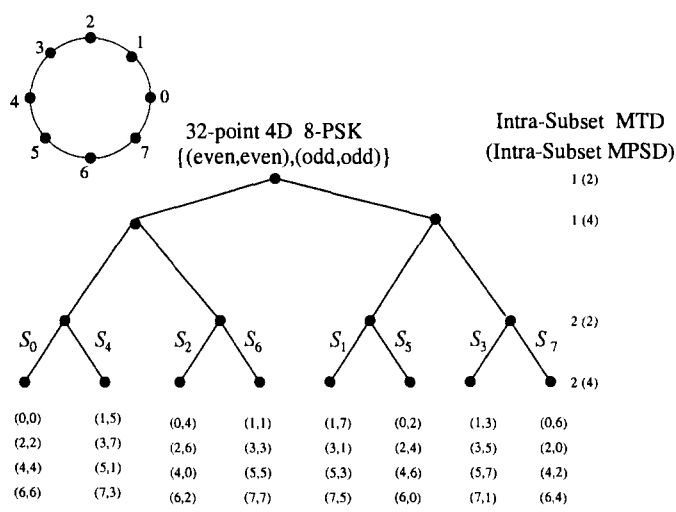
Figure 8.16. Partition of the 32-point 4-D 8-PSK constellation into 8 subsets.
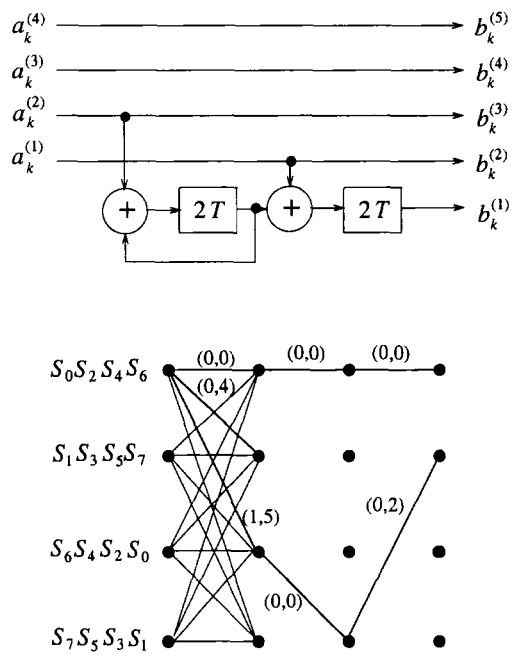

Figure 8.17. 4-D 4-state 8-PSK rate-2/3 trellis code with 2 bits/symbol.

tional Ungerboeck trellis codes. MTCM codes can be designed with parallel transitions, while still achieving an asymptotic bit error probability on fading channels that decays faster than an inverse linear function of $\bar{\gamma}_s$.
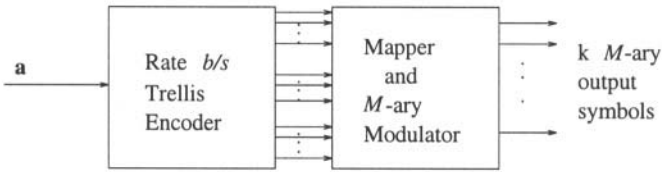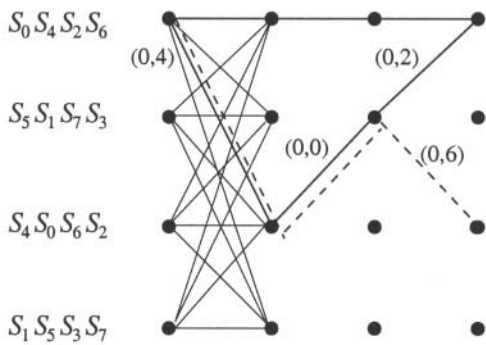
Figure 8.18.   MTCM encoder.

Figure 8.19.   Code trellis for the rate-4/5 4-state 8-PSK MTCM trellis code.

Fig. 8.19 shows the code trellis for the rate-4/6, 4-state, 8-PSK MTCM code reported in [84]. The signal point sets are obtained using the above method and they are the same as those used in the 4-D code and shown in Fig. 8.16. There are 16 paths emanating from each node and, hence, there are four parallel paths between nodes. This code has the same MTD and MPSD as the previous 4-D code. However, MDD1 = 5 and MDD2 = 5 and, therefore, the previous 4-D code remains a better choice since MDD1 is smaller.

5.1.3   2-D TRELLIS CODES

It is surprising that Ungerboeck's 2-D 8-state and 16-state 8-PSK trellis codes that were originally designed for AWGN channels, remain good for interleaved flat fading channels. In fact, Ungerboeck's 2-D, rate-2/3, 8-state 8-PSK trellis code shown in Fig. 8.20, has an MTD of two (the thick lines) and the corresponding product squared Euclidean distance is 8 which is better than the previous codes. Note that the shortest error event is not necessarily the minimum distance error event. In Fig. 8.20, the minimum squared Euclidean distance is 4.585, corresponding to an error event of length 3 (the dashed path). Note also that MDD1 = 3 and MDD2 = 3. It is obvious that the set partitioning for this code was intended to maximize the minimum squared Euclidean distance. Finally, we note that good 2-D TCM codes for interleaved
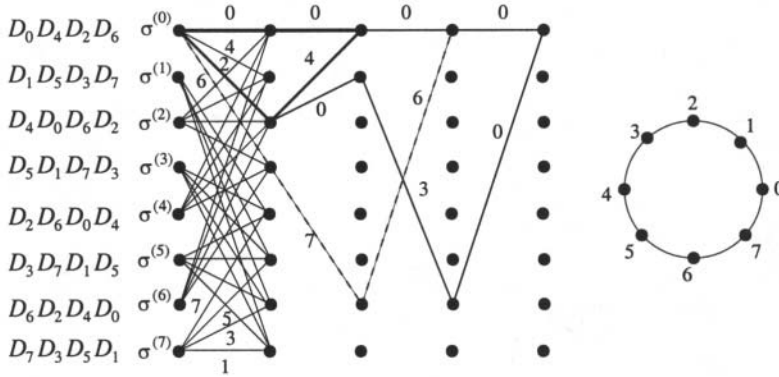
Figure 8.20.   Trellis diagram for Ungerboeck's rate-2/3, 8-state, 8-PSK trellis code.

flat fading channels will not have parallel transitions and the connectivity between the states will be as low as possible.

6.   CODED PERFORMANCE ON ISI CHANNELS

Fig. 8.21 shows a model for digital signaling on a *non-interleaved* fading ISI channel. Chapter 7.3.1 showed that the overall channel with D-branch diversity can be replaced with the model in Fig. 7.5, consisting of D $(L + 1)$-tap transversal filters, the outputs of which are corrupted by AWGN samples $\{\eta_k^d\}$ with $\sigma_{\eta_k^d}^2 = N_o \, \forall \, k, \, d$. For TCM, this leads to the equivalent discrete-time model shown in Fig. 8.22. As discussed in Chapter 7.3.1.1, the tap gains are modeled as uncorrelated complex Gaussian processes, and with 2-D isotropic scattering the tap gain vector $\mathbf{g}_d(k) = (g_{0,d}(k), g_{1,d}(k), \dots, g_{L,d}(k))^T$ has covariance matrix $\mathbf{\Phi}_{\mathbf{g}_d}(m) = J_0(2\pi f_m mT) \mathbf{\Sigma}_d^2$ where $J_0(\cdot)$ is the zero-order Bessel function of the first kind and f_m is the maximum Doppler frequency, and $\mathbf{\Sigma}_d = \mathrm{diag}[\sigma_{0,d}, \sigma_{1,d}, \dots, \sigma_{L,d}]$ and $\sigma_{i,d}^2 = \frac{1}{2}\mathrm{E}[|g_{i,d}|^2]$. Here, we assume the $g_{i,d}(k)$ have zero-mean so that the magnitudes $|g_{i,d}(k)|$ are Rayleigh distributed. Assuming that the branches are balanced, the average received branch bit energy-to-noise ratio is

$$\bar{\gamma}_c = \frac{\mathrm{E}[|x_k|^2] \sum_{i=0}^{L} \mathrm{E}[|g_{i,d}|^2]}{2nN_o} \tag{8.87}$$

where n is the number of bits per symbol.

As discussed earlier, the rate-$m/(m + r)$ linear convolutional encoder contains m shift registers and is characterized by a set of generator polynomials $\mathbf{g}_i^{(j)}$, $1 \le i \le m$, $1 \le j \le m + r$. The length of the ith shift register is ν_i and the total number of memory elements in the encoder is $\nu_T = \sum_{i=1}^{m} \nu_i$. Since both the trellis encoder and ISI channel are finite state machines, it follows
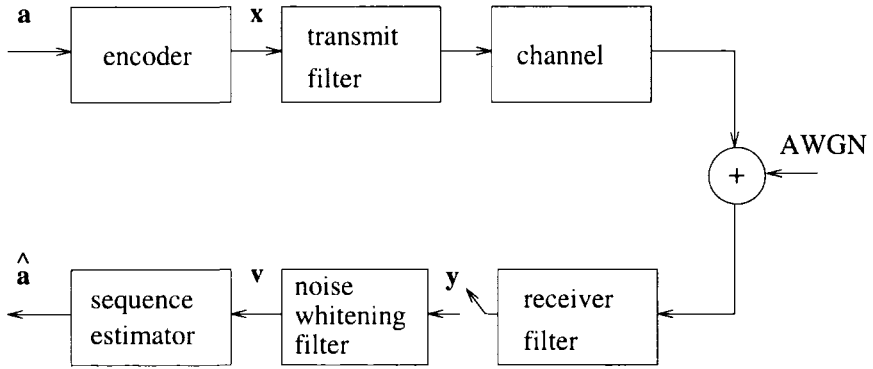
Figure 8.21. Digital communication on a non-interleaved fading ISI channel.
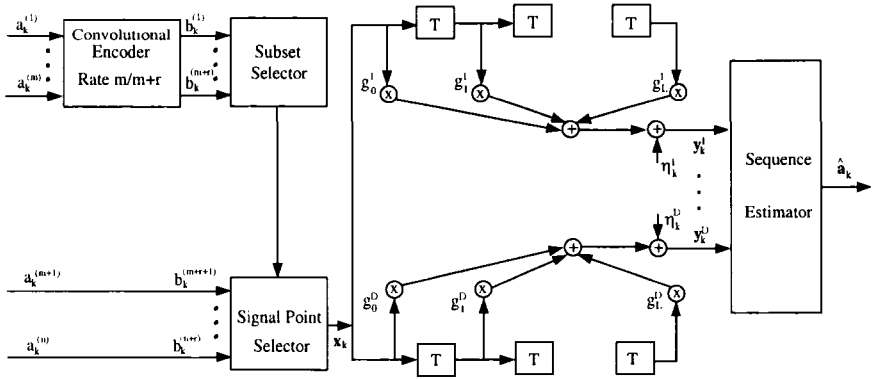
Figure 8.22. TCM-coded system with an equivalent discrete-time white noise channel model, from [303].

that the combined trellis encoder and ISI channel is also a finite state machine having **super-states**

$$\mathbf{s}_k^{(i)} = \left(\boldsymbol{\sigma}_k; x_{k-1}, \ldots, x_{k-L} \right) \ . \tag{8.88}$$

There are $N_S = 2^{\nu_T} 2^{nL} = 2^{\nu_S}$ super-states and $\boldsymbol{\sigma}_k$ is the encoder state defined in (8.28). An equivalent definition of the super-state is [50]

$$\begin{aligned} \mathbf{s}_k^{(i)} &= \left(\boldsymbol{\sigma}_{k-L}; \mathbf{a}_{k-1}, \ldots, \mathbf{a}_{k-L} \right) \\ &= \left(a_{k-1}^{(1)}, \cdots, a_{k-\nu_1-L}^{(1)}; \cdots; a_{k-1}^{(m)}, \cdots, a_{k-\nu_m-L}^{(m)}; \right. \\ &\qquad \left. a_{k-1}^{(m+1)}, \cdots, a_{k-L}^{(m+1)}; \cdots; a_{k-1}^{(n)}, \cdots, a_{k-L}^{(n)} \right) \ . \tag{8.89} \end{aligned}$$

The overall system also has state and trellis diagram descriptions. With MLSE, the Viterbi algorithm searches for the most likely path in the **super-trellis** based

on the sequence of received samples $\mathbf{v}$ at the output of the matched filter. An error event of length $\ell = k_2 - k_1$ occurs between epochs k_1 and k_2 in the super-trellis if the actual super-state $\mathbf{s}_k$ and the estimated super-state $\hat{\mathbf{s}}_k$ satisfy

$$s_{k_1} = \hat{s}_{k_1}, \quad s_{k_2} = \hat{s}_{k_2} \quad \text{and} \quad s_k \neq \hat{s}_k, \quad k_1 < k < k_2 . \tag{8.90}$$

Associated with every error event in the super-trellis is a pairwise error probability $P(\mathbf{x} \rightarrow \hat{\mathbf{x}})$, representing the probability that the receiver chooses sequence $\hat{\mathbf{x}}$ when $\mathbf{x}$ is transmitted.

6.1   TCM ON STATIC ISI CHANNELS

As shown in Chapter 7.6.1, the pairwise error probability is a function of the path distance matrix E defined in (7.160). The matrix E depends only on the trellis code and the length of the channel $L + 1$. Equations (7.163) and (eigenbound) provide a guideline for designing trellis codes for static ISI channels. The design criterion should be to maximize $\lambda_{\min}$ for the dominant error event. If the dominant error event has length ℓ, then this criterion implies that $\min\{r_\ell(0)\}$ is maximized, where $r_\ell(0)$ is the squared Euclidean distance between two allowable sequences of length ℓ in the super-trellis. The other design criterion should be to minimize the condition number $c(E)$. The matrix $\mathbf{E}$ is perfectly conditioned, or $c(E) = 1$, only when $r_\ell(i) = 0, i \neq 0$. In this case $\mathbf{E}$ has the form $r_\ell(0)\mathbf{I}_{L+1}$, where $\mathbf{I}$ is the identity matrix. AWGN channels of the form $\mathbf{g} = \mathbf{e}_i$, $i \in \{0, \ldots, L\}$ represent the eigenvectors of this matrix.

If the dominant error event of a trellis code has a perfectly conditioned path distance matrix $\mathbf{E}$, then the asymptotic performance of the code is the same for any channel vector g, including the AWGN channel. An interesting phenomenon occurs when the dominant error event associated with a trellis code does not have a perfectly conditioned path distance matrix. In this case the asymptotic performance of the trellis code over the channel described by the vector $\mathbf{g} = \mathbf{v}_{\max}$ is better than its asymptotic performance on an AWGN channel! The code has the worst performance on the channel described by the vector $\mathbf{g} = \mathbf{v}_{\min}$.

6.2   TCM ON NONINTERLEAVED FADING ISI
CHANNELS

As mentioned before, $c(E) = 1$ if and only if $r_\ell(i) = 0$, $i \neq 0$. It is impossible to obtain a code where $c(E) = 1$ for all error events. The cross terms $r_\ell(i), i \neq 0$ in the path distance matrix $\mathbf{E}$ of any error event will degrade the performance. The next section shows an example where the cross terms cause an error event with a smaller $r_\ell(0)$ to have a lower pairwise error probability than another error event having a larger $r_\ell(0)$. It is very difficult to control the cross terms of the dominant error events and, therefore, a less stringent criterion would be useful for predicting the performance of a trellis

code. We now show that if the squared Euclidean distance $\min\{r_\ell(0)\}$ does not increase linearly with ℓ, then a trellis code will not have good performance on a non-interleaved fading ISI channel. We require the following definition and two properties to develop this criterion.

Definition: Define $\Omega(k_1, k_1 + i)$ as the set of all distinct pairs of sequences in the code-trellis that originate from the same state at epoch k_1 and merge into the same state at epoch $k_1 + i$. Each pair of sequences in set $\Omega(k_1, k_1 + i)$ may also merge into the same state between epochs k_1 and $k_1 + i$ and possibly stay merged for at most $L - 1$ branches in the code-trellis and then diverge, thus forming one or more error events in the code-trellis.

Property 8.1: If an error event of length i having a squared Euclidean distance d^2 occurs in the code-trellis, then an error event of length $\ell = L + i$ having a squared Euclidean distance $r_{L+i}(0) = d^2$ occurs in the super-trellis. Conversely, if an error event occurs in the super-trellis between epochs k_1 and k_2, having length $k_2 - k_1 = L + i$ and a corresponding squared path distance $r_{L+i}(0)$, then there exists a pair of sequences $(\mathbf{x}_1, \mathbf{x}_2) \in \Omega(k_1, k_1 + i)$ in the code-trellis having a squared Euclidean distance equal to $r_{L+i}(0)$.

Proof: Using the definition of the error event in the super-trellis in (8.90), and using the two equivalent forms of the super-state (8.88) and (8.89), it follows directly that an error event in the code-trellis $\sigma_{k_1} = \hat{\sigma}_{k_1}$, $\sigma_{k_1+i} = \hat{\sigma}_{k_1+i}$, and $\sigma_k \neq \hat{\sigma}_k$ for $k_1 < k < k_1 + i$ implies that $s_{k_1} = \hat{s}_{k_1}$, $s_{k_1+i+L} = \hat{s}_{k_1+i+L}$, and $s_k \neq \hat{s}_k$ for $k_1 < k < k_1 + i + L$. It also follows that $\epsilon_k = 0$, $k \geq k_1 + i$ and, hence, $r_{L+i}(0) = d^2$.

Conversely, suppose that an error event occurs in the super-trellis between epochs k_1 and $k_1 + L + i$. It follows directly from (8.90), and the fact that $x_k = f_1(\sigma_k, \mathbf{a}_k)$ and $\sigma_{k+1} = f_2(\sigma_k, \mathbf{a}_k)$, that there exists a pair of sequences in the code-trellis $(\mathbf{x}_1, \mathbf{x}_2)$ that originate from the same state at epoch k_1, i.e., $\epsilon_{k_1} \neq 0$, and merge into the same state at epoch $k_1 + i$, i.e., $\epsilon_{k_1+i-1} \neq 0$. The two sequences may merge into the same state between epochs k_1 and $k_1 + i$, say at epoch $k_1 + m$ ($m < i$) and possibly stay merged for at most $L - 1$ branches; otherwise, $s_{k_1+L+m} = \hat{s}_{k_1+L+m}$ and the length of error event is shorter than $L + i$. Hence, $(\mathbf{x}_1, \mathbf{x}_2) \in \Omega(k_1, k_1 + i)$. It also follows from (7.162) that the squared Euclidean distance between $\mathbf{x}_1$ and $\mathbf{x}_2$ is equal to $r_{L+i}(0)$.

Property 8.2: Let $d_{\min}^2(i)$ denote the squared minimum distance of all error events in the code-trellis of length i. Then $\min\{r_{L+i}(0)\} \leq d_{\min}^2(i)$.

Proof: An error event in the super-trellis between epochs k_1 and $k_1 + L + i$ and the corresponding pair of sequences $(\mathbf{x}_1, \mathbf{x}_2)$ may result from either a single error event in the code-trellis between epochs k_1 and $k_1 + i$, or multiple error events of shorter lengths, e.g., i error events of length one (parallel transitions) or, in general, m error events of lengths $j_1, \ldots, j_m$. Note that m can take any value between one and i. Also, for any m, the lengths $j_1, \ldots, j_m$ can assume different values with the constraint $\sum j_m \leq i$ and $(\mathbf{x}_1, \mathbf{x}_2) \in \Omega(k_1, k_1 + i)$. It

follows that

$$\min\{r_{L+i}(0)\} = \min\left\{ \sum d^2_{\min}(j_m) \right\} \le d^2_{\min}(i) \qquad (8.91)$$

where the minimization goes over all $m \le i$ and all $j_1, \ldots, j_m$ with the above constraint. If there exists a single error event of length i in the code-trellis, then the theorem is immediate. Also, if there are no error events of length i in the code-trellis then, by definition, $d_{\min}(i) = \infty$ and the above inequality in (8.91) is satisfied.

If we treat the uncoded system as a trellis-coded system with a single state and parallel transitions, then, for two-tap channels $\min\{r_{i+L}(0)\} = i d^2_{\min}(0)$, $i = 1, 2, \ldots$, where $d^2_{\min}(0)$ is the minimum squared Euclidean distance in the signal constellation of the uncoded system. The important point is that $r_{L+i}(0)$ for the uncoded system grows linearly with the length of the error event. Properties 8.2 and 8.2 suggest that if a trellis code has a dense distance spectrum [284], then the set of $\min\{r_\ell(0)\}$, $\ell \ge L$ will also have a dense spectrum. Therefore, if the minimum squared Euclidean distance of a trellis code does not grow linearly with the length of the error events, then the uncoded system is expected to outperform the trellis-coded system for a non-interleaved fading ISI channel.

6.3   EXAMPLES

This section illustrates the above concepts by focusing on the 4-state 8-PSK and the 8-state 8-PSK Ungerboeck codes [330]. The corresponding uncoded system is 4-PSK. We have seen earlier that the 8-state code is suitable for interleaved flat fading channels, having MTD = 2 and MPSD = 8. Two ISI channels are considered i) a two-tap, T-space, static ISI channel and, ii) a two-tap, T-spaced, Rayleigh fading ISI channel.

6.3.1   STATIC ISI CHANNELS

For a two-tap static ISI channel, the path distance matrix is

$$\mathbf{E} = \begin{pmatrix} r_\ell(0) & r_\ell(1) \\ r_\ell^*(1) & r_\ell(0) \end{pmatrix} . \qquad (8.92)$$

Tables 8.1, 8.2, and 8.3, tabulate the values associated with matrix $\mathbf{E}$ for all error events of up to length 8 for the uncoded system and the two coded systems. Notice that the minimum squared Euclidean distance is $r_2(0) = 2.00$ for the uncoded system, $r_2(0) = 4.00$ for the 4-state trellis code, and $r_4(0) = r_5(0) = 4.59$ for the 8-state trellis code. Also, the matrix $\mathbf{E}$ associated with the minimum distance error event is perfectly conditioned for both the uncoded system and the 4-state trellis code, but not for the 8-state trellis code.

We now consider the coded performance for seven different channels with impulse responses listed in Table 8.4. Channels G, A1, and A2 were chosen

ℓ	$\min\{r_\ell(0)\}$	$|r_\ell(1)|$	λ_1	λ_2	$w(\mathbf{x}, \hat{\mathbf{x}})$	$P(\mathbf{x} \to \hat{\mathbf{x}})$
2	2.00	0.00	2.00	2.00	1	0.1705E-01
3	4.00	2.00	6.00	2.00	2	0.6786E-02
4	6.00	4.00	10.00	2.00	3	0.4238E-02
5	8.00	6.00	14.00	2.00	4	0.3081E-02
6	10.00	8.00	18.00	2.00	5	0.2420E-02
7	12.00	10.00	22.00	2.00	6	0.1993E-02
8	14.00	12.00	26.00	2.00	7	0.1694E-02

Table 8.1. Error events in the super-trellis for uncoded 4-PSK system over a two-tap channel

ℓ	$\min\{r_\ell(0)\}$	$|r_\ell(1)|$	λ_1	λ_2	$w(\mathbf{x}, \hat{\mathbf{x}})$	$P(\mathbf{x} \to \hat{\mathbf{x}})$
2	4.00	0.00	4.00	4.00	1	0.5528E-02
3	8.00	4.00	12.00	4.00	2	0.2033E-02
4	4.59	2.16	6.75	2.42	2	0.5261E-02
5	5.17	2.72	7.89	2.46	4	0.3831E-02
6	5.17	2.16	7.34	3.01	3	0.3790E-02
7	5.76	2.72	8.47	3.04	5	0.3173E-02
8	5.76	2.00	7.76	3.76	3	0.3227E-02

Table 8.2. Error events in the super-trellis for 4-state 8-PSK code over a two-tap channel

ℓ	$\min\{r_\ell(0)\}$	$|r_\ell(1)|$	λ_1	λ_2	$w(\mathbf{x}, \hat{\mathbf{x}})$	$P(\mathbf{x} \to \hat{\mathbf{x}})$
3	6.00	2.83	8.83	3.17	1	0.3301E-02
4	4.59	2.16	6.75	2.42	3	0.5106E-02
5	4.59	1.08	5.67	3.50	3	0.4556E-02
6	5.17	2.16	7.34	3.01	6	0.4094E-02
7	5.17	1.08	6.25	4.09	6	0.3654E-02
8	5.17	1.08	6.25	4.09	4	0.3654E-02

Table 8.3. Error events in the super-trellis for 8-state 8-PSK code over a two-tap channel

arbitrarily and have the best spectral characteristics. Channels B1 and B2 are equal to the eigenvectors associated with the minimum and maximum eigenvalues, respectively, for one of the length 3 and 4 error events in the uncoded system, and one of the error events of length 4 in the 8-state trellis code. Channels C1 and C2 are equal to the eigenvectors associated with the maximum and minimum eigenvalues, respectively, for some of the minimum distance error events that are associated with the 8-state trellis code.

Channel	g_0	g_1
G	0.9747	0.223
A1	0.866	0.500
A2	0.500	0.866
B1	-0.7071	0.7071
B2	0.7071	0.7071
C1	0.7071	-0.65333+j0.2705
C2	0.7071	0.65333-j0.2705

Table 8.4. Static ISI channels used to evaluate the performance of trellis codes.

Simulation results for the uncoded 4-PSK system, the 4-state trellis code, and the 8-state trellis code are shown in Figs. 8.23, 8.24, and 8.25, respectively. Although channel C2 has an in-band spectral null, making it perhaps the most difficult channel to equalize, the performance of uncoded 4-PSK on this channel is better than that on channels B1 and B2. The 4-state trellis code performs better on channels B1 and B2 than on channels C1 and C2, although channel B1 has a band-edge null while channel B1 does not. The 8-state trellis code performs better on channel C1 than C2. This makes sense because channel C1 is the eigenvector associated with λ_{max} for one of the dominant error events of the 8-state code. In general, Tables 8.1, 8.2, and 8.3 show that the coded systems have a larger λ_{min} than the uncoded systems. Therefore, it is reasonable that the coded systems have better performance than the uncoded system, although $c(\mathbf{E})$ for the dominant error event is greater for the 8-state code.

6.3.2 MULTIPATH FADING ISI CHANNELS

Consider a 2-tap fading ISI channel with matrix

$$\mathbf{A} = \Sigma_d \mathbf{E} \Sigma_d = \begin{pmatrix} \sigma_0^2 r_\ell(0) & \sigma_0 \sigma_1 r_\ell(1) \\ \sigma_0 \sigma_1 r_\ell^*(1) & \sigma_1^2 r_\ell(0) \end{pmatrix} \tag{8.93}$$

and corresponding eigenvalues

$$\lambda_{1,2} = r_\ell(0)\frac{\bar{E}}{2} \pm \frac{1}{2}\sqrt{r_\ell(0)^2 \left(\sigma_0^2 - \sigma_1^2\right)^2 + 4\sigma_0^2 \sigma_1^2 |r_\ell(1)|^2} \tag{8.94}$$

where $\bar{E} = \sum_{i=0}^{L} \sigma_{i\,d}^2$. It is obvious that $\lambda_1 = \lambda_2$ implies that $\sigma_0 = \sigma_1$ and $r_\ell(1) = 0$. For the case when $\sigma_0 = \sigma_1$, the eigenvalues are given by $\frac{\bar{E}}{2}[r_\ell(0) \pm |r_\ell(1)|]$.

Fig 8.26 plots the pairwise error probability of the 8-state code against the normalized energy in the first tap for different values of $r_\ell(0)$ and $|r_\ell(1)|$ as described in Table 8.3. Notice that the pairwise error probability is minimized for equal energy taps. Fig. 8.26 also shows how the pairwise error probability
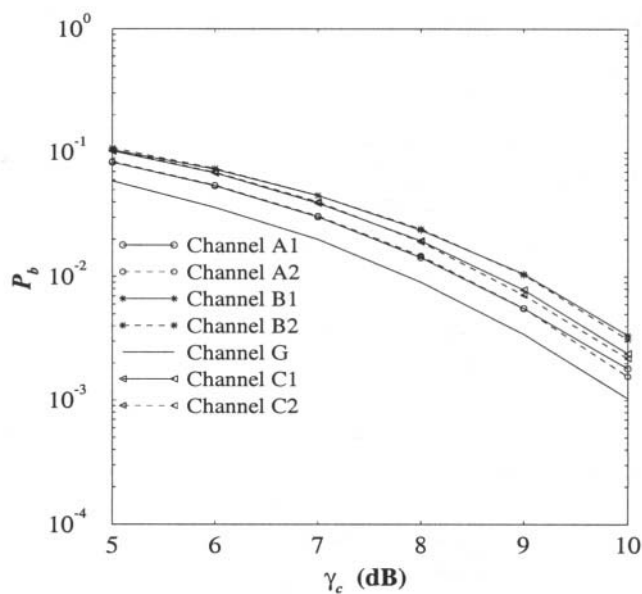
Figure 8.23.   Performance of uncoded 4-PSK over static ISI channels.
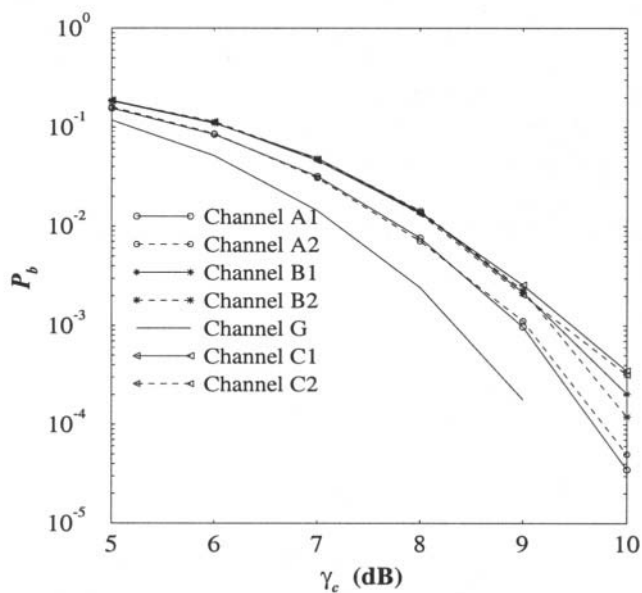
Figure 8.24.   Performance of 4-state 8-PSK TCM over static ISI channels.

decreases with an increase in the squared Euclidean distance, $r_\ell(0)$. Finally, Fig. 8.26 shows the effect of the eigenvalue spread. For example, the pairwise
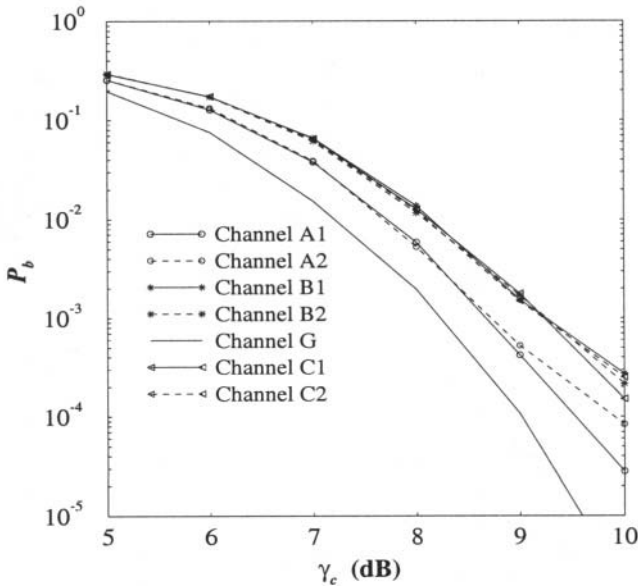
Figure 8.25.   Performance of 8-state 8-PSK TCM over static ISI channels.

error probability associated with the error event having $r_6(0) = 5.76$ and $|r_6(1)| = 1.08$ ($c = 1.46$) is better than that with $r_3(0) = 6.00$ and $|r_3(1)| = 2.83$ ($c = 2.79$), although the squared Euclidean distance is larger for the latter.

Fig. 8.27 plots the bit error probability of the uncoded 4-PSK system, and the 4- and 8-state 8-PSK trellis codes on a two-tap, T-spaced, fading channel with $\sigma_0 = \sigma_1$. The performance is completely reversed from that on an AWGN channel. The uncoded system outperforms either trellis-coded system and, moreover, the 4-state trellis code outperforms the 8-state trellis code. This behavior is consistent with the parameters listed in Tables 8.1, 8.2 and 8.3. Although the uncoded system has a smaller squared Euclidean distance, $r_\ell(0)$ for the uncoded systems grows faster with the length of the error events than either the 4- or 8-state trellis codes. By comparing the parameters of the 4- and 8-state trellis codes in Tables 8.2 and 8.3, respectively, it is not surprising that the 4-state trellis code outperforms the 8-state trellis code.

As a final example, consider the simple rate-1/2, 2-state, 4-PSK trellis code shown in Fig 8.28. The equivalent uncoded system is BPSK having a minimum squared Euclidean distance growth given by the values $\{d_{min}^2(i)/\bar{E}\} = \{4.0, 8.0, 12.0, 16.0, 20.0, 24.0, 28.0, \dots\}$. Table 8.5 lists the parameters of the code. Note that the code has a minimum squared Euclidean distance that grows linearly with the length of the error event but at a slower rate than that of the uncoded system. Fig. 8.29 shows the performance of the code. Unlike
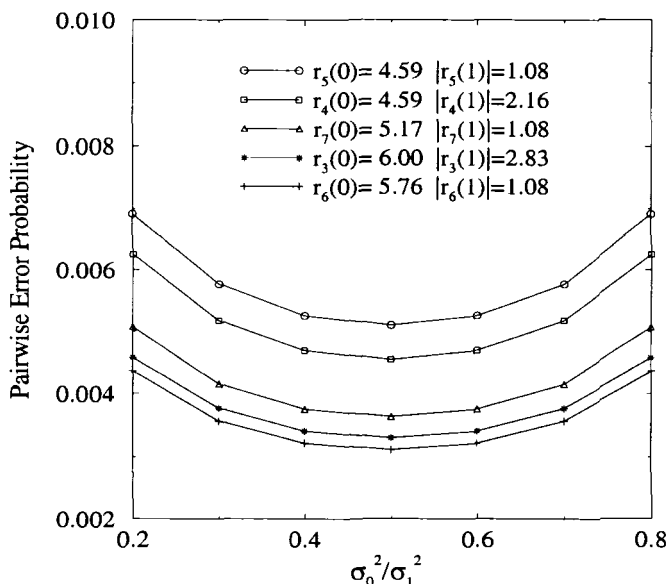
Figure 8.26. Pairwise error probability for several error events for the 8-state 8-PSK trellis code at $\bar{\gamma}_c = 10$ dB. See Table 8.3 for corresponding eigenvalues.

the Ungerboeck codes, the code at least offers slightly better performance than the equivalent uncoded system despite its simplicity.

6.4   EVALUATION OF UNION BOUNDS FOR TCM

The pair-state approach is one method for evaluating the error probability upper bound for TCM on intersymbol interference (ISI) channels [85, 37]. Unfortunately, there are $(N_S)^2$ pair-states, where N_S is the number of super-states. A simpler method that uses the transfer function of an $.N_S$-state error diagram has been proposed for linear filter channels [204]. Both of these techniques require a Chernoff bound on the pairwise error probability which can be loose, especially for fading channels.[2]   Here we describe a method for evaluating the union bound that uses an error-state diagram and a one-directional stack algorithm. The proposed method does not require the transfer function and, therefore, i) an exact expression for the pairwise error probability can be used yielding a tighter upper bound, and ii) the method is useful for large-state systems.

---

[2]The union bound may also be loose for fading channels.
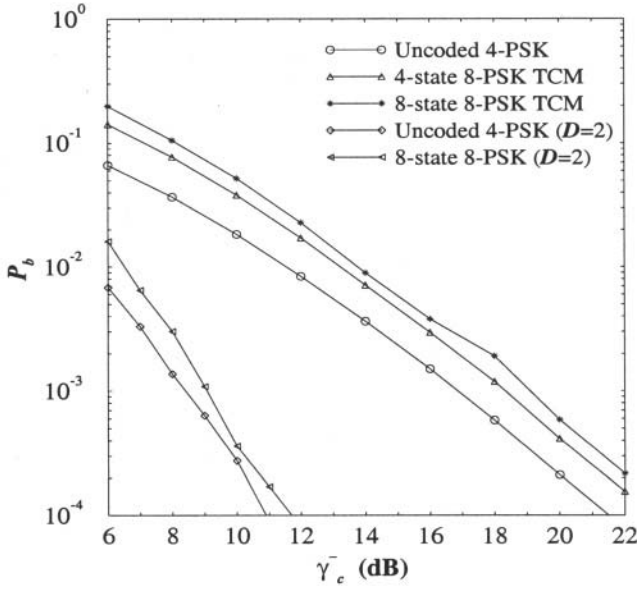
Figure 8.27.   Performance over a two-tap, T-spaced, fading ISI channel; $\sigma_0 = \sigma_1$.

The bit error probability for TCM on an ISI channel has the bound in (8.43), where $\mathbf{E}$ is the set of error sequences that correspond to all error events in the super-trellis at epoch j_1. For a static ISI channel, the pairwise error probability is given by (7.154) with the squared Euclidean path distance in (7.157). By following the same procedure as in Chapter 7.6.3, the parameters needed to evaluate the upper bound for a static ISI channel are the probability $P\{\mathbf{a}\}$, the number of bit errors $w_b(\mathbf{e})$, and the squared Euclidean path distance Δ^2 Likewise, for a fading ISI channel with equal diversity branches, the exact pairwise error probability is given by (7.181), and the parameters needed to evaluate the union bound are $P\{\mathbf{a}\}$, $w_b(\mathbf{e})$, and the matrix $\mathbf{A} \equiv \mathbf{A}_d$, where $\mathbf{A}_d$ is defined in (7.172).

The overall system has super states $\mathbf{s}_k^{(i)}$ for $i = 0, \cdots, N_S - 1$, where $\mathbf{s}_k^{(i)}$ is defined in (8.89). Define the **error state** as $\mathbf{v}_l = \mathbf{s}_k^{(i)} \oplus \mathbf{s}_k^{(j)}$ for some i and j. An error-state diagram can be constructed such that the initial and final nodes in the error-state diagram are zero-error states and each intermediate node represents a distinct non-zero error-state. A directed line from $\mathbf{v}_l$ to $\mathbf{v}_m$ indicates an allowable error-state transition $(\mathbf{v}_l, \mathbf{v}_m)$. There is one-to-one correspondence between the set $\mathbf{E}$ in (8.43) and the set of paths from the initial to final node in the error-state diagram.
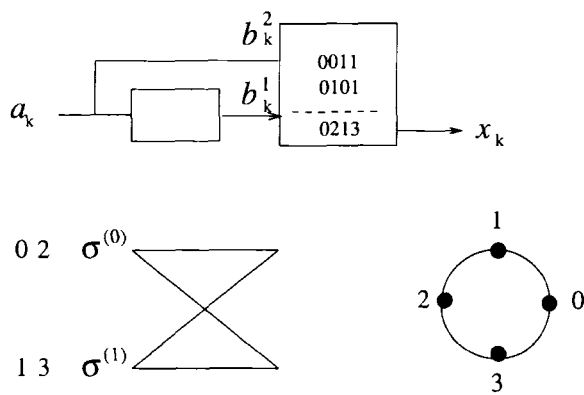
Figure 8.28. Simple rate-1/2, 2-state, 4-PSK trellis code.

i	$d_{\min}^2(i)/\bar{E}$	$w(\mathbf{x}, \hat{\mathbf{x}})$	N_{eq}
1	∞	0	0
2	6.0	1	1
3	8.0	2	1
4	10.0	3	1
5	12.0	4	1
6	14.0	5	1
7	16.0	6	1
8	18.0	7	1

Table 8.5. Error events in the code-trellis for the rate-1/2, 2-state, 4-PSK code in Fig. 8.28; N_{eq} is the number of error events of length i having a squared Euclidean distance of $d_{\min}^2(i)$.

To evaluate the union bound (8.43), we define an appropriate transition-gain for each transition in the error-state diagram as follows. Given an error-state transition $(\mathbf{v}_l, \mathbf{v}_m)$, each branch from $\mathbf{s}_k^{(i)}$ to $\mathbf{s}_{k+1}^{(j)}$ in the overall trellis diagram is assigned the appropriate branch distance (or branch distance matrix), $\mathbf{C}_{lm}(i,j)$ and number of bit errors $u_{lm}(i,j)$. This assignment can be conveniently described by an $N_S \times N_S$ transition matrix $\mathbf{B}_{lm} = [b_{lm}(i,j)]$ with elements

$$b_{lm}(i,j) \triangleq \begin{cases} Z_1^{u_{lm}} Z_2^{\mathbf{C}_{lm}(i,j)} & , \ \mathbf{s}_k^{(i)} \text{ to } \mathbf{s}_{k+1}^{(j)} \text{ transition possible} \\ 0 & , \ \text{otherwise} \end{cases} \quad (8.95)$$

Note that u_{lm} in (8.95) does not depend on the branch from $\mathbf{s}_k^{(i)}$ to $\mathbf{s}_{k+1}^{(j)}$, but only on the error-state transition $(\mathbf{v}_l, \mathbf{v}_m)$.

Consider the following simple example, consisting of a two-state 4-PAM trellis code with a two-tap channel $(L = 1)$. The encoder has generators $\mathbf{g}^{(1)} = (1, 0)$ and $\mathbf{g}^{(2)} = (0, 1)$, and the signal mapping is $x_k = 4a_k + 2a_{k-1} - 3$.
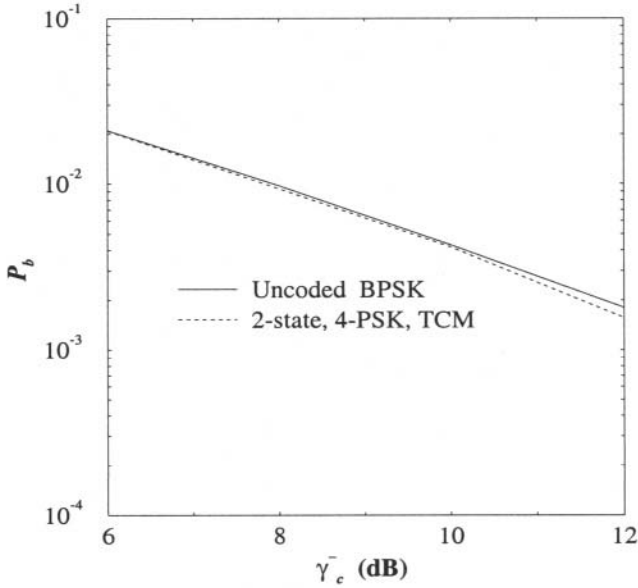
Figure 8.29. Performance of the rate-1/2, 2-state, 4-PSK trellis code on a two-equal-ray T-spaced fading ISI channel.

Fig. 8.30 shows the error-state diagram. As an example of how to obtain $\mathbf{B}_{lm}$, consider the error sequence $\{e_k = 1, \ e_{k+1} = 1, \ e_{k+2} = 1, \ e_{k+3} = 0, \ e_{k+4} = 0\}$ in Fig. 8.31. The error sequence corresponds to the path $\{\mathbf{v}_0, \ \mathbf{v}_2, \ \mathbf{v}_3, \ \mathbf{v}_3, \ \mathbf{v}_1, \ \mathbf{v}_4\}$ in the error-state diagram. Fig. 8.31 shows the super-trellis, along with the symbol error ε_k for $k_1 \leq k \leq k_1 + 4$. Note that all branches merging at the same node in Fig. 8.31 have the same symbol error ε_k. Given the pair $(\varepsilon_{k-1}, \ \varepsilon_k)$, the squared branch distance Δ_k^2 can be calculated by using (7.158) and (7.159), and the matrix $\mathbf{A}_k$ in (7.169) can be calculated from the branch distance matrix $\mathbf{E}_k$ in (7.159)[3]. For example, consider a static ISI channel with $g_0 = 1$ and $g_1 = 1$. The transition-gains $\mathbf{B}_{02}$ and $\mathbf{B}_{33}$ are, respectively,

$$
\mathbf{B}_{02} = Z_1 \cdot
\begin{bmatrix}
Z_2^{16} & 0 & Z_2^{16} & 0 \\
Z_2^{16} & 0 & Z_2^{16} & 0 \\
0 & Z_2^{16} & 0 & Z_2^{16} \\
0 & Z_2^{16} & 0 & Z_2^{16}
\end{bmatrix},
\tag{8.96}
$$

---

[3]In general, the squared branch distance and branch distance matrix are calculated using $(\varepsilon_{k-L}, \varepsilon_{k-L+1}, \cdots, \varepsilon_k)$.
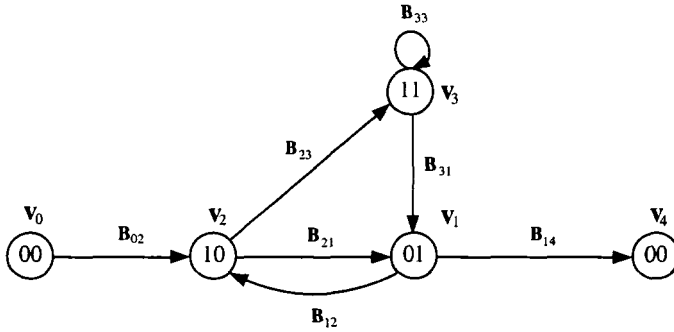
Figure 8.30. Error-state diagram for 2-state, 4-PAM, TCM on a two-tap ISI channel.

and

$$\mathbf{B}_{33} = Z_1 \cdot \begin{bmatrix} Z_2^{144} & 0 & Z_2^{16} & 0 \\ Z_2^{64} & 0 & Z_2^{0} & 0 \\ 0 & Z_2^{0} & 0 & Z_2^{64} \\ 0 & Z_2^{16} & 0 & Z_2^{144} \end{bmatrix} . \tag{8.97}$$

Likewise, for a two-tap fading ISI channel, the transition gain $\mathbf{B}_{02}$ is

$$\mathbf{B}_{02} = Z_1 \cdot \begin{bmatrix} Z_2^{\mathbf{A}_{02}} & 0 & Z_2^{\mathbf{A}_{02}} & 0 \\ Z_2^{\mathbf{A}_{02}} & 0 & Z_2^{\mathbf{A}_{02}} & 0 \\ 0 & Z_2^{\mathbf{A}_{02}} & 0 & Z_2^{\mathbf{A}_{02}} \\ 0 & Z_2^{\mathbf{A}_{02}} & 0 & Z_2^{\mathbf{A}_{02}} \end{bmatrix} \tag{8.98}$$

where

$$\mathbf{A}_{02} = \Sigma \begin{bmatrix} 16 & 0 \\ 0 & 0 \end{bmatrix} \Sigma . \tag{8.99}$$

For the error-state transition $(\mathbf{v}_l, \mathbf{v}_m)$, the **distance polynomial** [284] associated with the node $\mathbf{s}_k^{(i)}$ is the sum of the ith-row elements of the matrix $1/2^n \cdot \mathbf{B}_{lm}$, where we have assumed that the information vector $\mathbf{a}_k$ is transmitted with equal probability $1/2^n$. If the sum of row elements is the same for every row, then the matrix is **row-uniform** [204]. If the matrix $\mathbf{B}_{lm}$ is row-uniform for all error-state transitions, then the trellis code has the **uniform property.** For AWGN channels, many trellis codes including the Ungerboeck codes have the uniform property, meaning that the error probability can be evaluated by just considering the set of information sequences that originate from a particular state, say $\mathbf{s}_{k_1}^{(0)}$. However, for ISI channels the uniform property is lost and all possible information sequences must be considered. For example, the matrix $\mathbf{B}_{33}$ does not have the row-uniform property in our example system, although the trellis-code is uniform for AWGN channels. Nevertheless, some symmetry properties of the trellis code and trellis structure can still be
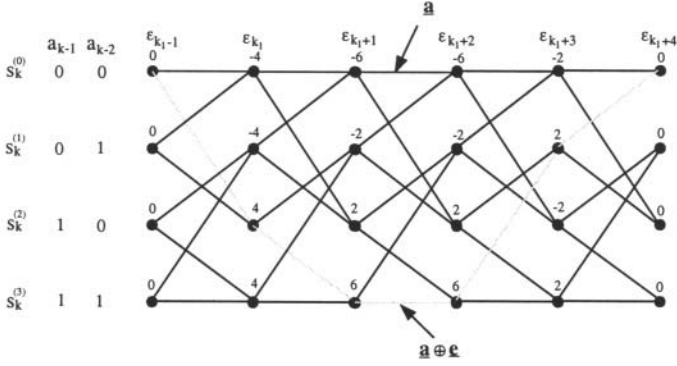
Figure 8.31. Super-trellis showing symbol errors for the error sequence $\{e_{k_1} = 1,\ e_{k_1+1} = 1,\ e_{k_1+2} = 1,\ e_{k_1+3} = 0,\ e_{k_1+4} = 0\}$.

exploited to simplify the performance analysis. Considering again the error sequence $\{e_k = 1,\ e_{k+1} = 1,\ e_{k+2} = 1,\ e_{k+3} = 0,\ e_{k+4} = 0\}$ in Fig. 8.31, the following observations can be made:

Observation 8.1:. Only half of the sequences originating from each state $s_{k_1}^{(i)}$ must be considered, because there is always a pair of correct sequences having the same probability P{a}, number of bit errors $w_b(\mathbf{e})$, and path distance (or path distance matrix), e.g., the information sequences $\mathbf{a}$ and $\mathbf{a} \oplus \mathbf{e}$. This symmetry property exists for every TCM system.

Observation 8.2:. Only the information sequences originating from states $s_{k_1}^{(0)}$ and $s_{k_1}^{(2)}$ must be considered, because for each information sequence originating from the state $s_{k_1}^{(1)}$ (or $s_{k_1}^{(3)}$) there always exists an information sequence originating from the state $s_{k_1}^{(0)}$ (or $s_{k_1}^{(2)}$) having the same set of parameters. This type of symmetry usually exists but depends on the particular trellis code. The algorithm discussed below exploits this type of symmetry by combining together all paths attached to the same node that have the same set of parameters.

The stack algorithm maintains an ordered stack where each entry represents one or more paths in the error-state diagram and contains the following information: terminal node, terminal state s_k, the number of branches H, $\sum_{(l,m)} u_{lm}$, $\sum_{(l,m)} \Delta_{lm}^2$ (or $\sum_{(l,m)} \mathbf{A}_{lm}$), and the *intermediate* bit error probability P_I. The set $\{(l, m)\}$ is the set of error-state transitions associated with the path under consideration. P_I is calculated according to

$$P_I = \frac{1}{n} \cdot \frac{1}{N_S \cdot 2^{nH}} \sum_{\{(l,m)\}} u_{lm} \cdot \hat{P}\left\{ \Gamma(\mathbf{a} \oplus \mathbf{e}) \geq \Gamma(\mathbf{a}_k) \middle| \mathbf{a} \right\} \qquad (8.100)$$

Initialization
r=0,sum=0,R,P_t

Top entry terminates at the final node ? — yes

sum=sum+M•P_t
r=r+M

no

Extend the top entry

For each extended path

P_t is in the stack ? — yes

Any entry with the same parameters ? — no

yes

Add the entry multiplicity

no

Insert the new path & reorder the stack

r > R or P_t < P_T — no

yes

Stop

Delete the top entry
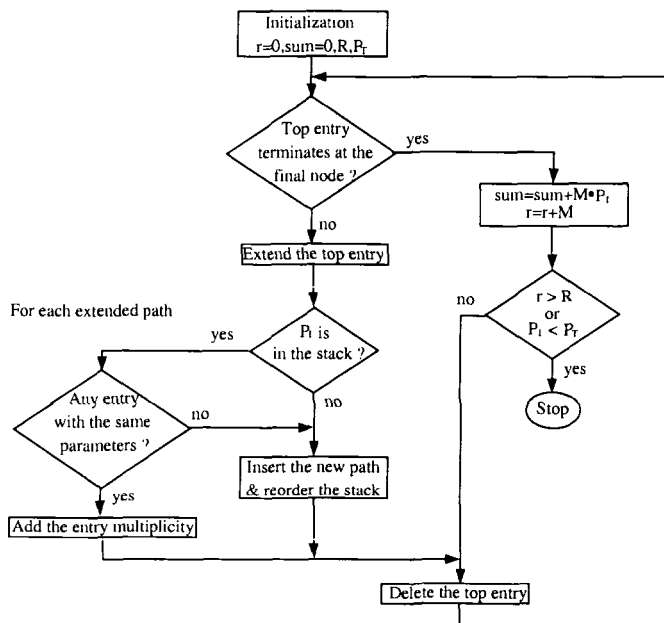
Figure 8.32. The stack algorithm.

where $\hat{P}\{\Gamma(\mathbf{a}\oplus\mathbf{e})\geq\Gamma(\mathbf{a})\,|\mathbf{a}\}$ is computed by using the squared path distance $\sum_{\{(l,m)\}}\Delta^2_{lm}$ for a static ISI channel and the matrix $\sum_{\{(l,m)\}}\mathbf{A}_{lm}$ for a fading ISI channel. The stack is ordered according to decreasing P_t. For a path from the initial to final node, we have $P\{\mathbf{a}\}=1/(N_S\cdot 2^{nH})$, $w_b(\mathbf{e})=\sum_{\{(l,m)\}}u_{lm}$, and $P\{\Gamma(\mathbf{a}\oplus\mathbf{e})\geq\Gamma(\mathbf{a})\,|\mathbf{a}\}=\hat{P}\{\Gamma(\mathbf{a}\oplus\mathbf{e})\geq\Gamma(\mathbf{a})\,|\mathbf{a}\}$. A key feature of the stack algorithm is that paths having the same terminal node, terminal state $\mathbf{s}_k$, number of branches $.H$, $\sum_{\{(l,m)\}}u_{lm}$, and $\sum_{\{(l,m)\}}\Delta^2_{lm}$ (or $\sum_{\{(l,m)\}}\mathbf{A}_{lm}$) are combined together, because from that point on they can be treated as a single path. These combinations reduce the computation required to evaluate the upper bound as discussed in Observation 8.2. The number of paths represented by a stack entry is called the path multiplicity, M. The detailed stack algorithm is shown in Fig. 8.32 an operates much the same as the stack algorithm described in Section 6.6.3.

Example 8.4

Consider a system that uses the 4-state 8-PSK Ungerboeck trellis code in Fig. 8.33 on a two-tap multipath-fading channel. In the simulations, the tap coefficients $\{g_{i,d}\}$ are generated by passing independent complex white Gaussian noise through a digital Butterworth filter having a 3-dB cut-off frequency equal to 0.4 Hz. The transmission rate is assumed to be 2400 symbols/sec and $\sigma^2=\frac{1}{2}E[|g_{i,d}|^2]=1$ for all i and d. Once again, the analytical results are
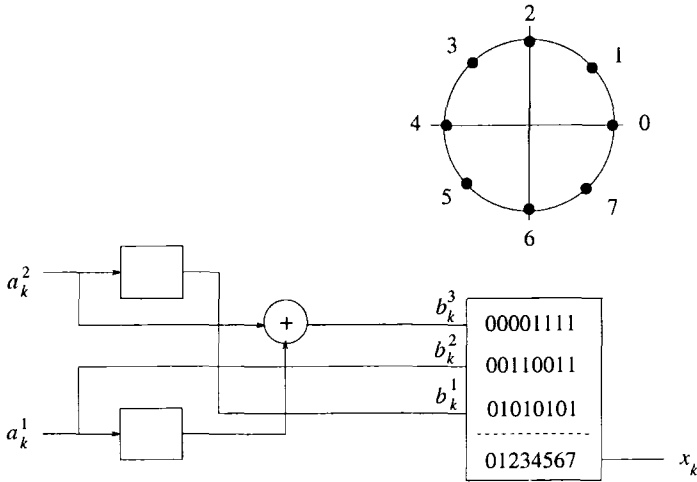
Figure 8.33. Encoder and signal mapping for the 4-state 8-PSK Ungerboeck code.

obtained by setting the threshold $P_T = 10^{-3} \cdot P_{\max}$, where $P_{\max}$ is the largest term in the upper bound. For TCM on multipath-fading channels, the average received bit energy-to-noise ratio per diversity branch is given by (8.87). Fig. 8.34 compares analytical and simulation results for this system. Without diversity ($D = 1$), the difference is about 2 dB for $P_b < 10^{-3}$. However, for two-branch diversity ($D = 2$), the difference is within 1 dB.

7. TURBO CODES

The principle of Turbo coding or concatenated coding is to construct long random-like codes that have a structure that permits practical decoding [26]. Turbo codes are interleaved concatenated codes that are constructed from simple component codes and pseudo-random interleavers. The interleaver makes the code appear random. Since the component codes are easy to decode, the overall code can be decoded by iteratively decoding the component codes. There are two basic types of Turbo codes depending on the type of concatenation, namely parallel concatenated codes and serial concatenated codes. The component codes can be either convolutional codes or block codes. Here we just consider convolutional component codes. Parallel concatenated convolutional codes (PCCCs) use recursive systematic convolutional (RSC) component codes. Serial codes. Serial concatenated convolutional codes (SCCCs) use a recursive or non-recursive convolutional outer code and a recursive convolutional inner code.
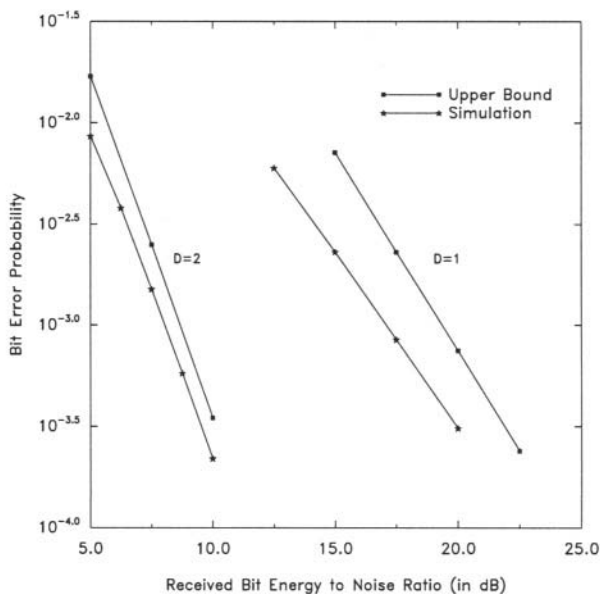
Figure 8.34. Comparison of analytical and simulation results for 4-state, 8-PSK, TCM on a two-equal-ray T-spaced fading ISI channel, from [303].

7.1 PCCC ENCODER

Fig. 8.35 shows a PCCC encoder structure which is a parallel concatenation of two RSC component codes[4]. The component codes must be recursive for reasons we will see later. If the component codes have rates $R_c^{(1)} = k/n_1$ and $R_c^{(2)} = k/n_2$, then the PCCC has code rate

$$R_T = \frac{R_c^{(1)} R_c^{(2)}}{R_c^{(1)} + R_c^{(2)}} = \frac{k}{n_1 + n_2} \ . \tag{8.101}$$

The input data sequence a is first encoded by RSC1. The feedforward and feedback generator polynomials of RSC1 are $\mathbf{g}^{(2)}(D) = g_0^{(2)} + g_1^{(2)} D + \ldots + g_\nu^{(2)} D^\nu$ and $\mathbf{g}^{(1)}(D) = g_0^{(1)} + g_1^{(1)} D + \ldots + g_\nu^{(1)} D^\nu$, respectively, where ν is the encoder memory. The outputs of RSC1 are the systematic component $\mathbf{b}_s = \{b_{s_k}\}$ and the parity component $\mathbf{b}_p^{(1)} = \{b_{p_k}^{(1)}\}$ defined by

$$b_{s_k} = a_k$$

$$b_{p_k}^{(1)} = \sum_{i=0}^{\nu} g_i^{(1)} d_{k-i}$$

---

[4] The parallel concatenation of more than two component codes is possible, but we will consider only two component codes for simplicity.
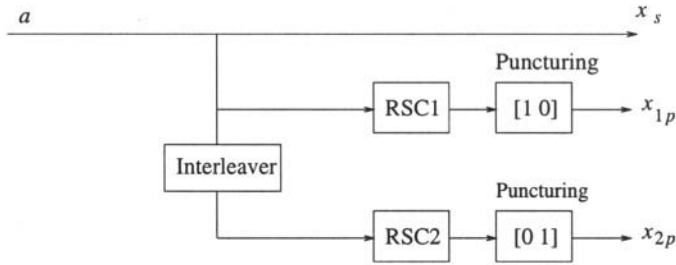
Figure 8.35. PCCC encoder.

where

$$d_k = a_k \oplus \sum_{i=1}^{\nu} g_i^{(2)} d_{k-i} \tag{8.102}$$

The data sequence a is interleaved by the interleaver π of size $.N = kN'$ into the sequence $\tilde{\mathbf{a}}$ and encoded using RSC2 to produce the parity sequence $\mathbf{b}_p^{(2)}$. The interleaving operation can be defined by a mapping $i \rightarrow \pi(i)$ of the input bit position i to output bit position $\pi(i)$. For example, the interleaver might perform the mapping

$$\{0, 1, 2, 3, \ldots, N-1\}_N \rightarrow \{23, 12, 6, 7, \ldots\}_N$$

For Turbo codes the choice of interleaver is crucial. In many cases, **random interleavers** are employed, where the interleaving mapping is completely random. In other cases, an S-random interleaver is used, where interleaver inputs that are separated by less than S positions, $|i - j| < S$, are interleaved into interleaver outputs that are separated by at least S positions, $|\pi(i) - \pi(j)| \geq S$.

A PCCC code word $\mathbf{b} = (\mathbf{b}_s, \mathbf{b}_p^{(1)}, \mathbf{b}_p^{(2)})$ is formed by the parallel concatenation (or interleaving) of the systematic component and the two parity sequences. If higher code rates are desired, then the parity outputs of the RSC component encoders can be punctured. A punctured Turbo code is obtained by starting with a rate-l/n Turbo code and deleting or puncturing specified parity-check symbols. For example, the puncturing pattering in Fig. 8.35 produces a rate-1/2 code from a rate-1/3 code[5]. Note that the systematic component of RSC2 is not transmitted. Tail bits are typically added to the data sequence to terminate RSC1 in the all-zeroes state while the trellis of RSC2 is left 'open'.

---

[5] A "1" in the puncturing pattern means that the bit is transmitted, while a "0" means that the bit is not transmitted

7.2   PCCC DECODER

The Turbo decoder is an iterative structure consisting of many identical stages, each consisting of two soft-input/soft-output SISO decoding units. The decoder structure for PCCCs is shown in Fig. 8.36. The (SISO) modules generate *a-posteriori* probabilities (APPs)

$$P(a_k|\mathbf{y}_s,\mathbf{y}_p^{(1)},\mathbf{y}_p^{(2)}) \tag{8.103}$$

or, for binary codes, log-likelihood ratios (LLRs)

$$L(a_k) = \log \frac{P(a_k = 1|\mathbf{y}_s,\mathbf{y}_p^{(1)},\mathbf{y}_p^{(2)})}{P(a_k = 0|\mathbf{y}_s,\mathbf{y}_p^{(1)},\mathbf{y}_p^{(2)})} \tag{8.104}$$

of each information bit a_k based on the received signal sequence $\mathbf{y}_s,\mathbf{y}_p^{(1)},\mathbf{y}_p^{(2)}$ and the *extrinsic* information passed between the two SISO modules.

The iterative decoding operation of parallel Turbo codes can be explained as follows, using LLRs as an example. At the mth iteration, $m \geq 1$, the LLRs generated by the SISO decoders for data bit a_k are

$$L_1^{(m)}(a_k) = L_{\text{sys}}(a_k) + L_{\text{ext2}}^{(m-1)}(a_k) + L_{\text{ext1}}^{(m)}(a_k) \tag{8.105}$$

$$\tilde{L}_2^{(m)}(a_k) = \tilde{L}_{\text{sys}}(a_k) + \tilde{L}_{\text{ext1}}^{(m)}(a_k) + L_{\text{ext2}}^{(m)}(a_k) \tag{8.106}$$

where $L_{\text{sys}}(a_k) = (2/\sigma^2)y_{s_k}$ is the LLR due to the systematic component, and $L_{\text{ext1}}^{(m)}(a_k)$ and $L_{\text{ext2}}^{(m)}(a_k)$ are the extrinsic information for each bit generated at the mth decoding stage by SISO1 and SISO2, respectively, and can be expressed as

$$L_{\text{ext1}}^{(m)}(a_k) = f(L_{\text{sys}}, L_{\text{ext2}}^{(m-1)}(a_k)) \tag{8.107}$$

$$L_{\text{ext2}}^{(m)}(a_k) = f(L_{\text{sys}}, L_{\text{ext1}}^{(m)}(a_k)) \tag{8.108}$$

where $f(\cdot)$ denotes the SISO decoding unit. The iterative procedure is started with initial condition $L_{\text{ext2}}^{(0)}(a_k) = 0$. The final bit decision for a_k is determined by the sign of $L_2^{(m)}(a_k)$.

A variety of SISO algorithms have been proposed to either generate or approximate the APPs or LLRs. Berrou *et al.* [36] and Robertson [283] used a modified version of the Bahl, Cocke, Jelinek and Raviv (BCJR) algorithm [25] to generate APPs. Hagenauer *et al.* [154], [156] introduced the soft-output Viterbi algorithm (SOVA) [155] to generate soft outputs based on the LLR.

As mentioned previously, Turbo codes can provide near Shannon limit performance. Fig. 8.37 shows the typical performance of a rate-1/2, 16-state, PCCC on an AWGN channel for different random interleaver sizes. Also included, is a 2^{16}-state convolutional code for comparison. Observe that a simple
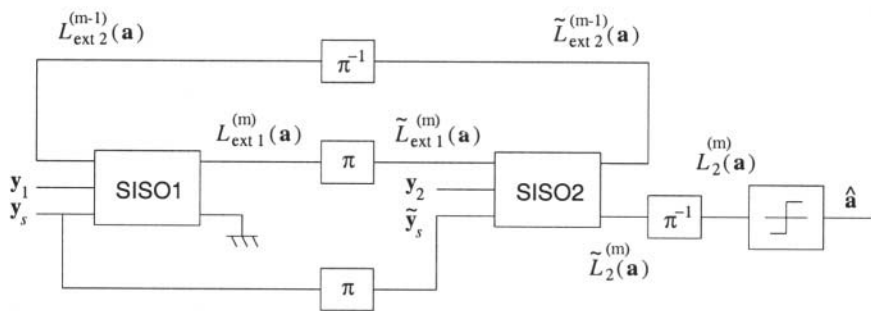
Figure 8.36.   PCCC decoder.
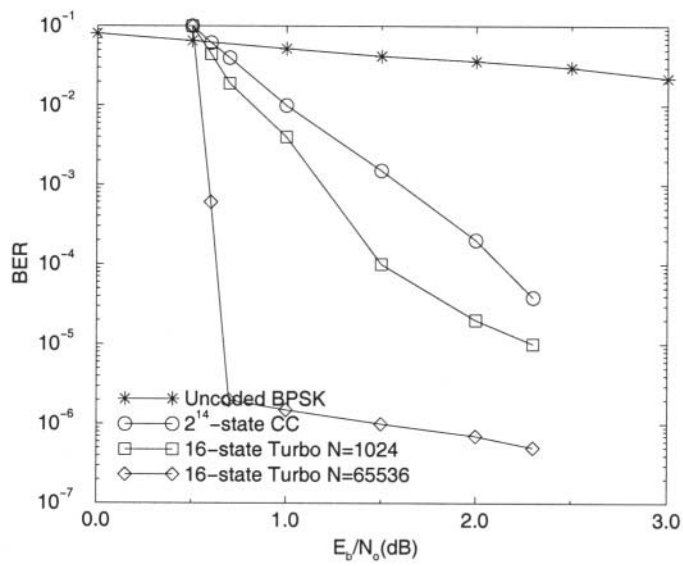
Figure 8.37.   Typical PCCC performance on an AWGN channel.

16-state PCCC can easily outperform a very complex 2^{16}-state convolutional code, at low E_b/N_o. At high E_b/N_o, the BER slope of PCCCs is shallow, loosely termed an "error floor." The error floor is not actually an error floor, but rather a change in the slope of the error rate curve due to the relatively small free distance of Turbo codes.
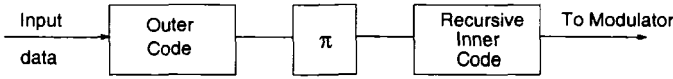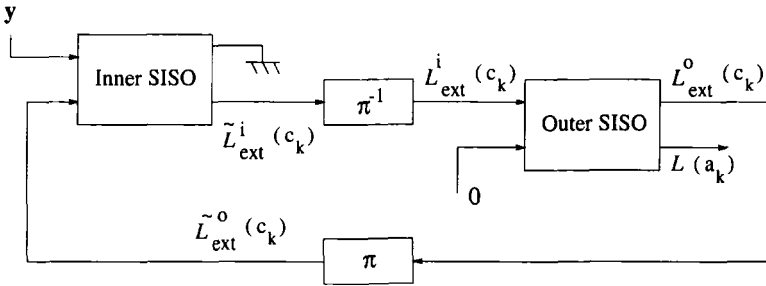
Figure 8.38. SCCC encoder.

Figure 8.39. SCCC decoder.

7.3 SCCC ENCODER AND DECODER

Fig. 8.38 shows a SCCC encoder which is a serial concatenation of two component codes separated by an interleaver. In a SCCC scheme, the input data sequence of length N', is first encoded by an outer convolutional code C_o with rate $R^o = k/p$. The output of C_o is interleaved using a pseudo-random interleaver of length $N = N'/R^o$, and then encoded using an inner convolutional code C_i with rate $R^i = p/n$. The SCCC has code rate

$$R_T = R_c^{(1)} R_c^{(2)} = (k/p)(p/n) = k/n \ . \tag{8.109}$$

The codewords of the outer and inner codes are referred to as outer and inner codewords, respectively. Consequently, the inner codewords are also the codewords of the SCCC. With SCCCs, the inner encoder must be recursive for reasons to be seen later. The outer code does not have to be recursive.

The structure of the SCCC decoder is shown in Fig. 8.39. It operates in an iterative fashion similar to the PCCC decoder. However, the SISO modules now produce APPs or LLRs for the information bits, a_k, and the code bits c_k from the outer coder.

7.4 WEIGHT DISTRIBUTION

It is sometimes useful to view PCCCs and SCCCs as equivalent block codes with input sequences of length $N' = N/k$ and $N' = Nk/p$, respectively, where N is the interleaver size. Like block codes, Turbo codes can be described by a distance spectrum (d, A_d), where A_d is the number of codewords of weight Hamming weight d. The conditional weight enumerating function (CWEF) of

a block code defined as [32]

$$A_w(z) \triangleq \sum_d A_{w,d} z^d \qquad (8.110)$$

where $A_{w,d}$ is the number of weight-d codewords having information-weight w. Note that $A_d = \sum_w A_{w,d}$. The smallest non-zero value of d is the free Hamming distance of the code, denoted by d_{free}. The union bound on the probability of bit error is

$$P_b(e) \leq \frac{1}{N'} \sum_w \sum_{d=d_{\text{free}}} w A_{w,d} P_2(d) \qquad (8.111)$$

where $P_2(d)$ is the pairwise error probability between two coded sequences separated by Hamming distance d.

To obtain a low $P_b(e)$ there are generally two approaches; we can either decrease $A_{w,d}$ or increase d_{free}. With convolutional codes, A_d increases rapidly with d^6. As a result, convolutional codes are said to have a "dense distance spectrum." Also, $A_d \propto N'$ with convolutional codes, due to their time invariant property. Hence, for convolutional codes a decrease in $P_b(e)$ is usually obtained by increasing d_{free}, which ultimately obtained by increasing the total encoder memory. Turbo codes take other approach by drastically decreasing A_d. This property is called "spectral thinning."

The spectral thinning property of Turbo codes can be explained intuitively as follows. Considering PCCCs, the total weight of a PCCC codeword is equal to the weight of the systematic and parity components

$$w(\mathbf{b}) = w(\mathbf{b}_s) + w(\mathbf{b}_p^{(1)}) + w(\mathbf{b}_p^{(2)}) \ . \qquad (8.112)$$

Consider for example a RSC with generator matrix $\left[1, \frac{1+D^2}{1+D+D^2}\right]$ and the random interleaver shown in Fig. 8.40. Certain input sequences a will lead to low output weights $w(\mathbf{b}_p^{(1)})$ from the first encoder RSC1. For example, the input sequence $a(D) = 1 + D^3$ produces the output $\mathbf{b}_p^{(1)}(D) = 1 + D + D^2 + D^3$ from the first encoder RSC1. However, the interleaved sequence $\tilde{\mathbf{a}}(D)$ will usually lead to a high output weight $w(\mathbf{b}_p^{(2)})$ from the second encoder RSC2. Consequently, most codewords have large weight. However, some input sequences that produce low weight codewords in one encoder, after interleaving will also produce low weight codewords in the other encoder. Therefore, there are a few codewords with small weight. For most random interleavers, this occurs with high probability [87]. At high E_b/N_o the error events corresponding

---

[6]It is important to realize that A_d is not equal to a_d (in our earlier discussion of convolutional codes), since the Turbo codewords can consist of multiple error events.
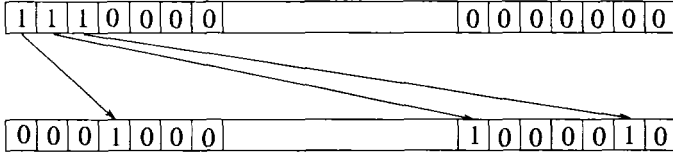
Figure 8.40. Random Turbo interleaver.

to these low-weight codewords dominate the BER performance with the result that the BER curves of PCCCs flatten at high E_b/N_o. This has been loosely termed as an "error floor" [32, 71].

In the sequel, convolutional codes, PCCC and SCCC are discussed simultaneously and, to avoid confusion, the quantities associated with them are distinguished by the superscripts c, T, and S, respectively.

For convolutional codes, every non-zero codeword corresponds to an error event or a concatenation of error events. The weight of a codeword equals the sum of the weights of the error events. Let $A^c_{w,d,i}$ denote the number of weight-d codewords having information weight-w and formed by the concatenation of i error events. Then, the number of weight-d codewords with information weight-w is $A^c_{w,d} = \sum_{i=1}^{n_{max}} A^c_{w,d,i}$, where n_{max} is the maximum number of possible error events for a length-N' input sequence.

The distance spectrum of Turbo codes is difficult to determine for a particular Turbo interleaver. Fortunately, Benedetto and Montorsi [32] solved this problem by introducing a hypothetical interleaver called *uniform interleaver* that permutes a given weight-w sequence onto any of the $\binom{N}{w}$ possible interleaved sequences with equal probability. The distance spectrum of a Turbo code with a uniform interleaver can be obtained by averaging the distance spectrum over all possible interleaver mappings. At least half the random interleavers are guaranteed to yield a weight distribution that is as good as the average weight distribution. Furthermore, most of the randomly generated interleavers have a weight distribution that is close to the average weight distribution. Hence, the typical performance of a Turbo code with a randomly chosen interleaver can be obtained from the average weight distribution with a uniform interleaver.

7.4.1 WEIGHT DISTRIBUTION OF PCCCS

With a uniform interleaver the number of weight-d Turbo codewords with weight-w input sequences is, for large N, [31]

$$A^T_{w,d} \approx \sum_{l=0}^{d} \sum_{n_1=1}^{n_{max}} \sum_{n_2=1}^{n_{max}} \frac{\binom{N}{n_1}\binom{N}{n_2}}{\binom{N}{w}} A^c_{w,l,n_1} A^c_{w,d-l,n_2} \qquad (8.113)$$

Using the approximation $\binom{N}{n} \approx \frac{N^n}{n!}$ gives

$$A_{w,d}^T \approx \sum_{l=0}^{d} \sum_{n_1=1}^{n_{max}} \sum_{n_2=1}^{n_{max}} \frac{w!}{n_1! \cdot n_2!} N^{n_1+n_2-w} A_{w,l,n_1}^c A_{w,d-l,n_2}^c \tag{8.114}$$

Observe that the multiplicity, $A_{w,d}^T$, of the PCCC codewords is inversely proportional to the interleaver length N. Consequently, increasing N results in very small multiplicity, a phenomenon called **spectral thinning**, and is the reason for the remarkable performance of Turbo codes. In contrast, we note that the time-invariant property of convolutional codes implies that $A_d^c \propto N$.

The uniform interleaver is hypothetical and impractical. For reasonably large interleaver sizes N, random interleavers perform very well [87]. To see why, consider a rate-1/3, 8-state, PCCC code where the RSC component encoders have generator matrices $\left[1, \frac{1+D^2}{1+D+D^2}\right]$. Since the component codes are recursive, all weight-1 input sequences produce infinite-weight output sequences. The minimum distance error event at the output of *each* RSC encoder corresponds to an input error sequence of the form $D^i(1 + D + D^2)$. However, the random interleaver permutes such sequences very effectively so that the output of the other encoder has high weight [87]. Weight-2 input error sequences to RSC1 of the form $.D^i(1 + D^3)$ will produce a finite-weight output sequence having the form $D^i(1 + D + D^2 + D^3)$. However, the random interleaver permutes these sequences into sequences which are *not* of the form $.D^j(1 + D^3)$ with high probability [87]. However, an occasional *bad mapping* occurs, where input sequences of the form $D^i(1 + D^3)$ are permuted into input sequences of the form $D^j(1 + D^3)$ for some i, j. This is illustrated in Fig. 8.41. Such input sequences produce low-weight outputs from both encoders and define the minimum Hamming distance of the PCCC code. The probability that an input sequence $D^i\mathbf{a}$ of weight-w is interleaved into a sequence $\tilde{\mathbf{a}}$ of the form $D^j\mathbf{a}$ for at least one pair i, j is proportional to N^{w-2} [87]. Hence, bad mappings are very likely to occur for weight-2 input sequences and very unlikely to occur for weight $w > 2$ input sequences. So the minimum distance error event corresponds to a weight-2 input sequence with very high probability. If the smallest weight RSC output corresponding to all weight-2 input sequences is d_{ceff^c}, then the free Hamming distance of the PCCC code is $d_{free}^T = 2 + 2d_{ceff}$. For our example PCCC code, the free distance is $d_{free} = 2 + 4 + 4 = 10$, which is rather small. This small free Hamming distance is typical of PCCCs precisely the reason for the so called BER and FER floor of PCCCs. Finally, we note that other types of interleavers, such as the S-random interleaver, are generally very difficult to analyze, but most of the above arguments are valid.

PCCCs inherently provide unequal error protection, because the bad interleaver mappings define certain bit positions are affected by th dominant error events. Such bad mappings affect only a very few bit positions, but they nev-
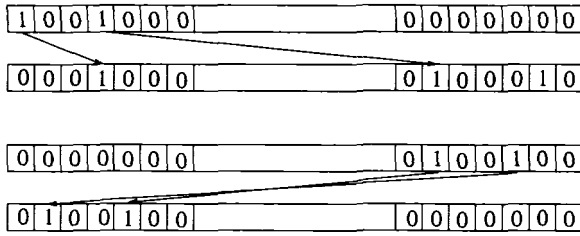
Figure 8.41. Bad random interleaver mappings.

ertheless result in a BER floor. In contrast, for convolutional codes all bit positions in the input sequence are affected by the same error events. Consequently, all bit positions are equally likely to be in error. So PCCCs are inherently unequal error protecting (UEP) codes.

It is instructive to understand how the expected number of bad mappings changes with the interleaver size, N. The total number of possible interleaver mappings for a block of N bits is $N!$ The number of *bad mappings,* where a sequence of the form $D^i(1 + D^3)$ is mapped into a sequence of the form $D^j(1 + D^3)$ is approximately $N \times 2 \times (N - 2)!$ The approximation is due to the fact that edge effects have been ignored which is a valid assumption for large N. Therefore, the probability that a sequence of the form $D^i(1 + D^3)$ is mapped onto a sequence of the form $D^j(1 + D^3)$ is

$$P(D^i(1 + D^3) \rightarrow D^j(1 + D^3)) = \frac{2N(N - 2)!}{N!} = \frac{2}{N - 1} . \qquad (8.115)$$

Assuming that the mappings for the different bit positions are independent and ignoring the edge effects, the distribution of the total number of such *bad* mappings k, in a block of length N, can be approximated by a binomial distribution for small k, i.e., [7]

$$P(\text{total number of bad mappings} = k) = \binom{N}{k} \left(\frac{2}{N - 1}\right)^k \left(1 - \frac{2}{N - 1}\right)^{N-k}$$

The mean number of *bad* mappings is $N\frac{2}{N-1}$, which converges to 2 for large N. Therefore, the mean number of data bits affected by *bad* mappings converges to 4 for large N, since the *bad* mappings correspond to weight-2 input error sequences.

---

[7]The case of large k is not of interest because the probability of many bad mappings is extremely small and therefore, does not contribute significantly to the mean of the distribution.

7.4.2 WEIGHT DISTRIBUTION OF SCCCS

Consider the serial concatenation system in Fig. 8.38. Let the input block length is N' bits. The length of the outer codeword and, therefore, the interleaver size and length of the input to the inner encoder is $N = N'/R^o = N'p/k$ bits. Under the assumption of a uniform interleaver, the number of weight-h code words that are generated by weight-w input sequences is [287]

$$A_{w,d}^S = \sum_{l=d_f^o}^{N} \sum_{n^o=1}^{n_M^o} \sum_{n^i=1}^{n_M^i} \frac{\binom{N'/p}{n^o}\binom{N/p}{n^i}}{\binom{N}{l}} A_{w,l,n^o}^{C_o} A_{l,d,n^i}^{C_i} \qquad (8.116)$$

where d_f^o is the minimum free distance of the outer code, and n_M^o and n_M^i refer to the maximum number of error events possible for the outer and inner codes, respectively. By using the approximation $\binom{N}{n} \approx \frac{N^n}{n!}$ [287]

$$A_{w,d}^S \approx \sum_{l=d_f^o}^{N} \sum_{n^o=1}^{n_M^o} \sum_{n^i=1}^{n_M^i} N^{n^o+n^i-l-1} \frac{l!}{p^{n^o+n^i}n^o!n^i!} \frac{1}{n} A_{w,l,n^o}^o A_{l,h,n^i}^i \qquad (8.117)$$

where w_m^o is the minimum-weight of all input sequences that will produce an error event for the outer code.

Observe from (8.117) that the contribution of each codeword to the BER is multiplied by the term $N^{n^o+n^i-l-1}$. Therefore, when $n^o + n^i - l - 1 < 0$, increasing N decreases the BER exponentially. This effect is called the interleaver gain. Consider a weight-l outer codeword which is a result of n^o error events of the outer code. If the inner encoder is non-recursive, then a weight-l outer codeword can result in a maximum of l error events (each '1' in the outer codeword can cause an error event). Therefore, n^i can be equal to l. In this case, the exponent of N will be $n^o - 1$ and, when $n^o > 1$, the exponent of N will be positive. Consequently, increasing N increases the contribution of such codewords to the final BER [287]. When $n^o = 1$, the exponent of N will be zero, implying that the interleaver does not impact the multiplicity of such codewords or, equivalently, no interleaving gain is possible.

When the inner encoder is recursive, only input sequences having weight-2 or greater can cause error events. Therefore, a weight-l outer codeword can cause at most $\lfloor l/2 \rfloor$ error events for the inner code. Consequently, the exponent of N is $n^o - \lceil l/2 \rceil - 1$. If all outer codewords corresponding to one error event of the outer code ($n^o = 1$) have weight $l > 2$ or, equivalently, the free distance of the outer code is greater than 2, the exponent of N is always negative. This implies that increasing N will always decrease the BER.
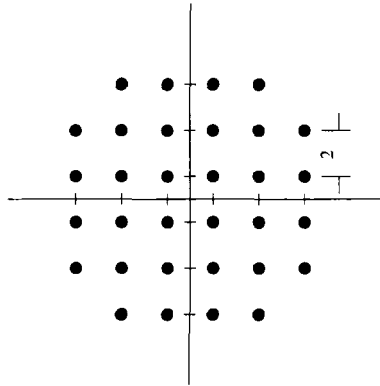
Problems

8.1. Consider a rate-1/3 convolutional code with generators $\mathbf{g}^{(1)} = (111)$, $\mathbf{g}^{(2)} = (111)$, and $\mathbf{g}^{(3)} = (101)$.

    a) Draw a block diagram of the encoder structure.

    b) Draw the state diagram and trellis diagram.

    c) Determine the output sequence corresponding to the input sequence 1110101.

8.2. The output of a rate-1/3 convolutional encoder with constraint length 3 to the input $\mathbf{a} = (1,1,0,\ldots)$ is $\mathbf{b} = (111, 110, 010,\ldots)$

    a) Determine the transfer function $T(D,N,L)$.

    b) Determine the number of paths through the state diagram or trellis that diverge from the all-zeroes state and remerge with the all-zeroes state 7 branches later.

    c) Determine the number of paths of Hamming distance 20 from the all zeroes sequence.

8.3. Consider the rate-1/3 code in Problem 8.1.

    a) Determine the transfer function $T(D, N, L)$ of the code. What is the free Hamming distance d_{free}?

    b) Assuming the use of BPSK signaling and an AWGN channel, derive a union-Chernoff bound on the decoded bit error probability with i) hard decision decoding and ii) soft decision decoding.

    c) Repeat part b) assuming an interleaved flat Rayleigh fading channel, where the receiver has perfect knowledge of the channel.

8.4. Consider the 8-PAM and 32-CROSS signal constellations in Fig. 8.42.

    a) Construct the partition chain as in Fig. 8.11 and compute the minimum Euclidean distance between signal points at each step in the partition chain.

    b) What is the average symbol energy for each of the signal constellations.

8.5. Consider the 2-state, rate-1/2, trellis encoder shown in Fig. 8.43. By using this encoder with a 4-PAM and 8-PAM signal constellation we can construct a TCM systems having bandwidth efficiencies of 1 bit/s/Hz and 2 bits/s/Hz, respectively.

    a) Determine the appropriate partitions for the signal constellation for the 2-state, 4-PAM and 8-PAM trellis codes.

    b) Construct and label the trellis diagrams for the 2-state 4-PAM and 8-PAM trellis codes.
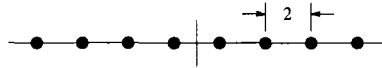
a) 8-PAM

b) 32-CROSS

Figure 8.42. Signal constellations for Problem 8.4.
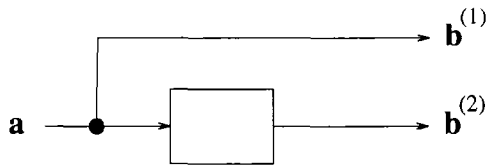
Figure 8.43. Trellis encoder for Problem 8.5.

   c) Determine the minimum Euclidean distance for each trellis code, and the asymptotic coding gain on an AWGN channel relative to the equivalent uncoded systems.

8.6. Construct and label the trellis diagram for a two-state MTCM system using 8-PSK. What is the asymptotic coding gain for this system on an AWGN channel relative to the equivalent uncoded system.

8.7. For the MTCM code shown in Fig. 8.19, show how the values of MDD1 and MDD2 are determined. Repeat for the 2-D code shown in Fig. 8.20.

8.8. To simplify the calculation of performance bounds a Chernoff bound is often imposed on the pairwise error probability.

   a) Derive the Chernoff bound on the pairwise error probability for an AWGN channel with soft decision decoding, given by (8.61).

b) Derive the Chernoff bound on the pairwise error probability for an AWGN channel with hard decision decoding, given by (8.63).

c) Derive the Chernoff bound on the pairwise error probability for an interleaved flat fading channel with soft decision decoding, given by (8.75).

8.9. Suppose that the 2-state, 4-PAM trellis code in Problem 8.5 is used on a 2-tap ISI channel characterized by the channel vector $\mathbf{g} = (g_0, g_1)$.

a) Construct the super-trellis diagram.

b) What is the minimum distance error event in the super-trellis?

c) Determine the condition number of the path distance matrix $\mathbf{E}$ for the minimum distance error event found in part b).

d) Determine the channel vectors that minimize and maximize the pairwise error probability.

Chapter 9

SPREAD SPECTRUM TECHNIQUES

Spread spectrum systems were originally developed for military applications, to provide antijam and low probability of intercept communications by spreading a signal over a large frequency band and transmitting it with a low power per unit bandwidth [86], [262], [306]. Recently, code division multiple access (CDMA) based on spread spectrum technology has been recognized as a viable alternative to both frequency division multiple access (FDMA) and time division multiple access (TDMA) for cellular systems. During the late 1980s and early 1990s, Qualcomm, Inc.'s efforts, along with those of many other organizations such as Motorola and AT&T, have lead to the North American IS-95 cellular standard [96]. A detailed description of the IS-95 CDMA cellular approach can be found in a number of papers, including those by Lee [193] and Gilhousen *et al.* [136]. The book by Viterbi [344] provides a good coverage of the spread spectrum concepts that form the foundation of the IS-95 CDMA cellular system.

Spread spectrum signals have the distinguishing characteristic that the bandwidth used is much greater than the message bandwidth. This band spread is achieved by using a spreading code or pseudo-noise (PN) sequence that is *independent* of the message and is known to the receiver. The receiver uses a synchronized replica of the PN sequence to despread the received signal allowing recovery of the message. Since the PN sequence is independent of the message, the bandwidth expansion does not combat additive white Gaussian noise (AWGN), unlike some other modulation techniques such as wide band analog FM. Nevertheless, the wide band character of spread spectrum signals can be utilized to mitigate the effects of interference and multipath fading.

While there are many different types of spread spectrum systems, the two predominant types are direct sequence (DS) spread spectrum and frequency hopped (FH) spread spectrum. DS spread spectrum achieves the band spread

by using the PN sequence to introduce rapid phase transitions into the carrier containing the data, while FH spread spectrum achieves the band spread by using the PN sequence to pseudo-randomly hop the carrier frequency throughout a large band. An excellent tutorial treatment of spread spectrum can be found in the books by Simon *et al.* [306] and Ziemer and Peterson [381]. Some of the early proposals that applied CDMA to cellular radio, such as the system proposed by Cooper and Nettleton [65], were based on FH spread spectrum while most of the recent standards, such as IS-95, favor DS spread spectrum. As a result, the focus of this chapter is on DS CDMA.

While it appears that any cellular system can be suitably optimized to yield a competitive spectral efficiency regardless of the multiple access technique being used, CDMA offers a number of advantages along with some disadvantages. The advantages of CDMA for cellular applications include i) universal one-cell frequency reuse, ii) narrow band interference rejection, iii) inherent multipath diversity in DS CDMA, iv) ability to exploit silent periods in speech voice activity, v) soft hand-off capability, vi) soft capacity limit, and vii) inherent message privacy. The disadvantages of CDMA include i) stringent power control requirements with DS CDMA , ii) hand-offs in dual mode systems, and iii) difficulties in determining the base station (BS) power levels for deployments that have cells of differing sizes, and iv) pilot timing.

This chapter begins with an introduction to DS and FH spread spectrum in Section 1., along with a comparison between these two types of spread spectrum systems. Such a comparison is important if we are to determine the best CDMA approach for a given environment. PN sequences are fundamental to all spread spectrum systems and are the subject of Section 2.. The remainder of the chapter concentrates on DS spread spectrum. Section 5. discusses the performance of point to point DS spread spectrum on frequency selective fading channels and shows how a RAKE receiver can be used to gain multipath diversity. Error probability upper and lower bounds and approximations are essential for predicting the performance of CDMA systems. Section 6. considers an accurate analysis of the error probability of DS CDMA on AWGN channels. Several Gaussian approximations to the error probability are derived. The chapter concludes with a performance evaluation of cellular DS CDMA. Unfortunately, DS CDMA cellular systems are very complex systems with intricate interactions between system functions. Therefore, the analytical evaluation of system capacity typically requires simplifying assumptions, while focusing on a particular parameter or effect. Usually we can obtain relative performance comparisons, while the true capacity of a suitably optimized CDMA system in a realistic deployment scenario remains elusive.
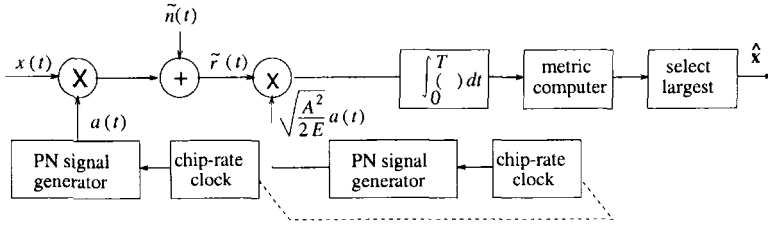
Figure 9.1. Simplified DS/QPSK system.

1. BASIC PRINCIPLES OF SPREAD SPECTRUM
1.1 DIRECT SEQUENCE (DS) SPREAD SPECTRUM

A simplified DS/QPSK spread spectrum system is shown in Fig. 9.1. The pseudo-random (PN) sequence generator produces a **spreading sequence a** = $\{ajt\}$, which is actually a periodic deterministic sequence with period N. This spreading sequence is used to generate the **spreading waveform**

$$a(t) = A \sum_k a_k h_c(t - kT_c) \ , \tag{9.1}$$

where $\mathbf{a} = \{a_k : a_k \in \{\pm 1 \pm j\}\}$ is a complex spreading sequence, T_c is the PN symbol or **chip** period, and $h_c(t)$ is a real chip amplitude shaping function having a peak amplitude of unity. The energy per chip is

$$E_c = A^2 \sigma_a^2 \int_{-\infty}^{\infty} h_c^2(t)dt = A^2 \int_{-\infty}^{\infty} h_c^2(t)dt \tag{9.2}$$

since $\sigma_a^2 = \frac{1}{2}\mathrm{E}[|a_k|^2] = 1$. Notice that spectral control is achieved with DS spread spectrum by shaping the PN chips rather than the data symbols.

The data symbol sequence $\{x_n\}$ is used to generate the waveform

$$x(t) = \sum_n x_n u_T(t - nT) \ , \tag{9.3}$$

where A is the amplitude, $\mathbf{x} = \{x_n : x_n \in \{\pm 1/\sqrt{2} \pm j/\sqrt{2}\}\}$ is the complex data symbol sequence, and T is the data symbol duration. It is necessary that T be an integer multiple of T_c, and the ratio $G = T/T_c$ is called the **processing gain,** defined here as the number of PN chips per data symbol. There are two types of spreading codes, distinguished by the relative values of N and G. A **short code** has $G = N$, so that each data symbol is spread by a full period of the spreading sequence. A **long code** has $G \ll N$, so that each data symbol is spread by a subsequence or partial period of the spreading sequence.

The DS/QPSK complex envelope, obtained by multiplying $a(t)$ and $x(t)$, is

$$\tilde{s}(t) = A \sum_n x_n h_n(t - nT) \tag{9.4}$$

where

$$h_n(t) = \sum_{k=0}^{G-1} a_{nG+k} h_c(t - kT_c) \ . \tag{9.5}$$

The **complex spreading** operation is illustrated in Fig. 9.2. Notice that the DS/QPSK signal can be thought of as a QPSK signal where the nth data symbol is shaped with the amplitude shaping pulse $h_n(t)$ in (9.5). For short codes $h_n(t)$ is the same for all data symbols. The advantage of complex spreading is a reduction in the peak-to-average ratio of the magnitude of the complex envelope. OQPSK should not be used with complex spreading, since it will actually increase the peak-to-average ratio. The complex envelope $\tilde{s}(t)$ is applied to a quadrature modulator to produce the bandpass waveform

$$
\begin{aligned}
s(t) \quad = \quad & A \sum_n \Big\{ \ (x_{I,n} h_{I,n}(t - nT) - x_{Q,n} h_{Q,n}(t - nT)) \cos(2\pi f_c t) \\
& - (x_{Q,n} h_{I,n}(t - nT) + x_{I,n} h_{Q,n}(t - nT)) \sin(2\pi f_c t) \Big\}
\end{aligned} \tag{9.6}
$$

where

$$
\begin{aligned}
h_n(t) \quad &= \quad h_{I,n}(t) + j h_{Q,n}(t) \tag{9.7}\\
x_n \quad &= \quad x_{I,n} + j x_{Q,n} \ . \tag{9.8}
\end{aligned}
$$

During the time interval $[nT, (n+1)T]$, the DS/QPSK complex envelope can assume one of the four possible values

$$\tilde{s}_i(t) = A h_n(t) x_i, \quad i = 1, \ldots, 4 \ . \tag{9.9}$$

Using the basis function

$$\phi_n(t) = \sqrt{\frac{A^2}{2E}} h_n(t) \tag{9.10}$$

where $E = G E_c$ is the symbol energy, we can write

$$\tilde{s}_i(t) = \sqrt{2E} x_i \phi_n(t), \quad i = 1, \ldots, 4 \tag{9.11}$$

and it follows that the complex DS/QPSK signal vectors are

$$\tilde{s}_i = \sqrt{2E} x_i, \quad i = 1, \ldots, 4 \ . \tag{9.12}$$

Notice that the basis function $\phi_n(t)$ is indexed with the baud epoch n.

Besides complex spreading, other types of PN spreading are possible. We could use **dual-channel quaternary spreading** as shown in Fig. 9.3. Usually this scheme is used with OQPSK modulation to reduce the peak-to-average
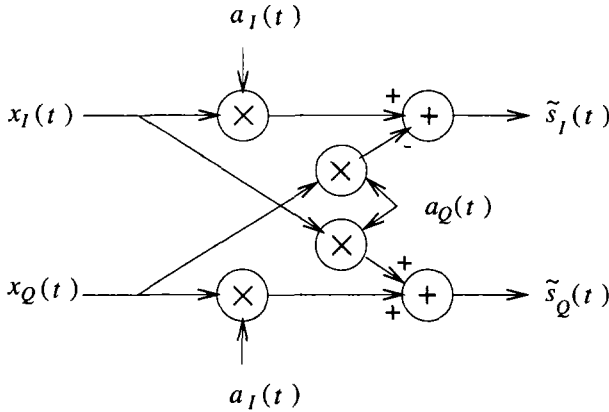
Figure 9.2.   Complex spreading.
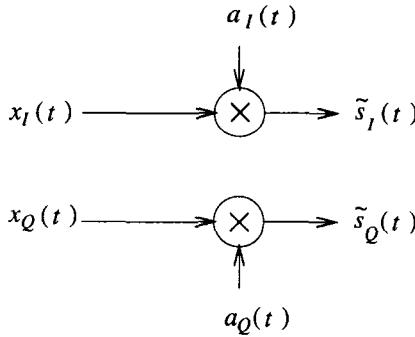
Figure 9.3.   Dual-channel quaternary spreading.

ratio of the magnitude of the complex envelope. If only one data sequence is to be transmitted, then we could use either **simple binary spreading** or **balanced quaternary spreading**, as shown in Fig. 9.4. Balanced quaternary spreading is known to be less sensitive to interference than simple binary spreading.

Fig. 9.1 also shows a simplified DS/QPSK receiver. In general, the DS spread spectrum receiver must perform three functions; synchronize with the incoming spreading sequence, despread the signal, and detect the data. Consider the received complex envelope in the time interval $[nT, (n + 1)T]$. This signal can be despread and detected by using the correlator detector in Fig. 5.2 or the matched filter detector in Fig. 5.3, where $\phi_n(t)$ is defined in (9.10). The output of the correlator or matched filter despreader/detector is

$$\tilde{r} = \tilde{s}_i + \tilde{n} \tag{9.13}$$

where $\tilde{n}$ is a zero-mean Gaussian random variable with variance $\frac{1}{2}E[|\tilde{n}|^2] = N_o$.

462
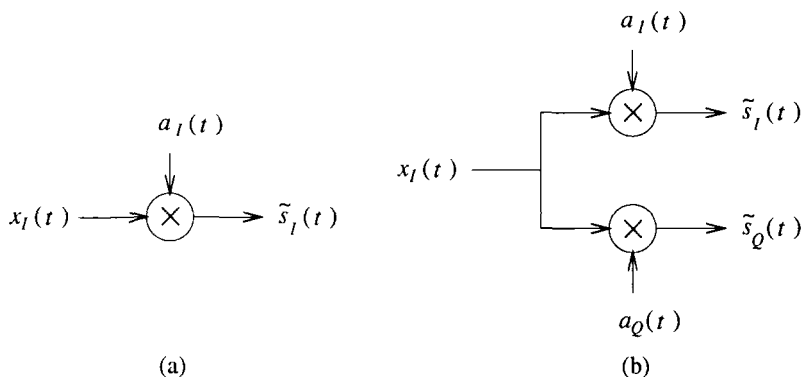
Figure 9.4. Spreading binary data using (a) simple binary spreading, and (b) balanced quaternary spreading.

The ML receiver observes $\tilde{r}$ and decides in favor of the signal vector $\tilde{s}_m$ that minimizes the squared Euclidean distance

$$\mu(\tilde{s}_m) = \|\tilde{r} - \tilde{s}_m\|^2 \ .\tag{9.14}$$

It follows that the bit error probability of DS/QPSK with Gray coding is identical to QPSK, given by

$$P_b = Q(\sqrt{2\gamma_b})\tag{9.15}$$

where $\gamma_b = E_b/N_o$ is the received bit energy-to-noise ratio. Note that spread spectrum signaling does nothing to improve the error rate performance on an AWGN channel. However, in the sequel we will show that spread spectrum signaling offers significant error rate performance gains against additive interference, multipath-fading, and other types of channel impairments.

1.2 FREQUENCY HOP (FH) SPREAD SPECTRUM

Frequency hopping spread spectrum systems hop the carrier frequency pseudo-randomly throughout a finite set of hop frequencies. The most common type of modulation with frequency hopping is orthogonal M-ary frequency shift keying (MFSK). The MFSK complex envelope is

$$\tilde{s}(t) = A \sum_n e^{jx_n\pi\Delta_f t} u_T(t - nT)\tag{9.16}$$

where Δ_f is the frequency separation, and $x_n \in \{\pm1, \pm3, \ldots, \pm M - 1\}$. A FH/MFSK waveform can be generated by using a digital frequency synthesizer whose inputs consist of the data sequence and the contents of a pseudo-noise sequence generator. A conceptual FH/MFSK spread spectrum system is shown in Fig. 9.5.
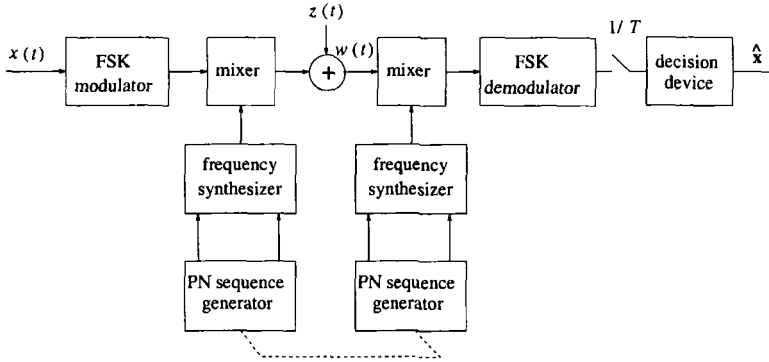
Figure 9.5.   Simplified FH system operating on an AWGN channel.

There are two basic types of FH spread spectrum modulation, fast frequency hopping (FFH) and slow frequency hopping (SFH). SFH systems transmit one or more (in general L) data symbols per hop. The SFH/MFSK complex envelope is

$$\tilde{s}(t) = A\sum_{n}\sum_{i=1}^{L} e^{jx_{nL+i}\pi\Delta_f t + 2\pi f_n t} u_T(t - (nL + i)T) \qquad (9.17)$$

where the first sum indexes the sequence of hop frequencies f_n, and the second sum indexes the vector of L data symbols $\mathbf{x}_n = (x_{nL+1}, x_{nL+2}, \ldots, x_{(n+1)L})$ that are transmitted at the nth hop.

FFH systems transmit the same data symbol on multiple (in general L) hop frequencies. If independent interference is experienced on each of the hop frequencies then a diversity gain is achieved. The FFH/MFSK complex envelope is

$$\tilde{s}(t) = A\sum_{n}\sum_{i=1}^{L} e^{jx_n\pi\Delta_f t + 2\pi f_{nL+i}t} u_{T/L}(t - (nL + i)T/L) \qquad (9.18)$$

where the first sum indexes the sequence of data symbols, x_n, and the second sum indexes the sequence of hop frequencies $\mathbf{f}_n = (f_{nL+1}, f_{nL+2}, \ldots, f_{(n+1)L})$ that are used for the nth data symbol.

With orthogonal MFSK the required frequency separation Δ_f depends on the type of detection that is used. Coherent detection requires a frequency separation $\Delta_f = 1/2T_s$, while non-coherent detection requires $\Delta_f = 1/T$ (see Problem 4.5). If coherent detection can be used, then the error probability of SFH/MFSK or FFH/MFSK on an AWGN channel is given by (5.103). However, FH/MFSK is often detected non-coherently because of the difficulty in achieving rapid carrier synchronization when the carrier frequency is hopped.

The error probability of SFH/MFSK on an AWGN channel with non-coherent square-law detection is given by (5.174). If FFH/MFSK is used on an AWGN channel, then the error probability assumes a more complicated form, (see [270]).

2.   SPREADING SEQUENCES

CDMA systems achieve their multiple-access capability by using large sets spreading sequences that are chosen to have three desirable attributes; i) the sequences are balanced so that each element of the sequence alphabet occurs with equal frequency, ii) the autocorrelations have small off-peak values, to allow for rapid sequence acquisition at the receiver and to minimize self interference due to multipath, iii) the cross-correlations are small at all delays, to minimize multiple-access interference.

Spreading sequences are often characterized in terms of their discrete-time correlation properties. Let $\mathbf{a}^{(k)}$ denote the kth complex spreading sequence[1]. For spread spectrum systems that employ short codes, each data symbol is spread by a full period of the spreading sequence. In this case the full period correlation properties are of interest. The **full period autocorrelation** of the sequence $\mathbf{a}^{(k)}$ is[2]

$$\phi_{k,k}(n) = \frac{1}{2N} \sum_{i=0}^{N-1} a_i^{(k)} a_{i+n}^{(k)*} \tag{9.19}$$

and the **full period cross-correlation** between the sequences $\mathbf{a}^{(k)}$ and $\mathbf{a}^{(m)}$ is

$$\phi_{k,m}(n) = \frac{1}{2N} \sum_{i=0}^{N-1} a_i^{(k)} a_{i+n}^{(m)*} \tag{9.20}$$

where N is the length or period of the spreading sequences.

The **aperiodic autocorrelation** of $\mathbf{a}^{(k)}$ is defined as

$$\phi_{k,k}^a(n) = \begin{cases} \frac{1}{2N} \sum_{i=1}^{N-n} a_{i+n}^{(k)} a_i^{(k)*} & , \ 0 \leq n \leq N-1 \\ \frac{1}{2N} \sum_{i=1}^{N+n} a_i^{(k)} a_{i-n}^{(k)*} & , \ -N+1 \leq n \leq 0 \\ 0 & , \ |n| \geq N \end{cases} \tag{9.21}$$

For spread spectrum systems that employ long codes, each data symbol is spread by only a portion of the spreading sequence. In this case, the partial period correlations are of interest. The **partial period auto- and cross-**

---

[1]The following development also applies to real spreading sequences.
[2]Throughout this section complex spreading sequences are assumed. For real spreading sequences, the correlation functions are similar but are normalized by N rather than $2N$.

correlations are

$$\phi_{k,k}^{p}(n) \;=\; \frac{1}{2G} \sum_{i=0}^{G-1} a_i^{(k)} a_{i+n}^{(k)^{*}} \tag{9.22}$$

$$\phi_{k,m}^{p}(n) \;=\; \frac{1}{2G} \sum_{i=0}^{G-1} a_i^{(k)} a_{i+n}^{(m)^{*}} \;. \tag{9.23}$$

The partial period correlations are not only a function of the delay n, but also depend upon the point in the sequence(s) where the summation actually starts. The partial period correlations are difficult to derive analytically, except for certain types of sequences. Therefore, we often resort to a statistical treatment under the assumption that the sequences are randomly generated, i.e., the sequence elements are chosen from the set $\{\pm 1, \pm j\}$ independently and with probability. For random sequences

$$\frac{1}{2}E[a_n^{(k)}] = 0 \qquad \frac{1}{2}E[|a_n^{(k)}|^2] = 1 \qquad \frac{1}{2}E[a_n^{(k)} a_n^{(m)^{*}}] = 0 \;. \tag{9.24}$$

Hence, the mean value of the partial period autocorrelation is

$$\mu_{\phi_{k,k}^{p}}(n) = E[\phi_{k,k}^{p}(n)] \;=\; \frac{1}{2G} \sum_{i=0}^{G-1} E[a_i^{(k)} a_{i+n}^{(k)^{*}}]$$

$$= \;\; \delta_{n,\ell N} \tag{9.25}$$

where

$$\delta_{n,\ell N} = \left\{ \begin{array}{ll} 1 & , \quad n = \ell N \\ 0 & , \quad n \neq \ell N \end{array} \right. \tag{9.26}$$

l an integer. The variance of the partial period autocorrelation is

$$\sigma_{\phi_{k,k}^{p}(n)}^{2} \;=\; E[|\phi_{k,k}^{p}(n)|^2] - \mu_{\phi_{k,k}^{p}(n)}^{2}$$

$$= \;\; \frac{1}{(2G)^2} \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} E[a_i^{(k)} a_{i+n}^{(k)^{*}} a_j^{(k)} a_{j+n}^{(k)^{*}}] - \mu_{\phi_{k,k}^{p}(n)}^{2}$$

$$= \;\; (1 - \delta_{n,\ell N})(1/G) \;. \tag{9.27}$$

Likewise, the mean and variance of the partial period cross-correlation are

$$\mu_{\phi_{k,m}^{p}(n)} \;=\; E[\phi_{k,m}^{p}(n)] = 0 \;, \;\; \forall n \tag{9.28}$$

$$\sigma_{\phi_{k,m}^{p}(n)}^{2} \;=\; E[|\phi_{k,m}^{p}(n)|^2] - \mu_{\phi_{k,m}^{p}(n)}^{2} = 1/G \;, \;\; \forall n \;. \tag{9.29}$$

2.1   SPREADING WAVEFORMS

The full period cross-correlation between two spreading waveforms $a^{(k)}(t)$ and $a^{(m)}(t)$ is

$$
\begin{aligned}
R_{k,m}(\tau) &= \frac{1}{T}\int_0^T a^{(k)}(t)\, a^{(m)*}(t+\tau)dt \\
&= \frac{1}{T}\sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} a_i^{(k)} a_j^{(m)*}\int_0^T h_c(t-iT_c)h_c(t+\tau-jT_c)dt \ .
\end{aligned}
$$

(9.30)

The integral in (9.30) is nonzero only where the chip pulses $h_c(t-iT_c)$ and $h_c(t+\tau-jT_c)$ overlap. Since the delay τ can assume any value let $\tau = \ell T_c + \delta$, where $\ell = \lfloor \tau/T_c \rfloor$ is an integer and $0 \le \delta < T_c$. If the chip pulses are chosen to have duration T_c and $\tau = \ell T_c + \delta$, then the chip pulses overlap only for $i = \ell + j$ and $i = \ell + j + 1$, so that

$$
\begin{aligned}
R_{k,m}(\tau) &= \frac{1}{N}\sum_{i=0}^{N-1} a_i^{(k)} a_{\ell+i}^{(m)*} \frac{1}{T_c}\int_0^{T_c-\delta} h_c(t')h_c(t'+\delta)dt' \\
&+ \frac{1}{N}\sum_{i=0}^{N-1} a_i^{(k)} a_{\ell+i+1}^{(m)*} \frac{1}{T_c}\int_{T_c-\delta}^{T_c} h_c(t')h_c(t'-T_c+\delta)dt'
\end{aligned}
$$

(9.31)

The **continuous-time partial autocorrelation functions** of the chip waveform $h_c(t)$ are defined as [272]

$$
R_h(\delta) = \frac{1}{T_c}\int_0^{T_c-\delta} h_c(t')h_c(t'+\delta)dt'
$$

(9.32)

$$
\hat{R}_h(\delta) = \frac{1}{T_c}\int_{T_c-\delta}^{T_c} h_c(t')h_c(t'-T_c+\delta)dt'
$$

(9.33)

allowing us to write

$$
R_{k,m}(\tau) = \phi_{k,m}(\ell)R_h(\delta) + \phi_{k,m}(\ell+1)\hat{R}_h(\delta)
$$

(9.34)

where $\phi_{k.m}(\ell)$ is the full period cross-correlation defined in (9.20). As an example, if $h_c(t) = u_{T_c}(t)$, then

$$
R_{k,m}(\tau) = \phi_{k,m}(\ell)\left(1 - \frac{\delta}{T_c}\right) + \phi_{k,m}(\ell+1)\frac{\delta}{T_c} \ .
$$

(9.35)

When $G < N$, the partial correlations in (9.22) and (9.23) must be used. In this case the cross-correlation in (9.34) becomes a random variable that (for random spreading sequences) has mean and variance

$$
\mu_{R_{k,m}(\tau)} = \mu_{\phi_{k,m}(\ell)}R_h(\delta) + \mu_{\phi_{k,m}(\ell+1)}\hat{R}_h(\delta) = 0
$$

(9.36)

$$\sigma^2_{R_{k,m}(\tau)} = \sigma^2_{\phi_{k,m}(\ell)} R_h^2(\delta) + \sigma^2_{\phi_{k,m}(\ell+1)} \hat{R}_h^2(\delta)$$

$$= \frac{1}{G} \left(R_h^2(\delta) + \hat{R}_h^2(\delta) \right) . \qquad (9.37)$$

Likewise, the autocorrelation is also a random variable that (for random spreading sequences) has mean and variance

$$\mu_{R_{k,k}(\tau)} = \mu_{\phi_{k,k}(\ell)} R_h(\delta) + \mu_{\phi_{k,k}(\ell+1)} \hat{R}_h(\delta)$$

$$= \begin{cases} R_h(\delta), & \ell = iG \\ \hat{R}_h(\delta), & \ell + 1 = iG \\ 0, & \text{elsewhere} \end{cases} \qquad (9.38)$$

$$\sigma^2_{R_{k,k}(\tau)} = \sigma^2_{\phi_{k,k}(\ell)} R_h^2(\delta) + \sigma^2_{\phi_{k,k}(\ell+1)} \hat{R}_h^2(\delta)$$

$$= \begin{cases} R_h^2(\delta), & \ell = iG \\ \hat{R}_h^2(\delta), & \ell + 1 = iG \\ 1/G, & \text{elsewhere} \end{cases} \qquad (9.39)$$

where i is an integer.

2.2   *M*-SEQUENCES

One very well known class of spreading sequences are the maximal length sequences or *m*-sequences. As shown in Fig. 9.6, an *m*-sequence $\tilde{\mathbf{a}} = \{a_k\}$, $a_k \in \{0,1\}$, is generated by using a linear feedback shift register (LFSR) of length m. The sequence $\mathbf{a} = \{a_k\}$ is obtained by using the level shift $a_k = 2\tilde{a}_k - 1$. The feedback or connection polynomial is a primitive polynomial of degree m over GF(2), having the form

$$p(x) = 1 \oplus p_1 x \oplus p_2 x^2 \oplus p_3 x^3 \oplus \cdots \oplus p_{m-1} x^{m-1} \oplus x^m \qquad (9.40)$$

where $p_i \in \{0,1\}$ and $\oplus$ denotes modulo 2 addition. Tables of primitive polynomials, $p(x)$, are tabulated in many texts, e.g., [199]. Notice that $p_0 = 1$, since this represents the feedback connection tap. Also, $p_m = 1$; otherwise, if $p_m = 0$ we are effectively using a shift register of length less than m.

Maximal length sequences are by definition the longest sequences that can be generated by an LFSR of a given length. For a shift register of length m, a sequence of of length $N = 2^m - 1$ is generated. As a *m*-sequence generator cycles through one full period of length $N = 2^m - 1$, the contents of the *m*-stage shift register go through all possible $2^m - 1$ non-binary *m*-tuples values or states. The all-zeroes state is the only forbidden *m*-tuple, since the LFSI would lock in this state.

The *m*-sequences have many remarkable properties, and every full period of an *m*-sequence satisfies some important randomness properties. First, the
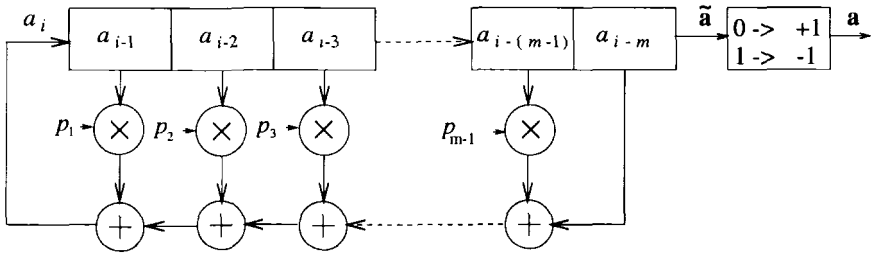
Figure 9.6. *m*-sequence generator.

sequence is balanced with 2^{m-1} ones and $2^{m-1} - 1$ zeros. A run is defined as a string of consecutive zeros or ones, and a sequence can be characterized in terms of its run length distribution. For *m*-sequences the number of runs of length P, n_P, is

$$n_P = \begin{cases} 2^{m-P-1} , & P = 1, 2, \ldots, m-1 \\ 1 , & P = m \end{cases} \tag{9.41}$$

The full period autocorrelation of an *m*-sequence is

$$\phi(n) = \begin{cases} 1 & , \ n = \ell N \\ -1/N & , \ n \neq \ell N \end{cases} . \tag{9.42}$$

For large values of N, $\phi(n) \approx \delta(n)$ so that m-sequences are almost ideal when viewed in terms of their full period autocorrelation. For a rectangular chip shaping function $h_c(t) = u_{T_c}(t)$, the corresponding spreading waveform $a(t)$ has autocorrelation function

$$R(\tau) = \phi(\ell) \left(1 - \frac{\delta}{T_c}\right) + \phi(\ell + 1)\frac{\delta}{T_c} . \tag{9.43}$$

This function is plotted in Fig. 9.7.

The mean and variance of the partial period autocorrelation of an *m*-sequence can be obtained in a straight forward fashion by replacing the expectations in (9.25) and (9.27) with averages over all possible starting positions. This gives

$$\mu_{\phi(n)} = \begin{cases} 1 & , \ n = \ell N \\ -1/G & , \ n \neq \ell N \end{cases} \tag{9.44}$$

$$\sigma^2_{\phi(n)} = \begin{cases} 0 & , \ n = \ell N \\ \frac{1}{G}\left(1 + \frac{1}{N}\right)\left(1 - \frac{G}{N}\right) & , \ n \neq \ell N \end{cases} . \tag{9.45}$$

Unfortunately, *m*-sequences also have a number of undesirable properties. First, the number of *m*-sequences that can be generated by a LFSR of length m
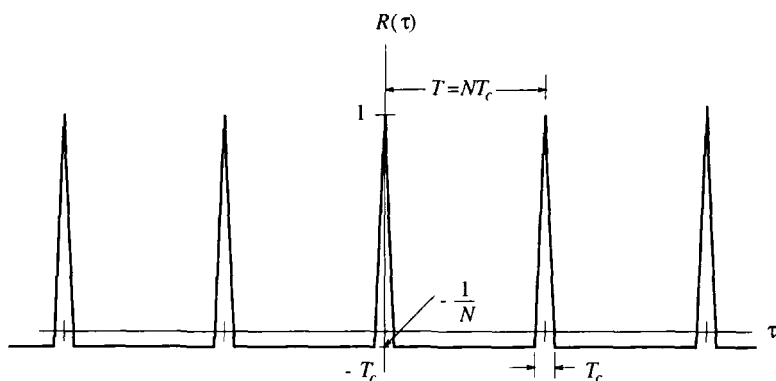
Figure 9.7. Typical full period autocorrelation function of an m-sequence spreading waveform.

is equal to the number of primitive polynomials of degree m over GF(2), and is given by $\Phi(2^m - 1)/m$, where $\Phi(\cdot)$ is the Euler Totient function

$$\Phi(n) = n \prod_{p|n} \left(1 - \frac{1}{p}\right) \tag{9.46}$$

where the product is over all primes p that divide n. Hence, there are relatively few m-sequences for a given shift register length m. Second, only for certain values of m, do there exist a few pairs of m-sequences with low full period cross-correlations. In general, m-sequences do not have good cross-correlation properties. Consider the full period cross-correlation $\phi_{k,m}(n)$ between two m-sequences $\mathbf{a}^{(k)}$ and $\mathbf{a}^{(m)}$. Let us define the average full period cross-correlation

$$\theta = \frac{1}{N} \sum_{n=0}^{N-1} \phi_{k,m}(n) \tag{9.47}$$

The value of θ depends on the particular pair of m-sequences that are selected. The best and worst case values of θ are shown in Table 9.1. Notice that the worst case full period cross-correlations are very large even for long sequence lengths.

2.3   GOLD SEQUENCES

A set of **Gold sequences** [141] consists of $2^m + 1$ sequences each with a period of $N = 2^m - 1$ that are generated by using a **preferred pair** of m-sequences obtained as follows. Let GF(2^m) be an extension field of GF(2). Let a be a primitive Nth root of unity in the extension field GF(2^m), where $N = 2^m - 1$. Let $p_1(x)$ and $p_2(x)$ be a pair of primitive polynomials over GF(2) each having degree m such that $p_1(\alpha) = 0$ and $p_2(\alpha^d) = 0$ for some integer d.

m	N	Number of m-sequences	θ Worst	θ Best
5	31	6	0.35	0.29
6	63	6	0.36	0.24
7	127	18	0.32	0.13
8	255	16	0.37	0.12
9	511	48	0.22	0.06
10	1023	60	0.37	0.06
11	2047	176	0.14	0.03
12	4095	144	0.34	0.03

Table 9.1.   Best and worst case average cross-correlations for m-sequences.

Consider the case when $m \neq 0 \bmod 4$. If $d = 2^h + 1$ or $d = 2^{2h} - 2^h + 1$ and if $e = \mathrm{GCD}(m, h)$ is such that m/e is odd, then $p_1(x)$ and $p_2(x)$ constitute a preferred pair of polynomials. Note that $p_2(x)$ may not be unique. For example, with $m = 5$, both $h = 1$ and $h = 2$ will work, so that we can choose $p_2(x^3) = 0$ or $p_2(x^5) = 0$. To find the corresponding polynomials we can refer to Peterson's table of irreducible polynomials [261]. The two m-sequences $\mathbf{a}^{(1)}$ and $\mathbf{a}^{(2)}$ that are generated by using $p_1(x)$ and $p_2(x)$ are known as a preferred pair of m-sequences. Their cross-correlation function is three-valued with the values $\{-1, -t(m), t(m) - 2\}$ where

$$t(m) = \begin{cases} 2^{(m+1)/2} + 1 & , \quad m \text{ odd} \\ 2^{(m+2)/2} + 1 & , \quad m \text{ even} \end{cases} . \tag{9.48}$$

By using the preferred pair of sequences $\mathbf{a}^{(1)}$ and $\mathbf{a}^{(2)}$, we can construct a set of Gold sequences by taking the sum of $\mathbf{a}^{(1)}$ with all cyclically shifted versions of $\mathbf{a}^{(2)}$ or vice versa. A typical Gold sequence generator is shown in Fig. 9.8, where the preferred pair of polynomials are $p_1(x) = 1 + x^2 + x^5$ and $p_2(x) = 1 + x + x^2 + x^4 + x^5$. This above procedure yields N new sequences each with period $N = 2^m - 1$. These sequences along with the original two sequences gives a set of $2^m + 1$ sequences.

It is important to note that all the $2^m + 1$ Gold sequences are balanced with 2^{m-1} ones and $2^{m-1} - 1$ zeros. In fact, it can be shown that only $2^m - 2^{m-e} - 1$ of the Gold sequences are balanced. The balanced Gold sequences are the most desirable. With the exception of the preferred pair of sequences $\mathbf{a}^{(1)}$ and $\mathbf{a}^{(2)}$, the Gold sequences are not m-sequences and, therefore, their autocorrelations are not two-valued. However, Gold sequences have three-valued off-peak autocorrelations and cross-correlations, with possible values $\{-1, -t(m), t(m) - 2\}$, where $t(m)$ is defined in (9.48).
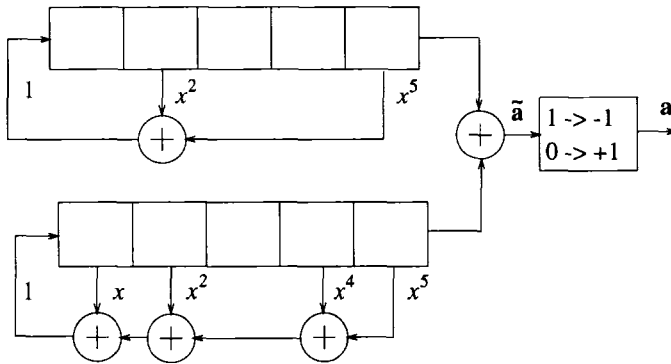
Figure 9.8. A Gold sequence generator with $p_1(x) = 1 + x^2 + x^5$ and $p_2(x) = 1 + x + x^2 + x^4 + x^5$. This sequence generator can produce 32 Gold sequences of length 31.

m	N	Number m sequences	Peak cross correlation	m-sequence $\phi_{max}/\phi(0)$	$t(m)$	Gold sequence $t(m)/\phi(0)$
3	7	2	5	0.71	5	0.71
4	15	2	9	0.60	9	0.60
5	31	6	11	0.35	9	0.29
6	63	6	23	0.36	17	0.27
7	127	18	41	0.32	17	0.13
8	255	16	95	0.37	33	0.13
9	511	48	113	0.22	33	0.06
10	1023	60	383	0.37	65	0.06
11	2047	176	287	0.14	65	0.03
12	4095	144	1407	0.34	129	0.03

Table 9.2. Peak cross-correlation of m-sequences and Gold sequences.

2.4   KASAMI SEQUENCES

The construction of **Kasami sequences** proceed as follows [177], [178]. Let m be even. Let $p_1(x)$ be a primitive polynomial over the binary field GF(2) with degree m and a as a root, and let $p_2(x)$ be the irreducible minimal polynomial of α^d where $d = 2^{m/2} + 1$. Once again, these polynomials can be identified using Peterson's table of irreducible polynomials [261]. Let $\mathbf{a}^{(1)}$ and $\mathbf{a}^{(2)}$ represent the two m-sequences of periods $2^m - 1$ and $2^{m/2} - 1$ that are generated by $p_1(x)$ and $p_2(x)$, respectively. The set of Kasami sequences is generated by using the two m-sequences in a fashion similar to the generation of Gold sequences, i.e., the set of Kasami sequences consists of the long sequence $\mathbf{a}^{(1)}$ and the sum of $\mathbf{a}^{(1)}$ with all $2^{m/2} - 1$ cyclic shifts of the short sequence $\mathbf{a}^{(2)}$. The number of Kasami sequences in the set is $2^{m/2}$, each having period $N = 2^m - 1$.
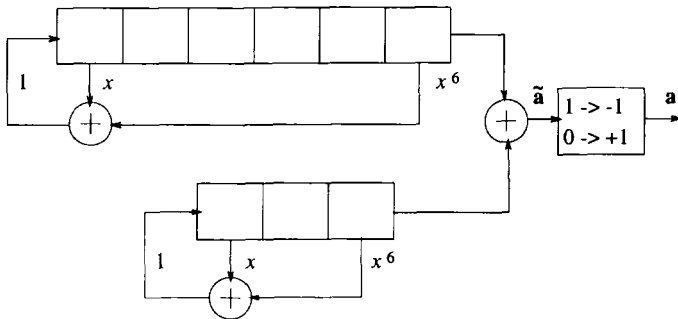
Figure 9.9. A Kasami sequence generator with $p_1(x) = 1 + x + x^6$ and $p_2(x) = 1 + x + x^3$. This sequence generator can produce 8 Kasami sequences of length 63.

In fact, this set is known as the *small set* of Kasami sequences. A typical Kasami sequence generator is shown in Fig. 9.9 with generator polynomials $p_1(x) = 1 + x + x^6$ and $p_2(x) = 1 + x + x^3$.

Like Gold sequences, the off-peak autocorrelation and cross-correlation functions of Kasami sequences are also three-valued, however, the possible values are $\{-1, -s(m), s(m) - 2\}$ where

$$s(m) = 2^{m/2} + 1 \ . \tag{9.49}$$

2.5 BARKER SEQUENCES

Barker sequences exist for lengths 2, 3, 4, 5, 7, 11, and 13, given as follows:

$$
\begin{aligned}
\mathbf{a} &= (+1 - 1) \\
\mathbf{a} &= (+1 + 1 - 1) \\
\mathbf{a} &= (+1 + 1 - 1 + 1) \\
\mathbf{a} &= (+1 + 1 + 1 - 1 + 1) \\
\mathbf{a} &= (+1 + 1 + 1 - 1 - 1 + 1 - 1) \\
\mathbf{a} &= (+1 + 1 + 1 - 1 - 1 - 1 + 1 - 1 - 1 + 1 - 1) \\
\mathbf{a} &= (+1 + 1 + 1 + 1 + 1 - 1 - 1 + 1 + 1 - 1 + 1 - 1 + 1)
\end{aligned}
$$

The mirror images (or time reversed) sequences are also Barker sequences. Barker sequences of other lengths do not exist.

Barker sequences are specially designed sequences that have almost ideal aperiodic autocorrelation functions, defined in (9.21). For the Barker sequences

$$\phi_{k,k}^a(n) = \begin{cases} 1 & , \quad n = 0 \\ 0, 1/N, \text{ or } -1/N & , \quad 1 \leq |n| \leq N - 1 \end{cases} \tag{9.50}$$

2.6 WALSH-HADAMARD SEQUENCES

Walsh-Hadamard sequences are obtained by selecting as sequences the rows of a **Hadamard matrix H**$_M$. For $M = 2$ the Hadamard matrix is

$$\mathbf{H}_2 = \begin{bmatrix} +1 & +1 \\ +1 & -1 \end{bmatrix} . \tag{9.51}$$

Larger Hadamard matrices are obtained by using the recursion

$$\mathbf{H}_{2M} = \begin{bmatrix} \mathbf{H}_M & \mathbf{H}_M \\ \mathbf{H}_M & -\mathbf{H}_M \end{bmatrix} . \tag{9.52}$$

For example,

$$\mathbf{H}_8 = \begin{bmatrix}
+1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\
+1 & -1 & +1 & -1 & +1 & -1 & +1 & -1 \\
+1 & +1 & -1 & -1 & +1 & +1 & -1 & -1 \\
+1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 \\
+1 & +1 & +1 & +1 & -1 & -1 & -1 & -1 \\
+1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 \\
+1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 \\
+1 & -1 & -1 & +1 & -1 & +1 & +1 & -1
\end{bmatrix} . \tag{9.53}$$

The rows in the Hadamard matrix define the Walsh-Hadamard sequences, and have the property that they are all orthogonal to each other.

The Walsh-Hadamard sequences can be used for orthogonal spreading, also called orthogonal CDMA, where the users are distinguished by assigning them different Walsh-Hadamard sequences, and the data symbols are sent by using simple binary spreading as shown in Fig. 9.4.. With orthogonal CDMA, the data symbols of the different users must be synchronized to within a small fraction of a chip period. This is because the Walsh-Hadamard sequences have very poor cross-correlations at non-zero lags. In fact, some of the Walsh-Hadamard sequences are just cyclic shifts of each other. Finally, multipath will also destroy the orthogonality of the received waveforms, because the Walsh Hadamard sequences have large off-peak autocorrelation values even at small lags.

2.6.1 ORTHOGONAL AND BI-ORTHOGONAL MODULATION

The Walsh Hadamard sequences can be used for modulation rather than spreading. There are several possibilities. One is M-ary orthogonal modulation, where $k = \log_2 M$ bits are used to select one of the M orthogonal waveforms for transmission. The signals can be detected coherently or non-coherently as discussed in Chapters 4 and 5. Another possibility is a variant of biorthogonal modulation, where each row of the Hadamard matrix is used to

$c_{4,1} = (1,1,1,1)$

$c_{2,1} = (1,1)$

$c_{1,1} = (1)$

$c_{4,2} = (1,1,-1,-1)$

$c_{4,3} = (1,-1,1,-1)$

$c_{2,2} = (1,-1)$

$c_{4,4} = (1,-1,-1,1)$

SF=1	SF=2	SF=4	(spreading factor)
$R = R_{max}$	$R = \dfrac{R_{max}}{2}$	$R = \dfrac{R_{max}}{4}$	(data rate)
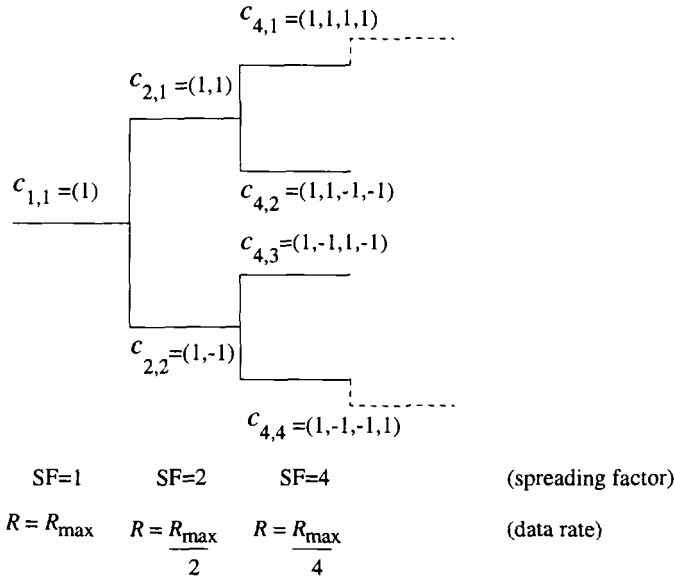
Figure 9.10. Construction of orthogonal spreading codes with different spreading factors.

send one bit of information. In this case M bits are sent at one time. This type of modulation requires coherent detection.

2.7 VARIABLE LENGTH ORTHOGONAL CODES

In multimedia applications it is necessary to support a variety of data services ranging from low to very high bit rates. Quite often these services are used concurrently and they all use the same spread bandwidth. Consider a system where each data symbol in the highest bit rate service $R = R_{max}$ is spread by an orthogonal sequence of length $N = 2^m$. Then the data symbols in a service with bit rate $R = R_{max}/2^k$ must be spread by a sequence of length 2^{m+k}. One way to achieve orthogonality between spreading sequences with different spreading factors is to use tree structured orthogonal codes. The construction of these codes is illustrated in Fig. 9.10. Tree-structured orthogonal codes are generated recursively according to the following:

$$
\mathbf{c}_{2n} = \begin{bmatrix} \mathbf{c}_{2n,1} \\ \mathbf{c}_{2n,2} \\ \vdots \\ \mathbf{c}_{2n,2n} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} \mathbf{c}_{n,1} & \mathbf{c}_{n,1} \\ \mathbf{c}_{n,1} & -\mathbf{c}_{n,1} \end{bmatrix} \\ \vdots \\ \begin{bmatrix} \mathbf{c}_{n,n} & \mathbf{c}_{n,n} \\ \mathbf{c}_{n,n} & -\mathbf{c}_{n,n} \end{bmatrix} \end{bmatrix} \tag{9.54}
$$

where $\mathbf{c}_{2n}$ is an orthogonal code set of size $2n$. The orthogonality properties are similar to Walsh-Hadamard sequences. In fact the set of sequences is identical, and only their order is different.

A code can be assigned for use if and only if no other code either on the path from the specific code to the root of the tree, or on the subtree produced by the specific code, is already being used. Hence, the total number of available codes is not fixed, but depends on the rate or spreading factor of each physical channel.

2.8 COMPLEMENTARY CODE KEYING (CCK)

Complementary codes have the property that the sum of their aperiodic autocorrelation functions are zero for all delays except zero delay. That is,

$$\frac{1}{M} \sum_{k=1}^{M} \phi_{k,k}^{a}(n) = \delta(n) \ . \tag{9.55}$$

A variety of constructions exist for complementary codes and two examples are given here.

The IEEE 802.11b standard uses CCK. For 11 Mb/s transmission length-8 sequences are used. The eight complex chip values for CCK code words are

$$\mathcal{C} = \left\{ e^{j(\phi_1 + \phi_2 + \phi_3 + \phi_4)}, e^{j(\phi_1 + \phi_3 + \phi_4)}, e^{j(\phi_1 + \phi_2 + \phi_4)}, \right.$$
$$\left. -e^{j(\phi_1 + \phi_4)}, e^{j(\phi_1 + \phi_2 + \phi_3)} e^{j(\phi_1 + \phi_3)} - e^{j(\phi_1 + \phi_2)} e^{j(\phi_1)} \right\} \tag{9.56}$$

where the phases $\{\phi_1, \phi_2, \phi_3, \phi_4\}$ are QPSK phases. The phases $\phi_2, \phi_3,$ and ϕ_4 each take on 4 different values, leading to a code alphabet of size 64. The phase ϕ_1 is differentially encoded across successive codewords. Since each of the 4 phases $\phi_1 - -\phi_4$ represents 2 bits of information, 8 bits are transmitted per codeword. The chip rate for IEEE 802.11 is 11 Mchips/s, so that the resulting bit rate is 11 Mb/s.

The IEEE 802.11b standard for 5.5 Mb/s transmission is similar but uses CCK with length-4 sequences. The complex chip values for the CCK code words are

$$\mathcal{C} = \left\{ e^{j(\phi_1 + \phi_2 + \phi_3)}, e^{j(\phi_1 + \phi_3)}, e^{j(\phi_1 + \phi_2)}, -e^{j(\phi_1)} \right\} \tag{9.57}$$

where, again, the phases $\{\phi_1, \phi_2, \phi_3\}$ are QPSK phases.

3. POWER SPECTRAL DENSITY OF DS SPREAD SPECTRUM SIGNALS

We seen earlier that the DS/QPSK signal can be thought of as a QPSK signal where the nth data symbol is shaped with the amplitude shaping pulse in (9.5).

For uncorrelated zero-mean data symbols, the results in Chapter 4 showed that the power spectral density (psd) of the complex envelope is (c.f. 4.206)

$$S_{\tilde{s}\tilde{s}}(f) = \frac{A^2}{T}\sigma_x^2 |H_a(f)|^2 \tag{9.58}$$

where $h_a(t)$ is the amplitude shaping pulse. In the case of a short code, the amplitude shaping pulse is

$$h_a(t) = \sum_{k=0}^{N-1} a_k h_c(t - kT_c) \ . \tag{9.59}$$

Taking the Fourier transform of $h_a(t)$ gives

$$H_a(f) = H_c(f) \sum_{k=0}^{N-1} a_k e^{-j2\pi fkT_c} \tag{9.60}$$

and

$$|H_a(f)|^2 = |H_c(f)|^2 \sum_{k=0}^{N-1}\sum_{\ell=0}^{N-1} a_k a_\ell^* e^{-j2\pi f(k-\ell)T_c} \tag{9.61}$$

The above expression can be put in a more convenient form by using the aperiodic autocorrelation defined in (9.21). It can be shown that

$$|H_a(f)|^2 = |H_c(f)|^2 \, 2N\Phi_{k,k}(f) \tag{9.62}$$

where $\Phi_{k,k}(f)$ is the discrete-time Fourier transform (DTFT) of the aperiodic autocorrelation function, defined by

$$\Phi_{k,k}(f) = \sum_{n=-N+1}^{N-1} \phi_{k,k}^a(n) e^{-j2\pi fnT_c} \ . \tag{9.63}$$

Using $T = NT_c$ and $\sigma_x^2 = \frac{1}{2}\mathrm{E}[|x_i|^2] = 1/2$, we can write

$$S_{\tilde{s}\tilde{s}}(f) = \frac{A^2}{T_c} |H_c(f)|^2 \, \Phi_{k,k}(f) \tag{9.64}$$

Observe that the psd depends on both $|H_c(f)|$ and $\Phi_{k,k}(f)$. Suppose the spreading sequence has an ideal "thumbtack" aperiodic autocorrelation function

$$\phi_{k,k}^a(n) = \begin{cases} 1, & n = 0 \\ 0, & n \neq 0 \end{cases} \tag{9.65}$$

Then $\Phi_{k,k}(f) = 1$ and

$$S_{\tilde{s}\tilde{s}}(f) = \frac{A^2}{T_c} |H_c(f)|^2 \tag{9.66}$$
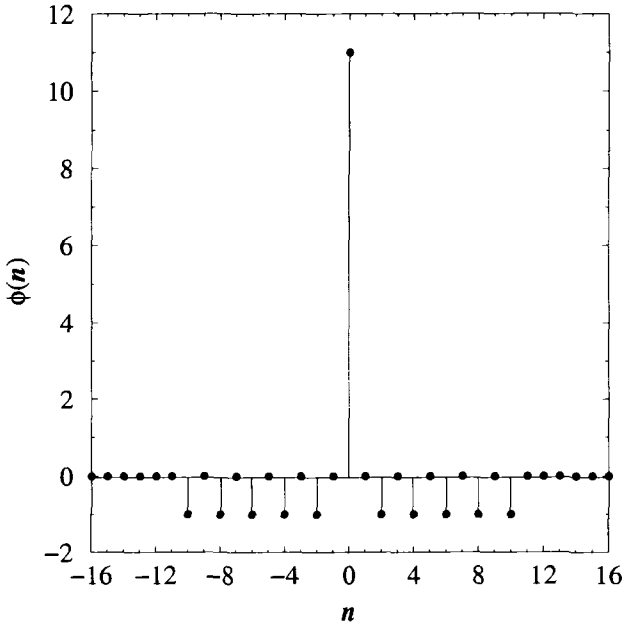
Figure 9.11.   Aperiodic autocorrelation function for the length-11 Barker sequence.

In this case, the psd depends only on the chip shaping response $|H_c(f)|$. For example, if $h_c(t) = u_{T_c}(t)$, then $H_c(f) = T_c\text{sinc}(fT_c)$ and $S_{\tilde{s}\tilde{s}}(f) = A^2 T_c \text{sinc}^2(fT_c)$. Unfortunately, no spreading sequences having the ideal aperiodic autocorrelation function in (9.65) exist for any non-trivial length.

Consider the following two spreading sequences

$$\mathbf{a}^{(1)} = (-1+1-1-1+1-1-1-1+1+1+1) \tag{9.67}$$

$$\mathbf{a}^{(2)} = (+1-1-1+1-1-1-1+1+1+1+1-1+1-1+1)$$

The first is a length-11 Barker sequence and the second is a length-15 *m*-sequence. The aperiodic autocorrelation functions for these sequences are shown in Figs. 9.11 and 9.12, respectively. The corresponding power spectral densities with a the rectangular chip shaping function $h_c(t) = u_{T_c}(t)$ are plotted in Figs. 9.13 and 9.14, respectively. Notice that the aperiodic autocorrelation of the m-sequence deviates significantly from the ideal function in (9.65). This leads to spectral peaks and nulls in Fig. 9.14. For wireless local area networks (LAN) that operate in unlicensed bands, such spectral peaks are highly undesirable. The length-11 Barker sequence is seen to provide a much smoother psd without any large peaks. For this reason, the length-11 Barker sequence has been chosen for the IEEE 802.11 wireless (LAN) specification.
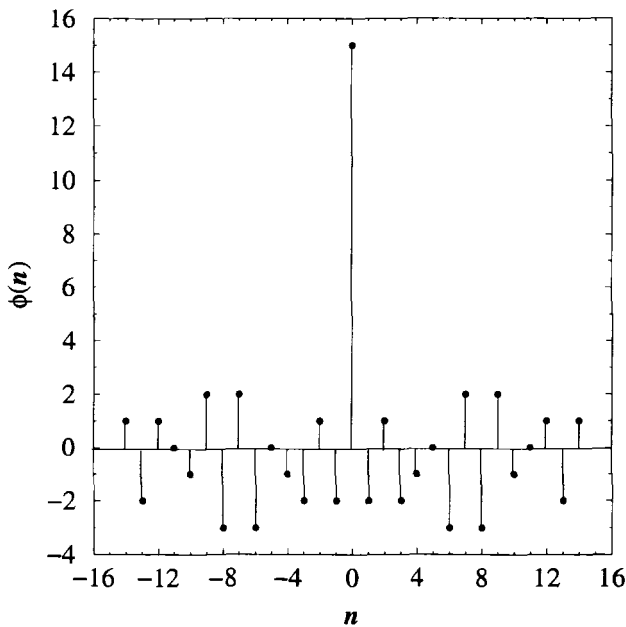
Figure 9.12. Aperiodic autocorrelation function for the length-15 m-sequence.

It is interesting to note that complementary code keying is used, then the psd depends on the DTFT of the average aperiodic autocorrelation function in (9.55). In this case,

$$\frac{1}{M} \sum_{k=1}^{M} \Phi_{k,k}(f) = 1$$

and the psd has the ideal form in (9.66).

Finally, if a long code is used, then the power spectrum must be obtained by averaging over all possible spreading code subsequences of length G. Usually, this will result in a "smoother" power density spectrum.

4. PERFORMANCE OF DS/QPSK IN TONE INTERFERENCE

Spread spectrum systems must often operate in the presence of narrowband interfering signals. In the United States, commercial spread spectrum systems operate in the unlicensed ISM (Instrumentation, Scientific, and Medical) bands according to FCC Part 15 spectral etiquette rules. The ISM bands are characterized by sources of narrowband interference. Military systems are often jammed with narrowband interference. Here we consider the effect of continuous wave (CW) tone interference on the performance of DS/QPSK.
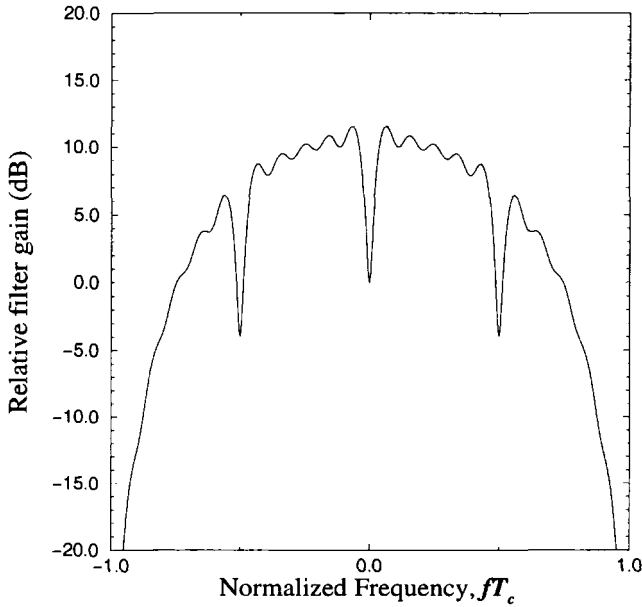
Figure 9.13.   Psd with the length-11 Barker sequence.

Consider a DS/QPSK system with dual-channel quaternary spreading as shown in Fig. 9.3. The bandpass DS/QPSK waveform is

$$
\begin{aligned}
s(t) \;=\; & A \sum_n \{ x_{I,n} h_{I,n}(t - nT) \cos(2\pi f_c t) \\
& - x_{Q,n} h_{Q,n}(t - nT) \sin(2\pi f_c t) \}
\end{aligned} \tag{9.68}
$$

where A is the amplitude. During time interval $[nT, (n + 1)T]$ the transmitted quaternary data symbol is $x_n = (x_{I,n}, x_{Q,n})$, $x_{I,n}, x_{Q,n} \in \{+1/\sqrt{2}, -1/\sqrt{2}\}$ and the spreading waveforms are

$$
h_{I,n}(t) \;=\; \sum_{k=0}^{G-1} a_{I,nG+k} h_c(t - nT_c) \tag{9.69}
$$

$$
h_{Q,n}(t) \;=\; \sum_{k=0}^{G-1} a_{Q,nG+k} h_c(t - nT_c) \tag{9.70}
$$

With dual-channel quaternary spreading, the energy per modulated symbol is
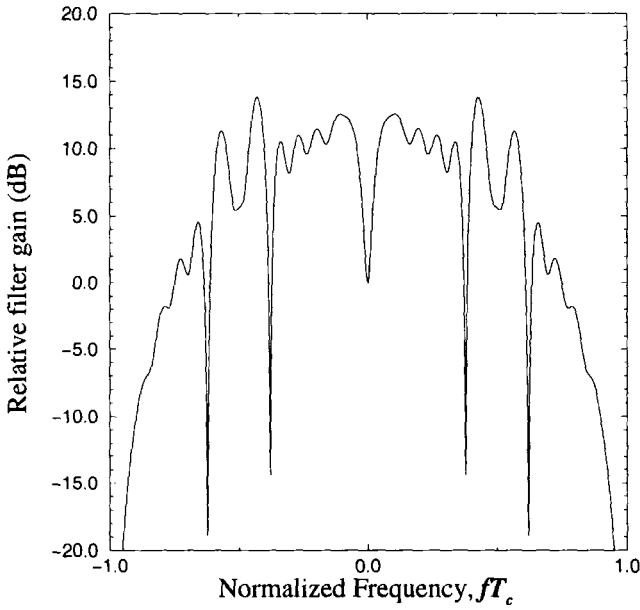
Figure 9.14. Psd with the length-15 m-sequence.

$$
\begin{aligned}
E &= \int_0^T s^2(t)dt \\
&= A^2 \int_0^T \left\{ x_{I,n}^2 h_{I,n}^2(t) \cos^2 2\pi f_c t + x_{Q,n}^2 h_{Q,n}^2(t) \sin^2 2\pi f_c t \right\} \\
&= \frac{A^2}{4} \int_0^T \left\{ h_{I,n}^2(t) + h_{Q,n}^2(t) \right\} \\
&= \frac{A^2}{4} \sum_{k=0}^{G-1} \left(a_{I,nG+k}^2 + a_{Q,nG+k}^2 \right) \int_0^{T_c} h_c^2(t)dt \\
&= G\frac{A^2}{2} \int_0^{T_c} h_c^2(t)dt \\
&= GE_c
\end{aligned}
\tag{9.71}
$$

where

$$
E_c = \frac{A^2}{2} \int_0^{T_c} h_c^2(t)dt
\tag{9.72}
$$

is the energy per PN chip. Note that (9.72) and (9.2) differ by a factor of 2, because (9.2) assumes complex spreading while (9.72) assumes quadra-

ture spreading. This can be seen by comparing the energy of the bandpass waveforms in (9.68) and (9.68) over the interval $[nT, (n+1)T]$.

The received bandpass signal in the presence of tone interference and additive white Gaussian noise (AWGN) is

$$r(t) = s(t) + n(t) + J(t) \tag{9.73}$$

where $n(t)$ is AWGN with two-sided power spectral density $N_o/2$ and $J(t)$ is the tone interference of the form

$$J(t) = A_J \cos(2\pi f_J t + \theta) \tag{9.74}$$

where A_J is the tone amplitude, f_J is its frequency, and θ is a random phase uniformly distribution on the interval $[-\pi, \pi]$. The tone energy in a time interval of duration T is

$$E_J = \frac{A_J^2 T}{2} . \tag{9.75}$$

The received signal is despread and processed with the quadrature demodulator shown in Fig. 9.15 to generate the decision variables Z_I and Z_Q. To derive the values of Z_I and Z_Q we consider the signal, noise, and interference separately. During the time interval $[nT, (n+1)T]$ the contribution of the *signal term* to Z_I and Z_Q is

$$
\begin{aligned}
Z_I(s) &= \int_0^T \tilde{s}_I(t) \cdot \sqrt{\frac{A^2}{4E}} h_{I,n}(t) dt \\
&= \int_0^T A x_{I,n} h_{I,n}(t) \cdot \sqrt{\frac{A^2}{4E}} h_{I,n}(t) dt \\
&= x_{I,n} A \sqrt{\frac{A^2}{4E}} \int_0^T h_{I,n}^2(t) dt \\
&= x_{I,n} \sqrt{E}
\end{aligned}
\tag{9.76}
$$

where we have used (9.71). Likewise

$$Z_Q(s) = x_{Q,n} \sqrt{E} . \tag{9.77}$$

The contribution of the *AWGN term* to Z_I and Z_Q is

$$Z_I(n) = \int_0^T \tilde{n}_I(t) \sqrt{\frac{A^2}{4E}} h_{I,n}(t) dt \tag{9.78}$$

$$Z_Q(n) = \int_0^T \tilde{n}_Q(t) \sqrt{\frac{A^2}{4E}} h_{Q,n}(t) dt . \tag{9.79}$$
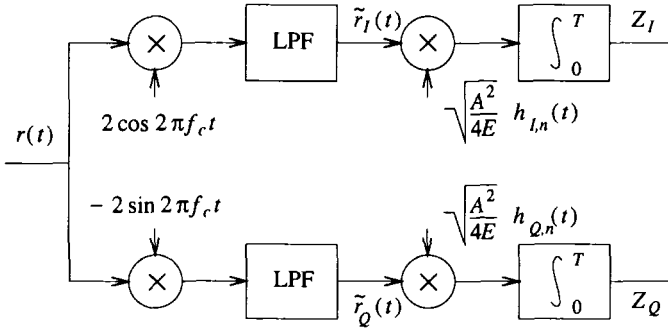
Figure 9.15. Quadrature demodulator for DS/QPSK.

It can be shown that $Z_I(n)$ **and** $Z_Q(n)$ are independent zero-mean Gaussian random variables with variance $N_o/2$.

Finally, the contribution of the *tone interference term* to Z_I **and** Z_Q can be calculated as follows:

$$
\begin{aligned}
Z_I(J) &= \int_0^T J(t) \cdot 2\cos(2\pi f_c t) \cdot \sqrt{\frac{A^2}{4E}} h_{I,n}(t)\,dt \\
&= \int_0^T A_J \cos(2\pi f_J t) \cdot 2\cos(2\pi f_c t) \cdot \sqrt{\frac{A^2}{4E}} h_{I,n}(t)\,dt \\
&= A_J \sqrt{\frac{A^2}{4E}} \int_0^T 2h_{I,n}(t)\cos(2\pi f_c t)\cos(2\pi f_J t + \theta)\,dt \\
&= A_J \sqrt{\frac{A^2}{4E}} \int_0^T h_{I,n}(t)\left\{\cos(2\pi\Sigma_f t + \theta) + \cos(2\pi\Delta_f t + \theta)\right\}\,dt
\end{aligned}
$$

$$(9.80)$$

where

$$\Sigma_f = f_c + f_J \qquad (9.81)$$
$$\Delta_f = f_J - f_c \; . \qquad (9.82)$$

Using $A_J = \sqrt{2E_J/T}$, we can write

$$Z_I(J) = \sqrt{E_J} \sqrt{\frac{A^2 T}{2E}} \frac{1}{T} \int_0^T h_{I,n}(t)\left\{\cos(2\pi\Sigma_f t + \theta) + \cos(2\pi\Delta_f t + \theta)\right\}\,dt$$

$$(9.83)$$

Finally, using

$$E = G\frac{A^2}{2}\int_0^{T_c} h_c^2(t)\,dt \qquad (9.84)$$

we can write

$$Z_I(J) = \sqrt{E_J/\bar{h}_c}\frac{1}{T}\int_0^T h_{I,n}(t)\left\{\cos(2\pi\Sigma_f t + \theta) + \cos(2\pi\Delta_f t + \theta)\right\}dt \ .$$
(9.85)

where

$$\bar{h}_c = \frac{1}{T_c}\int_0^{T_c} h_c(t)dt \ .$$
(9.86)

Using further trigonometric identities we can write

$$
\begin{aligned}
Z_I(J) \ = \ & \sqrt{E_J/\bar{h}_c}\left\{\cos\theta\frac{1}{T}\int_0^T h_{I,n}(t)\left(\cos(2\pi\Sigma_f t) + \cos(2\pi\Delta_f t)\right)dt\right. \\
& \left. -\sin\theta\frac{1}{T}\int_0^T h_{I,n}(t)\left(\sin(2\pi\Sigma_f t) + \sin(2\pi\Delta_f t)\right)dt\right\} \ .
\end{aligned}
$$
(9.87)

In a similar fashion

$$
\begin{aligned}
Z_I(J) \ = \ & \int_0^T J(t)\cdot 2\sin(2\pi f_c t)\cdot\sqrt{\frac{A^2}{4E}}h_{Q,n}(t)dt \\
= \ & \sqrt{E_J/\bar{h}_c}\left\{\cos\theta\frac{1}{T}\int_0^T h_{Q,n}(t)\left(\sin(2\pi\Delta_f t) - \sin(2\pi\Sigma_f t)\right)dt\right. \\
& \left. +\sin\theta\frac{1}{T}\int_0^T h_{Q,n}(t)\left(\cos(2\pi\Delta_f t) - \sin(2\pi\Sigma_f t)\right)dt\right\} \ .
\end{aligned}
$$
(9.88)

Combining the signal, noise, and tone interference terms

$$
\begin{aligned}
Z_I &= Z_I(s) + Z_I(n) + Z_I(J) \\
Z_Q &= Z_Q(s) + Z_Q(n) + Z_Q(J) \ .
\end{aligned}
$$
(9.89)

It follows that Z_I and Z_Q are independent Gaussian random variables with variance $N_o/2$ and means

$$
\begin{aligned}
\mathrm{E}[Z_I] &= x_{I,n}\sqrt{E} + I_I\sqrt{E_J/\bar{h}_c} \\
\mathrm{E}[Z_Q] &= x_{Q,n}\sqrt{E} + I_Q\sqrt{E_J/\bar{h}_c}
\end{aligned}
$$
(9.90)

where

$$
\begin{aligned}
I_I \ = \ & \cos\theta\frac{1}{T}\int_0^T h_{I,n}(t)\left(\cos(2\pi\Sigma_f t) + \cos(2\pi\Delta_f t)\right)dt \\
& -\sin\theta\frac{1}{T}\int_0^T h_{I,n}(t)\left(\sin(2\pi\Sigma_f t) + \sin(2\pi\Delta_f t)\right)dt
\end{aligned}
$$
(9.91)

$$
\begin{aligned}
I_Q \ = \ & \cos\theta\frac{1}{T}\int_0^T h_{Q,n}(t)\left(\sin(2\pi\Delta_f t) - \sin(2\pi\Sigma_f t)\right)dt \\
& +\sin\theta\frac{1}{T}\int_0^T h_{Q,n}(t)\left(\cos(2\pi\Delta_f t) - \sin(2\pi\Sigma_f t)\right)dt \ .
\end{aligned}
$$
(9.92)

Error probability with a short code:. For the purpose of illustration assume a rectangular chip shaping pulse $h_c(t) = u_{T_c}(t)$ so that $\bar{h}_c = 1$ in (9.90), and assume a short code ($G = N$) so that each data symbol is spread by the same sequence. Furthermore, assume that the same spreading sequence is used on the inphase and quadrature channels so that

$$h(t) = h_{I,n}(t) = h_{Q,n}(t) = \sum_{k=0}^{G-1} a_k u_{T_c}(t - kT_c) \qquad (9.93)$$

$$(9.94)$$

It follows that

$$
\begin{aligned}
I_I &= \cos\theta \frac{1}{NT_c} \int_0^{NT_c} \sum_{k=0}^{N-1} a_k u_{T_c}(t - kT_c) \\
&\quad \times (\cos(2\pi\Sigma_f t) + \cos(2\pi\Delta_f t))\, dt \\
&\quad - \sin\theta \frac{1}{NT_c} \int_0^{NT_c} \sum_{k=0}^{N-1} a_k u_{T_c}(t - kT_c) \\
&\quad \times (\sin(2\pi\Sigma_f t) + \sin(2\pi\Delta_f t))\, dt \\
&= \frac{1}{N} \sum_{k=0}^{N-1} a_k \left\{ \cos\theta \int_k^{k+1} (\cos(2\pi\Sigma_f T_c t) + \cos(2\pi\Delta_f T_c t))\, dt \right. \\
&\quad \left. - \sin\theta \int_k^{k+1} (\sin(2\pi\Sigma_f T_c t) + \sin(2\pi\Delta_f T_c t))\, dt \right\} . \qquad (9.95)
\end{aligned}
$$

Likewise

$$
\begin{aligned}
I_Q &= \frac{1}{N} \sum_{k=0}^{N-1} a_k \left\{ \cos\theta \int_k^{k+1} (\sin(2\pi\Delta_f T_c t) - \sin(2\pi\Sigma_f T_c t))\, dt \right. \\
&\quad \left. + \sin\theta \int_k^{k+1} (\cos(2\pi\Delta_f T_c t) - \sin(2\pi\Sigma_f T_c t))\, dt \right\} . \qquad (9.96)
\end{aligned}
$$

Fortunately, the above integrals exist in closed form. Defining

$$\alpha = 2\pi\Sigma_f T_c \qquad (9.97)$$
$$\beta = 2\pi\Delta_f T_c \qquad (9.98)$$

we have

$$
\begin{aligned}
I_I = \frac{1}{N} \sum_{k=0}^{N-1} a_k \Bigg\{ \cos\theta \left(\frac{\sin((k+1)\alpha) - \sin(k\alpha)}{\alpha} \right. \\
+ \left. \frac{\sin((k+1)\beta) - \sin(k\beta)}{\beta} \right) \\
- \sin\theta \left(\frac{\cos(k\alpha) - \cos((k+1)\alpha)}{\alpha} \right. \\
+ \left. \frac{\cos(k\beta) - \cos((k+1)\beta)}{\beta} \right) \Bigg\}
\end{aligned}
$$

$$(9.99)$$

and

$$
\begin{aligned}
I_Q = \frac{1}{N} \sum_{k=0}^{N-1} a_k \Bigg\{ \cos\theta \left(\frac{\cos(k\beta) - \cos((k+1)\beta)}{\beta} \right. \\
- \left. \frac{\cos(k\alpha) - \cos((k+1)\alpha)}{\alpha} \right) \\
+ \sin\theta \left(\frac{\sin((k+1)\beta) - \sin(k\beta)}{\beta} \right. \\
- \left. \frac{\cos(k\alpha) - \cos((k+1)\alpha)}{\alpha} \right) \Bigg\} \quad .
\end{aligned}
$$

$$(9.100)$$

Due to the random phase of the tone interferer, the tone interference circularly symmetric, similar to the AWGN. This allows us to rotate the signal constellation for the purpose of calculating the bit error probability. The rotated constellation is shown in Fig. 9.16. In the absence of tone interference, the probability of correct symbol reception is

$$P(c) = (1 - P_b)^2 \tag{9.101}$$

where

$$P_b = Q\left(\sqrt{2\gamma_b}\right) \tag{9.102}$$

is the bit error probability, and $\gamma_b = E_b/N_o$ is the received bit energy-to-noise ratio.

The probability of correct reception is

$$P_{C|b_0 b_1} = (1 - P_{b1})(1 - P_{b2}) \quad . \tag{9.103}$$

When tone interference is present, the error probability depends on the transmitted symbol and the interference impairment I_I and I_Q. Referring to Fig. 9.16

$$P_{C|00} = (1 - P_{b1})(1 - P_{b2}) \tag{9.104}$$
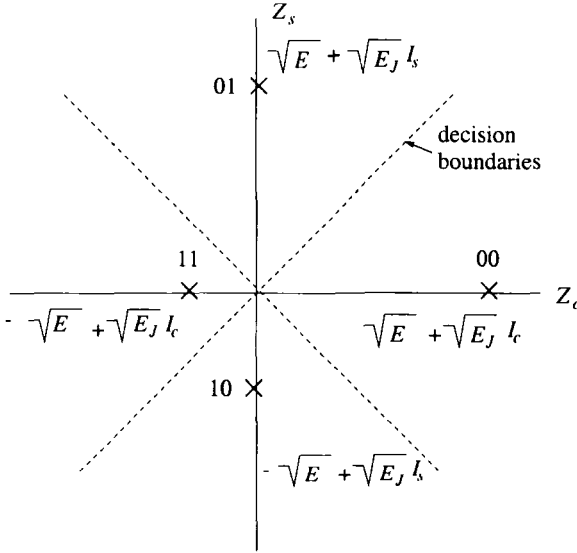
Figure 9.16.   QPSK signal constellation with tone interference.

where

$$P_{b1} = P_{b2} = Q\left(\sqrt{\frac{2E_b}{N_o}\left(1 + 2\sqrt{E_J/E}I_I + (E_J/E)^2 I_I^2\right)}\right) \ . \quad (9.105)$$

Hence, we can write

$$P_{b|00} = Q\left(\sqrt{\frac{2E_b}{N_o}\left(1 + 2\sqrt{E_J/E}I_I + (E_J/E)^2 I_I^2\right)}\right) \ . \quad (9.106)$$

In a similar fashion,

$$P_{b|01} = Q\left(\sqrt{\frac{2E_b}{N_o}\left(1 + 2\sqrt{E_J/E}I_Q + (E_J/E)^2 I_Q^2\right)}\right) \quad (9.107)$$

$$P_{b|11} = Q\left(\sqrt{\frac{2E_b}{N_o}\left(1 - 2\sqrt{E_J/E}I_I + (E_J/E)^2 I_I^2\right)}\right) \quad (9.108)$$

$$P_{b|10} = Q\left(\sqrt{\frac{2E_b}{N_o}\left(1 - 2\sqrt{E_J/E}I_Q + (E_J/E)^2 I_Q^2\right)}\right) \ . \quad (9.109)$$

Since all symbols are equally likely, the bit error probability is

$$P_b = \frac{1}{4}(P_{b|00} + P_{b|11} + P_{b|10} + P_{b|01}) \ . \quad (9.110)$$
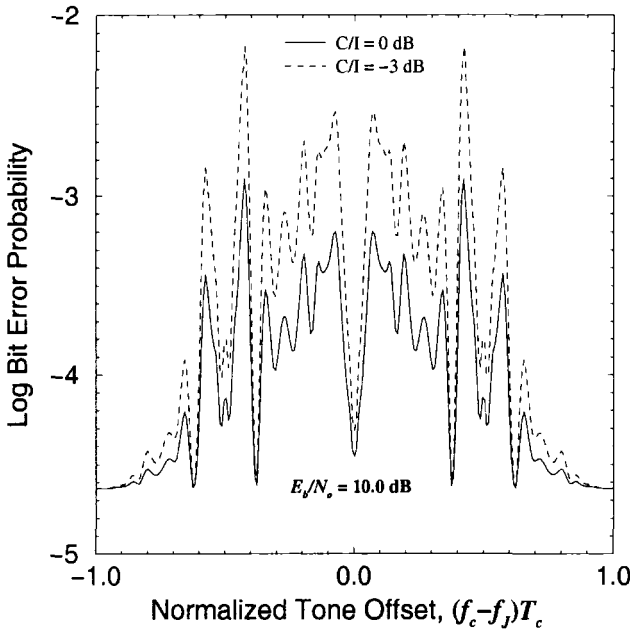
Figure 9.17.   Bit error probability with length-15 m-sequence.

Observe that the bit error probability depends on I_I and I_Q. However I_I and I_Q are random variables due to the random phase θ. Therefore, the bit error probability must calculated by averaging (9.110) over random phase of the tone interferer.

Fig. 9.17 shows the bit error probability when the length-15 m-sequence, $\mathbf{a}^{(2)}$ in (9.67) is used as a short code ($G = 15$). Fig 9.17 arbitrarily assumes that $f_c = 280$ MHz, and $T_c = 191 \times 10^{-9}$. Observe that the bit error probability varies greatly with the frequency of the tone interferer. It is interesting to note that an interfering at the carrier frequency f_c is not the worst case. Also, the bit error probability is seen to exhibit an error floor due to the AWGN.

Fig. 9.18 shows the bit error probability when the length-11 Barker sequence $\mathbf{a}^{(1)}$ in (9.67) is used as a short code ($G = 11$). Observe that the length-11 Barker sequence generally has worse performance for the same E/E_J than the length-15 m-sequence, except at frequencies where the length-15 m-sequence is highly sensitive to tone interference. This is because the length-11 Barker sequence has a lower processing gain compared to the length-15 m-sequence.

Fig. 9.19 inverts Fig. 9.17 and plots the E/E_J required to achieve a bit error rate of 10^{-6} with the length-15 m-sequence in the presence of a single tone interferer and AWGN. Likewise, Fig. 9.20 inverts Fig. 9.18 for the length-11
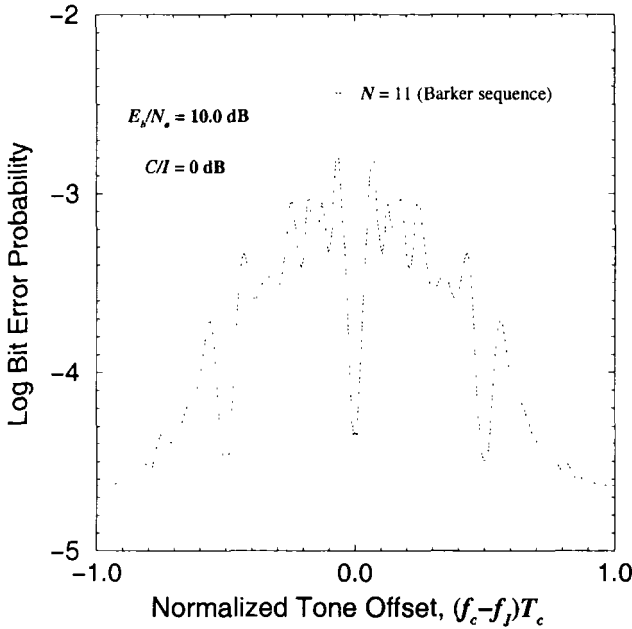
Figure 9.18. Bit error probability with the length-11 Barker sequence.

Barker sequence. Observe that the sensitivity to tone interference is much less with the Barker sequence.

The sensitivity of the error probability to the frequency of the tone interferer can be explained as follows. The data symbols on the inphase and quadrature channels are spread by using the amplitude shaping pulse

$$h(t) = \sum_{k=0}^{N-1} a_k h_c(t - kT_c) \tag{9.111}$$

where $\{x_k\}_{k=0}^{N-1}$ is the periodic spreading sequence of length N. After quadrature demodulation the receiver employs a correlator or matched filter detector with impulse response[3]

$$
\begin{aligned}
h_r(t) &= h^*(NT_c - t) \\
&= \sum_{k=0}^{N-1} a_{N-k} h_c(t - kT_c) \ . \tag{9.112}
\end{aligned}
$$

---

[3] We assume the usual case where $h_c(-t) = h_c(t)$.
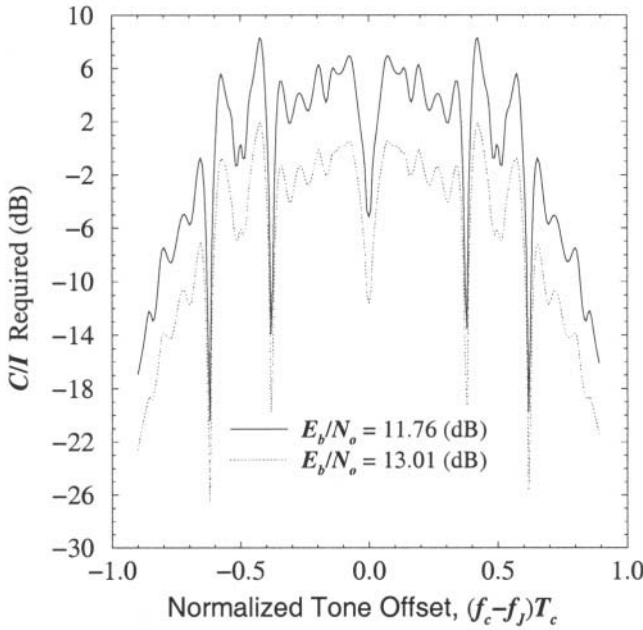
Figure 9.19. Required C/I to achieve 10^{-6} bit error rate with a length-15 m-sequence.

The filter $h_r(t)$ has transfer function

$$
\begin{aligned}
H_r(f) &= \int_{-\infty}^{\infty} h_r(t)e^{-j2\pi ft}dt \\
&= \int_{-\infty}^{\infty} \sum_{k=0}^{N-1} a_{N-k}h_c(t-kT_c)e^{-j2\pi ft}dt \\
&= \sum_{k=0}^{N-1} a_{N-k} \int_{-\infty}^{\infty} h_c(t-kT_c)e^{-j2\pi ft}dt \\
&= H_c(f) \sum_{k=0}^{N-1} a_{N-k}e^{-j2\pi fkT_c} \\
&= H_c(f)A(f) \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad (9.113)
\end{aligned}
$$

where

$$
A(f) = \sum_{k=0}^{N-1} a_{N-k}e^{-j2\pi fkT_c} \quad . \quad\quad\quad (9.114)
$$

For a rectangular chip shaping function $h_c(t) = u_{T_c}(t)$,

$$
H_c(f) = \mathrm{Sa}(\pi fT_c)e^{-j\pi fT_c} \quad\quad\quad (9.115)
$$
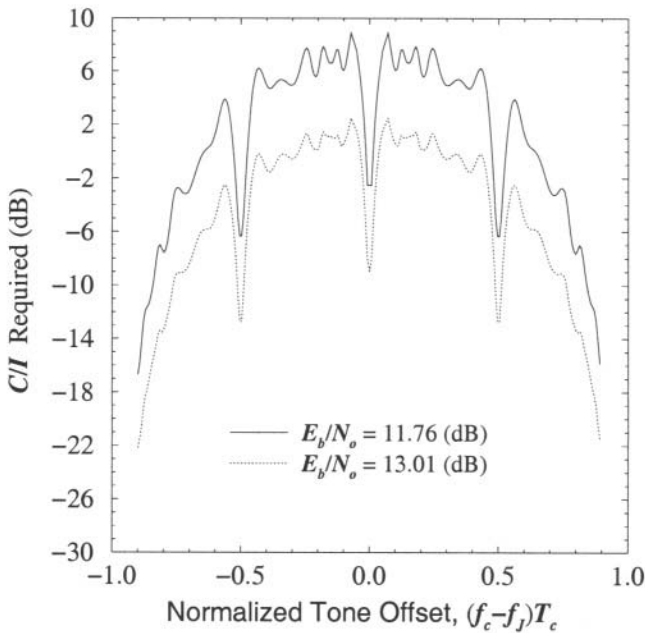
Figure 9.20. Required C/I to achieve 10^{-6} bit error rate with a length-11 Barker sequence.

The corresponding amplitude response $|H_r(f)|$ for the length-15 *m*-sequence and the length-11 Barker sequence are identical in form to the corresponding transmitted psds shown in Fig. 9.14 and 9.13, respectively. The frequencies where $|H_r(f)|$ is has the highest relative gain are the exact same frequencies where the tone interferer causes a large error probability. If the length of the short code is increased, the sensitivity to tone interference will not necessarily diminish. To make the receiver less sensitive to tone interference, we must ensure that the period autocorrelation function in (9.21) is a close to ideal as possible. In other words, the power spectrum $\Phi_{k,k}(f)$ defined in (9.63) is as flat as possible. Although some types of sequences, such as Gold and Kasami sequences, have excellent cross-correlation properties, there aperiodic autocorrelation functions are usually far from ideal. Their corresponding power spectrum $\Phi_{k,k}(f)$ is typically full of peaks and nulls. The Barker sequences have the best aperiodic autocorrelation properties and will result in the least senstivity to tone interference.

Error probability with a long code:. With a long code each data symbol is spread with a subsequence of a long PN sequence. In this case, the error probability must be averaged over the starting phase of the PN subsequence that is used to spread each data symbol.

Fig. 9.21 shows the effect of using a long PN sequence. Three cases are considered; a length-31 *m* sequence with generator polynomial $1 + x^2 + x^5$, a length-127 *m*-sequence with generator polynomial $1 + x + x^7$, and a length-2047 *m*-sequence with generator polynomial $1 + x + x^{11}$. Fig 9.22 shows the length-63 *m*-sequence with generator polynomial $1 + x + x^6$. The processing gain in each case is $G = 15$ chips/symbol. For the length-63 *m*-sequence, 15 and 63 have a common factor of 3 and, therefore, there are three different sets of subsequences to consider.

Observe that the bit error probability with a long code is less sensitive to the tone frequency as compared to a short code. For sequence lengths of 127 and 2047, the bit error probability is maximized when $f_J = f_c$. For all three sequence lengths, there are still some spectral irregularities, because the length of the shift register (5, 7, and 11) that is used to generate the PN sequence is less than the processing gain (15). Hence, the data symbols are not spread will all possible binary N-tuples, thus leading to the irregularities observed in Fig. 9.21. It is interesting to note that the length-2047 *m*-sequence seems to be more sensitive to an on-carrier tone than the length-127 *m*-sequence. The reason is that the length-15 subsequences of the length-127 *m*-sequence tend to be more balanced (equal number of -1's and 1's) than the length-15 subsequences of the length-2047 *m*-sequence.

Finally, comparison of Figs. 9.17 and 9.21 leads to the observation that the bit error probability with the short length-15 PN sequence is worse than that realized with a long PN sequence (e.g., the length-127 *m*-sequence) only at 4 narrow ranges of tone frequencies.

5.   DS SPREAD SPECTRUM ON FREQUENCY-SELECTIVE FADING CHANNELS

Suppose that the DS complex envelope $\tilde{s}(t)$ is strictly bandlimited to a bandwidth of W/2 Hz, by using for example spectral raised cosine pulse shaping. Since the low-pass signal $\tilde{s}(t)$ is band-limited to $|f| \le W/2$, the sampling theorem can be invoked and $\tilde{s}(t)$ can be completely described by the set of complex samples $\{\tilde{s}(n/W)\}_{n=-\infty}^{\infty}$. The sampled version of $\tilde{s}(t)$ is

$$\tilde{s}_\delta(t) \;=\; \sum_{n=-\infty}^{\infty} \tilde{s}\left(\frac{n}{W}\right) \delta\left(t - \frac{n}{W}\right) \tag{9.116}$$

$$=\; \tilde{s}(t) \sum_{n=-\infty}^{\infty} \delta\left(t - \frac{n}{W}\right) \; . \tag{9.117}$$

Taking the Fourier transform of both sides of (9.117) gives

$$\tilde{S}_\delta(f) \;=\; \tilde{S}(f) * W \sum_{n=-\infty}^{\infty} \delta(f - nW)$$
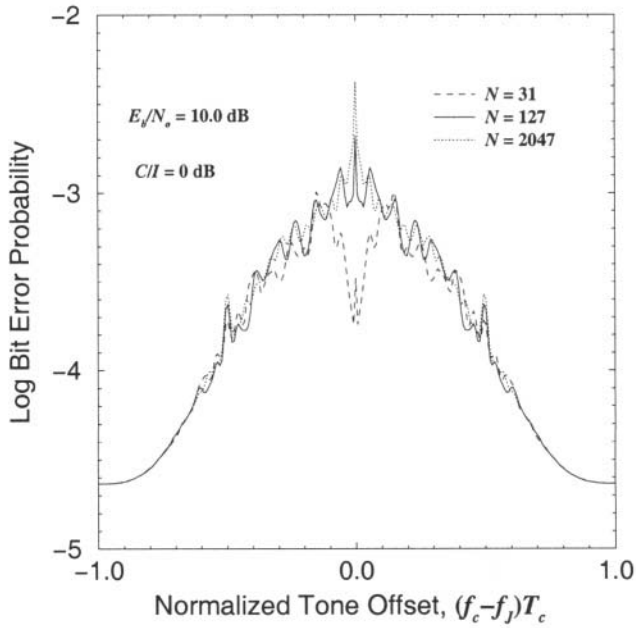
Figure 9.21. Bit error probability with length-31, 127 and 2047 *m*-sequences.

$$= W \sum_{n=-\infty}^{\infty} \tilde{S}(f) * \delta(f - nW)$$

$$= W \sum_{n=-\infty}^{\infty} \tilde{S}(f - nW) . \qquad (9.118)$$

From (9.118) we can see that

$$\tilde{S}(f) = \frac{1}{W} \tilde{S}_\delta(f) \qquad 0 \le |f| \le W/2 . \qquad (9.119)$$

Another useful expression can be obtained by taking the Fourier transform of both sides of (9.116) giving

$$\tilde{S}_\delta(f) = \sum_{n=-\infty}^{\infty} \tilde{s}\left(\frac{n}{W}\right) e^{-j2\pi nf/W} . \qquad (9.120)$$

Combining (9.119) and (9.120) gives

$$\tilde{S}(f) = \frac{1}{W} \sum_{n=-\infty}^{\infty} \tilde{s}\left(\frac{n}{W}\right) e^{-j\pi nf/W} \qquad 0 \le |f| \le W/2 . \qquad (9.121)$$

$E_s/N_o = 10.0$ dB

$C/I = 0$ dB

—— Phase 1
··· Phase 2
- - - Phase 3
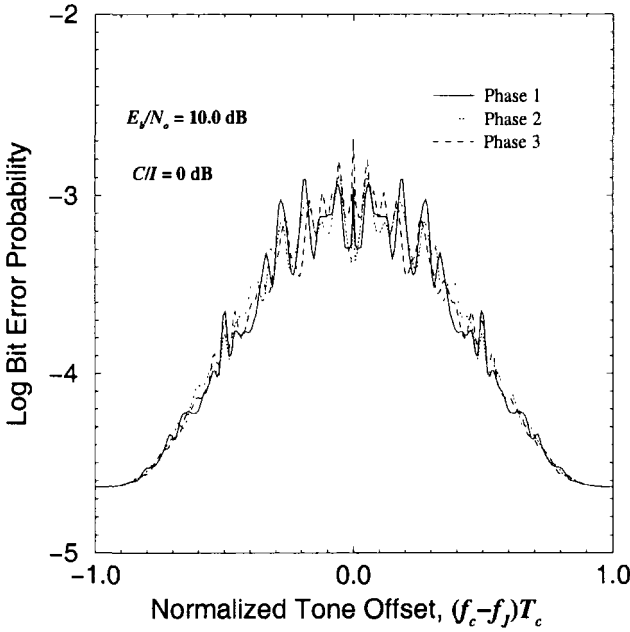
Figure 9.22.   Bit error probability with length-63 m-sequence.

If the low-pass signal $\tilde{s}(t)$ is transmitted over a multipath fading channel with time-variant transfer function $T(f, t)$, the received (noiseless) complex envelope is

$$\tilde{r}(t) = \int_{-\infty}^{\infty} \tilde{S}(f)T(f,t)e^{j2\pi ft}df \ . \tag{9.122}$$

Substituting $\tilde{S}(f)$ from (9.121) gives

$$
\begin{aligned}
\tilde{r}(t) &= \frac{1}{W} \sum_{n=-\infty}^{\infty} \tilde{s}\left(\frac{n}{W}\right) \int_{-\infty}^{\infty} T(f,t)e^{-j2\pi f(t-n/W)}df \\
&= \frac{1}{W} \sum_{n=-\infty}^{\infty} \tilde{s}\left(\frac{n}{W}\right) g\left(t - \frac{n}{W}, t\right) \\
&= \frac{1}{W} \sum_{n=-\infty}^{\infty} \tilde{s}\left(t - \frac{n}{W}\right) g\left(\frac{n}{W}, t\right)
\end{aligned}
\tag{9.123}
$$

where $g(\tau, t)$ is the time-variant impulse response of the channel. By defining

$$g_n(t) = \frac{1}{W} g\left(\frac{n}{W}, t\right) \tag{9.124}$$
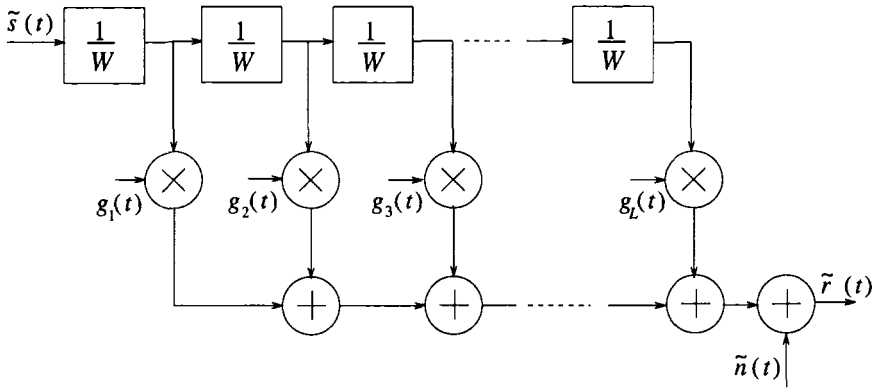
Figure 9.23. Tapped delay line model of a frequency selective fading channel, from [270].

the noiseless received complex envelope can be written as

$$\tilde{r}(t) = \sum_{n=-\infty}^{\infty} g_n(t)\tilde{s}\left(t - \frac{n}{W}\right) \qquad (9.125)$$

and it follows that the complex low-pass impulse response of the channel is

$$g(t, \tau) = \sum_{n=-\infty}^{\infty} g_n(t)\delta\left(\tau - \frac{n}{W}\right) . \qquad (9.126)$$

For WWSUS channels, the $\{g_n(t)\}$ in (9.124) are independent complex Gaussian random processes. For all practical purposes, the channel will be causal with an impulse response that is nonzero over a time interval of duration $T_{\max}$. In this case, $g_n(t) = 0$, $n \leq 0, n > L$, where $L = \lfloor T_{\max}/W \rfloor + 1$ and $\lfloor x \rfloor$ is the smallest integer greater than x. It follows that the channel impulse response is

$$g(t, \tau) = \sum_{n=1}^{L} g_n(t)\delta\left(\tau - \frac{n}{W}\right) . \qquad (9.127)$$

In conclusion, the frequency selective fading channel can be modeled as an L-tap, $1/W$-spaced, tapped delay line with tap gain vector

$$\mathbf{g}(t) = (g_1(t), g_2(t), \dots, g_L(t))$$

as shown in Fig. 9.23. It should be emphasized that the channel vector $\mathbf{g}(t)$ is *not* the same as the channel vector $\mathbf{g}_T(t)$ associated with the T-spaced discrete-time white noise channel model in Chapter 7.

If ideal Nyquist chip amplitude pulse shaping is used such that $h_c(t) = S_a(\pi t/T_c)$, then $W = 1/T_c$ and the channel can be represented as a T_c-spaced

or chip-spaced tapped delay line. Such a model is very convenient because it leads to a simplified analysis. However, if any other pulse shape is used, such as a raised cosine pulse, then the tapped delay line channel model in Fig. 9.23 is not T_c-spaced, e.g., a raised cosine pulse with $\beta = 1$ (or 100% excess bandwidth) results a $T_c/2$-spaced tapped delay line. Moreover, the 1/W-spaced tapped delay line model was derived under the assumption of a strictly band-limited (non-causal) chip shaping pulse $h_c(t)$. Any time-limited (causal) chip shaping pulse leads to a spectrum $\tilde{S}(f)$ that is not band-limited and, therefore, the underlying assumptions in deriving the 1/W-spaced tapped delay line model are violated. Very often, the channel is simply modeled as consisting of uncorrelated T_c-spaced rays i.e.,

$$g(t,\tau) = \sum_{n=-\infty}^{\infty} g_n(t)\delta\left(\tau - nT_c\right) \quad . \tag{9.128}$$

However, in reality the channel rays will not be T_c-spaced. We have seen in Chapter 2.3.4 that the net effect of non-T_c-spaced rays is to introduce correlations into the taps of the T_c-spaced model. Such correlations are complicated to handle in an analytical sense.

5.1   RAKE RECEIVER

A variety of receiver structures can be used to detect DS spread spectrum signals. For DS CDMA where multiple users share the same band, there are two broad types of detectors. The first is a **conventional detector** and is based on the use of correlators or matched filters. With a conventional detector the other user interference, or multiple-access interference is treated as additional unwanted noise. The second is a **multiuser detector,** that uses co-channel demodulation principles to simultaneously detect all the signals that are present. In this section, we concentrate on conventional detectors for DS spread spectrum on multipath fading channels.

A simple type of conventional detector uses the autocorrelation properties of the spreading sequences to reject the multipath interference [133], [132]. Sometimes this is called a **multipath rejection** receiver. Another approach exploits the autocorrelation properties of the spreading sequences to resolve the multipath components and combine them together to obtain a diversity advantage. Since the diversity is obtained from the multipath channel it is sometimes called **multipath diversity.**

To develop the multipath diversity receiver, suppose that one of M possible waveforms having complex envelopes $\tilde{s}_m(t), m = 0, \ldots, M-1$ are transmitted at each baud epoch. With the frequency-selective fading channel shown in

Fig. 9.23, the corresponding received complex envelope is

$$
\begin{aligned}
\tilde{r}(t) &= \sum_{\ell=1}^{L} g_\ell(t)\tilde{s}_m\left(t - \frac{\ell}{W}\right) + \tilde{n}(t) \\
&= \hat{s}_m(t) + \tilde{n}(t)
\end{aligned} \tag{9.129}
$$

where

$$
\hat{s}_m(t) = \sum_{\ell=1}^{L} g_\ell(t)\tilde{s}_m\left(t - \frac{\ell}{W}\right) \ . \tag{9.130}
$$

As discussed in Section 5.2, the maximum likelihood coherent receiver employs a correlator or matched filter to the possible received pulses $\hat{s}_m(t)$ to compute the metrics

$$
\begin{aligned}
\mu(m) &= \text{Re}\left\{\int_0^T \tilde{r}(t)\hat{s}_m^*(t)dt\right\} - E_{\hat{m}} \\
&= \text{Re}\left\{\int_0^T \tilde{r}(t)\sum_{\ell=1}^{L} g_\ell^*(t)\tilde{s}_m^*(t - \ell/W)dt\right\} - E_{\hat{m}} \tag{9.131}
\end{aligned}
$$

where $E_{\hat{m}}$ is energy in the received pulse $\hat{s}_m(t)$. The receiver chooses the index m that maximizes $\mu(m)$.

The receiver described by (9.131) correlates the received complex envelope $\tilde{r}(t)$ with delayed versions of the possible waveforms $\tilde{s}_m(t)$, followed by maximal ratio combining. This leads to the receiver structure shown in Fig. 9.24. By changing the variable of integration in (9.131) an alternate form of the RAKE receiver can be obtained as shown in Fig. 9.25. In this case the waveform $\tilde{s}_m(t)$ is correlated with delayed versions of the received complex envelope $\tilde{r}(t)$. This receivers in Figs. 9.24 and 9.25 were first derived by Price and Green [269], and are commonly called RAKE receivers due to their similarity to the ordinary garden rake.

Error Probability of DS/BPSK with a RAKE Receiver. Consider DS/BPSK signaling with a short PN code $(G = N)$. The two possible DS/BPSK waveforms that are transmitted at each baud epoch have the complex envelopes

$$
\tilde{s}_0(t) = -\tilde{s}_1(t) = Ah(t) \tag{9.132}
$$

where

$$
h(t) = \sum_{k=0}^{N-1} a_k h_c(t - nT_c) \ . \tag{9.133}
$$

With DS/BPSK the received waveforms $\hat{s}_m(t)$ have equal energy so the bias term $E_{\hat{m}}$ in (9.131) is not needed. Assume that $\tilde{s}_0(t)$ is transmitted. Then
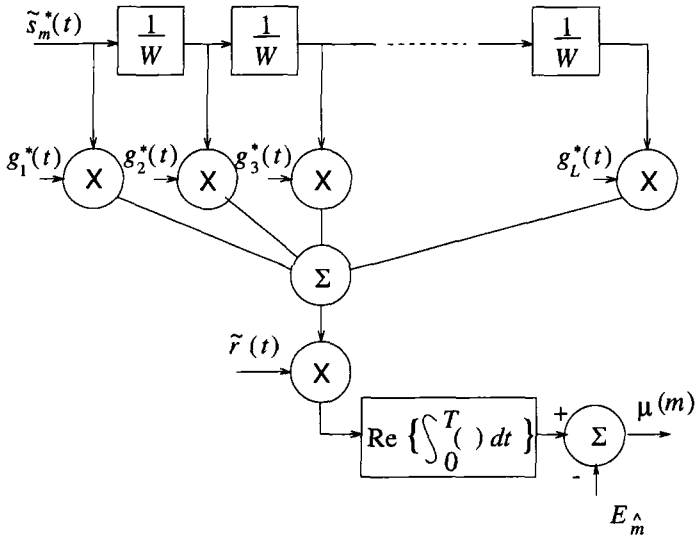
Figure 9.24. RAKE receiver for DS/QPSK signals.
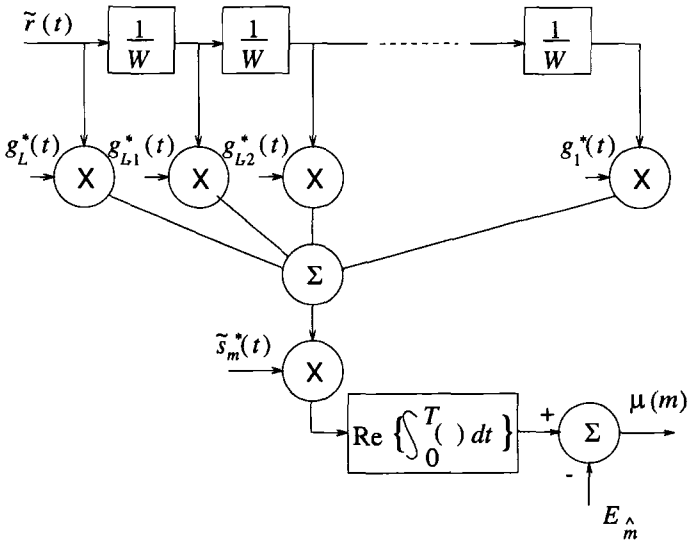
Figure 9.25. Alternate form of RAKE receiver for DS/QPSK signals.

using (9.131)

$$\mu = \sum_{m=1}^{L} \sum_{\ell=1}^{L} \mathrm{Re} \left\{ g_m g_\ell^* \int_0^T \tilde{s}_0(t - m/W)\tilde{s}_0^*(t - \ell/W)dt \right\} + \tilde{n} \quad (9.134)$$

498

where

$$\tilde{n} = \sum_{m=1}^{L} \mathrm{Re}\left\{ g_m^* \int_0^T \tilde{n}(t)\tilde{s}_0^*(t - m/W)dt \right\} \qquad (9.135)$$

and $g_m = \alpha_m e^{j\phi_m}$. The random variable $\tilde{n}$ is Gaussian with zero-mean and variance

$$\sigma_{\tilde{n}}^2 = 2EN_o \sum_{m=1}^{L} \alpha_m^2 . \qquad (9.136)$$

In general, the integral in (9.134) is a complicated function of the spreading sequence and chip amplitude shaping pulse that is used. However, certain cases lead to useful insight. For example, suppose the ideal Nyquist pulse $h_c(t) = S_a(\pi t/T_c)$ with bandwidth $W = 1/T_c$ is used. Strictly speaking this pulse is non-causal so the limits of integration in (9.134) must be from $-\infty$ to ∞. This leads to[4]

$$\begin{aligned}
I &\equiv \int_{-\infty}^{\infty} \tilde{s}_0(t - m/W)\tilde{s}_0^*(t - \ell/W)dt \\
&= A^2 \sum_{k=0}^{N-1} \sum_{j=0}^{N-1} a_k a_j \int_{-\infty}^{\infty} h_c(t - (m+k)T_c)h_c(t - (\ell + j)T_c)dt \\
&= A^2 \sum_{k=0}^{N-1} a_k a_{m+k-\ell} \int_{-\infty}^{\infty} h_c^2(t)dt \\
&= 2E_c N \phi_{aa}(m - \ell) = 2E\phi_{aa}(m - \ell) \qquad (9.137)
\end{aligned}$$

where the second last step follows under the assumption of a short code. Therefore, (9.134) becomes

$$\mu = 2E \sum_{m=1}^{L} \alpha_m^2 + 2E \sum_{m=1}^{L} \sum_{\substack{\ell=1 \\ \ell \neq m}}^{L} \mathrm{Re}\left\{ g_m g_\ell^* \right\} \phi_{aa}(m - \ell) + \tilde{n} . \qquad (9.138)$$

The second term in the above expression is a **self interference** that arises from the non-ideal autocorrelation properties of the spreading sequence.

To demonstrate the effect of the self interference, assume a WSSUS Rayleigh fading channel and consider the random variable

$$\begin{aligned}
Y_{m,\ell} &= \mathrm{Re}\left\{ g_m g_\ell^* \right\} \\
&= (\alpha_m \cos \phi_m)(\alpha_\ell \cos \phi_\ell) + (\alpha_m \sin \phi_m)(\alpha_\ell \sin \phi_\ell) .(9.139)
\end{aligned}$$

---

[4]Since DS/BPSK signaling is used the spreading sequence **a** is real with autocorrelation function $\phi_{aa}(n) = E[a_i a_{i+n}]$.

Define the new random variables

$$X_{I,k} = \alpha_k \cos \phi_k \qquad X_{Q,k} = \alpha_k \sin \phi_k \ . \qquad (9.140)$$

Then

$$\alpha_k = \sqrt{X_{I,k}^2 + X_{Q,k}^2} \qquad \phi_k = \mathrm{Tan}^{-1} \frac{X_{I,k}}{X_{Q,k}} \ . \qquad (9.141)$$

Therefore,

$$Y_{m,\ell} = X_{I,m} X_{I,\ell} + X_{I,m} X_{I,\ell} \qquad (9.142)$$

Since the $X_{I,k}$ and $X_{I,k}$ are independent zero-mean Gaussian random variables with variance σ_k^2, $Y_{m,\ell}$ has the Laplacian density

$$p_{Y_{m,\ell}}(y) = \frac{1}{2\sigma_m \sigma_\ell} \exp\left\{ -\frac{|y|}{\sigma_m \sigma_\ell} \right\} \ . \qquad (9.143)$$

Making the substitution for $Y_{m,\ell}$ and rearranging the sum in the second term in (9.138) gives

$$\mu = 2E \sum_{m=1}^{L} \alpha_m^2 + 4E \sum_{k=1}^{L-1} \sum_{i=1}^{L-1-k} Y_{i,i+k} \phi_{aa}(k) + \tilde{n} \ . \qquad (9.144)$$

It is difficult to evaluate the effect of the self-interference exactly, because the $Y_{m,\ell}$ are non-Gaussian and correlated. However, the self interference due to multipath can be minimized by using spreading codes that have small auto-correlation sidelobes in the time intervals during which delayed signals with significant power are expected. For large delays, the stringent requirements on the autocorrelation function can be relaxed. For asynchronous CDMA appli-cations, the spreading codes still must have small cross-correlation sidelobes over all delays. It is easy to find reasonably large sets of sequences that satisfy these properties. For example, a set of $2^m + 1$ Gold sequences can be generated of length $2^m - 1$. Of these $2^m + 1$ sequences, $2^{m-n+1} + 1$ will have their first autocorrelation off-peak ($t_m - 2$ or t_m) at least n chip durations from the main autocorrelation peak. Consequently, these $2^{m-n+1} + 1$ sequences will introduce negligible self interference if they are used on a channel having an impulse response whos length does not exceed nT_c seconds.

If the spreading sequences have an ideal autocorrelation function, i.e., $\phi_{aa}(n - m) = \delta_{nm}$, then there is no self interference and (9.144) becomes

$$\mu = 2E \sum_{m=1}^{L} \alpha_m^2 + \tilde{n} \ . \qquad (9.145)$$

The probability of bit error is

$$P_b(\gamma_b) = Q\left(\sqrt{2\gamma_b} \right) \qquad (9.146)$$

where γ_b is the received bit energy-to-noise ratio given by

$$\gamma_b = \frac{1}{\sigma_{\tilde{n}}^2}\left(2E\sum_{m=1}^{L}\alpha_m^2\right)^2 = \sum_{m=1}^{L}\gamma_m \qquad (9.147)$$

where

$$\gamma_m = \frac{\alpha_m^2 E}{N_o} \ . \qquad (9.148)$$

With Rayleigh fading, each of the γ_m are exponentially distributed with density function

$$p(\gamma_m) = \frac{1}{\bar{\gamma}_m}\exp\left\{-\frac{\gamma_m}{\bar{\gamma}_m}\right\} \qquad (9.149)$$

where $\bar{\gamma}_m$ is the average received bit energy-to-noise ratio for the kth channel tap. To compute the density of γ_b, first note that the characteristic function of γ_m is

$$\psi_{\gamma_m}(jv) = \frac{1}{1 - jv\bar{\gamma}_m} \qquad (9.150)$$

so that the characteristic function of γ_b is

$$\psi_{\gamma_b}(jv) = \prod_{m=1}^{L}\frac{1}{1 - jv\bar{\gamma}_m} \qquad (9.151)$$

By using a partial fraction expansion and taking the inverse characteristic function, the density of γ_b is

$$p_{\gamma_b}(x) = \sum_{m=1}^{L}\frac{A_m}{\bar{\gamma}_m}\exp\left\{-\frac{x}{\bar{\gamma}_k}\right\} \qquad (9.152)$$

where

$$A_m = \prod_{\substack{i=1 \\ i\neq m}}^{L}\frac{\bar{\gamma}_m}{\bar{\gamma}_m - \bar{\gamma}_i} \ . \qquad (9.153)$$

Therefore, with Rayleigh fading the average probability of bit error is

$$\begin{aligned}
P_b &= \int_0^{\infty}Q\left(\sqrt{2x}\right)p_{\gamma_b}(x)dx \\
&= \frac{1}{2}\sum_{m=1}^{L}A_m\left[1 - \sqrt{\frac{\bar{\gamma}_m}{1 + \bar{\gamma}_m}}\right] \qquad (9.154)
\end{aligned}$$

In order to proceed further the $\bar{\gamma}_m$ must be specified. One plausible model assumes an exponentially decaying power delay profile, e.g.,

$$\bar{\gamma}_m = Ce^{-k/\varepsilon} \qquad (9.155)$$
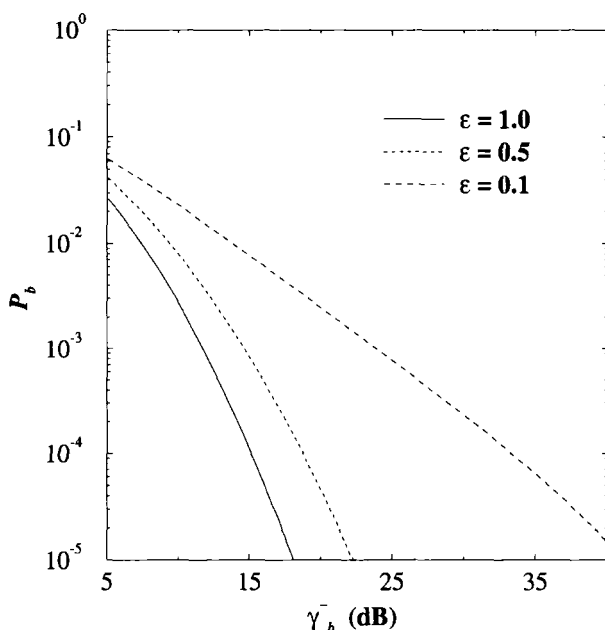
Figure 9.26. Bit error probability with a RAKE receiver for DS/BPSK signaling on a multipath fading channel. The channel has $L = 4$ taps and a 4-tap RAKE receiver is used.

where ε controls the delay spread and C is chosen to satisfy the constraint $\sum_{m=1}^{L} \bar{\gamma}_m = \bar{\gamma}_b$. Solving for C yields

$$\bar{\gamma}_m = \frac{(1 - e^{-1/\varepsilon})e^{-k/\varepsilon}}{e^{-1/\varepsilon} - e^{-(L+1)/\varepsilon}} \bar{\gamma}_b \qquad (9.156)$$

The probability of bit error is plotted in Fig. 9.26 for $L = 4$ and various values of ε. For small ε, the channel is not dispersive and very little multipath diversity is obtained. However, as ε becomes large the channel becomes more dispersive and a greater diversity gain is achieved.

Finally, we note that the number of taps actually used in the RAKE receiver can be less than the channel length L. However, such a RAKE receiver will not capture all the received signal energy and suffer from some loss in performance.

6.   ERROR PROBABILITY FOR DS CDMA ON AWGN CHANNELS

DS CDMA systems achieve multiple-access capability by assigning each user a unique PN spreading sequence. In general, however, the transmissions from the different users are not synchronized and arrive at the intended receiver with different amplitudes, delays and phases. The exact error probability with

a conventional correlation detector will depend on the particular spreading sequences that are employed and will also be a function of the random amplitudes, delays, and phases of the signals that arrive at the intended receiver. Unfortunately, the exact error probability is difficult to derive and evaluate and, therefore, a variety of upper and lower bounds, and Gaussian approximations to the probability of error have been suggested in the literature.

Suppose that K users simultaneously access the channel using DS/BPSK signaling with a short spreading code of length N. The transmitted complex envelope for the ith user is

$$\tilde{s}^{(i)}(t) = A \sum_n x_n^{(i)} h^{(i)}(t - nT) \tag{9.157}$$

where

$$h^{(i)}(t) = \sum_{k=0}^{N-1} a_k^{(i)} h_c(t - kT_c) \tag{9.158}$$

and $\mathbf{a}^{(i)} = \{a_k^{(i)}\}$ and $\mathbf{x}^{(i)} = \{x_n^{(i)}\}$ are the ith user's spreading and data sequences, respectively. The data symbols $x_n^{(i)}$ are independent, random variables chosen from the set $\{-1, +1\}$ with equal probability. In practice, the spreading sequences $\mathbf{a}^{(i)}$ are carefully chosen to have good correlation properties, e.g., Gold sequences or Kasami sequences. However, many of the error probability approximations in the literature assume random spreading sequences, where the error probability is obtained averaging over the ensemble of all possible spreading sequences, including those where multiple users are assigned the same spreading sequence.

In general the signals from the various transmitters will arrive at the intended receiver with different power levels. However, DS CDMA systems must be power controlled such that all signals arrive at the intended receiver with the same power level. Power control is needed to combat the near-far effect, where strong signals will capture the receiver and mask out the weaker signals. Under the assumption of perfect power control and a frequency non-selective channel, the received complex envelope is[5]

$$\tilde{r}(t) = \sum_{i=1}^{K} e^{j\phi_i} \tilde{s}^{(i)}(t - \tau_i) + \tilde{n}(t) \tag{9.159}$$

where the $\{\tau_i\}$ and $\{\phi_i\}$ are the random delays and carrier phases of the received signals. This leads to the model shown in Fig. 9.27.

In this section, we consider the performance with an ideal correlation receiver, where the composite received signal is multiplied by a synchronized

---

[5]Here we assume the normalization $\alpha = 1$.
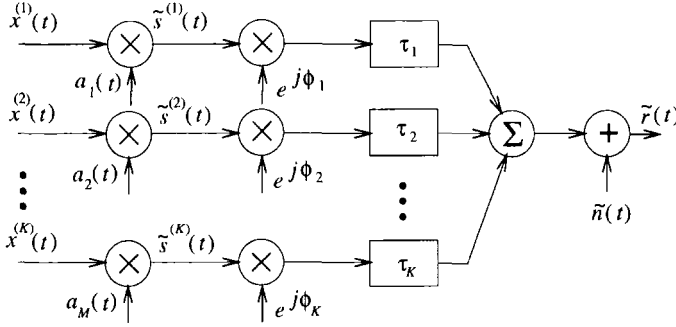
Figure 9.27. DS CDMA signaling on an AWGN channel.

replica of the spreading sequence of the intended transmission; since the two sequences cancel, the desired data sequence can be obtained at the output of the correlator. Because of symmetry, we only need to consider the receiver that is matched to the first transmitter. Furthermore, since only the relative delays and phases are important, we can set $\tau_1 = 0$ and $\phi_1 = 0$, and assume that the remaining ϕ_i are uniformly distributed on $[0, 2\pi)$ and τ_i are uniformly distributed on $[0,T)$.

The decision variable at the output of the correlator in Fig. 9.1 has been derived by Lehnert and Pursley [195] and is given by

$$\mu_n = \sqrt{2E} \left[x_n^{(1)} + \frac{1}{N} \sum_{k=2}^{K} B_{k,1}(x_n^{(k)}, x_{n-1}^{(k)}, \tau_k) \cos \phi_k \right] + \tilde{n}_I \qquad (9.160)$$

where $\tilde{n}_I$ is a zero-mean Gaussian random variable with variance N_o

$$B_{k,1}(x_n^{(k)}, x_{n-1}^{(k)}, \tau_k) = x_{n-1}^{(k)} R_{k,1}^p(\tau_k) + x_n^{(k)} \hat{R}_{k,1}^p(\tau_k) \qquad (9.161)$$

and $R_{k,m}^p(\tau)$ and $\hat{R}_{k,m}^p(\tau)$ are the continuous-time partial cross-correlation functions of $a^{(k)}(t)$ and $a^{(m)}(t)$, defined by

$$R_{k,m}^p(\tau) = \frac{1}{T} \int_0^\tau a^{(k)}(t - \tau) a^{(m)*}(t) dt \qquad (9.162)$$

$$\hat{R}_{k,m}^p(\tau) = \frac{1}{T} \int_\tau^T a^{(k)}(t - \tau) a^{(m)*}(t) dt \ . \qquad (9.163)$$

The functions $R_{k,m}^p(\tau)$ and $\hat{R}_{k,m}^p(\tau)$ can be expressed in terms of the discrete aperiodic cross-correlation function $\phi_{k,m}^a(n)$ and the continuous-time partial chip autocorrelation functions $R_h(\delta)$ and $\hat{R}_h(\delta)$ in (9.33) as

$$R_{k,m}^p(\tau) = \phi_{k,m}^a(\ell - N) \hat{R}_h(\delta) + \phi_{k,m}^a(\ell + 1 - N) R_h(\delta) \quad (9.164)$$

$$\hat{R}_{k,m}^p(\tau) = \phi_{k,m}^a(\ell) \hat{R}_h(\delta) + \phi_{k,m}^a(\ell + 1) R_h(\delta) \qquad (9.165)$$

where $\ell = \lfloor \tau/T_c \rfloor$ and $\delta = \tau - \ell T_c$. Note that δ is uniform on $[0, T_c)$ and ℓ is uniform on the set $\{0, 1, \ldots, N-1\}$. Combining (9.161), (9.164), and (9.165) gives

$$
\begin{aligned}
B_{k,1}(x_n^{(k)}, x_{n-1}^{(k)}, \tau_k) &= \left[x_{n-1}^{(k)} \phi_{k,1}^a(\ell_k - N) + x_n^{(k)} \phi_{k,1}^a(\ell_k) \right] \hat{R}_h(\delta_k) \\
&+ \left[x_{n-1}^{(k)} \phi_{k,1}^a(\ell_k + 1 - N) + x_n^{(k)} \phi_{k,1}^a(\ell_k + 1) \right] R_h(\delta_k)
\end{aligned}
\tag{9.166}
$$

To proceed any further requires information about the aperiodic cross-correlation functions of the spreading sequences being used, as well as the chip amplitude shaping function. For the special case of random spreading sequences and a rectangular chip shaping function $h_c(t) = u_{T_c}(t)$, Morrow and Lehnert [233] have shown that

$$
B_{k,1}(x_n^{(k)}, x_{n-1}^{(k)}, \tau_k) = P_k \delta_k + Q_k(1 - \delta_k) + X_k + Y_k(1 - 2\zeta_k) .
\tag{9.167}
$$

where δ_k is uniform on $[0,1)$, P_k and Q_k are uniform on $\{-1, +1\}$, and X_k and Y_k are discrete random variables having the probability distribution functions

$$
p_{X_k}(i) = \frac{1}{2^{-A}} \binom{A}{\frac{i+A}{2}}, \quad i \in \{-A, -A+2, \ldots, A-2, A\}
\tag{9.168}
$$

$$
p_{Y_k}(i) = \frac{1}{2^{-B}} \binom{B}{\frac{i+B}{2}}, \quad i \in \{-B, -B+2, \ldots, B-2, B\}
\tag{9.169}
$$

The quantities A and B are related to

$$
C \triangleq N\phi_{1,1}^a(1) = \sum_{j=0}^{N-2} a_j^{(1)} a_{j+1}^{(1)}
\tag{9.170}
$$

by

$$
A = \frac{N - 1 + C}{2}
\tag{9.171}
$$

$$
B = \frac{N - 1 - C}{2}
\tag{9.172}
$$

where $\phi_{1,1}^a(1)$ is the aperiodic cross-correlation of the spreading sequence of the first user as defined in (9.21). The parameter B is the number of chip boundaries in one period of the sequence $\mathbf{a}^{(1)}$ at which a transition to a different value occurs. For random spreading sequences, C has the probability distribution

$$
p_C(i) = \frac{1}{2^{N-1}} \binom{N-1}{\frac{i+N-1}{2}}, \quad i \in \{-N+1, -N+3, \ldots, N-3, N-1\} .
\tag{9.173}
$$

6.1   STANDARD GAUSSIAN APPROXIMATION

The **standard Gaussian approximation** assumes that the multiple access interference

$$I = \frac{1}{N} \sum_{k=2}^{K} W_k \cos \phi_k \qquad (9.174)$$

with

$$W_k = B_{k,1}(x_n^{(k)}, x_{n-1}^{(k)}, \tau_k) \qquad (9.175)$$

can be modeled as a Gaussian random variable with a distribution that is completely specified by its mean and variance. The approximation is obtained by conditioning the multiple access interference on the random set of parameters $\{\zeta_k, \phi_k, B\}$ followed by ensemble averaging. It is not difficult to show that

$$E[Z_k|C] = 0 \qquad (9.176)$$

where $Z_k \in \{P_k, Q_k, X_k, Y_k\}$ since the conditional density functions for P_k, Q_k, X_k, and Y_k are symmetrical about zero. Hence $E[W_k] = 0$ and finally $E[I] = 0$.

To compute the variance of the multiple access interference, it is convenient to define the vectors $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_K)$ and $\boldsymbol{\phi} = (\phi_1, \phi_2, \ldots, \phi_K)$. The variance of the multiple access interference is

$$
\begin{aligned}
\sigma_I^2 &= \mathrm{E}\left[I^2 | \boldsymbol{\delta}, \boldsymbol{\phi}, B \right] \\
&= \frac{1}{N^2} \sum_{k=2}^{K} \mathrm{E}\left[W_k^2 | \delta_k, B \right] \mathrm{E}\left[\cos^2 \phi_k | \phi_k \right] \\
&= \frac{1}{2N^2} \sum_{k=2}^{K} (1 + \cos(2\phi_k)) \mathrm{E}\left[W_k^2 | \delta_k, B \right] \qquad (9.177)
\end{aligned}
$$

Since all the Z_k are independent it follows that

$$\mathrm{E}\left[W_k^2 | \delta_k, B \right] = 2(2B + 1)(\delta_k^2 - \delta_k) + N \qquad (9.178)$$

so that

$$\sigma_I^2 = \frac{1}{2N^2} \sum_{k=2}^{K} (1 + \cos(2\phi_k)) 2(2B + 1)(\delta_k^2 - \delta_k) + \frac{1}{N} . \qquad (9.179)$$

If the intended sequence is known, then B is known. For random sequences $E[B] = (N - 1)/2$ giving

$$\sigma_I^2 = \frac{1}{N} \sum_{k=2}^{K} (1 + \cos(2\phi_k))(\delta_k^2 - \delta_k + 1/2) . \qquad (9.180)$$

Several possibilities can be examined from here, including the following two important cases.

Chip and Phase Asynchronous Signals:. The interfering signals are characterized by δ uniform on $[0,1)$ and ϕ uniform on $[0, 2\pi)$ so that $E[\delta_k^2 - \delta_k] = -1/6$ and $E[\cos(2\phi)] = 0$. In this case, $\sigma_I^2 = (K-1)/3N$. Hence, the decision variable in (9.160) can be interpreted as Gaussian random variable with mean $x_n^{(1)} \in \{-1, +1\}$ and variance $(K-1)/3N$ leading to the probability of bit error

$$P_b = Q\left(\sqrt{\frac{3N}{K-1}}\right) \quad . \tag{9.181}$$

The carrier to interference ratio *C/I* can be defined as the carrier power divided by the total noise power

$$\frac{C}{I} = \frac{1}{K-1} \quad . \tag{9.182}$$

By comparing (9.181) with the probability of bit error for binary signaling on an AWGN channel, i.e., $P_b = Q(\sqrt{2\gamma_b})$ we see that the C/I and the effective bit energy-to-noise ratio, γ_b, are related by

$$\gamma_b = \frac{2N}{3}\frac{C}{I} \quad . \tag{9.183}$$

Chip and Phase Synchronous Signals:. The interfering signals have $\delta_k = 0$ and $\phi_k = 0$ so that $\sigma_I^2 = (K-1)/N$ and

$$P_b = Q\left(\sqrt{\frac{N}{K-1}}\right) \quad . \tag{9.184}$$

For chip and phase synchronous signals C/I and the effective γ_b are related by

$$\gamma_b = 2N\frac{C}{I} \quad . \tag{9.185}$$

Coherent addition of interfering signals yields worst case interference with random spreading sequences. The orthogonal Walsh-Hadamard sequences are less random (secure) but yield zero correlation (better performance) under this condition.

The standard Gaussian approximation can be quite inaccurate when the number of simultaneous users *K* is small or the processing gain *N* is large. To circumvent this deficiency a number of improved approximations have been developed.

6.2   IMPROVED GAUSSIAN APPROXIMATION

An improved Gaussian approximation can be obtained by averaging the conditional probability of error over the variance of the multiple access inter-

ference. That is

$$P_b = \int_0^\infty Q\left(\frac{1}{\sqrt{\sigma_I^2}}\right) p_{\sigma_I^2}(\psi)d\psi \tag{9.186}$$

where $p_{\sigma_I^2}(\psi)$ is the pdf of the multiple-access variance σ_I^2. From (9.179)

$$\sigma_I^2 = \frac{1}{N^2}\sum_{k=2}^K L_k \tag{9.187}$$

where $L_k = U_k V_k$ and

$$U_k = (1 + \cos(2\phi_k)) \tag{9.188}$$
$$V_k = (2B+1)(\delta_k^2 - \delta_k) + N/2 \ . \tag{9.189}$$

Note that the $\{U_k\}$ are independent and the $\{V_k\}$ are conditionally independent given B. By using the results in [186] and [117, pp. 79-82], [186, pg. 123,244] the conditional pdf of L_k is

$$p_{L|B}(\ell) = \frac{1}{2\pi\sqrt{\tilde{B}z}}\log\left|\frac{\sqrt{N-\ell}+\sqrt{\tilde{B}}}{\sqrt{N-\ell}-\sqrt{\tilde{B}}}\right| \tag{9.190}$$

where $\tilde{B} = B + 1/2$. Since the L_k are independent and identically distributed, the density of σ_I^2 is obtained by taking the $(K-2)$-fold convolution and removing the condition on B, i.e.,

$$p_{\sigma_I^2}(\psi) = \frac{1}{N^2}\mathrm{E}_B\left[p_{L|B}(\ell) * \cdots * p_{L|B}(\ell)\right] \ . \tag{9.191}$$

This improved Gaussian approximation has been shown to be much more accurate that the standard Gaussian approximation [233]. However, the $(K-2)$-fold convolution in (9.191) must be obtained numerically followed by an additional numerical integration for computing the probability of error. Hence, the utility of this improved Gaussian approximation is limited.

6.3   SIMPLIFIED GAUSSIAN APPROXIMATION

Another simpler but still accurate Gaussian approximation has been derived by Holtzman [168]. To describe this method, let $P(\theta)$ be any function of a random variable θ having mean μ and variance σ^2. The using a Taylor series expansion about the mean μ gives

$$P(\theta) = P(\mu) + (\theta - \mu)P'(\mu) + \frac{1}{2}(\theta - \mu)^2 P''(\mu) + \cdots \tag{9.192}$$

Taking expectations

$$\mathrm{E}[P(\theta)] \approx P(\mu) + \frac{1}{2}P''(\mu)\sigma^2 \ . \tag{9.193}$$

Instead of using the Taylor series expansion, we can start with differences (Stirling's formula) and write

$$P(\theta) = P(\mu) + (\theta - \mu)\frac{P(\mu + h) - P(\mu - h)}{2h}$$
$$+ \frac{1}{2}(\theta - \mu)^2 \frac{P(\mu + h) - 2P(\mu) + P(\mu - h)}{h^2} + \cdots (9.194)$$

Taking expectations

$$E[P(\theta)] \approx P(\mu) + \frac{1}{2}\frac{P(\mu + h) - 2P(\mu) + P(\mu - h)}{h^2}\sigma^2 . \qquad (9.195)$$

Holtzman [168] has shown that $h = \sqrt{3}\sigma$ yields good results so that

$$E[P(\theta)] \approx \frac{2}{3}P(\mu) + \frac{1}{6}P(\mu + \sqrt{3}\sigma) + \frac{1}{6}P(\mu - \sqrt{3}\sigma) . \qquad (9.196)$$

To apply the above result, we let μ and σ^2 be the mean and variance of σ_I^2 in (9.187). Then

$$\mu = \frac{K - 1}{N^2}E[L_k]$$
$$= \frac{K - 1}{N^2}\left(\frac{N}{2} - \frac{E[B]}{3} - \frac{1}{6}\right)$$
$$= (K - 1)/3N \qquad (9.197)$$

where the last line assumes ensemble averaging with random spreading sequences. The variance is

$$\sigma^2 = \frac{K - 1}{N^4}\left(E[L_k^2] - E^2[L_k] + (K - 2)\text{cov}(L_j, L_k)\right) \quad \text{for any } j \neq k.$$
$$= \frac{K - 1}{N^4}\left(\frac{23}{360}N^2 + \left(\frac{1}{20} + \frac{K - 2}{36}\right)N - \frac{1}{20} - \frac{K - 2}{36}\right) . \qquad (9.198)$$

This yields

$$P_b \approx \frac{2}{3}Q\left(\sqrt{\frac{3N}{K - 1}}\right) + \frac{1}{6}Q\left(\frac{1}{\sqrt{(K - 1)/3N + \sqrt{3}\sigma}}\right)$$
$$+ \frac{1}{6}Q\left(\frac{1}{\sqrt{(K - 1)/3N - \sqrt{3}\sigma}}\right) . \qquad (9.199)$$

The above calculations are very simple and lead to quite accurate results for all values of K and N. Fig. 9.28 compares the standard and simplified improved
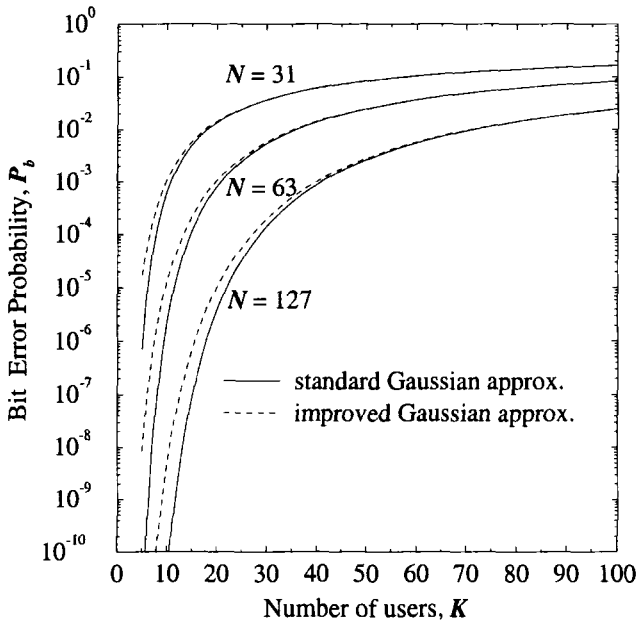
Figure 9.28. Bit error probability against the number of users and various processing gains. The standard Gaussian approximation is shown to under estimate the error probability for small numbers of users.

Gaussian approximations for various processing gains and number of simultaneous users. Note that the standard Gaussian approximation under estimates the error probability for small numbers of users. In this case, the improved Gaussian approximation should be used. However, the accuracy of the standard Gaussian approximation improves when the number of simultaneous users increases.

Note that the above approximations assume an AWGN channel. For frequency selective fading channels, the approximations must be modified to account for the effects of self interference, multipath interference, and envelope fading. In this case the complex low-pass received signal is

$$\tilde{r}(t) = \sum_{i=1}^{K} \sum_{k=l}^{L_i} g_{i,k} \tilde{s}^{(i)}(t - \tau_i - k/W) + \tilde{n}(t) \qquad (9.200)$$

where $g_{i,k}$ is the complex gain associated with the ith user and the kth channel path. A variety of conventional RAKE demodulator structures can be employed to gain a diversity advantage, including the maximal ratio scheme discussed in Section 5.1.

Problems

9.1. Suppose that a DS/BPSK spread spectrum signal is corrupted by a single, phase-asynchronous, interfering tone at the carrier frequency. The received low-pass waveform is

$$\tilde{r}(t) = \tilde{s}(t) + \tilde{i}(t)$$

where $\tilde{s}(t)$ is defined in (9.4) and

$$\tilde{i}(t) = A_i e^{j\phi}$$

where ϕ is an arbitrary phase offset. Assume the use of a short Gold code (of arbitrary length). Compute the probability of bit error with a simple correlation detector.

9.2. The generator polynomials for constructing "Gold-like" code sequences of length $N = 7$ are

$$\begin{aligned} p_1(x) &= 1 + p + p^3 \\ p_2(x) &= 1 + p^2 + p^3 \ . \end{aligned}$$

The sequences are "Gold-like" because $p_1(x)$ and $p_2(x)$ are not a preferred pair and, therefore, will not actually generate a set of Gold sequences. However, the procedure used to construct the set sequences is similar to that used to construct Gold sequences. Generate all the "Gold-like" codes of length 7 and determine the cross-correlation functions of one sequence with each of the others.

9.3. (computer exercise) Write a computer program to generate a set of Gold sequences of length 127.

  a) Plot the mean and variance of the partial period autocorrelation as a function of the processing gain $10 \le G \le 20$ for this set of Gold codes.

  b) Repeat part a) for the partial period cross-correlation.

9.4. (computer exercise) Consider a DS/BPSK CDMA system that uses length-31 Gold codes. The *kth* user spreads their binary data by using the Gold code

$$\mathbf{a}^{(k)} = (a_1^{(k)}, a_2^{(k)}, \dots, a_{31}^{(k)})$$

The complex envelope of the modulated waveform for the *k*th user is

$$\tilde{s}^k(t) = A \sum_n x_n^{(k)} h^{(k)}(t - nT)$$

where

$$h^{(k)}(t) = \sum_{j=1}^{31} a_j^{(k)} u_{T_c}(t - jT_c)$$

$\{x^{(k)}\}$ is uncorrelated zero-mean binary data sequence, T_c is the chip duration and $T = 31\,T_c$ is the data bit duration.

a) Pick one of the length-31 Gold codes at random for the kth user and write down the sequence. Plot the aperiodic autocorrelation function of sequence.

b) Plot the power spectrum of the kth user $S_{\tilde{s}\tilde{s}}^{(k)}(f)$ against the normalized frequency fT_b.

c) Repeat parts a) and b) for a randomly chosen length-63 Gold code for the kth user.

9.5. Plot the continuous-time partial autocorrelation functions of the chip waveform, $R_h(\delta)$ and $\hat{R}_h(\delta)$ in (9.32) and (9.33), respectively, as a function of the fractional chip delay δ for the following chip shaping pulses:

$$h_a(t) = \begin{cases} u_{T_c}(t) & \text{non-return-to-zero} \\ \sin(\pi t/T_c)u_{T_c}(t) & \text{half-sinusoid} \\ 1 - 2|t - T_c/2|/T_c u_{T_c}(t) & \text{triangular} \end{cases}$$

9.6. Consider the set of Walsh-Hadamard sequences of length 16. Determine full period autocorrelation $\phi_{k,k}(n)$ for this set of sequences. Tabulate your results in the $k \times n$ matrix

$$\rho = [\rho_{k,n}]_{k \times n}$$

where $\rho_{k,n} = \phi_{k,k}(n)$.

9.7. A wireless LAN system uses biorthogonal modulation based on the use of length-8 Walsh-Hadamard codewords. Following the discussion leading to (4.81), a set of 16 biorthogonal signals is constructed according to

$$\begin{aligned} \tilde{s}_m(t) &= A \sum_{k=1}^{8} h_{m_k} h_c(t - kT_c), & m = 1, \dots, 8 \\ &= -\tilde{s}_m(t), & m = 9, \dots, 16 \end{aligned}$$

where T_c is the "chip duration" and $T = 8T_c$ is the symbol duration. Assume an uncorrelated data sequence and assume that all 16 signals are used with equal probability.

a Assuming that $h_c(t) = u_{T_c}(t)$, find the power density spectrum of the complex envelope $S_{\tilde{s}\tilde{s}}(f)$.

b Plot the power spectrum $S_{\tilde{s}\tilde{s}}(f)$ against the normalized frequency fT_b, where T_b is the *bit* duration.

9.8. (computer exercise) Suppose that a DS/BPSK spread spectrum system uses an m-sequence of length 127 as a long spreading code. The generator polynomial for the m-sequence is

$$p(x) = 1 + x^3 + x^7$$

The processing gain G is equal to 10, i.e., there are 10 chips per modulated symbol.

a) Assuming that $h_c(t) = u_{T_c}(t)$, find the power density spectrum of the complex envelope $S_{\tilde{s}\tilde{s}}(f)$. Note that you must average over all possible subsequences of length 10.

b) Plot the power spectrum $S_{\tilde{s}\tilde{s}}(f)$ against the normalized frequency fT_b, where T_b is the *bit* duration. Comment on the effect of using a long code on the power spectrum.

9.9. A spread spectrum system transmits a binary data sequence $\mathbf{x} = \{x_k\}$, $x_i \in \{-1, +1\}$, by using the following length-3 short code spreading sequence

$$\mathbf{a} = (-1, +1, +1)$$

a) Compute and plot the full period autocorrelation of the sequence **a**.

b) Compute and plot the aperiodic autocorrelation of the sequence **a**.

c) Suppose the chip shaping function is

$$h_a(t) = \sin\left(\frac{\pi t}{T_c}\right) u_{T_c}(t)$$

What is the transmitted power density spectrum?

d) At which frequencies is the receiver most sensitive to tone interference?

9.10. Suppose that the multipath intensity profile of a channel is given by

$$\phi_g(\tau) = \frac{P}{\mu_\tau} e^{-\tau/\mu_\tau} \ .$$

a) What is the average delay and delay spread of the channel?

b) Suppose DS/BPSK spread spectrum is used on the channel. The receiver employs a two-tap RAKE receiver (assume ideal Nyquist pulses and maximal ratio combining). The tap spacing of the RAKE tapped delay line is equal to the chip duration T_c. Neglecting self-interference, write down an expression for the probability of bit error in terms of the average delay of the channel and the average received bit energy-to-noise ratio.

c) If the bit error probability for a nondispersive channel $(\mu_\tau = 0)$ is 10^{-3}, what is the value of delay spread μ_τ that will reduce the bit error probability from 10^{-3} to 10^{-4}?

9.11. A multipath fading channel has the multipath intensity profile

$$\phi_g(\tau) = \frac{P}{\mu_\tau} e^{-\tau/\mu_\tau}$$

Suppose that DS/BPSK spread spectrum is used on this channel. The receiver employs a 3-tap, T_c-spaced, RAKE receiver with selective diversity combining. Assume ideal Nyquist pulses and the use of spreading sequences having an ideal autocorrelation function. Find the probability of error in terms of the average received bit energy-to-noise ratio.

9.12. Consider the perfectly power controlled DS/BPSK CDMA system that is analyzed in Section 6.. Determine the probability of bit error with a standard Gaussian approximation for the following cases;

a) Chip synchronous and phase asynchronous.

b) Chip asynchronous and phase synchronous.

This page intentionally left blank

Chapter 10

TDMA CELLULAR ARCHITECTURES

This chapter considers the architectural aspects of TDMA digital cellular systems. Regardless of the chosen access method the ultimate goal is to achieve high capacity while satisfying quality of Service (QoS) expectations. An architecture must also be defined so as to be flexible to accommodate system growth. Some schemes for high are flexible enough to be applied to any air interface. But regardless of what air interface is chosen the objective is to implement a system that permits easy installation and growth.

Microcells are a straight forward solution to achieving high capacity. However, as the microcells are introduced, a mixed cell architectures naturally evolves, consisting of both overlaid macrocells and underlaid microcells. Such an arrangement is called a hierarchical architecture. Hierarchical architectures can be implemented for both TDMA and CDMA systems. When microcells are introduced a key issue is the partitioning of the frequency resources among the hierarchical layers. The most attractive hierarchical systems are those that do not partition the system resources among the hierarchical layers. If the entire spectral allocation is used in each hierarchical layer, then *both* high capacity and high flexibility can be achieved. CDMA systems employ universal frequency reuse, but require sophisticated power control algorithms if the spectrum is to be shared between hierarchical layers. TDMA systems can avoid this through careful frequency planning so that the frequency resuse constraint is not violated when microcells are introduced.

Macrodiversity architectures are another method for achieving high capacity, where the same signal is received by, and perhaps transmitted by, multiple BSs. Macrodiversity is an effective method for combatting shadow and envelope fading. In fact, cellular handoff algorithms implement macrodiversity. The soft handoff techniques used in CDMA sytems are a well known method for realizing macrodiversity. In TDMA systems that use hard handoff algorithms

will not yield as much macrodiversity gain due to latencies in the order of a few seconds in the hard handoff algorithms. The requirement for hard handoff in TDMA systems arises a result of the non-universal frequency reuse. However, if dynamic channel assignment (DCA) techniques are used, then TDMA systems can realize benfits from macrodiversity similar to those obtained in CDMA systems. DCA techniques are considered in detail in Chapter 13.

Section 1.1 discusses the use of directional antennas for controling co-channel interference (CCI).

1. CELL SECTORING

1.1 CELL SECTORING WITH WIDE-BEAM DIRECTIONAL ANTENNAS

One of the simplest methods for controling CCI is to use directional antennas at the BSs. On the forward channel, directional antennas reduce the generation of CCI by transmitting the signals to the MSs with a narrower angle-of-arrival (AoA) spread than omni-directional antennas. On the reverse channel, directional antennas reduce the effect of the CCI because they respond to CCI that is generated with a narrower AoA spread about the MS. Here we consider two types of directional antennas; conventional wide-beam directional antennas, and switched beam antennas.

Consider a uniform deployment of hexagonal cells, where the BSs employ omni-directional antennas. Suppose that we ignore the effects of shadowing and multipath-fading, and assume the simple path loss model in (2.226), such that the received desired signal power at distance d is

$$\mu_{\Omega_p} = \Omega_t G_T G_R \left(\frac{h_b h_m}{d^2}\right)^2 \tag{10.1}$$

where Ω_t is the transmit power G_T and G_R are the transmit and receiver antenna gains, respectively, and h_b and h_m are the heights of the BS and MS antennas, respectively. As illustrated in Fig. 10.1, the worst case forward channel CCI situation occurs when the MS is located at the comer of a cell, furthest from its serving BS . There are six first-tier co-channel BSs, two each at (approximate) distances of $D - R$, D, and $D + R$. If the values of Ω_t, G_T and h_b are assumed the same for all BS antennas, then it follows that the worst case C/I is

$$
\begin{aligned}
\Lambda &= \frac{1}{2} \frac{R^{-\alpha}}{(D-R)^{-\alpha} + D^{-\alpha} + (D+R)^{-\alpha}} \\
&= \frac{1}{2} \frac{1}{\left(\frac{D}{R}-1\right)^{-\alpha} + \left(\frac{D}{R}\right)^{-\alpha} + \left(\frac{D}{R}+1\right)^{-\alpha}} \ .
\end{aligned}
\tag{10.2}
$$
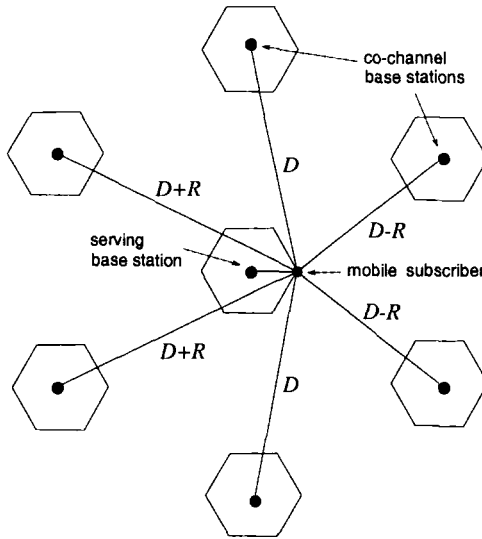
Figure 10.1. Worst case CCI situation on the forward channel.

With a path loss exponent $\beta = 3.5$, the worst case C/I is

$$(\Lambda)_{\text{dB}} = \begin{cases} 14.3 \text{ dB} & \text{for } N = 7 \\ 9.2 \text{ dB} & \text{for } N = 4 \\ 6.3 \text{ dB} & \text{for } N = 3 \end{cases}$$

The minimum allowable cluster size is determined by the minimum C/I requirement of the radio receiver. Unfortunately, the above worst case C/I values may be too small to yield acceptable performance, especially when we account for shadowing and multipath-fading.

Sectoring is a very common method that is employed in cellular systems to improve the C/I performance, whereby the cells are divided into radial sectors with wide-beam directional BS antennas. Cellular systems are quite often deployed with 120°, and sometimes 60°, cell sectors. An *N*-cell reuse cluster with 120° sectors yields an *N/3N* reuse plan (*N* cells and *3N* sectors). As shown in Fig. 10.2, 120° cell sectoring reduces the number of first-tier co-channel interferers from six to two. The two first tier interferers are located at approximate distances of D and $D + 0.7R$. The resulting worst case C/I is

$$\begin{aligned} \Lambda &= \frac{R^{-\alpha}}{D^{-\alpha} + (D + 0.7R)^{-\alpha}} \\ &= \frac{1}{\left(\frac{D}{R}\right)^{-\alpha} + \left(\frac{D}{R} + 0.7\right)^{-\alpha}} \end{aligned} \qquad (10.3)$$
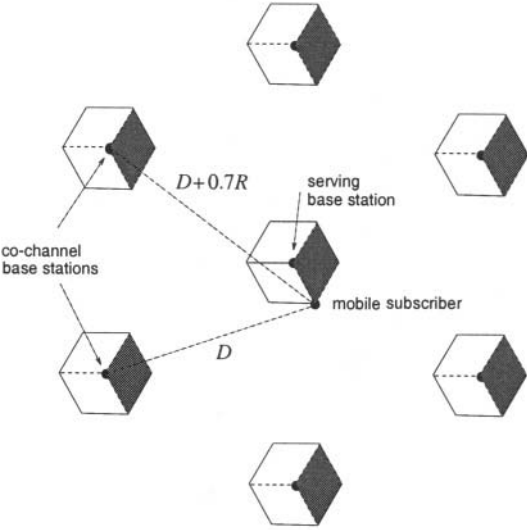
Figure 10.2. Worst case CCI situation on the forward channel with 120° cell sectoring.

Hence

$$(\Lambda)_{dB} = \begin{cases} 21.1 \text{ dB} & \text{for } N = 7 \\ 17.1 \text{ dB} & \text{for } N = 4 \\ 15.0 \text{ dB} & \text{for } N = 3 \end{cases} \tag{10.4}$$

For $N = 7$, 120° sectoring yields a 6.8 dB C/I gain over the case when omnidirectional antennas are used.

To derive a benefit from sectoring, the carriers that are assigned to each cell must be patitioned into disjoint sets, such that each sector uses a disjoint set of carriers. This finer partitioning of the carriers results in a loss in trunking efficiency, which we will quantify in the next section. Hence, cell sectoring improves the C/I performance at the cost of trunking efficiency.

1.2 SECTORING WITH SWITCHED-BEAM ANTENNAS

Switched-beam antennas can be used in place conventional wide-beam directional antennas, to improve both the coverage and system capacity. Switched-beam antennas are a simple type of smart antenna where multiple antenna beams are used within each cell sector, and beam steering is achieved through a simple beam selection mechanism. Switched-beam antennas are more practical than other types of smart antennas such as phased array antennas, because no complicated multi-beam beamforming is needed and no significant changes to existing cellular systems that use conventional wide-beam directional antennas are required.

Switched-beam smart antennas are based on the retro-targeting concept. The selection of the activated receive beam is based on a simple Received Signal Strength Indicator (RSSI) and SAT tone[1]. Forward channel transmissions are over the best received beam, i.e., the same beam is used for both reception and transmission. Beam forming is accomplished by using physically directive antenna elements to create aperture, and thus gain [35]. If the received C/I falls below some preset level, then the BS then switches to the best available beam for both reception and transmission. The directive nature of the narrow-beam ensures that the average level of CCI experienced by any given link is much less than that experienced when conventional wide-beam directional antennas are used [312], thus offering substantial performance advantages [281].

Spatial diversity is typically not used when a smart antenna system is deployed at a BS, simply because the existing physical tower structures prevent it. Angular diversity is a possibility, but it is not effective for macrocellular applications with their characteristically small AOA spreads. Our analysis of switched-beam smart antennas begins with the assumption of stationary or slowly moving MSs, where the radio link quality depends on the instantaneous received envelope due to Rayleigh fading. The pdf of the received signal power *s* due to the combined effect of log-normal shadowing and Rayleigh fading can be modeled by the composite log-normal Gamma distribution in (2.222). We have already seen that the composite log-normal Gamma distribution can be approximated by a log-normal distribution with mean and variance given by (2.223).

Comparisons will be made with an AMPS reference system, where the BSs use 120° sectoring with two-branch spatial diversity and selective combining. The received signal having the best quality (determined by RSSI) is selected for output. Assume that the branches experience independent Rayleigh fading[2]. Furthermore, assume that the shadowing experienced on all antenna branches of the same BS are perfectly correlated. Then the pdf of the conditional received squared envelope with two-branch microscopic selection diversity is

$$p_{\alpha^2|\Omega_p}(x) = \frac{2}{\Omega_p}e^{-x/\Omega_p}\left(1 - e^{-x/\Omega_p}\right) \ . \tag{10.5}$$

By averaging over the log-normal shadowing, the pdf of the composite received squared envelope is

---

[1] The detection of SAT tone can prevent *beam falsing* where the system is spoofed into thinking that the desired MS is located in a different beam when a strong co-channel interferer is present on another beam.
[2] As shown in Chapter 2.1.5.1, a separation of 5 wavelengths will still result in a branch correlation of about 0.7, making the performance of the reference system optimistic [173].

$$p_{\alpha_c^2}(x) = \int_0^\infty \frac{2}{\Omega_p} e^{-x/\Omega_p}(1 - e^{-x/\Omega_p})$$

$$\times \frac{\xi}{\sqrt{2\pi}\sigma_\Omega \Omega_p} \exp\left\{-\frac{(10\log_{10}\Omega_p - \mu_{\Omega_p})^2}{2\sigma_\Omega^2}\right\} d\Omega_p \quad (10.6)$$

where $\xi = 10/\ln 10$. Appendix 10A shows that the pdf $p_{\alpha_c^2}(x)$ can be approximated by a purely log-normal distribution with mean and variance given by

$$\mu_{(\text{dBm})} = \xi^{-1}[\ln 2 - C] + \mu_{\Omega_p\ (\text{dBm})}$$
$$\sigma^2 = \xi^{-2}[\zeta(2,1) - 2(\ln 2)^2] + \sigma_\Omega^2 . \quad (10.7)$$

where $\xi = (\ln 10)/10$, $\psi(\cdot)$ is the Euler psi function and $\zeta(\cdot,\cdot)$ is Riemann's zeta function as defined in Appendix 10A.

The MS antennas are assumed to be omni-directional. With omni-directional BS antennas, there are six first-tier co-channel interferers for both the forward and reverse channels. The number of first-tier interferers is reduced to two with 120° sectoring. With a switched-beam smart antenna, the number of first-tier co-channel interferers on the forward channel is a random variable ranging from 0 to 6, due to the narrow-beam directional antennas and the dependency of the activated beam on the MS location. If there are N_I co-channel interferers each with mean μ_i and variance σ_i^2 (in natural units), then the total interfering power is approximately log-normal. For our purpose, the mean and variance of the approximate log-normal distribution is obtained by using Fenton-Wilkinson method as described in Section 3.1.1. Finally, if the CCI from the antenna sidelobes is ignored, there is at most one interferer on the reverse channel when the smart antenna beamwidth is less than 40°.

1.3 TRUNKPOOL TECHNIQUES

In switched-beam smart antenna systems, the narrow-beam directional antennas are analogous to cell sectoring that can reduce unnecessary spillage of radiation [217] and mitigate the effects of channel time dispersion [227]. Higher antenna gains also can be achieved because of narrow antenna beamwidths. However, switched-beam smart antennas will have more frequent handoffs (due to inter-sector handoffs) that result in reduced trunking efficiency. To overcome the trunking efficiency degradation caused by narrow beam sectoring, **sector-trunkpool** and **omni-trunkpool** load sharing schemes are suggested.

Fig. 10.3 shows a switched-beam smart antenna with 4 azimuthal elements (beams) per 120° degree sector, i.e., 30° beam widths. With a sectored-trunkpool arrangement, all the channels assigned to a 120° sector are shared by all four beams within that sector. Each sector antenna acts as a common aperture for one of four beams. No handoffs are needed unless the MS crosses sector
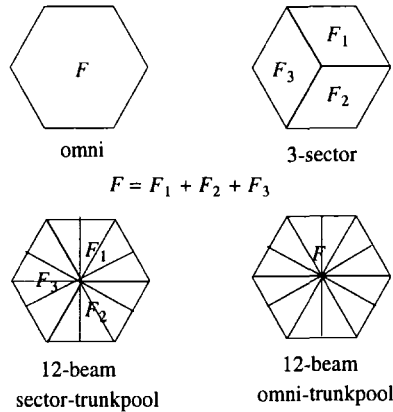
Figure 10.3.   Trunkpool schemes for switched-beam antennas.

or cell boundaries. In this case, the trunking efficiency will remain the same as the reference system, where each wide-beam sector has a unique channel assignment. This concept can be extended to the omni-trunkpool technique, where any of the channels assigned to a cell can be assigned to any one of the activated beams. In this case, no handoffs are needed unless a MS crosses a cell boundary.

Usually, the trunking efficiency is measured by the *channel usage efficiency* (or loading factor) [313]

$$\eta_T = \rho(1 - P_{\mathrm{B}})/m \qquad \text{Erlangs/channel} \qquad (10.8)$$

where ρ is the offered traffic, P_{B} is the blocking probability, and m is the number of channels. From the Erlang-B formula under the blocked-calls-cleared assumption, P_{B} can be shown to be[3]

$$P_{\mathrm{B}} = \frac{\rho^m/m!}{\sum_{k=0}^{m} \rho^k/k!} \quad . \qquad (10.9)$$

With AMPS, each service provider has 416 duplex channels, consisting of 395 traffic channels and 21 control channels. Fig. 10.4 shows the channel usage efficiency for different trunkpool techniques. With the omni-trunkpool technique, the channel usage efficiency is increased 31.2% as compared to a 7-cell reuse reference system when $P_B = 0.01$. In contrast, channel usage efficiency is increased only 17.4% when the frequency reuse cluster size is reduced from 7 to 4 cells. Therefore, the omni-trunkpool technique is helpful for increasing the trunking efficiency when smart antenna systems are employed.

---

[3]The Erlang B formula assumes an infinite subscriber population, and ignores handoff traffic.
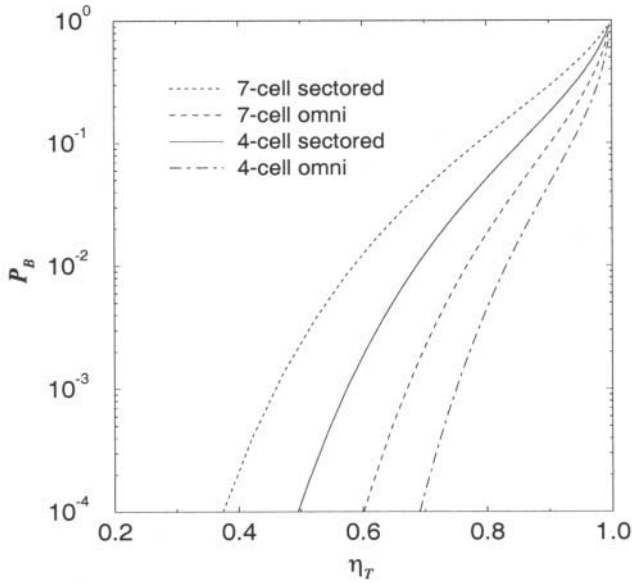
Figure 10.4. Channel usage efficiency with various trunkpool schemes.

1.4 CELLULAR PERFORMANCE WITH SWITCHED-BEAM ANTENNAS

Our performance evaluation begins with the following assumptions:

1. Each cell is circular in shape under a hexagonal cell layout. The MSs are uniformly distributed within a cell.

2. Only the first tier of co-channel interferers is considered.

3. The system utilization is assumed to be 100% (worst case).

4. The 120° directional antennas used in the reference system have perfect directivity, i.e., there are no sidelobes. However, sidelobes adjacent to the main beam are considered for the switched-beam smart antenna system. The smart antenna front-to-back ratio is 30 dB, and the adjacent sidelobe attenuation is 12 dB.

Because power control is employed in the existing AMPS system, two different cases are considered: power control and no power control. Practical power control algorithms usually react to the total received signal strength (C+I). However, for simplicity, the power control algorithm we consider reacts to the desired received signal strength C only[4]. The following notation will

---

[4]Since the required CIR is 17 dB in the AMPS system, the interference power I can be safely neglected.

be used to distinguish between different frequency reuse factors, trunkpool techniques and antenna beam-widths, when switched-beam smart antennas are used:

sma - [*reuse cluster size*] - [*trunkpool type*] - [*antenna beamwidth*]

where

- *reuse cluster size:* 7-cell or 4-cell

- *trunkpool type:* omni or sectored

- *antenna beamwidth:* 30° (12 elements) or 15° (24 elements).

1.4.1 REVERSE CHANNEL

When power control is not employed, all MSs are assumed to transmit with the same power. The CIR at the serving BS is

$$\Lambda = \frac{10^{\frac{\varsigma_d}{10}}}{\sum_i 10^{\frac{\varsigma_i}{10}}} \tag{10.10}$$

where the subscripts d and i index the desired signals and interfering signals, respectively. The random variables ς_d and ς_i are Gaussian distributed, and their means and variances can be derived from (10.7) and (2.223) for the reference system and the smart antenna system, respectively. It is noted that the reference system has two-branch selection diversity in the reverse channel. The number of co-channel interferers depends on the position of the desired and co-channel MSs. The CIR will vary as a function of the activated beams in the co-channel cells. When power control is employed, we assume that the power received at the serving BS from each MS is maintained at a constant level C. The power transmitted by the ith mobile in the jth cell, MS_{ij}, at distance d_{ij} to its serving BS is P_{ij}. The power received at BS_j is

$$C = P_{ij} 10^{\frac{\varsigma_{ij}}{10}} \quad . \tag{10.11}$$

MS_{ij} is also at distance d_{i0} to the reference BS, BS_0, and will generate CCI with a power equal to

$$I_{i0} = P_{ij} 10^{\frac{\varsigma_{i0}}{10}} = C 10^{\frac{\varsigma_{i0} - \varsigma_{ij}}{10}} = C 10^{\frac{\varsigma_i}{10}} \quad . \tag{10.12}$$

Then the C/I at BS_0 is

$$\Lambda = \frac{C}{C \sum_i 10^{\frac{\varsigma_i}{10}}} = \frac{1}{\sum_i 10^{\frac{\varsigma_i}{10}}} \quad . \tag{10.13}$$

Assuming that the path loss follows a fourth law with distance, the mean and variance of ς_i are

$$\mu_i = \mu_{i0} - \mu_{ij} = 10 \log_{10} \left(\frac{d_{i0}}{d_{ij}} \right)^4$$

$$\sigma_i^2 = \sigma_{i0}^2 + \sigma_{ij}^2 . \tag{10.14}$$

1.4.2   FORWARD CHANNEL

The major difference between the forward and reverse channels is the number of co-channel interferers. The calculation of the C/I is similar, but no antenna diversity is included for reference system. The means and variances of the log-normal random variables ς are calculated from (2.223).

1.4.3   PERFORMANCE CRITERIA AND RESULTS

Two criteria will be used to evaluate the performance of the switched-beam smart antenna system:

Criterion 1: The area-averaged probability, P_1, that the received C/I exceeds a target value, Λ_{th}.

Criterion 2: The percentage of the cell area, P_2, where the received C/I exceeds a target value, Λ_{th}, 75% of the time.

Criterion 1 can be treated as a global performance measure. However, bad locations will be masked from the area-averaged performance by good locations. This is particularly true when the performance is non-homogeneous over the cell area. In this case, Criterion 2 is useful. The performance with Criterion 1 is plotted in Figs. 10.5 and 10.6. A significant performance improvement is observed with switched-beam smart antenna systems, especially for the forward channel. For example, with a sma-7-omni-30° system, Criterion 1 yields an improvement of at least 5 dB compared to the reference system. The major factors affecting the area-averaged C/I with switched beam smart antennas are the antenna beamwidth and the frequency re-use factor. The trunkpool techniques do not play an important role.

Figs. 10.7 and 10.8 show the performance with Criterion 2. In this case there is only about 2 dB improvement with respect to the reference system. Trunkpool techniques are shown have significant effect in the performance with Criterion 2. The performance with the sectored trunkpool is better than the omni-trunkpool. The sma-4-sec-30° system is worse than the reference system.

To explain the difference between performance Criterion 1 and 2 more clearly, Figs. 10.9 and 10.10 plot the points during a Monte Carlo simulation where the C/I is less than 17 dB more than 25% time. These points are called
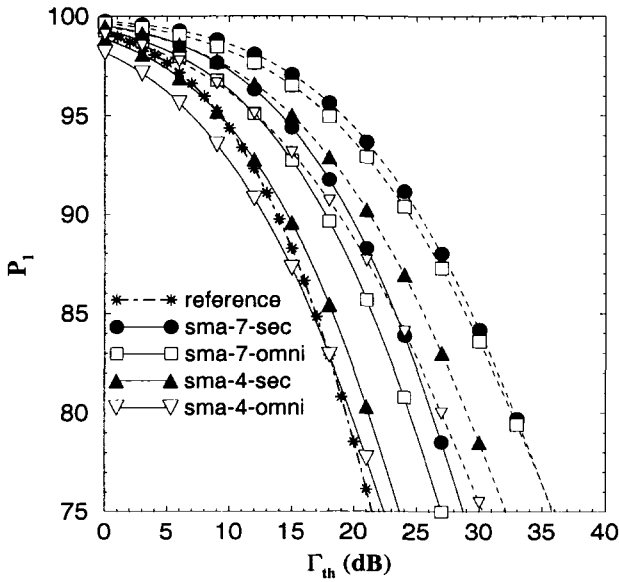
Figure 10.5. Area-averaged probability, P_1, for the reverse channel, with $30°$ antenna beamwidths (solid), and $15°$ antenna beamwidths (dashed); $\sigma = 8$ dB.
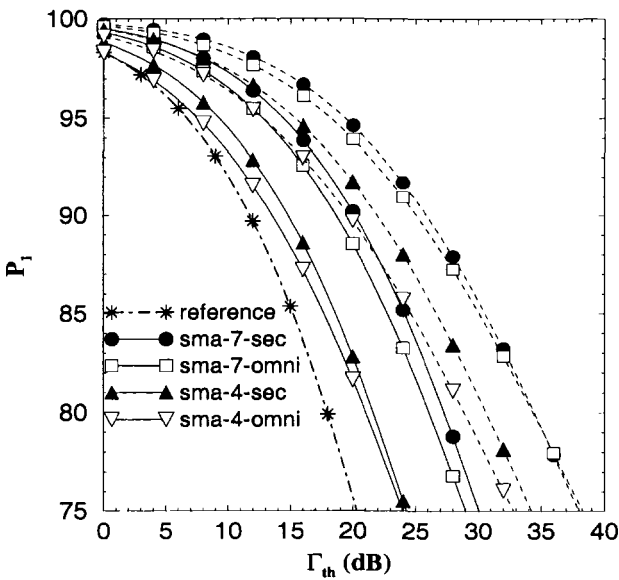
Figure 10.6. Area-averaged probability, P_1, for the forward channel, with $30°$ antenna beamwidths (solid), and $15°$ antenna beamwidths (dashed); $\sigma = 8$ dB.
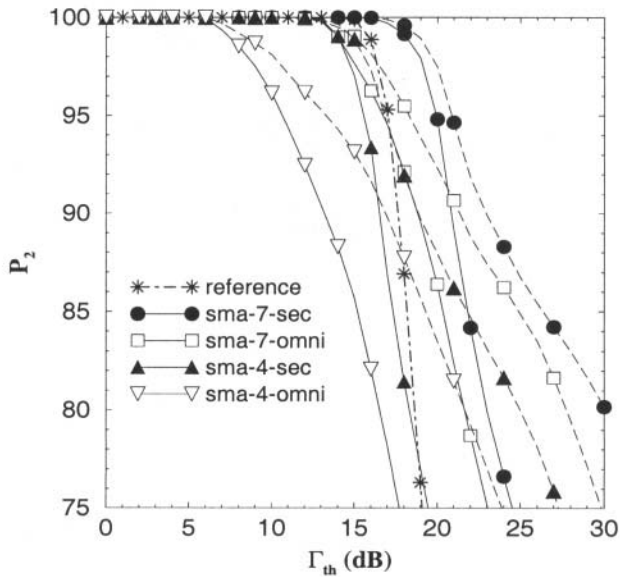
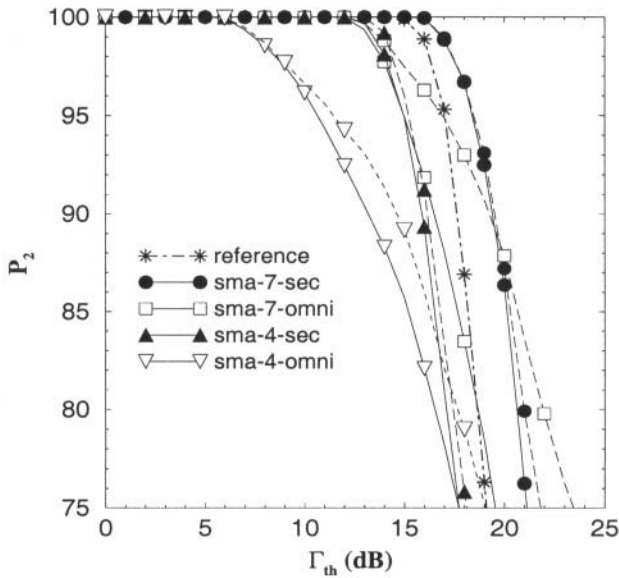*Figure 10.7.* Percentage of the cell area, P_2, for the reverse channel, where the C/I exceeds the target value, Λth, 75% of the time with 30° antenna beamwidths (solid), and 15° antenna beamwidths (dashed); $\sigma = 8$ dB.

**bad points**[5]. With power control, bad points can occur anywhere within the cell and not just on the cell boundary. This is because the power control algorithm only controls the power of the desired signal. When a MS is close to its serving BS the power levels are reduced. Hence, more interference will be experienced on both the forward and reverse channels when the co-channel interferers are close to this BS yet far from their own serving BSs. When switched-beam smart antennas are employed, the performance with Criterion 1 can be improved and yet the performance with Criterion 2 may change little.

Another interesting phenomenon shown in Fig. 10.9 is that the reverse channel bad points with switched-beam smart antennas are concentrated in radial sectors called **bad areas,** i.e., the C/I improvement is not uniform over the entire cell area. We call this is the **cart-wheel** effect. However, it is not present in the forward channel. Fig. 10.11 replots Fig. 10.7 based only on the bad areas. As shown in Fig. 10.11, even the sma-4-sec-15° system is worse than the reference system. Bad areas will always exist in systems using switched-beam smart antennas no matter how narrow the antenna beamwidth is and regardless of whether or not power control is used. Of course, the number of bad points is reduced when the antenna beamwidth is decreased. To mitigate the cart-wheel

---

[5]The location of bad points are not fixed. Their locations vary with the locations of the co-channel interferers.
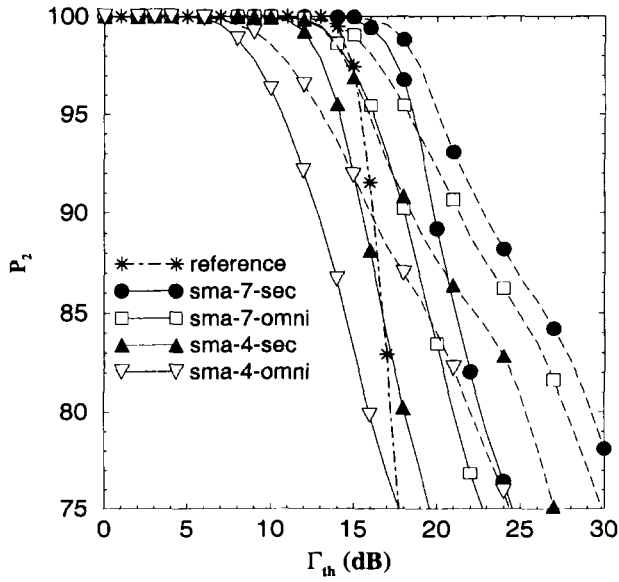
Figure 10.8. Percentage of the cell area, P_2, for the forward channel, where the C/I exceed the target value, Λ_{th}, 75% of the time with 30° antenna beamwidths (solid), and 15° antenna beamwidths (dashed); $\sigma = 8$ dB.
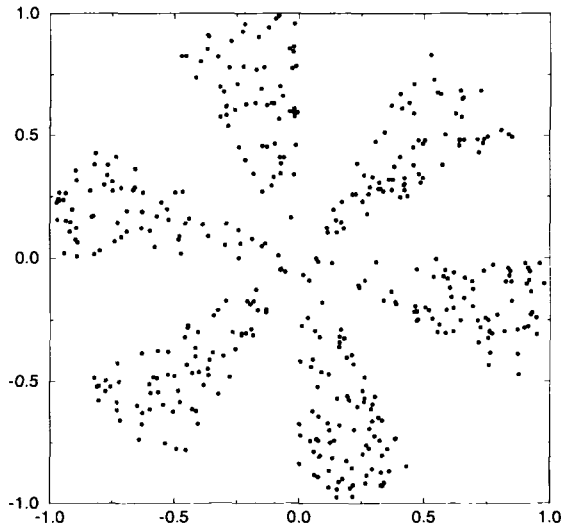
Figure 10.9. Reverse channel bad points for a 'sma-7-omni-30°' system with power control.
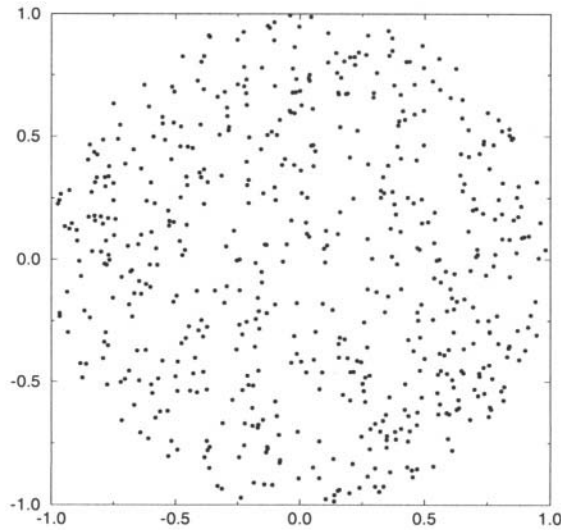
Figure 10.10. Forward channel bad points for a 'sma-7-omni-30°' system with power control.

effect, one possibility is to rotate some switched beam cells with respect to others to distribute the bad points. However, this may be effective only when the rotation degree is larger than the AOA spread of the signals. Another approach is to use dynamic channel assignment to avoid using the same channel when a potential co-channel interferer is nearby.

2.   CONVENTIONAL CELL SPLITTING

Conventional cell splitting is a straight forward process of introducing new, smaller, cells into an existing cellular deployment. By doing so, the cellular system can be tailored to meet traffic growth. To illustrate conventional cell splitting, consider the uniform grid of hexagonal cells shown in Fig. 10.12. If heavy traffic loading is experienced at the midpoint between two the cells labeled **1**, then a split cell labeled **1'** is introduced at that location. The area of the split cell is 1/4 of the area of the parent cells. Additional split cells can be introduced to accommodate traffic loading in other locations throughout the system area. For example, the split cell **2'** can be located at the midpoint between the **2** cells.

Because the split cells are smaller, the transmit power can be reduced. To estimate the transmit power requirements in the split cells, we note that the received power for a MS located at the corner of a parent cell is

$$\Omega(R_o) = A\Omega_o R_o^{-\beta} \qquad (10.15)$$

Figure 10.11. Percentage of the cell area, P_2, for the reverse channel, where the C/I exceeds the target value, Λ_{th}, 75% of the time with 30° antenna beamwidths (solid), and 15° antenna beamwidths (dashed); $\sigma = 8$ dB.
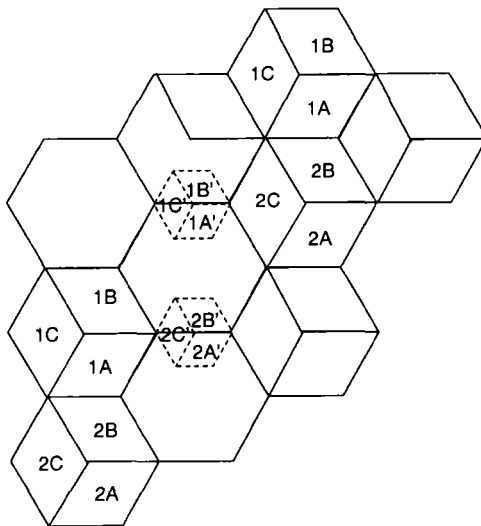
Figure 10.12. Conventional cell splitting is used to accommodate an increased traffic load by introducing smaller cells.

while the received power at the boundary of a split cell is

$$\Omega(R_s) = A\Omega_s R_s^{-\beta} \ . \tag{10.16}$$

where Ω_o and Ω_s, and R_o and R_s, are the transmit power and cell radius associated with the parent cells and split cells, respectively. To keep the received power associated with a MS located on the cell boundary constant, the required transmitter power must be adjusted according to the path loss model. Assuming the model in (1.7), the new transmit power can be obtained as follows:

$$\Omega_s = \Omega_o \left(\frac{R_s}{R_o}\right)^{-\beta} \ . \tag{10.17}$$

If $\beta = 4$, then $\Omega_s = \Omega_o/16$, since $R_s = R_o/2$. Hence, the split cells can reduce their transmit power levels by 12 dB.

After introducing the split cells, changes in the frequency plan are required to avoid violations of the reuse constraint. A very straight forward approach is **channel segmenting**, where the channel sets in the co-channel cells are divided into two groups; the split cells are assigned one group of channels, while the parent co-channel cells are assigned the other group of channels. Unfortunately, this arrangement sacrifices trunking efficiency because the parent cells cannot use the channels assigned to the split cells. Furthermore, if the parent cells are already near capacity, then segmentation of the channels in these cells will require the introduction of more split cells. Hence, a propagation of splitting occurs throughout the system area, requiring the installation of a large number of additional cell sites. Therefore, channel segmenting is not a good option.

Another solution is shown in Fig. 10.13, where **overlaid** inner cells are introduced into the parent cells. Once again, the channels sets are divided into two groups. MSs located within the overlaid inner cells and the split cells use one group of channels, while MSs located within the outer cells use the other group of channels. Whenever a MS moves between the inner and outer areas of a cell a hand-off must be executed, to avoid violations of the co-channel reuse constraint.

2.1   REUSE PARTITIONING

Halpern [157] suggested an overlay/underlay scheme based on the concept of **reuse partitioning,** where multiple co-channel reuse factors are used in the same deployment. An inner cell is created within each of the existing cells as shown in Fig. 10.14. For the example in Fig. 10.14, channels are assigned to the inner and outer cells according to a 3-cell and 7-cell reuse plan, respectively, although other reuse plans could be used. Channels that are assigned to the inner and outer cells can only be used by MSs located within the inner and outer cells, respectively. Handoffs are required when a MS crosses the boundary between
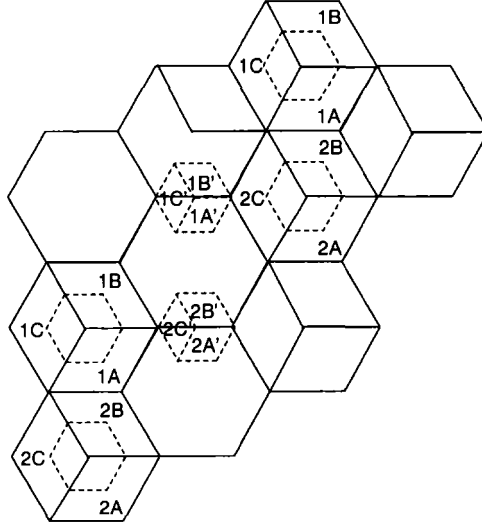
Figure 10.13.   Overlaid inner cells can be used to maintain the frequency reuse constraint when cell splitting is used.

an inner and outer cell. The reduced radii of the inner cells leads to an increase in cell capacity. To quantify this increase let

$$
\begin{aligned}
R_i &= \text{radius of the inner cells.} \\
R_o &= \text{radius of the outer cells.} \\
D_i &= \text{reuse distance for the inner cells.} \\
D_o &- \text{reuse distance for the outer cells.}
\end{aligned}
$$

Suppose that an acceptable link quality requires a co-channel reuse factor $D_i/R_i = D_o/R_o = 4.6$. If a 7-cell and 3-cell reuse cluster is used for the outer and inner cells, respectively, then $D_i/R_o = 3$ and

$$
\frac{D_i/R_i}{D_i/R_o} = \frac{4.6}{3} \ . \tag{10.18}
$$

Hence, the inner and outer cell radii are related by $R_i = 0.65 R_o$ and, therefore, the inner and outer cell areas are related by $A_i = (0.65)^2 A_o = 0.43 A_o$. If a total of N_T channels are available, then $0.43 N_T$ channels should be assigned to the inner cells and $0.57 N_T$ channels assigned to the outer cell area (assuming a homogenous traffic distribution throughout the system area). The resulting cell capacity is

$$
N_\mu = 0.57 N_T/7 + 0.43 N_T/3 = 0.225 N_T \ \text{channels/cell} \ . \tag{10.19}
$$

532
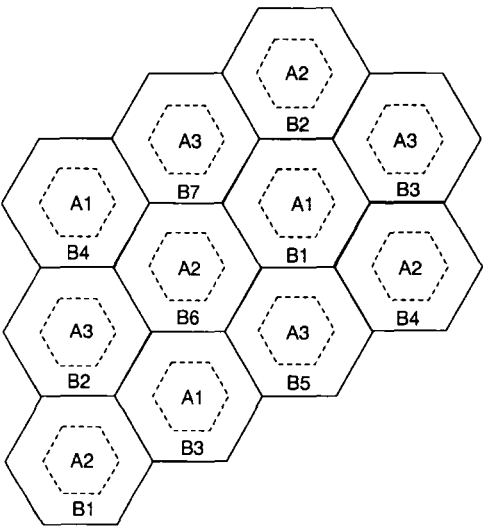
Figure 10.14. Reuse partitioning can be used to increase the channel reuse efficiency, from [157].

On the other hand, with a conventional 7-cell reuse plan

$$N_\mu = N_T/7 \ \text{channels/cell} \ . \tag{10.20}$$

Hence, an improvement of 1.575 in cell capacity is realized.

2.1.1 CELL SPLITTING WITH REUSE PARTITIONING

Cell splitting can also be used with reuse partitioning. An example is shown in Fig. 10.15 where a split cell is added between the parent **B2** cells. The split cell also uses reuse partitioning. To maintain the C/I at an acceptable level, some of the channels in the **B2** cells are moved to the inner cells and are denoted by **B2'**. Furthermore, the closest co-channel inner cells **A1** must have their channels partitioned in a similar fashion. Thus we see a drawback when using cell splitting with the reuse partitioning scheme – the cells must be divided into many concentric rings that use disjoint channel sets, and handoffs must occur when a MS crosses the boundary between two rings.

3. CLUSTER PLANNED HIERARCHICAL ARCHITECTURE

One drawback of conventional cell splitting and reuse partitioning is that the split cells and overlaid cells can only be introduced at specific locations in the cellular deployment. Unfortunately, these locations may not necessary correspond to the hot spot areas that are experiencing the highest traffic growth. We now describe a TDMA hierarchical architecture based on the concept of
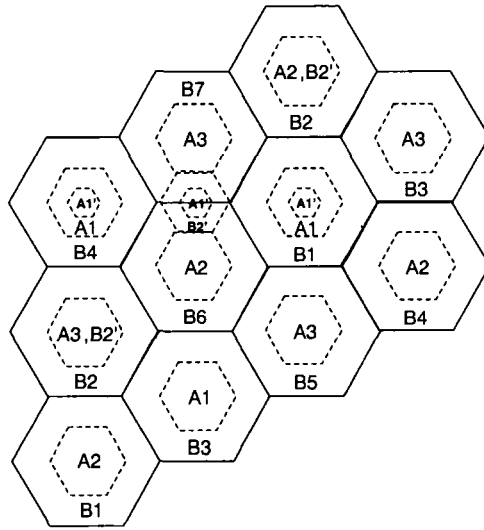
Figure 10.15.   Cell splitting can be used in combination with reuse partitioning.

cluster planning, where macrocells and microcells reuse the same frequencies. Moreover, the microcells can be gradually and extensively deployed at any location to increase the capacity throughout the entire service area. With these flexibilities, the cluster planning approach allows the smooth evolution of existing macrocellular systems into a hierarchical mixed cell architecture.

3.1   SYSTEM ARCHITECTURE

A traditional 7/21 frequency reuse system is shown in Fig. 10.16. The channels are partitioned into 21 sets and each set is reused in a diamond-shaped sector with an adequate distance of separation. Unfortunately, the interfering regions for each channel cover the whole service area. This widely distributed CCI from the macrocells makes it impossible to reuse the same channel frequencies in the microcells.

Cluster planning can be used to change the conventional sectored arrangement into one having some areas of very low interference for a specified set of carriers. The basic cluster planning procedure is as follows:

Cluster planning procedure:

1. Assign the same channels to each cell site as in the traditional 7/21 frequency reuse plan shown in Fig. 10.16.

2. Divide the macrocell reuse clusters into three groups as shown in Fig. 10.17.

3. Let the first group be the reference group.
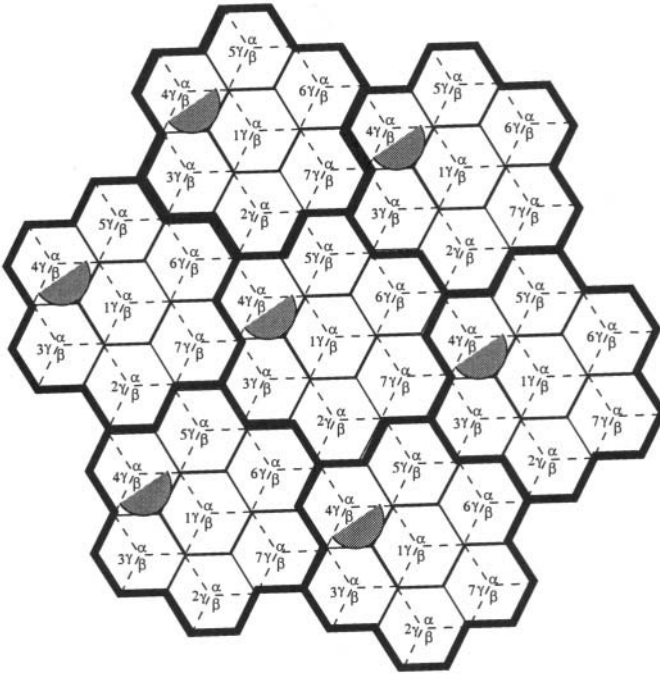
Figure 10.16. Traditional 7/21 frequency reuse plan, from [350].

4. Rotate the channel sets of each cell in the second group 120° *clockwise* with respect to the first group.

5. Rotate the channel sets of each cell in the third group 120° *counter-clockwise* with respect to the first group.

The cluster planning procedure creates low-interference regions outside the areas of the designated macrocell sectors for each channel set. These low-interference regions are called **micro-areas.** Fig. 10.18 shows the result of rotating the sectors. We see that zones $A \sim F$ have a very low interference for channel set 4_β, since they are located in the back-lobe areas of the macrocell sectors using channel set 4_β. Thus microcells can be introduced in these areas by using channel set 4_β.

3.2   UNDERLAID MICROCELL PLANNING ALGORITHM

In the cluster-planned hierarchical architecture, microcells are located in micro-areas where certain macrocell channel sets can be reused. To have the greatest flexibility in selecting the microcell BS locations, it is important to identify all possible micro-areas and the associated channels sets that can be
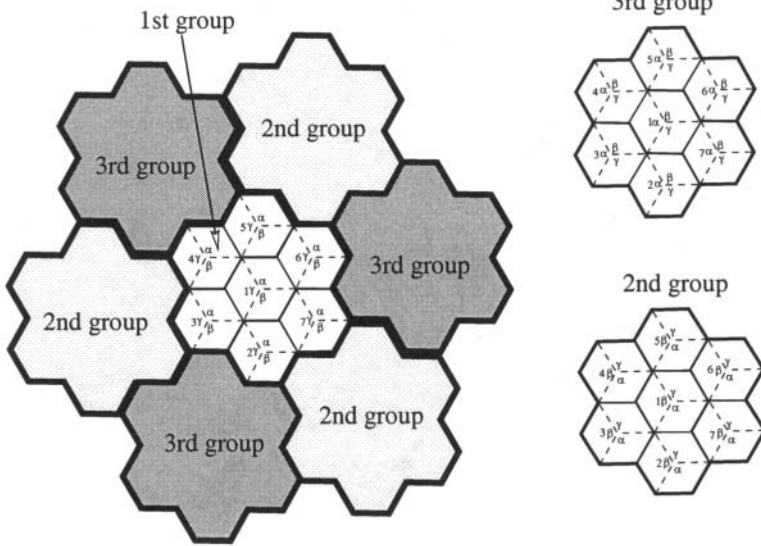
Figure 10.17.   Proposed 7/21 frequency reuse plan with cluster planning, from [350].

used by microcells that are deployed in these areas. In the cluster-planned architecture, the front-lobe areas of the directional antennas are used by the macrocells, while the back-lobe areas of the directional antennas are used by the microcells. In a conventional frequency reuse system (see for example Fig. 10.16, the back-lobe area of each channel set will still encounter some first-ring interferers. To protect the back-lobe areas from the first tier interferers, we rotate the sectors through the cluster planning procedure. Cluster planning creates low-interference micro-areas as shown in Fig. 10.18, that lie in the back-lobe areas of the first-tier interferers. For ease of indexing, a micro-area denotes a region of three adjacent macrocell sectors, each of which belongs to different BS. Fig. 10.19 shows an example of a micro-area. Each micro-area has an **interference neighborhood,** M, defined as the 18 neighboring macrocell sectors that surround the micro-area.

The following algorithm systematically determines the channels that can be used in each micro-area. Let c_i^j represent the channel set in sector $i = \alpha, \beta, \gamma$, of the cell site c, where $c = 1, \ldots, 7$. The superscript $j = 1, 2, 3$ indexes the three groups of rotated clusters.

- Given a desired micro-area and a corresponding interference neighborhood, M, let

$$\Theta = \left\{ c_i^j \in M \right\}$$

denote the union of channel sets c_i^j in the interference neighborhood M.

Micro-area A~F
(using channel set 4 β)
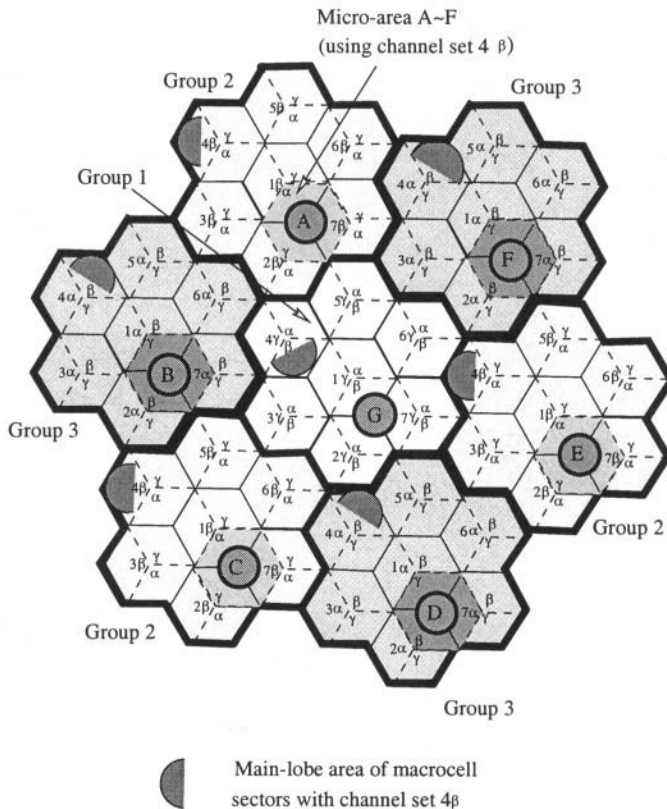
Figure 10.18. Microcells can reuse low-interference macrocell channels in the proposed hierarchical architecture. The macrocell channel set 4_β can be reused in the micro-area A $\sim$ F, from [350].

- From Θ, construct a 3 x 3 *indicator matrix* $\mathbf{B}_c = [b_{ij}]$ for BSs $c = 1,\ldots,7$, where

$$b_{ij} = \begin{cases} 1 & \text{if the channel set } c_i^j \in M; \\ 0 & \text{otherwise} \end{cases}$$

- If the indicator matrix $\mathbf{B}_c$ for some cell site c has a row of ones and two rows of zeroes, then the zero-rows of $\mathbf{B}_c$ indicate the low-interference macrocell channel sets for the micro-area.

Example 9.1 —— According to Fig. 10.19, the interference neighborhood for micro-area A is

$$\Theta = \left\{ 1_\alpha^2, 1_\beta^2, 1_\gamma^2, 2_\alpha^2, 2_\beta^2, 2_\gamma^2, 3_\alpha^2, 3_\alpha^3, 3_\gamma^2, \right.$$
$$\left. 4_\alpha^1, 4_\alpha^2, 4_\alpha^3, 5_\alpha^1, 5_\alpha^2, 5_\gamma^1, 6_\alpha^2, 6_\beta^2, 6_\beta^3, 7_\alpha^2, 7_\beta^2, 7_\gamma^2 \right\}$$
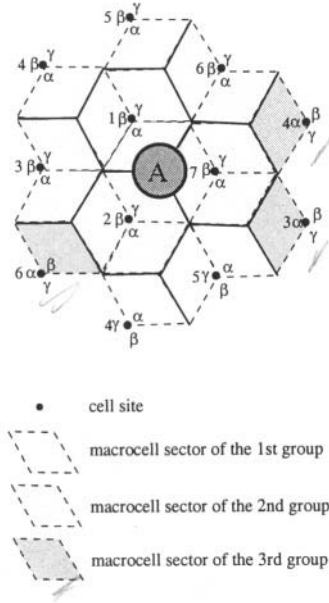
Figure 10.19. Interference neighborhood for micro-area A in Fig. 10.18, from [350].

Then the indicating matrices are

$$\mathbf{B}_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix} ; \mathbf{B}_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix} ; \mathbf{B}_3 = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

$$\mathbf{B}_4 = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} ; \mathbf{B}_5 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} ; \mathbf{B}_6 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\mathbf{B}_7 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix} .$$

Examining the indicating matrix $\mathbf{B}_c, c = 1,\dots,7$, we find that $\mathbf{B}_4$ is the only matrix having a row of ones and two rows of zeroes; the second and the third rows of $\mathbf{B}_4$ are the zero rows. Based on the above algorithm, the low-interference macrocell channel sets for micro-area A are 4_β and 4_γ.

To see if other micro-areas can be defined in the proposed system architecture, consider the system in Fig. 10.20 having 100 micro-areas defined over the service area. By applying the above algorithm, the available macrocell channel sets for each micro-area are listed in Table 10.1. Note that the micro-areas are capable of reusing two macrocell channel sets and microcells can be deployed throughout the whole service area.

538
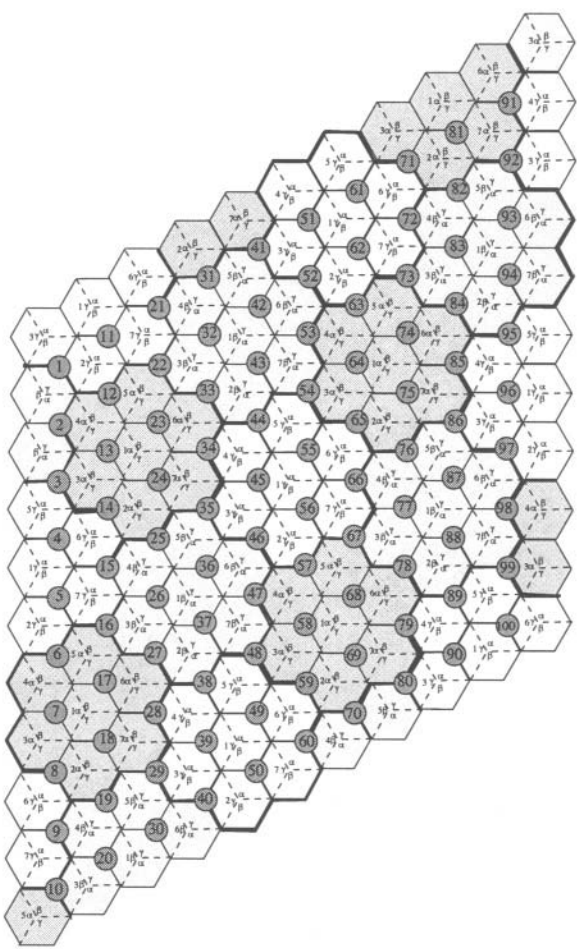
Figure 10.20. Channel reuse in the proposed 3-sector cellular system, where each micro-area consists of three sectors that belong to three different macrocells, from [350].

Since a micro-area consists of 3 macrocell sectors, each macrocell area has 5 available channel sets – 3 assigned to macrocells and 2 assigned to microcells. Within each micro-area the microcells are deployed according to a conventional frequency reuse plan. The microcells could use omnidirectional antennas or sectored antennas. Let C_μ represent the number of the microcell clusters that are deployed in a micro-area. Since each microcell cluster can reuse two sets of low-interference macrocell channels as shown in the above example, the cell capacity can be increased by factor of $1 + 2 \times C_\mu/3$ times. Later we will show that $C_\mu = 6$ is possible and, hence, giving a capacity increase of 7 times.

Zone	Channel Set	Zone	Channel Set	Zone	Channel Set	Zone	Channel Set
1	$7_\alpha, 7_\beta$	26	$6_\alpha, 6_\gamma$	51	$6_\alpha, 6_\beta$	76	$3_\alpha, 3_\beta$
2	$5_\beta, 5_\gamma$	27	$7_\alpha, 7_\gamma$	52	$7_\alpha, 7_\beta$	77	$6_\alpha, 6_\gamma$
3	$1_\beta, 1_\gamma$	28	$5_\alpha, 5_\beta$	53	$5_\beta, 5_\gamma$	78	$7_\alpha, 7_\gamma$
4	$2_\beta, 2_\gamma$	29	$1_\alpha, 1_\beta$	54	$1_\beta, 1_\gamma$	79	$5_\alpha, 5_\beta$
5	$4_\alpha, 4_\gamma$	30	$2_\alpha, 2_\beta$	55	$2_\beta, 2_\gamma$	80	$1_\alpha, 1_\beta$
6	$3_\alpha, 3_\gamma$	31	$3_\alpha, 3_\beta$	56	$4_\alpha, 4_\gamma$	81	$4_\alpha, 4_\beta$
7	$6_\beta, 6_\gamma$	32	$6_\alpha, 6_\gamma$	57	$3_\alpha, 3_\gamma$	82	$3_\alpha, 3_\beta$
8	$7_\beta, 7_\gamma$	33	$7_\alpha, 7_\gamma$	58	$6_\beta, 6_\gamma$	83	$6_\alpha, 6_\gamma$
9	$5_\alpha, 5_\gamma$	34	$5_\alpha, 5_\beta$	59	$7_\beta, 7_\gamma$	84	$7_\alpha, 7_\gamma$
10	$1_\alpha, 1_\gamma$	35	$1_\alpha, 1_\beta$	60	$5_\alpha, 5_\gamma$	85	$5_\alpha, 5_\beta$
11	$4_\alpha, 4_\gamma$	36	$2_\alpha, 2_\beta$	61	$2_\beta, 2_\gamma$	86	$1_\alpha, 1_\beta$
12	$3_\alpha, 3_\gamma$	37	$4_\beta, 4_\gamma$	62	$4_\alpha, 4_\gamma$	87	$2_\alpha, 2_\beta$
13	$6_\beta, 6_\gamma$	38	$3_\beta, 3_\gamma$	63	$3_\alpha, 3_\gamma$	88	$4_\beta, 4_\gamma$
14	$7_\beta, 7_\gamma$	39	$6_\alpha, 6_\beta$	64	$6_\beta, 6_\gamma$	89	$3_\beta, 3_\gamma$
15	$5_\alpha, 5_\gamma$	40	$7_\alpha, 7_\beta$	65	$7_\beta, 7_\gamma$	90	$6_\alpha, 6_\beta$
16	$1_\alpha, 1_\gamma$	41	$1_\alpha, 1_\beta$	66	$5_\alpha, 5_\gamma$	91	$5_\alpha, 5_\beta$
17	$2_\alpha, 2_\gamma$	42	$2_\alpha, 2_\beta$	67	$1_\alpha, 1_\gamma$	92	$1_\alpha, 1_\beta$
18	$4_\alpha, 4_\beta$	43	$4_\beta, 4_\gamma$	68	$2_\alpha, 2_\gamma$	93	$2_\alpha, 2_\beta$
19	$3_\alpha, 3_\beta$	44	$3_\beta, 3_\gamma$	69	$4_\alpha, 4_\beta$	94	$4_\beta, 4_\gamma$
20	$6_\alpha, 6_\gamma$	45	$6_\alpha, 6_\beta$	70	$3_\alpha, 3_\beta$	95	$3_\beta, 3_\gamma$
21	$5_\alpha, 5_\gamma$	46	$7_\alpha, 7_\beta$	71	$7_\beta, 7_\gamma$	96	$6_\alpha, 6_\beta$
22	$1_\alpha, 1_\gamma$	47	$5_\beta, 5_\gamma$	72	$5_\alpha, 5_\beta$	97	$7_\alpha, 7_\beta$
23	$2_\alpha, 2_\gamma$	48	$1_\beta, 1_\gamma$	73	$1_\alpha, 1_\gamma$	98	$5_\beta, 5_\gamma$
24	$4_\alpha, 4_\beta$	49	$2_\beta, 2_\gamma$	74	$2_\alpha, 2_\gamma$	99	$1_\beta, 1_\gamma$
25	$3_\alpha, 3_\beta$	50	$4_\alpha, 4_\gamma$	75	$4_\alpha, 4_\beta$	100	$2_\beta, 2_\gamma$

Table 10.1.   The macrocell channel sets that can be used in the underlaid microcells, from [350].

3.3   PERFORMANCE ANALYSIS OF CLUSTER PLANNED ARCHITECTURE

Propagation Model and System Assumptions:.   Our analysis uses a simple modification of the path loss model in (10.1)[6]

$$\mu_{\Omega_p} = \frac{\Omega_t (h_b h_m)^2}{d^\beta} \quad , \tag{10.21}$$

where μ_{Ω_p} and Ω_t are the received and transmitted powers, h_b and h_m are the BS and MS antenna heights, respectively, d is the radio path length, and β is the path loss exponent. Although (10.21) is more suitable for a macrocell

---

[6]Here we incorporate the antenna gains in the transmitted power.

environment than a microcell environment, it is still characteristic of the path loss experienced by the microcell links at locations that are well outside of the microcells. In other words, the model is applicable when considering the CCI that is generated by distant microcells.

CCI: In our two-tiered hierarchical architecture, four types of CCI must be considered; macrocell-to-macrocell, microcell-to-microcell, macrocell-to-microcell and microcell-to-macrocell CCI. Adjacent channel interference should also be considered.

Antennas: The macrocell BSs are assumed to use 120° wide-beam directional antennas, while microcell BSs use omni-directional antennas. It is possible to improve the C/I performance by sectoring the microcells as well, but we do not consider this here. The MSs use omni-directional antennas.

Uplink power control: We adopt the power control scheme used in IS-54/136 and AMPS [95]. The transmitted power of a Class IV IS-54/136 portable handset is adjusted in six levels from -22 dBW to -2 dBW in steps of 4 dB. Downlink power control is not required in the proposed architecture. Before proceeding, we first clarify our notation. When the subscripts M and μ are used, they refer to macrocells and microcells, respectively; when m and b are used, they denote the MS and BS, respectively; when d and u are used, they indicate the downlink and uplink, respectively.

3.3.1 MACROCELL PERFORMANCE

Section 3.2 showed that the cluster planning technique creates low interference regions, thereby allowing the microcells to reuse macrocell frequencies. However, some macrocells will experience higher interference after rotating the sectors. This is the cost of cluster planning. To evaluate the influence of the sector rotations on the macrocell performance, consider both the conventional macrocellular system in Fig. 10.16 and the proposed hierarchical cellular system in Fig. 10.18 without the underlaid microcells. Fig. 10.21 shows the simulation results of the uplink C/I performance for both systems, assuming that the MSs are uniformly distributed in each sector and they transmit with the maximum power. We consider the uplink case because the downlink performance is usually better than the uplink performance. With respect to a 90% coverage probability, one can observe that the sector rotation technique creates low interference regions at the cost of about 3.1 dB, 3.3 dB, and 3.5 dB of C/I degradation for path loss exponent $\beta = 3.6, 3.8,$ and 4.0, respectively. Even after sector rotations, the macrocells can maintain a C/I greater than 20 dB over 90% of the coverage area. In the following, we will further include the effect of underlaid microcells when analyzing the performance of the proposed hierarchical cellular system. For ease of analysis, we hereafter adopt the worst case scenario where a MS is situated on a cell boundary.
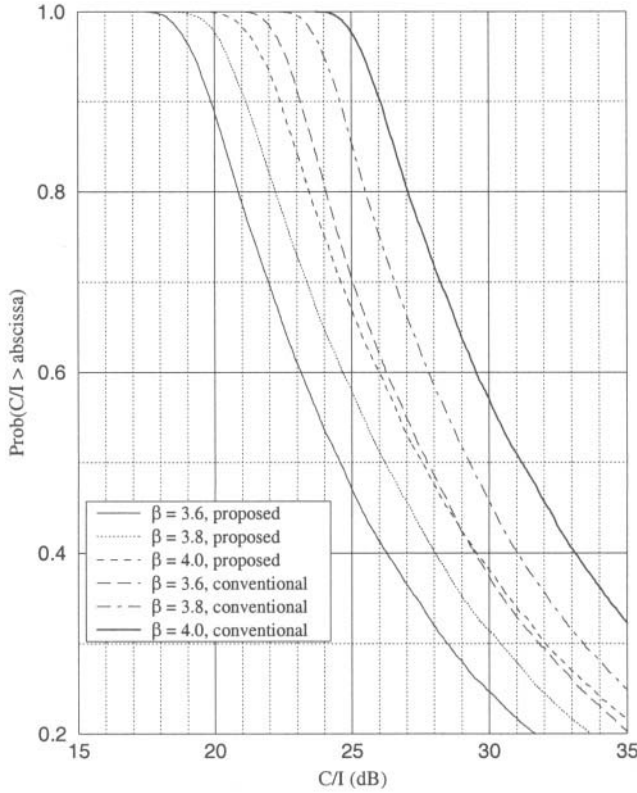
Figure 10.21. Comparison of the uplink C/I performance of conventional macrocells and the proposed hierarchical cellular system without the underlaid microcells for different path loss exponent β, .

Downlink C/I analysis:. By applying (10.21) with $\beta = 4$, we express the C/I received by the MS at the macrocell boundary as

$$\frac{C_M^d}{I_M^d + J_{\mu M}^d} = \frac{\Omega_{t,b}^M \dfrac{(h_b^M h_m)^2}{R_M^4}}{\displaystyle\sum_{i=1}^{N_M} \frac{\Omega_{t,b}^M (h_b^M h_m)^2}{D_i^4} + \sum_{j=1}^{Z_\mu} \sum_{k=1}^{C_\mu} \frac{\Omega_{t,b}^\mu (h_b^\mu h_m)^2}{d_{jk}^4}}$$

(10.22)

where

$$
\begin{aligned}
C_M^d &= \quad \text{MS received power from the desired macrocell BS} \\
I_M^d &= \quad \text{downlink macrocell-to-macrocell CCI} \\
J_{\mu M}^d &= \quad \text{downlink microcell-to-macrocell CCI}
\end{aligned}
$$

$$
\begin{aligned}
\Omega_{t,b}^{M} &= \text{macrocell BS transmitted power} \\
\Omega_{t,b}^{\mu} &= \text{microcell BS transmitted power} \\
N_M &= \text{the number of macrocell co-channel interferers} \\
Z_u &= \text{the number of interfering micro-areas} \\
C_\mu &= \text{the number of microcell clusters in a micro-area} \\
D_i &= \text{MS distance to the } i\text{-th interfering macrocell BS} \\
d_{j,k} &= \text{MS distance to the } k\text{-th interfering} \\
&\qquad \text{microcell BS in the } j\text{-th micro-area} \\
h_b^{M} &= \text{macrocell BS antenna height} \\
h_b^{\mu} &= \text{microcell BS antenna height} \\
h_m &= \text{MS antenna height} \\
R_M &= \text{macrocell radius}
\end{aligned}
$$

Referring to Fig. 10.20 and Table 10.1, we examine the downlink interference when a macrocell MS using channel set 1_β is located at the macrocell boundary near micro-area 56. One can find that the macrocell-to-macrocell downlink interference I_M^d mainly comes from two first-tier macrocell BSs located near micro-areas 77 and 68 with distances $[D_1, D_2] = [4, 3.61]R_M$, respectively. However, because the objective of cluster planning is to carefully manage the C/I, the performance may be sensitive to the C/I. Consequently, we also consider the three second-tier interfering BSs located near micro-areas 11, 17 and 62, located at distances $[D_3, D_4, D_5] = [8.89, 8.89, 8.72]R_M$, respectively. For the micro-cell-to-macrocell downlink interference $J_{\mu M}^d$, one can find six interfering micro-areas 35, 48, 54, 80, 86, and 99 in the first tier with distances of $[\overline{d_1}, \overline{d_2}, \overline{d_3}, \overline{d_4}, \overline{d_5}, \overline{d_6}] = [3, 4.58, 3.46, 6, 5.2, 6.25]R_M$, respectively. The second-tier interfering micro-areas 3, 29, 41, and 92 have distances $[\overline{d_7}, \overline{d_8}, \overline{d_9}, \overline{d_{10}}] = [7.55, 9, 7.94, 12]R_M$, respectively. We assume that each micro-area has C_μ microcell reuse clusters, and each of these clusters has K_μ microcells. Through the channel selection algorithm in Section 3.2, each micro-area is assigned two macrocell channel sets. We further partition these two sets of channels into K_μ groups and then assign each group to the K_μ microcells in each cluster. In this manner, a macrocell channel set is used C_μ times in a micro-area. For ease of analysis, we assume that the distance $\overline{d_j}$ approximates $d_{j,k}$, where $\overline{d_j}$ is the distance from a macrocell MS to the center of the j-th interfering micro-area and $d_{j,k}$ is defined following (10.22). In our example, the microcell BS antenna height is one third of macrocell BS antenna height, i.e., $h_b^{\mu}/h_b^{M} = 1/3$, although this ratio can be easily varied. With the
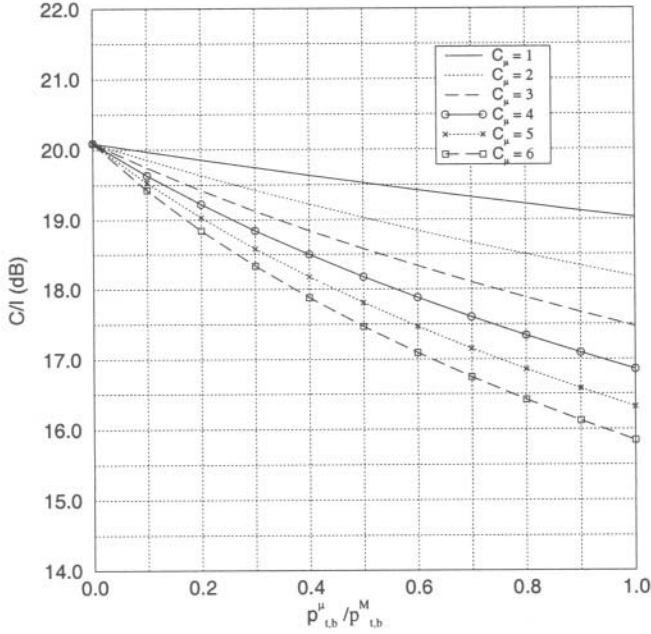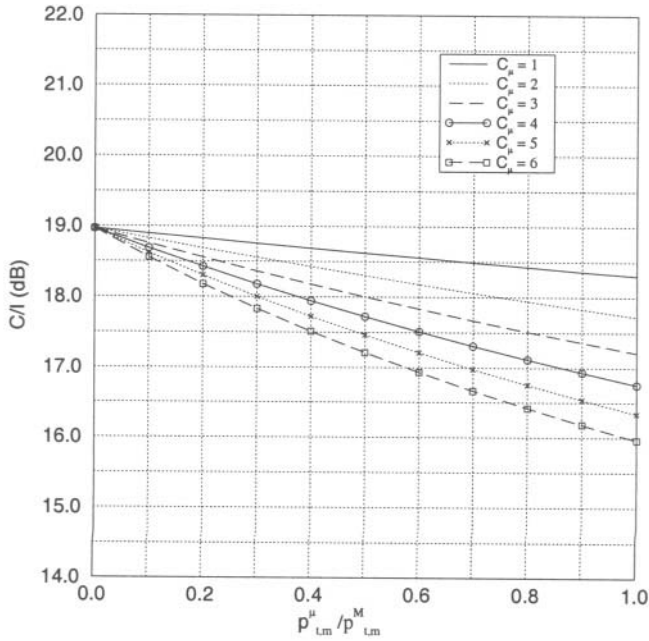
Figure 10.22. Macrocell downlink C/I performance against $\Omega_{t,b}^{\mu}/\Omega_{t,b}^{M}$ for different values of C_{μ}, where $\Omega_{t,b}^{\mu}/\Omega_{t,b}^{M}$ is the microcell to microcell BS transmit power, and C_{μ} is the number of microcell clusters in a micro-area, from [350].

above assumptions in (10.22),

$$\frac{C_M^d}{I_M^d + J_{\mu M}^d} = \frac{1}{1.02875 \times 10^{-2} + C_{\mu}\left(\dfrac{\Omega_{t,b}^{\mu}}{\Omega_{t,b}^{M}}\right) \times 2.79449 \times 10^{-3}} . \quad (10.23)$$

We show the downlink C/I performance in terms of C_{μ} and $\Omega_{t,b}^{\mu}/\Omega_{t,b}^{M}$ in Fig. 10.22 with consideration of only first-tier interfering BSs and in Table 10.2 with both first- and second-tier interfering BSs. Observe that $C/I \geq 18$dB for $C_{\mu} = 6$ and $\Omega_{t,b}^{\mu}/\Omega_{t,b}^{M} \leq 0.3$. In other words, the channel set 4_{β} can be reused six times in the micro-area while still keeping the worst case macrocell downlink C/I greater than 18 dB. The reuse increases even further if the required C/I is smaller than 18 dB. Furthermore, by comparing the results in Table 10.2 with Fig. 10.22, one can find that the second-tier interfering BSs only degrade the C/I by an additional 0.5 dB over the first-tier interfering BSs.

	$C_M^d/(I_M^d + J_{\mu M}^d)$ (dB)			
$\dfrac{\Omega_{t,b}^{\mu}}{\Omega_{t,b}^{M}}$	$C_\mu = 1$	$C_\mu = 2$	$C_\mu = 4$	$C_\mu = 6$
0	19.88	19.88	19.88	19.84
0.1	19.76	19.64	19.42	19.22
0.2	19.65	19.42	19.02	18.65
0.3	19.53	19.22	18.65	18.15
0.4	19.43	19.02	18.31	17.70
0.5	19.32	18.83	17.99	17.29
0.6	19.22	18.65	17.69	16.91
0.7	19.12	18.48	17.42	16.57
0.8	19.02	18.30	17.16	16.25
0.9	18.93	18.14	16.91	15.96
1.0	18.83	17.99	16.68	15.68

Table 10.2. Downlink C/I performance for overlaying macrocells, where $h_b^{\mu}/h_b^{M} = 1/3$.

Uplink CCI analysis:. By modifying (10.22) slightly, we can formulate the uplink C/I as

$$\frac{C_M^u}{I_M^u + J_{\mu M}^u} = \frac{\Omega_{t,m}^M \dfrac{(h_b^M h_m)^2}{R_M^4}}{\displaystyle\sum_{i=1}^{N_M} \frac{\Omega_{t,m}^M (h_b^M h_m)^2}{D_i^4} + \sum_{j=1}^{Z_\mu} \sum_{k=1}^{C_\mu} \frac{\Omega_{t,m}^\mu (h_b^M h_m)^2}{d_{jk}^4}}$$

(10.24)

where

$$
\begin{aligned}
C_M^u &= \quad \text{macrocell BS received power from the desired MS} \\
I_M^u &= \quad \text{uplink macrocell-to-macrocell interference} \\
J_{\mu M}^u &= \quad \text{uplink microcell-to-macrocell interference} \\
\Omega_{t,m}^M &= \quad \text{macrocell MS transmitted power} \\
\Omega_{t,m}^\mu &= \quad \text{microcell MS transmitted power}
\end{aligned}
$$

and where the remaining parameters have already been defined following (10.22). With directional antennas, the macrocell BSs experience fewer interfering micro-areas in the uplink direction as compared with the downlink direction. Consider the macrocell sector that is assigned with channel set 2γ

and located near micro-area 37. This macrocell sector encounters two first-tier and four second-tier macrocell interfering MSs at distances

$$[D_1, D_2, D_3, D_4, D_5, D_6] = [3.61, 3.61, 8.54, 8.19, 8.19, 7.81]R_m \ ,$$

and interfering micro-areas 23, 55, 61, 68, 74, 100, (i.e. $Z_\mu = 6$) at distances

$$[d_1, d_2, d_3, d_4, d_5, d_6] = [7.0, 7.0, 14.7, 5.3, 11.5, 9.53]R_m \ .$$

We can ignore the effect of the three other interfering micro-areas 4, 17, 49 because they are located in the back-lobe area of the sector using channel set 2γ. By substituting the above values into (10.24), the uplink C/I performance for this example becomes

$$\frac{C_M^u}{I_M^u + J_{\mu M}^u} = \frac{1}{1.2677 \times 10^{-2} + C_\mu \left(\dfrac{\Omega_{t,m}^\mu}{\Omega_{t,m}^M}\right) \times 2.11 \times 10^{-3}} \ . \qquad (10.25)$$

Fig. 10.23 shows the results. Suppose that a worst case target C/I of 18 dB is chosen. Then it is observed that the C/I is greater than 18 dB for $C_\mu = 1 \sim 6$ if

$$\frac{\Omega_{t,m}^\mu}{\Omega_{t,m}^M} \le 0.2 \ . \qquad (10.26)$$

Note that we obtained (10.26) under the assumption that the interfering macrocell MSs are on the cell boundary and are transmitting with the maximum power. Thus (10.26) can be used to determine the maximum microcell MS's transmitted power. For example, consider an IS-54/136 Class IV portable handset (that adjusts its transmitted power in six levels from -22 dBW to -2 dBW). Then (10.26) implies that the maximum microcell MS transmitted power is -9 dBW, which is still in the operational range of the Class IV terminal. The implication is that the requirement in (10.26) can be fulfilled by the current uplink power control scheme in the IS-54/136 system without changing the MS transmitted power specification.

3.3.2 MICROCELL PERFORMANCE

We now show how the microcell size should be chosen to achieve the required C/I performance.

Downlink microcell size:. A feasible microcell size should satisfy two conditions: (i) C-criterion: a MS will receive stronger power, C, at the microcell boundary than at the macrocell boundary; (ii) C/I-criterion: the C/I at the microcell boundary equals or exceeds that at the macrocell boundary.

Figure 10.23. Macrocell uplink C/I performance against $\Omega_{t,m}^\mu/\Omega_{t,m}^M$ for different values of C_μ, where $\Omega_{t,m}^\mu/\Omega_{t,m}^M$ is the ratio of the transmitted power of the microcell MS to that of the macrocell MS, and C_μ is the number of microcell clusters in a micro-area, from [350].

C-criterion: From the path loss model in (10.21), the microcell radius R_μ can be calculated as

$$R_\mu \leq \left[\left(\frac{\Omega_{t,b}^\mu}{\Omega_{t,b}^M}\right)\left(\frac{h_b^\mu}{h_b^M}\right)^2\right]^{1/4} R_M \ , \tag{10.27}$$

where R_M, h_b^μ, h_b^M, $\Omega_{t,b}^\mu$, and $\Omega_{t,b}^M$ are defined in (10.22).

S/I-criterion: The S/I received by the MS at the microcell boundary can be written as

$$\frac{C_\mu^d}{I_\mu^d + J_{M\mu}^d} = \frac{\Omega_{t,b}^\mu \frac{(h_{\mu,b}h_m)^2}{R_\mu^4}}{\sum\limits_{i=1}^{C_\mu-1}\frac{\Omega_{t,b}^\mu(h_{\mu,b}h_m)^2}{D_{\mu_i}^4} + \sum\limits_{i=1}^{N_{M_f}}\frac{\Omega_{t,b}^M(h_{M,b}h_m)^2}{D_{M_f,i}^4} + \frac{1}{\eta}\left(\sum\limits_{i=1}^{N_{M_b}}\frac{\Omega_{t,b}^M(h_{M,b}h_m)^2}{D_{M_b,i}^4}\right)}$$

$$\tag{10.28}$$

where the parameters $\Omega_{t,b}^\mu$, $\Omega_{t,b}^M$, h_b^M, h_b^μ, C_μ, and h_m are already defined in (10.22) and

$$
\begin{aligned}
C_\mu^d &= \text{MS received power from its desired microcell BS} \\
I_\mu^d &= \text{downlink microcell-to-microcell interference} \\
J_{M\mu}^d &= \text{downlink macrocell-to-microcell interference} \\
N_{Mf} &= \text{the number of main-lobe macrocell interferers} \\
N_{Mb} &= \text{the number of back-lobe macrocell interferers} \\
D_{Mf,i} &= \text{MS distance to the } i\text{-th main-lobe interfering BS} \\
D_{Mb,j} &= \text{MS distance to the } j\text{-th back-lobe interfering BS} \\
D_{\mu,i} &= \text{MS distance to the } i\text{-th microcell interfering BS} \\
R_M &= \text{macrocell radius} \\
R_\mu &= \text{microcell radius} \\
\eta &= \text{the front-to-back ratio of the directional antenna in macrocells}
\end{aligned}
$$

Let $(C/I)_{\text{req}}$ denote the required C/I. Then (10.28) becomes

$$
\frac{R_\mu}{R_M} \le \left[\frac{(C/I)_{\text{req}}^{-1}}{\left[\displaystyle\sum_{i=1}^{C_\mu-1} \left(\frac{1}{\widehat{D_{\mu i}}}\right)^4 + \left(\displaystyle\sum_{i=1}^{N_{Mf}}\left(\frac{1}{\widehat{D_{Mf,i}}}\right)^4 + \frac{1}{\eta}\displaystyle\sum_{j=1}^{N_{Mb}}\left(\frac{1}{\widehat{D_{Mb,j}}}\right)^4\right)\left(\frac{\Omega_{t,b}^M}{\Omega_{t,b}^\mu}\right)\left(\frac{h_{M,b}}{h_{\mu,b}}\right)^2\right]} \right]^{\frac{1}{4}}
$$

$$(10.29)$$

where $\widehat{D_{Mf,i}} = D_{Mf,i}/R_M$, $\widehat{D_{Mb,j}} = D_{Mb,j}/R_M$, and $\widehat{D_{\mu i}} = D_{\mu i}/R_M$, are the normalized distances of interferers with respect to macrocell radius R_M. Our studies assume that the microcells and macrocells have similar shapes, and that the microcell clusters are adjacent to each other in a given micro-area. Suppose the distances from a microcell MS to its interfering microcell BSs are equal and close to the microcell co-channel reuse distance D_μ (i.e., $D_{\mu,i} = D_\mu$, for $i = 1, \cdots, C_\mu$). Then

$$
\frac{D_\mu}{R_\mu} = \sqrt{3K_\mu} \ , \tag{10.30}
$$

where K_μ denotes the microcell cluster size. With C_μ microcell clusters and K_μ microcells inside each cluster, a micro-area has in total $C_\mu K_\mu$ microcells. Suppose that taken together they are smaller than the area of a macrocell. Then

$$
\frac{R_M}{R_\mu} \ge \sqrt{C_\mu K_\mu} \ . \tag{10.31}
$$

Substituting (10.30) (10.31) into (10.29), we get

$$\frac{R_\mu}{R_M} \le \left[\frac{(C/I)_{\text{req}}^{-1}}{\frac{(C_\mu - 1)C_\mu^2}{9} + \left(\sum_{i=1}^{N_{M_f}} \left(\frac{1}{\widehat{D_{M_{f,i}}}} \right)^4 + \frac{1}{\eta} \sum_{j=1}^{N_{M_b}} \left(\frac{1}{\widehat{D_{M_{b,j}}}} \right)^4 \right) \left(\frac{\Omega_{t,b}^M}{\Omega_{t,b}^\mu} \right) \left(\frac{h_b^M}{h_b^\mu} \right)^2} \right]^{\frac{1}{4}}$$

(10.32)

Notice that we consider N_{M_b} back-lobe macrocell interferers in (10.32). The back-lobe interference from the macrocell BSs can be ignored for the macrocell MS, but for the microcell MS, this kind of interference may be relatively strong compared to the received signal strength from the *low-powered* microcell BS. For the same reason, the macrocell interferers in the second ring are considered here.

Example 9.2———————————————————Referring to Fig. 10.20 and Table 10.1, micro-area 56 can be assigned channel sets $[4_\alpha, 4_\gamma]$. Take channel set 4_γ as an example. Micro-area 56 will experience three first-tier back-lobe interferers ($N_{M_b} = 3$), each of which has the following distance

$$[\widehat{D_{M_{b,1}}}, \widehat{D_{M_{b,2}}}, \widehat{D_{M_{b,3}}}] = [2.65, 2.65, 2.65]$$

(10.33)

to the center of micro-area 56. Three main-lobe interfering macrocells in the second tier are located near micro-areas 25, 79, 64 with the distances of

$$\left[\widehat{D_{M_{f,1}}}, \widehat{D_{M_{f,2}}}, \widehat{D_{M_{f,3}}} \right] = [5.29, 5.29, 5.29] \quad .$$

(10.34)

Furthermore, three main-lobe interfering macrocell BSs in the third tier are located near micro-areas 13, 70, and 85 with distances of

$$\left[\widehat{D_{M_{f,4}}}, \widehat{D_{M_{f,5}}}, \widehat{D_{M_{f,6}}} \right] = [7.0, 7.0, 7.0] \quad .$$

(10.35)

It is also important to determine if there exist interfering microcell BSs from neighboring micro-areas. Fig. 10.20 and Table 10.1 shows one feature of the proposed system architecture; the adjacent micro-areas are assigned different macrocell channel sets. For instance, micro-area 56 in Fig. 10.20 is assigned channel sets $[4_\alpha, 4_\gamma]$, while the neighboring micro-areas 45, 46, 55, 57, 66, and 67 use channel sets $[6_\alpha, 6_\beta]$, $[7_\alpha, 7_\beta]$, $[2_\beta, 2_\gamma]$, $[3_\alpha, 3_\gamma]$, $[5_\alpha, 5_\gamma]$, $[1_\alpha, 1_\gamma]$. Obviously, when considering the interfering microcell BSs, a microcell MS will only be affected by the interfering microcell BSs in the same micro-area. Assume that each micro-area consists of C_μ microcell clusters. Then a MS will experience the interference from the remaining $C_\mu - 1$ microcell BSs, excluding the desired one. Substituting (10.33), (10.34), and (10.35) into

(10.32), one can obtain

$$
\frac{R_\mu}{R_M} \leq \left[\frac{(C/I)_{\text{req}}^{-1}}{\frac{(C_\mu - 1)C_\mu^2}{9} + \left(5.0803 \times 10^{-3} + 0.0608 \times \frac{1}{\eta}\right) \left(\frac{\Omega_{t,b}^M}{\Omega_{t,b}^\mu}\right) \left(\frac{h_b^M}{h_b^\mu}\right)^2} \right]^{\frac{1}{4}}
$$

$$(10.36)$$

(a) $C_\mu = 1$: We first consider the special case where only one microcell is installed in each micro-area. This situation may occur with initial microcell deployment. Fig. 10.24 shows the effect of the front-to-back ratio η on the microcell radius, whereby $(C/I)_{\text{req}} = 18$ dB and $h_b^\mu / h_b^M = 1/3$. If the C/I- and C-criterion result in different microcell radii, then the smallest one must be chosen. From Fig. 10.24, one can observe that if front-to-back ratio $\eta \geq 10$ dB, the microcell radius is determined by the C-criterion, but when $\eta \leq 5$ dB, the C/I-criterion dominates the C-criterion. For instance, in the case of $\eta = 10$ dB and $\Omega_{t,b}^\mu / \Omega_{t,b}^M = 0.4$, one can obtain $R_\mu \leq 0.5 R_M$ by the C/I-criterion and $R_\mu \leq 0.46 R_M$ by the C-criterion, respectively. We must satisfy the more stringent requirement and, therefore, the microcell radius is $0.46 R_M$. In this example, one can see that a larger front-to-back ratio η does not imply a larger microcell size, since the C-criterion, which is independent of η, will dominate the C/I-criterion when η is large.

(b) $C_\mu \geq 2$: Next, we consider the case where many microcells are deployed in each microarea. Fig. 10.25 shows the downlink microcell size against $\Omega_{t,b}^\mu / \Omega_{t,b}^M$ for different values of C_μ, where $\Omega_{t,b}^\mu / \Omega_{t,b}^M$ is the ratio of the transmitted power of the microcell BS to that of the macrocell BS, and C_μ is the number of microcell clusters in a micro-area. It is observed that if $C_\mu \geq 3$, $\Omega_{t,b}^\mu / \Omega_{t,b}^M$ has little effect on the downlink microcell size. This is because the interference from the microcells, I_μ^d, will dominate the macrocell interference, J_μ^d, when the number of co-channel microcells $(C_\mu - 1)$ becomes large in a given micro-area. In other words, if a large number of microcells are installed, the C/I-criterion will become a dominating factor in determining the microcell size. In the case of $C_\mu = 6$, for example, one should follow the C/I-criterion to get $R_\mu \leq 0.165 R_M$ from Fig. 10.25.

Uplink microcell size:. Similar to the previous analysis for the downlink microcell size, the uplink microcell size is derived from the C/I analysis. More
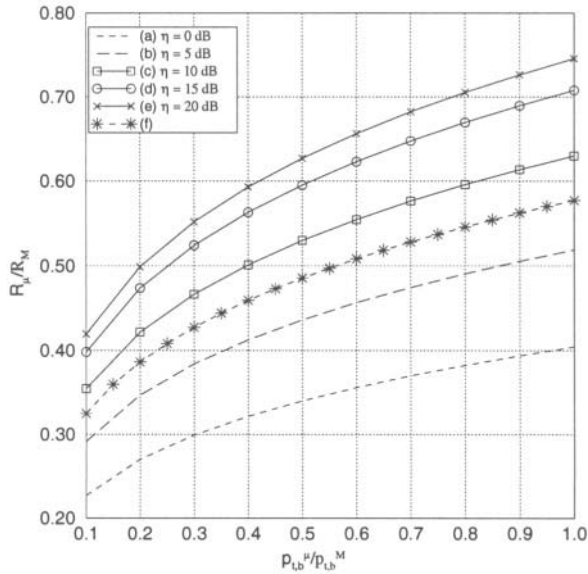
Figure 10.24. Effect of front-to-back ratio η on the microcell radius based on downlink microcell C/I performance analysis, where R_μ/R_M and $\Omega^\mu_{t,b}/\Omega^M_{t,b}$ are the cell radius ratio and transmitted power ratio of microcells over macrocells, respectively. With $(C/I)_{req} = 18$ dB and $h^\mu_b/h^M_b = 1/3$, curves (a) ~ (e) are obtained by C/I-criterion for $\eta = 0, 5, 10, 15, 20$ dB, respectively, while curve (f) is obtained by C-criterion, from [350].

specifically,

$$\frac{C^u_\mu}{I^u_\mu + J^u_{M\mu}} = \frac{\Omega^\mu_{t,m}\dfrac{(h^\mu_b h_m)^2}{R^4_{\mu,up}}}{\displaystyle\sum_{i=1}^{C_\mu-1}\frac{\Omega^\mu_{t,m}(h^\mu_b h_m)^2}{D^4_{\mu i}} + \sum_{i=1}^{N_{M_{f,i}}}\frac{\Omega^M_{t,m}(h^b_M h_m)^2}{D^4_{M i}}}$$

(10.37)

where the parameters $\Omega^\mu_{t,b}$, $\Omega^M_{t,b}$, C_μ h^M_b, $D_{\mu,i}$, R_M, h^μ_b, and h_m have been defined in (10.22) and (10.28) and

$$
\begin{aligned}
C^u_\mu &= \text{microcell BS received power from the desired microcell MS}\\
I^u_{\mu M} &= \text{uplink microcell-to-macrocell interference}\\
J^u_{M\mu} &= \text{uplink macrocell-to-microcell interference}\\
N_{M,f} &= \text{the number of macrocell interfering MSs}\\
D_{M,i} &= \text{BS distance to the } i\text{-th interfering macrocell MS}\\
R_{\mu,up} &= \text{uplink microcell radius}
\end{aligned}
$$
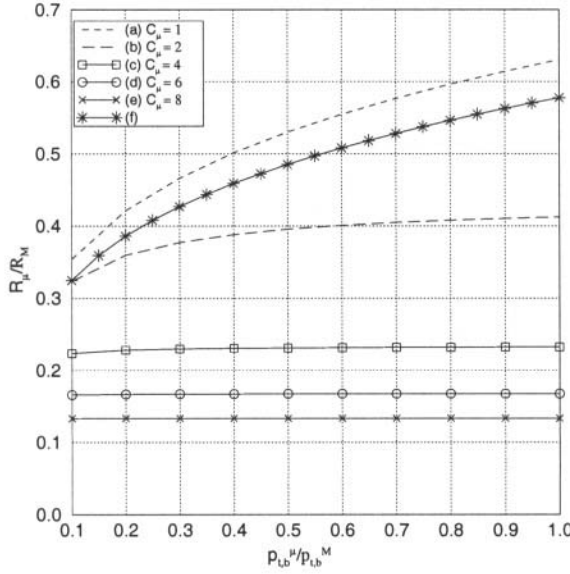
Figure 10.25. Downlink microcell radius R_μ against $\Omega_{t,b}^\mu/p_{t,b}^M$ for different values of C_μ in the case $\eta = 10$ dB, $(C/I)_{\text{req}} = 18$ dB, and $h_b^\mu/h_b^M = 1/3$, whereby the microcell radius is normalized with respect to the macrocell radius R_M; $\Omega_{t,b}^\mu/\Omega_{t,b}^M$ represents the ratio of the transmitted power of microcell BS to that of macrocell BS; and C_μ is the number of clusters in a micro-area; η is the front-to-back ratio of the directional antenna; $h_b^\mu/h_{t,b}^M$ is the ratio of the microcell BS antenna to the macrocell BS antenna. Curves (a) $\sim$ (e) are obtained by C/I-criterion for $C_\mu = 1, 2, 4, 6, 8$, while curve (f) is obtained by C-criterion, from [350].

Let $D_{M,i} = \widehat{D_{M,i}} R_M$ and $(C/I)_{\text{req}}$ denote the required C/I for a microcell BS. Using the same assumptions for getting (10.29), one can simplify (10.37) as

$$
\frac{R_{\mu,up}}{R_M} \leq \left[\frac{(C/I)_{\text{req}}^{-1}}{\dfrac{(C_\mu - 1) C_\mu^2}{9} + \left(\displaystyle\sum_{i=1}^{N_{M_f}} \left(\frac{1}{\widehat{D_{M_f,i}}} \right)^4 \right) \left(\dfrac{\Omega_{t,m}^M}{\Omega_{t,m}^\mu} \right)} \right]^{\frac{1}{4}}
$$

(10.38)

We have shown that when the number of microcell clusters C_μ becomes large, the downlink microcell size is insensitive to the interference from the macrocells. This is also true for determining the uplink microcell size. This will be shown by a later example. When microcell interference dominates the
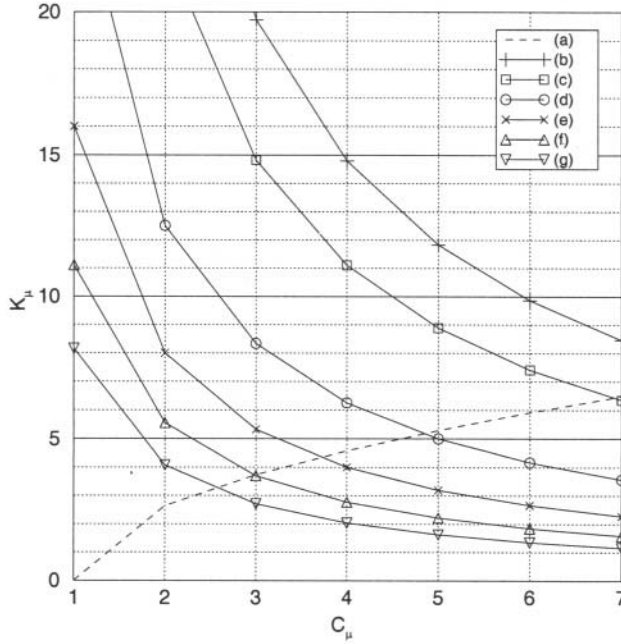
Figure 10.26. K_μ against C_μ with R_μ/R_M as a parameter, whereby K_μ is the microcell cluster size, C_μ is the number of clusters in a micro-area, and R_μ/R_M is the ratio of the microcell radius to the macrocell radius. Curve (a) represents the lower bound of K_μ, while curves (b) $\sim$ (g) represent the upper bound of K_μ for $R_\mu/R_M = 0.13, 0.15, 0.20, 0.25, 0.30, 0.35$, respectively, from [350].

performance, (10.38) can be approximated as

$$\frac{R_{\mu,up}}{R_M} \leq \left[\frac{1}{(C/I)_{\text{req}} \dfrac{(C_\mu - 1)\, C_\mu^2}{9}} \right]^{\frac{1}{4}} . \tag{10.39}$$

By combining (10.30) (10.31) (10.39), we obtain upper and lower bounds on K_μ as

$$\frac{1}{3}\sqrt{(C/I)_{\text{req}}(C_\mu - 1)} \leq K_\mu \leq \frac{1}{C_\mu}\left(\frac{R_M}{R_\mu}\right)^2 . \tag{10.40}$$

The relation between K_μ and C_μ with R_μ/R_M as a parameter is shown in Fig. 10.26.

Example 9.3————————Consider again micro-area 56 in Fig. 10.20. Referring to Table 10.1, micro-area 56 can be assigned channel sets $[4_\alpha, 4_\gamma]$. Take channel set 4_α for example. The worst case occurs when interfering macrocell MSs transmit maximum power, i.e., at the macrocell boundary. For

the example considered, the three first-tier interfering macrocell MSs near micro-areas 45, 47, 77 are located at distances of $[\widehat{D_{M,1}}, \widehat{D_{M,2}}, \widehat{D_{M,3}}] = [2.0, 2.0, 2.0]$, respectively; the three second-tier interfering macrocell MSs near micro-areas 26, 53, 89 are located at distances $[\widehat{D_{M,4}}, \widehat{D_{M,5}}, \widehat{D_{M,6}}] = [4.36, 4.36, 4.36]$, respectively; the three third-tier interfering macrocell MSs near micro-areas 32, 38, and 98 are located at distances $[\widehat{D_{M,7}}, \widehat{D_{M,8}}, \widehat{D_{M,9}}] = [6.0, 6.0, 6.0]$, respectively. Substituting these values into (10.38) and letting $(S/I)_{\text{req}} = 18$ dB, we show in Fig. 10.27 the ratio of microcell radius to macrocell radius R_μ / R_M against $\Omega_{t,m}^\mu / \Omega_{t,m}^M$ for different values of C_μ, where $\Omega_{t,m}^\mu / \Omega_{t,m}^M$ is the ratio of the transmitted power of the microcell MS to that of the macrocell MS, and C_μ is the number of the microcell clusters in a micro-area. It is shown that as C_μ increases, microcell size becomes insensitive to $\Omega_{t,m}^\mu / \Omega_{t,m}^M$.

Suppose our objective is to implement six microcell clusters in each macro-area (i.e. $C_\mu = 6$) and still maintain $(C/I)_{\text{req}} = 18$ dB. We first need to know the feasible cluster size K_μ and the microcell radius. From Fig. 10.26, we obtain $K_\mu = 7$ and $R_\mu = 0.15 \times R_M$. Then from Fig. 10.27, we find the transmitted power for a microcell MS should be at least 0.017 times that for a macrocell MS. Consider an interfering macrocell MS which is an IS-54/136 Class IV portable handset transmitting at -2 dBW. The microcell MS transmitted power should be larger than - 20 dBW in this case. Recall the transmitted power of an IS-54/136 Class IV portable handset ranges from -22 dBW to -2 dBW. Consequently, the current IS-54/136 Class IV portable handset can be used in both the macrocells and microcells of the cluster planned architecture without changing the handset transmit power specification.

3.3.3   ADJACENT CHANNEL INTERFERENCE ANALYSIS

To avoid excessive adjacent channel interference, it is desirable not to use the same channel sets in adjacent sectors. We will first review a frequency plan designed to avoid adjacent channel interference in the conventional macrocellular system. Then we will show that the same plan works for the cluster-planned hierarchical architecture. As shown in Fig. 10.16, a traditional 7/21 macrocellular system has 21 sectors. If the forward and reverse links each have 10 MHz of available spectrum, and the channel bandwidth is 30 KHz, then a total of 333 carriers can be assigned to the 21 sectors. A frequency plan that avoids adjacent channel interference is shown in Table 10.3 [190]. Each row in the table represents a frequency set that is designated to a sector. This scheme separates any two carriers assigned to adjacent sectors by seven carriers.

If the frequency plan in Table 10.3 is applied to the cluster-planned architecture in Fig. 10.20, there is no adjacent channel interference between macrocell
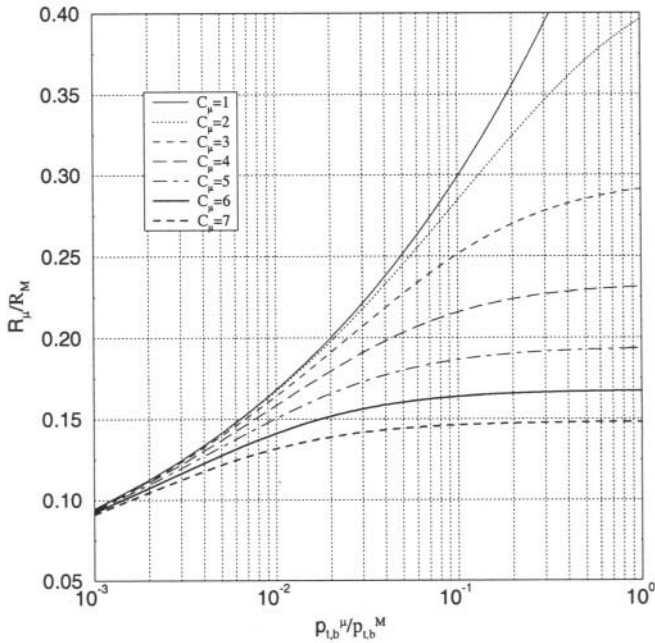
Figure 10.27. Uplink microcell radius R_μ against $\Omega_{t,m}^\mu/\Omega_{t,m}^M$ for different values of C_μ, where the microcell radius is normalized by the macrocell radius R_M, $\Omega_{t,m}^\mu/\Omega_{t,m}^M$ is the ratio of the transmitted power of the microcell MS to that of the macrocell MS, C_μ is the number of microcell clusters in a micro-area, and $(C/I)_{\text{req}} = 18$ dB, from [350].

sectors. Even with the addition of underlaid microcells, a 2-carrier separation is maintained between the carriers assigned to the microcells and the co-site macrocells withing a micro-area. For example, referring to Fig. 10.20 and Table 10.1, the channel set $[4\alpha, 4\gamma]$ is assigned to micro-area 56. The co-site macrocell sectors that use channel set $1\beta, 2\alpha$, and 7γ have at least a 2-carrier separation. This feature is valid for all the micro-areas with channel assignment of Table 10.1.

4.   MACRODIVERSITY ARCHITECTURES

Microscopic diversity techniques are used to combat the effects of envelope fading. Macrodiversity, or a large-scaled space diversity, has long been recognized as an effective tool to combat shadowing [173, 190], although it is effective against envelope fading as well. A macrodiversity system serves a mobile station (MS) simultaneously by several base stations (BSs). At any time, the BS with the best quality measure is chosen to serve the MS. The criterion for branch (or BS) selection is a key issue when designing a macrodiversity

1α	1	22 43 64 85 106 127 148 169 190 211 232 253 274 295 316
2α	2	23 44 65 86 107 128 149 170 191 212 233 254 275 296 317
3α	3	24 45 66 87 108 129 150 171 192 213 234 255 276 297 318
4α	4	25 46 67 88 109 130 151 172 193 214 235 256 277 298 319
5α	5	26 47 68 89 110 131 152 173 194 215 236 257 278 299 320
6α	6	27 48 69 90 111 132 153 174 195 216 237 258 279 300 321
7α	7	28 49 70 91 112 133 154 175 196 217 238 259 280 301 322
1β	8	29 50 71 92 113 134 155 176 197 218 239 260 281 302 323
2β	9	30 51 72 93 114 135 156 177 198 219 249 261 282 303 324
3β	10 31 52 73 94	115 136 157 178 199 220 250 262 283 304 325
4β	11 32 53 74 95	116 137 158 179 200 221 251 263 284 305 326
5β	12 33 54 75 96	117 138 159 180 201 222 252 264 285 306 327
6β	13 34 55 76 97	118 139 160 181 202 223 253 265 286 307 328
7β	14 35 56 77 98	119 140 161 182 203 224 254 266 287 308 329
1γ	15 36 57 78 99	120 141 162 183 204 225 255 267 288 309 330
2γ	16 37 58 79 100	121 142 163 184 205 226 256 268 289 310 331
3γ	17 38 59 80 101	122 143 164 185 206 227 257 269 290 311 332
4γ	18 39 60 81 102	123 144 165 186 207 228 258 270 291 312 333
5γ	19 40 61 82 103	124 145 166 187 208 221 251 271 292 313
6γ	20 41 62 83 104	125 146 167 188 209 222 252 272 293 314
7γ	21 42 63 84 105	126 147 168 189 210 223 253 273 294 315

Table 10.3. Frequency management plan for avoiding adjacent channel interference.

system. Usually, the branch selection is based on the local mean power rather than the instantaneous power [173, 376, 322, 3, 4, 350], because the branch selection algorithm cannot react to the rapidly varying instantaneous signal power. Here we focus on *local – mean – based* branch selection schemes.

Previous studies on macrodiversity systems have evaluated the co-channel interference performance with shadowing only [371], [34], [347] and shadowed Rayleigh fading channels [348]. The co-channel interference performance was also discussed in [202], but it was assumed that the branch selection was based on the instantaneous signal power. The error rate performance of macrodiversity systems has been analyzed in Gaussian noise with both

shadowing and Rayleigh (or Nakagami) fading [376], [322], [3], [4], [323]. However, these papers did not consider co-channel interference. The analysis in [351] carries this further by considering the effect of Ricean fading on a local-mean-based macrodiversity system and by considering the correlation effect of the wanted signal at different branches of a macrodiversity system.

4.1 PROBABILITY OF CO-CHANNEL INTERFERENCE OUTAGE

We now consider an analytical model for calculating the probability of co-channel interference outage, O_I, for an L-branch local-mean-based macro-diversity system with log-normal shadowing. Our model assumes that the local mean envelope power of the desired signal, $\Omega_{d,k}$, is available for each branch k, where $k = 1, \ldots, L$. In practice, the desired signal power is mixed with the total interference power for each branch $\Omega_{I,k}$, so that $\Omega_{d,k} + \Omega_{I,k}$ is actually measured. However, the difference is small for large $\Omega_{d,k}/\Omega_{I,k}$. If the branch having the largest $\Omega_{d,k}$ is selected, then the local-mean envelope power of the selected branch is

$$\Omega_d = \max(\Omega_{d,1}, \Omega_{d,2}, \cdots, \Omega_{d,L}) \quad . \tag{10.41}$$

Let $F_k(x)$ and $p_k(x)$ denote the cumulative distribution function (cdf) and the pdf of $\Omega_{d,k}$, respectively. If the $\Omega_{d,k}$ are *independent* random variables with the pdf in (2.200), then Ω_d has the pdf $p_{\Omega_d}(y) = L\left[F_k(y)\right]^{L-1} p_k(y)$. The probability of co-channel interference outage is

$$
\begin{aligned}
O_I &= \Pr\left[\Omega_d/\Omega_I < \Lambda_{\text{th}}\right] \\
&= 1 - \int_0^\infty \left[\int_{-\infty}^{x/\Lambda_{\text{th}}} \Omega_I(y)dy\right]\Omega_d(x)dx \quad ,
\end{aligned}
\tag{10.42}
$$

where Ω_d and Ω_I are the total powers of the desired and interfering signals for the selected branch with pdfs $p_{\Omega_d}(x)$ and $p_{\Omega_I}(y)$, respectively, and Λ_{th} is the threshold C/I.

The interfering signals add noncoherently so that the total interference power on the kth branch is $\Omega_{I,k} = \sum_{i=1}^{N_I} \Omega_{I,k,i}$, where N_I is the number of interferers and $\Omega_{I,k,i}$ is the power of the ith interferer on the kth branch. It is widely accepted that $\Omega_{I,k}$ can be approximated by a log-normal random variable with area mean power $\mu_{\Omega_{I,k}}$ and standard deviation $\sigma_{\Omega_{I,k}}$. As discussed in Chapter 3.1, the parameters $\sigma_{\Omega_{I,k}}$ and $\mu_{\Omega_{I,k}}$ can be calculated by using a variety of methods, including the Fenton-Wilkinson and Schwartz-and-Yeh methods.

If the $\{\Omega_{I,k}\}_{k=1}^n$ are independent and identically distributed (iid), and the $\{\Omega_{d,k}\}_{k=1}^L$ are also iid and independent of the $\{\Omega_{I,k}\}_{k=1}^n$, then [371, 347]

$$O_I = 1 - L \int_0^\infty \left[\int_0^{x/\Lambda_{th}} \frac{1}{\sqrt{2\pi}\sigma_{\Omega_I}\xi y} \right.$$

$$\times \exp\left[\frac{-\left(10\log_{10}y - \mu_{\Omega_I \text{ (dB)}}\right)^2}{2\sigma_{\Omega_I}^2} \right] dy \Bigg]$$

$$\times \left[1 - Q\left(\frac{10\log_{10}x - \mu_{\Omega_d \text{ (dB)}}}{\sigma_{\Omega_d}} \right) \right]^{L-1}$$

$$\times \frac{1}{\sqrt{2\pi}\sigma_{\Omega_d}\xi x} \exp\left[\frac{-\left(10\log_{10}x - \mu_{\Omega_d \text{ (dB)}}\right)^2}{2\sigma_{\Omega_d}^2} \right] dx \quad (10.43)$$

where σ_{Ω_d} and $\mu_{\Omega_d \text{ (dB)}}$ are the shadowing standard deviation and area mean power of the desired signal on the kth diversity branch, respectively.

For ease of evaluation, we let $w = (10\log_{10}x - \ln\mu_{\Omega_d \text{ (dB)}})/\sqrt{2}\sigma_{\Omega_d}$ and transform (10.43) into a Hermite integration form. That is,

$$O_I = 1 - \int_{-\infty}^\infty g(w)\exp(-w^2)dw \simeq 1 - \sum_{i=1}^n g(w_i)h_i , \qquad (10.44)$$

where

$$g(w) = \frac{L}{\sqrt{\pi}} \left[1 - Q\left(\frac{\sqrt{2}\sigma_{\Omega_d}w + \xi[\mu_{\Omega_d \text{ (dB)}} - \mu_{\Omega_I \text{ (dB)}} - \Lambda_{th(dB)}]}{\sigma_{\Omega_I}} \right) \right]$$

$$\times \left[1 - Q\left(\sqrt{2}w \right) \right]^{L-1} , \qquad (10.45)$$

and w_i and h_i are the roots and weight factors of the nth-order Hermite polynomial, respectively [1].

4.2   SHADOW CORRELATION

Until now, we have assumed independent shadowing on the macrodiversity branches. However, in many cases the macrodiversity branches will be correlated. Define

$$\Omega_d = (\Omega_{d,1}, \Omega_{d,2}, \cdots, \Omega_{d,L}) . \qquad (10.46)$$

For a correlated L-branch macrodiversity system, the joint pdf of Ω_d is [78]

$$p_{\Omega_d}(\mathbf{z}) = \frac{\exp\left[-\frac{1}{2}\mathbf{Y}^T\mathbf{M}^{-1}\mathbf{Y}\right]}{\sqrt{(2\pi)^L \det(\mathbf{M})}\xi^L z_1 \cdots z_L} \qquad (10.47)$$

where $\mathbf{z} = (z_1, \cdots, z_L)$, $\mathbf{Y}^T = [y_1, \cdots, y_L]$ denotes the transpose of column vector

$$\mathbf{Y} = \begin{bmatrix} 10\log_{10}(z_1) - \mu_{\Omega_{d,1}} \text{ (dB)} \\ \vdots \\ 10\log_{10}(z_L) - \mu_{\Omega_{d,L}} \text{ (dB)} \end{bmatrix} \tag{10.48}$$

and $\mu_{\Omega_{d,1}}$ (dB)$, \cdots, \mu_{\Omega_{d,L}}$ (dB) are the area means of each diversity branch. The covariance matrix M is expressed as

$$\mathbf{M} = \begin{bmatrix} \sigma_{\Omega_1}^2 & \cdots & \nu_{1L} \\ \vdots & \ddots & \vdots \\ \nu_{L1} & \cdots & \sigma_{\Omega_L}^2 \end{bmatrix} \tag{10.49}$$

where σ_Ω is the shadowing standard deviation and ν_{ij} is the covariance of $\Omega_{d,i}$ (dB) and $\Omega_{d,j}$ (dB)

$$\nu_{ij} = \mathrm{E}\left[\left(\Omega_{d,i} \text{ (dB)} - \mu_{\Omega_{d,i}} \text{ (dB)} \right) \left(\Omega_{d,j} \text{ (dB)} - \mu_{\Omega_{d,j}} \text{ (dB)} \right) \right] . \tag{10.50}$$

It is convenient to define $\mathbf{N} = \mathbf{M}^{-1}$ and express the matrix multiplication in (10.47) in the form

$$\mathbf{Y}^T\mathbf{N}\mathbf{Y} = \sum_{i=1}^{L} N_{ii}y_i^2 + 2\sum_{i=1}^{L-1}\sum_{j=i+1}^{L} N_{ij}y_iy_j \tag{10.51}$$

where N_{ij} is the element in the ith row and jth column.

According to (10.41), (10.47), and (10.51), the probability that $\Omega_d < y$ is

$$\Pr(\Omega_d < y) = \int_0^y \cdots \int_0^y \frac{1}{\sqrt{(2\pi)^L \det(\mathbf{M})}\xi^L z_1 \cdots z_L}$$

$$\exp\left[-\frac{1}{2}\left(\sum_{i=1}^{L} N_{ii}y_i^2 + 2\sum_{i=1}^{L-1}\sum_{j=i+1}^{L} N_{ij}y_iy_j \right) \right] d\mathbf{z} \tag{10.52}$$

where N_{ij} and y_i are defined in (10.51) and (10.48), respectively.

The key for obtaining the probability of co-channel interference outage of the local-mean-based macrodiversity system is to find the pdf of the combiner output power, $p_{\Omega_d}(y)$. Unlike the uncorrelated case where there exists a closed-form expression for $p_{\Omega_d}(y)$, one can not easily get a simple closed formula for the joint distribution of more than two mutually correlated log-normal random variables. However, for $L = 2$ and $\mu_{\Omega_{d,j}}$ (dB) $= \mu_{\Omega_d}$ (dB)

$$
p_{\Omega_d}(z) = \frac{1}{\sqrt{2\pi \det(\mathbf{M})}\xi^2} \left\{ \frac{1}{\sqrt{N_{22}}} \exp\left[-\frac{z^2}{2}\left(N_{11} - \frac{N_{12}}{N_{22}}\right)\right] \right.
$$

$$
\left[1 - Q\left(\left(\sqrt{N_{22}} + \frac{N_{12}}{\sqrt{N_{22}}}\right)z\right)\right]
$$

$$
+ \frac{1}{\sqrt{N_{11}}} \exp\left[-\frac{z^2}{2}\left(N_{22} - \frac{N_{12}}{N_{11}}\right)\right]
$$

$$
\left. \left[1 - Q\left(\left(\sqrt{N_{11}} + \frac{N_{12}}{\sqrt{N_{11}}}\right)z\right)\right] \right\}
\tag{10.53}
$$

where $z = (10\log_{10}y - \mu_{\Omega_d} \text{ (dB)})$, Consider the following covariance matrix $\mathbf{M}$

$$
\mathbf{M} = \begin{bmatrix} \sigma^2 & \nu \\ \nu & \sigma^2 \end{bmatrix},
\tag{10.54}
$$

and

$$
\mathbf{N} = \mathbf{M}^{-1} = \frac{1}{\sigma^4 - \nu^2}\begin{bmatrix} \sigma^2 & -\nu \\ -\nu & \sigma^2 \end{bmatrix}.
\tag{10.55}
$$

By substituting (10.55) into (10.53), we express the pdf of the output local-mean power of the dual macrodiversity system as

$$
p_{\Omega_d}(y) = \frac{\sqrt{2}}{\sqrt{\pi}\sigma\xi^2 y}\left\{1 - Q\left[\left(\frac{1-r}{\sqrt{1-r^2}}\right)\left(\frac{10\log_{10}y - \mu_{\Omega_d} \text{ (dB)}}{\sigma}\right)\right]\right.
$$

$$
\left. \exp\left[-\frac{(10\log_{10}y - \mu_{\Omega_d} \text{ (dB)})^2}{2\sigma^2}\right]\right\}
\tag{10.56}
$$

where the correlation coefficient r is defined as $r = \nu/\sigma^2$. Combining (10.42) and (10.56), gives

$$
O_I = 1 - \int_{-\infty}^{\infty} g(w)\exp(-w^2)dw \simeq 1 - \sum_{i=1}^{n} g(w_i)h_i,
\tag{10.57}
$$

where

$$
g(w) = \frac{2}{\sqrt{\pi}}\int_{-\infty}^{\infty}\left[1 - Q\left(\frac{\sqrt{2}\sigma_d w + [\mu_{\Omega_d} \text{ (dB)}} - \mu_{\Omega_I} \text{ (dB)} - \Lambda_{\text{th (dB)}}]}{\sigma_{\Omega_I}}\right)\right]
$$

$$
\times \left[1 - Q\left(\frac{1-r}{\sqrt{1-r^2}}\sqrt{2}w\right)\right].
\tag{10.58}
$$

4.3 NUMERCIAL EXAMPLES

Consider a cellular system with nine cells per cluster. In this case, two co-channel interferers are at *5.2R*, where R is the cell radius. Assume the mobile
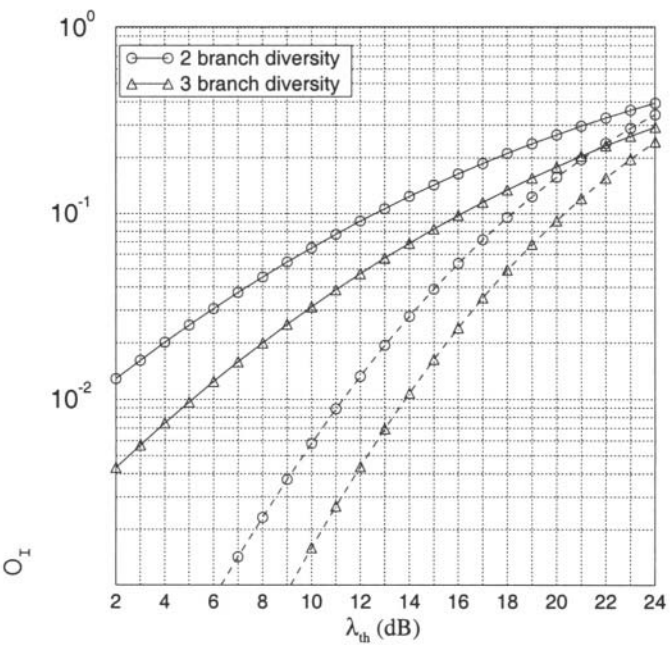
Figure 10.28. Probability of co-channel interference outage, O_I, against the required threshold, Λ_{th}, at the receiver for the local-mean-based macrodiversity system, where the solid lines (———) denote the case for shadowing standard deviation $\sigma = 10$ dB and the dashed lines ($- - -$) for $\sigma = 6$ dB; $a = b = 2$, $g = 0.15R$; two interferers are located at a distance of 5.2R.

unit is on the boundary of the cell at a distance of R to the BS. Consider a dual slope path loss model with $a = b = 2$ and $g = 0.15\,R$ in (2.248).

Fig. 10.28 shows the probability of co-channel interference outage performance, while Tab. 10.4 lists the threshold Λ_{th} and diversity gain (D.G.) in terms of 5 and 10 % co-channel interference outage probabilities. Diversity gain here is defined as the additional C/I (in dB) that is required by a system without diversity to produce the same probability of co-channel interference outage. Some general observations can be made: 1) a higher shadowing spread leads to a higher diversity gain and a lower required threshold Λ_{th}; **2)** the diversity gain per branch is decreased as the number of diversity branches is increased; 3) the diversity gain increases with the requirement of the system, e.g., the diversity gain for a 5 % outage probability is higher than that for a 10 % outage probability.

We evaluate the effects of correlation coefficient r on a 2-branch macrodiversity system with various σ; $\sigma = 6$ dB in Fig. 10.29 and $\sigma = 10$dB in Fig. 10.30. With respect to a 10 % outage, Tab. 10.5 lists Λ_{th} with different r. Observe

L	$O_I = 5\%$		$O_I = 10\%$	
	λ_{th}	D. G.	λ_{th}	D. G.
1	10.96	-	13.69	-
2	15.78	4.82	18.12	4.43
3	17.97	7.01	20.46	11.78
4	19.41	8.45	21.80	13.13

Table 10.4.   Macrodiversity gain (D. G.) and the threshold λ_{th} of C/I set at the receiver in terms of 5 % and 10 % probability of co-channel interference outage, O_I, over a pure shadowing channel; $\sigma = 6$ dB.
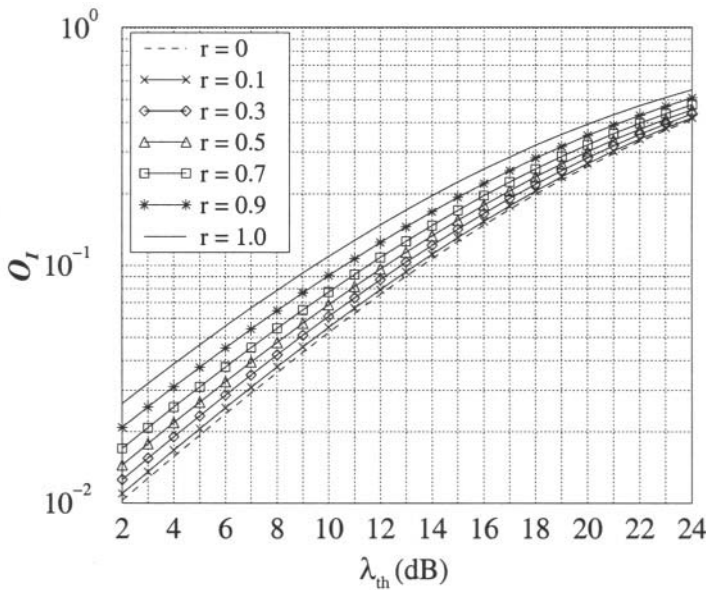
Figure 10.29.   Effect of branch correlation coefficient r on the local-mean-based macrodiversity system with $\sigma = 6$ dB; $a = b = 2$, $g = 0.15$R; two interferers are located at a distance of 5.2R.

that as r approaches unity, the diversity gain becomes zero. Furthermore, for $r = 0.7$, the diversity gain will be reduced to about 50 % of the gain when $r = 0$.

APPENDIX 10.A: Derivation of Equation (10.7)

The conditional pdf of the received squared envelope, α^2, is

$$p_{\alpha^2|\Omega_p}(x) = \frac{2}{\Omega_p}e^{-x/\Omega_p}(1 - e^{-x/\Omega_p}) \ . \tag{10-10.A.1}$$
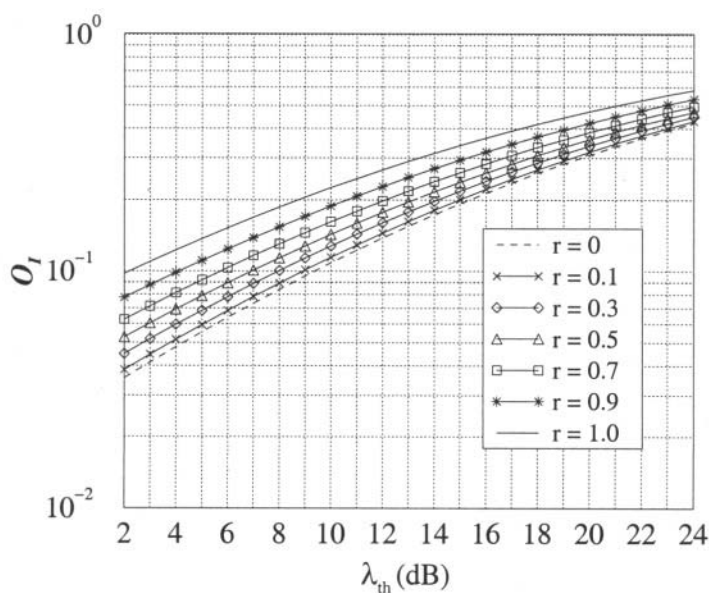
Figure 10.30. Effect of branch correlation coefficient r on the local-mean-based macrodiversity system with $\sigma = 10$ dB; $a = b = 2$, $g = 0.15R$; two interferers are located at a distance of 5.2R.

r	$\sigma = 6$ dB		$\sigma = 10$ dB	
	λ_{th}	D. G.	λ_{th}	D. G.
0	13.54	3.99	9.24	6.96
0.1	13.39	3.84	8.93	6.65
0.3	12.78	3.23	7.94	5.66
0.5	12.23	2.68	7.00	4.72
0.7	11.46	1.91	5.67	3.39
0.9	10.51	1.02	4.12	1.84
1.0	9.55	–	2.28	–

Table 10.5. Effects of branch correlation on a 2-branch macrodiversity.

Averaging over distribution of log-normal shadowing yields the composite pdf for squared envelope

$$p_{\alpha_c^2}(x) = \int_0^\infty \frac{2}{w} e^{-x/w}\left(1 - e^{-x/w}\right) \qquad (10\text{-}10.\text{A}.2)$$

$$\times \frac{1}{\sqrt{2\pi}\xi\sigma_\Omega w} \exp\left\{-\frac{\left(10\log_{10}w - \mu_{\Omega_p\,(dBm)}\right)^2}{2\sigma_\Omega^2}\right\} dw$$

The mean of the approximate log-normal distribution is

$$
\begin{aligned}
\mu_{(\text{dBm})} &= \mathrm{E}[10\log_{10}\alpha_c^2] \\
&= \int_0^\infty \int_0^\infty (10\log_{10}x)\frac{2}{w}e^{-x/w}(1-e^{-x/w}) \\
&\quad \times \frac{1}{\sqrt{2\pi}\xi\sigma_\Omega w}\exp\left\{-\frac{(10\log_{10}w-\mu_{\Omega_{p\,(\text{dBm})}})^2}{2\sigma_\Omega^2}\right\}\,dx\,dw \\
&= \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma_\Omega\xi^2 w^2}\exp\left\{-\frac{(10\log_{10}w-\mu_{\Omega_{p\,(\text{dBm})}})^2}{2\sigma_\Omega^2}\right\} \\
&\quad \times \int_0^\infty 2\ln(x)e^{-x/w}\left(1-e^{-x/w}\right)\,dx\,dw \ . \quad (10\text{-}10.\text{A}.3)
\end{aligned}
$$

From [147, 4.352.1], the inner integral becomes

$$
\int_0^\infty 2\ln(x)e^{-x/w}(1-e^{-x/w})dx = w[\ln(2)+\ln(w)-C] \quad (10\text{-}10.\text{A}.4)
$$

where $C \simeq 0.5772$ is *Euler's constant*. Hence,

$$
\mu_{(\text{dBm})} = \xi^{-1}[\ln(2)-C] + \mu_{\Omega_{p\,(\text{dBm})}} \quad (10\text{-}10.\text{A}.5)
$$

In a similar fashion, the mean square value is

$$
\begin{aligned}
\mathrm{E}[(10\log_{10}(\alpha_c^2))^2] &= \int_0^\infty \int_0^\infty (10\log_{10}x)^2\frac{2}{w}e^{-x/w}(1-e^{-x/w}) \\
&\quad \times \frac{1}{\sqrt{2\pi}\sigma_\Omega\xi w}\exp\left\{-\frac{(10\log_{10}w-\mu_{\Omega_{p\,(\text{dBm})}})^2}{2\sigma_\Omega^2}\right\}\,dx\,dw \\
&= \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma_\Omega\xi^3 w^2}\exp\left\{-\frac{(10\log_{10}w-\mu_{\Omega_{p\,(\text{dBm})}})^2}{2\sigma_\Omega^2}\right\} \\
&\quad \times \int_0^\infty 2(\ln x)^2 e^{-x/w}(1-e^{-x/w})dx\,dw \ . \quad (10\text{-}10.\text{A}.6)
\end{aligned}
$$

From [147, 4.358.2], the inner integrals become

$$
\begin{aligned}
\int_0^\infty 2(\ln x)^2 e^{-x/w}dx &= 2w([\ln w - C] + \zeta(2,1)) \\
\int_0^\infty (\ln x)^2 e^{-2x/w}dx &= w([\ln w - \ln(2) - C]^2 + \zeta(2,1)) \quad (10\text{-}10.\text{A}.7)
\end{aligned}
$$

where

$$
\zeta(2,m) = \sum_{k=0}^\infty \frac{1}{(m+k)^2} \quad (10\text{-}10.\text{A}.8)
$$

is *Reimann's zeta function*. Finally, the variance is

$$
\begin{aligned}
\sigma^2 &= \mathrm{E}[(10\log_{10}\alpha_c^2)^2] - \mathrm{E}^2[10\log_{10}\alpha_c^2] \\
&= \xi^{-2}[\zeta(2,1) - 2(\ln 2)^2] + \sigma_\Omega^2 \ . \quad (10\text{-}10.\text{A}.9)
\end{aligned}
$$

Problems

10.1. Consider a cellular system that uses a 7-cell reuse cluster without cell sectoring.

    a) Show graphically the worst case CCI situation for the reverse channel.

    b) Ignoring envelope fading and shadowing and assuming the simple path loss model in (1.7), calculate the worst case carrier-to-interference ratio in terms of the co-channel reuse factor D/R.

    c) Repeat parts a) and b) if $120°$ cell sectoring is used.

10.2. One method for improving the capacity of a cellular system employs a *two-channel bandwidth* scheme as suggested by Lee [192], where a hexagonal cell is divided into two concentric hexagons as shown in Fig. 10.A.I below. The inner hexagon is serviced by 15 kHz channels, while the outer hexagon is serviced by 30 kHz channels. Suppose that the 30 kHz channels require $\Lambda = 18$ dB to maintain an acceptable radio link quality, while the 15 kHz channels require $\Lambda = 24$ dB.

Assume a fourth-law path loss model and suppose that the effects of envelope fading and shadowing can be ignored. Consider the mobile-to-base link and suppose that there are six co-channel interferers at distance D from the BS. For a 7-cell reuse cluster, it follows that the worst case carrier-to-interference ratio, Λ, when a mobile station (MS) is located at distance d from the BS is $\Lambda = (D/d)^4/6$. Hence, $\Lambda = 18$ dB requires $D/R_o = 4.6$, and $\Lambda = 24$ dB requires $D/R_i = 6.3$, where R_i and R_o are the radii of the inner and outer cells, respectively.

    a) Use the values of D/R_i and D/R_o to determine the ratio of the inner and outer cell areas, A_i/A_o.

    b) Let N_i and N_o be the number of channels that are allocated to the inner and outer portions of each cell, and assume that the channels are assigned such that $N_i/N_o = A_i/(A_o - A_i)$, Determine the increase in capacity (as measured in channels per cell) over a conventional *one-channel bandwidth* system that uses only 30 kHz channels.

10.3. It has been suggested by [192] that the two-channel bandwidth scheme in Problem 10.2 can be combined with Halpern's reuse partitioning scheme. In this case, 15 kHz channels are used in the inner cells and 30 kHz channels are used in the outer cells. In order to have adequate performance in the inner or low bandwidth ring we must have $D_i/R_i = 6.3$, while the outer higher bandwidth ring can use $D_o/R_o = 4.6$.

Compute the increase in capacity (as measured in channels per cell) that will result from using this scheme, as compared to a conventional system using a 7-cell reuse cluster.
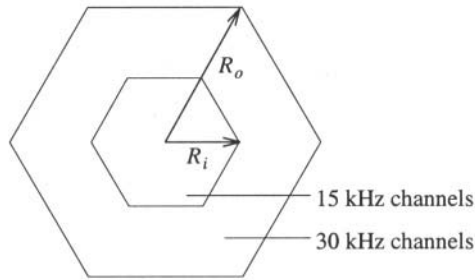
Figure 10.A.1. Cell division with two channel bandwidth scheme.

10.4. In Section 1.1 the worst case forward channel carrier-to-interference ratio was calculated by considering only the first tier of co-channel interferers. Calculate the amount of interference from the second tier of co-channel interferers. Is it reasonable to neglect this interference?

10.5. Microcells are characterized by very erratic propagation environments. This problem is intended to illustrate the imbalance in the forward and reverse channel carrier-to-interference ratio that could occur in a street microcell deployment. Consider the scenario shown in Fig. 10.A.2, that consists of two co-channel **BSs**, BS_1 and BS_2, communicating with two co-channel **MSs**, MS_1 and MS_2. Neglect the effects of shadowing and multipath, and assume that the NLOS corner path loss model in (2.214). Suppose that a $= 2$, $b = 4$, and $g = 150$ m. Plot Λ at BS_1, BS_2, MS_1, and MS_2 as MS_2 moves from A to C. When plotting your results, assume a received power level of 1 dBm at a distance of one meter.
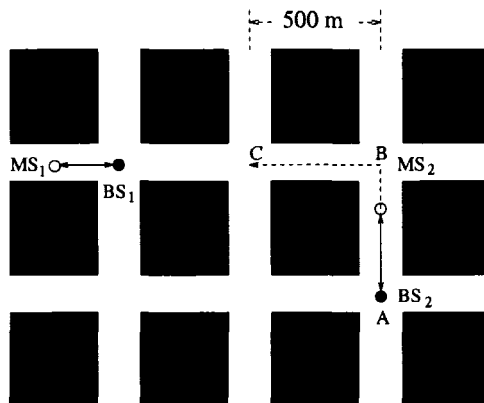
Figure 10.A.2. Microcellular propagation environment for Problem 10.5.

This page intentionally left blank

Chapter 11

CDMA CELLULAR ARCHITECTURES

CDMA is an attractive proposition for increasing cellular system capacity in dense urban areas, due to its many inherent benefits like the ability to mitigate multipath fading and interference, universal frequency reuse, soft handoff capability, and the ability to exploit voice activity detection. Numerous authors have investigated the capacity and performance of CDMA cellular systems for a propagation environment characterized by path loss and shadowing, including Gilhousen *et al.* [136], Kudoh and Matsumoto [183], and Newson and Heath [247]. Mokhtar and Gupta [228] considered reverse channel capacity on shadowed Nakagami fading channels, where the desired and interfering signals have the same fading statistical characteristics.

CDMA systems must use reverse channel power control; otherwise, the link performance will suffer from the **near-far effect,** a condition where the transmissions received from distant MSs experience excessive interference from nearby MSs. The IS-95 reverse link employs a fast closed-loop power control algorithm to combat variations in the received signal power due to path loss, shadowing, and fast envelope fading (at low Doppler frequencies). A large number of power control algorithms have been suggested in the literature. Ariyavistitakul and Chang [15] proposed a fast signal-to-interference ratio (SIR) based feedback power control algorithm that can mitigate both multipath fading and shadowing. For our purpose, we consider a simple closed-loop reverse channel power control scheme that equalizes the received power C from all MSs that are served by the same BS cell sector.

Power control is also useful on the forward channel of CDMA systems for combating the **corner effect,** a condition where a MS experiences a decrease in received signal strength and an increase in multiple-access interference as it exits a cell corner. Various "power balancing" schemes have been proposed to balance the BS transmit power for each MS [136] , [46]. Chang and Ren

[46] have compared power balancing and mobile assisted SIR-based forward channel power control algorithms. They have shown the former to be better than the latter. Zorzi [382] has analyzed some simplified power control algorithms in the absence of shadowing. For our purpose, we will assume a slow open-loop power balancing algorithm.

1. CAPACITY OF CELLULAR CDMA

CDMA cellular systems typically employ **universal frequency reuse,** where the bandwidth is shared by all the cells and transmissions are distinguished through the assignment of unique spreading sequences. For such systems, multiple-access interference from neighboring cells must be carefully accounted for. The propagation path loss associated with these interfering signals is relatively small compared to those found in narrow-band and mid-band TDMA systems that employ frequency reuse plans.

With cellular CDMA systems, any technique that reduces multiple-access interference translates into a capacity gain. Since cellular CDMA systems use speech coding, the multiple-access interference can be reduced by using voice activity detection along with variable rate speech transmission. This technique reduces the rate of the speech coder when silent periods are detected in the speech waveform. Voice activity detection has often been cited as an advantage of CMDA systems over TDMA systems. However, TDMA systems can also benefit from voice activity detection and discontinuous transmission, through a reduction in the level of co-channel interference.

Cell sectoring is another very effective method for reducing multiple-access interference, where each cell is sectored by using directional antennas. With 120° cell sectors, multiple-access interference on the reverse channel will only arise from MSs that are located in the shaded area of Fig. 11.1, where only the adjacent cells are shown. Likewise, multiple-access interference on the forward channel is generated by BSs that are transmitting to MSs located in the shaded regions of Fig. 11.2, where again only the adjacent cells are shown. In either case, 120° cell sectoring reduces the multiple-access interference by roughly a factor of three (on average); we say on average because the MSs are randomly distributed throughout the plane. Further improvements can be gained by using simple switched beam smart antenna systems with 30° or 15° sectors. A straight forward application of these antenna systems reduces the multiple access interference by a factor of 12 and 24, respectively, over a system using omni-directional antennas.

Our analysis of cellular CDMA starts with a cellular layout described by a uniform plane of hexagonal cells of radius R. Each cell contains a centrally located BS with 120° cell sectors. It is further assumed that the MSs are uniformly distributed throughout the system area with a density of K MSs per
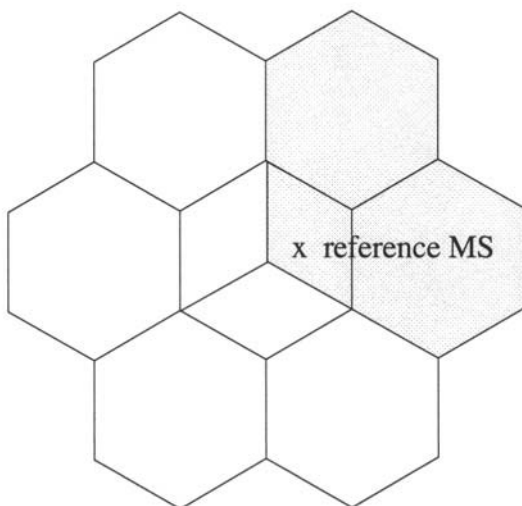
Figure 11.1.   Reverse channel transmissions from MSs located in the shaded area will cause multiple-access interference with the reverse channel transmission from the reference MS.

cell sector. For hexagonal cells of radius R, this yields a subscriber density of

$$\rho = \frac{2K}{3\sqrt{3}R^2} \text{ per unit area} .\tag{11.1}$$

The effects of voice activity detection can be modeled by assuming that each transmitter is independently active with probability p, so that the number of active transmitters in each cell has a (K,p) binomial distribution. The average number of active transmitters in a cell sector is Kp.

The standard Gaussian approximation in Chapter 9.6.1 has been extensively employed in the literature for the performance prediction of cellular CDMA systems. For random spreading sequences, we have seen that the standard Gaussian approximation for a power controlled chip and phase asynchronous reverse channel of a CDMA system predicts a bit signal-to-noise ratio of $\gamma_b = 3G/(2(K-1))$, where G is the processing gain and K is the number of simultaneously received signals. This assumes the use of a coherent correlation receiver with bit-by-bit decisions. If the signals are chip and phase synchronous as is the case in the forward channel of a CDMA cellular system, then the standard Gaussian approximation yields a bit signal-to-noise ratio of $\gamma_b = G/(2(K-1))$. However, it is important to realize that this expression assumes random spreading sequences. If orthogonal spreading codes such as Walsh-Hadamard codes are used on the forward channel as is the case with the IS-95 system [96], then the multiple-access interference from the serving BS is effectively zero unless the channel delay spread destroys the orthogonality of
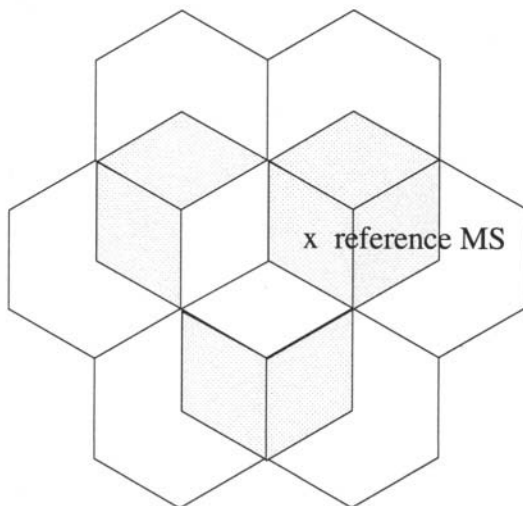
Figure 11.2. Forward channel transmissions to MSs located in the shaded areas will cause multiple-access interference with the forward channel transmission to the reference MS.

the received signals. Under this condition out-of-cell interference dominates the performance.

1.1 REVERSE LINK CAPACITY

Perfect reverse channel power control maintains a constant received power C at a BS for all MSs served by that BS. The *jth* MS located in cell i is denoted by MS_{ij}. The power transmitted by MS_{ij}, located at distance d_{ij} from its serving BS, BS_i, is P_{ij}. The received power at BS_i is

$$C = P_{ij}10^{\frac{\varsigma_{ij}}{10}} \tag{11.2}$$

where ς_{ij} is a random variable due to shadowing and fading. MS_{ij} is also at distance d_{i0} to the reference BS, BS_0, and will produce an out-of-cell interference equal to

$$
\begin{aligned}
\frac{I_o(ij)}{C} &= 10^{\varsigma_{i0}/10} \cdot \left(\frac{1}{10^{\varsigma_{ij}/10}}\right) \\
&= 10^{(\varsigma_{i0}-\varsigma_{ij})/10} \leq 1 \ . \tag{11.3}
\end{aligned}
$$

The first term is due to path loss and shadowing to BS_0, while the second term is the effect of the power control to compensate for the corresponding attenuation to BS_j. Note that $I_o(ij)/C$ is always less than unity; otherwise the MS would execute a handoff to the BS which makes it less than unity.

For our purpose, we assume a shadowed Nakagami fading channel, where the received signal power has the composite *Gamma*-log-normal pdf in (2.187).

The composite pdf is approximated by a purely log-normal pdf with mean and standard deviation given by (2.188). Hence, the random variables ς_{ij} and ς_{i0} are treated as Gaussian random variables with means and variances, respectively,

$$\mu_{ij \text{ (dBm)}} = \xi^{-1}[\psi(m_{ij}) - \ln(m_{ij})] + \mu_{\Omega_p \text{ (dBm)}}(ij) \qquad (11.4)$$

$$\mu_{i0 \text{ (dBm)}} = \xi^{-1}[\psi(m_{i0}) - \ln(m_{i0})] + \mu_{\Omega_p \text{ (dBm)}}(i0)$$

$$\sigma_{ij}^2 = \xi^{-2}\zeta(2, m_{ij}) + \sigma_\Omega^2$$

$$\sigma_{i0}^2 = \xi^{-2}\zeta(2, m_{i0}) + \sigma_\Omega^2$$

where $\xi = (\ln 10)/10$, m_{ij} and m_{i0} are the Nakagami shape factors, and σ_Ω is the shadow standard deviation. The parameters $\mu_{\Omega_p \text{ (dBm)}}(ij)$ and $\mu_{\Omega_p \text{ (dBm)}}(i0)$ are determined by the path loss. Using the simple path loss model in (1.7), their difference is

$$\mu_{\Omega_p \text{ (dBm)}}(ij) - \mu_{\Omega_p \text{ (dBm)}}(i0) = 10\beta \log_{10}(d_{ij}/d_{i0}) \ . \qquad (11.5)$$

The total out-of-cell interference-to-signal ratio is equal to

$$\frac{I_o}{C} = \int\int \chi I_o(ij)\Phi_{0j}\rho dA \qquad (11.6)$$

where

$$\Phi_{0j} = \begin{cases} 1, & \text{if } 10^{(\varsigma_{i0}-\varsigma_{ij})/10} < 1 \\ 0, & \text{otherwise} \end{cases} \qquad (11.7)$$

ρ is user density over the area *A,* and χ is the voice activity variable

$$\chi = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{with probability } 1-p \end{cases} \ . \qquad (11.8)$$

The total out-cell interference I_o can be modeled as a Gaussian random variable by invoking the central limit theorem. The mean of the total out-of-cell interference-to-carrier ratio is

$$E[I_o/C] = \int\int E[\chi 10^{(\varsigma_{i0}-\varsigma_{ij})/10}\Phi_{0j}]\rho dA$$

$$= \int\int p E[10^{(\varsigma_{i0}-\varsigma_{ij})/10}\Phi_{0j}]\rho dA \qquad (11.9)$$

Let $x = \varsigma_{i0} - \varsigma_{ij}$ and define

$$\mu_{x \text{ (dB)}} = \mu_{i0 \text{ (dBm)}} - \mu_{ij \text{ (dBm)}}$$

$$\sigma_x^2 = \sigma_{i0}^2 + \sigma_{ij}^2 \ . \qquad (11.10)$$

Then the inner expectation in (11.9) is

$$
\begin{aligned}
E[10^{(\varsigma_{i0}-\varsigma_{ij})/10}\Phi_{0j}] &= \int_{-\infty}^{0} e^{\xi x} \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left\{-\frac{(x-\mu_{x\,(dB)})^2}{2\sigma_x^2}\right\} dx \\
&= \exp\left\{\frac{\xi^2\sigma_x^2}{2}+\xi\mu_{x\,(dB)}\right\} \frac{1}{\sqrt{2\pi}\sigma_x} \\
&\quad \times \int_{-\infty}^{0} \exp\left\{-\left(\frac{x}{\sqrt{2}\sigma_x}-\frac{\xi\sigma_x}{\sqrt{2}}-\frac{\mu_{x\,(dB)}}{\sqrt{2}\sigma_x}\right)^2\right\} dx \\
&= \exp\left\{\frac{\xi^2\sigma_x^2}{2}+\xi\mu_{x\,(dB)}\right\} \\
&\quad \times \left[1-Q\left(-\xi\sigma_x-\frac{\mu_{x\,(dB)}}{\sigma_x}\right)\right] \quad .
\end{aligned}
$$

$$(11.11)$$

Therefore,

$$
\begin{aligned}
E[I_o/C] &= p\int\int \exp\left\{\frac{\xi^2\sigma_x^2}{2}+\xi\mu_{x\,(dB)}\right\} \\
&\quad \times \left[1-Q\left(-\xi\sigma_x-\frac{\mu_{x\,(dB)}}{\sigma_x}\right)\right]\rho dA \quad .
\end{aligned}
$$

$$(11.12)$$

In a similar fashion,

$$
\begin{aligned}
E[(I_o/C)^2] &= \int\int E[\chi^2 10^{(\varsigma_{i0}-\varsigma_{ij})/5}\Phi_{0j}^2]\rho dA \\
&= p\int\int\int_{-\infty}^{0} e^{2\xi x} \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left\{-\frac{(x-\mu_{x\,(dB)})^2}{2\sigma_x^2}\right\} dx\rho dA \\
&= p\int\int \exp\left\{2\xi^2\sigma_x^2+2\xi\mu_{x\,(dB)}\right\} \\
&\quad \left[1-Q\left(-2\xi\sigma_x-\frac{\mu_{x\,(dB)}}{\sigma_x}\right)\right]\rho dA
\end{aligned}
$$

$$(11.13)$$

Finally, the variance of I_o/C is

$$
\begin{aligned}
Var[I_o/C] &= p\int\int \exp\left\{2\xi^2\sigma_x^2+2\xi\mu_{x\,(dB)}\right\} \\
&\quad \left[1-Q\left(-2\xi\sigma_x-\frac{\mu_{x\,(dB)}}{\sigma_x}\right)\right]\rho dA \\
&\quad -E^2[I_o/C] \quad .
\end{aligned}
$$

$$(11.14)$$

The integrals in (11.12) and (11.14) must be numerically evaluated over the random mobile locations in the area *A,* as defined by the set of shaded sectors in Fig. 11.1.

With perfect power control, the in-cell interference I_{in} is

$$I_{\text{in}} = C \sum_{i=1}^{K-1} \chi_i \tag{11.15}$$

where χ_i is a Bernoulli random variable equal to 1 with probability p and 0 with probability $1 - p$. Let $I = I_0 + I_{\text{in}}$ be the total interference. Then the probability that the received γ_b at a BS is below a required value, $\gamma_{b \text{ req}}$, is

$$
\begin{aligned}
P_{\text{out}} &= \text{P}(\gamma_b < \gamma_{b \text{ req}}) \\
&= \text{P}((I/C)_{\gamma_b} > (I/C)_{\gamma_b \text{ req}}) \\
&= \text{P}\left(\sum_{i=1}^{K-1} \chi_i + I_0/C > (I/C)_{\gamma_b \text{ req}}\right) \\
&= \text{P}\left(I_0/C > (I/C)_{\gamma_b \text{ req}} - \sum_{i=1}^{K-1} \chi_i\right) \\
&= \sum_{k=0}^{K-1} \text{P}\left(I_0/C > (I/C)_{\gamma_b \text{ req}} - k \,\bigg|\, \sum_{i=1}^{K-1} \chi_i = k\right) \text{P}\left(\sum_{i=1}^{K-1} \chi_i = k\right) \\
&= \sum_{k=0}^{K-1} \binom{K-1}{k} p^k (1-p)^{K-1-k} Q\left(\frac{(I/C)_{\gamma_b \text{ req}} - k - \text{E}[I_0/C]}{\sqrt{\text{Var}[I_0/C]}}\right)
\end{aligned}
\tag{11.16}
$$

In all of our numerical results, we assume a chip rate of $R = 1.25$ Mchips/s and a source symbol rate of 8 kb/s, yielding a processing gain of $G = 156.25$. We further assume a voice activity factor of $p = 3/8$. Fig. 11.3 shows the reverse channel capacity for different $\gamma_{b \text{ req}}$ and shadow standard deviations. The reverse channel capacity is greatly increased by a reduction in $\gamma_{b \text{ req}}$ and slightly reduced when the shadow standard deviation is increased. Fig. 11.4 shows the reverse channel capacity with different Nakagami shape factors for the desired and interfering signals. Observe that a change in the Nakagami shape factor m_I of interfering signals has very little effect on the reverse channel capacity. Fig. 11.5 further illustrates the effect of fading and shadowing on the reverse channel capacity. As expected, shadowing and fading have relatively little impact on the reverse channel capacity, since these components of the received signal are power controlled. Therefore, fading and shadowing variations only affect the out-of-cell interference.
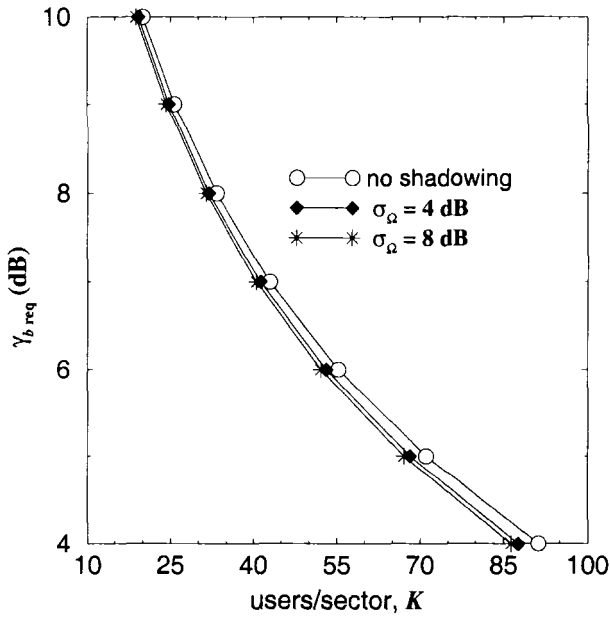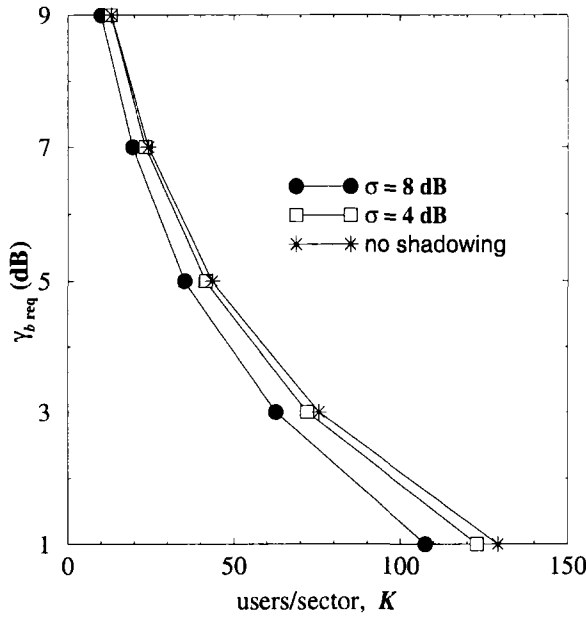
Figure 11.3. Reverse channel capacity with $P_{\text{out}} = 10^{-2}$ for different $\gamma_{b \text{ req}}$; the Nakagami shape factors are $m_d = 1$ and $m_I = 1$ (Rayleigh fading), and σ_Ω is the shadow standard deviation.

The ratio of the mean out-of-cell interference to the mean in-cell interference ratio is

$$\theta = \frac{\text{E}[I_{\text{o}}]}{\text{E}[I_{\text{in}}]} = \frac{\text{E}[I_{\text{o}}/C]}{\text{E}[I_{\text{in}}/C]} = \frac{\text{E}[I_{\text{o}}/C]}{pK} \tag{11.17}$$

With a 4th-order path loss exponent, Newson and Heath [247] showed that $\theta = 0.5$ when no fading and shadowing are considered and $\theta \approx 0.66$ when shadowing is considered with $\sigma_\Omega = 8$ dB. This translates into a frequency reuse efficiency f, defined as the ratio of mean in-cell interference to the total mean interference, of 0.66 and 0.38, respectively. Table 11.1 tabulates the corresponding values of θ and $f = 1/(1 + \theta)$ for the CDMA cellular system under consideration for different propagation conditions. The calculations only consider the first tier of interfering cells. Observe that the frequency reuse efficiency decreases with the shadow standard deviation, σ_Ω, and slightly increases when m_d increases or m_I decreases.

To show that the values of θ and f in Table 11.1 do not depend on the number of users per cell, K, we derive the cdf of the out-of-cell interference to the in-cell interference $I_{\text{o}}/I_{\text{in}}$ for the reverse channel as
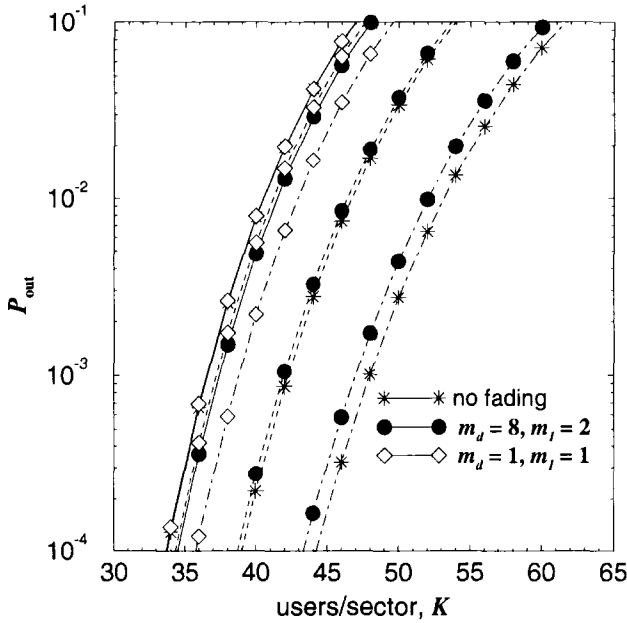
Figure 11.4. Reverse channel capacity for different propagation environments with $\gamma_{b \text{ req}} = 8.76$ dB. Solid lines denote $\sigma_\Omega = 0$ dB, dotted lines denote $\sigma_\Omega = 4$ dB, and dot-dashed lines denote $\sigma_\Omega = 8$ dB.

$$
\begin{aligned}
\mathrm{P}(I_\mathrm{o}/I_\mathrm{in} < z) &= \mathrm{P}\left(\frac{I_\mathrm{o}/C}{I_\mathrm{in}/C} < z \right) \\
&= 1.0 - \mathrm{P}(I_\mathrm{o}/C > z \, I_\mathrm{in}/C) \\
&= 1.0 - \left\{ \sum_{k=0}^{K-1} \binom{K-1}{k} \eta^k (1-\eta)^{K-1-k} \right. \\
&\quad \left. \times Q\left(\frac{kz - \mathrm{E}[I_\mathrm{o}/C]}{\sqrt{\mathrm{Var}[I_\mathrm{o}/C]}} \right) \right\}
\end{aligned}
$$

(11.18)

Fig. 11.6 plots the distribution of $I_\mathrm{o}/I_\mathrm{in}$ (in dB) with different shadow standard deviations. Although the distribution varies with K, the mean value $\mathrm{E}[I_\mathrm{o}/I_\mathrm{in}]$ remains almost the same, i.e., all the curves cross at the 50% point. This implies that the values of θ and f in Table 11.1 do not depend on K.

Extensions of the above results to include the effects of soft handoff have been provided in [343]. Soft handoff was shown to improve coverage by a factor of 2 to 2.5 in cell area, i.e., the number of BSs can be reduced by this factor. It was also shown to increase the reverse channel capacity by a factor better than 2.
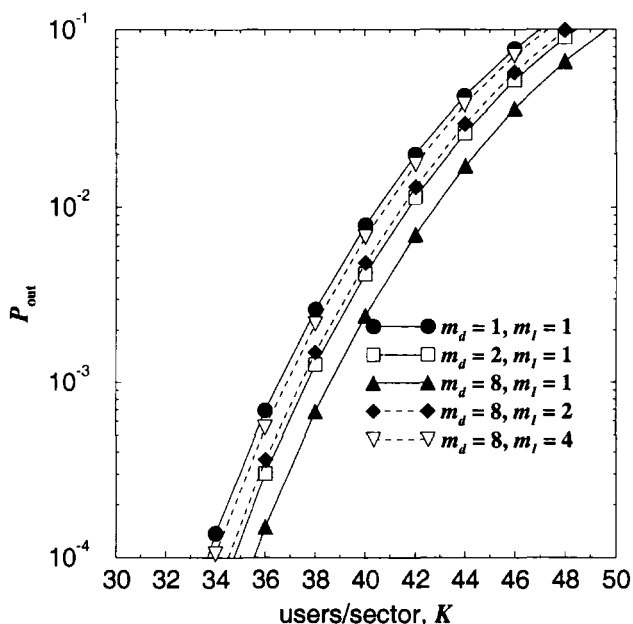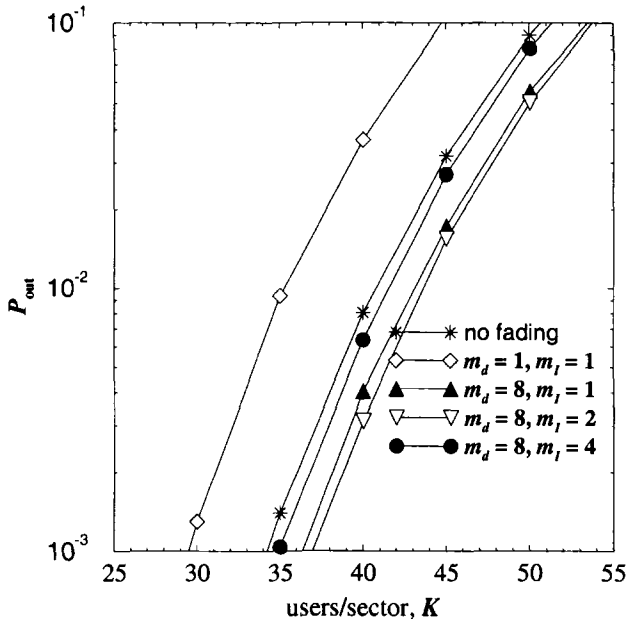
Figure 11.5. Reverse channel capacity for different Nakagami shape factors with $\gamma_{b\ req}$ = 8.76 dB, σ_Ω = 8 dB.

m_d	m_I	σ_Ω (dB)	θ	$f = 1/(1 + \theta)$
8	8	8	60.14%	62.45%
8	4	8	58.83%	62.96%
8	2	8	56.20%	64.02%
8	1	8	51.11%	66.1i8%
4	1	8	52.36%	65.63%
2	1	8	54.90%	64.56%
1	1	8	59.73%	62.61%
1	1	10	57.34%	63.56%
1	1	6	60.16%	62.44%
1	1	4	57.82%	63.36%
1	1	no shadowing	51.76%	65.89%
no fading	no fading	no shadowing	21.81%	82.10%
no fading	no fading	4	39.42%	71.73%
no fading	no fading	8	60.02%	62.49%
no fading	no fading	10	61.65%	61.86%

Table 11.1. Ratio of the mean of out-of-cell interference to the mean in-cell interference ratio, θ, and frequency reuse efficiency, f, under different propagation conditions.
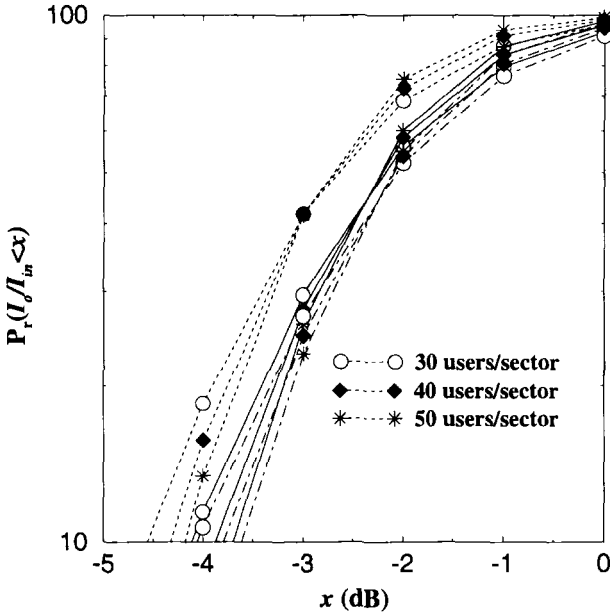
Figure 11.6. Distribution of out-of-cell interference to in-cell interference, I_o/I_{in}, for the reverse channel. Solid lines denote $\sigma_\Omega = 0$ dB, dotted lines denote $\sigma_\Omega = 4$ dB, and dot-dashed lines denote $\sigma_\Omega = 8$ dB; $m_d = 1$, $m_I = 1$ (Rayleigh fading).

1.2 FORWARD LINK CAPACITY

For the forward channel, a pilot signal is transmitted from each BS. The pilot signal is a spread spectrum signal that causes interference in every cell, thereby reducing the capacity. However, this is offset by a decrease in $\gamma_{b\ req}$ due to coherent modulation. With forward channel balancing power control, the mobile measures the received signal and periodically transmits the measurement to its serving BS [136]. When the total power requested by mobiles is below the maximum allowable transmit power, the BS will reduce its transmit power, thereby reducing interference; otherwise, the BS will redistribute the power from the forward links with good quality to those with poor quality.

In the worst case situation, each BS always transmits with the maximum allowable power P_{max}. From (11.2), the γ_b at the ith mobile under this condition is

$$\gamma_{b,i} = \frac{3G}{2(I/C)_i} = \frac{3G}{2(\sum_{j=0}^{M} C_{T_j} - \delta\phi_i C_{T_0})/\delta\phi_i C_{T_0}} \tag{11.19}$$

$$C_{T_j} = P_{max}10^{\frac{\varsigma_j}{10}} \tag{11.20}$$

where M is the number of surrounding BSs that are included in the calculation, C_{T_j} is the received power from BS_j, $1 - \delta$ is the fraction of the total power allocated to the pilot, and the weighting factor ϕ_i is the fraction of the remaining power allocated to the ith mobile. Note that (11.19) assumes the worst case condition of chip and phase asynchronous random spreading sequences. As in [136] our results assume that 20% of the total BS transmit power is allocated to the pilot signal. Once again, the ς_i are Gaussian random variables due to shadow and fading variations, with means and variances obtained from (2.188).

The BS distributes its transmit power proportionally according to the needs of each mobile within its cell. This is accomplished by first obtaining the required ϕ_i for each mobile, $(\phi_i)_{\text{req}}$, by setting $\gamma_b = \gamma_{b\text{ req}}$ in (11.19). To account for the voice activity, we then calculate the modified weighting factor [183]

$$\overline{\phi_i} = \frac{(\phi_i)_{\text{req}}}{\sum_{\substack{j=1 \\ j \neq i}}^{K} \chi_j(\phi_j)_{\text{req}} + (\phi_i)_{\text{req}}} \ . \tag{11.21}$$

The power balancing scheme in [46] does the same thing, except that the voice activity factors, χ_j, are not considered. The outage probability then becomes

$$\begin{aligned} P_{\text{out}} &= \text{P}\left(\gamma_b < \gamma_{b\text{ req}}\right) \\ &= \text{P}\left(\overline{\phi_i} < (\phi_i)_{\text{req}}\right) \ . \end{aligned} \tag{11.22}$$

Numerical results can be obtained from the last equation in (11.22) by using Monte Carlo simulation techniques to account for the random user locations, and shadow and fading variations. For each set of MS locations and propagation conditions, we first determine the required fraction of power, $(\phi_i)_{\text{req}}$, needed to meet the $\gamma_{b\text{ req}}$ requirement. Afterwards, we find if the actual power allocation for each MS, $\overline{\phi_i}$, is sufficient.

Fig. 11.7 shows how the forward channel capacity depends on $\gamma_{b\text{ req}}$ and the shadow standard deviation. Shadowing has a slightly stronger effect on forward channel capacity compared to the reverse channel. Fig. 11.8 shows the forward channel capacity for various Nakagami shape factors. The Nakagami shape factor also plays a significant role in forward channel capacity, and overly optimistic capacity estimates will be obtained if fading is neglected.

1.3   IMPERFECT POWER CONTROL

Any power control algorithm will inevitably be subject to some degree of error. It has been experimentally verified that the power control error (in dB) can be modeled as a zero-mean Gaussian random variable with variance σ_E^2 [183], [247]. For the reverse channel, (11.2) has the modified form

$$C10^{\varsigma_{E_j}/10} = P_{ij}10^{\varsigma_d/10} \tag{11.23}$$

Figure 11.7. Forward channel capacity with $P_{\text{out}} = 10^{-2}$ for different $\gamma_{b\,\text{req}}$; the Nakagami shape factors are $m_d = 1$ and $m_I = 1$ (Rayleigh fading), and σ_Ω is the shadow standard deviation.

where ς_{E_j} is the power control error. The mean and variance of $x = \varsigma_{i0} - \varsigma_{ij}$ with imperfect power control are similar to (11.10), but have the form

$$\mu_x \text{ (dB)} = \mu_{i0}\text{ (dBm)} - \mu_{ij}\text{ (dBm)} \qquad \sigma_x^2 = \sigma_{i0}^2 + \sigma_{ij}^2 + \sigma_E^2 \ . \qquad (11.24)$$

With imperfect power control, the in-cell interference experienced by MS_{0i} at the reference BS, BS_0, is $I_{\text{in}} = C\kappa$, where

$$\kappa = \sum_{\substack{j=1 \\ j\neq i}}^{K} 10^{\varsigma_{E_j}/10} \chi_j \ . \qquad (11.25)$$

Then

$$
\begin{aligned}
P_{\text{out}} &= \text{P}(\gamma_b < \gamma_{b\,\text{req}}) \\
&= \text{P}(I_o/C + I_{\text{in}}/C > (I/C)_{\gamma_b\,\text{req}}) \\
&= \text{P}(I_o/C > (I/C)_{\gamma_b\,\text{req}} - \kappa) \\
&= \sum_{k=0}^{K-1} \binom{K-1}{k} p^k (1-p)^{K-1-k} \\
&\quad \times \int_0^\infty \text{P}\left(I_o/C > (I_o/C)_{\gamma_b\,\text{req}} - \kappa | \kappa\right) p(\kappa|k)d\kappa \ . \quad (11.26)
\end{aligned}
$$

Figure 11.8. Forward channel capacity for different Nakagami shape factors with $\gamma_{b\ req} = 6.76$ dB, $\sigma_\Omega = 8$ dB.

Note that the conditional pdf of k given $k, p(\kappa|k)$, is approximately log-normal. The log-normal approximation can be calculated using the Fenton-Wilkonson and Schwartz-Yeh methods discussed in Chapter 3. Observe from Fig. 11.9 that the reverse channel capacity is dramatically decreased as the power control error increases. For $P_{out} = 0.01$ and power control errors of $\sigma_E = 1$dB, 2dB, and 3 dB, the reverse channel capacity is decreased by 24%, 50%, and 68%, respectively.

To consider the effect of power control error on the forward channel, (11.21) becomes

$$\overline{\phi_i} = \frac{\phi_i 10^{\varsigma_{E_i}/10}}{\sum_{\substack{j=1 \\ j\neq i}}^{K} \chi_j \phi_j 10^{\varsigma_{E_j}/10} + \phi_i 10^{\varsigma_{E_i}/10}} \tag{11.27}$$

Fig. 11.9 shows that the forward channel capacity is reduced by 31%, 64%, and 83% for $\sigma_E = 1.0$ dB, 2.0 dB, and 3.0 dB, respectively. Note that imperfect power control has a more severe effect on the forward channel than the reverse channel for the same propagation conditions.

2. ERROR PROBABILITY WITH RAKE RECEPTION

Consider a CDMA cellular system that uses 120° cell sectoring. The reference MS, MS_j is located in Cell Sector 0. Assume chip and phase asynchronous
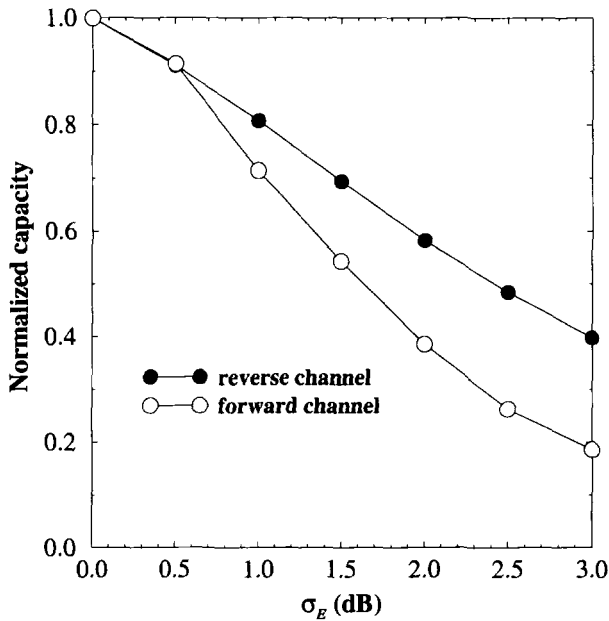
Figure 11.9. Forward and reverse channel capacity with imperfect power control. The capacity is normalized with respect to the capacity with perfect power control; σ_E is the standard deviation of the power control error, $m_d = 1$, $m_I = 1$, and $\sigma_\Omega = 8$ dB.

signaling and a large number of simultaneous users with random spreading codes. Under such conditions, the standard Gaussian approximation in Chapter 9.6 applies and the multiple access interference, I, in (9.174) is Gaussian distributed with variance $\sigma_I^2 = (K-1)/3N$, where K is the number of active transmissions and N is the processing gain (assuming a short spreading code). The total impairment in (9.160) due to multiple access interference and Gaussian noise is $\tilde{I} = \sqrt{2E}I + \tilde{n}_I$ and has variance $\sigma_{\tilde{I}}^2 = (K-1)2E/3N + N_o$, where E is the bit energy. Hence, the error probability is

$$P_b = \left(\sqrt{\frac{6NE}{2E(K-1) + 3NN_o}} \right) \tag{11.28}$$

Assuming an error proability expression of the form $P_b = Q(\sqrt{2\gamma_b})$, the equivalent bit energy-to-noise ratio is

$$\gamma_b = \frac{3NE}{2E(K-1) + 3NN_o} \tag{11.29}$$

However, in a multipath-fading environment the above expression must be modified. The instantaneous equivalent bit energy-to-noise ratio that is associated

with $\text{MS}_{\hat{j}}$ and the path ℓ path is

$$R_{\hat{j},\ell} = \frac{3N\lambda_\ell^{\hat{j}(0)}}{2\sum_{c,i,n\in U}\lambda_n^{i(c)} + 3NN_o} \tag{11.30}$$

where $\lambda_n^{i(c)} = |g_n^{i(c)}|^2 E$ and $g_n^{i(c)}$ is the nth channel tap gain that is associated with MS_i in Cell Sector c. Very often the thermal noise N_o can be neglected in deference to the typically dominant effect of the multiple-access interference. The set U in (11.30) is defined as

$$U \triangleq \bigcup_{c\in A} U_c \tag{11.31}$$

where

$$U_c = \{(c,i,n) : \chi_{ci} = 1,\ 1 \le i \le K,\ 1 \le n \le L,\ \text{and}\ (c,i,n) \ne (0,\hat{j},n)\} \tag{11.32}$$

and

$$\chi_{ci} = \begin{cases} 1 & \text{if transmitter } i \text{ in Cell Sector } c \text{ is active} \\ 0 & \text{otherwise} \end{cases} \tag{11.33}$$

In (11.32) the assumption is made that the self-interference due to multipath can be neglected. Note that the multipath increases the level of multiple access by increasing the size of the set U_c in (11.32).

The mean interference power that is received at the serving BS from all out-of-cell MSs is a constant. If there are a total of K_S MSs per cell, the with the use of cell sectoring and voice activity gating, $K_S p/3$ MSs will be actively transmitting in each cell sector on average. Of course the number of active MSs is binomially distributed, but our simplified analysis will use the average number of active MSs. With power control, the ratio of the *average* received power received over path ℓ to the *average* received total noise power is approximately

$$\bar{\gamma}_\ell \approx \frac{3N\Gamma_\ell^{\hat{j}(0)}}{2\,(K_S p/3 - 1)\sum_{n=1}^{L}\Gamma_n^{i(0)} + 2\sum_{c\ne 0,i,n\in U}\Gamma_n^{i(c)}}. \tag{11.34}$$

The second term in the denominator of (11.34) sums the average out-of-cell multiple-access interference. It must be determined by careful study of the particular CDMA deployment. For uniform hexagonal cells, we have seen earlier (see Table 11.1) that this term is approximately 50% of the first term in the denominator. However, for microcells this term can be as large as 100% [166], depending on the cell layout, user spatial distribution and propagation

path loss exponent. Assuming a frequency reuse efficiency f, (11.34) becomes

$$\bar{\gamma}_\ell \approx \frac{3fN\Gamma_\ell^{\hat{j}(0)}}{2\left(K_Sp/3 - 1\right)\sum_{n=1}^{L}\Gamma_n^{i(c)}} \quad . \tag{11.35}$$

The values of $\Gamma_\ell^{\hat{j}(0)}$ and $\Gamma_n^{i(c)}$ depend, among other things, on the delay spread of the channel. If the delay spread exhibits an exponential decay, then it follows from (9.156) that

$$\Gamma_n^{i(0)} = \frac{(1 - e^{-1/\varepsilon})e^{-n/\varepsilon}}{e^{-1/\varepsilon} - e^{-(L+1)/\varepsilon}} \cdot \Gamma_t \tag{11.36}$$

where

$$\Gamma_t = \sum_{n=1}^{L}\Gamma_n^{i(c)} \tag{11.37}$$

is the total power received from each MS.

Since the received signals from the interfering users are Rayleigh faded, they can be treated as zero-mean complex Gaussian random processes. Furthermore, the multiple-access interference consists of a large number of uncorrelated interferers so that it can be treated as a stationary Gaussian random process. However, we must still account for the fading fluctuations in the envelope of the desired signal.

The error probability depends on the type of diversity combining and detection being used. We assume that an M-tap RAKE receiver with maximal ratio combining. In general, $M \neq L$, where L is the number of taps in the tapped delay line channel model. The receivers are assumed to use D-branch spatial diversity so there are DM replicas of the desired signal that are available for processing. The instantaneous received bit energy-to-noise ratio that is associated with path ℓ and antenna m, $\gamma_{m,\ell}$, is exponentially distributed with density

$$p_\gamma(x) = \frac{1}{\bar{\gamma}_\ell}e^{-x/\bar{\gamma}_\ell} \tag{11.38}$$

where we have assumed that identical antenna elements so that $\bar{\gamma}_{m,\ell} = \bar{\gamma}_\ell$ with $\bar{\gamma}_\ell$ given by (11.35).

2.1 MAXIMAL RATIO COMBINING

This section presents a simplified performance analysis of cellular CDMA. Although there are many different performance aspects, we focus here on bit error probability of cellular CDMA with RAKE reception. While the multipath allows us to gain a diversity advantage by employing a RAKE receiver, it also has the undesirable effect of accentuating the effect of the multiple-access interference. Hence, we are interested in assessing the diversity gains that we can expect.

Following the treatment of maximal ratio combining in Chapter 6.3, the instantaneous *processed* bit energy-to-total noise ratio is defined as

$$\gamma_{\text{p}} = \sum_{m=1}^{D} \sum_{k=1}^{M} \gamma_{m,k} \ . \tag{11.39}$$

With coherent BPSK or QPSK signaling, the bit error probability conditioned on γ_{p} is

$$P_b(\gamma_{\text{p}}) = Q(\sqrt{2\gamma_{\text{p}}}) \ . \tag{11.40}$$

Since the $\gamma_{m,k}$ are statistically identical with respect to the index m, it follows that the characteristic function of γ_{p} is

$$\begin{aligned} \psi_{\gamma_{\text{p}}}(s) &= \prod_{k=1}^{M} \frac{1}{(1 - s\bar{\gamma}_k)^D} \\ &= \sum_{k=1}^{M} \sum_{j=1}^{D} \frac{A_{kj}}{(1 - s\bar{\gamma}_k)^j} \end{aligned} \tag{11.41}$$

where

$$A_{kj} = \frac{1}{(D-j)!(-\bar{\gamma}_k)^{D-j}} \left\{ \frac{\mathrm{d}^{D-j}}{\mathrm{d}s^{D-j}} (1 - s\bar{\gamma}_k)^D \psi_{\gamma_{\text{p}}}(s) \right\} \Bigg|_{s=1/\bar{\gamma}_k} \ . \tag{11.42}$$

It follows that the density of γ_{p} is

$$p_{\gamma_{\text{p}}}(x) = \sum_{k=1}^{M} \sum_{j=1}^{D} A_{kj} \frac{1}{(j-1)!(\bar{\gamma}_k)^j} x^{j-1} e^{-x/\bar{\gamma}_k} \ . \tag{11.43}$$

Therefore, the average bit error probability becomes

$$\begin{aligned} P_b &= \int_0^{\infty} P_b(x) p_{\gamma_{\text{p}}}(x) \, dx \\ &= \sum_{k=1}^{M} \sum_{j=1}^{D} A_{kj} \int_0^{\infty} Q(\sqrt{2x}) \frac{1}{(j-1)!(\bar{\gamma}_k)^j} x^{j-1} e^{-x/\bar{\gamma}_k} \, dx \\ &= \sum_{k=1}^{M} \sum_{j=1}^{D} A_{kj} \left(\frac{1 - \mu_k}{2}\right)^j \sum_{n=0}^{j-1} \binom{j-1+n}{n} \left(\frac{1 + \mu_k}{2}\right)^n \end{aligned} \tag{11.44}$$

where

$$\mu_k = \sqrt{\frac{\bar{\gamma}_k}{1 + \bar{\gamma}_k}} \ . \tag{11.45}$$
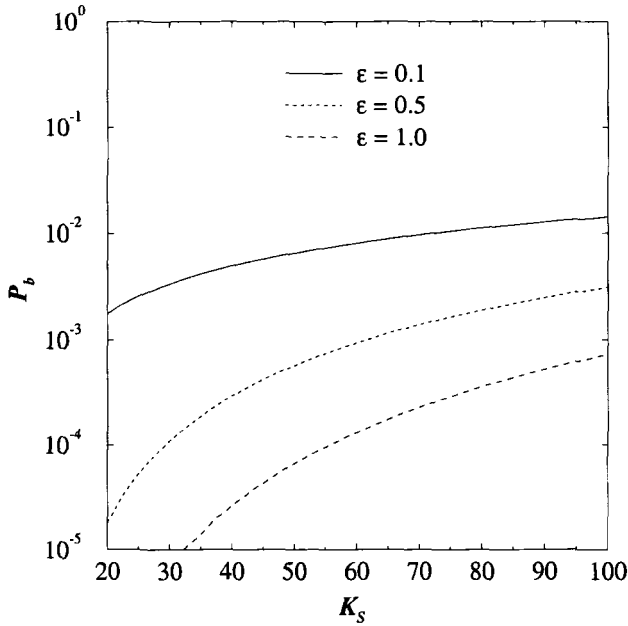
Figure 11.10. Bit error probability against the number of MS per cell, K_S. A 4-tap RAKE receiver is used without antenna diversity on a 4-tap channel. The voice activity factor is $p = 0.5$ and the frequency reuse efficiency is $f = 2/3$. Values of $\varepsilon = 1.0, 0.5$ and 0.1 are shown.

It is useful to express the performance as a function of the total average received bit energy-to-noise ratio per antenna branch, $\bar{\gamma}_c$, defined as

$$\bar{\gamma}_c \triangleq \sum_{k=1}^{L} \bar{\gamma}_k$$

$$= \frac{3Nf}{2(K_S p/3 - 1)} \tag{11.46}$$

where the last step is obtained by using (11.35)–(11.37). Note that $\bar{\gamma}_p/D \le \bar{\gamma}_c$. Finally,

$$\bar{\gamma}_k = \frac{3Nf}{2(K_S p/3 - 1)} \cdot \frac{(1 - e^{-1/\varepsilon})e^{-k/\varepsilon}}{e^{-1/\varepsilon} - e^{-(L+1)/\varepsilon}} . \tag{11.47}$$

Fig. 11.10 shows the CDMA reverse channel performance with $D = 1$, $L = 4$, $M = 4$, $p = 0.5$, $f = 2/3$ and various ε. For cellular CDMA systems an error probability on the order of 10^{-2} to 10^{-3} is deemed acceptable. Note that the error probability increases as the channel becomes less dispersive (smaller ε) because the RAKE receiver cannot gain a diversity advantage. In order to prevent poor performance in a nondispersive channel environment, antenna diversity can be used. Fig. 11.11 shows the performance under conditions
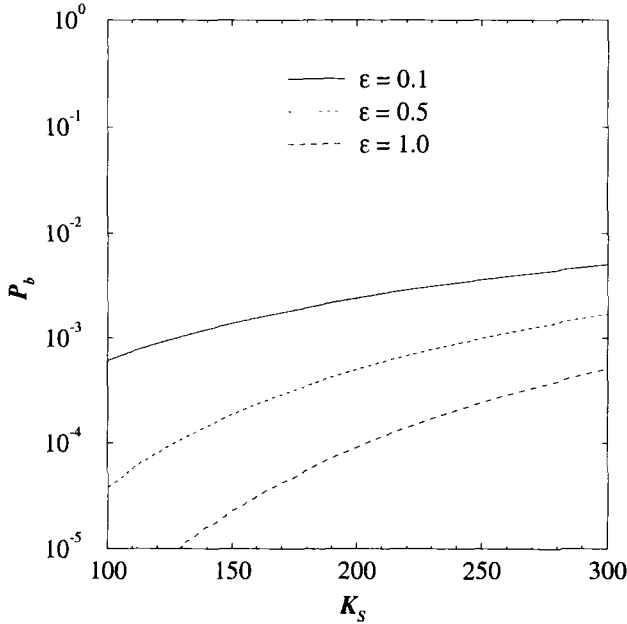
Figure 11.11. Bit error probability against the number of MS per cell $K_S = 3K$. A 4-tap RAKE receiver is used with 2-branch antenna diversity on a 4-tap channel. The voice activity factor is $p = 0.5$ and the frequency reuse efficiency is $f = 2/3$. Values of $\varepsilon = 1.0, 0.5$ and 0.1 are shown.

identical to those in Fig. 11.10 except that 2-branch antenna diversity is used, with independently faded branches. Note the scale change on the abssisa in Figs. 11.10 and 11.ll.

Problems

11.1. Consider a CDMA cellular system where there are 5 in-cell interferers. Each interferer is independently active with probability p, and is characterized by a power control error ς_{E_j}. The power control errors, ς_{E_j} (in dB) are independent zero-mean Gaussian random variables with variance σ_E^2. Hence, the in-cell interference is

$$I_{\text{in}} = C \sum_{j=1}^{5} 10^{\varsigma_{E_j}/10} \chi_j$$

where

$$\chi_j = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p \end{cases}$$

a) The in-cell interference can be approximated as a log-normal random variable conditioned on the number of active interferers

$$k = \sum_{j=1}^{5} \chi_j$$

Determine the mean and variance of the log-normal approximation as a function of k for $\sigma_E = 1$, 2, and 3 dB. Use the Fenton-Wilkinson approach.

b) Assuming that the value of $(I_{in})_{dB}$ is Gaussian when conditioned on the number of active interferers, write down and expression for the pdf of $(I_{in})_{dB}$.

11.2. (computer exercise) The purpose of this problem is to determine the relative contribution of the first-, second-, and third-tier cells to the out-of-cell interference in a CDMA cellular system. Also, we wish to determine the impact of the path loss exponent on the out-of-cell interference.

Consider a CDMA cellular system characterized by log-normal shadowing with a shadow standard deviation σ_Ω dB and inverse β power path loss. Neglect envelope fading. All other factors such as base-station antenna heights, cell sizes, etc. are uniform. The ratio of the mean out-of-cell interference to mean in-cell interference is

$$\theta = \frac{E[I_o/C]}{pK}$$

where

$$E[I_o/C] = p \int \int \exp\left\{\frac{\xi^2\sigma_x^2}{2} + \xi\mu_x\right\} \left[1 - Q\left(-\xi\sigma_x - \frac{\mu_x}{\sigma_x}\right)\right] \rho \, dA$$

and

$$\mu_x = \mu_{i0} - \mu_{ij}$$
$$\sigma_x^2 = 2\sigma_\Omega^2$$

For regular hexagonal cells, the subscriber density is

$$\rho = \frac{2K}{3\sqrt{3}R^2}$$

Hence,

$$\theta = \frac{2}{3\sqrt{3}R^2} \int \int \exp\left\{\frac{\xi^2\sigma_x^2}{2} + \xi\mu_x\right\} \left[1 - Q\left(-\xi\sigma_x - \frac{\mu_x}{\sigma_x}\right)\right] \rho \, dA$$

where the double integral is over the two-dimensional out-of-cell area.

a) Consider the first tier of interfering cells. For the case of $\sigma_\Omega = 8$ dB, calculate θ when $\beta = 3$ and $\beta = 4$.

b) Repeat part a), but this time consider only the *second* tier of interfering cells.

c) Finally, repeat part a), for the *third* tier of interfering cells.

d) What conclusions can you draw?

11.3. CDMA systems use soft handoff, where the transmissions to/from multiple base stations are combined to give a macro-diversity advantage. Suppose that the receive bit energy-to-noise ratio for branch i, denoted by γ_i, has the probability density

$$p_{\gamma_i}(x) = \frac{1}{\bar{\gamma}_i} e^{-x/\bar{\gamma}_i}$$

where $\bar{\gamma}_i = E[\gamma_i]$.

a) The reverse link uses selection macro-diversity such that

$$\gamma_b^s = \max\{\gamma_1, \gamma_2, \ldots, \gamma_L\}$$

An outage occurs if $\gamma_b^s < \gamma_{th}$. What is the probability of outage?

b) The forward link uses maximal ratio combining such that

$$\gamma_b^{mr} = \gamma_1 + \gamma_2 + \cdots + \gamma_L$$

Again, an outage occurs if $\gamma_b^{mr} < \gamma_{th}$. What is the probability of outage if $\bar{\gamma}_1 = \bar{\gamma}_2 = \cdots = \bar{\gamma}_L$?

c) For $L = 2$ and an outage probability of 10^{-4} what is the *difference* in the required γ_{th} (in units of decibels) with selection and maximal ratio combining, again assuming that $\bar{\gamma}_1 = \bar{\gamma}_2$?

Chapter 12

LINK QUALITY MEASUREMENT
AND HANDOFF INITIATION

When a new call arrives, mobile station (MS) must be connected to a suitable base station(s) (BSs). Also, when a MS traverses a cell boundary an **intercell handoff** is required so that an acceptable link quality can be maintained without causing unnecessary co-channel and adjacent channel interference. Failure to handoff a MS at an established cell boundary also tends to increase blocking, because some cells will carry more traffic than planned. Sometimes an **intracell handoff** is desirable when the link with the serving BS is affected by excessive interference, while another link with the same BS can provide better quality. The handoff process consists of two stages: i) link quality evaluation and handoff initiation, ii) allocation of radio and network resources.

In general, cellular systems with smaller cell sizes require faster and more reliable link quality evaluation and handoff algorithms. Labedz [185] has shown that the number of cell boundary crossings is inversely proportional to the cell size. Furthermore, Nanda [245] has shown that the handoff rate increases with only the square-root of the call density in macrocells, but linearly with the call density in microcells. Since the MS has a certain probability of handoff failure each time a handoff is attempted, it is clear that handoff algorithms must become more robust and reliable as the cell sizes decrease.

One of the major tasks in a cellular system is to monitor the link quality and determine when handoff is required. If a handoff algorithm does not detect poor signal quality fast enough, or makes too many handoffs, then capacity is diminished due to excessive co-channel interference and/or an undue switching load. A variety of parameters such as bit error rate (BER) [66], carrier-to-interference ratio (C/I) [121], distance [222], [102], traffic load, signal strength [222], [149], [150], [238], [52], [340], and various combinations of these fundamental schemes have been suggested for evaluating the link quality and deciding when a handoff should be performed. Of these, temporal

averaging signal strength based handoff algorithms that measure the received carrier plus interference power (C+I) have received the most attention due to their simplicity and good performance in macrocellular systems. However, spectrally efficient cellular systems are interference limited and a large C+I does not necessary imply a large C/I. Since radio link quality is more closely related to C/I than to C+I, it is apparent that C/I based handoff algorithms are highly desirable for microcellular systems with their characteristically erratic propagation environments. A discussion of C/I measurement techniques is included in this chapter.

Based on the roles that the BSs and MSs perform in the process of link quality evaluation and handoff initiation, there are three categories of handoff algorithms. The first is a **network-controlled handoff** (NCHO) algorithm which has been widely used in first generation analog cellular systems, such as AMPS. With a NCHO algorithm, the link quality is only monitored by the serving BS and the surrounding BSs. The handoff decision is made under the centralized control of a mobile telephone switching office (MTSO). Typically, NCHO algorithms only support only intercell handoffs, have handoff network delays of several seconds, and have relatively infrequent updates of the link quality estimates from the alternate BSs.

The second type of handoff algorithm is the **mobile-assisted handoff** (MAHO) algorithm which is widely used in many second generation digital cellular systems, such as IS-54 and GSM. MAHO algorithms use both the serving BS and the MS to measure link quality of the serving BS; however, link quality measurements of the alternate BSs are only obtained by the MS. The MS periodically relays the link quality measurements back to the serving BS, and the handoff decision is still made by the serving BS along with the MTSO. MAHO algorithms typically support both intracell and intercell handoffs, have network delays on the order of one to two seconds, an use relatively frequent updates of the link quality measurements.

The third type of handoff algorithm is a **mobile-controlled handoff** (MCHO) algorithm, a decentralized strategy that is used in some of the more recent digital cordless telephone systems, such as DECT. With MCHO algorithms the link quality with the serving BS is measured by both the serving BS and the MS. Like a MAHO algorithm, the measurements of link quality for alternate BSs are done at the MS, and both intracell and intercell handoffs are supported. However, unlike the MAHO algorithms, the link measurements at the serving BS are relayed to the MS, and the handoff decision is made by the MS. MCHO algorithms typically have the lowest handoff network delays (usually about 100 ms) and are the most reliable.

Once the handoff process is initiated, handoff algorithms can also differ in the way that a call transferred to a new link. Handoff algorithms can be categorized into forward and backward types. **Backward handoff** algorithms initiate the

handoff process through the serving BS, and no access to the "new" channel is made until the control entity of the new channel has confirmed the allocation of resources. The advantage of backward algorithms is that the signaling information is transmitted through an existing radio link and, therefore, the establishment of a new signaling channel is not required during the initial stages of the handoff process. The disadvantage is that the algorithm may fail in conditions where the link quality with the serving BS is rapidly deteriorating. This type of handoff is used in most of the TDMA cellular systems such as GSM. **Forward handoff** algorithms initiate the handoff process via a channel the target BS without relying on the "old" channel during the initial phase of the handoff process. The advantage is a faster handoff process, but the disadvantage is a reduction in handoff reliability. This type of handoff is used in digital cordless telephone systems such as DECT. Handoff can also be distinguished according to **hard handoffs** and **soft handoffs.**

Hard handoffs:. With hard handoffs, a MS can connect to only one BS at a time. An absolute (binary) decision is made to initiate and execute a handoff without making a number of simultaneous connections among candidate BSs. The handoff is initiated based on a hysteresis imposed on the current link. The target BS is already selected prior to executing the handoff based on link measurements and the active connection is transferred to the target BS instantly. The connection experiences a brief interruption during the actual transfer because MS can only connect to one BS at a time. Hard handoff does not take advantage of diversity gain opportunity during handoff where the signals from two or more BSs arrive at comparable strengths. Yet, it is a simple and inexpensive way to implement handoff. This type of handoff is used in most TDMA cellular systems such as IS-54, PDC, and GSM.

Signal strength based hard handoff algorithms have been optimized by minimizing two conflicting design criteria; the handoff delay and the mean number of handoffs between BSs. It is important to keep the handoff delay small to prevent dropped calls and to prevent an increase in co-channel interference due to distortion of the cell boundaries. Likewise, it is important to keep the mean number of handoffs between BSs along a handoff route at a reasonably low value to prevent excessive loading and resource consumption on the network. Several authors [238], [340], [222], [150] have applied these (or similar) design criteria while adjusting two important design parameters; the required average signal strength difference, or **hysteresis** H, between the BSs before a hard handoff is initiated, and the temporal window length T over which the signal strength measurements are averaged. The handoff hysteresis prevents excessive handoffs due to "ping-ponging" between BSs. The best choice of T and H depends on the propagation environment. Usually, the averaging interval T is
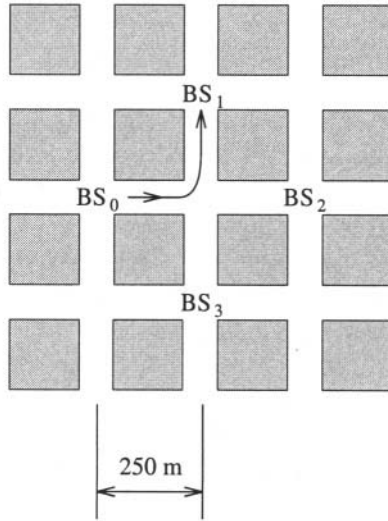
Figure 12.1. Typical NLoS handoff scenario. The MS rounds the corner, losing the LoS from BS_0 and gaining the LoS from BS_1. The frequency reuse plan is specified by using the different numbers at the BSs, from [18].

chosen to correspond to 20 to 40 wavelengths, and the hysteresis H is chosen on the order of the shadow standard deviation.

Murase [238] studied the tradeoff between the hysteresis and window length for line-of-sight (LoS) and non line-of-sight (NLoS) hard handoffs. For LoS handoffs, the MS always maintains a LoS with both the serving and target BS. This would be the case, for example, when a MS traverses along a route from BS_0 to BS_2 in Fig. 12.1. NLoS handoffs, on the other hand, arise when the MS suddenly loses the LoS component with the serving BS while gaining a LoS component with the target BS. This phenomenon is called the **corner effect** [238], [52] since it occurs while turning corners in urban microcellular settings like the one shown in Fig. 12.1 where the MS traverses along a route from BS_0 to BS_1. In this case, the average received signal strength with the serving BS can drop by 25-30 dB over distance as small as 10 m [238].

Corner effects may also cause link quality imbalances on the forward and reverse channels due to the following mechanism. Quite often the co-channel interference will arrive via a NLoS propagation path. Hence, as a MS rounds a corner, the received signal strength at the serving BS suffers a large decrease while the NLoS co-channel interference remains the same, i.e., the corner effect severely degrades the C/I on the reverse channel. Meanwhile, the corner will cause the same attenuation to both the desired and interfering signals that are received at the MS. Therefore, unless there are other sources of co-channel

interference that become predominant as the MS rounds the corner, the C/I on the forward channel will remain about the same.

If the handoff requests from rapidly moving MSs in microcellular networks are not processed quickly, then excessive dropped calls will occur. Fast temporal based hard handoff algorithms can partially solve this problem, where short temporal averaging windows are used to detect large, sudden, drops in signal strength [238]. However, the shortness of a temporal window is relative to the MS velocity and, furthermore, a fixed time averaging interval makes the hard handoff performance sensitive to velocity with the best performance being achieved at only a particular velocity. Velocity adaptive handoff algorithms can overcome these problem, and are known to be robust to the severe propagation environments that are typical of urban microcellular networks [18].

Soft handoffs:. With soft handoffs, a MS can connect to a number of candidate BSs during a handoff process. Eventually, the handoff is completed when the MS selects the best candidate BS as the target. Soft handoff is more careful in selecting the target BS, because the target BS needs to be the best candidate (provide the strongest signal) from among the available BSs. During the handoff process, soft handoff further enhances the system performance through diversity reception. Unlike hard handoff, the necessary link quality measurements for handoff are done by the MS, where it constantly monitors the pilot signals from surrounding BSs. Soft handoff is a type of Mobile Assisted Handoff. However, all these advantages do not imply that soft handoff is without its weaknesses. Soft handoff is complex and expensive to implement. Also, forward interference actually increases with soft handoff since several BSs, instead of one, can connect to the same MS. This increase in forward interference can become a problem if the handoff region is large, such that there are many MSs in soft handoff mode.

Soft handoff has a special importance in CDMA based systems, due to its close relationship to **power control**. CDMA systems are interference-limited meaning their capacities are closely related to the amount of interference the systems can tolerate. Due to its effective frequency reuse factor of one, a CDMA system cell is affected by, not only interference within its own cell, but also interference from its neighboring cells also. To alleviate level of interference, and thus increase the capacity and quality, CDMA systems employ power control. Power control attempts to solve the near/far problem by adjusting transmit power so that the target C/I is evenly satisfied. The fundamental idea behind power control is to restrain from transmitting more power than necessary in order to limit interference. With power control, each MS and BS is disciplined to transmit just enough power to meet the target C/I level. But in order for power control to work effectively, the system must ensure that each MS is connected to the BS with the strongest signal at all time, otherwise a
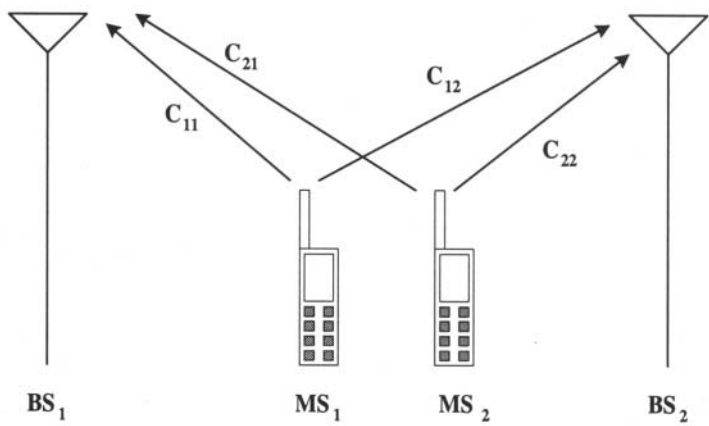
Figure 12.2. Soft handoff example, from [273].

positive power feedback problem can instablize the entire system. For example, consider a simple system consisting of two BSs and two MSs as shown in Fig. 12.2. Let us assume that each MS must satisfy a target C/I = 1. Let C_{11}, C_{12}, C_{21} and C_{22} be equal to 5, 6, 7 and 4, respectively. With soft handoff, each MS connects to the best available BS; MS_1 connects to BS_2 and MS_2 connects to BS_1. Then $C/I_1 = C_{12}/C_{22} = 6/4$ and $C/I_2 = C_{21}/C_{11} = 7/5$, and C/I_1 and C/I_2 both satisfy the target C/I. However without soft handoff, the system can no longer guarantee that the MSs are connected to the best BSs. Assume that MS_1 is connected to BS_1 and MS_2 is connected to BS_2. Then, $C/I_1 = C_{11}/C_{21} = 5/7$ and $C/I_2 = C_{22}/C_{12} = 4/6$. Since both C/I_1 and C/I_2 fail to meet the target C/I, power control will attempt to increase the C/Is by increasing the MS transmit powers. But for the given setting, increasing the MS transmit powers also increases the respective interference levels and C/Is continue to stay below the target C/I, causing a positive power feedback effect. Soft handoff does indeed ensure that each MS is served by the best BS a majority of the time. For this reason it is a *required* feature in power controlled systems like CDMA.

Although the best handoff algorithm is the one that maximizes the capacity of the network, there are many criterion to judge the performance of a handoff algorithm. These include the probability of handoff initiation, probability of dropped call, the mean number of handoff requests as a MS traverses over a handoff route, and the delay before a handoff is initiated after a MS crosses an established cell boundary. These quantities depend on the measure of link quality and the propagation environment. Finally, network parameters such as the probabilities of new call blocking, the probability of forced termination, and handoff queuing time are important. Note that we may also wish to distinguish

between dropped calls that are due to a failed handoff mechanism, and forced terminations that are due to the lack of an unavailable channel in the target cell after successful initiation of the handoff process.

The remainder of this chapter is organized as follows. Section 1. presents several different types of signal strength based handoff algorithms. This is followed by a detailed treatment of spatial signal strength averaging in Section 3.. Guidelines are developed on the window averaging length that is needed so that Ricean fading can be neglected in analog and sampled averaging. These guidelines are necessary for local mean and velocity estimation. Section 4. motivates the need for velocity adaptive handoff algorithms and presents three velocity estimators. The velocity estimators are compared in terms of their sensitivity to Rice factor, directivity, and additive Gaussian noise. In Section 5., the velocity estimators are incorporated into a velocity adaptive handoff algorithm. Section 6. provides an analytical treatment of conventional signal strength based hard handoff algorithms while Section 7. does the same for soft handoffs. In Section 8., methods are discussed for C/I measurements in TDMA cellular systems. Finally 9. wraps up with some concluding remarks.

1. SIGNAL STRENGTH BASED HARD HANDOFF ALGORITHMS

Traditional mobile assisted handoff algorithms use signal strength estimates that are obtained by calculating time averages of the received squared envelope, $< |\tilde{r}_i(t)|^2 >$, from N neighboring BSs, BS_i, $i = 0, \ldots, N-1$. A MS is reconnected to an alternate BS whenever the signal strength estimate of the target BS exceeds that of the serving BS by at least H dB. For example, a handoff is performed between two BSs, BS_0 and BS_1, when

$$
\begin{aligned}
Y_1(n) &> Y_0(n) + H \quad \text{if the serving BS is } \text{BS}_0 \\
Y_0(n) &> Y_1(n) + H \quad \text{if the serving BS is } \text{BS}_1
\end{aligned}
\tag{12.1}
$$

where H denotes the hysteresis (in dB), and $Y_0(n)$ and $Y_1(n)$ are the estimated mean signal strengths (in dBm) of BS_0 and BS_1, given by

$$
Y_0(n) = \frac{1}{N} \sum_{k=n-N+1}^{n} |r_0(k)|^2_{(\text{dB})}
\tag{12.2}
$$

$$
Y_1(n) = \frac{1}{N} \sum_{k=n-N+1}^{n} |r_1(k)|^2_{(\text{dB})}
\tag{12.3}
$$

respectively, where $|\tilde{r}_i(kT_s)|^2_{(\text{dB})}$ is the kth sample of the squared envelope (in dBm), T_s is the sampling period, and N is the window length.

Many other variations of signal strength based handoff algorithms have been suggested in the literature. In one variation, handoffs are also triggered when

the measured signal strength of the serving BS drops below a threshold. For example, a handoff could be performed between BS_0 and BS_1 when

$$
\begin{aligned}
Y_1(n) &> Y_0(n) + H \text{ and } Y_0(n) > \Omega_L, \text{ if the serving BS is } BS_0 \\
Y_1(n) &> Y_0(n) \text{ and } Y_0(n) < \Omega_L, \text{ if the serving BS is } BS_0 \\
Y_0(n) &> Y_1(n) + H \text{ and } Y_1(n) > \Omega_L, \text{ if the serving BS is } BS_1 \\
Y_0(n) &> Y_1(n) \text{ and } Y_1(n) < \Omega_L, \text{ if the serving BS is } BS_1 \quad (12.4)
\end{aligned}
$$

This scheme encourages a handoff whenever the received signal strength from the serving BS drop below the threshold Ω_L, thereby reducing the probability of dropped call.

Another variation discourages handoffs when the received signal strength from the serving BS exceeds another threshold Ω_U. For example, a handoff is performed between BS_0 and BS_1 when

$$
\begin{aligned}
Y_1(n) &> Y_0(n) + H \text{ and } \Omega_L < Y_0(n) < \Omega_U, \text{ if the serving BS is } BS_0 \\
Y_1(n) &> Y_0(n) \text{ and } Y_0(n) < \Omega_L, \text{ if the serving BS is } BS_0 \\
Y_0(n) &> Y_1(n) + H \text{ and } \Omega_L < Y_1(n) < \Omega_U, \text{ if the serving BS is } BS_1 \\
Y_0(n) &> Y_1(n) \text{ and } Y_1(n) < \Omega_L, \text{ if the serving BS is } BS_1 \quad (12.5)
\end{aligned}
$$

This scheme avoids unnecessary handoffs, thereby reducing the network load and network delay.

Direction biased handoff algorithms have also been suggested for improving the handoff performance in urban microcells [23]. These algorithm incorporate moving direction information into the handoff algorithm to encourage handoffs to BSs that the MS is approaching, and to discourage handoffs to BSs that the MS is moving away from. Let BS_s denote the serving BS. A direction biased handoff algorithm can be defined by grouping all the BSs being considered as handoff candidates, including BS_s, into two sets based on their direction information. Define

$$
\begin{aligned}
\mathcal{A} &:= \text{ the set of BSs the MS is approaching} \quad (12.6) \\
\mathcal{R} &:= \text{ the set of BSs the MS is moving away from.} \quad (12.7)
\end{aligned}
$$

By introducing an encouraging hysteresis H_e, and a discouraging hysteresis H_d, a direction biased handoff algorithm requests a handoff to BS_j if $BS_j \in \mathcal{R}$ and

$$
\begin{aligned}
Y_j(n) &> X_s(n) + H, \text{ if } BS_s \in \mathcal{R} \\
Y_j(n) &> X_s(n) + H_d, \text{ if } BS_s \in \mathcal{A} \quad (12.8)
\end{aligned}
$$

or if $BS_j \in \mathcal{A}$ and

$$Y_j(n) > X_s(n) + H_e, \text{ if } BS_s \in \mathcal{R}$$
$$Y_j(n) > X_s(n) + H, \text{ if } BS_s \in \mathcal{A} . \tag{12.9}$$

To encourage handoffs to BSs in $\mathcal{A}$ and discourage handoffs to BSs in $\mathcal{R}$, the hysteresis values should satisfy $H_e \leq H \leq H_d$. When equality holds, the algorithm reduces to the conventional method described in (12.1). Good values for H_e, H, and H_d depend on the propagation environment and BS layout. In general, a direction biased handoff algorithm can maintain a lower mean number of handoffs and handoff delay, and provide better cell membership properties.

2. PILOT-TO-INTERFERENCE RATIO BASED SOFT HANDOFF ALGORITHMS

In CDMA based systems each BS transmits pilot signal to assist soft handoff [96]. In synchronous CDMA systems, all BSs use the same pilot code and the BS are distinguished by using different phase shifts of the same pilot. In asynchronous CDMA systems, each cell is allocated a distinct scrambling code. In any case, the MSs use the pilot signals to initiate and complete handoffs. Each pilot signal strength is measured by its pilot-to-interference ratio (PIR), which is the ratio of received pilot energy per chip to total interference spectral density:

$$\text{PIR} = \frac{E_c}{I_t} \tag{12.10}$$

An *active set* refers to the set of BSs to which a MS is connected at any given time. The active set contains a single BS most of time, but additional BSs are added to the set during soft handoff.

Soft handoffs are initiated based on a hysteresis imposed on the PIRs. An upper threshold, T_{add}, determines the pilot signal level for which qualifying BSs are added to the active set, whereas a lower threshold, T_{drop}, determines when the weak pilot BSs are dropped from the active set. The handoff margin, the difference between T_{add} and T_{drop}, is an indicator of how long a soft handoff will take on average. A wider margin results in a longer average soft handoff duration. Fig. 12.3 shows how changes in the handoff parameters affects the handoff region. Reducing T_{drop} and T_{add} expands the cell boundaries and thus increases the soft handoff region.

The soft handoff margin and thresholds are very important parameters in determining system performance, and need to be carefully optimized for a given situation. Allowing more MSs to be in soft handoff mode will decreases reverse link interference, by allowing more MSs to benefit from macrodiversity. That is, the MSs in handoff are connected to the best available link and, therefore, do not transmit excessive power. However, the increase in the number of MSs in soft handoff and the increase in the average handoff duration can
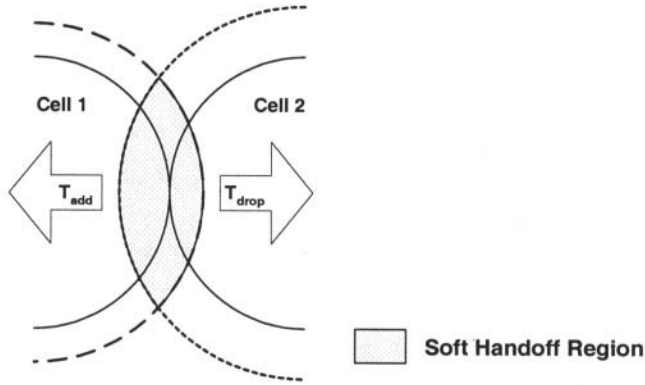
Figure 12.3.   Soft handoff boundaries and region.

increase system complexity and tie up already scarce system resources. Also, as mentioned previously, soft handoff increases the forward link interference by allowing multiple BSs to transmit to one MS. The challenge is to optimized the handoff parameters so that the capacity and quality of service requirements are satisfied, while keeping the operational cost and system complexity down. Other important soft handoff parameters include the T_{drop} timer and the ratio between the handoff region and total cell area. The T_{drop} timer is the length of time that a signal level must remain below T_{drop} to drop a BS from the active set.

3.   SIGNAL STRENGTH AVERAGING

The received squared envelopes $|\tilde{r}(t)|^2$ are affected by Ricean fading, log-normal shadowing, and path loss attenuation. Here we assume a narrow-band system with flat fading, although the techniques that are described in the sequel can be extended to other systems with some modification. For middle-band TDMA systems, it is likely that the necessary signal strength information can be obtained from the adaptive equalizer or channel estimator. Likewise, for wide-band CDMA systems, the tap weightings in a RAKE receiver could be used to estimate the received signal strength.

Two Ricean fading models are considered. The first model assumes that the received bandpass signal is

$$r(t) = g_I(t) \cos 2\pi f_c t - g_Q(t) \sin 2\pi f_c t \qquad (12.11)$$

where f_c is the carrier frequency, and $g_I(t)$ and $g_Q(t)$ are independent Gaussian random processes with variance b_0 and means $m_I(t) = m_I$ and $m_Q(t) = m_Q$, respectively. As discussed in Chapter 2.1.2 the envelope $\alpha(t) = |g_I(t) + jg_Q(t)|$, is Ricean distributed with Rice factor $K = s^2/(2b_0)$, where $s^2 =$

$m_I^2 + m_Q^2$. The second Ricean fading due to Aulin [16], assumes that

$$
\begin{aligned}
m_I(t) &= s \cdot \cos(2\pi f_m \cos\theta_0 t + \phi_0) \\
m_Q(t) &= s \cdot \sin(2\pi f_m \cos\theta_0 t + \phi_0)
\end{aligned}
\tag{12.12}
$$

where $f_m \cos\theta_0$ and ϕ_0 are the Doppler shift and angle offset of the LoS signal, respectively. Once again, the envelope $\alpha(t)$ is Ricean distributed with Rice factor $K = s^2/(2b_0)$. Both models are equivalent for Rayleigh fading ($K = 0$).

As suggested in Chapter 2.4.1, the spatial correlation of the log-normal shadowing can be effectively described by the negative exponential model

$$
\phi_{\Omega_{(dBm)}\Omega_{(dBm)}}(kT) = \sigma_\Omega^2 \zeta_D^{(vT/D)|k|}
\tag{12.13}
$$

where σ_Ω is the shadow standard deviation (typically between 4 and 12 dB), and ζ_D is the spatial shadow correlation between two points separated by D m.

For LoS propagation we assume the two-slope path loss model given by (2.248). For NLoS propagation we use the model in (2.250) yielding, for example, the signal strength profile in Fig. 12.4.

Time averaging $< |\tilde{r}_i(t)|^2 >$ and hysteresis H reduce the effect of fading and shadowing variations that would otherwise cause large numbers of unnecessary handoffs. Short spatial windows average over the fades while longer spatial windows average over the shadows as well. The effect of the spatial window length on handoffs is well documented in the literature [222], [150], [238], [340]. However, for the development of fast microcellular handoff algorithms, new guidelines must be developed so that spatial averaging can be used effectively for reducing the effects of fading in microcells.

3.1   CHOOSING THE PROPER WINDOW LENGTH

One method for determining the proper window length is to use analog averaging. The following development extends the original work of Lee [194] by incorporating Aulin's Ricean fading model. With Lee's multiplicative model, the squared-envelope of the composite signal at position y is

$$
\hat{\alpha}_c^2(y)^2 = \alpha^2(y) \cdot \Omega_p(y)
\tag{12.14}
$$

where $\alpha^2(y)$ is a non-central chi-square random variable with 2 degrees of freedom (Ricean fading), and $\Omega_p(y)$ is a log-normal random variable (log-normal shadowing). If the local mean is constant with distance, then $\Omega_p(y) = \Omega_p$. Assuming ergodicity, an integral spatial average of $\hat{\alpha}_c^2(y)$ can be used to estimate the local mean Ω_p, i.e.,

$$
\bar{\Omega}_p = \frac{1}{2L} \int_{x-L}^{x+L} \hat{\alpha}_c^2(y) dy = \frac{\Omega_p}{2L} \int_{x-L}^{x+L} \alpha^2(y) dy
\tag{12.15}
$$

600
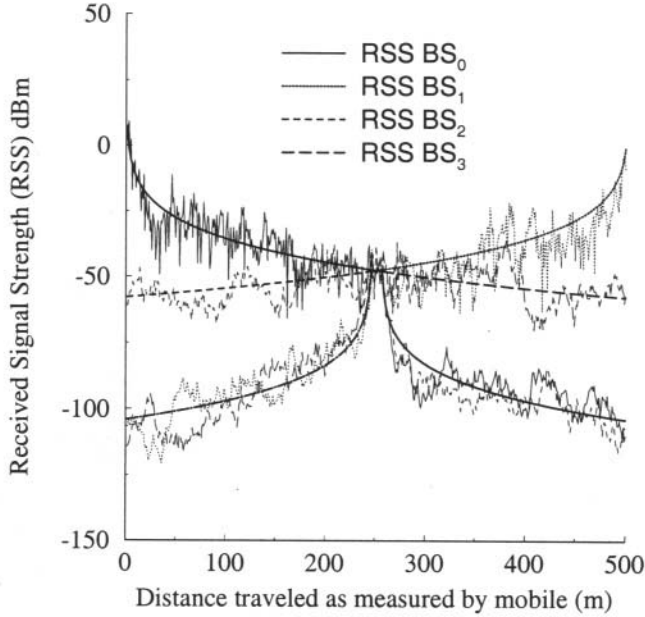
Figure 12.4. Typical average received signal strength (thick lines) and instantaneous received signal strength (thin lines) for a NLoS handoff scenario. For each particular BS, the received signal strength is shown when the MS is connected to that particular BS and the MS moves along the route in Fig. 12.1, from [18].

where the second equality holds since $\Omega_p(y)$ is constant over the spatial interval $(x-L, x+L)$. The accuracy of the estimate can be determined from the variance of (12.15), calculated as [189]

$$\sigma_{\hat{\Omega}_p}^2 = \frac{1}{L} \int_0^{2L} \left(1 - \frac{\ell}{2L}\right) \mu_{\alpha^2\alpha^2}(\ell)d\ell \tag{12.16}$$

where $\mu_{\alpha^2\alpha^2}(\ell) = \mathrm{E}[\alpha^2(y)\alpha^2(y+\ell)] - \mathrm{E}[\alpha^2(y)]\mathrm{E}[\alpha^2(y+\ell)]$ is the *spatial* autocovariance of the squared envelope, and E[x] denotes the ensemble average of x. Aulin [16] derived $\mu_{\alpha^2\alpha^2}(\ell)$ as (c.f. 2.83),

$$\mu_{\alpha^2\alpha^2}(\ell) = \left(\frac{\Omega_p}{K+1}\right)^2 J_0(2\pi\ell/\lambda_c)\left[J_0(2\pi\ell/\lambda_c) + 2K\cos\left(2\pi\ell\cos(\theta_0)/\lambda_c\right)\right] \tag{12.17}$$

where $J_0(\,\cdot\,)$ is the zero-order Bessel function of the first kind, K is the Rice factor, λ_c is the carrier wavelength, and θ_0 is the angle that the specular component makes with the MS direction of motion. The spatial autocovariance of the squared envelope can be obtained directly from Fig. 2.14 by using the time-distance transformation $f_m\tau = \ell/\lambda_c$.

Substituting (12.17) into (12.16) yields

$$\sigma_{\bar{\Omega}_p}^2 = \left(\frac{\Omega_p}{K+1}\right)^2 \frac{1}{L} \int_0^{2L} \left(1 - \frac{\ell}{2L}\right)$$

$$\times J_0(2\pi\ell/\lambda_c) \left[J_0(2\pi\ell/\lambda_c) + 2K\cos\left(2\pi\ell\cos(\theta_0)/\lambda_c\right)\right] d\ell \quad. \tag{12.18}$$

As desired, $\sigma_{\bar{\Omega}_p}^2 \to 0$ as $L \to \infty$. If L is large, then $\bar{\Omega}_p$ can be considered Gaussian since it is the summation of many independent random variables. However, if $\sigma_{\bar{\Omega}_p}^2$ is relatively large compared to Ω_p (due to small L or small Ω_p), then it is more appropriate to treat $\bar{\Omega}_p$ as a non-central chi-square random variable. In this case, it may be more appropriate to approximate $\bar{\Omega}_p$ as a log-normal random variable which has the same general shape as a non-central chi square distribution (i.e., zero at the origin with an infinitely long tail) [142], [143].

Proceeding under the assumption that $\bar{\Omega}_p$ is approximately Gaussian, the 1σ spread can be calculated to measure the accuracy of the estimator, where

$$1\sigma \text{ spread} = 10 \cdot \log_{10} \frac{\Omega_p + \sigma_{\bar{\Omega}_p}}{\Omega_p - \sigma_{\bar{\Omega}_p}} \tag{12.19}$$

with the interpretation that $\text{Prob}(|\bar{\Omega}_{p\,(dB)} - \Omega_{p\,(dB)}| \leq 1\,\sigma \text{ spread}) = 0.68.$[1] Observe from (12.18) and (12.19) that the accuracy of the local mean estimate depends on *K, L,* and θ_0. Fig. 12.5 shows the 1σ spread when $\theta_0 = 60^\circ$ for various values of *K*. In general, $\bar{\Omega}_p$ approaches Ω_p with increasing *K*. However, the angle θ_0 also affects the accuracy as shown in Fig. 12.6. When $\theta_0 = 90^\circ$ the 1 σ spread is minimized, resulting in the best estimate of the local mean. Conversely, the worst estimates occur for small θ_0 (in the neighborhood of 10° in Fig. 12.6). The actual angle that the maximum occurs is a function of *L,* and it can easily be shown that the 1 σ spread has a local minimum at $\theta_0 = 0^\circ$ and global minimum at $\theta_0 = 90^\circ$ for all *L*. In any case, the required spatial averaging distance for local mean estimation in microcells depends on *K* and θ_0.

3.2   CHOOSING THE PROPER NUMBER OF SAMPLES TO AVERAGE

Most practical signal strength estimators use samples of the signal strength rather than analog averaging. We must determine the required number and spacing of samples that should be used, to sufficiently mitigate the effects of fading. Consider the sampled squared composite envelope

$$\alpha_c^2[i] \triangleq \alpha_c^2(iS) \tag{12.20}$$

---

[1]The probability of lying within one standard deviation of the mean of a Gaussian random variable is 0.68.
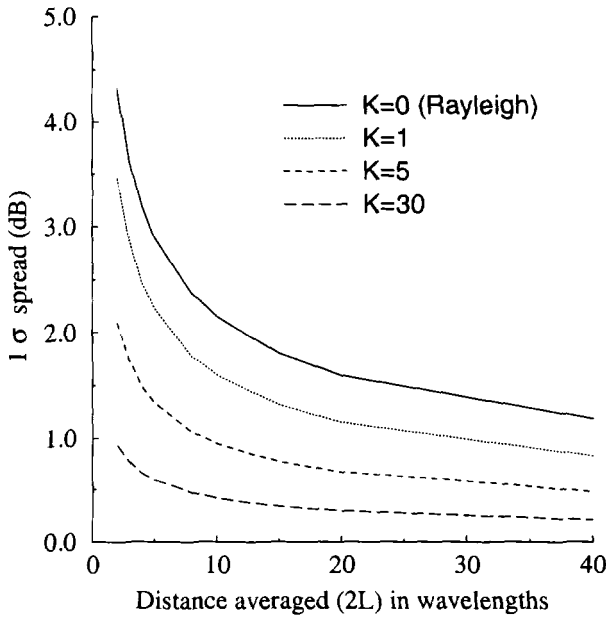
Figure 12.5. Dependency of the 1σ spread for squared-envelope samples on the averaging distance (2L) and Rice factor K when $\theta_0 = 60°$, from [18].
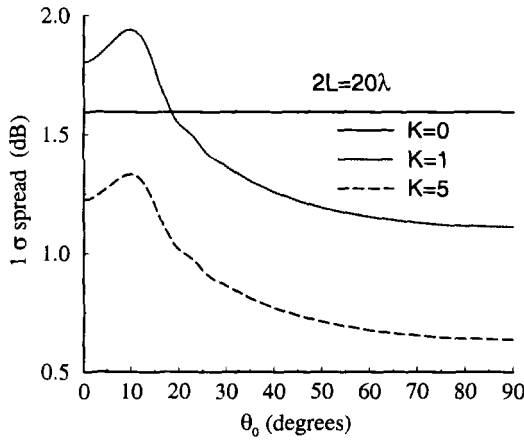
Figure 12.6. Dependency of the 1σ spread on the specular angle θ_0 mod 90°, from [18].

where S is the spatial sampling period, and i is an integer. Then the unbiased estimate

$$\bar{\Omega}_p = \frac{1}{N} \sum_{i=0}^{N-1} \alpha_c^2[i] \tag{12.21}$$

can be used to determine an estimate of the local mean squared-envelope Ω_p. As with analog averaging, the variance of this estimate can be used to measure its accuracy, where

$$\sigma_{\hat{\Omega}_p}^2 = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} E\left[\alpha_c^2[i]\alpha_c^2[j]\right] - (E[\bar{\Omega}_p])^2 \ . \tag{12.22}$$

By using (12.17) along with the symmetric properties of the autocovariance, (12.22) becomes

$$
\begin{aligned}
\sigma_{\hat{\Omega}_p}^2 &= \frac{\mu_{\alpha^2\alpha^2}(0)}{N} + 2\sum_{j=1}^{N-1}\left(\frac{N-j}{N^2}\right)\mu_{\alpha^2\alpha^2}(Sj) \\
&= \left(\frac{\Omega_p}{K+1}\right)^2\left[\frac{1+2K}{N} + 2\sum_{j=1}^{N-1}\left(\frac{N-j}{N^2}\right)J_0(2\pi Sj/\lambda_c)\right. \\
&\qquad \left.\times\left[J_0(2\pi Sj/\lambda_c) + 2K\cos\left(2\pi Sj\cos(\theta_0)/\lambda_c\right)\right]\right] \tag{12.23}
\end{aligned}
$$

where S is measured in wavelengths (λ_c). Note that $\sigma_{\hat{\Omega}_p}$ depends on N, K, S, and θ_0. Fortunately, the effect of each parameter is nearly independent of the others. Fig. 12.7 illustrates the relationship between S and K for $\theta_0 = 0°$, where $N = \lceil 20\lambda_c/S \rceil$ so that the averages are over $20\lambda_c$ (and $\lceil x \rceil$ denotes the smallest integer greater than or equal to x). Increasing N for a fixed S will increase the spatial averaging distance, thereby lowering the $1\ \sigma$ spread in a manner similar to analog averaging in Fig. 12.5. The discontinuities in Fig. 12.7 are due to the $\lceil x \rceil$ function. Observe that if $S < 0.5\lambda_c$ then the discrete local mean estimate is approximately equivalent to the estimate from analog averaging ($\theta_0 = 0°$ in Fig. 12.6) over the same spatial distance. Similar to Fig. 12.6, we also observe that small Rice factors, e.g., $K = 0.1$ and $K = 1$, at $\theta_0 = 0°$ increase the $1\ \sigma$ spread. The spikes at $0.5\lambda_c$ and $1\lambda_c$, correspond to the location of the first lobe of the autocovariance function given by (12.17) and plotted in Fig 2.13.

Although we often assume $\theta_0 = 0°$ in our treatment, Fig. 12.8 shows the relationship between the $1\ \sigma$ spread and S, for $K = 1$, $N = \lceil 20\lambda_c/S \rceil$, and several values of θ_0. Increasing θ_0 generally lowers the $1\ \sigma$ spread except for some small angles as shown in Fig. 12.6; it also shifts the spike at $0.5\lambda_c$ to the right, because the first sidelobe of (12.17) shifts as θ_0 increases.

To summarize, the spatial averaging distance that is needed to sufficiently reduce the effects of fading depends on K and θ_0. If sample averaging is used, then the sample spacing should be less than $0.5\lambda_c$. As a rule of thumb, a spatial averaging distance of 20 to 40 λ_c should be sufficient for most applications.
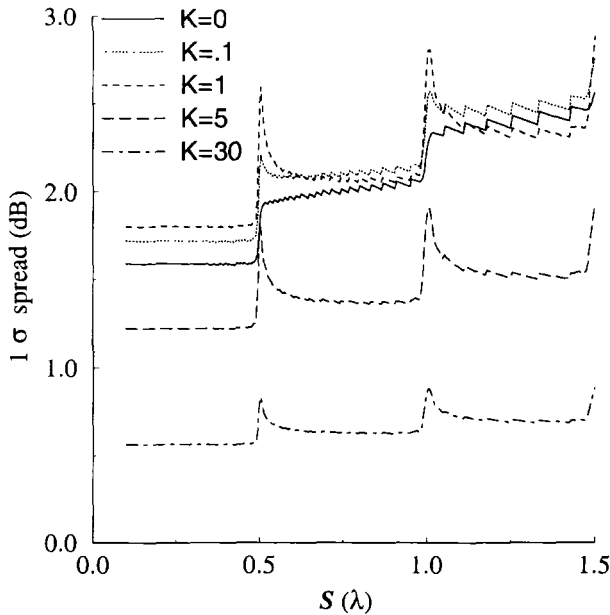
Figure 12.7. 1σ spread versus S for various K, with $\theta_0 = 0°$, distance averaged $= 20\lambda_c$, from [18].

4. VELOCITY ESTIMATION IN CELLULAR SYSTEMS

Temporal based handoff algorithms can yield poor handoff performance in microcells due to the diverse propagation environment and the wide range of MS velocities. Consider the NLoS handoff scenario shown in Fig. 12.1, where a MS traveling from BS_0 has a Ricean faded log-normal shadowed LoS signal from BS_0, and a Rayleigh faded log-normal shadowed NLoS signal from BS_1, until it rounds the corner where the situation is suddenly reversed. The loss (gain) of the LoS component causes a rapid decrease (increase) in the signal strength. Effective handoff algorithms for this scenario should use short temporal averaging window and a large hysteresis, so that rapid changes in the mean signal strength are detected and unnecessary handoffs are prevented [238]. Unfortunately, temporal averaging with a short fixed window length gives optimal handoff performance for only a single velocity. For example, consider again the handoff scenario in Fig. 12.1 along with the received signal strength profile in Fig. 12.4. Assume log-normal shadowing with $\sigma_\Omega=6$ dB and choose D so that $\phi_{\Omega_{(dB)}\Omega_{(dB)}}(d) = 0.1\sigma_\Omega^2$ at $d = 30$ m in (12.13). The simulation of a 2.27 s non-overlapping temporal power averaging handoff algorithm with a hysteresis $H = 8$ dB has a handoff performance shown by the
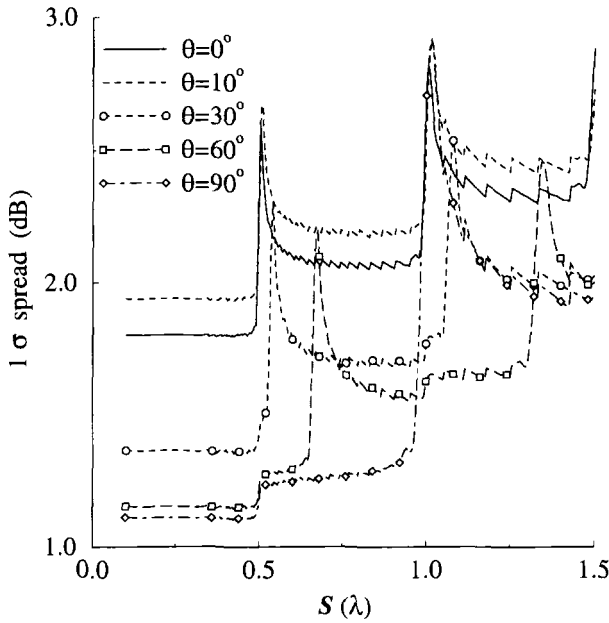
Figure 12.8. 1 σ spread versus S (λ_c) for various θ_0, $K = 1$, distance averaged $= 20\lambda_c$, from [18].

lines in Fig. 12.14.[2] The handoff performance is evaluated by the mean number of handoffs, averaged over 1000 runs, versus the distance from BS_0 where 50% (and 90%) of the MSs have made a handoff to BS_1, i.e., $\Pr(BS_1) = 0.5$ and $\Pr(BS_1) \approx 0.9$ at the abscissa. This distance gives a measure of the handoff delay, assuming that handoffs will occur between BS_0 and BS_1 only.

Fig 12.14 only shows the handoff request delay, while in a real system the network delay should also be included. However, the performance of a velocity adaptive handoff algorithm can still be evaluated without knowledge of the network delay. For example, suppose that the receiver threshold is −90 dBm. Also, assume that a good handoff algorithm should have at least 90 % of the MSs handed off before a distance d_{cutoff}, where d_{cutoff} is chosen as that distance where the mean signal strength is $2\sigma_\Omega$ above −90 dBm. If $\sigma_\Omega = 6$ dB and the data from Fig. 12.4 is used, then a signal strength of $-90 + 12 = -78$ dBm occurs at 283 m for BS_0. Hence, if the velocity adaptive handoff algorithm can adapt to the point at 5 km/h in Fig. 12.14, corresponding to handoff requests at a distance 262 m, and the maximum speed of a MS turning the corner is

---

[2] A 2.27 s window corresponds to a $20\lambda_c$ spatial window at a velocity of 5 km/h, assuming a carrier frequency of 1.9 GHz. Section 5. further details the simulation.

40 km/h (40/3.6 m/s), then a maximum network delay of

$$\text{Max Network Delay} = \frac{3.6 \text{ s}}{40 \text{ m}} (283 - 262) \text{ m} = 1.89 \text{ s}$$

can be tolerated. For some of the newer cellular standards, e.g., GSM, this network delay is tight but acceptable, implying the usefulness of velocity adaptive handoff algorithms discussed here. In the above example the 5 km/h point on the curve in Fig. 12.14 was chosen as the desired operating point, because the best handoff performance occurs near the knee of the curve where the mean number of handoffs and handoff delay are jointly minimized. Other hysteresis and window lengths could possibly result in better performance. However, the settings used here ($H = 8$ dB and a $20\lambda_c$ spatial window) are adequate to illustrate the usefulness of velocity adaptive handoff algorithms.

Some cellular system proposals have suggested the deployment of micro-cells along with "umbrella" macrocells for accommodating high speed MSs. Velocity estimation will be necessary for these systems along with a macro-to-microcell and micro-to-macrocell handoff scheme. Alternatively, if a low network handoff delay can be achieved, then a velocity adaptive handoff algorithm can maintain good link quality without the need for umbrella macrocells.

4.1 LEVEL CROSSING RATE ESTIMATORS

It is well known that the zero crossing rates of the quadrature components $g_I(t)$ and $g_Q(t)$ and the level crossing rates of the envelope $|g(t)| = |g_I(t) + g_Q(t)|$ of a received sinusoid in noise, are functions of the MS velocity as discussed in Chapter 2.1.4. The envelope **level crossing rate** (LCR) is defined as the average number of positive going crossings per second, a signal makes of a predetermined level R. Likewise, the **zero crossing rate** (ZCR) is defined as the average number of positive going zero crossings a signal makes per second.

Assuming fading model in (12.11), the means m_I and m_Q can be subtracted from the inphase and quadrature components and the ZCR of the resulting signals be used to estimate the velocity. Rice gives the ZCR of $g_I(t) - m_I$ or $g_Q(t) - m_Q$ as(c.f. 2.103)

$$L_{\text{ZCR}_1} = \frac{1}{\pi} \sqrt{\frac{b_2}{b_0}} \tag{12.24}$$

and the envelope LCR with respect to the level R as, (2.90) and (2.91),

$$L_R = \int_0^\infty \dot{\alpha} p(R, \dot{\alpha}) d\dot{\alpha} = \frac{R(2\pi)^{-3/2}}{\sqrt{Bb_0}} \int_{-\pi}^\pi d\theta \int_0^\infty \dot{\alpha} d\dot{\alpha} \tag{12.25}$$

$$\exp\left\{ -\frac{1}{2Bb_0} \left[B\left(R^2 - 2Rs\cos\theta + s^2\right) + (b_0\dot{\alpha} + b_1 s \sin\theta)^2 \right] \right\}$$

where $p(R, \dot{\alpha})$ is the joint probability density function of the envelope α (evaluated at $r = R$) and the slope of the envelope $\dot{\alpha}$, and $B = b_0 b_2 - b_1^2$. From (2.93), the b_n are equal to

$$b_n = (2\pi)^n b_0 \int_0^{2\pi} \hat{p}(\theta)(f_m \cos\alpha - f_q)^n d\theta + (2\pi)^n \int_{-B_w/2}^{B_w/2} \frac{N_o}{2} f^n df \quad (12.26)$$

where $2b_0$ is the scatter power, v is the velocity, λ_c is the carrier wavelength, $f_m = v/\lambda_c$ is the maximum Doppler frequency, $f_s = f_c + f_q$ is the frequency of the specular or LoS component, and $\hat{p}(\theta)$ is the continuous AoA distribution of the *scatter* component of the arriving plane waves [173]. The second term in (12.26) is due to additive bandpass Gaussian noise, centered at f_c, with a two-sided power spectral density of $N_o/2$ watts/Hz and a noise bandwidth of B_w Hz, resulting in a total power of $N_o B_w$ watts. For the special case when $\theta_0 = 0°$ and there is 2-D isotropic scattering, then $\hat{p}(\theta) = 1/(2\pi)$, $-\pi \le \theta \le \pi$ and (12.26) can be written as

$$b_n = (2\pi)^n \frac{b_0}{\pi} \int_{-f_m}^{f_m} \frac{f^n}{\sqrt{f_m^2 - f^2}} df + (2\pi)^n \int_{-B_w/2}^{B_w/2} \frac{N_o}{2} f^n df \ . \quad (12.27)$$

With Aulin's Ricean fading model with the means of $g_I(t)$ and $g_Q(t)$ defined in (12.12), the ZCR of $g_I(t)$ or $g_Q(t)$ is [282]

$$L_{ZCR_2} = L_{ZCR_1} \left[e^{-\gamma} I_0(\beta) + \frac{b^2}{2\gamma} I_e \left(\frac{\beta}{\gamma}, \gamma \right) \right], \quad (12.28)$$

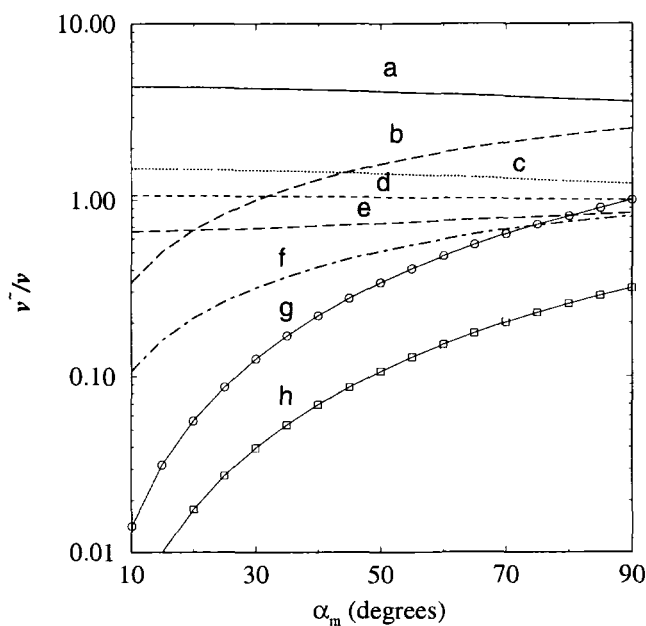where $I_0(x)$ is the zero-order modified Bessel function of the first kind, and

$$\gamma = \frac{a^2 + b^2}{4}, \quad \beta = \frac{a^2 - b^2}{4}, \quad a = \sqrt{2K}$$

$$I_e(k, x) = \int_0^x e^{-u} I_0(ku) du, \quad b = 2\pi f_m \cos(\theta_0) \sqrt{\frac{2K b_0}{b_2}} \ . (12.29)$$

Two-dimensional isotropic scattering is a reasonable assumption for macrocells. However, in microcells the scattering is often non-isotropic. Nevertheless, one approach is to derive the velocity estimators under the assumption of 2-D isotropic scattering with no additive noise, and afterwards study the effects of the mismatch caused by non-isotropic scattering and noise. Using (12.24) along with the definition for the b_n in (12.27) with $N_o = 0$ gives

$$L_{ZCR_1} = \sqrt{2} v/\lambda_c \quad (12.30)$$

and (c.f. 2.99)

$$L_R = (v/\lambda_c) \sqrt{2\pi(K + 1)} \rho e^{-K-(K+1)\rho^2} I_0 \left(2\rho \sqrt{K(K + 1)} \right) \quad (12.31)$$

where $\rho = R/R_{\text{rms}}$, where $R_{\text{rms}} = \sqrt{\Omega_p}$ is the rms signal level. Likewise, for $\theta_0 = 0°$ and Aulin's Ricean fading model with the means in (12.12) we have

$$\gamma = \frac{3K}{2}, \quad \beta = -\frac{K}{2}, \quad a = \sqrt{2K}, \quad b = 2\sqrt{K} \qquad (12.32)$$

and L_{ZCR_2} reduces to

$$L_{\text{ZCR}_2} = (v/\lambda_c)\sqrt{2}\left[e^{-3K/2}I_0\left(-\frac{K}{2}\right) + \frac{4}{3}I_e\left(-\frac{1}{3}, \frac{3K}{2}\right)\right] \quad . \qquad (12.33)$$

Clearly, each of the above level crossing rate estimators is proportional to the velocity v and, hence, can be used as a velocity estimator. However, it remains to be seen if they are robust to K, non-isotropic scattering, additive noise, and other factors. We first consider the robustness with respect to K and treat the other factors afterwards.

L_{ZCR_1} is not affected by K. Fig. 2.14 in compared the level crossing rate L_R for different K with the conclusion that the LCR around $\rho = 0$ dB is roughly independent of K. This attractive property suggests that the level crossing rate can be used to provide a velocity estimate that is robust to K. Consequently, the steps for using the LCR (or ZCR) of $|g(t)|$ (or $g_I(t)$ or $g_Q(t)$), for velocity estimation are; determine R_{rms} (or m_I or m_Q), estimate the number of crossings per second $\hat{L}_{R_{\text{rms}}}$ (or $\hat{L}_{\text{ZCR}_1}$), and use (12.31) to solve for v, with $\rho = 1$ and $K = 0$ (or 12.30 for ZCR$_1$). Thus, the following velocity estimators are robust with respect to K assuming the Ricean fading model in (12.11):

$$\hat{v}_{\text{ZCR}_1} \approx \frac{\lambda_c\hat{L}_{\text{ZCR}}}{\sqrt{2}}, \quad \hat{v}_{\text{LCR}} \approx \frac{\lambda_c\hat{L}_{R_{\text{rms}}}}{\sqrt{2\pi}e^{-1}} \quad . \qquad (12.34)$$

Fig. 12.9 shows the effect of K and θ_0 on L_{ZCR_2}. Notice that if the angle of the specular component is $\theta_0 = 0°$ or $180°$, then L_{ZCR_2} can have up to 40% relative error. Consequently, a non-zero value of K should be chosen as default to minimize the effect of K. Choosing $K \approx 0.61$ yields a maximum error of at most 20% which is quite acceptable for urban LoS velocity adaptive handoff applications. In this case, the velocity estimate from (12.33) becomes

$$\hat{v}_{\text{ZCR}_2} \approx \frac{\lambda_c\hat{L}_{\text{ZCR}_2}}{1.2\sqrt{2}} \quad . \qquad (12.35)$$

4.2   COVARIANCE APPROXIMATION METHODS

A velocity estimator has been proposed by Holtzman and Sampath that relies upon an estimate of the autocovariance between faded samples $r[i]$, where the $r[i]$ can be envelope, squared-envelope, or log-envelope samples [169], [290].
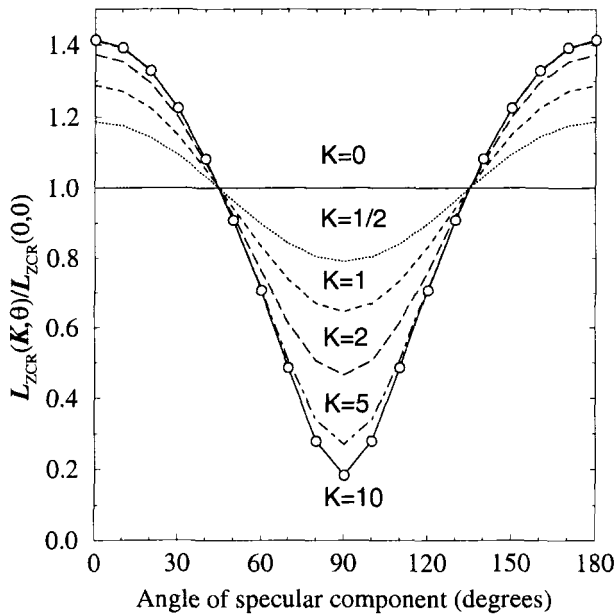
Figure 12.9.   Normalized zero crossing rate versus K and θ_0, from [18].

With this method, referred to here as the **covariance method** (COV), the statistic

$$V = \frac{1}{N} \sum_{k=1}^{N} (r[k+\tau] - r[k])^2 \qquad (12.36)$$

is calculated. If N is large and ergodicity applies, then V can be replaced by the ensemble average

$$E[V] = 2\mu_{rr}(0) - 2\mu_{rr}(\tau) \qquad (12.37)$$

where $\mu_{rr}(\tau)$ denotes the autocovariance of $r[k]$. The general form for $\mu_{rr}(\tau)$, assuming squared-envelope samples, can be derived from [16] and [290] as

$$
\begin{aligned}
\mu_{rr}(\tau) \;=\; & 4a(\tau)\left[a(\tau) + s^2 \cdot \cos(2\pi f_m \tau \cos\theta_0)\right] + 4c^2(\tau) \\
& + \frac{2N_o a(\tau)\sin(B_w \pi \tau)}{\pi\tau} \\
& + \frac{4K N_o b_0^2 \cos(2\pi f_m \tau \cos\theta_0)\sin(B_w \pi \tau)}{\pi\tau} \\
& + \frac{N_o{}^2 \sin(B_w \pi \tau)^2}{\pi^2 \tau^2} \qquad (12.38)
\end{aligned}
$$

610

where [16]

$$a(\tau) \quad = \quad b_0 \int_0^{2\pi} \hat{p}(\theta) \cos\left(2\pi f_m \tau \cos\theta\right) d\theta \tag{12.39}$$

$$c(\tau) \quad = \quad b_0 \int_0^{2\pi} \hat{p}(\theta) \sin\left(2\pi f_m \tau \cos\theta\right) d\theta \; . \tag{12.40}$$

This estimator depends on $\hat{p}(\theta)$ and, hence, is also a function of the scattering environment. Like the LCR estimator, we first assume isotropic scattering without additive noise to derive a velocity estimator and afterwards evaluate the effect of non-isotropic scattering and noise.[3] If the channel is characterized by isotropic scattering and squared-envelope samples are used, then using (12.17) gives

$$\mathrm{E}[V] = \overline{V} \quad = \quad 2\left(\frac{\Omega_p}{K+1}\right)^2 \left[(1+2K) \right. \tag{12.41}$$
$$\left. - J_0(2\pi f_m \tau)\left[J_0(2\pi f_m \tau) + 2K \cos(2\pi f_m \tau \cos\theta_0)\right)\right]\right]$$

which is dependent on K and θ_0. If $\mu_{rr}(0)$ is known exactly, then the bias with respect to K can be eliminated for small τ by the normalization [290]

$$\frac{\overline{V}}{\mu_{rr}(0)} \approx (2\pi v \tau_t / \lambda_c)^2 \frac{1 + 2K + K\cos(2\theta_0)}{(1+2K)} \tag{12.42}$$

so that [290]

$$\hat{v}_{\mathrm{COV}} \approx \frac{\lambda_c}{2\pi\tau_t} \sqrt{\frac{\overline{V}}{\mu_{rr}(0)}} \tag{12.43}$$

where τ_t is the sample spacing in seconds/sample.

In large co-channel interference situations it may be preferable to modify the above scheme since the empirical average in (12.36), and in particular $\mu_{rr}(0)$, is sensitive to co-channel interference as shown in [182]. Consequently, defining,

$$U(\tau) = \frac{1}{N}\sum_{k=1}^{N} r[k+\tau]r[k] - \left(\frac{1}{N}\sum_{k=1}^{N} r[k]\right)^2 \tag{12.44}$$

and $V_2 = 2U(\tau_1) - 2U(\tau_2)$, yields

$$\mathrm{E}[V_2] = 2\mu_{rr}(\tau_1) - 2\mu_{rr}(\tau_2) \tag{12.45}$$

so that $\mathrm{E}[V_2]/\mu_{rr}(0)$ is equal to (12.42) with $\tau_t = \tau_2^2 - \tau_1^2$, and a result similar to (12.43) follows.

---

[3] Only isotropic scattering was considered in [169], [290]

Whether V or V_2 is used, $\mu_{rr}(0)$ is never known exactly and must be estimated by the MS in the same way that m_I, m_Q, and R_{rms} must be estimated in the ZCR and LCR methods, respectively. Consequently, to actually use (12.43) it must be shown or verified that

$$v \propto E\left[\sqrt{\frac{\overline{V}}{\mu_{rr}(0)}}\right] . \tag{12.46}$$

This is analytically difficult, but simulation results in Section 5. suggest that (12.43) is a useful approximation to (12.46).

It is also shown in Appendix 10A that

$$\lim_{\tau \to 0} \hat{v}_{\text{COV}} = \lim_{\tau \to 0} \frac{\lambda_c}{2\pi\tau}\sqrt{\frac{\overline{V}}{\mu_{rr}(0)}} = v\sqrt{\frac{1 + 2K + K\cos(2\theta_0)}{1 + 2K}} . \tag{12.47}$$

It follows from (12.43) and (12.47) that K and θ_0 cause at most 20% error in v [290], thus providing a velocity estimator that is reasonably robust with respect to K.

4.3   VELOCITY ESTIMATOR SENSITIVITY

To illustrate the sensitivity of the velocity estimators, the ratio of the corrupted velocity estimate to the ideal velocity estimate is used. For the LCR and ZCR velocity estimators with the fading model in (12.11) we have

$$\frac{\tilde{v}}{v} = \frac{\tilde{L}_{R_{\text{rms}}}(\tilde{b}_0, \tilde{b}_1, \tilde{b}_2)}{L_{R_{\text{rms}}}(b_0, b_1, b_2)} \tag{12.48}$$

and

$$\frac{\tilde{v}}{v} = \frac{\tilde{L}_{\text{ZCR}_1}}{L_{\text{ZCR}_1}} = \sqrt{\frac{\tilde{b}_2}{\tilde{b}_0} \cdot \frac{b_0}{b_2}} \tag{12.49}$$

where $\tilde{v}$ denotes the corrupted velocity estimate, and $\tilde{L}_{R_{\text{rms}}}(\tilde{b}_0, \tilde{b}_1, \tilde{b}_2)$ and $L_{R_{\text{rms}}}(b_0, b_1, b_2)$ are given by (12.25) with the appropriate values of b_n and $b_{n'}$, respectively. Little simplification results for the LCR method in general. However, when $K = 0$ (12.49) simplifies to [173]

$$\frac{\tilde{v}}{v} = \sqrt{\frac{\tilde{b}_2/\tilde{b}_0 - \tilde{b}_1^2/\tilde{b}_0^2}{b_2/b_0 - b_1^2/b_0^2}} . \tag{12.50}$$

For Aulin's fading model in (12.12) the sensitivity of the ZCR is

$$\frac{\tilde{v}}{v} = \frac{\tilde{L}_{\text{ZCR}_2}}{L_{\text{ZCR}_2}} = \frac{\tilde{L}_{\text{ZCR}_1}}{L_{\text{ZCR}_1}} \frac{\left[e^{-\tilde{\gamma}}I_0(\tilde{\beta}) + \frac{\tilde{b}^2}{2\tilde{\gamma}}I_e\left(\frac{\tilde{\beta}}{\tilde{\gamma}}, \tilde{\gamma}\right)\right]}{\left[e^{-\gamma}I_0(\beta) + \frac{b^2}{2\gamma}I_e\left(\frac{\beta}{\gamma}, \gamma\right)\right]} \tag{12.51}$$

where $\tilde{\beta}$, $\tilde{\gamma}$, and $\tilde{b}$ are given by (12.29) using (12.26) where appropriate. Likewise, for the covariance method using squared envelope samples we have

$$
\frac{\tilde{v}}{v} = \frac{\frac{\lambda_c}{2\pi\tau}\sqrt{\frac{2\tilde{\mu}_{rr}(0)-2\tilde{\mu}_{rr}(\tau)}{\tilde{\mu}_{rr}(0)}}}{\frac{\lambda_c}{2\pi\tau}\sqrt{\frac{2\mu_{rr}(0)-2\mu_{rr}(\tau)}{\mu_{rr}(0)}}} .
\tag{12.52}
$$

4.3.1   EFFECT OF THE SCATTERING DISTRIBUTION

Here we study the sensitivity of the velocity estimators to the scattering distribution by using four different non-isotropic scattering models. With the first model S1, plane waves arrive from one direction only with a varying degree of directivity as might happen when signals are channeled along a city street. The probability density of the scatter component of the arriving plane waves as a function of angle of arrival has the form in (directive), where the vehicle motion is in the direction of $\theta = 0^\circ$, and θ_m determines the directivity of the incoming plane waves. Fig. 2.7 shows a polar plot of $\hat{p}(\theta)$ for $\theta_m = 30^\circ, 60^\circ$ and 90°. The second model S2, assumes that the plane waves can arrive from either the front ($\theta = 0^\circ$) or back ($\theta = 180^\circ$), which may be typical for city streets that dead end at another street. In this case $\frac{\hat{p}(\theta)}{2}$ and $\frac{\hat{p}(\theta-\pi)}{2}$ are combined to form the distribution versus angle of arrival. The resulting density is similar to Fig. 2.7 but with lobes extending in both the 0° and 180° directions. The third and fourth models S3 and S4, respectively, are similar to S1 and S2 except that the distributions are rotated by 90°, so that the plane waves tend to arrive perpendicular to the direction of travel. This may occur when a MS passes through a street intersection. The effect of the scattering distribution is determined for the cases when the velocity estimator has been designed for i) isotropic scattering and, ii) scattering model S1 with $\theta_m = 90^\circ$. The scattering model that the velocity estimator has been designed for will determine the values of b_0, b_1 and b_2 in the denominators of (12.48)–(12.51), while the values of $\tilde{b}_0$, $\tilde{b}_1$ and $\tilde{b}_2$ depend on the scattering environment that is actually present. The effect of non-isotropic scattering on the COV estimate (12.52) can be found from the results in Appendix 10A with $N_o = 0$, or by using small values of τ in (12.52). Here we chose the latter with $\tau = 1/50$.

Fig. 12.10 shows the effect of the scattering distribution on each of the velocity estimators. Due to the very large number of possible scenarios, only the most significant results are plotted in Fig. 12.10 and curves similar (but not equal) to the plotted curves are simply asterisked in the accompanying table. Velocity estimators with the subscript "d" in Fig. 12.10 correspond to those that are designed for scattering model S1 with $\theta_m = 90^\circ$. By using

Figure 12.10. Non-isotropic scattering effects. — corresponds to a curve that had insufficient precision to be reported. Superscript * denotes that the curve is approximately equal the letter curve, from [18].

	Scattering model							
	1		2		3		4	
Method	$K = 0$	$K = 1$	$K = 0$	$K = 1$	$K = 0$	$K = 1$	$K = 0$	$K = 1$
LCR	h	—	c^*	c^*	f	f	f	f
LCR_d	g	—	a	c	b	f^*	b	f^*
ZCR_1	c^*	c^*	c^*	c^*	f	f	f	f
ZCR_{1d}	d^*	d^*	d^*	d^*	f^*	f^*	f^*	f^*
ZCR_2	c^*	d^*	c^*	d^*	f	d^*	f	d^*
ZCR_{2d}	d^*	d	d^*	d	f^*	d^*	f^*	d^*
COV	h	d^*	c^*	d^*	f	e^*	f	e^*
COV_d	g	d^*	a	d^*	b	e	b	e

Fig. 12.10 the relative robustness of the various velocity estimators to the scattering distribution has been summarized by the ranking in Table 12.1.

Curve	Rank	$\tilde{v}/v$ at $\theta_m = 10°$	$\tilde{v}/v$ at $\theta_m = 90°$
d	Excellent	1.06	1.0
e , c	Very Good	.66 , 1.52	.85 , 1.24
b , f	Good	.34 , .107	2.6 , .82
a , g , h	Poor	4.5 , .014 , .004	3.6 , 1.0 , .32

Table 12.1.   Robustness to the scattering distribution.

In urban situations, robustness with respect to scattering models S1 and S2 is important. The LCR and COV methods are very sensitive to the directivity in scattering model S1 when $K = 0$ as shown by curve "h". This sensitivity can be partially mitigated by using the velocity estimators LCR_d and COV_d that have been designed for scattering model S1 with $\theta_m = 90°$ as shown by curve "g." However, the price for increased robustness to scattering model S1 is the increased sensitivity of LCR_d to scattering models S2, S3, and S4 when $K = 0$. Fortunately, the presence of even a small specular component ($K = 1$) reduces the sensitivity as seen in COV ($K = 1$) and LCR_d. In contrast a specular component does not reduce the sensitivity of the LCR estimator in scattering models S2, S3, and S4, because $b_1 = 0$, and therefore the ratio of the crossing rates in (12.50) depends on $\ddot{b}_2$ and b_2 and is independent of K. Results are not shown for LCR or LCR_d with $K = 1$ scattering model S1, due to numerical difficulties in calculating (12.25) for small θ_m. For large $\theta_m > 80°$ the results obtained were very close to curve "d."

The ZCR velocity estimator is generally more robust than the LCR and COV methods. The presence of a small specular component improves robustness to the scattering distribution as seen in ZCR_2, $ZCR_{2,d}$ (ZCR_1 and ZCR_{1_d} are independent of K). Also, velocity estimators that have been designed for scattering model S1 with $\theta_m = 90°$ perform slightly better than those designed for isotropic scattering. However, the improvement obtained by using these velocity estimators must be weighed against the relative error that will be introduced if the scattering is actually isotropic. For LCR_d and COV_d, $\tilde{v}/v = .316$ and for ZCR_d $\tilde{v}/v = 1.15$. Since all the velocity estimators seem to have some sensitivity to the scattering distribution, and sensitivity is greatly reduced when $K > 0$, we conclude that those designed for isotropic scattering should be adequate.

In summary, for very directive situations where the plane waves arrive from either the front or back but not both, the ZCR, COV_d, or LCR_d methods are the

most robust. If the plane waves arrive from both the front and back, then all the velocity estimators with the exception of LCR_d are reasonably robust. The sensitivity to directivity is reduced when a specular component is present. In the unlikely event that $K = 0$ and plane waves arrive from the perpendicular direction with high directivity, all methods will have a significant bias. Finally, another method for overcoming the sensitivity to the scattering distribution is to obtain velocity estimates from signals arriving from a distant cell or an umbrella cell, since they will experience isotropic scattering.

4.3.2 EFFECTS OF ADDITIVE GAUSSIAN NOISE

Since the effect of the scattering distribution has already been established, the sensitivity to additive white Gaussian noise (AWGN) is determined by using (12.48) to (12.52) with isotropic scattering. With AWGN the rms value of the received signal is $\tilde{R} = \sqrt{s^2 + 2b_0 + N_o B_w}$, and the values of $\tilde{b}_n$ and b_n in (12.48) are

$$
\begin{aligned}
\tilde{b}_0 &= b_0 + \frac{N_o B_w}{2} \\
\tilde{b}_2 &= 2b_0 \left(\pi f_m\right)^2 + \frac{N_o B_w^3 \pi^2}{6} \quad b_2 = 2b_0 \left(\pi f_m\right)^2
\end{aligned}
\tag{12.53}
$$

For the LCR velocity estimator (12.48) with (12.25) become, after considerable algebra,

$$
\begin{aligned}
\frac{\tilde{v}}{v} &= \left(1 + \frac{K+1}{6\gamma_S} \left(\frac{B_w}{f_m}\right)^2\right)^{\frac{1}{2}} \frac{\sqrt{\gamma_S(\gamma_S + 1)}}{\gamma_S + K + 1} \\
&\times \frac{I_0\left(\frac{2\sqrt{\gamma_S(\gamma_S+1)K(K+1)}}{\gamma_S+K+1}\right)}{I_0\left(2\sqrt{K(K+1)}\right)} \\
&\times \exp\left\{2K + 1 - \frac{\gamma_S(2K+1) + K + 1}{\gamma_S + K + 1}\right\},
\end{aligned}
\tag{12.54}
$$

where

$$
\gamma_S \triangleq \frac{s^2 + 2b_0}{N_o B_w} = \frac{2\sigma^2(K+1)}{N_o B_w}
\tag{12.55}
$$

is defined as the signal-to-noise ratio. Likewise, for the ZCR velocity estimator ZCR_1 (12.49) becomes

$$
\frac{\tilde{v}}{v} = \sqrt{\frac{\gamma_S + \left(\frac{B_w}{f_m}\right)^2 \frac{K+1}{6}}{\gamma_S + K + 1}}.
\tag{12.56}
$$

For Aulin's fading model in (12.12), the effect of AWGN on ZCR_2 can be obtained from (12.51) with $\tilde{L}_{ZCR_1}/L_{ZCR_1}$ in (12.56),

$$\tilde{\beta} = \frac{K}{2}\left(1 - 2\cos^2(\theta_0)\left[\frac{\gamma_S + K + 1}{\gamma_S + \frac{K+1}{6}\left(\frac{B_w}{f_m}\right)^2}\right]\right) \tag{12.57}$$

$$\tilde{\gamma} = \frac{K}{2}\left(1 + 2\cos^2(\theta_0)\left[\frac{\gamma_S + K + 1}{\gamma_S + \frac{K+1}{6}\left(\frac{B_w}{f_m}\right)^2}\right]\right) \tag{12.58}$$

$$\tilde{b} = 2\sqrt{K}\cos(\theta)\sqrt{\frac{\gamma_S + K + 1}{\gamma_S + \frac{K+1}{6}\left(\frac{B_w}{f_m}\right)^2}} \tag{12.59}$$

and a, b, γ, and β given by (12.32).

In [290], the effect of AWGN on the COV velocity estimator has been derived as a function of $\tau > 0$. Here we provide a closed form analytic result for the effect of AWGN on the COV velocity estimate for the limiting case when $\tau \to 0$. The limiting case is important for comparisons to AWGN effects on level crossing rate estimators, and since (12.42) is only valid for small τ. Consequently, the $\lim_{\tau \to 0} \tilde{v}/v$ in (12.52) is found, and afterwards, the effect of $\tau > 0$ in (12.52) is compared. It is shown in Appendix 10A that

$$\lim_{\tau \to 0} \tilde{v}/v = \frac{\sqrt{\zeta}}{\sqrt{\frac{(1+2K+K\cos(2\theta_0))}{(1+2K)}}} \tag{12.60}$$

where ζ is given by (12-12.A.4), with $a(0) = b_0, a'(0) = c(0) = c'(0) = c''(0) = 0$ and $a''(0) = 2b_0(\pi f_m)^2$ for isotropic scattering.

It is apparent from (12.54), and (12.56)–(12.60) that the effect of AWGN depends on K, B_w, γ_S, v and θ_0. For a practical system, the bandwidth B_w can be chosen as the maximum expected Doppler frequency over the range of velocities. However, a smaller B_w in reference to the actual maximum Doppler frequency f_m will result in velocity estimates that are less sensitive to noise. Therefore, a better approach is to use the velocity estimate $\hat{v}$ to continuously adjust B_w to be just greater than the current maximum Doppler frequency, i.e., $B_w \gtrsim \hat{v}/\lambda_c$. Fig. 12.11 shows the effect of AWGN on each of the velocity estimators with respect to K, γ_S, and v, assuming $\theta_0 = 0°$ (head-on LoS specular component). A value of $B_w = 357$ Hz is chosen which allows speeds up to 100 km/h at $f_c = 1.9$ GHz. For $K = 0$, AWGN has the same effect on all the velocity estimators. For larger velocities, e.g., 20 km/h, the bias becomes insignificant because B_w/f_m is small. However, for small velocities, e.g., 1 km/h, a very large B_w/f_m results in a significant bias. As mentioned above, this slow speed bias can be reduced by adapting the filter bandwidth B_w.
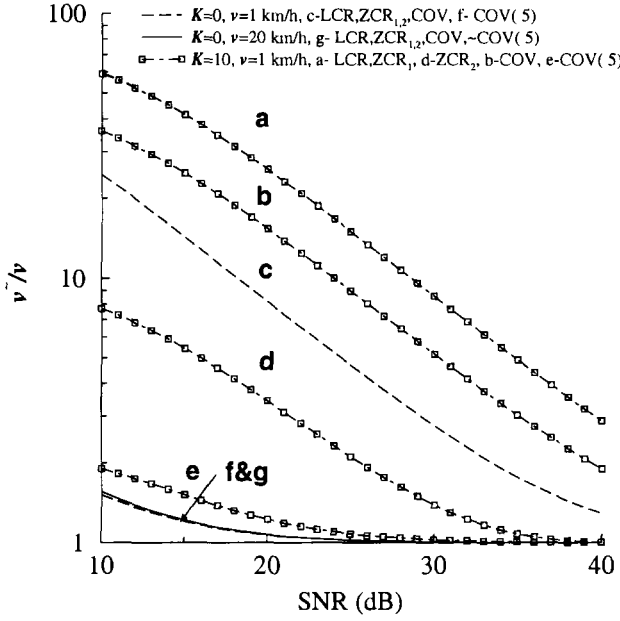
Figure 12.11. The effect of AWGN on the velocity estimates. $COV(.5) \Rightarrow 2\pi v \tau / \lambda_c = .5$, from [18].

It must also be remembered that Fig. 12.11 shows the worst case performance of the COV method as $\tau \to 0$. Any $\tau > 0$ will reduce the bias of the COV method due to AWGN. For example, if $2\pi v \tau / \lambda_c = 0.5$ in (12.52) then a large reduction in the effect of AWGN is realized, as shown by the curves labeled COV(.5) in Fig. 12.11. However, the accuracy of the COV velocity estimate itself improves with smaller τ, so that increasing τ for reduced noise sensitivity must be weighed against the reduced accuracy of the velocity estimate itself. This will be discussed further in the next section.

5.   VELOCITY ADAPTIVE HANDOFF ALGORITHMS

To study other velocity adaptive handoff issues we now assume $K = 0$, isotropic scattering, and no AWGN.

A velocity adaptive handoff algorithm must adapt the temporal window over which the mean signal strength estimates are taken by either keeping the sampling period constant and adjusting the number of samples per window, or vice versa. Here, we assume the latter. To reduce the variance in the velocity estimate, a sum of weighted past velocity estimates is performed using an exponential average of past estimates, i.e.,

$$\check{v}(n) = a\check{v}(n-1) + (1-a)\hat{v}(n) \qquad (12.61)$$

where a controls the weighting of past estimates used in the average, and $\hat{v}(n)$ is the current velocity estimate. The accuracy of the velocity estimates will be affected by the window length W_l used to obtain the velocity estimates (not to be confused with the window length over which the signal strengths are averaged), and the number of samples per wavelength N_λ.

To show the effect of parameters a, N_λ, and W_l, simulation of the NLoS handoff scenario shown in Figs. 12.1 and 12.4 was performed. The path loss was assumed to follow the two-slope model in Section 3. with $a = 2$, $b = 2$, and $g = 150$ m in (2.212). Drastic path loss at the corner was assumed to take effect 5 m into the corner, so that the MS moving from BS_0 to BS_1 would experience the corner effect at 255 m from BS_0. The corner effect was modeled by choosing the average received signal strength at 255 m as the initial signal strength in a new LoS path loss model with $a = 2$, $b = 2$ and $g = 150$ m as before. Correlated log-normal shadows were used having a standard deviation of $\sigma_\Omega = 6$ dB and D in (12.13) set so that shadows decorrelated to $0.1\sigma_\Omega^2$ at 30 m. The instantaneous signal strength samples were affected by Rayleigh fading using the Jakes' simulator presented in Section 2.3.2. Samples were taken of the log-envelope and appropriately converted to envelope or squared envelope samples for the velocity estimator under study. Two-branch antenna diversity was assumed, so that the $\hat{v}(n)$ in (12.61) represent the average estimate out of the diversity branches at position n.

As mentioned previously, Fig. 12.14 shows the performance of a temporal handoff algorithm with $H = 8$ dB, signal strength averaging over 2.27 s, and overlapping windows by $2.27/2 = 1.135$ s. Slightly better temporal handoff performance can probably be obtained by fine tuning these values. However, for purposes of studying the velocity adaptive algorithms it is sufficient to maintain $H = 8$ dB and adapt to some point near the knee of the performance curve. Consequently, the velocity estimators were designed to adapt to the 5 km/h operating point which corresponds to signal strength window averages over $\approx 20\lambda_c$ with a window overlap of $\approx 10\lambda_c$.

A total of 1000 runs were made from BS_0 to BS_1, and the 95% confidence intervals were calculated for i) the velocity at 100 m, ii) the corner at 255 m, and iii) the probability of being assigned to BS_0 at 255 m. This resulted in a 95% confidence interval spread of $\hat{v} \pm .5$ km/h and $\Pr(BS_0) \pm .025$. Likewise, the mean number of handoff values had a 95% confidence interval spread of approximately .05 (mean number of handoffs $\pm .025$).

5.1   EFFECT OF N_λ

To examine the effect of N_λ, assume that $a = .1$ and $W_l = 10\lambda_c$ for the LCR, ZCR and COV velocity estimators, and assume that the MS traverses the NLoS handoff route in Fig. 12.1 at 30 km/h. Furthermore, assume that the velocity estimators are initialized to 5 km/h, and that the MS is measuring signals from

BS_0 and BS_1 only. Fig. 12.12 shows the effect of N_λ on the velocity estimate, in the first 90 m of the call as the MS moves from BS_0 to BS_1, in terms of the response time and final velocity estimate. The LCR velocity estimator requires a higher sampling density than the COV or ZCR methods and its final velocity and response time to an incorrect startup value (5 km/h) improve dramatically when N_λ is increased from 10 to 30 samples/wavelength. For $N_\lambda = 30$ the COV method shows a slight overshoot in the initial convergence, a characteristic seen with all the velocity estimators as the sampling density is increased. It is interesting to note that for $N_\lambda = 10$ samples/wavelength $2\pi(v\tau)/\lambda_c = 2\pi 0.1\lambda_c/\lambda_c = .628$ and the final COV velocity estimate is close to the actual 30 km/h with a reasonable response time. This fact, along with the results of the Section 4.3.2 where $2\pi v\tau/\lambda_c = .5$ confirm that the effects of AWGN can be mitigated by using a larger sample spacing without drastically affecting the velocity estimate. We also note that the simulations used an estimate of the rms value R_{rms} in the LCR method and an estimate of the variance $\mu_{rr}(0)$ in the COV method. Thus the practicality of the velocity estimators that have been derived assuming perfect knowledge of these values is confirmed. Although not shown here, the Rice factor K was also found to have little effect thus confirming the claimed robustness of the estimators. Over the 1000 runs, the ZCR had the smallest velocity variance followed by the COV and LCR methods, respectively.

5.2   CORNER EFFECTS AND SENSITIVITY TO A AND W_L

The sharp downward spike at the corner (255 m) for the LCR velocity estimate in Fig. 12.13 is typical of the corner effects on the velocity estimators. The effect is caused by a sudden change in path loss which lowers the local mean estimate in the LCR method thus yielding fewer level crossings per second. This corner effect is apparent, although less acute in the ZCR and COV methods due to their quick adaptability. The LCR and ZCR methods may exhibit a drop in estimated velocity when the average signal strength changes abruptly. Although not shown here, the COV method has an upward bias with an abrupt increase in the average signal strength, and a downward bias when the opposite occurs. These corner effect properties could possibly be exploited to provide a combined corner detecting velocity adaptive handoff algorithm [17].

Larger values of a reduce the variance of the velocity estimate while sacrificing response time. Smaller values of a provide faster startup convergence and more sensitivity to corner effects.

Although a velocity window length W_l less than $20\lambda_c$ will increase the variance of the velocity estimates, it is beneficial for reducing the corner effect on the velocity estimator, as shown for the LCR method. Although not shown,
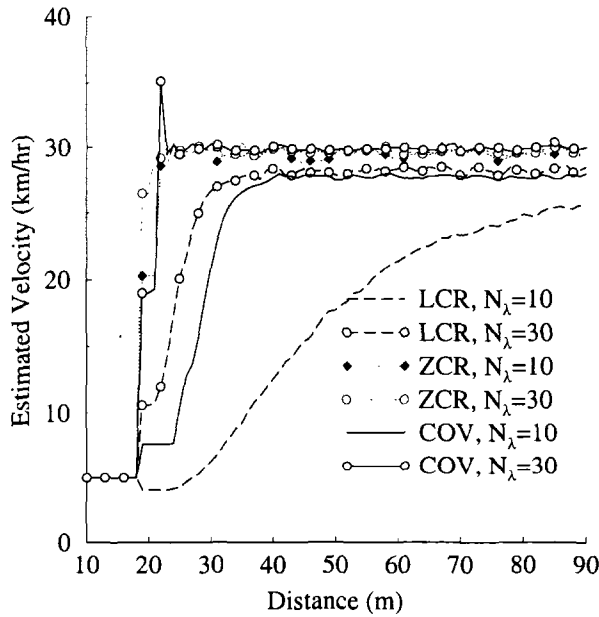
Figure 12.12. The effect of N_λ on the mean response time to a change in velocity. $a = .1$, $W_l = 10\lambda_c$, from [18].

the same is true for the ZCR and COV methods. The ZCR curve with $W_l = 20\lambda_c$ and $a = 0.5$ shows an overshoot in the initial convergence. This arises because the $W_l = 20\ \lambda_c$ windows that are used to obtain the velocity estimates overlap by $10\lambda_c$. Hence, part of the velocity estimate is derived from the previous window which may have a different sampling period due to adaptation. Note that we have used overlapped windows because they result in less handoff delay. Thus, it is probably better for initial startup to derive velocity estimates from the non-overlapped portions of the signal strength windows.

5.3   VELOCITY ADAPTIVE HANDOFF PERFORMANCE

Now that the effect of each parameter has been determined, the performance of the velocity adapted handoff algorithm is shown by the various symbols in Fig. 12.14 for a MS traveling at 30 km/h. The estimators, were selected to adapt to the 5 km/h operating point, the algorithm parameters were chosen as a $= .1$, $W_l = 10\lambda_c$ with an initial startup velocity of 5 km/h. The mean number of handoffs were found to have a 95% confidence interval with a span of about 0.05 (mean number of handoffs ± 0.025) about the mean that is plotted. The velocity adaptive handoff algorithm performs very well by maintaining the desired operating point near the 5 km/h point.
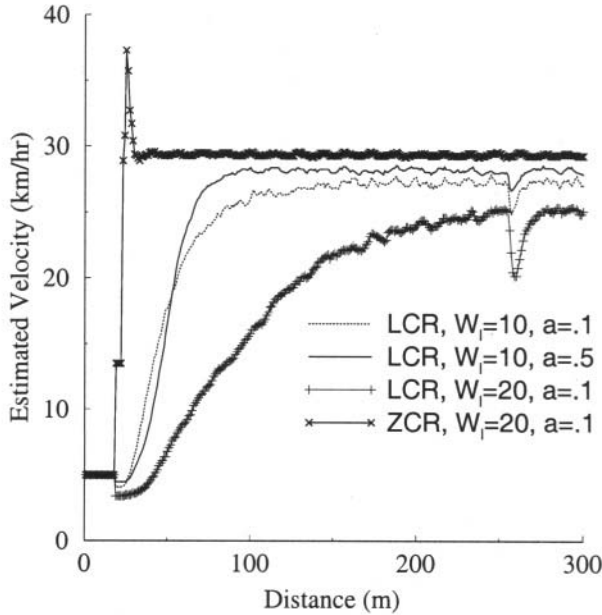
Figure 12.13. The effect of a and W_l on the mean response and corner effects, from [18].

6.   HARD HANDOFF ANALYSIS

The classical signal strength based hard handoff algorithm compares signal strength averages measured over a time interval T (seconds), and executes a hand-off if the average signal strength of the target BS is at least H (dB) larger than that of the serving BS [150], [18], [339], [338]. The analytical computation of the handoff characteristics for this classical signal strength based handoff algorithm is generally intractable. However, for the case when the average signal strength decays smoothly along a handoff route and the handoff hysteresis H is not too small compared to the shadow standard deviation, Vijayan and Holtzman [339], [338] have developed an analytical method to characterize the performance of the classical signal strength based handoff algorithm. They have also extended their results to include handoff algorithms that use absolute measurements [379], similar to the one in (12.4).

Consider the case of a MS moving at a constant velocity along a straight line between two BSs, BS_0 and BS_1, that are separated by a distance of D meters. We neglect envelope fading under the assumption that the received signal strength estimates are averaged by using a window with an appropriate length as explained in Section 3.. In any case, however, the signal strength estimates will respond to path loss and shadowing variations. Considering the effects of path loss and shadowing, the signal levels $\Omega_{0\ (dB)}(d)$ and $\Omega_{1\ (dB)}(d)$
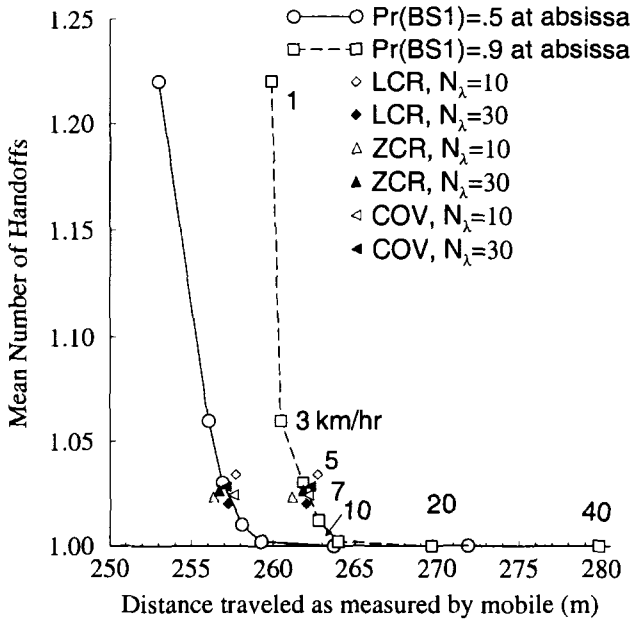
Figure 12.14. Handoff performance of a 2.27 second averaging handoff algorithm in comparison with a velocity adaptive handoff algorithm using the LCR, ZCR or covariance method for velocity control. $H = 8$ dB, $\sigma_\Omega = 6$ dB, from [18].

that are received from BS_0 and BS_1, respectively, are (1.7)

$$\Omega_{0\,(dB)}(d) = \Omega_{(dB)}(d_0) - 10\beta\log_{10}(d/d_0) + \epsilon_{0\,(dB)} \qquad (12.62)$$

$$\Omega_{1\,(dB)}(d) = \Omega_{(dB)}(d_0) - 10\beta\log_{10}((D-d)/d_0) + \epsilon_{1\,(dB)} \qquad (12.63)$$

where d is the distance between BS_0 and the MS. The parameters $\epsilon_{0\,(dB)}$ and $\epsilon_{1\,(dB)}$ are independent zero-mean Gaussian random processes with variance σ_Ω^2, reflecting a log-normal shadowing model. The signal strength measurements are assumed to be averaged by using an exponential averaging window with parameter d_{av} so that the averaged signal levels from the two BSs are, respectively,

$$\overline{\Omega}_{0\,(dB)}(d) = \frac{1}{d_{av}} \int_0^d \Omega_{0\,(dB)}(d-x)e^{-x/d_{av}}\,dx \qquad (12.64)$$

$$\overline{\Omega}_{1\,(dB)}(d) = \frac{1}{d_{av}} \int_0^d \Omega_{1\,(dB)}(d-x)e^{-x/d_{av}}\,dx \;. \qquad (12.65)$$

To describe the signal strength based handoff algorithm, let

$$x(d) = \overline{\Omega}_{0\,(dB)}(d) - \overline{\Omega}_{1\,(dB)}(d) \qquad (12.66)$$
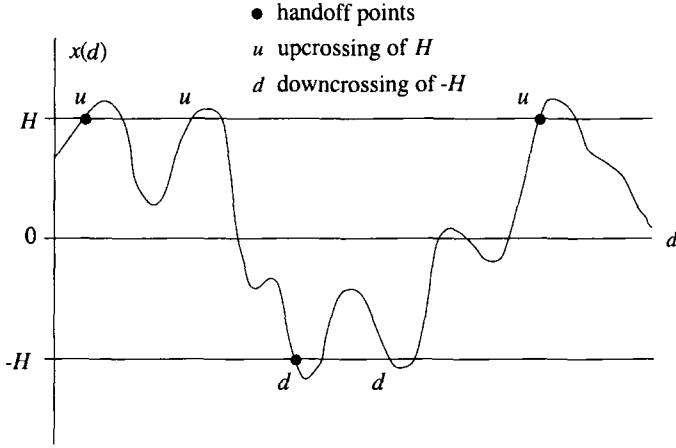
Figure 12.15. Handoff initiation points with their associated hysteresis level crossings. The MS is moving from BS_0 to BS_1 and assumed to be communicating with BS_1 at the beginning of the interval shown, from [339].

denote the difference between the averaged signal strength estimates for BS_0 and BS_1. Consider the crossings of $x(d)$ with respect to the hysteresis levels $\pm H$ (dB) as illustrated in Fig. 12.15. A handoff is triggered if $x(d)$ has a down-crossing at $-H$ (dB) given that the last level crossing was an up-crossing at H (dB), or if $x(d)$ has an up-crossing at H (dB) given that the last level crossing was a down-crossing at $-H$ (dB). Vijayan and Holtzman verified that the two point processes, up-crossings of H (dB) and down-crossings of $-H$ (dB), can be modeled as independent Poisson processes under the assumption that $x(d)$ is a stationary *zero-mean* Gaussian random process, i.e., changes in the mean are ignored and the MS is moving along the boundary between two cells [339], This result also applies when $x(d)$ has non-zero mean, but in this case the up-crossing and down-crossing rates are not equal. The Poisson assumption is asymptotically true for large H, but has been shown to hold true for H values of practical interest, i.e., those on the order of the shadow standard deviation σ_Ω [339].

The handoff analysis proceeds by dividing up a handoff route into small spatial intervals of length d_s, such that only one level crossing is likely to occur within each interval. The probability of handoff at distance $d = nd_s$ is [339]

$$p_{ho}(n) = p_d(n)p_{lu}(n) + p_u(n)(1 - p_{lu}(n)) \qquad (12.67)$$

where $p_u(n)$ and $p_d(n)$ is the probability of an up-crossing or down-crossing in the nth interval, and $p_{lu}(n)$ is the probability that the last event was an up-crossing. In other words, the MS was assigned to BS_0 at the beginning of the nth interval. This can happen in one of two mutually exclusive ways; i) there is

an up-crossing but no down-crossing in the $(n - 1)$th interval, and ii) there are no crossings in the the $(n - 1)$th interval, and the last event before the $(n - 1)$th interval was an up-crossing. By assuming $p_{lu}(1) = 1$, the following recursive equation for $p_{lu}(n)$ can be derived as a function of $p_u(n - 1)$, $p_d(n - 1)$ and $p_{lu}(n - 1)$ [339]

$$p_{lu}(n) = p_u(n-1)(1-p_d(n-1))+(1-p_u(n-1))(1-p_d(n-1))p_{lu}(n-1) \ . \tag{12.68}$$

As detailed in [339], the probabilities $p_d(n)$ and $p_u(n)$ are functions of the mean $\mu_x(d)$, variance $\sigma_x^2(d)$, and variance of the derivative of $x(d)$, $\sigma_{\dot{x}}^2(d)$. These in turn are functions of the statistics of $\Omega_{0\,(\mathrm{dB})}(d)$ and $\Omega_{1\,(\mathrm{dB})}(d)$, which depend on the path loss and shadowing. Austin and Stüber have shown how these statistics depend on the co-channel interference [22]. We will first evaluate the statistics of $\Omega_{0\,(\mathrm{dB})}(d)$ and $\Omega_{1\,(\mathrm{dB})}(d)$ and afterwards derive the appropriate expressions for $p_d(n)$ and $p_u(n)$.

As discussed in Chapter 3, co-channel interference is usually assumed to add on a power basis [295], [264]. Hence, in the presence of N co-channel interferers the signals received from BS_0 and BS_1 are, respectively,

$$\Omega_{0\,(\mathrm{dB})}(d) \;=\; 10 \log_{10} \left(\sum_{k=0}^{N} \Omega_{0,k(\mathrm{dB})}(d) \right) \tag{12.69}$$

$$\Omega_{1\,(\mathrm{dB})}(d) \;=\; 10 \log_{10} \left(\sum_{k=0}^{N} \Omega_{1,k(\mathrm{dB})}(d) \right) \tag{12.70}$$

where $\Omega_{0,0(\mathrm{dB})}(d)$ and $\Omega_{1,0(\mathrm{dB})}(d)$ are the power of the desired signals from BS_0 and BS_1, respectively, and $\Omega_{0,k(\mathrm{dB})}(d)$ and $\Omega_{1,k(\mathrm{dB})}(d)$ $k = 1, \ldots, N_I$ are the powers of the interfering co-channel signals received at the same BSs. Once again, the $\Omega_{0,k(\mathrm{dB})}(d)$ and $\Omega_{1,k(\mathrm{dB})}(d)$ are log-normally distributed. As discussed in Chapter 3.1, the sum of log-normal random variables can be approximated by another log-normal random variable and, hence, $\Omega_{0\,(\mathrm{dB})}(d)$ and $\Omega_{1\,(\mathrm{dB})}(d)$ remain Gaussian. Here we consider the approximations suggested by Fenton [295], [264], and Schwartz and Yeh [295].

Following the notation in Chapter 3.1, define $\hat{\Omega} = \xi\Omega_{(\mathrm{dB})}$, where $\xi = (\mathrm{In}\ 10)/10 = 0.23026$. If the N_I interferers for BS_0 have means $\mu_{\hat{\Omega}_{0,k}}(d)$ and variance $\sigma_{\hat{\Omega}}^2$, then the mean and variance of $\hat{\Omega}_0(d)$I using the Fenton-Wilkinson approach are

$$\mu_{\hat{\Omega}_0}(d) \;=\; \frac{\sigma_{\hat{\Omega}}^2 - \sigma_{\hat{\Omega}_0}^2(d)}{2} + \ln \left(\sum_{k=0}^{N_I} e^{\mu_{\hat{\Omega}_{0,k}}(d)} \right) \tag{12.71}$$

$$\sigma_{\hat{\Omega}_0}^2(d) = \ln\left((e^{\sigma_{\hat{\Omega}}^2} - 1) \frac{\sum_{k=0}^{N_I} e^{2\mu_{\hat{\Omega}_{0,k}}(d)}}{\left(\sum_{k=0}^{N_I} e^{\mu_{\hat{\Omega}_{0,k}}(d)}\right)^2} + 1 \right) \quad (12.72)$$

where the conversion of $\mu_{\hat{\Omega}_0}(d)$ and $\sigma_{\hat{\Omega}_0}^2(d)$ to units of decibels is $\mu_{\Omega_0}(d) = \xi^{-1}\mu_{\hat{\Omega}_0}(d)$, and $\sigma_{\Omega_0}^2(d) = \xi^{-2}\sigma_{\hat{\Omega}_0}^2(d)$, respectively. Schwartz and Yeh's approach is an recursive technique that combines only two log-normal variates at a time. For example, combining $\hat{\Omega}_{0,0}(d)$ and $\hat{\Omega}_{0,1}(d)$ gives the intermediate result

$$\mu_{\hat{\Omega}_0}(d) = \mu_{\hat{\Omega}_{0,0}}(d) + G_1 \quad (12.73)$$
$$\sigma_{\hat{\Omega}_0}^2(d) = \sigma_{\hat{\Omega}}^2 - G_1^2 - 2\sigma_{\hat{\Omega}}^2 G_3 + G_2 \quad (12.74)$$

where G_1, G_2, and G_3 are defined by (3.20), (3.23), and (3.24), respectively. The final values of $\mu_{\hat{\Omega}_0}(d)$ and $\sigma_{\hat{\Omega}_0}^2(d)$ are obtained by recursion.

By using either approach the mean $\mu_x(d)$ can be determined. Since $x(d)$ is modeled as a Gaussian random process, the probabilities $p_d(n)$ and $p_u(n)$ can be computed by using the same procedure used to determine the envelope level crossing rates in Section 2.1.4. In particular,

$$p_u(n) = d_s \int_0^\infty \dot{x} p(H, \dot{x}) d\dot{x}$$
$$p_d(n) = d_s \int_{-\infty}^0 |\dot{x}| p(H, \dot{x}) d\dot{x} \quad (12.75)$$

where $p(H, \dot{x})$ is the joint density function of $x(kd_s)$ and its derivative $\dot{x}(kd_s)$. Since $x(kd_s)$ and $\dot{x}(kd_s)$ are independent Gaussian random variables

$$p_u(kd_s) = \frac{d_s}{\sqrt{2\pi b_0}} \exp\left\{ -\frac{(H - \mu_x(kd_s))^2}{2b_0} \right\} \quad (12.76)$$

$$\times \left[\mu_{\dot{x}}(kd_s) Q\left(-\frac{\mu_{\dot{x}}(kd_s)}{\sqrt{b_2}} \right) + \sqrt{\frac{b_2}{2\pi}} \exp\left\{ \frac{-\mu_{\dot{x}}^2(kd_s)}{2b_2} \right\} \right]$$

where, from (2.79)

$$b_0 = \sigma_x^2(d) = 2 \int_0^\infty \hat{S}_{xx}(f) df \quad (12.77)$$

$$b_2 = \sigma_{\dot{x}}^2(d) = 2(2\pi)^2 \int_0^\infty f^2 \hat{S}_{xx}(f) df \quad (12.78)$$

and $\hat{S}_{xx}(f)$ is the power spectrum of $x(d)$ that includes the effect of co-channel interference. Likewise,

$$p_d(kd_s) = \frac{d_s}{\sqrt{2\pi b_0}}\exp\left\{-\frac{(H+\mu_x(kd_s))^2}{2b_0}\right\} \tag{12.79}$$

$$\times \quad \left[-\mu_{\dot{x}}(kd_s)Q\left(\frac{\mu_{\dot{x}}(kd_s)}{\sqrt{b_2}}\right) + \sqrt{\frac{b_2}{2\pi}}\exp\left\{-\frac{\mu_{\dot{x}}^2(kd_s)}{2b_2}\right\}\right]$$

The autocovariance of $\Omega_{0\ (\mathrm{dB})}(d)$ or $\Omega_{1\ (\mathrm{dB})}(d)$ (equal to the shadow auto-correlation) *without* co-channel interference is modeled by

$$\mu_{\Omega_{(\mathrm{dB})}\Omega_{(\mathrm{dB})}}(d) = \sigma_{\Omega}^2\exp(-|d_1-d_2|/d_0) \tag{12.80}$$

where $d = d_1 - d_2$, and d_0 controls the decorrelation with distance. Let $\tilde{\mu}_{\Omega_{(\mathrm{dB})}\Omega_{(\mathrm{dB})}}(d)$ denote the same function when co-channel interference is present. The value $\mu_{\Omega_{(\mathrm{dB})}\Omega_{(\mathrm{dB})}}(0)$ can be accurately approximated by using either (12.72) or (12.74)). An approximation of $\hat{\phi}_{\Omega_{(\mathrm{dB})}\Omega_{(\mathrm{dB})}}(d)$ for $d > 0$ can be obtained by substituting σ_{Ω}^2 in (12.80) with the value obtained in (12.72) or (12.74). The accuracy of this approximation was tested through the simulation of mutually uncorrelated log-normal interferers, each having the shadow auto-covariance in (12.80) with $\sigma_{\Omega} = 6$ dB and $d_0 = 20$ m. Fig. 12.16 shows the results and verifies that the proposed approximation of $\tilde{\mu}_{\Omega_{(\mathrm{dB})}\Omega_{(\mathrm{dB})}}(d)$ is fairly accurate. Also, very accurate modeling of $\tilde{\mu}_{\Omega_{(\mathrm{dB})}\Omega_{(\mathrm{dB})}}(d)$ is not essential in handoff analysis [338].

Using the above approximation gives

$$\hat{S}_{xx}(f) = \frac{2(\sigma_{\Omega_0}^2(d) + \sigma_{\Omega_1}^2(d))d_0}{(1+d_0^2(2\pi f)^2)(1+d_{\mathrm{av}}^2(2\pi f)^2)} \tag{12.81}$$

so that

$$\sigma_x^2(d) = \frac{(\sigma_{\Omega_0}^2(d) + \sigma_{\Omega_1}^2(d))d_0}{d_0 + d_{\mathrm{av}}}, \quad \sigma_{\dot{x}}^2(d) = \frac{\sigma_x^2(d)}{d_{\mathrm{av}}\,d_0}\ . \tag{12.82}$$

6.1   SIMULATION RESULTS

Consider a MS traversing from BS_0 to BS_1 separated by 1000 m with two co-channel interferers as shown in Fig. 12.17. Assume a square-law path loss with distance (used here to accentuate the co-channel interference effects), $d_{\mathrm{av}} = 10$ m, $d_0 = 20$ m, and choose $\sigma_{\Omega} = 4$ dB so that both the Fenton and Schwartz and Yen log-normal approximations are accurate. Fig. 12.18 compares analytical and simulation results for the handoff probabilities in the presence and absence of co-channel interference. Note that the presence of co-channel interference actually lowers the probability of handoff. Schwartz and
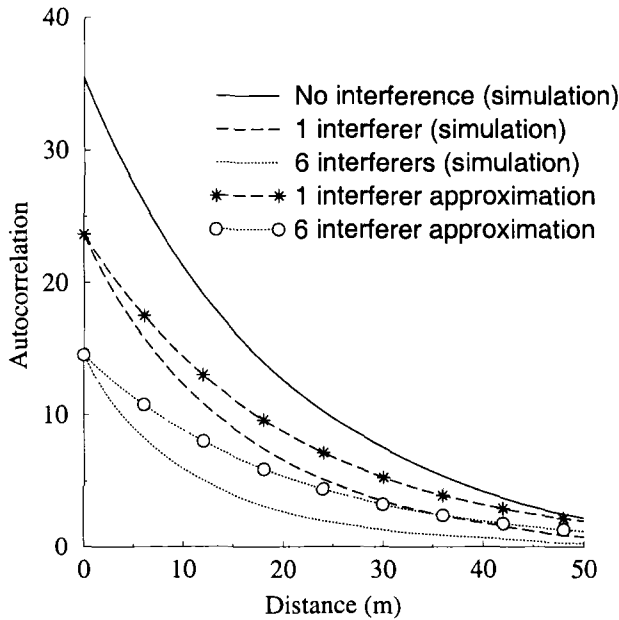
Figure 12.16. Shadow autocorrelation with and without co-channel interference, from [22].
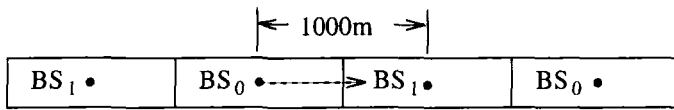
Figure 12.17. Base station layout, MS route (dotted line), and location of co-channel interferers.

Yeh's method leads to an accurate prediction of the handoff probabilities while Fenton's method does not lead to as much accuracy. Finally, the accuracy of the prediction of handoff probabilities leads us to conclude that the assumptions made for $\tilde{\phi}_{\Omega_{(dB)}\Omega_{(dB)}}(d)$ were reasonable.

7. SOFT HANDOFF ANALYSIS

Zhang and Holtzman have extended their method in [339], [338] to analyze soft handoff in CDMA systems [380]. This section present the soft handoff analysis in [380] with a small modification. Some simplifying assumptions are made in [380], and we change one of the assumptions to present a soft handoff analysis in a different perspective.

Once again, the system model consists of two BSs, BS_0 and BS_1, separated by a distance of D meters. Considering the effects of path loss and shadowing, the signal levels $\Omega_{0\ (dB)}(d)$ and $\Omega_{1\ (dB)}(d)$ that are received from BS_0 and BS_1,
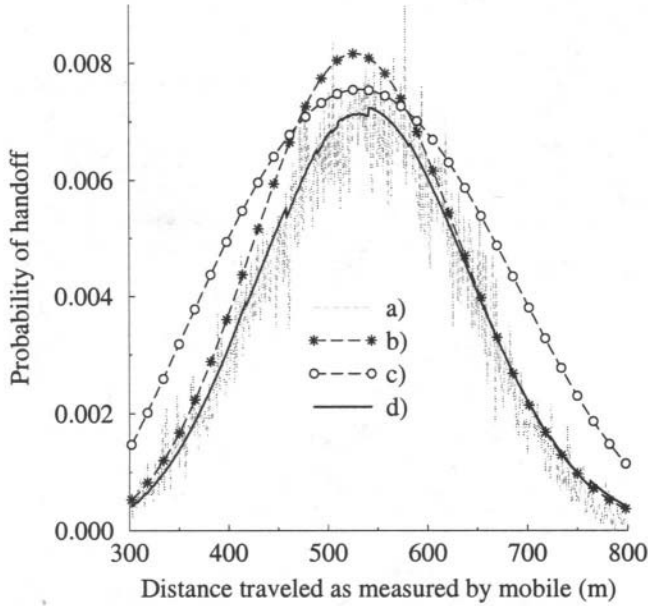
Figure 12.18. Simulation vs. analytical model performance. a) Simulation of LoS handoff with co-channel interference, b) handoff analysis model in the absence of co-channel interference, c) handoff analysis model including co-channel interference and using the Fenton-Wilkinson log-normal approximation, d) handoff analysis model including co-channel interference and using the Schwartz and Yeh log-normal approximation, from [22].

are given by (12.62) and (12.63), respectively. Shadow correlations are again described by (12.80), where σ_Ω is the shadow standard deviation, $d = d_1 - d_2$, and d_0 controls the shadow decorrelation with distance.

Zhang and Holtzman make some simplifying assumptions i) the use of pilot strength in handoff decisions rather than E_c/I_o, and ii) the use of relative thresholds instead of absolute thresholds. Here we modify the analysis to incorporate the absolute thresholds which give us another set of results. A more realistic analysis can include both absolute and relative thresholds.

The MS moves at a constant speed v, and the pilot signal strengths are sampled at every T seconds. The MS location changes by $d_s = vT$ during every sampling interval. Let $\Omega_{0\,(\text{dB})}(n) = \Omega_{0\,(\text{dB})}(nd_s)$ and $\Omega_{1\,(\text{dB})}(n) = \Omega_{1\,(\text{dB})}(nd_s)$, where $n = 1, 2, \cdots, D/d_s$. If BS_0 is not in the active set at epoch $n - 1$, then the probability BS_0 will be in the active set at epoch n is

$$P_{N \to \text{BS}_0}[n] = P[\Omega_{0\,(\text{dB})}(n) > T_{\text{add}} | \Omega_{0\,(\text{dB})}(n - 1) < T_{\text{add}}] \ . \qquad (12.83)$$

Similarly, the probability BS_0 does not join the active set at epoch n given BS_0 is not in the active set at epoch $n - 1$ is

$$\begin{aligned} \mathrm{P}_{N \to N}[n] &= \mathrm{P}[\Omega_{0 \text{ (dB)}}(n) < T_{\text{add}} | \Omega_{0 \text{ (dB)}}(n-1) < T_{\text{add}}] \\ &= 1 - \mathrm{P}_{N \to \mathrm{BS}_0}[n]. \end{aligned} \quad (12.84)$$

A BS will be dropped from the active set if its pilot strength drops below T_{drop} for consecutive M samples. Therefore, the probability that BS_0 is dropped from the active set at epoch n given that it is in the active set at epoch $n - 1$ is

$$\begin{aligned} \mathrm{P}_{\mathrm{BS}_0 \to N}[n] &= \mathrm{P}[\Omega_{0 \text{ (dB)}}(n-M) < T_{\text{drop}} | \Omega_{0 \text{ (dB)}}(n-M-1) > T_{\text{drop}}] \\ &\quad \prod_{k=n-M+1}^{n} \mathrm{P}[\Omega_{0 \text{ (dB)}}(k) < T_{\text{drop}} | \Omega_{0 \text{ (dB)}}(k-1) < T_{\text{drop}}] \\ \mathrm{P}_{\mathrm{BS}_0 \to \mathrm{BS}_0}[n] &= 1 - \mathrm{P}_{\mathrm{BS}_0 \to N}[n]. \end{aligned} \quad (12.85)$$

Similar probabilities associated with BS_1 can be computed by simply replacing $\Omega_{0 \text{ (dB)}}(n)$ with $\Omega_{1 \text{ (dB)}}(n)$. Then,

$$\begin{aligned} \mathrm{P}_{\mathrm{BS}_0}[n] &= \mathrm{P}_{\mathrm{BS}_0}[n-1]\mathrm{P}_{\mathrm{BS}_0 \to \mathrm{BS}_0}[n] + (1 - \mathrm{P}_{\mathrm{BS}_0}[n-1])\mathrm{P}_{N \to \mathrm{BS}_0}[n] \\ \mathrm{P}_{\mathrm{BS}_1}[n] &= \mathrm{P}_{\mathrm{BS}_1}[n-1]\mathrm{P}_{\mathrm{BS}_1 \to \mathrm{BS}_1}[n] + (1 - \mathrm{P}_{\mathrm{BS}_1}[n-1])\mathrm{P}_{N \to \mathrm{BS}_1}[n] \end{aligned} \quad (12.86)$$

where $\mathrm{P}_{\mathrm{BS}_0}[n]$ and $\mathrm{P}_{\mathrm{BS}_1}[n]$ are the probabilities that the active set contains BS_0 and BS_1 at epoch n, respectively. Since the MS starts from BS_0 to BS_1, we give the following initial condition:

$$\mathrm{P}_{\mathrm{BS}_)}[0] = 1, \;]\mathrm{P}_{\mathrm{BS}_1}[D/d_s] = 1 \;. \quad (12.87)$$

The MS's active connection will be dropped if the active set does not contain any BS. Therefore, the outage probability is simply

$$\mathrm{P}_{\text{out}}[n] = (1 - \mathrm{P}_{\mathrm{BS}_0}[n])(1 - \mathrm{P}_{\mathrm{BS}_1}[n]) \;. \quad (12.88)$$

Another performance indicator is the expected number of BSs in the active set, N_{BS}:

$$N_{BS} = \frac{1}{D/d_s} \sum_{k=1}^{D/d_s} (\mathrm{P}_{\mathrm{BS}_0}[k] + \mathrm{P}_{\mathrm{BS}_1}[k]) \;. \quad (12.89)$$

As N_{BS} increases, the network overhead that is required to manage the soft handoffs will also increase.

7.1   SIMULATION RESULTS

The derived conditional probabilities can easily be computed using the jointly Gaussian pdf with the shadow correlation in (12.80). Then, $\mathrm{P}_{\mathrm{BS}_0}[n]$ and $\mathrm{P}_{\mathrm{BS}_1}[n]$ are computed iteratively. The parameters used in our simulation are as follows:

- path loss exponent $\beta = 4$

- shadow standard deviation $\sigma_\Omega = 8$ dB

- shadow decorrelation distance $d_0 = 20$ m

- sampling distance $d_s = 2$ m

- drop threshold timer $M = 5$.

Fig. 12.19 shows the effect of different handoff parameter settings on the probability of each BS belonging to the active set as a function of the MS location. Relaxing the handoff parameters (lower thresholds) increases the assignment probabilities. Lowering T_{add}, has the effect of admitting BSs to the active set more easily while lowering T_{drop} causes BSs to stay in the active set longer. Therefore, relaxing the soft handoff parameters allows more MSs be in soft handoff mode.

Fig. 12.20 shows the outage probability as a function of distance. It is assumed that an outage occurs when the active set does not contain any BS. Observe that the outage probability decreases as the handoff parameters are relaxed. It is important to understand that this result does not account for other outage possibilities such as the failure to meet the C/I requirement. It does, however, suggest that relaxing the handoff parameters benefits the system performance by preventing unnecessary dropped calls due to unoptimized parameter settings. But, as previously discussed, soft handoff is an expensive proposition to the system. As shown in Table 12.2, relaxing handoff parameters increases the average number of BSs in the active set. In turn the system requires more overhead resource to manage soft handoff. Therefore, the challenge is to optimize the soft handoff parameters setting to maximize the system performance while minimizing the system resource usage.

T_{add}	T_{drop}	N_{BS}
-116 dB	-123 dB	1.23
-120 dB	-123 dB	1.35
-120 dB	-126 dB	1.45

Table 12.2. The expected number of BSs in the active set

Finally, CDMA systems such as IS-95 use both absolute and relative thresholds to manage soft handoff. T_{add} is in absolute scale while T_{drop} is in relative scale. An absolute T_{add} ensures that every BS that is able to contribute to the diversity gain is included in the active set with high probability. The effect
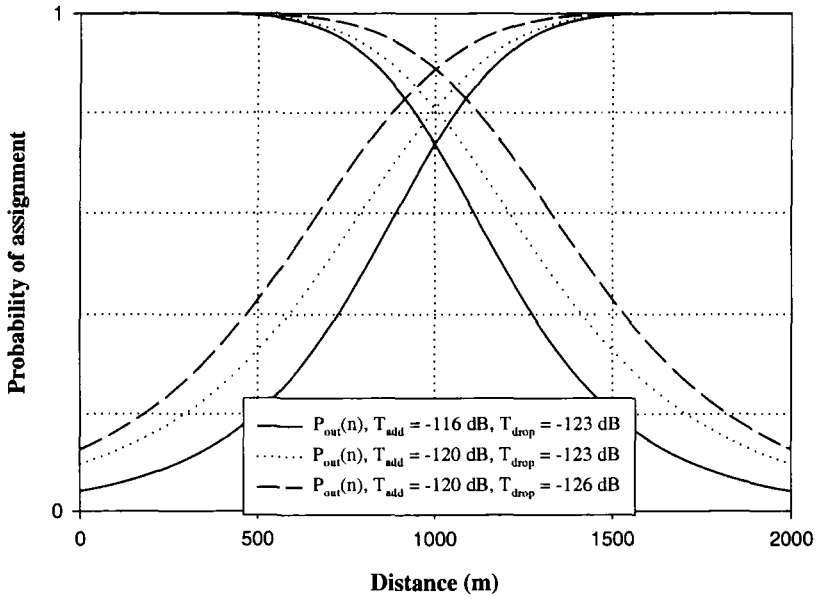
Figure 12.19. $P_{BS_0}(n)$ and $P_{B_1}(n)$ against the MS location.

of a relative T_{drop} is to drop a BS from the active set only when its link has deteriorated far below the best link. The analysis presented here assumes that both thresholds are absolute, while the analysis in [380] uses relative thresholds only. Although a more accurate analysis can be obtained by incorporating the two analysis together, the basic findings and observations are not expected to change significantly.

8. CIR-BASED LINK QUALITY MEASUREMENTS

Cellular radio resource allocation algorithms have been developed for hand-offs [121], dynamic channel assignment [246], [145], and power control [14], [15], under the assumption that the MSs and/or BSs have access to real time measurements of the received carrier-to-interference plus noise ratio C/(I+N). However, very little literature has appeared on methods for measuring C/(I+N). Kozono [182] suggested a method for measuring co-channel interference (CCI) in AMPS, by separating two terms at different frequencies which are known functions of the signal and interference. Yoshida [315] suggested a method for in-service monitoring of multipath delay spread and CCI for a QPSK sig-
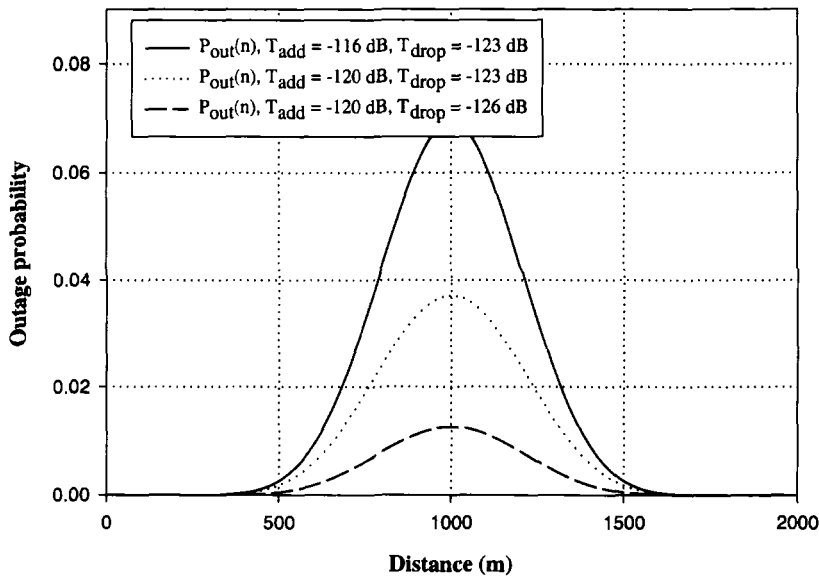
Figure 12.20. Outage probability vs. the location of MT

nal. He reported that the CCI can be monitored provided that the delay spread is negligible compared to the symbol duration. Sollenberger [307] used the eye-opening as a measure of signal quality.

In this section we present a technique for estimating (S+I+N) and S/(I+N) that could be used in signal quality based resource allocation algorithms in TDMA cellular systems [19], [21]. Section 8.1 presents the discrete channel model. Estimation methods for the received (I+N) and C/(I+N) are then derived in Section 8.1 whose accuracy is only a function of the symbol error statistics. These estimators are evaluated by software simulation for an IS-54 frame structure in Section 8.2.

8.1   DISCRETE-TIME MODEL FOR SIGNAL QUALITY ESTIMATION

As shown in Section 6.3.1, the overall channel consisting of the transmit filter, channel, matched filter, sampler (and noise whitening filter) can be modeled by a

T-spaced, $L + 1$-tap, transversal filter[4]. The overall discrete-time is described by the channel vector $\mathbf{g} = [g_0, g_1, \ldots, g_L]^T$, where T denotes transposition. Let $\mathbf{v} = [v_0, \ldots, v_M]^T$ denote the received signal vector consisting of M samples, where $v_k = \sum_{i=0}^{L} g_i x_{k-i}$. Assuming that the channel does not change significantly over a block of $L + M + 1$ symbols, the received vector $\mathbf{v}$ can be written as

$$\mathbf{v} = \mathbf{X}\mathbf{g} + \mathbf{w} \tag{12.90}$$

where $\mathbf{X}$ is an $(M+1) \times (L+1)$ Toeplitz matrix consisting of the transmitted symbols of the form

$$\mathbf{X} = [x_{i,j}] = \begin{bmatrix} x_0 & x_{-1} & \cdots & x_{-L} \\ x_1 & x_0 & \cdots & x_{1-L} \\ \vdots & \vdots & \cdots & \vdots \\ x_M & x_{M-1} & \cdots & x_{M-L} \end{bmatrix} \tag{12.91}$$

and $\mathbf{w} = [w_0, \ldots, w_M]^T$ is a vector consisting of the samples of the received interference plus noise.

8.1.1   ESTIMATION OF (I+N)

An (I+N) or C/(I+N) estimator requires a method for separating $\mathbf{f}$ and $\mathbf{w}$ from the observation of $\mathbf{y}$. Consider the situation where $L > M$, so that $\mathbf{A}$ has more rows than columns. Then, there exists a vector $\mathbf{c} = [c_0, \ldots, c_M]^T$ in the *null space* of $\mathbf{X}$ such that $\mathbf{c}^T\mathbf{X} = \mathbf{0}$. If $\mathbf{X}$ is known, then $\mathbf{c}$ can be easily determined. Then

$$\mathbf{c}^T\mathbf{y} = \mathbf{0} + \mathbf{c}^T\mathbf{w} \tag{12.92}$$

and, therefore, $\mathbf{g}$ and $\mathbf{w}$ are completely separated from the observation $\mathbf{y}$. However, with the exception of the training and perhaps the color code sequences, $\mathbf{X}$ is not known exactly because the data symbols comprising $\mathbf{X}$ must be obtained from decisions. Therefore the matrix of decisions $\hat{\mathbf{X}}$ must be used instead, where $\hat{\mathbf{A}} = \mathbf{A} + \boldsymbol{\Delta}$ and $\boldsymbol{\Delta} = [\delta_{i,j}]$ is the matrix of symbol errors. Nevertheless, a vector $\hat{\mathbf{c}}$ can still be found in the *null space* of $\hat{\mathbf{X}}$ so that

$$\hat{\mathbf{c}}^T\mathbf{y} = \hat{\mathbf{c}}^T\mathbf{X}\mathbf{g} + \hat{\mathbf{c}}^T\mathbf{w} \; . \tag{12.93}$$

Hence, an (I+N) estimate can be obtained from

---

[4]If rate 2/T sampling is used, then the overall channel is a t/2-spaced, $2L + 1$-tap, transversal filter.

$$
\begin{aligned}
\hat{\sigma}_{I+N}^2 &= \mathrm{E}\left[\frac{\mathbf{y}^H \hat{\mathbf{c}}^* \hat{\mathbf{c}}^T \mathbf{y}}{\|\hat{\mathbf{c}}\|^2}\right] \\
&= \mathrm{E}\left[\frac{\mathbf{g}^H \boldsymbol{\Delta}^H \hat{\mathbf{c}}^* \hat{\mathbf{c}}^T \boldsymbol{\Delta}\mathbf{g}}{\|\hat{\mathbf{c}}\|^2}\right] + \mathrm{E}\left[\frac{\mathbf{w}^H \hat{\mathbf{c}}^* \hat{\mathbf{c}}^T \mathbf{w}}{\|\hat{\mathbf{c}}\|^2}\right] \\
&= \left(\sum_{i=0}^{L}\sum_{j=0}^{M}\sum_{k=0}^{L}\sum_{\ell=0}^{M} \mathrm{E}\left[\frac{\delta_{i,j}^* f_j^* \hat{c}_i^* \delta_{k,\ell} f_\ell \hat{c}_k}{\|\hat{\mathbf{c}}\|^2}\right]\right. \\
&\quad \left. + \sum_{i=0}^{M}\sum_{j=0}^{M} \mathrm{E}\left[\frac{\hat{c}_i \hat{c}_j^*}{\|\hat{\mathbf{c}}\|^2}\right] \mathrm{E}[w_i w_j^*]\right)
\end{aligned}
\tag{12.94}
$$

where H is the Hermitian transpose, and where the second equality is obtained by using $\hat{\mathbf{A}} = \mathbf{A} + \boldsymbol{\Delta}$ along with the reasonable assumption that $\mathbf{w}$ has zero mean and is uncorrelated with $\mathbf{A}$ and $\mathbf{f}$. It is also reasonable to assume that the symbol errors are independent with a constant variance, i.e.,

$$
\mathrm{E}[|\delta_{j,i}|^2] = \sigma_\Delta^2 \ . \tag{12.95}
$$

We then have

$$
\hat{\sigma}_{I+N}^2 = \sigma_F^2 \sigma_\Delta^2 + \sum_{i=0}^{M}\sum_{j=0}^{M} \mathrm{E}\left[\frac{\hat{c}_i \hat{c}_j^*}{\|\hat{\mathbf{c}}\|^2}\right] \mathrm{E}[w_i w_j^*] \tag{12.96}
$$

where $\sigma_F^2 = \sum_{i=0}^{L} \sigma_f^2(i)$ and $\sigma_f^2(i)$ is the variance of the ith channel tap.

To determine $\mathrm{E}[w_i w_j^*]$, define the vector $\mathbf{w}$ as

$$
\mathbf{w} = \sum_{k=1}^{N_I} \mathbf{B}(k)\mathbf{g}(k) + \mathbf{n} \tag{12.97}
$$

where $\mathbf{B}(k)$ is an $(M+1) \times (L+1)$ matrix consisting of the symbols from the kth interferer with associated channel tap vector $\mathbf{g}(k)$, N is the number of interferers, and $\mathbf{n}$ is the vector of additive white Gaussian noise samples. The elements of $\mathbf{w}$ are

$$
w_i = \sum_{k=1}^{N_I}\sum_{\ell=0}^{L} b(k)_{i,\ell}\, g(k)_\ell + n_i \ , \quad i = 0, \dots, M \tag{12.98}
$$

where $\mathbf{B}(k) = [b(k)_{i,\ell}]$ and $\mathbf{g}(k) = [g(k)_\ell]$. We now assume that the data symbols have zero mean, the data sequences comprising the $\mathbf{B}(k)$ matrices for the interferers are both uncorrelated and mutually uncorrelated, and the n_i are independent zero mean Gaussian random variables with variance σ_n^2. Then

$E[w_i w_j^*] = 0$ for $i \neq j$ and

$$
\begin{aligned}
\sigma_w^2 = E[|w_i|^2] &= E\left[\sum_{k=1}^{N_I}\sum_{\ell=0}^{M} |b(k)_{i,\ell}|^2 \, |g(k)_\ell|^2 + |n_i|^2\right] \\
&= \sum_{k=1}^{N_I} \sigma_b^2 \sum_{\ell=0}^{L} \sigma_{g(k)_\ell}^2 + \sigma_n^2 \\
&= \sigma_I^2 + \sigma_n^2 \\
&= \sigma_{I+N}^2
\end{aligned}
\tag{12.99}
$$

where σ_b^2 denotes the symbol variance of the interferers, $\sigma_{g(k)_\ell}^2$ denotes the variance of the ℓth channel tap gain associated with the kth interferer, and σ_I^2 denotes the total interference power. Using this result, (12.96) becomes

$$
\hat{\sigma}_{I+N}^2 = \sigma_F^2 \sigma_\Delta^2 + \sigma_w^2 = \sigma_F^2 \sigma_\Delta^2 + \sigma_{I+N}^2 \;.
\tag{12.100}
$$

In practice, the ensemble averaging in (12.94) must be replaced by an empirical average over P independent output vectors $\mathbf{y}_i$ so as to provide the unbiased estimate

$$
\hat{\sigma}_{I+N}^2 = \frac{1}{P}\sum_{i=1}^{P} \frac{\mathbf{y}_i^H \hat{\mathbf{c}}_i^* \hat{\mathbf{c}}_i^T \mathbf{y}_i}{\|\hat{\mathbf{c}}_i\|^2} \;.
\tag{12.101}
$$

8.1.2   ESTIMATION OF C/(I+N)

A C/(I+N) estimator can be formed by using $\hat{\sigma}_{I+N}^2$, and one possibility is as follows. The total received signal power from the desired signal, interfering signals, and noise is

$$
\begin{aligned}
\sigma_{C+I+N}^2 &= \frac{1}{L+1} E\left[\mathbf{y}^H \mathbf{y}\right] \\
&= \frac{1}{L+1} E\left[\mathbf{g}^H \mathbf{A}^H \mathbf{A} \mathbf{g} + \mathbf{w}^H \mathbf{w}\right] \\
&= \frac{1}{L+1}\left(\sum_{j=0}^{L} \sigma_f^2(j) \sum_{i=0}^{M} |a_{i,j}|^2 + (L+1)\sigma_w^2\right)
\end{aligned}
\tag{12.102}
$$

where the second equality follows from the assumption that $\mathbf{w}$ has zero mean, and the third equality requires that either the elements of the data sequence comprising the $\mathbf{A}$ matrix or the channel taps are uncorrelated. Once again, when $|a_{i,j}|^2 = \sigma_a^2$ (a constant) then

$$
\sigma_{C+I+N}^2 = \sigma_a^2 \sigma_F^2 + \sigma_w^2 = \sigma_C^2 + \sigma_w^2 \;.
\tag{12.103}
$$

Using (12.100) and assuming that $\sigma_{\mathbf{\Delta}}^2$ is small yields the C/(I+N) estimate

$$\widehat{\mathrm{CIR}} = \left(\frac{\sigma_{C+I+N}^2}{\hat{\sigma}_{I+N}^2} - 1 \right) \approx \frac{\sigma_C^2}{\sigma_{I+N}^2} \ . \tag{12.104}$$

The above approximation becomes exact when $\mathbf{A}$ is known exactly. Finally, by replacing ensemble averages with empirical averages we obtain the C/(I+N) estimate

$$\widehat{\mathrm{CIR}} = \frac{\frac{1}{L+1} \sum_{i=1}^{P} \mathbf{y}_i^H \mathbf{y}_i}{\sum_{i=1}^{P} \frac{\mathbf{y}_i^H \mathbf{c}_i^* \mathbf{c}_i^T \mathbf{y}_i}{\|\hat{\mathbf{c}}_i\|^2}} - 1 \ . \tag{12.105}$$

8.2 TRAINING SEQUENCE BASED C/(I+N) ESTIMATION

The bursts in TDMA cellular systems contain known training and color code sequences. These sequences are used for BS and sector identification, sample timing, symbol synchronization, and channel estimation. As mentioned in the previous section, the (I+N) and C/(I+N) estimators will only work well when $\sigma_{\mathbf{\Delta}}^2$ is small. Fortunately, if the (I+N) and C/(I+N) estimators are used during the training and color code sequences[5], $\sigma_{\mathbf{\Delta}}^2 = 0$.

The (I+N) and C/(I+N) estimators of the previous section were evaluated through the software simulation of an IS-54 [95] system. The baud rate is 24,300 symbols/s and each frame is composed of 6 bursts of 162 symbols so that the frame rate is 25 frames/s. The MS is assumed to have correctly determined the serving BS, i.e., the color code is known, and is monitoring its half rate channel (one burst per frame). Therefore, the known symbols within a burst consist of the 14 symbol training sequence at the beginning of the burst, and a 6 symbol color code sequence in the middle of the burst as shown in Fig. 1.5. For simulation purposes, a two-equal-ray T-spaced Rayleigh fading channel was chosen. The channel taps were assumed to be uncorrelated, although tap correlation will not affect the proposed algorithms because the various estimates depend only on the sum of the tap variances $\sigma_F^2 = \sum_{i=1}^{M} \sigma_f^2(i)$. Shadowing is assumed to remain constant over the estimates and is therefore neglected. Finally, it is assumed that the receiver can correctly synchronize onto each of the received bursts, i.e., perfect sample timing is assumed.

Four consecutive symbols were used to form a 3×2 Toeplitz non-symmetric matrix $\mathbf{A}$. Let $\{y_1(i), \ldots, y_{14}(i)\}$ denote the 14 received symbols corresponding to the training sequence of the ith frame, and $\{y_{15}(i), \ldots, y_{21}(i)\}$ the 6 received symbols in the color code. From the training sequence 4 estimates of

---

[5]The color code is known provided the MS has correctly determined its serving BS.

(I+N) and C/(I+N) were formed by using the following 4 sets

$$\{\{y_1(i), \quad \dots, y_4(i)\}, \{y_5(i), \dots, y_8(i)\},$$
$$\{y_9(i), \dots, y_{12}(i)\}, \{y_{11}(i), \dots, y_{14}(i)\}\}$$

(12.106)

where the fourth set shares two symbols with the third set. Likewise, 2 estimates of (I+N) and C/(I+N) were formed from the 6 symbol color code sequence by using the 2 sets

$$\{\{y_{15}(i), \dots, y_{18}(i)\}, \{y_{17}(i), \dots, y_{21}(i)\}\} \qquad (12.107)$$

which share two common symbols. Although the (I+N) and C/(I+N) estimators in (12.101) and (12.105) assume independent $\mathbf{y}_j(i), j = 1,\dots, 21$, the additional estimates of (I+N) and C/(I+N) which use overlapped symbols at the ends of the training and color code sequences was found to improve the I+N and C/(I+N) estimates. The channel tap gains associated with the interferers were assumed to be constant during known symbols. Additive white Gaussian noise at 20 dB below the interference power was also included.

To evaluate the performance of the (I+N) estimator, we define the average absolute percentage error between the (I+N) estimate and the true interference plus noise power as

$$\frac{\left|\hat{\sigma}_{I+N}^2 - \sigma_{I+N}^2\right|}{\sigma_{I+N}^2} \times 100 \ . \qquad (12.108)$$

Fig. 12.21 depicts the average absolute percentage error over 500 independent averages for a specified averaging time (s), MS velocity *(v),* and number of interferers (N_I). Since the interference plus noise estimator is compared against σ_{I+N}^2 under the assumption that the fading has been averaged out, it is natural to expect the estimator to perform worse for lower MS velocities when the averaging length is short, as Fig. 12.21 illustrates. Nevertheless, the presence of multiple interferers can improve the estimate, since with multiple interferers it is less likely that the total interference power will be small due to fading.

Likewise, Fig. 12.22 depicts the average absolute percentage error between the (C+I+N) estimate, $\hat{\sigma}_{C+I+N}^2$, and the true total received power, σ_{C+I+N}^2. As before, the MS velocity has a large effect on the estimator performance. Also, the C/I has a minor effect. However, in contrast to the (I+N) estimator, the number of interferers has little effect for C/I between 5 and 20 dB and, hence, variations in the number of interferers are not shown in Fig 12.22.

Fig. 12.23 depicts performance of the C/(I+N) estimator for an actual C/I of 5 dB. Only the performance with C/I = 5 dB is shown, since the estimator was found insensitive to C/I variations when the actual C/I was between 5 and 20 dB. For high speed MS, C/(I+N) can be estimated to within 2 dB in less
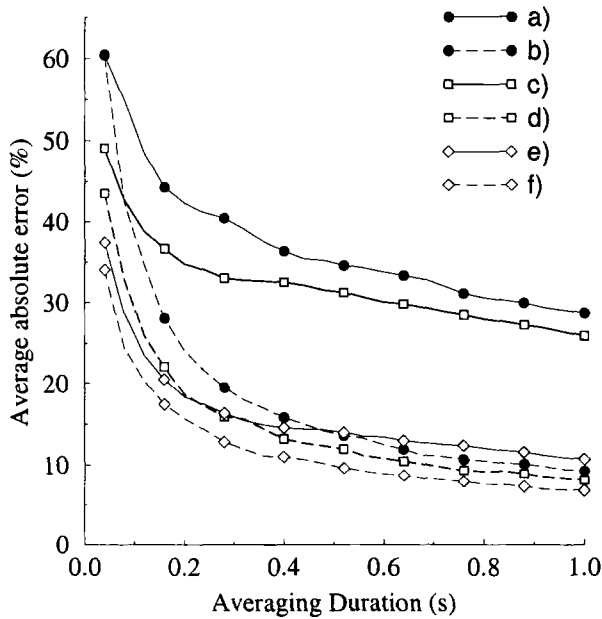
Figure 12.21. Average absolute percent error of the (I+N) estimator against the averaging time, from [21], [19]. The frame duration is 40 ms. Legend: a) $N_I = 1$, $v = 5$ km/h, b) $N_I = 1$, $v = 100$ km/h, c) $N_I = 2$, $v = 5$ km/h, d) $N_I = 2$, $v = 100$ km/h, e) $N_I = 6$, $v = 5$ km/h, f) $N_I = 6$, $v = 100$ km/h.

than a second. A slight improvement is also obtained when the MS uses two slots per frame (a full rate channel) as shown in Fig. 12.24.

9. SUMMARY

This chapter has provided a detailed discussion of local mean estimation in microcells and presented three velocity estimators that can be used for adaptive signal strength window averaging. The accuracy of local mean estimation in microcells was shown to depend on the Rice factor, the angle of the specular component, and the averaging length. For sample averaging, sample spacings less than $.5\lambda_c$ should be used. All three velocity estimators are relatively insensitive to the Rice factor under isotropic scattering. The LCR and COV velocity estimators are highly sensitive to non-isotropic scattering, whereas the ZCR estimator is reasonably robust. However, as is likely in urban microcells, the presence of a specular component can significantly reduce non-isotropic scattering biases. When $K = 0$, AWGN has the same effect on each of the three methods. However, when $K \neq 0$ and infinitely small sample spacing is used, the best performance is achieved with the ZCR, COV, and LCR methods, in that order. With larger sample spacings, the COV method is able to show
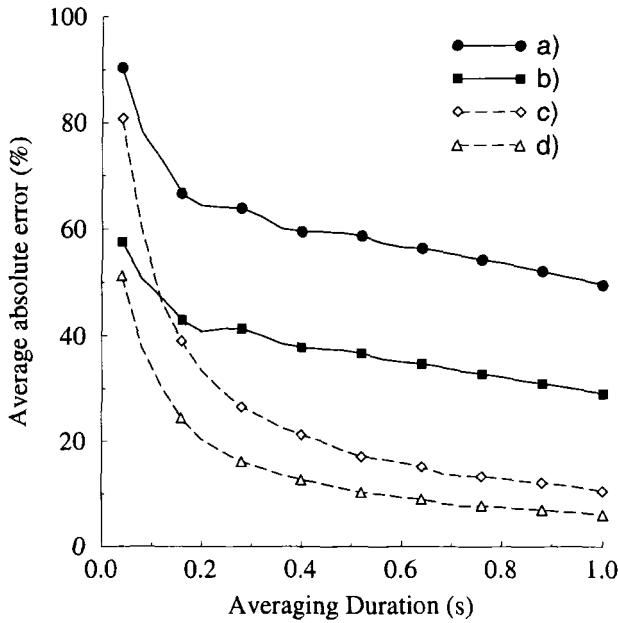
Figure 12.22. Average absolute percent error of the (C+I+N) estimator against the averaging time, from [19], [21]. The frame duration is 40 ms. Legend: a) $N_I = 1$, $v = 5$ km/h, C/I = 5 dB, b) $N_I = 1$, $v = 5$ km/h, C/I = 20 dB, c) $N_I = 1$, $v = 100$ km/h, C/I = 5 dB, d) $N_I = 1$, $v = 100$ km/h, C/I = 20 dB.

a greatly reduced sensitivity. To reduce AWGN effects, an adaptive filter bandwidth with respect to the maximum Doppler frequency and/or increasing the sampling period should be used. Increasing the sampling density reduces the bias in the final velocity estimate and improves the rate of convergence to changes in velocity or propagations effects such as the corner effect. The ZCR method has the fastest mean response time followed by the COV and LCR methods. All the velocity estimators are biased by the corner effect. Averaging the velocity estimates over several windows gives a slower initial convergence but reduces prolonged biases in the velocity estimates when a MS turns a corner. Shorter window lengths can also be used for faster adaptation. Each of the velocity estimators can successfully maintain good handoff performance over a wide range of MS velocities in a typical NLoS handoff scenario.

An analytical technique has been discussed for evaluating handoff performance, where the handoff rates can be studied in terms of the level crossings of the averaged signal level process.

Signal quality estimation techniques were examined for multipath fading channels having co-channel interference and additive Gaussian noise in TDMA cellular systems. Estimators for (C+I+N), (I+N) and C/(I+N) have been devel-
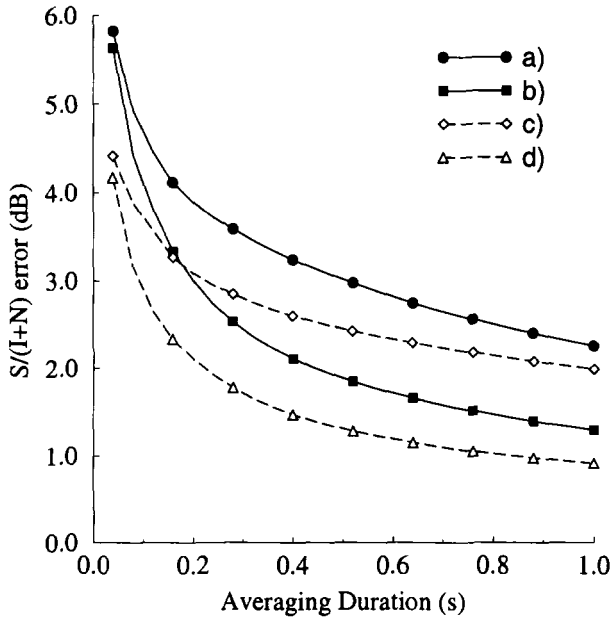
Figure 12.23. Average error of the C/(I+N) estimator against the averaging time, from [19], [21]. The frame duration is 40 ms. Legend: a) $N_I = 1$, $v = 5$ km/h, C/I = 5 dB, b) $N_I = 1$, $v = 100$ km/h, C/I = 5 dB, c) $N_I = 6$, $v = 5$ km/h, C/I = 5 dB, d) $N_I = 6$, $v = 100$ km/h, C/I = 5 dB.

oped whose accuracy is only a function of the symbol error statistics. These estimators have been applied to a cellular TDMA system, where knowledge of the training and color code sequences is used to form the estimates. Simulation results show that C/(I+N) can be estimated to within 2 dB in less than a second for high speed MSs.

APPENDIX 12.A: Derivation of Equations (12.47) and (12.60)

The limit in (12.60) can be written as

$$
\lim_{\tau \to 0} \frac{\frac{\lambda}{2\pi\tau}\sqrt{\frac{2\tilde{\mu}_{rr}(0)-2\tilde{\mu}_{rr}(\tau)}{\tilde{\mu}_{rr}(0)}}}{\frac{\lambda}{2\pi\tau}\sqrt{\frac{2\mu_{rr}(0)-2\mu_{rr}(\tau)}{\mu_{rr}(0)}}} = \frac{\lim_{\tau \to 0}\frac{\lambda}{2\pi\tau}\sqrt{\frac{2\tilde{\mu}_{rr}(0)-2\tilde{\mu}_{rr}(\tau)}{\tilde{\mu}_{rr}(0)}}}{\lim_{\tau \to 0}\frac{\lambda}{2\pi\tau}\sqrt{\frac{2\mu_{rr}(0)-2\mu_{rr}(\tau)}{\mu_{rr}(0)}}} . \qquad (12\text{-}12.A.1)
$$

Note that the limit of the denominator gives (12.47) and is a special case of the numerator limit with $N_o = 0$. To find the numerator limit the following property can be used [314]

If a function $f(\tau)$ has a limit as τ approaches a, then

$$
\lim_{\tau \to a} \sqrt[n]{f(t)} = \sqrt[n]{\lim_{\tau \to a} f(t)} \qquad (12\text{-}12.A.2)
$$
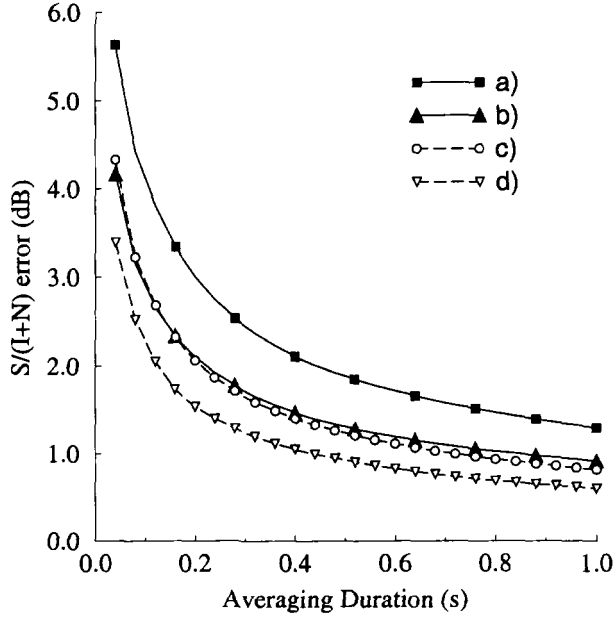
Figure 12.24. Average error of the C/(I+N) estimator for half rate and full rate channels against the averaging time, from [19], [21]. The frame duration is 40 ms. Legend: a) $N_I = 1$, $v = 100$ km/h, C/I = 5 dB, half rate channel, b) $N_I = 6$, $v = 100$ km/h, C/I = 5 dB, half rate channel, c) $N_I = 1$, $v = 100$ km/h, C/I = 5 dB, full rate channel, d) $N_I = 6$, $v = 100$ km/h, C/I = 5 dB, full rate channel.

provided either τ is an odd positive integer or n is an even positive integer and $\lim_{\tau \to a} f(\tau) > 0$.

Therefore, if the limit

$$\zeta = \lim_{\tau \to 0} f^2(\tau) = \lim_{\tau \to 0} \frac{\lambda^2}{(2\pi\tau)^2} \frac{2\tilde{\mu}_{rr}(0) - 2\tilde{\mu}_{rr}(\tau)}{\tilde{\mu}_{rr}(0)} \qquad (12\text{-}12.A.3)$$

exists and is positive, the solution to $(12 - 12.A.1)$ will be readily determined. It is apparent that L'Hôpital's Rule should be applied to determine the limit in $(12 - 12.A.3)$. After substituting $\tilde{\mu}_{rr}(\tau)$ from (12.38) and applying L'Hôpital's Rule four times, the limit is found as

$$\begin{aligned}
\zeta &= \lambda^2 \left(B_w^4 \, N_o{}^2 \, \pi^2 + 3 \, B_w \, K \, N_o \, (2\pi f_m)^2 \, \sigma^2 \right. \\
&\quad \left. + 2 \, B_w^3 \, K \, N_o \, \pi^2 \, \sigma^2 + 2 \, B_w^3 \, N_o \, \pi^2 \, a(0) \right) \\
&\quad \times \left[6 \, \pi^2 \left(B_w \, N_o + 2 \, a(0) \right) \left(B_w \, N_o + 4 \, K \, \sigma^2 + 2 \, a(0) \right) \right]^{-1} \\
&\quad + \lambda^2 \left(6 \, K \, (2\pi f_m)^2 \, \sigma^2 \, a(0) + 3 \, B_w \, K \, N_o \, (2\pi f_m)^2 \, \sigma^2 \, \cos(2\,\theta_0) \right.
\end{aligned}$$

$$+6\,K\,(2\pi f_m)^2\,\sigma^2\,a(0)\,\cos(2\,\theta_0)\Big)$$

$$\times\Big[6\,\pi^2\,(B_w\,N_o+2\,a(0))\,\big(B_w\,N_o+4\,K\,\sigma^2+2\,a(0)\big)\Big]^{-1}$$

$$+\frac{\lambda^2\,\big(-12\,a'(0)^2-12\,c'(0)^2-6\,B_w\,N_o\,a''(0)-12\,K\,\sigma^2\,a''(0)\big)}{6\,\pi^2\,(B_w\,N_o+2\,a(0))\,(B_w\,N_o+4\,K\,\sigma^2+2\,a(0))}$$

$$+\frac{\lambda^2\,(-12\,a(0)\,a''(0)-12\,c(0)\,c''(0))}{6\,\pi^2\,(B_w\,N_o+2\,a(0))\,(B_w\,N_o+4\,K\,\sigma^2+2\,a(0))}\qquad\text{(12-12.A.4)}$$

where $a(\tau)$ and $c(\tau)$ are given by (12.40) and $x'(0)$ denotes the derivative of $x(t)$ evaluated at 0. Consequently, $a(0)=b_0=\sigma^2$, $a'(0)=c(0)=0$, and

$$a''(0)\;=\;b_0(2\pi f_m)^2\int_0^{2\pi}\hat{p}(\theta)\cos^2\theta\,d\theta$$

$$c'(0)\;=\;b_0 2\pi f_m\int_0^{2\pi}\hat{p}(\theta)\cos\theta\,d\theta\;.\qquad\text{(12-12.A.5)}$$

Using these, and the fact that $(1+\cos(2\theta))/2\;=\;\cos^2(\theta)$, (12-12.A.4) is positive for all θ under all scattering scenarios mentioned here. Consequently, applying theorem 12-12.A.2, the limit of the numerator of $(12-12.A.1)$ is the square root of $(12-12.A.4)$, which if desired can easily be put in terms of the signal-to-noise ratio γ_S using

$$\sigma=\sqrt{\frac{\gamma_S N_o B_w}{2(K+1)}}\qquad\text{(12-12.A.6)}$$

from (12.55). The denominator of $(12-12.A.1)$, which is also (12.47) is obtained by assuming isotropic scattering and no noise, so that $a(0)=\sigma^2$, $a'(0)=c(0)=c'(0)=c''(0)=0$, $a''(0)=2b_0(\pi f_m)^2$, and $N_o=0$ in $(12-12.A.4)$. After taking the square root, the result is

$$\lim_{\tau\to 0}\frac{\lambda}{2\pi\tau}\sqrt{\frac{2\mu_{rr}(0)-2\mu_{rr}(\tau)}{\mu_{rr}(0)}}=v\sqrt{\frac{(1+2\,K+K\,\cos(2\,\theta_0))}{(1+2\,K)}}\;.$$

$$\text{(12-12.A.7)}$$

Problems

12.1. Suppose that a MS is traveling along a straight line from BS_1 to BS_2, as shown in Fig. 12.A.1. The BSs are separated by distance D, and the MS is at distance r from BS_1 and distance $D-r$ from BS_2. Ignore the effects of fading and assume that the signals from the two BSs experience independent log-normal shadowing. The received signal power (in decibels) at the MS

from each BS has the Gaussian density in (1.5), where the propagation path loss is described by

$$\mu_{\Omega_k} = \Omega_{(dB)}(d_o) - 10\beta \log_{10}(d/d_o)$$

- A handoff from BS_1 to BS_2, or vice versa, can *never* occur if $|\Omega_{1\ (dB)} - \Omega_{2\ (dB)}| < H$ but *may or may not* occur otherwise.
- A handoff from BS_1 to BS_2 will occur if the MS is currently assigned to BS_1 *and* $\Omega_{2\ (dB)} \geq \Omega_{1\ (dB)} + H$.

a) Find an expression for the probability that a handoff can never occur from BS_1 to BS_2, or vice versa.

b) Given that the MS is currently assigned to BS_1 what is the probability that a handoff will occur from BS_1 to BS_2.
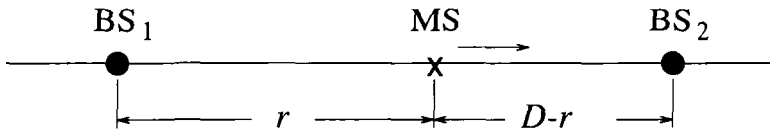
Figure 12.A.1. MS traversing from BS_0 to BS_1 along a handoff route.

12.2. A freeway with a speed limit of 120 km/h passes through a metropolitan area. If the average call duration is 120 s

a) What will be the average number of handoffs in a cellular system that uses omnidirectional cells having a 10 km radius.

b) Repeat part a) for a cellular system that uses 120° sectored cells having a 1 km radius.

12.3. Derive equation (12.23).

12.4. Derive equation (12.25).

12.5. Derive equation (12.38).

12.6. Derive equation (12.54).

This page intentionally left blank

Chapter 13

CHANNEL ASSIGNMENT TECHNIQUES

There are many methods of allocating a channel upon a new call arrival or handoff attempt. A good channel allocation algorithm is the one that yields high spectral efficiency for a specified **grade of service** (including link quality, probability of new call blocking, and the probability of forced termination) and given degree of computational complexity. It also keeps the planned cell boundaries intact, allocates a channel to a MS quickly, maintains the best speech quality for the MS at any instant, and relieves undesired network congestion. As shown in Fig. 13.1, there are three basic types of channel assignment algorithms, **fixed, flexible,** and **dynamic** [318].

Fixed channel assignment (FCA). is typically used by girst generation macrocellular systems where disjoint subsets of the available channels are permanently allocated to the cells in advance according to their *estimated* traffic loads. The cells are arranged in tessellating reuse clusters whose size is determined by the co-channel reuse constraint. For example, the North American AMPS system typically uses a 7-cell reuse cluster with $120°$ sectoring. The 12.5 MHz bandwidth allocation for AMPS can support a total of 416 two-way channels, 21 of which are control channels (one for each sector in a cluster), leaving a total of 395 traffic channels. This yields an allocation of 56 channels/cell with uniform FCA.

FCA provides adequate capacity performance in macrocellular systems that are characterized by stationary and homogeneous traffic, and a predictable propagation environment. In this case the channel resources can be allocated statically, since the call blocking probabilities can be predicted with reasonable certainty. Under conditions of nonstationary and nonhomogeneous traffic, however, FCA is spectrally inefficient because the channels are literally fixed to the cells. A new call or handoff arrival that finds all channels busy in a cell
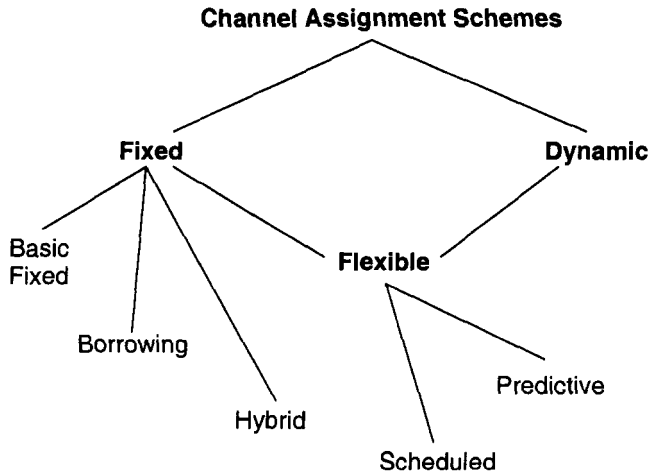
Figure 13.1. Basic classifications of channel assignment schemes, from [318].

will be blocked even though there may be several idle channels in the adjacent cells that could service the call. These blocking probabilities can be reduced by using various schemes that borrow channels from neighboring cells. The most basic scheme is simple borrowing, where a MS can be allocated a channel from a neighboring cell, provided that it does not degrade the link quality of other calls by introducing excessive co-channel interference. Once a channel is borrowed, all other cells that are within the co-channel reuse distance are prohibited from using the channel. The efficiency of this borrowing strategy tends to degrade in heavy traffic and the channel utilization is worse than FCA. This problem can be partially solved by using a hybrid channel assignment scheme, where the channels assigned to a cell are divided into two groups; the channels in one group are owned by the cell, while the channels in the other group may be borrowed. There are several variations of this theme. The ratio of the number of owned-to-borrowable channels can be dynamically varied to compensate for traffic changes.

In microcellular systems the propagation environment is highly erratic, and the traffic is characterized by spatial and temporal variations. Furthermore, the decreased cell sizes imply an increase in handoff traffic, since a call may be handed off several times before its natural completion. Because of these properties, the channel assignment problem in microcellular and macrocellular networks is fundamentally different. The uneven nature of the traffic and the larger volume of handoff attempts in microcellular networks demand careful attention. Furthermore, a microcellular channel assignment strategy has to be fast, because the handoffs must be serviced quickly due to the small cell sizes and propagation anomalies such as the street corner effect.

Dynamic channel assignment (DCA). is one well known solution to the microcellular channel assignment problem, where the dynamic nature of the strategy permits adaptation to spatial and temporal traffic variations while the distribution of control reduces the required computation and communication among base stations (BSs), thereby reducing system latencies. DCA schemes have no exclusive relationship between cells and channels, and in their most general form they allow any cell to use any channel that does not violate the co-channel reuse constraint. DCA schemes are known to outperform FCA under conditions of light nonstationary traffic. However, under conditions of heavy traffic FCA usually provides better performance, because the DCA schemes often yield an inefficient arrangement of channels. Although DCA has clear benefits, the cost can be quite high because it not only requires increased computation and communication among BSs but also an increased number of radio ports at the BSs; in the extreme case each BS must have the ability to use all channels simultaneously.

Practical DCA schemes differ in degree of network planning and the required communication among BSs. Centralized DCA schemes require centralized control with system-wide channel information. The extreme example is maximum packing (MP) [105], where a new call or handoff arrival is blocked only if there is no global rearrangement of calls to channels that will accommodate the service need. Unfortunately, the enormous computation and communication among cells render centralized DCA schemes impractical. In fact, the number of channel rearrangements required between two subsequent arrivals in a two dimensional network with MP can increase without bound with the number of cells in the network [279].

Fully decentralized DCA schemes are the other extreme andrequire no network planning or communication among BSs [111], [255], [8]. These schemes are ideal for cordless telephone systems that use MCHO, such as DECT. They often rely upon the passive monitoring of idle channels at each BS, allowing the cells to acquire any idle channel that is deemed to provide a sufficient carrier-to-interference ratio (C/I).

Decentralized DCA schemes require limited communication among local clusters of BSs. One DCA scheme is dynamic resource acquisition (DRA) [244]. With DRA, the channel (or carrier) that is acquired due to a new call arrival or handoff is chosen to minimize a cost function, and the channel (or carrier) that is released due a call completion or handoff is chosen to maximize a reward function. The cost and reward functions can be selected to maximize the spectral efficiency of the cellular network for a specified grade of service. The computation of the cost and reward functions for a given cell depends on the usages of the channels (or carriers) in the set of surrounding cells called the DRA neighborhood [244]. Another distributed DCA scheme is simple dynamic channel assignment (SDCA) [356]. SDCA performs slightly worse

than DRA, but requires communication among a smaller set of cells called the interference neighborhood [356],

Decentralized and fully decentralized schemes are not without their problems. These include service interruption, deadlock, and instability. A service interruption occurs when a channel allocation causes an existing link to fall below the threshold C/I. The interrupted mobile station (MS) then tries to find an alternate link and if unsuccessful a service termination occurs. This is known as deadlock. A sequence of successive interrupts, or rippling effect, caused by channel allocations is called an instability.

DCA schemes also have the advantage of assigning the same channel to a MS moving from one cell to another provided that the level of co-channel interference is tolerable, while FCA must conduct a handoff with a channel change because the same channel is not available in adjacent cells. Handoffs without channel changes are attractive because they can eliminate the need for channel searching and ultimately relieve the BSs from extra computation. More important, this mechanism is essential for supporting macrodiversity TDMA cellular architectures where the signal from a MS can be simultaneously received by two or more BS yielding a diversity improvement against shadow (and fading) variations. Such architectures provide the same benefit as soft handoff in CDMA systems.

Flexible channel assignment, algorithms combine aspects of fixed and dynamic channel assignment schemes. Each cell is assigned a fixed set of channels, but a pool of channels is reserved for flexible assignment. The assignment of flexible channels can be either scheduled or predictive [316]. Scheduled assignment schemes rely on known changes in traffic patterns. The flexible channels are assigned to the cells on a scheduled basis to account for these foreseeable changes in traffic patterns. With predictive assignment, the traffic load is continuously or periodically measured at every BS, and the flexible channels are assigned to the cells according to these measurements.

Forced terminations are generally perceived to severely degrade the quality of service. For this reason, **handoff priority** schemes are usually employed to allocate channels to handoff requests more readily than to new call arrivals. Handoff priority reduces the probability of forced termination at the expense of a (slight) increase in the probability of new call blocking. Practical cellular systems are designed to have a probability of new call blocking less that 5%, with a probability of forced termination perhaps an order of magnitude smaller.

The use of **guard channels** is one method of achieving handoff priority, where the channels are divided into two groups; one group is for new calls and handoff requests, and the other group is reserved for handoff requests only [170]. Another method is to queue the handoff requests (but not the new call arrivals) [170], [129]. This method can be combined with guard channels.

This chapter is intended to introduce the various approaches to cellular channel assignment. Unfortunately, most channel assignment schemes are quite detailed and founded largely on ad hoc principles. Furthermore, the channel assignment schemes are almost always evaluated by using detailed simulations with a variety of assumptions concerning the mobile radio environment, e.g., cellular topology and reuse factors, traffic patterns, propagation factors, mobility, etc.. The combination of these factors makes a systematic comparison of the various DCA schemes largely infeasible and a true consensus of the best scheme cannot be attained. Therefore, we will briefly outline some of the many different DCA schemes, followed by a detailed evaluation of a few specific schemes that serve to illustrate the basic concepts.

Throughout the chapter various performance measures will be used to evaluate the channel assignment schemes, including the following

- Probability of new call blocking, P_b, defined as

$$P_b = \frac{\text{number of new calls blocked}}{\text{number of new call arrivals}} \ .$$

- Probability of forced termination, P_f, defined as

$$P_f = \frac{\text{number of handoff calls blocked}}{\text{number of handoff attempts}} \ .$$

- Grade of service, GOS, defined as

$$\text{GOS} = \frac{P_b R_N}{(R_N + R_H)} + \frac{P_f R_H}{(R_N + R_H)}$$

  where R_N and R_H are the new call and handoff arrival rates, respectively.

- Channel changing rate, R_C, defined as

$$R_C = \frac{\text{number of channel changes}}{\text{number of handoffs}} \ .$$

The remainder of this chapter begins with an overview of some important DCA schemes. These include the fully centralized Maximum Packing (MP) and MAXMIN DCA strategies in Section 1.. Decentralized DCA strategies, such as First Available (FA), Nearest Neighbor (NN), and Dynamic Resource Acquisition (DRA) are discussed in Section 2.. Fully decentralized DCA schemes are the topic of Section 3., including Channel Segregation (CS) and Minimum Interference (MI), along with aggressive and timid strategies. Hybrid FCA/DCA schemes are the subject of Section 4.. The important class of borrowing schemes are the topic of Section 5., including Borrowing with Channel

Ordering (BCO), Borrowing with Directional Locking (BDCL), and Compact Pattern based DCA (CPDCA). Finally, our overview of DCA schemes wraps up with a treatment of Directed Retry (DR) and Directed Handoff (DH), Moving Direction (MD) strategies, reduced transceiver coverage, reuse partitioning, and handoff priority.

Following our results in [356], Section 10. provides some detailed and instructive examples of distributed DCA schemes for TDMA microcellular systems. In particular, two DCA strategies are presented that accommodate handoff queueing. An aggressive DCA policy with handoff queueing is also considered where a cell may be forced to terminate calls in progress in order to accommodate handoff requests in neighboring cells. The conditions for forced termination are carefully determined to ensure a performance improvement over a timid policy.

1.   CENTRALIZED DCA

Centralized DCA schemes require system-wide information and control for making channel assignments. As expected, centralized DCA schemes can theoretically provide the best performance. However, the enormous amount of computation and communication among BSs leads to excessive system latencies and renders centralized DCA schemes impractical. Nevertheless, centralized DCA schemes often provide a useful benchmark to compare the more practical decentralized DCA schemes.

1.1   MAXIMUM PACKING (MP)

The Maximum Packing (MP) algorithm was originally presented by Everitt and Macfadyen in 1983 [106]. With the MP policy a call is blocked only if there is no global rearrangement of calls to channels that will accommodate the call. Accomplishing this task requires a controller with system-wide information along with the ability to perform call rearrangements. The MP policy has the ability to serve all calls in a network with the minimum number of channels. Equipped with the capability, MP can yield the lowest new call blocking and forced termination probabilities of any DCA scheme under any traffic conditions.

Kelly [180] presented an interesting and enlightening analytical approach to MP DCA, by modeling the MP policy as a circuit-switched network. This allows some very powerful and well known network analysis tools to be applied to the analysis of MP DCA. The analysis ignores situations where the MS is moving from one cell to another or out of the service area, i.e., the handoff and roaming problem. Upon a call arrival in a particular cell, the MP policy checks to see if all reuse clusters that contain that cell have at least one channel available. If so, then the call can be accommodated through channel rear-

Figure 13.2.   Five cell deployment with MP

rangements; otherwise, the call is blocked. For example, consider the simple system consisting of five cells shown in Fig. 13.2. In this example, co-channel cells must be separated by at least two cells so there are three reuse clusters; $CL_1 = (1,2,3)$, $CL_2 = (2,3,4)$, and $CL_3 = (3,4,5)$. When a call arrives in cell 2, it can be accommodated if there is at least one channel available in clusters CL_1 and CL_2.

The stochastic model for MP uses the following definitions:

$$
\begin{aligned}
\mathcal{R} &= \text{set of cells in the system.} \\
K &= |\mathcal{R}| = \text{number of cells in the system.} \\
N_T &= \text{total number of channels available.} \\
n_i &= \text{number of calls in progress in cell } i. \\
\mathbf{n} &= (n_i, i \in \mathcal{R}) = \text{state vector.} \\
\mathcal{S} &= \text{set of admissible states.} \\
\rho_i &= \text{traffic load in cell } i.
\end{aligned}
$$

The set of admissible states depends on the particular cell layout. Let J be the number of complete or partial reuse clusters $CL_i, i = 1, \ldots, J$ that can be defined such that i) each reuse cluster differs by at least one cell, i.e., they are not totally overlapping, and ii) all cells are contained in at least one such cluster. For the example in Fig. 13.2, $J = 3$. Now let $\mathbf{A} = [a_{ij}]_{J \times K}$ be the **demand matrix**, where $a_{ij} = 1$, if $i = j$ and if cell j is in the same cluster as cell i; otherwise, $a_{ij} = 0$. For the example in Fig. 13.2

$$
\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix} . \tag{13.1}
$$

Matrix $\mathbf{A}$ tabulates the channel requirements for servicing calls that arrive in each of the cells. For example, a call arrival in cell 2 requires that a channel be available in CL_1 and CL_2 but not in CL_3 and, therefore, $a_{1,2} = a_{2,2} = 1$ and $a_{3,2} \approx 0$. Finally, let $N_i, i = 1, \ldots, J$ be the number of channels that are available in CL_i, $N_i \leq N_T$, and $\mathbf{N} = (N_1, \ldots, N_J)$. Then the set of admissible states is then given by

$$
\mathcal{S} = \{\mathbf{n} : \mathbf{A}\mathbf{n}^T \leq \mathbf{N}\} . \tag{13.2}
$$

It is well known (e.g., [179]) that $\mathbf{n}$ has the steady-state distribution

$$\pi(\mathbf{n}) = G(\mathbf{N}) \prod_{i \in \mathcal{R}} \frac{\rho_i^{n_i}}{n_i!}, \quad \mathbf{n} \in \mathcal{S} \tag{13.3}$$

where $G(\mathbf{N})$ is the normalizing constant

$$G(\mathbf{N}) = \left(\sum_{\mathbf{n} \in \mathcal{S}} \prod_{i \in \mathcal{R}} \frac{\rho_i^{n_i}}{n_i!} \right)^{-1}. \tag{13.4}$$

Then the steady-state probability that a call arrival in cell i is blocked is

$$B_i = 1 - \frac{G(\mathbf{N})}{G(\mathbf{N} - \mathbf{A}\mathbf{e}_i^T)} \tag{13.5}$$

where $\mathbf{e}_i$ is a length K vector with a '1' at position i and '0' elsewhere. Even though B_i appears to have a compact closed form expression, the computation of $G(\mathbf{N})$ is prohibitive except for very simple cases. Therefore, approximate methods are usually employed. One approximation assumes that the availability of channels in the clusters CL_i are *independent* events. This independence assumption leads to

$$B_i \approx 1 - \prod_{j \in \mathrm{CL}_i} (1 - E_j) \tag{13.6}$$

where the $E_j, j = 1, \ldots, J$ solve the nonlinear equations

$$E_j = E \left(\sum_{r \in \mathcal{R}} a_{jr} \rho_r \prod_{i \in r - \{j\}} (1 - E_i), N_j \right), \quad j = 1, \ldots, J \tag{13.7}$$

with

$$E(\rho, N) = \frac{\rho^N}{N!} \left(\sum_{n=0}^{N} \frac{\rho^n}{n!} \right)^{-1} \tag{13.8}$$

being the Erlang-B formula. Kelly [180] showed that there is an unique solution to the above nonlinear equations. The intuitive notion behind this approximation is that when $a_{jr} = 1$ the call arrivals of rate ρ_r in cell r are thinned by a factor of $1 - E_i$ by each cluster $\mathrm{CL}_i, i \in r - \{j\}$ before being offered to CL_j.

1.2   MAXMIN SCHEME

The MAXMIN scheme was introduced by Goodman *et al.* [145]. With the MAXMIN scheme, a MS is assigned a channel that maximizes the minimum C/I that any MS will experience in the system at the time of assignment. Assuming

that the link quality depends on the average received C/I, the C/I of MS_i at its serving BS is

$$\Lambda(\mathbf{d})_{(dB)} = \Omega(d_i)_{(dB)} - 10 \log_{10} \sum_{k \in I} 10^{\Omega(d_k)_{(dB)}/10} \ . \qquad (13.9)$$

where the $\Omega(d_j)_{(dB)}$ are independent Gaussian random variables with the density in (1.5) and (1.6), and d_j is the distance between MS_j and the BS for MS_i. The set I consists of all MSs other than MS_i that are using the same channel. A MS that requires service is assigned the channel j that gives

$$\max_{j \in C} \min_{i \in S} \{\Lambda_i\} \qquad (13.10)$$

where i and j index the set of MSs and channels, respectively, C is the set of channels that are available at the BS corresponding to the MS that requires service, Λ_i is the C/I of MS_i at its BS, and S is the set of all MS in service including the MS that requires service. We have already seen methods for C/I monitoring in Section 10.6.

2.   DECENTRALIZED DCA
2.1   FIRST AVAILABLE (FA) AND NEAREST NEIGHBOR (NN)

In 1972, Cox and Reudnik [72] proposed four basic decentralized DCA algorithms and compared them to FCA for the case of linear highway macrocells: **First Available** (FA), **Nearest Neighbor** (NN), **Nearest Neighbor+1** (NN+1), and **Mean Square** (MSQ). All four schemes allow a BS to acquire any idle channel that is not being used in its **interference neighborhood,** defined as the set of surrounding cells that can interfere with the BS. The schemes differ in the way that the channel selected should more than one channel be available for acquisition. The FA scheme acquires the first available channel found in the search. Assuming that a channel can be reused D_N cells away without causing excessive co-channel interference, the NN policy acquires the channel that is being used by the nearest BS at distance D_N or greater. The NN+1 policy acquires the channel that is in use at the nearest BS at distance $D_N + 1$ or greater with the goal of allowing more MSs to retain the same channel as they cross cell boundaries. The MSQ policy seeks to assign the available channel that minimizes the mean square of the distances among all BSs using the same channel. The DCA schemes were shown to outperform the FCA schemes in terms of probabilities of new call blocking and forced termination, except under conditions of heavy traffic. Of these four DCA schemes, the NN policy performs the best.
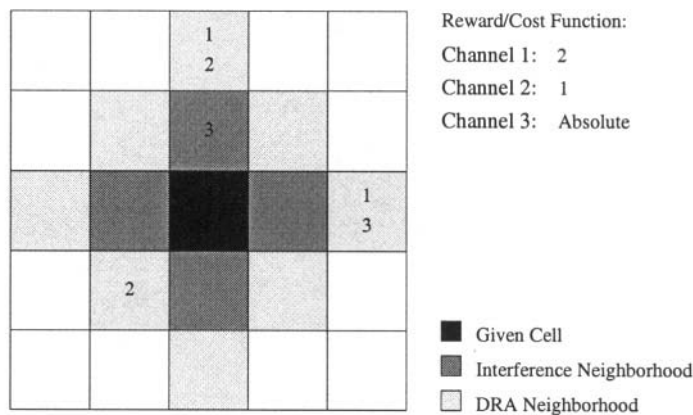
Figure 13.3. The DRA (light shaded) and interference (dark shaded) neighborhoods of a cell under consideration (shaded black). The numbers within each cell indicate active channels.

2.2   DYNAMIC RESOURCE ACQUISITION (DRA)

Nanda and Goodman [244], have proposed a distributed DCA strategy called Dynamic Resource Acquisition (DRA). When a channel must be selected for acquisition or release by a BS, DRA calculates a reward/cost function for each channel. The reward associated with a channel release is the number of cells in the interference neighborhood of the BS that could acquire the channel after it is released. When a channel is released, the busy channel giving the largest reward is selected. Channel rearrangements may be required to do this. The cost associated with a channel acquisition is the number of cells in the interference neighborhood of the BS that would be deprived from using the acquired channel. When a channel is acquired, the available channel having the smallest cost is selected. In the event of a tie in the reward/cost function, the released/acquired channel is chosen randomly.

As described in [244], the calculation of the reward/cost function requires channel usage information from all the cells within the **DRA neighborhood** of a BS. The DRA neighborhood of a BS is the set of cells whose interference neighborhoods overlap with the interference neighborhood of that BS. Any cell outside the DRA neighborhood of a BS will not affect the calculation of the reward/cost function associated with that BS. Fig. 13.3 illustrates the reward/cost functions associated with three carriers for a 2-D grid of square cells. The cell under consideration is shaded black. In case of a carrier acquisition, Channel 2 would be selected by the given BS since it has the smallest cost. Channel 3 could not be selected, because it would violate the co-channel reuse constraint. If Channels 1 and 2 are active in the given cell and a carrier is to be released, then Channel 1 would be selected since it gives the largest reward.

3. FULLY DECENTRALIZED DCA

3.1 CHANNEL SEGREGATION (CS)

Akaiwa and Andoh [9] proposed a distributed adaptive self-organizing DCA strategy whereby the BSs use Channel Segregation (CS) to develop *favorite* channels through an evolutionary process that is based on the criteria of eliminating unnecessary interference. Their scheme has been developed for TDMA systems with the assumption that each BS can access any channel by tuning a carrier frequency and selecting a time slot. CS also accounts for the effect of unaccessible channels where a call can be blocked in a cell even when there are idle channels because of the restriction placed on the number of different carrier frequencies that may be simulatneously used, i.e., the BS has a finite number of radio ports each of which can be tuned to only one frequency.

A flowchart of the CS algorithm is shown in Fig. 13.4. Each BS ranks the channels according to a priority function *P(i)*, where a large *P(i)* corresponds to a high priority, e.g., in [9] $P(i) = N_s/N_t$, where N_s is the number of successful uses of the channel plus the number of accesses to the channel when it is idle but unaccessible, and N_t is the total number of trials for the channel. When a call arrives, the BS senses the highest priority channel from the list of channels it is not currently using. If the channel is sensed idle, then the channel is checked for accessibility. If accessible, it is acquired and its priority is increased; otherwise, its prioity is increased and the BS recursively senses the next highest priority channel that it is not currently using. If all channels are sensed busy, then the call is blocked. Akaiwa and Andoh [9] demonstrated by simulation that the CS policy outperforms FCA and the FA DCA algorithm.

The steps within the dashed box in Fig. 13.4 are a modification so that the original CS algorithm developed by Akaiwa [8] for FDMA systems can be applied to TDMA systems. Simulation results show that this modification achieves the goal of gathering channels with the highest priorities onto the same carrier frequency, thus reducing the probability of call blocking due to the unavailability of a BS transceiver.

3.2 CHANNEL SEGREGATION WITH VARIABLE THRESHOLD

Another channel segregation scheme has been proposed by Hanabe *et al.* [160] that uses prioritized orderings with a variable interference threshold. The channels are ranked from highest to lowest according to their priority values. Each BS measures the interference levels of its currently unused channels. For each channel, the priority value is decreased if the interference level is higher than a predefined threshold and the threshold for that channel is decreased. Likewise, the priority value is increased if the interference level is lower than
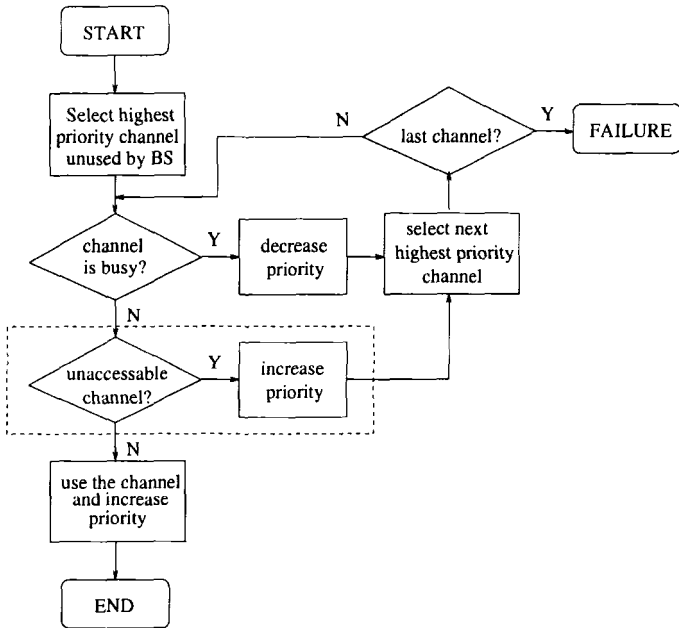
Figure 13.4. Channel segregation (CS) algorithm.

a predefined threshold and the threshold for that channel is increased. Hanabe *et al.* do not clearly define their priority function. However, it is likely that the priority function is defined as the ratio of number of times that the interference level of a channel is less than the threshold to the total number of times the channel is sensed. The interference threshold is varied depending on the ranking of the channel in the priority list. For example, the particular threshold that Hanabe *et al.* propose is as follows

$$T(k) = \begin{cases} T_o + 15 + S, & k = 1, 2, \ldots N_1 \\ T_o + 10 + S, & k = 9, 10, \ldots N_2 \\ T_o + 5 + S, & k = 17, 18, \ldots N_3 \end{cases} \qquad (13.11)$$

For the example in [160], $N_1 = 8$, $N_2 = 16$, and $N_3 = 72$, where N_3 is the total number of channels in the system, $T(k)$ *is* the threshold, T_o is the minimum required C/I, and S is a constant margin. Upon call arrival, the highest priority idle channel that meets the C/I threshold is chosen. If no suitable channels are available, the call is blocked. If the C/I drops below the required level during a call, a handoff procedure is initiated. In Hanabe *et al.'s* scheme, handoffs are not prioritized and are treated the same as new call arrivals.

The rationale for using a variable threshold $T(k)$ for each channel in (13.11) can be answered by examining the case where the thresholds are fixed. Allocation of a high priority channel with a fixed threshold is more likely to cause

interference since the C/I thresholds for all channels are the same. The reason is that a higher priority channel will be acquired over a lower priority channel if both channels exceed the C/I threshold, regardless of whether or not a lower priority channel would cause less interference to neighboring cells. Thus the allocation of the higher priority channel may cause service interruptions, dead-locks, and instability. For the case of a variable threshold, a higher threshold is assigned to higher priority channels to reduce the probability of co-channel interference, and a lower threshold is assigned to lower priority channels to decrease the probability of blocking.

3.3 MINIMUM INTERFERENCE (MI) SCHEMES

Schemes based on Minimum Interference (MI) have been presented by Goodman *et al.* [145]. The basic MI scheme has been incorporated into the CT-2 and DECT systems. With these schemes, the MS signals the BS with the strongest paging signal for a channel. The BS measures the interference level on all channels that it is not already using. The MS is then assigned the channel with the minimum interference. This policy coupled with mobile controlled handoff (MCHO) guarantees good performance. Variations of the MI scheme have been proposed that differ in the order in which MSs are assigned channels. These include **Random Minimum Interference** (RMI), **RMI with Reassignment** (RMIR), and **Sequential Minimum Interference** (SMI). The RMI scheme serves calls in the order that they arrive. The RMIR scheme serves the call requests according to the RMI scheme, but afterwards each MS is reassigned according to the MI policy. The order of reassignments is random. Those MSs initially denied service try again to acquire a channel. The procedure is repeated a fixed number of times. The SMI algorithm assigns channels according to the MI scheme but in a sequential order. In [145] linear microcells are considered and the sequence that is followed is to serve a MS only after all MS to its left have had a chance to be served. This, however, requires some co-ordination between BSs and the extension to 2-D schemes is not obvious. Goodman *et al.* [145] showed that the probability of blocking decreases with FCA, RMI, RMIR, SMI, in that order.

3.4 AGGRESSIVE AND TIMID DCA STRATEGIES

Distributed self organizing DCA algorithms that use aggressive and timid strategies were first introduced by Cimini and Foshini [55]. These simple autonomous DCA algorithms can self-organize with little loss in capacity compared to the best globally coordinated channel selection algorithm. In their paper, two classes of algorithms were studied; **timid** where a MS acquires a channel only if the channel is free of interference, and **aggressive** where a MS can acquire a channel even if it is not free of interference. The studies in

[55] showed that a linear array of cells could self organize its placement of a single channel to **saturate** the array from random starting arrangements. An array is saturated when no additional cells can use a channel without violating the co-channel reuse constraint. Channel usage in the array organizes itself according to the DCA policy. The performance of the algorithm is measured in terms of the **saturation density,** defined as the ratio of the number of cells using a particular channel to the number of cells in the arrary. Timid algorithms which require no call rearrangements have been shown to have saturation densities that compare favorably with FCA, while the aggressive algorithms have higher saturation densities at the expense of some instability. This is due to a **simulated annealing** mechanism where an instability perturbs a system so as to escape a local optimum in an attempt to reach the global optimum.

The saturation densities can be derived for the case of linear and hexagonal planar cells with R-cell buffering[1]. For linear cells, the maximum and minimum saturation densities are $C_{\max} = 1/(R+1)$ and $C_{\min} = 1/(2R+1)$. The saturation density can also be obtained for the random placement of a channel in a linear array. In this case, cells sequentially acquire the channel; the next cell to acquire the channel is chosen uniformly from those cells not already using the channel that could use the channel without violating the co-channel reuse constraint. The derivation of the saturation density in this case is quite lengthy but leads to the result [57]

$$C_{\mathrm{ran}} = \int_0^1 \exp\left\{ 2 \sum_{i=0}^{R-1} \frac{(v^{i+1}-1)}{i+1} \right\} dv \ . \tag{13.12}$$

The saturation density can also be obtained as a function of the traffic load ρ as [57]

$$C(\rho)\left(1 - C(\rho)R\right)^R = \rho\left(1 - (R+1)C(\rho)\right)^{R+1} \tag{13.13}$$

which has a unique solution in the interval $0 < C(\rho) < 1/(R+1)$. For hexagonal planar cells the minimum and maximum saturation densities are

$$C_{\min} = \frac{1}{1 + 3R(R+1)} \tag{13.14}$$

$$C_{\max} = \begin{cases} \frac{4}{3(R+1)^2} & , \quad R \text{ even} \\ \frac{4}{1+3(R+1)^2} & , \quad R \text{ odd} \end{cases} \ . \tag{13.15}$$

However, expressions for C_{ran} and $C(\rho)$ for the hexagonal planar array are unknown.

---

[1]The reuse factor N is related to the number of buffer rings R as follows. For linear cells $N = R + 1$. For hexagonal planar cells, $N = i^2 + ij + j^2$, where for R odd $i = j = (R+1)/2$, and for R even $i = R/2$ and $j = R/2 + 1$.

R	*Linear*	*Planar hexagonal*
1	0.864	0.693
2	0.825	0.658
3	0.804	0.627

Table 13.1. Normalized channel utilizations $\delta = C_{\mathrm{ran}}/C_{\mathrm{max}}$ for the timid DCA algorithm, from [56].

For the case of a single channel the blocking probability has the exact form [57]

$$P_b = 1 - C(\rho)/\rho \; . \tag{13.16}$$

For the case of multiple channels, Cimini *et al.* have derived a very accurate approximation for the call blocking performance of timid algorithms. They also derived lower bounds on the call blocking performance of aggressive algorithms [56]. If a total of N_T channels are available, the effective number of channels available for use in a reuse cluster of size N is δN, where δ is called the **normalized channel utilization** defined as the saturation density that is achieved with a particular algorithm C to the maximum possible saturation density C_{max}. Values for δ are tabulated in Table 13.1. For FCA, each cell has $m = N_T/N$ available channels and the blocking probability can be obtained from the Erlang-B formula $P_b = E(\rho, m)$ in (13.8), where ρ is the traffic load per cell. For the case of the timed algorithm a call is blocked if all channels are use in the interference neighborhood. To approximate the blocking probability for the timid algorithm, we replace ρ by $N\rho$ and m by δN_T, i.e., $P_b = E(N\rho, \delta N_T)$. To lower bound the blocking probability with an aggressive algorithm, we replace ρ by $N\rho$ as before, and m by N_T ($\delta = 1$), so that $P_b > E(N\rho, N_T)$. The performance of a practical aggressive algorithm will lie somewhere between the timid algorithm and the aggressive bound. Finally, we note that the blocking probability with an aggressive algorithm includes the calls that are blocked and the calls that are dropped because the aggressive algorithm has taken the channel and another suitable channel cannot be found.

4. HYBRID FCA/DCA SCHEMES

DCA schemes perform very well under light non-stationary non-homogeneous traffic. However, under conditions of uniformly heavy traffic FCA outperforms most of the DCA schemes, except perhaps MP. As a result of this behavior efforts have been directed toward hybrid FCA/DCA schemes that are intended to provide a compromise between FCA and DCA. Cox and Reudink [73] introduced a hybrid scheme, called **Dynamic Channel Reassignment** (DCR)

where each cell is assigned number of fixed channels, while the remaining channels are available for DCA. Fixed channels are used first to accommodate call requests. Calls that cannot be serviced by the fixed channels are offered to the dynamically assigned channels. The dynamic channel that is selected can be obtained by using any of the elementary schemes such as FA, NN, NN+1, etc. Upon a call completion on a fixed channel, DCR executes a search to determine if a call nominally assigned to a dynamic channel can be reassigned to the newly released fixed channel.

5.   BORROWING SCHEMES

Engel and Peritsky [101] introduced an FCA scheme with **borrowing.** The channels that are assigned to each BS are divided into two sets, fixed and borrowable. The fixed channels can only be used by the BS they are assigned to, while the remaining channels can be borrowed by a neighboring BS if necessary. Calls are serviced by using the fixed channels whenever possible. If necessary a channel is borrowed from a neighboring cell to service the call provided that the use of the borrowed channel does not violate the co-channel reuse constraint. The channel is borrowed from the neighboring BS having the largest number of available channels for borrowing. Improvements on this scheme were also proposed by Engel and Peritsky, where a call being serviced by a borrowed channel is transferred to a fixed channel whenever a fixed channel becomes available. The same idea was proposed by Anderson [13]. Scheduled and predictive channel assignment schemes have also been proposed, where the ratio owned to borrowable channels is dynamically varied according to the traffic conditions.

5.1   BORROWING WITH CHANNEL ORDERING (BCO)

Elnoubi *et. al.* [100] proposed a channel borrowing strategy that makes use channel orderings, called Borrowing with Channel Ordering (BCO). A group of channels is initially assigned to each cell according to a fixed channel assignment; these channels are called nominal channels and are arranged in an ordered list. The call arrival policy for BCO is illustrated by the flow chart in Fig. 13.5. Upon a call arrival in a cell, the BS searches for an available nominal channel nearest to the beginning of the channel ordering. If a nominal channel is available it is assigned to the call; otherwise, the BS attempts to borrow a channel from the adjacent cell having the largest number of channels available for borrowing. A channel is available for borrowing if it is unused in the adjacent cell and the other two co-channel cells. To illustrate this point, refer to Fig. 13.7. If cell Bl borrows a channel c from cell $A1$, then cells A1, A2, and A3 are **locked** from using channel c since their use of channel c will
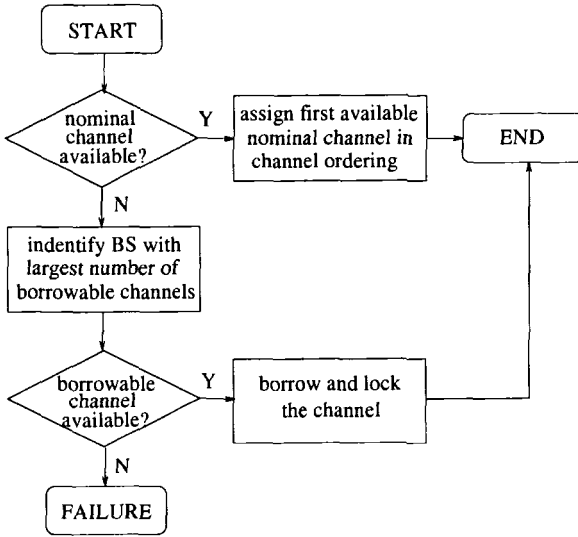
Figure 13.5.   Call arrival policy for the BCO algorithm.

violate the co-channel reuse constraint. Being blocked, channel c can neither be used to service a call in these three cells nor borrowed from these three cells. Finally, when a channel is borrowed from an adjacent cell, the available channel appearing nearest to the end of the channel ordering of the adjacent cell is selected. If no channels are available for borrowing, the call is blocked.

The call departure policy for BCO is illustrated in Fig. 13.6. When a call terminates on a borrowed channel, the borrowed channel is released in the three cells where it is locked. When a call terminates on a nominal channel and there are calls in progress with the same BS on borrowed channels, then the channel that is borrowed from the adjacent cell with the largest number of lent channels is released in the three cells where it is locked and its associated call is reassigned to the newly idle nominal channel. Finally, if a call completes on a nominal channel and there are no calls in progress with the same BS on borrowed channels, the call occupying the nominal channel nearest to the end of the channel ordering is reassigned to the newly idle nominal channel.

Kuek and Wong [184] introduced a DCA scheme called **Ordered Dynamic Channel Assignment/Reassignment** (ODCAR) that also combines channel ordering with channel rearrangements. The differences between the BCO and ODCAR schemes are very minor and quite subtle. BCO borrows a channel from the adjacent cell having the largest number of available *channels for borrowing,* while ODCAR borrows a channel from the adjacent cell having the largest number of available *nominal channels* that it could use to service its own calls. When a call completes on a nominal channel and there are calls
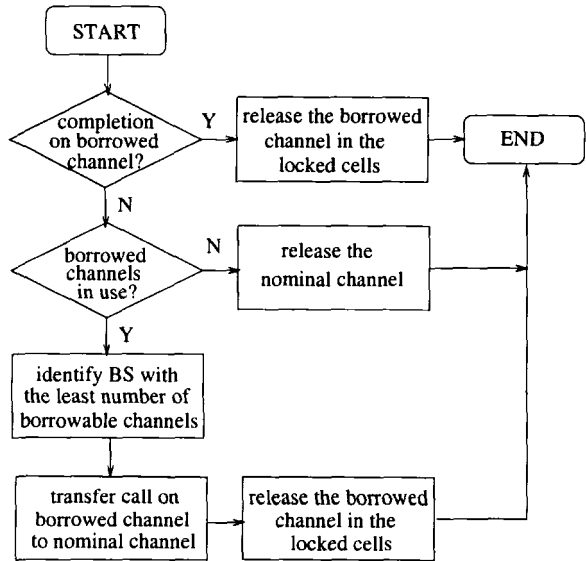
Figure 13.6. Call departure policy for the BCO algorithm.

in progress with the same BS on borrowed channels, then BCO releases the channel that is borrowed from the adjacent cell with the *largest number of lent channels* while ODCAR releases the channel that is borrowed from the adjacent cell with *line fewest number of nominal channels.* Finally, when a call completes on a borrowed channel BCO simply releases the channel in the three cells where it is locked, while ODCAR again releases the borrowed channel from the adjacent cell with the fewest number of nominal channels.

5.2 BORROWING WITH DIRECTIONAL LOCKING

Zhang and Yum [378] introduced a new scheme called Borrowing with DireCtional Locking (BDCL) and compared it with borrowing with channel ordering (BCO). Referring to the $N = 7$ cell reuse pattern in Fig. 13.7, the BCO scheme operates as follows. If cell B1 borrows a channel c from cell A1, then cells A1, A2, and A3 are **locked** from using channel c since their use of channel c would violate the co-channel reuse constraint. In the BDCL scheme, instead of locking channel c in cell A3 in all directions, channel c only needs to be locked in directions 1, 2, and 3. Cells that lie in the other three directions from A3, say B2, can freely borrow channel c from cell A3 without violating the co-channel reuse constraint. Whether or not channel c may be borrowed from A3 depends, however, on its locking conditions in A4, A5, and A6. Should the channel happen to be locked in A4, A5, or A6 but the cell locking is beyond B2's interference neighborhood, then B2 could still borrow channel c. This scheme
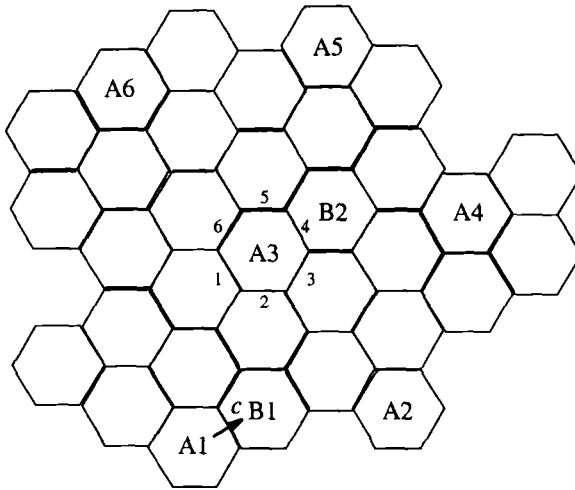
Figure 13.7.   Principle of borrowing with directional channel locking (BDCL).

increases the number of channels available for borrowing over the straight BCO scheme. Furthermore, the BDCL scheme uses channel rearrangements similar to the channel ordering scheme proposed by Elnoubi *et al.* [100], except that the directional locking mechanism is accounted for. Zhang and Yum [378] concluded that the BDCL scheme outperforms the BCO and FCA schemes in terms of blocking probabilities when the cells have nonuniform but stationary traffic loads.

5.3   BORROWING WITHOUT LOCKING

A borrowing scheme, Channel Borrowing Without Locking (CBWL), has been proposed by Jiang and Rappaport [174] that does not require channel locking by using borrowed channels with a reduced power level to limit inter-ference with co-channel cells. This allows the channel to be reused in all cells except the cell from which it has been borrowed. However, it also implies that channels can only be accessed in part of the borrowing cell. To determine if a channel can be borrowed with enough signal strength, the BS broadcast a **borrowed channel sensing signal** with the same reduced power of a borrowed channel.

The CBWL scheme divides the channels into six groups that can be lent to the neighboring cells, such that channels in the ith group can only be lent to the ith adjacent cell. This principle of **directional lending** is illustrated in Fig. 13.8, where channels in the group A1 can be borrowed by MSs in all of the B cells. Because of the reduced power level of borrowed channels, the MSs in the B cells that borrow the group A1 channels will be concentrated along the
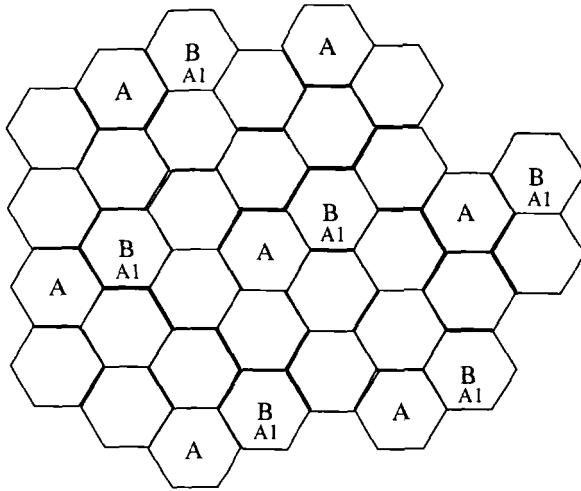
Figure 13.8. Principle of channel borrowing without locking (CBWL).

A-B cell boundaries. The CBWL scheme reduces the BS complexity because each BS does not need to have the capability of transmitting and receiving on all the channels assigned to its neighboring cells, but only a fraction of them in each cell. Furthermore, the division of borrowable channels into six groups limits co-channel interference so that locking is not required.

Various forms of channel rearrangements can be used enhance the scheme. For example, if cell B wishes to borrow a channel from cell A, the call is blocked if all the channels in group A1 of cell A are busy. However, it may be possible for cell A to transfer one of the calls to another group, say A2, to accommodate the borrow request. If this is not possible, cell A could itself borrow a channel from an adjacent cell to free up a channel to lend to cell B. Many other types of rearrangement policies are also possible.

5.4   COMPACT PATTERN BASED DCA

Yeung and Yum [373] introduced Compact Pattern based DCA (CPDCA), that attempts to dynamically keep the co-channel cells of any channel to a **compact pattern,** where a compact pattern of a network is the channel allocation pattern with the minimum average distance between co-channel cells. In other words, CPDCA attempts to increase spectral efficiency by keeping all channels at their minimum co-channel reuse distance. CPDCA accomplishes this task in two stages; i) channel acquisitions where an optimal idle channel is assigned to the MS, and ii) channel packing for the restoration of the compact patterns upon the release of a compact channel. Channel packing is achieved by reassigning at most one call per channel release.

Channels are assigned by using system-wide call arrival rate information to assign a channel that has a compact pattern that will yield the largest reduction in the overall system blocking probability. If a compact pattern is not available, the most optimal non-compact pattern is selected. If a call completes on a compact channel, CPDCA attempts to reassign a call on a noncompact channel to the newly idle compact channel. If no such call exists, CPDCA reassigns a call on the compact pattern that is least utilized to the newly idle channel. The first step minimizes the number of noncompact channels being used, while the second step packs the ongoing calls onto complete compact patterns. Yeung and Yun have shown their CPDCA scheme to outperform BDCL.

6.   DIRECTED RETRY AND DIRECTED HANDOFF

Everitt [104] introduced the **directed retry** (DR) and **directed handoff** (DH) channel assignment algorithms. If a BS does not have an idle channel available to service a call with the DR policy, the MS tries to acquire an idle channel in any other cell that can provide a satisfactory signal quality. DR exploits the overlapping nature of cells in a practical cellular system, where some percentage of MSs can establish a suitable link with more than one BS. DH also exploits the overlapping nature of cells to direct some of the ongoing calls in a heavily loaded cell to an adjacent cell that is carrying a relatively light load. Both the DR and DH schemes can be used in conjunction with either FCA or DCA, and Everitt concluded that FCA and maximum packing DCA in conjunction with the combination of DR and DH offer about the same performance. Therefore, FCA in conjunction with DR and DH is the preferred scheme, since an improvement over FCA can be gained without the added complexity of DCA.

7.   MOVING DIRECTION STRATEGIES

The moving direction (MD) strategy, proposed by Okada and Kubota, exploits information about the MS movement to reduce forced terminations and channel changes [250], [251]. The strategy attempts to assign an available channel from among those channels already assigned to MSs elsewhere in the service area that are moving in the same direction as the MS under consideration. Sets of MSs moving in the same direction are formed. When a MS enters a cell, a MS from the same set is probably leaving a cell. This allows both MSs to retain the same channel, thus reducing both the number of changes changes and probability of forced termination. This method is particularly useful for highway microcell deployment, where the traffic direction is highly predictable. Okada and Kubota compared the MD strategy with Cox and Reudnick's FA, NN, and NN+1 strategies [251]. The MD strategy was shown to offer the lowest channel changing rate and the lowest probability of forced termination.

However, the NN strategy provided a slightly lower probability of new call blocking.

A variation of the MD scheme considers both **Speed and Moving Direction** (SMD) [249]. MS are divided into two classes; high speed MS (HSMS) who are traveling at 50 to 60 km/h and low speed MS (LSMS) who are traveling at 0 to 4 km/h. To reduce the probability of forced termination and channel changing rate, SMD uses the MD policy for its HSMS call requests. For the LSMS, the NN strategy is employed since LSMSs do not experience forced terminations or channel changes as frequently as HSMSs. Again, the SMD scheme was shown to outperform the FA, NN, and NN+1 policies in terms of channel changing rates and the probability of forced termination. Finally, we mention that a variety of velocity estimation techniques are available as discussed in Section 10.3. Moving direction information can be obtained by using the past signal strength history in LOS environments [23] or the sign of the Doppler.

8. REDUCED TRANSCEIVER COVERAGE

Takeo *et al.* [317] proposed a scheme where nonuniform traffic is handled by adjusting the BS transmit power level of the control channel according to the traffic variance for every control period. Since the MS uses the control channel to determine which BS to connect to, the effective cell size is dynamically varied. Highly loaded cells decrease the transmit power to shrink the cell sizes, while lightly loaded cells increase the transmit power to enlarge the cell sizes. This scheme may cause some unwanted side effects, for example, handoffs can occur even for stationary MSs. The experimental results in [317] suggest that the call blocking probability increases in proportion to a decrease in the control period and, therefore, frequent updating of the control channel power should be avoided. Takeo *et al.* [317] did not address the problem that arises when many adjacent cells are heavily loaded, a potentially deleterious situation since it may result in coverage gaps within a particular reuse cluster.

8.1 REUSE PARTITIONING

Reuse partitioning employs a two-level cell plan where clusters of size M are overlaid on clusters of size N, $N > M$. Fig. 9.20 shows a FCA scheme using reuse partitioning with $M = 3$ and $N = 9$. As discussed in Section 9.5.1, reuse partitioning divides the available channels into two sets; one set can be used by the inner cells only, while the other set can be used by both the inner and outer cells. Reuse partitioning uses rearrangements so that whenever possible MSs in the inner cells are assigned channels allocated for use in the inner cells only.

An **autonomous reuse partitioning** (ARP) scheme has been suggested by Kanai [176]. With this scheme an *identical* ordering of channels is given to

all BSs. Upon call arrival, the channels are checked in order and the first one exceeding a C/I threshold for both the forward and reverse links is acquired. If no channels are available the call is blocked. The advantage of using a fixed ordering is that the channels higher in the ordering are used more frequently and, hence, have higher interference levels. This enables each BS to acquire channels with minimum C/I margins without the need for sorting channels according to their interference levels. The algorithm is self organizing in the sense that channels high in the ordering (with high interference levels) are allocated to MS that are close to a BS (with strong received signal levels). Channels low in the ordering tend to be allocated to MSs that are far from a BS with weak received signal levels.

Another scheme, **self-organized reuse partitioning** (SORP) has been proposed by Furukawa and Yoshihiko [125]. The BSs allocate channels by measuring the power levels transmitted from the MSs. This method relies upon a table at each BS that contains, for each channel, the average transmit power for MSs using the channel in its cell and all the surrounding cells. The table is updated with each call arrival and the update information is shared among the BSs. When a call arrives, the BS obtains the output power of the calling MS and assigns that channel with the corresponding average transmit power that is closest to that of the calling MS. The channel is acquired if available; otherwise the second closest candidate is examined, and so on. As a result of this procedure in each BS, channels that correspond to the same power are grouped autonomously for self-organizing reuse partitioning. The SORP scheme was shown to offer about the same blocking probability as the ARP scheme, but SORP requires less time to search for a channel and generally provides a higher C/I.

9. HANDOFF PRIORITY

Since the forced termination of a call in progress is worse than the blocking of a new call, it is important to consider **handoff priority** in the design of a channel assignment strategy. This is especially important in microcellular systems with their increased number of handoffs. Two possible methods of achieving handoff priority are to use **guard channels** where a fraction of the channels are reserved for handoff requests only [170], and **handoff queueing** where a handoff request from a MS is placed in a queue with the target BS while the MS maintains a radio link with its serving BS [170], [129]. Both methods are known to decrease the probability of dropped calls. However, queueing does this with a smaller increase in the probability of new call blocking.

Handoff queueing exploits the time interval that the MS spends in the handoff region, i.e., between the time when the handoff request is generated and the time when the call will be dropped due to a degradation in link quality. The simplest queueing scheme uses a first in first out (FIFO) policy. More elaborate

queueing schemes use measurement based priority, where the queue is ranked according to the measured link quality of the MSs in the queue [318]. MSs with the lowest link quality are placed in the highest priority class, and the handoff queue is sorted continuously according to the priority classes.

10. EXAMPLE DCA SCHEMES FOR TDMA SYSTEMS

We assume a TDMA system with **carrier groupings,** where the calls are packed into TDMA carriers such that each cell acquires the minimum number of carriers required to carry the calls. This packing may require channel rearrangements when the channels are released. A benefit of carrier groups is a reduction in the computation required to make decisions regarding acquisitions and releases. This reduction in complexity reduces the time required to select a channel, thus lowering the probability of dropped call.

Whenever a channel is needed a TDMA DCA scheme follows a strategy which, if necessary, selects a carrier for acquisition according to a **carrier acquisition criterion.** Likewise, when a channel is released another strategy is followed which, if necessary, selects a carrier to be released according to a **carrier release criterion.** The flow charts in Figs. 13.9 and 13.10 illustrate the general procedure for acquiring and releasing channels and carriers. The shaded blocks are steps that support handoff queueing and will be discussed later in the chapter.

ACQUIRE CARRIER Policy (non-queueing case).   As Fig. 13.9 shows, the following policy is executed upon a new call or handoff arrival:

1. If at least one idle channel is available among the already acquired carriers, then assign an idle channel to the call; otherwise attempt to acquire a new carrier according to the carrier acquisition criterion.

    (a) If the carrier acquisition is successful, then assign one channel of the newly acquired carrier to the call; otherwise block the call.

RELEASE CARRIER Policy (non-queueing case).   As Fig. 13.10 shows, the following policy is executed upon a call completion or a handoff departure:

1. If the channel release will not yield an idle carrier, then no carrier is released; otherwise a carrier is selected for release according to the carrier release criterion.

    (a) The call that occupies the carrier selected for release is reassigned to the newly idle channel, and the selected carrier is released.
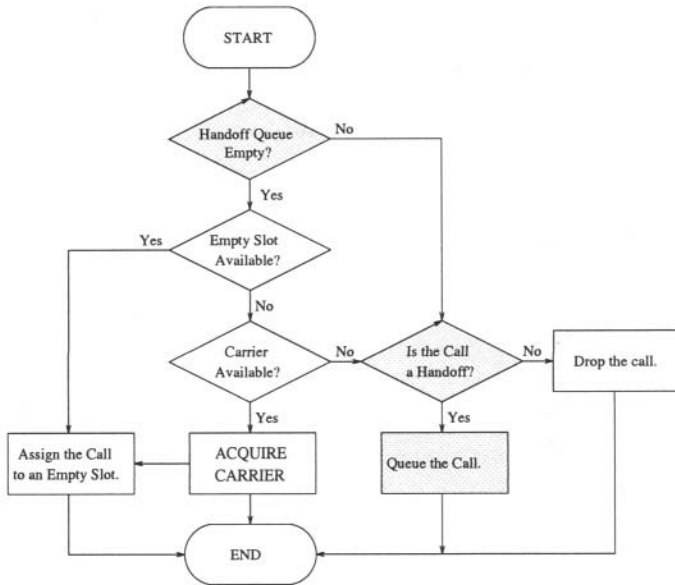
Figure 13.9. ACQUIRE CARRIER policy that is executed upon a new call or handoff arrival.
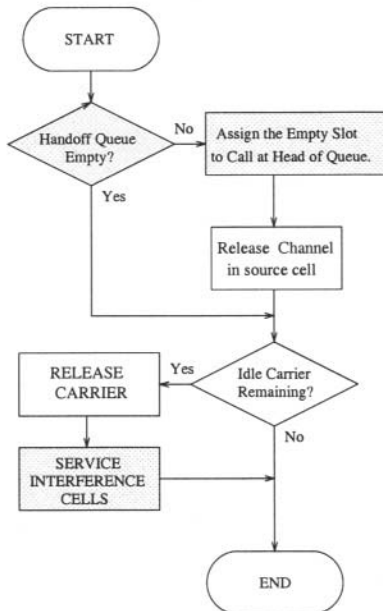
Figure 13.10. RELEASE CARRIER policy that is executed upon a call completion or handoff departure.

10.1 THE SIMPLE DCA (SDCA) STRATEGY

Elnoubi *et al.* [100] proposed the BCO strategy that makes use of different channel orderings in each cell. Here, we consider a channel assignment strategy that uses **carrier orderings** rather than channel orderings [356]. However, unlike the BCO strategy, the carriers are not explicitly divided into nominal and borrowed sets with a specified rule for borrowing carriers. In our scheme, each cell has its own carrier ordering, and no two cells separated less than the frequency reuse distance have the same carrier ordering. The orderings are designed so that carriers occurring near the beginning of a cell's carrier ordering occur near the end of the carrier orderings of the cells in its interference neighborhood. For example, suppose there are 9 available carriers with a 3-cell reuse cluster. Three different carrier orderings are necessary to ensure that cells within the frequency reuse distance have distinct carrier orderings. For example, the following carrier orderings will do.

$$
\begin{aligned}
A: & \quad \{1,4,7,8,9,5,6,2,3\} \\
B: & \quad \{2,5,8,9,7,6,4,3,1\} \\
C: & \quad \{3,6,9,7,8,4,5,1,2\}
\end{aligned}
\qquad (13.17)
$$

These carrier orderings are obtained by first listing the 9 available carriers column-wise until they are all assigned. Then columns 4 and 5 are permutations of the 3rd column, columns 6 and 7 are permutations of the 2nd column, and columns 8 and 9 are permutations of the 1st column. Notice that Carriers 1, 2, and 3 each appear first in one of the orderings and appear near the end of the other two orderings.

The carrier selection criterion is as follows. When a carrier is needed in a cell, the available carrier occurring nearest to the beginning of the cell's carrier ordering is selected. If there are no available carriers, then the carrier acquisition fails. When a carrier is released in a cell, the busy carrier occurring nearest to the end of the cell's carrier ordering is selected. This may require a rearrangement of calls within a cell to carriers that are closer to the beginning of the cell's carrier ordering. The above strategy is hereafter referred to as the **simple dynamic channel assignment** (SDCA) strategy, because of the simplicity of the carrier selection criterion.

Note that the SDCA scheme does not need an exchange of information within the interference neighborhoods. The busy/idle status of carriers can be determined by passive non-intrusive monitoring at each BS.

10.2 A QUEUEING DCA STRATEGY

Forced terminations of calls in progress are worse than blocking of new calls. Forced terminations or handoff blocking occurs when an active call crosses a cell boundary, and the target cell cannot accommodate the additional call. As

described in [170], [129], one way to establish **handoff priority** is to queue the handoff attempts. If the target cell is momentarily unable to accommodate the additional call, the MS maintains its link with the source cell and enters a queue in the target cell. A queue failure occurs when either the signal level drops below some threshold before the call can be serviced by the target cell, the time spent in the queue exceeds a time-out interval, or the queue overflows. A queue success occurs when a channel becomes available and the queue is non-empty. The newly available channel is then assigned to the call at the head of the queue, and a channel within the source cell is released. Here we combine handoff queueing with DCA.

In a DCA strategy, there are two ways for a channel to become available in a cell. Either a call terminates (due to a handoff or completion) or a carrier is released somewhere in the interference neighborhood thus allowing the carrier to be acquired by the cell. When a cell releases a carrier, there may be multiple cells in its interference neighborhood that could acquire the released carrier to service their queued calls. However, the frequency reuse constraint will be violated if all these cells acquire the carrier. To determine which cells may acquire the carrier, we may assume that each cell has a subset of carriers designated as **owned carriers.** The owned carriers are a subset at the beginning of the carrier orderings. Owned carriers are distributed so that no two cells separated less than the frequency reuse distance share any owned carriers. The remaining carriers are designated as **borrowed carriers.** Considering the previous example in (13.17) where 9 carriers were distributed among three carrier orderings, the owned and borrowed carrier orderings are

$$
\begin{array}{ccc}
 & \underline{\text{Owned}} & \underline{\text{Borrowed}} \\
A: & \{1,4,7\} & \{8,9,5,6,2,3\} \\
B: & \{2,5,8\} & \{9,7,6,4,3,1\} \\
C: & \{3,6,9\} & \{7,8,4,5,1,2\}
\end{array}
\qquad (13.18)
$$

Cells tend to use their owned carriers before borrowing carriers from other cells. When a cell releases a borrowed carrier, the cells in the interference neighborhood that own the released carrier are given the first opportunity to service their handoff queues. If any of these cells have queued calls, then they can acquire the carrier without violating the frequency reuse constraint. If none of the owner cells in the interference neighborhood acquire the released carrier, then some of the remaining cells in the interference neighborhood may acquire the carrier to service their handoff queues.

A strategy combines DCA and handoff queueing is described below along with the flow charts in Figs. 13.9, 13.10 , 13.11, and 13.12.

ACQUIRE CARRIER Policy (queueing case).   Referring to Fig. 13.9, the following policy is executed upon a new call arrival or handoff attempt:

1. If the handoff queue is not empty, then either queue the handoff call or drop the new call; otherwise

   (a) If there is at least one idle channel, then assign an idle channel to the call; otherwise try to acquire a according to the carrier acquisition criterion.

      i. If the carrier acquisition is successful, then assign one channel of the newly acquired carrier to the call; otherwise either queue the handoff call or drop the new call.

RELEASE CARRIER Policy (queueing case). Referring to Fig. 13.10, the following policy is executed upon a call completion, a handoff, or a failure from the handoff queue of an adjacent cell:

1. If the handoff queue is not empty then assign the newly available channel to the call at the head of the handoff queue. The channel currently serving the call is released according to the RELEASE CARRIER policy.

2. If the channel release will yield an idle carrier, then a carrier is selected for release according to the carrier release criterion. The call that occupies the carrier selected for release is reassigned to the newly idle channel, the selected carrier is released, and the SERVICE INTERFERENCE CELLS policy is executed.

SERVICE INTERFERENCE CELLS Policy. Referring to Fig. 13.11, the following policy is executed whenever a carrier is released:

1. If a borrowed carrier is released then any owner cell in the interference neighborhood that has a non-empty handoff queue and can acquire the released carrier without violating the reuse constraint, will acquire the carrier and service its handoff queue according to the SERVICE HANDOFF QUEUE policy.

2. After the owner cells are given the opportunity to service their queues, the remaining cells in the interference neighborhood are given the opportunity to service their handoff queues by using the SERVICE HANDOFF QUEUE policy.

SERVICE HANDOFF QUEUE Policy. Referring to Fig. 13.12, whenever a carrier is acquired in a cell having a non-empty handoff queue, the following policy is executed:

1. An empty slot is assigned to the call at the head of the handoff queue, and the channel currently serving the call is released by using the RELEASE CARRIER policy.
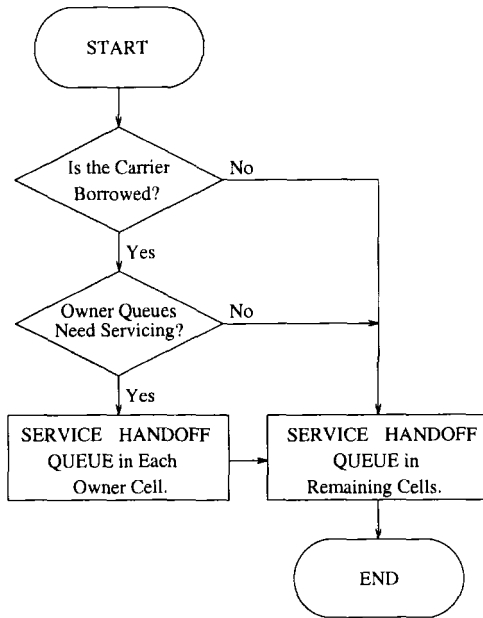
Figure 13.11.   SERVICE INTERFERENCE CELLS Policy.

(a) Step 1. is executed until either all of the available slots are filled or the handoff queue is empty.

10.3   AN AGGRESSIVE DCA STRATEGY

DCA strategies increase trunking efficiency by assigning channels to cells as they are needed. Care must be taken to avoid a poor allocation of channels; otherwise capacity will suffer. With SDCA, the carriers are acquired and released according to a carrier acquisition and release criteria that attempts to maximize capacity by favoring tightly packed arrangements of *co-carrier* cells. This strategy is similar to the 2-D RING strategy in [173] and suffers from the same problem; when a carrier is selected for acquisition, multiple carriers must be available for the carrier acquisition criteria to yield any advantage. The more carriers available for each selection process the better. At high traffic loads, very few carriers may be available for acquisition. In fact there may be only one or none, in which case there is no choice. Under such conditions, carriers tend to be assigned where they can be, rather than where they should be, and capacity suffers [105]. Under such conditions DCA strategies usually perform *worse* than FCA.

As discussed in Section 3.4 the performance at high traffic loads can be improved by using an aggressive policy where, under certain conditions, a cell
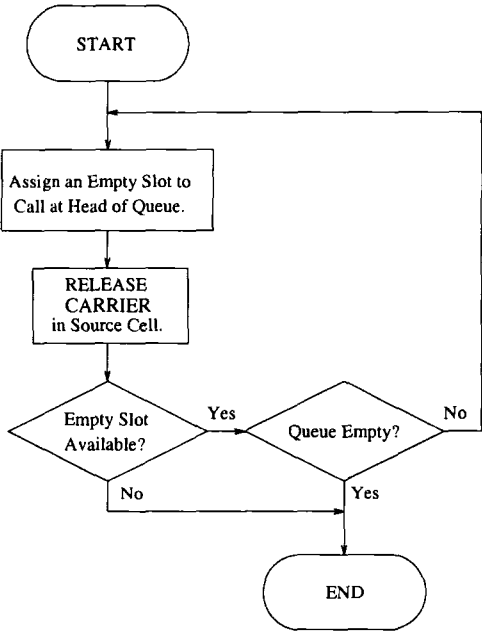
Figure 13.12.   SERVICE HANDOFF QUEUE Policy.

that cannot acquire a carrier may force a surrounding cell to give up a carrier so that it may service one or more calls. Thus, a cell can actually *take* a carrier from its neighbors if none are otherwise available, according to the following TAKE CARRIER policy.

TAKE CARRIER Policy.   Referring to Fig. 13.13, the following policy is executed when a call is in jeopardy due to a handoff failure (in the no queueing case), a queue failure, or the execution of the TAKE CARRIER policy in another cell:

1. The entire set of carriers is examined. If all carriers are being used, then the TAKE CARRIER policy fails; otherwise, all unused carriers are examined, and the number of calls within the interference neighborhood that will be placed in jeopardy by taking each carrier is calculated. Note that the number of jeopardized calls must be determined at each cell in the interference neighborhood. The carrier that will place the fewest number of calls in jeopardy is selected. In event of a tie, the carrier appearing earliest in the carrier ordering is selected. Let C_j be the number of calls that will be placed in jeopardy by taking the selected carrier.

2. The number of calls that will be serviced by taking the selected carrier, C_s is calculated. For a handoff attempt and no queueing, $C_s = 1$; for a queue

failure, C_s is the minimum of the number of queued calls C_q and the number of slots per carrier N_s; for a carrier that is lost to another cell executing the TAKE CARRIER policy, C_s ranges from 1 to $.N_s$.

3. If the selected carrier is owned and $C_j > C_s$ or if the selected carrier is not owned and $C_j \geq C_s$, then the TAKE CARRIER policy fails; otherwise all cells in the interference neighborhood that are currently using the selected carrier are told to release it.[2]

4. Each cell in the interference neighborhood releases the selected channel.

5. The selected channel is taken. For a handoff attempt (no queueing), the handoff is completed. For a queue failure, the SERVICE HANDOFF QUEUE policy is executed. For a carrier that is lost to another cell executing the TAKE CARRIER policy, the slots of the taken carrier are assigned to the calls in jeopardy.

6. Each cell that was forced to release the selected carrier executes the SER-VICE INTERFERENCE CELLS policy.

7. Each cell that was forced to release the selected channel and still has calls in jeopardy after the cell taking the carrier services its queue, executes the ACQUIRE CHANNEL policy.

   (a) If the carrier acquisition is successful, then the cell executes the SER-VICE HANDOFF QUEUE policy; otherwise it executes the TAKE CARRIER policy.

   (b) If a cell from which a carrier was taken cannot obtain a new carrier, it must drop some of its calls. Queued calls are dropped first because they are in greater danger of being dropped than active calls. If more calls must be dropped after dropping the queued calls, then active calls are dropped until there are no excess calls.

Note that the TAKE CARRIER policy is only executed if the SCDA carrier acquisition criteria fails to acquire a carrier. Unlike SCDA, the TAKE CAR-RIER policy acquires carriers that place the fewest number of calls in jeopardy. From a practical standpoint it is important to note that the aggressive SCDA strategy requires communication among BSs in the interference neighborhood to execute the TAKE CARRIER policy.

---

[2]When there is no queueing $C_s = 1$ and, therefore, only an owned carrier can be taken that will not place more than one call in jeopardy.
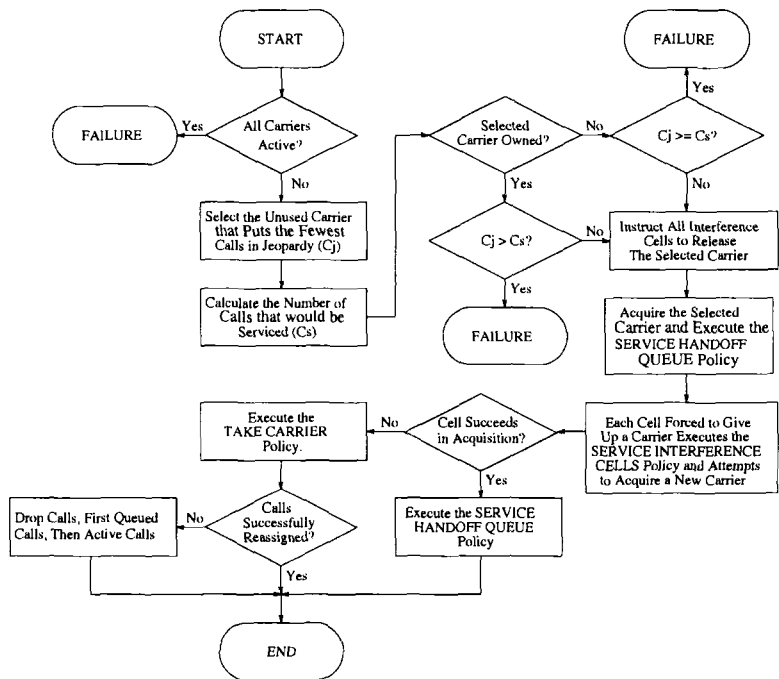
Figure 13.13. TAKE CARRIER Policy.

10.4 SIMULATION MODEL, RESULTS, AND DISCUSSION

Consider a microcellular environment consisting of a rectangular grid of intersecting streets, as shown in Fig. 13.14. It is assumed that MS traffic flowing off an edge of the grid wraps around to the opposite edge. However, the interference neighborhoods of each cell do not wrap around. If two cells are on opposite edges of the grid, such that MSs leaving one cell enter the other, they may simultaneously use the same carrier since they are not spatially adjacent.

Line-of-sight co-channel cells must be separated by at least 3 cells. There are no reuse constraints on non line-of-sight co-channel cells, due to the corner effect. The frequency reuse factor is 4, meaning that the set of carriers must be divided into 4 subsets for FCA, and for SDCA there must be 4 different carrier orderings. The interference neighborhood and cell reuse pattern is shown in Fig. 13.15.

To account for the uneven distribution of teletraffic in the microcellular environment the identical active-dormant Markov model from [244] is used, but modified to account for handoff queueing. The model is Markovian so that all events occur with exponentially distributed interarrival times. However, the
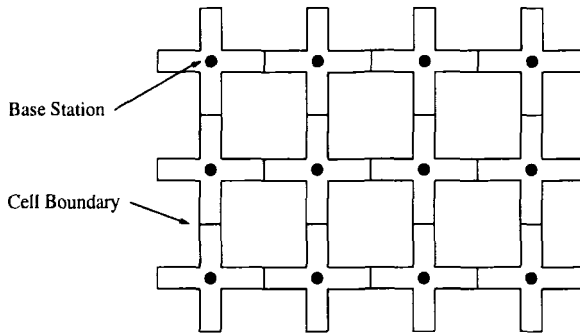
Figure 13.14.   Cells and BSs in an urban microcellular environment.

1	2	3	4	1	2	3	4	1
4	1	2	3	4	1	2	3	4
3	4	1	2	3	4	1	2	3
2	3	4	1	2	3	4	1	2
1	2	3	4	1	2	3	4	1
4	1	2	3	4	1	2	3	4
3	4	1	2	3	4	1	2	3
2	3	4	1	2	3	4	1	2
1	2	3	4	1	2	3	4	1

Figure 13.15.   Interference neighborhood and cell reuse pattern.

parameters of the distributions change with time to reflect the time-varying nature of the model. The state of cell i at any time can be described by the following parameters

New call arrival rate: λ_i
Number of active calls: $N_{\text{active},i}$
Number of queued calls: $N_{\text{queued},i}$

New Call Arrivals:.   Call arrivals in cell i are Poisson with rate λ_i. This parameter is binary valued, where $\lambda \in \{\lambda_{ACT}, \lambda_{DOR}\}$. These two new call arrival rates correspond to two different cell modes, active and dormant. The arrivals of new calls in different cells are assumed to be independent, so the global call arrival rate is

$$\Lambda = \sum_i \lambda_i \ .$$

(13.19)

Call Completions:. The duration of each call is exponentially distributed with mean μ. In cell i there are $N_{\text{active},i}$ active calls and $N_{\text{queued},i}$ queued calls, any of which could be completed at any time. These calls are assumed to be independent, so the call completion rate in cell i is

$$r_{c,i} = \frac{N_{\text{active},i} + N_{\text{queued},i}}{\mu} \ . \tag{13.20}$$

The completion of calls in different cells are assumed to be independent. Therefore, the global call completion rate is:

$$r_c = \sum_i r_{c,i} \ . \tag{13.21}$$

Handoff Attempts. A handoff is attempted whenever an active call crosses a cell boundary and needs to be serviced by the target cell. To determine the handoff rate, it is assumed that each call is handed off an average of h times over its duration. Since the traffic flows wrap around the grid edges, the handoff calls are uniformly distributed to one of the four neighboring cells. Queued calls can be safely assumed to never cross a cell boundary, because the time required to traverse a cell will be much longer than the maximum time allowed in the handoff queue. Therefore, queued calls do not contribute to the handoff rate. The handoff rate in cell i is

$$r_{h,i} = \frac{h N_{\text{active},i}}{\mu} \ . \tag{13.22}$$

Call handoffs in different cells are assumed to be independent, so the global handoff attempt rate is

$$r_h = \sum_i r_{h,i} \ . \tag{13.23}$$

Mode Transitions. Each cell remains in its current mode for duration D, where D is exponentially distributed with mean $1/\overline{D}$. If the cell is in active mode, then $\overline{D} = \overline{D}_{ACT}$, and if the cell is in dormant mode, then $\overline{D} = \overline{D}_{DOR}$. If there are N_{ACT} active cells and N_{DOR} dormant cells, then the global active-to-dormant and dormant-to-active transition rates are, respectively,

$$r_{ACT \to DOR} = \frac{N_{ACT}}{\overline{D}_{ACT}} \tag{13.24}$$

$$r_{DOR \to ACT} = \frac{N_{DOR}}{\overline{D}_{DOR}} \tag{13.25}$$

The probability of a cell being in the active mode is

$$P_{ACT} = \frac{\overline{D}_{ACT}}{\overline{D}_{ACT} + \overline{D}_{DOR}} \ . \tag{13.26}$$

As the simulation progresses, five types of events are generated: new call arrivals, call completions, handoff attempts, active-to-dormant mode transitions, and dormant-to-active mode transitions. All events occur independently. Therefore, five random times are generated and the next event corresponds to the one with the minimum time. Once an event is selected, the event must be randomly assigned to a cell. The probability of cell i being selected for each type of event is

New call arrival: $\qquad\qquad \lambda_i/\Lambda$

Call completion: $\qquad\qquad r_{c,i}/r_c$

Handoff Attempt: $\qquad\qquad r_{h,i}/r_h$

ACT $\to$ DOR transition: $\begin{cases} 1/N_{ACT} & \text{, if cell } i \text{ is active} \\ 0 & \text{, if cell } i \text{ is dormant} \end{cases}$

DOR $\to$ ACT transition: $\begin{cases} 1/N_{DOR} & \text{, if cell } i \text{ is dormant} \\ 0 & \text{, if cell } i \text{ is active} \end{cases}$

The active to dormant traffic ratio $R_{ACT/DOR} = \lambda_{ACT}/\lambda_{DOR}$ specifies the ratio of the new call arrival rates in the active and dormant cells. To complete the model, we specify the offered traffic per cell, ρ. Then the active and dormant call arrival rates are:

$$\lambda_{ACT} = \frac{\rho R_{ACT/DOR}}{\mu(1 + P_{ACT}(R_{ACT/DOR} - 1))} \qquad (13.27)$$

$$\lambda_{DOR} = \frac{\rho}{\mu(1 + P_{ACT}(R_{ACT/DOR} - 1))} \qquad (13.28)$$

where P_{ACT} is as defined in (13.26). The parameters used in the simulations are as follows

Number of cells:	144 (12 x 12 square)
Total number of carriers:	40
Number of slots per carrier:	3
Number of channels per cell (FCA):	30
Number of owned carriers per cell (DCA):	10
Average call duration:	120s
Average number of handoffs per call:	3
Average duration of the ACTIVE mode:	60s
Average duration of the DORMANT mode:	600s
ACTIVE to DORMANT traffic ratio:	5
Offered traffic:	0-50 Erlangs
Queue size:	10
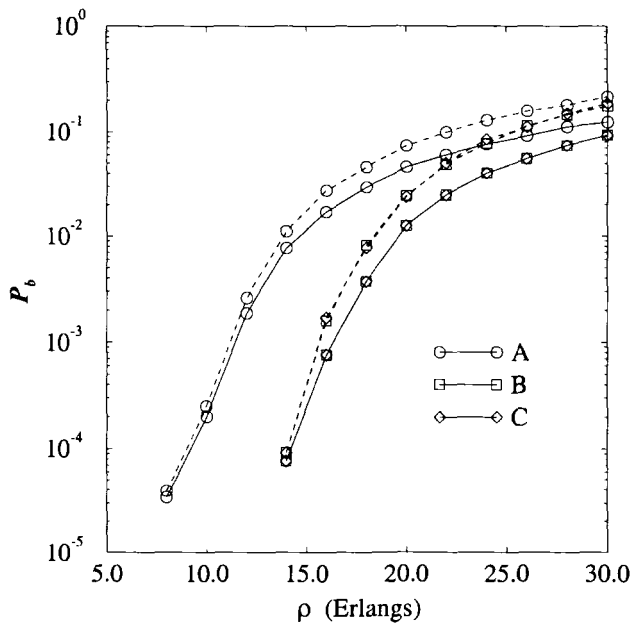
Figure 13.16. Probability of new call blocking versus offered traffic without queueing (solid) and 5-second handoff queue (dashed). Legend: A = FCA, B = SDCA, C = Aggressive SDCA, from [356].

Figs. 13.16 through 13.17 compare the probability of new call blocking and the probability of forced termination for the FCA, SDCA, and aggressive SDCA strategies. Results are shown without handoff queueing and with a 5-second handoff queue. Observe from Fig. 13.16 that a substantial reduction in the probability of new call blocking is achieved by using SDCA as compared to FCA. Handoff queueing causes a sight increase in the probability of new call blocking because handoff calls are given priority over new calls when a channel has been released and is available for acquisition. Aggressive SDCA also causes a very slight increase in the probability of new call blocking over non-aggressive SDCA. Fig. 13.17 shows that both handoff queueing and SDCA significantly lower the probability of forced termination. Aggressive SDCA tends to be more effective than the non-aggressive SCDA when these schemes are combined with handoff queueing.

Carrier Acquisitions. It is useful to determine the increase in the rate of carrier acquisitions that results from using aggressive SDCA. Fig. 13.18 plots the cell carrier acquisition rate for aggressive and non-aggressive SDCA. Notice that the carrier acquisition rates at lower traffic loadings are almost the same. At higher traffic loadings hand-off queueing has the largest effect on the carrier
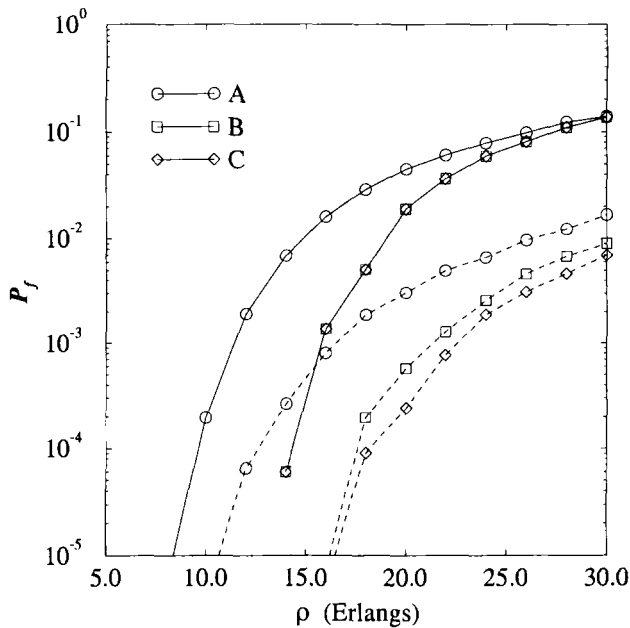
Figure 13.17. Probability of forced termination versus offered traffic without queueing (solid) and a 5-second handoff queue (dashed). Legend: A = FCA, B = SDCA, C = Aggressive SDCA, from [356].

acquisition rate. However, aggressive SDCA causes only a very slight increase in the carrier acquisition rate over non-aggressive SDCA.

The results presented here have been obtained under the assumption that the interference (and DRA) neighborhoods are symmetrical (cell A interferes biconditionally with cell B) and the average traffic loading is identical for all cells. This is not true of a practical system and, therefore, preassigned carrier orderings should not be used. In an actual microcellular system an adaptive, self-organizing algorithm for ordering of carriers and the selection of owned carriers is preferable. Also, an adaptive aggressive strategy may be employed that uses current performance (e.g., the current new call blocking and forced termination probabilities) and perhaps forward-looking strategies [105] to make a more informed decision when to *take* a carrier. It is expected that some performance deterioration will result over the ideal symmetrized case presented here, because of the aforementioned network asymmetries and the finite convergence rate of the adaptive algorithms.

Finally, the channel assignment strategies as described do not take into consideration the arrangement of calls on the carriers. In reality, the C/I is not the same for each channel. A more effective strategy also arranges the calls in order to combat the unpredictable signal and interference variations
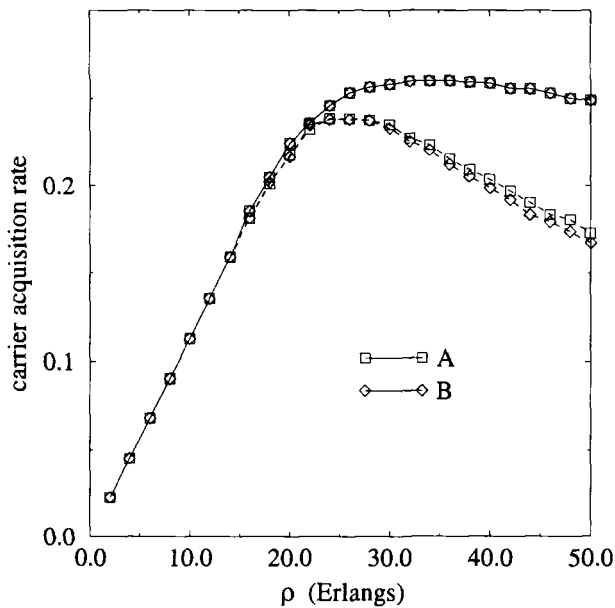
Figure 13.18. Carrier acquisitions per second (per cell) versus offered traffic without queueing (solid) and a 5-second handoff queue (dashed). Legend: A = SDCA, B = Aggressive SDCA, from [356].

present in microcells. Finally, the use of hand-off queueing will exaggerate the cell boundaries thereby causing increased co-channel interference. Unlike the purely statistical model that is used here, the study of these issues will require explicit models for the mobility of MSs and the radio propagation environment.

11.   CONCLUDING REMARKS

Although it is very difficult to arrive at a consensus as to what the best channel assignment algorithm is, an effective DCA algorithm should possess distributed control mechanisms, handoff prioritization, high channel utilization, and stability. Unfortunately, there is no single DCA algorithm that combines all these features and the *best* solution is sure to depend on the service area characteristics. For example, cordless phones require a fully decentralized algorithm while urban microcells should allow some limited communication among BSs. Although some very interesting DCA schemes have been proposed in the literature, not all the issues have been sufficiently addressed to make them practical. Many of the current systems either require too much computation and communication among BSs, yield low channel utilization, or exhibit instability. As a result, DCA will be an active area of research for some time.

In general, the analytical treatment of DCA algorithms is quite difficult and few results have appeared in the literature. Most of DCA algorithms are derived on an ad hoc basis and evaluated by computer simulation. The development of new analytical tools is important for systematic development and will yield valuable insight into the performance of new DCA algorithms.

Much of the existing literature has separated the handoff problem from the channel assignment problem. However, these two problems linked and it is desirable that they receive unified treatment. For example, one performance measure for a handoff algorithm is the mean number of handoffs against the handoff delay. However, such an analysis usually proceeds under the assumption that a channel will always be available for a handoff. Clearly, this is not the case in practice. Most channel assignment schemes are designed for a single application, i.e., voice services. However, future systems will have to support a variety of multimedia applications that have different GOS requirements and require different types and amounts of network resources including channel resources, delay, etc. The channel assignment problem for multimedia services is an open area for research.

Problems

11.1. Suppose that the maximum packing (MP) policy is used with the system shown in Fig. 13.2. Suppose that 10 channels are available for use within each of the three reuse clusters $\text{CL}_i, i = 1, 2, 3$.

a) Compute the number of admissible states $|\mathcal{S}|$.

b) By using the approximation in (13.6) compute the approximate blocking probabilities for each cell assuming a traffic load of $\rho = 2$ Erlangs in each cell.

c) Compare the blocking probabilities in part b) with FCA for the same traffic load.

11.2. Show that the maximum and minimum saturation densities for a linear array of cells is $C_{\max} = (R+1)^{-1}$ and $C_{\min} = (2R+1)^{-1}$.

11.3. Show that the maximum and minimum saturation densities for a planar array of cells are

$$C_{\min} = \frac{1}{1 + 3R(R+1)}$$

$$C_{\max} = \begin{cases} \frac{4}{3(R+1)^2} & , \quad R \text{ even} \\ \frac{4}{1+3(R+1)^2} & , \quad R \text{ odd} \end{cases}.$$

11.4. Consider a linear array of cells with $R = 1$ and a total of $N_T = 24$ channels. Plot the blocking probability, P_b, against the offered traffic per

cell, ρ, with FCA, timid DCA, and aggressive DCA. What conclusions can you make?

11.5. Derive equation (13.12).

Appendix A
Probability and Random Processes

The theory of probability and random processes is essential in the design and performance analysis of communication systems. This Appendix presents a brief review of the basic concepts of probability theory and random processes. It is intended that most readers have already had some exposure to probability and random processes, so that this Appendix is intended to provide a brief overview. A very thorough treatment of this subject is available in a large number of textbooks, including [256], [196].

1. CONDITIONAL PROBABILITY AND BAYES' THEOREM

Let A and B be two events in a sample space S. The **conditional probability** of A given B is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{A.1}$$

provided that $P(B) \neq 0$. If $P(B) = 0$, then $P(A|B)$ is undefined. There are several special cases.

- If $A \cap B = \emptyset$, then events A and B are **mutually exclusive**, i.e., if B occurs then A could not have occurred and $P(A|B) = 0$.

- If $B \subset A$, then knowledge that event B has occurred implies that event A has occurred and so $P(A|B) = 1$.

- If A and B are **statistically independent**, then $P(A \cap B) = P(A)P(B)$ and so $P(A|B) = P(A)$.

There is a strong connection between mutually exclusive and independent events. It may seem that mutually exclusive events are independent, but just

the exact opposite is true. Consider two events A and B with $P(A) > 0$ and $P(B) > 0$. If A and B are mutually exclusive, then $A \cap B = 0$ and $P(A \cap B) = 0 \neq P(A)P(B)$. Therefore, mutually exclusive events with non-zero probability cannot be independent. Thus disjointness of events is a property of the events themselves, while independence is a property of their probabilities.

In general, the events $A_i, i = 1,\ldots, n$, are independent if and only if for all collections of k distinct integers $(i_1, i_2, \ldots, i_k)$ chosen from the set $(1, 2, \ldots, n)$, we have

$$P\left(A_{i_1} \bigcap A_{i_2} \bigcap \cdots \bigcap A_{i_k}\right) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k})$$

for $2 \leq k \leq n$.

In summary

- If $A_i, i = 1, \ldots, n$ is a sequence of mutually exclusive events, then

$$P\left(\bigcup_{i=1}^{n}\right) = \sum_{i=1}^{n} P(A_i) . \tag{A.2}$$

- If $A_i, i = 1, \ldots, n$ is a sequence of independent events, then

$$P\left(\bigcap_{i=1}^{n}\right) = \prod_{i=1}^{n} P(A_i) . \tag{A.3}$$

Total Probability. The collection of sets $\{B_i\}$, $i = 1, \ldots, n$ forms *a partition* of the sample space S if $B_i \bigcap B_j = \emptyset$, $i \neq j$ and $\bigcup_{i=1}^{n} B_i = S$. For any event $A \subset S$ we can write

$$A = \bigcup_{i=1}^{n} (A \bigcap B_i) . \tag{A.4}$$

That is, every element of A is contained in one and only one B_i. Since $(A \bigcap B_i) \bigcap (A \bigcap B_j) = \emptyset$, $i \neq j$, the sets $A \bigcap B_i$ are mutually exclusive. Therefore,

$$
\begin{aligned}
P(A) &= \sum_{i=1}^{n} P(A \bigcap B_i) \\
&= \sum_{i=1}^{n} P(A|B_i)P(B_i) . \tag{A.5}
\end{aligned}
$$

This last equation is often referred to as total probability.

Bayes' Theorem. Let the events B_i, $i = 1, \ldots, n$ be mutually exclusive such that $\bigcup_{i=1}^{n} B_i = S$, where S is the sample space. Let A be an event with nonzero probability. Then as a result of conditional probability and total probability:

$$
\begin{aligned}
P(B_i|A) &= \frac{P(B_i \cap A)}{P(A)} \\[2mm]
&= \frac{P(A|B_i)P(B_i)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)} \quad .
\end{aligned}
$$

a result known as Bayes' theorem.

2. MEANS, MOMENTS, AND MOMENT GENERATING FUNCTIONS

The kth **moment** of a random variable, $E[X^k]$, is defined as

$$
E[X^k] \triangleq \begin{cases} \sum_{x_i \in R_X} x_i^k p_X(x_i) & \text{if } X \text{ is discrete} \\[3mm] \int_{R_X} x^k p_X(x) \mathrm{d}x & \text{if } X \text{ is continuous} \end{cases} \tag{A.6}
$$

where $p_X(x_i) \triangleq P(X = x_i)$ is the **probability distribution function** of X, and $p_X(x) \triangleq P(X = x)$ is the **probability density function** (pdf) of X. The kth **central moment** of the random variable X is $E[(X - E[X])^k]$. The **variance** is the second central moment.

The **moment generating function or characteristic function** of a random variable X is

$$
\psi_X(jv) \triangleq E[e^{jvX}] = \begin{cases} \sum_{x_i \in R_X} e^{jvx_i} p_X(x_i) & \text{if } X \text{ is discrete} \\[3mm] \int_{R_X} e^{jvx} p_X(x) \mathrm{d}x & \text{if } X \text{ is continuous} \end{cases} \tag{A.7}
$$

where $j = \sqrt{-1}$. Note that the continuous version is a Fourier transform, except for the sign in the exponent. Likewise, the discrete version is a z-transform, except for the sign in the exponent.

The probability distribution and probability density functions of discrete and continuous random variables, respectively, can be obtained by taking the inverse transforms of the characteristic functions, i.e.,

$$
p_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \psi_X(jv) e^{-jvx} \, dv \tag{A.8}
$$

and

$$
p_X(x_k) = \frac{1}{2\pi} \oint_C \psi_X(jv) e^{-jvx_k} \, dv \quad . \tag{A.9}
$$

The **cumulative distribution function** (cdf) of a random variable X is defined as

$$F_X(x) \triangleq P(X \leq x) = \begin{cases} \sum_{x_i \leq x} p_X(x_i) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{x} p_X(x)dx & \text{if } X \text{ is continuous} \end{cases} \quad . \quad \text{(A.10)}$$

The **complementary distribution function** (cdfc) is defined as

$$F_X^c(x) \triangleq 1 - F_X(x) \quad . \quad \text{(A.11)}$$

3.   SOME USEFUL PROBABILITY DISTRIBUTIONS
3.1   DISCRETE DISTRIBUTIONS

Binomial Distribution.   Let X be a **Bernoulli random variable** such that $X = 0$ with probability $1 - p$ and $X = 1$ with probability p. Although X is a discrete random random variable with an associated probability distribution function, it is possible to treat X as a continuous random variable with a pdf by using dirac delta functions. In this case, the pdf of X has the form

$$p_X(x) = (1 - p)\delta(x) + p\delta(x - 1) \quad . \quad \text{(A.12)}$$

Let $Y = \sum_{i=1}^{n} X_i$, where the X_i are independent and identically distributed with density $p_X(x)$. Then the random variable Y is an integer from the set $\{0, 1, \ldots, n\}$ and the probability distribution of Y is

$$p_Y(k) = P(Y = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \ldots, n \quad . \quad \text{(A.13)}$$

The random variable Y also has the pdf

$$p_Y(y) = \sum_{k=0}^{n} \binom{n}{k} p^k (1 - p)^{n-k} \delta(y - k) \quad . \quad \text{(A.14)}$$

Poisson Distribution.   The random variable X has a **Poisson distribution** if

$$p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, \ldots, \infty \quad \text{(A.15)}$$

Geometric Distribution.   The random variable X has a **geometric distribution** if

$$p_X(k) = (1 - p)^{k-1} p, \quad k = 1, 2, \ldots, \infty \quad . \quad \text{(A.16)}$$

3.2   CONTINUOUS DISTRIBUTIONS

Many communication systems are affected by additive Gaussian noise. Therefore, the Gaussian distribution and various functions of Gaussian distributions play a central role in the characterization and analysis of communication systems.

Gaussian Distribution.   A Gaussian random variable has the pdf

$$p_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{A.17}$$

where $\mu = \mathrm{E}[X]$ is the mean and $\sigma^2 = \mathrm{E}[(X-\mu)^2]$ is the variance. Sometimes we use the shorthand notation $X \sim N(\mu, \sigma^2)$ meaning that X is a Gaussian random variable with mean μ and variance σ^2. The random variable X is said to have a standard normal distribution if $X \sim N(0,1)$.

The cumulative distribution function (cdf) of X is

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy \ . \tag{A.18}$$

The cdf of a standard normal distribution defines the Q function

$$Q(x) \triangleq \int_{x}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \tag{A.19}$$

and the cdfc defines the Φ function

$$\Phi(x) \triangleq 1 - Q(x) \ . \tag{A.20}$$

If X is a non-standard normal random variable, $X \sim N(\mu, \sigma^2)$, then

$$F_X(x) \;=\; \Phi\left(\frac{x-\mu}{\sigma}\right) \tag{A.21}$$

$$F_X^c(x) \;=\; Q\left(\frac{x-\mu}{\sigma}\right) \ . \tag{A.22}$$

Quite often the cumulative distribution function of a Gaussian random variable is described in terms of the **error function** erf(x) and the **complementary error function** erfc(x), defined by

$$\mathrm{erfc}(x) \triangleq \frac{2}{\sqrt{\pi}} \int_{x}^{\infty} e^{-y^2} dy \triangleq 1 - \mathrm{erf}(x) \ . \tag{A.23}$$

The complementary error function and Q function are related as follows

$$\mathrm{erfc}(x) \;=\; 2Q(\sqrt{2}(x)) \tag{A.24}$$

$$Q(x) \;=\; \frac{1}{2}\mathrm{erfc}\left(\frac{x}{\sqrt{2}}\right) \ . \tag{A.25}$$

Rayleigh Distribution. Let $X_1 \sim N(0, \sigma^2)$ and $X_2 \sim N(0, \sigma^2)$ be independent normal random variables. The random variable $R = \sqrt{X_1^2 + X_2^2}$ is said to be Rayleigh distributed. To find the pdf and cdf of R first define

$$V = \text{Tan}^{-1}\left(\frac{X_2}{X_1}\right) .$$

Then

$$\begin{aligned} X_1 &= R\cos V \\ X_2 &= R\sin V . \end{aligned}$$

By using a bivariate transformation of random variables

$$p_{RV}(r, v) = p_{X_1 X_2}(r\cos v, r\sin v)\,|J(r, v)|$$

where

$$J(r, v) = \begin{vmatrix} \frac{\partial x_1}{\partial r} & \frac{\partial x_1}{\partial v} \\ \frac{\partial x_2}{\partial r} & \frac{\partial x_2}{\partial v} \end{vmatrix} = \begin{vmatrix} \cos v & r\sin v \\ \sin v & r\cos v \end{vmatrix} = r(\cos^2 v + \sin^2 v) = r$$

Since

$$p_{X_1 X_2}(x_1, x_2) = \frac{1}{2\pi\sigma^2} e^{-\frac{x_1^2 + x_2^2}{2\sigma^2}}$$

we have

$$p_{RV}(r, v) = \frac{r}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}} . \tag{A.26}$$

It follows that the marginal pdf of R is

$$\begin{aligned} p_R(r) &= \int_0^{2\pi} p_{RV}(r, v)\,dv \\ &= \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} \qquad r \geq 0 . \end{aligned} \tag{A.27}$$

The cdf is

$$F_R(r) = 1 - e^{-\frac{r^2}{2\sigma^2}} \quad r \geq 0 . \tag{A.28}$$

Rice Distribution. Let $X_1 \sim N(\mu_1, \sigma^2)$ and $X_2 \sim N(\mu_2, \sigma^2)$ be independent normal random variables with non-zero means. The random variable $R = \sqrt{X_1^2 + X_2^2}$ has a Rice distribution or is said to be Ricean distributed. To find the pdf and cdf of R again define $V = \text{Tan}^{-1}(X_2/X_1)$. Then by using a bivariate transformation $J(r, v) = r$ and

$$p_{RV}(r, v) = r \cdot p_{X_1 X_2}(r\cos v, r\sin v) . \tag{A.29}$$

However,

$$
\begin{aligned}
p_{X_1 X_2}(x_1, x_2) &= \frac{1}{2\pi\sigma^2} \exp\left\{ -\frac{(x_1 - \mu_1)^2 + (x_2 - \mu_2)^2}{2\sigma^2} \right\} \\
&= \frac{1}{2\pi\sigma^2} \exp\left\{ -\frac{x_1^2 + x_2^2 + \mu_1^2 + \mu_2^2 - 2(x_1\mu_1 + x_2\mu_2)}{2\sigma^2} \right\} .
\end{aligned}
$$

Hence,

$$
p_{RV}(r, v) = \frac{1}{2\pi\sigma^2} \exp\left\{ -\frac{r^2 + \mu_1^2 + \mu_2^2 - 2r(\mu_1 \cos v + \mu_2 \sin v)}{2\sigma^2} \right\} .
$$

Now define $s \triangleq \sqrt{\mu_1^2 + \mu_2^2}$ and $t \triangleq \mathrm{Tan}^{-1}\mu_2/\mu_1$, $0 \le t \le 2\pi$, so that $\mu_1 = s \cos t$ and $\mu_2 = s \sin t$. Then

$$
\begin{aligned}
p_{RV}(r, v) &= \frac{1}{2\pi\sigma^2} \exp\left\{ -\frac{r^2 + s^2 - 2rs(\cos t \cos v + \sin t \sin v)}{2\sigma^2} \right\} \\
&= \frac{1}{2\pi\sigma^2} \exp\left\{ -\frac{r^2 + s^2 - 2rs \cos(v - t)}{2\sigma^2} \right\} .
\end{aligned}
$$

The marginal pdf of R is

$$
P_R(r) = \frac{r}{\sigma^2} e^{-\frac{(r^2 + s^2)}{2\sigma^2}} \cdot \frac{1}{2\pi} \int_0^{2\pi} e^{-\frac{rs}{\sigma^2} \cos(v - t)} dv . \tag{A.30}
$$

The zero order modified Bessel function of the first kind is defined as

$$
I_0(x) \triangleq \frac{1}{2\pi} \int_0^{2\pi} e^{-x \cos\theta} d\theta . \tag{A.31}
$$

Therefore,

$$
P_R(r) = \frac{r}{\sigma^2} e^{-\frac{(r^2 + s^2)}{2\sigma^2}} \cdot \frac{1}{2\pi} I_0\left(\frac{rs}{\sigma^2}\right), \qquad r \ge 0 . \tag{A.32}
$$

The cdf of R is

$$
\begin{aligned}
F_R(r) &= \int_0^r p_R(r) dr \\
&= 1 - Q\left(\frac{s}{\sigma}, \frac{r}{\sigma}\right)
\end{aligned}
$$

where $Q(a, b)$ is called the Marcum Q-function.

Central Chi-Square Distribution. Let $X \sim N(0, \sigma^2)$ and $Y = X^2$. Then it can be shown that

$$
p_Y(y) = \frac{p_X(\sqrt{y}) + p_X(-\sqrt{y})}{2\sqrt{y}}
$$

$$
= \frac{1}{\sqrt{2\pi y}\sigma} e^{-\frac{y}{2\sigma^2}} \qquad y \geq 0 \ .
$$

The characteristic function of Y is

$$
\psi_Y(jv) = \int_{-\infty}^{\infty} e^{jvy} p_Y(y) dy
$$

$$
= \frac{1}{\sqrt{1 - j2v\sigma^2}} \ . \tag{A.33}
$$

Now define the random variable $Y = \sum_{i=1}^{n} X_i^2$, where the X_i are independent and $X_i \sim N(0, \sigma^2)$. Then

$$
\psi_Y(jv) = \frac{1}{(1 - j2v\sigma^2)^{n/2}} \ . \tag{A.34}
$$

Taking the inverse transform gives

$$
p_Y(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \psi_Y(jv) e^{-jvy} dv
$$

$$
= \frac{1}{(2\sigma^2)^{n/2}\Gamma(n/2)} y^{n/2-1} e^{-\frac{y}{2\sigma^2}}, \quad y \geq 0 \ .
$$

where $\Gamma(k)$ is the Gamma function and

$$
\Gamma(k) = \int_0^{\infty} u^{k-1} e^{-u} du = (k-1)!
$$

if k is a positive integer. If n is even (which is usually the case in practice) and we define $m = n/2$, then the pdf of Y defines the central chi-square distribution

$$
p_Y(y) = \frac{1}{(2\sigma^2)^m (m-1)!} y^{m-1} e^{-\frac{y^2}{2\sigma^2}} \qquad y \geq 0 \tag{A.35}
$$

and the cdf of Y is

$$
F_Y(y) = 1 - e^{-\frac{y}{2\sigma^2}} \sum_{k=0}^{m-1} \frac{1}{k!} \left(\frac{y}{2\sigma^2} \right)^k \qquad y \geq 0 \ . \tag{A.36}
$$

The **exponential distribution** is a special case of the central chi-square distribution when $m = 1$. In this case

$$
p_Y(y) = \frac{1}{2\sigma^2} e^{-\frac{y}{2\sigma^2}}
$$

$$
F_Y(y) = 1 - e^{-\frac{y}{2\sigma^2}} \ . \tag{A.37}
$$

Non-Central Chi-Square Distribution.   Let $X \sim N(\mu, \sigma^2)$ and $Y = X^2$. Then

$$p_Y(y) = \frac{p_X(\sqrt{y}) + p_X(-\sqrt{y})}{2\sqrt{y}}$$

$$= \frac{1}{\sqrt{2\pi y}\sigma} e^{-\frac{(y+\mu^2)}{2\sigma^2}} \cosh\left(\frac{\sqrt{y}\mu}{\sigma^2}\right), \quad y \geq 0 .$$

The characteristic function of Y is

$$\psi_Y(jv) = \int_{-\infty}^{\infty} e^{jvy} p_Y(y) dy$$

$$= \frac{1}{\sqrt{1 - j2v\sigma^2}} \exp\left\{\frac{j\mu^2 v}{1 - j2v\sigma^2}\right\} .$$

Now define the random variable $Y = \sum_{i=1}^{n} X_i^2$, where the X_i are independent normal random variables and $X_i \sim N(\mu_i, \sigma^2)$. Then

$$\psi_Y(jv) = \frac{1}{(1 - j2v\sigma^2)^{n/2}} \exp\left\{\frac{jv \sum_{i=1}^{n} \mu_i^2}{1 - j2v\sigma^2}\right\} .$$

Taking the inverse transform gives

$$p_Y(y) = \frac{1}{2\sigma^2} \left(\frac{y}{s^2}\right)^{\frac{n-2}{4}} e^{-\frac{(s^2+y)}{2\sigma^2}} I_{n/2-1}\left(\sqrt{y}\frac{s}{\sigma^2}\right) \quad y \geq 0$$

where

$$s = \sum_{i=1}^{n} \mu_i^2$$

and $I_k(x)$ is the modified Bessel function of the first kind and order k, defined by

$$I_k(x) \triangleq \frac{1}{2\pi} \int_0^{2\pi} e^{x \cos \theta} \cos(n\theta) d\theta .$$

If n is even (which is usually the case in practice) and we define $m = n/2$, then the pdf of Y is

$$p_Y(y) = \frac{1}{2\sigma^2} \left(\frac{y}{s^2}\right)^{\frac{m-1}{2}} e^{-\frac{(s^2+y)}{2\sigma^2}} I_{m-1}\left(\sqrt{y}\frac{s}{\sigma^2}\right), \quad y \geq 0 \qquad \text{(A.38)}$$

and the cdf of Y is

$$F_Y(y) = 1 - Q_m\left(\frac{s}{\sigma}, \frac{\sqrt{y}}{\sigma}\right) \qquad \text{(A.39)}$$

where $Q_m(a, b)$ is called the generalized Q-function.

Multivariate Gaussian Distribution. Let $X_i \sim N(\mu_i, \sigma_i^2), i = 1, \ldots, n$, be correlated Gaussian random variables covariances

$$
\begin{aligned}
\mu_{X_i X_j} &= E\left[(X_i - \mu_i)(X_j - \mu_j)\right] \\
&= E[X_i X_j] - \mu_i \mu_j, \quad 1 \le i, j \le n .
\end{aligned}
$$

Let

$$
\begin{aligned}
\mathbf{X} &= (X_1, X_2, \ldots, X_n)^T \\
\boldsymbol{\mu}_X &= (\mu_1, \mu_2, \ldots, \mu_n)^T \\
\boldsymbol{\Lambda} &= \begin{bmatrix} \mu_{X_1 X_1} & \cdot & \cdot & \cdot & \cdot & \mu_{X_1 X_n} \\ \vdots & & & & & \vdots \\ \mu_{X_n X_1} & \cdot & \cdot & \cdot & \cdot & \mu_{X_n X_n} \end{bmatrix}
\end{aligned}
$$

where $\mathbf{X}^T$ is the transpose of $\mathbf{X}$. Then the joint pdf of $\mathbf{X}$ defines the multivariate Gaussian distribution

$$
p_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} |\boldsymbol{\Lambda}|^{1/2}} \exp\left\{ -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_X)^T \boldsymbol{\Lambda}^{-1}(\mathbf{x} - \boldsymbol{\mu}_X) \right\} \qquad \text{(A.40)}
$$

where $|\boldsymbol{\Lambda}|$ is the determinant of $\boldsymbol{\Lambda}$.

4.   UPPER BOUNDS ON THE CDFC

Several different approaches can be used to upper bound the area under the tails of a probability density function including the Chebyshev and Chernoff bounds.

Chebyshev Bound.   The Chebyshev bound is derived as follows. Let X be a random variable with mean μ_X, variance σ_X^2, and pdf $p_X(x)$.   Then the variance of X is

$$
\begin{aligned}
\sigma_X^2 &= \int_{-\infty}^{\infty} (x - \mu_X)^2 p_X(x) dx \\
&\ge \int_{|x - \mu_X| \ge \delta} (x - \mu_X)^2 p_X(x) dx \\
&\ge \delta^2 \int_{|x - \mu_X| \ge \delta} p_X(x) dx \\
&= \delta^2 P(|X - \mu_X| \ge \delta) .
\end{aligned}
$$

Hence,

$$
P(|X - \mu_X| \ge \delta) \le \frac{\sigma_X^2}{\delta^2} . \qquad \text{(A.41)}
$$

The Chebyshev bound is straightforward to apply but it tends to be quite loose.

Chernoff Bound. The Chernoff bound is more difficult to compute but is much tighter than the Chebyshev bound. To derive the Chernoff bound we use the following inequality

$$u(x) \le e^{\lambda x}, \qquad \forall \, x \text{ and } \forall \, \lambda \ge 0$$

where $u(x)$ is the unit step function. Then,

$$
\begin{aligned}
P(X \ge 0) &= \int_0^\infty p_X(x)dx \\
&= \int_{-\infty}^\infty u(x)p_X(x)dx \\
&\le \int_{-\infty}^\infty e^{\lambda x}p_X(x)dx \\
&= E[e^{\lambda x}] .
\end{aligned}
$$

The **Chernoff bound parameter** $\lambda > 0$ can be optimized to give the tightest upper bound. This can be accomplished by setting the derivative to zero

$$\frac{d}{d\lambda}E[e^{\lambda x}] = E\left[\frac{d}{d\lambda}e^{\lambda x}\right] = E[xe^{\lambda x}] = 0 .$$

Let $\lambda^* = \arg\min_{\lambda \ge 0} E[e^{\lambda x}]$ be the solution to the above equation. Then

$$P(X \ge 0) \le E[e^{\lambda^* x}] . \tag{A.42}$$

Example   A.1───
Let X_i, $i = 1, \ldots, n$ be independent and identically distributed random variables with density

$$p_X(x) = p\delta(x - 1) + (1 - p)\delta(x + 1) .$$

Let

$$Y = \sum_{i=1}^n X_i .$$

Then

$$
\begin{aligned}
P(Y \ge 0) &= P\left(\lceil n/2\rceil \text{ or more of the } X_i \text{ are ones }\right) \\
&= \sum_{k=\lceil n/2\rceil}^n \binom{n}{k} p^k(1 - p)^{n-k} .
\end{aligned}
$$

For $n = 10$ and $p = 0.1$

$$P(Y \ge 0) = 0.0016349 . \tag{A.43}$$

Chebyshev Bound

To compute the Chebyshev bound we first determine the mean and variance of Y.

$$
\begin{aligned}
\mu_Y &= n\mathrm{E}[X_i] \\
&= n[p - 1 + p] \\
&= n(2p - 1) \ .
\end{aligned}
$$

$$
\begin{aligned}
\sigma_Y^2 &= n\sigma_X^2 \\
&= n\left(\mathrm{E}[X_i^2] - \mathrm{E}^2[X_i]\right) \\
&= n\left(1 - (2p - 1)^2\right) \\
&= n\left(1 - 4p^2 + 4p - 1\right) \\
&= 4np(1 - p) \ .
\end{aligned}
$$

Hence,

$$
\mathrm{P}(|Y - \mu_Y| \geq \mu_Y) \leq \frac{\sigma_Y^2}{\mu_Y^2} = \frac{4np(1 - p)}{n^2(2p - 1)^2} \ .
$$

Then by symmetry

$$
\begin{aligned}
\mathrm{P}(Y \geq 0) &= \frac{1}{2}\mathrm{P}(|Y - \mu_Y| \geq \mu_Y) \\
&\leq \frac{2p(1 - p)}{n(2p - 1)^2} \ .
\end{aligned}
$$

For $n = 10$ and $p = 0.1$

$$
P(Y \geq 0) \leq 0.028125 \ . \tag{A.44}
$$

Chernoff Bound

The Chernoff bound is given by

$$
\begin{aligned}
P(Y \geq 0) &\leq \mathrm{E}[e^{\lambda y}] \\
&= \left(\mathrm{E}[e^{\lambda x}]\right)^n \ .
\end{aligned}
$$

However,

$$
\mathrm{E}[e^{\lambda x}] = pe^\lambda - (1 - p)e^{-\lambda} \ .
$$

To find the optimal Chernoff bound parameter we solve

$$
\frac{d}{d\lambda}\mathrm{E}[e^{\lambda x}] = pe^\lambda - (1 - p)e^{-\lambda} = 0
$$

giving

$$\lambda^* = \ln\left(\sqrt{\frac{1-p}{p}}\right) \ .$$

Hence,

$$\begin{aligned} \mathrm{P}(Y \geq 0) &\leq \left(\mathrm{E}[e^{\lambda^* x}]\right)^n \\ &= [4p(1-p)]^{n/2} \ . \end{aligned}$$

For $n = 10$ and $p = 0.1$

$$P(Y \geq 0) \leq 0.0060466 \ .$$

Notice that the Chernoff bound is much tighter that the Chebyshev bound in this case.

5. RANDOM PROCESSES

A random process, or stochastic process, $X(t)$, is an ensemble of sample functions $\{X_1(t), X_2(t), \ldots, X_\xi(t)\}$ together with a probability rule which assigns a probability to any event associated with the observation of these functions. Consider the set of sample functions shown in Fig. A. 1. The sample function x_i corresponds to the sample point s_1 in the sample space and, occurs with probability $\mathrm{P}(s_1)$. The number of sample functions, ξ, in the ensemble may be finite or infinite. The function $X_i(t)$ is deterministic once the index i is known. Sample functions may be defined at discrete or continuous instants in time, and their values (parameters) at these time instants may be discrete or continuous in time also.

Suppose that we observe all the sample functions at some time instant t_1. and their values form the set of numbers $\{X_i(t_1)\}, i = 1, 2, \ldots, \xi$. Since $X_i(t_1)$ occurs with probability $\mathrm{P}(s_i)$, the collection of numbers $\{X_i(t_1)\}, i = 1, 2, \ldots, \xi$ forms a random variable, denoted by $X(t_1)$. By observing the set of waveforms at another time instant t_2 we obtain a different random variable $X(t_2)$. A collection of n such random variables, $X(t_1)$, $X(t_2)$, ..., $X(t_n)$, has the joint cdf

$$F_{X(t_1), \ldots, X(t_n)}(x_1, \ldots, x_n) = \mathrm{P}(X(t_1) < x_1, \ldots, X(t_n) < t_n) \ .$$

A more compact notation can be obtained by defining the vectors

$$\begin{aligned} \mathbf{x} &= (x_1, x_2, \ldots, x_n) \\ \mathbf{X}(t) &= (X(t_1), X(t_2), \ldots, X(t_n)) \ . \end{aligned}$$
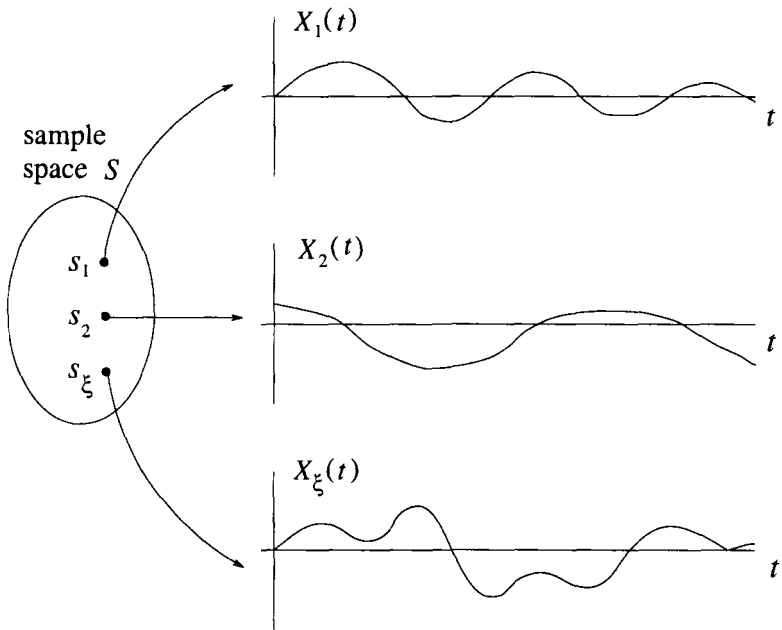
Figure A.1. Ensemble of sample functions for a random process.

Then the joint cdf and joint pdf are, respectively,

$$F_{\mathbf{X}(t)}(\mathbf{x}) = P(\mathbf{X}(t) \le \mathbf{x}) \tag{A.45}$$

$$p_{\mathbf{X}(t)}(\mathbf{x}) = \frac{\partial^n F_{\mathbf{X}(t)}(\mathbf{x})}{\partial x_1 \partial x_2 \cdots \partial x_n} . \tag{A.46}$$

A random process is **strictly stationary** if and only if the joint density function $p_{\mathbf{X}(t)}(\mathbf{x})$ is invariant under shifts of the time origin. In this case, the equality

$$p_{\mathbf{X}(t)}(\mathbf{x}) = p_{\mathbf{X}(t+\tau)}(\mathbf{x}) \tag{A.47}$$

holds for all sets of time instants $\{t_1, t_2, \ldots, t_n\}$ and all time shifts τ. Many important random processes that are encountered in practice are strictly stationary.

5.1   MOMENTS AND CORRELATION FUNCTIONS

To describe the moments and correlation functions of a random process, it is useful to define the following two operators

$$E[\,\cdot\,] \triangleq \text{ensemble average}$$

$$< \,\cdot\, > \triangleq \text{time average} .$$

The ensemble average of a random process at time t is

$$\mu_X(t) \equiv \mathrm{E}[X(t)] = \int_{-\infty}^{\infty} x p_{X(t)}(x) dx \quad . \tag{A.48}$$

The time average of a random process is

$$< X(t) >= \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} X(t) dt \quad . \tag{A.49}$$

In general, the time average $< X(t) >$ is also a random variable, because it depends on the particular sample function that is selected for time averaging.

The **autocorrelation** of a random process $X(t)$ is defined as

$$\phi_{XX}(t_1, t_2) = \mathrm{E}[X(t_1)X(t_2)] \quad . \tag{A.50}$$

The **autocovariance** of a random process $X(t)$ is defined as

$$\begin{aligned} \mu_{XX}(t_1, t_2) &= \mathrm{E}[(X(t_1) - \mu_X(t_1))(X(t_2) - \mu_X(t_2))] \\ &= \phi_{XX}(t_1, t_2) - \mu_X(t_1)\mu_X(t_2) \quad . \end{aligned} \tag{A.51}$$

A random process is strictly stationary only if

$$\mathrm{E}[X^n(t)] = \mathrm{E}[X^n] \quad \forall\, t, n \quad .$$

Hence, for a strictly stationary random process

$$\begin{aligned} \mu_X(t) &= \mu \\ \sigma_X^2(t) &= \sigma_X^2 \\ \phi_{XX}(t_1, t_2) &= \phi_{XX}(t_1 - t_2) \equiv \phi_{XX}(\tau) \\ \mu_{XX}(t_1, t_2) &= \mu_{XX}(t_1 - t_2) \equiv \mu_{XX}(\tau) \end{aligned}$$

where $\tau = t_1 - t_2$.

If a random process satisfies the following conditions

$$\begin{aligned} \mu_X(t) &= \mu_X \\ \mu_{XX}(t_1, t_2) &= \mu_{XX}(\tau) \end{aligned}$$

then it is said to be **wide sense stationary**. Note that a strictly stationary random process is always wide sense stationary, but the converse may not be true. However, a Gaussian random process is strictly stationary if it is wide sense stationary. The reason is that a joint Gaussian density of the vector $\mathbf{X}(t) = (X(t_1), X(t_2), \dots, X(t_n))$ is completely described by the means and covariances of the $X(t_i)$.

Properties of $\phi_{XX}(\tau)$. The autocorrelation function, $\phi_{XX}(\tau)$, of a stationary random process satisfies the following properties.

1. $\phi_{XX}(0) = E[X^2(t)]$

2. $\phi_{XX}(\tau) = \phi_{XX}(-\tau)$

3. $|\phi_{XX}(\tau)| \le \phi_{XX}(0)$

4. $\phi_{XX}(\infty) = E^2[X(t)] = \mu_X^2$

5. If $X(t) = X(t+\tau)$, then $\phi_{XX}(\tau) = \phi_{XX}(\tau+T)$, i.e., if $X(t)$ is periodic, then $\phi_{XX}(\tau)$ is periodic.

A random process is **ergodic** if for all $g(\mathbf{X})$ and $\mathbf{X}$

$$
\begin{aligned}
E[g(\mathbf{X})] &= \int_{-\infty}^{\infty} g(\mathbf{X}) P_{\mathbf{X}(t)}(\mathbf{x}) d\mathbf{x} \\
&= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} g[\mathbf{X}(t)] dt \\
&= \; < g[\mathbf{X}(t)] > \; . \qquad\qquad\text{(A.52)}
\end{aligned}
$$

For a random process to be ergodic, it must be strictly stationary. However, not all strictly stationary random processes are ergodic. A random process is **ergodic in the mean** if

$$
P(< X(t) > = \mu_X) = 1
$$

and ergodic in the autocorrelation if

$$
P(< X(t+\tau)X(t) > = \phi_{XX}(\tau)) = 1 \; .
$$

Example A.2

Consider the random process

$$
X(t) = A\cos(2\pi f_c t + \Theta)
$$

where A and f_c are constants, and

$$
p_\Theta(\theta) = \begin{cases} 1/(2\pi), & 0 \le \theta \le 2\pi \\ 0, & \text{elsewhere} \end{cases} \; .
$$

The mean of $X(t)$ is

$$
\mu_X(t) = E_\Theta[A\cos(2\pi f_c t + \theta)] = 0 = \mu_X
$$

and autocorrelation of X (t) is

$$
\begin{aligned}
\phi_{XX}(t_1, t_2) &= \mathrm{E}_\Theta[X(t_1)X(t_2)] \\
&= \mathrm{E}_\Theta[A^2 \cos(2\pi f_c t_1 + \theta) \cos(2\pi f_c t_2 + \theta)] \\
&= \frac{A^2}{2} \mathrm{E}_\Theta[\cos(2\pi f_c t_1 + 2\pi f_c t_2 + 2\theta)] \\
&\quad + \frac{A^2}{2} \mathrm{E}_\Theta[\cos(2\pi f_c(t_1 - t_2))] \\
&= \frac{A^2}{2} \cos 2\pi f_c(t_1 - t_2) \\
&= \frac{A^2}{2} \cos 2\pi f_c \tau, \quad \tau = t_1 - t_2 \;.
\end{aligned}
$$

It is clear that this random process is wide sense stationary.
   The time-average mean of $X(t)$ is

$$
< X(t) >= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} A \cos(2\pi f_c t + \theta)dt = 0
$$

and the time average autocorrelation of $X(t)$ is

$$
\begin{aligned}
&< X(t+\tau)X(t) > \\
&= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} A^2 \cos(2\pi f_c t + 2\pi f_c \tau + \theta) \cos(2\pi f_c t + \theta)dt \\
&= \lim_{T\to\infty} \frac{A^2}{4T} \int_{-T}^{T} A^2 \left[\cos(2\pi f_c \tau) \cos(4\pi f_c t + 2\pi f_c \tau + \theta)\right] dt \\
&= \frac{A^2}{2} \cos(2\pi f_c \tau) \;.
\end{aligned}
$$

By comparing the ensemble and time average mean and autocorrelation, we can conclude this random process is ergodic in the mean and ergodic in the autocorrelation.

Example A.3———————————————————————————————
   In this example we show that $|\phi_{XX}(\tau)| \le \phi_{XX}(0)$. This inequality can be established through the following steps.

$$
\begin{aligned}
0 &\le \mathrm{E}[(X(t+\tau) \pm X(t))^2] \\
&= \mathrm{E}[X^2(t) + X^2(t+\tau) \pm X(t+\tau)X(t)] \\
&= \mathrm{E}[X^2(t)] + \mathrm{E}[X^2(t+\tau)] \pm \mathrm{E}[X(t+\tau)X(t)] \\
&= 2\mathrm{E}[X^2(t)] \pm \mathrm{E}[X(t+\tau)X(t)] \\
&= 2\phi_{XX}(0) \pm 2\phi_{XX}(\tau) \;.
\end{aligned}
$$

Therefore,

$$\pm \phi_{XX}(\tau) \leq \phi_{XX}(0)$$
$$|\phi_{XX}(\tau)| \leq \phi_{XX}(0) \ .$$

Example A.4────────────────────────────────────

Consider the random process

$$Y(t) = X \cos t, \quad X \sim N(0, 1) \ .$$

In this example we will find the probability density function of $Y(0)$, the joint probability density function of $Y(0)$ and $Y(\pi)$, and determine whether or not $Y(t)$ is stationary.

1. To find the probability density function of $Y(0)$, note that

$$Y(0) = X \cos 0 = X \ .$$

Therefore,

$$p_{Y(0)}(y_0) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y_0^2}{2}} \ .$$

2. To find the joint density of $Y(0)$ and $Y(\pi)$, note that

$$Y(0) = X = -Y(\pi) \ .$$

Therefore

$$p_{Y(0)|Y(\pi)}(y_0|y_\pi) = \delta(y_0 + y_\pi)$$

and

$$p_{Y(0)Y(\pi)}(y_0, y_\pi) = p_{Y(0)|Y(\pi)}(y_0|y_\pi) p_{Y(\pi)}(y_\pi)$$
$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{y_\pi^2}{2}} \delta(y_0 + y_\pi) \ .$$

3. To determine whether or not $Y(t)$ is stationary, note that

$$E[Y(t)] = E[X] \cos t = 0$$
$$E[Y^2(t)] = E[X^2] \cos^2 t \ .$$

Since the second moment varies with time, this random process is not stationary.

5.2 CROSSCORRELATION AND CROSSCOVARIANCE

Consider two random processes $X(t)$ and $Y(t)$. The **crosscorrelation** of $X(t)$ and $Y(t)$ is

$$\phi_{XY}(t_1, t_2) = E[X(t_1)Y(t_2)] \tag{A.53}$$
$$\phi_{YX}(t_1, t_2) = E[Y(t_1)X(t_2)] . \tag{A.54}$$

The correlation matrix of $X(t)$ and $Y(t)$ is

$$\phi(t_1, t_2) = \begin{bmatrix} \phi_{XX}(t_1, t_2) & \phi_{XY}(t_1, t_2) \\ \phi_{YX}(t_1, t_2) & \phi_{YY}(t_1, t_2) \end{bmatrix} . \tag{A.55}$$

The **crosscovariance** of $X(t)$ and $Y(t)$ is

$$\mu_{XY}(t_1, t_2) = E[(X(t_1) - \mu_X(t_1))(X(t_2) - \mu_X(t_2))]$$
$$= \phi_{XY}(t_1, t_2) - \mu_X(t_1)\mu_X(t_2) . \tag{A.56}$$

The covariance matrix of $X(t)$ and $Y(t)$ is

$$\mu(t_1, t_2) = \begin{bmatrix} \mu_{XX}(t_1, t_2) & \mu_{XY}(t_1, t_2) \\ \mu_{YX}(t_1, t_2) & \mu_{YY}(t_1, t_2) \end{bmatrix} . \tag{A.57}$$

If $X(t)$ and $Y(t)$ are each wide sense stationary and jointly wide sense stationary, then

$$\phi(t_1, t_2) = \phi(t_1 - t_2) = \phi(\tau) \tag{A.58}$$
$$\mu(t_1, t_2) = \mu(t_1 - t_2) = \mu(\tau) \tag{A.59}$$

where $\tau = t_1 - t_2$.

Properties of $\phi_{XY}(\tau)$. The crosscorrelation function $\phi_{XY}(\tau)$ has the following properties.

1. $\phi_{XY}(\tau) = \phi_{YX}(-\tau)$

2. $|\phi_{XY}(\tau)| \leq \frac{1}{2}[\phi_{XX}(0) + \phi_{YY}(0)]$

3. $|\phi_{XY}(\tau)|^2 \leq \phi_{XX}(0)\phi_{YY}(0)$ if $X(t)$ and $Y(t)$ have zero mean.

Classifications of Random Processes. Two random processes $X(t)$ and $Y(t)$ are said to be

- **uncorrelated** if and only if $\mu_{XY}(\tau) = 0$.

- **orthogonal** if and only if $\phi_{XY}(\tau) = 0$.

- **statistically independent** if and only if

$$p_{\mathbf{X}(t)\mathbf{Y}(t+\tau)}(\mathbf{x},\mathbf{y}) = p_{\mathbf{X}(t)}(\mathbf{x})p_{\mathbf{Y}(t+\tau)}(\mathbf{y}) \ .$$

Furthermore, if $\mu_X = 0$ or $\mu_Y = 0$, then

$$\text{uncorrelated} \quad \longleftrightarrow \quad \text{orthogonal}$$
$$\text{statistically independent} \quad \longrightarrow \quad \text{uncorrelated} \ .$$

Example A.5

Find the autocorrelation function of the random process

$$Z(t) = X(t) + Y(t)$$

where $X(t)$ and $Y(t)$ are wide sense stationary random processes. The autocorrelation function is

$$\begin{aligned}
\phi_{XX}(\tau) &= \mathrm{E}[Z(t+\tau)Z(t)] \\
&= \mathrm{E}\left[(X(t+\tau) + Y(t+\tau))(X(t) + Y(t))\right] \\
&= \phi_{XX}(\tau) + \phi_{YX}(\tau) + \phi_{XY}(\tau) + \phi_{YY}(\tau) \ .
\end{aligned}$$

If $X(t)$ and $Y(t)$ are uncorrelated, then

$$\phi_{YX}(\tau) = \phi_{XY}(\tau) = \mu_X\mu_Y$$

and

$$\phi_{ZZ}(\tau) = \phi_{XX}(\tau) + \phi_{YY}(\tau) + 2\mu_X\mu_Y \ .$$

If $X(t)$ and $Y(t)$ are uncorrelated and zero-mean, then

$$\phi_{ZZ}(\tau) = \phi_{XX}(\tau) + \phi_{YY}(\tau) \ .$$

Example A.6

Can the following be a correlation matrix for two jointly wide sense stationary zero-mean random processes?

$$\boldsymbol{\phi}(\tau) = \begin{bmatrix} \phi_{XX}(\tau) & \mu_{XY}(\tau) \\ \mu_{YX}(\tau) & \mu_{YY}(\tau) \end{bmatrix} = \begin{bmatrix} A^2\cos\tau & 2A^2\cos(3\tau/2) \\ 2A^2\cos(3\tau/2) & A^2\sin 2\tau \end{bmatrix} \ .$$

Note that the following two conditions are violated.

1. $|\phi_{XY}(\tau)| \leq \frac{1}{2}[\phi_{XX}(0) + \phi_{YY}(0)]$

2. $|\phi_{XY}(\tau)|^2 \leq \phi_{XX}(0)\phi_{YY}(0)$ if $X(t)$ and $Y(t)$ have zero mean.

Therefore, $\boldsymbol{\phi}(\tau)$ is not a valid correlation matrix.

5.3   COMPLEX-VALUED RANDOM PROCESSES

A complex-valued random process is given by

$$Z(t) = X(t) + jY(t)$$

where $X(t)$ and $Y(t)$ are real random processes.

Autocorrelation Function.   The autocorrelation function of a complex-valued random process is

$$
\begin{aligned}
\phi_{ZZ}(t_1, t_2) &= \frac{1}{2} E[Z(t_1) Z^*(t_2)] \\
&= \frac{1}{2} E\left[(X(t_1) + jY(t_1)) \left(X(t_2) - jY(t_2) \right) \right] \\
&= \frac{1}{2} \{ \phi_{XX}(t_1, t_2) + \phi_{YY}(t_1, t_2) \\
&\quad + \{ j[\phi_{YX}(t_1, t_2) - \phi_{XY}(t_1, t_2)] \} \quad .
\end{aligned}
\tag{A.60}
$$

If $Z(t)$ is wide sense stationary, then

$$\phi_{ZZ}(t_1, t_2) = \phi_{ZZ}(t_1 - t_2) = \phi_{ZZ}(\tau) \quad .$$

Crosscorrelation Function.   Consider two complex-valued random processes

$$
\begin{aligned}
Z(t) &= X(t) + jY(t) \\
W(t) &= U(t) + jV(t) \quad .
\end{aligned}
$$

The crosscorrelation function of $Z(t)$ and $W(t)$ is

$$
\begin{aligned}
\phi_{ZW}(t_1, t_2) &= \frac{1}{2} E[Z(t_1) W^*(t_2)] \\
&= \frac{1}{2} \{ \phi_{XV}(t_1, t_2) + \phi_{YV}(t_1, t_2) \\
&\quad + j[\phi_{YU}(t_1, t_2) - \phi_{XV}(t_1, t_2)] \} \quad .
\end{aligned}
\tag{A.61}
$$

If $X(t)$, $Y(t)$, $U(t)$, and $V(t)$ are pairwise wide sense stationary random processes, then

$$\phi_{ZW}(t_1, t_2) = \phi_{ZW}(t_1 - t_2) = \phi_{ZW}(\tau) \quad . \tag{A.62}$$

The crosscorrelation of a complex wide sense stationary random process satisfies the following property

$$\phi_{ZW}^*(\tau) \;=\; \frac{1}{2}E[Z^*(t_1)W(t_1-\tau)]$$

$$=\; \frac{1}{2}E[Z^*(\hat{t}_1+\tau)W(\hat{t}_1)]$$

$$=\; \frac{1}{2}E[W(\hat{t}_1)Z^*(\hat{t}_1+\tau)]$$

$$=\; \phi_{WZ}(-\tau) \;. \tag{A.63}$$

It also follows that

$$\phi_{XX}^*(\tau) \;=\; \phi_{XX}(-\tau) \tag{A.64}$$

$$\phi_{XX}(\tau) \;=\; \phi_{XX}^*(-\tau) \;. \tag{A.65}$$

5.4 POWER SPECTRAL DENSITY

The power spectral density (psd) of a random process $X(t)$ is the Fourier transform of the autocorrelation function, i.e.,

$$\Phi_{XX}(f) \;=\; =\int_{-\infty}^{\infty}\phi_{XX}(\tau)e^{-j2\pi f\tau}d\tau \tag{A.66}$$

$$\phi_{XX}(\tau) \;=\; \int_{-\infty}^{\infty}\Phi_{XX}(f)e^{j2\pi f\tau}df \;. \tag{A.67}$$

If $X(t)$ is real, then $\phi_{XX}(\tau)$ is real and even. Therefore, $\Phi_{XX}(-f) = \Phi_{XX}(f)$ meaning that $\Phi_{XX}(f)$ is also real and even. If $X(t)$ is complex, then $\phi_{XX}(\tau) = \phi_{XX}^*(-\tau)$, and $\Phi_{XX}^*(f) = \Phi_{XX}(f)$ meaning that $\Phi_{XX}(f)$ is real but not necessarily even.

The power, P, in a random process $X(t)$ is

$$P \;=\; E[X^2(t)]$$

$$=\; \phi_{XX}(0)$$

$$=\; \int_{-\infty}^{\infty}\Phi_{XX}(f)df$$

a result known as **Parseval's theorem**.

The **cross power spectral density** between two random processes $X(t)$ and $Y(t)$ is

$$\Phi_{XY}(f) = \int_{-\infty}^{\infty}\phi_{XY}(\tau)e^{j2\pi f\tau}d\tau \;. \tag{A.68}$$

If $X(t)$ and $Y(t)$ are both real random processes, then

$$\phi_{XY}(\tau) = \phi_{YX}(\tau)$$

and

$$\Phi_{XY}(f) = \Phi_{YX}(-f) \;.$$

If $X(t)$ and $Y(t)$ are complex random processes, then

$$\phi_{XY}^*(\tau) = \phi_{YX}(-\tau)$$

and

$$\Phi_{XY}^*(f) = \Phi_{YX}(f) \ .$$

5.5 RANDOM PROCESSES FILTERED BY LINEAR SYSTEMS

Consider the linear system with impulse response $h(t)$, shown in Fig. A.2. Suppose that the input to the linear system is a wide sense stationary random process $X(t)$, with mean μ_X and autocorrelation $\phi_{XX}(\tau)$. The input and output are related by the convolution integral

$$Y(t) = \int_{-\infty}^{\infty} h(\tau)X(t-\tau)d\tau \ .$$

Hence,

$$Y(f) = H(f)X(f) \ .$$

The output mean is

$$\mu_Y = \int_{-\infty}^{\infty} h(\tau)\mathrm{E}[X(t-\tau)]d\tau = \mu_X \int_{-\infty}^{\infty} h(t\tau)d\tau = \mu_X H(0) \ .$$

The output autocorrelation is

$$
\begin{aligned}
\phi_{YY}(\tau) &= \mathrm{E}[Y(t+\tau)Y(t)] \\
&= \mathrm{E}\left[\int_{-\infty}^{\infty} h(\alpha)X(t+\tau+\alpha)\int_{-\infty}^{\infty} h(\beta)X(t-\beta)d\beta d\alpha\right] \\
&= \int_{-\infty}^{\infty} \infty_{-\infty}^{\infty} h(\alpha)h(\beta)\phi_{XX}(\tau-\alpha+\beta)d\beta d\alpha \\
&= \int_{-\infty}^{\infty} h(\beta)\int_{-\infty}^{\infty} h(\alpha)\phi_{XX}(\tau+\beta-\alpha)d\alpha d\beta \\
&= \left\{\int_{-\infty}^{\infty} h(\beta)\phi_{XX}(\tau+\beta)d\beta\right\} * h(\tau) \\
&= h(-\tau) * \phi_{XX}(\tau) * h(\tau) \ .
\end{aligned}
$$

Taking transforms, the output psd is

$$
\begin{aligned}
\Phi_{YY}(f) &\approx H(f)H^*(f)\Phi_{XX}(f) \\
&\approx |H(f)|^2 \Phi_{XX}(f) \ .
\end{aligned}
$$
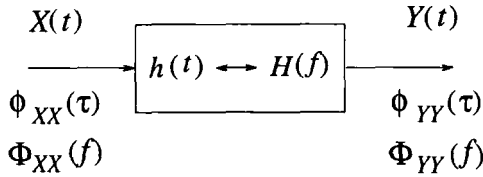
Figure A.2. Random process through a linear system.

Example A.7

Consider the linear system shown in Fig. A.2. In this example we will find the crosscorrelation between the input and output random processes, $X(t)$ and $Y(t)$, respectively. The crosscorrelation $\phi_{YX}(\tau)$ is given by

$$
\begin{aligned}
\phi_{YX}(\tau) &= \mathrm{E}[Y(t+\tau)X(t)] \\
&= \mathrm{E}\left[\int_{-\infty}^{\infty} h(\alpha)X(t+\tau-\alpha)X(t)d\alpha\right] \\
&= \int_{-\infty}^{\infty} h(\alpha)\mathrm{E}[X(t+\tau-\alpha)X(t)]d\alpha \\
&= \int_{-\infty}^{\infty} h(\alpha)\phi_{XX}(t-\alpha)d\alpha \\
&= h(\tau) * \phi_{XX}(\tau) \ .
\end{aligned}
$$

Also,

$$
\Phi_{YX}(f) = H(f)\Phi_{XX}(f) \ .
$$

Example A.8

Suppose that $X(t)$ is a Gaussian random process with mean μ_X and covariance function $\mu_{XX}(\tau)$. In this example we find the joint density of $X_1 \equiv X(t_1)$ and $X_2 \equiv Y(t_2)$. If a Gaussian random process is passed through a linear filter, then the output process is also Gaussian. Hence, X_1 and X_2 have joint Gaussian density function as defined in (A.40) that is completely described in terms of their means and covariances.

Step 1: Obtain the mean and covariance matrix of X_1 and X_2.
The crosscovariance of X_1 and X_2 is

$$
\begin{aligned}
\mu_{X_2 X_1}(\tau) &= \mathrm{E}\left[(Y(t+\tau) - \mu_Y)(X(t) - \mu_X)\right] \\
&= \mathrm{E}\left[Y(t+\tau)X(t)\right] - \mu_Y \mu_X \ .
\end{aligned}
$$

Now $\mu_Y = H(0)\mu_X$. Also, from the previous example

$$
\begin{aligned}
\mathrm{E}[Y(t+\tau)X(t)] &= \int_{-\infty}^{\infty} h(\alpha)\phi_{XX}(\tau-\alpha)d\alpha \\
&= \int_{-\infty}^{\infty} h(\alpha)[\mu_{XX}(\tau-\alpha)+\mu_X^2]d\alpha \\
&= \int_{-\infty}^{\infty} h(\alpha)\mu_{XX}(\tau-\alpha)d\alpha + H(0)\mu_X^2 \quad.
\end{aligned}
$$

Therefore,

$$
\mu_{X_2 X_1}(\tau) = \int_{-\infty}^{\infty} h(\alpha)\mu_{XX}(\tau-\alpha)d\alpha = h^*(\tau) * \mu_{XX}(\tau) \quad.
$$

Also

$$
\begin{aligned}
\mu_{X_1 X_2}(\tau) &= \mu_{Z_2 Z_1}(-\tau) \\
\mu_{X_1 X_1}(\tau) &= \mu_{XX}(\tau) \\
\mu_{X_2 X_2}(\tau) &= h(\tau) * h(-\tau) * \mu_{XX}(\tau) \quad.
\end{aligned}
$$

Hence, the covariance matrix is

$$
\Lambda = \begin{bmatrix} \mu_{X_1 X_1}(0) & \mu_{X_1 X_2}(\tau) \\ \mu_{X_2 X_1}(\tau) & \mu_{X_2 X_2}(0) \end{bmatrix} = \begin{bmatrix} \mu_{XX}(0) & h(\tau) * \mu_{XX}(-\tau) \\ h(\tau)\mu_{XX}(\tau) & h(\tau) * h(-\tau) * \mu_{XX}(\tau) \end{bmatrix}
$$

Step 2: Write the joint density function of X_1 and X_2.
Let

$$
\begin{aligned}
\mathbf{X} &= (X_1, X_2)^T \\
\boldsymbol{\mu}_X &= (\mu_X, \mu_Y)^T = (\mu_X, H(0)\mu_X)^T \quad.
\end{aligned}
$$

Then

$$
P_{\mathbf{X}}(\mathbf{x}) = \frac{1}{2\pi|\Lambda|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu}_X)'\Lambda^{-1}(\mathbf{z}-\boldsymbol{\mu}_X)\right\} \quad.
$$

5.6   DISCRETE-TIME RANDOM PROCESSES

Let $X_n \equiv X(n)$, where n is an integer time variable, be a discrete-time random process. Then the mth moment of X_n is

$$
\mathrm{E}[X_n^m] = \int_{-\infty}^{\infty} x_n^m p_X(x_n)dx_n \quad. \tag{A.69}
$$

The autocorrelation of X_n is

$$
\phi(n,k) = \frac{1}{2}\mathrm{E}[X_n X_k^*] = \frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_n x_k^* p_X(x_n)p_X(x_k)dx_m dx_k \tag{A.70}
$$

and the autocovariance is

$$\mu(n, k) = \phi(n, k) - \mathrm{E}[X_n]\mathrm{E}[X_k] \ . \tag{A.71}$$

If X_n is a wide sense stationary random process, then

$$\phi(n, k) \equiv \phi(n - k) \tag{A.72}$$
$$\mu(n, k) \equiv \mu(n - k) = \phi(n - k) - \mu_X^2 \ . \tag{A.73}$$

The psd of a discrete random process is

$$\Phi(f) = \sum_{n=-\infty}^{\infty} \phi(n)e^{-j2\pi fn} \tag{A.74}$$

where

$$\phi(n) = \int_{-1/2}^{1/2} \Phi(f)e^{j2\pi fn}df \ . \tag{A.75}$$

Note that $\Phi(f) = \Phi(f + k)$ for any integer k.

Consider a discrete-time linear time-invariant system with impulse response $h_n \equiv h(n)$. The input, X_n, and output, Y_n, are related by the convolution sum

$$Y_n = \sum_{-\infty}^{\infty} h_k X_{n-k} \ . \tag{A.76}$$

The output mean is

$$\mu_Y = \mathrm{E}[Y_n] = \sum_{k=-\infty}^{\infty} h_k \mathrm{E}[X_{n-k}] = \mu_X \sum_{k=-\infty}^{\infty} h_k = \mu_X H(0)$$

and the output autocorrelation is

$$
\begin{aligned}
\phi_{YY}(k) &= \frac{1}{2}\mathrm{E}[Y(n + k)Y^*(n)] \\
&= \frac{1}{2}\mathrm{E}\left[\sum_{\ell=-\infty}^{\infty} h_\ell X^*(n - \ell) \sum_{m=-\infty}^{\infty} h_m X(n + k - m) \right] \\
&= \frac{1}{2} \sum_{\ell=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} h_\ell h_m \mathrm{E}[X(n + k - m)X^*(n - \ell)] \\
&= \sum_{\ell=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} h_\ell h_m \phi_{XX}(k + \ell - m) \ .
\end{aligned}
$$

The output psd is

$$
\begin{aligned}
\Phi_{YY} &= \sum_{k=-\infty}^{\infty} \phi_{YY}(k) e^{-j2\pi f k} \\
&= \sum_{k=-\infty}^{\infty} \sum_{\ell=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} h_\ell h_m \phi_{XX}(k + \ell - m) e^{j2\pi f k} \\
&= \sum_{n=-\infty}^{\infty} \sum_{\ell=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} h_\ell h_m \phi_{XX}(n) e^{-j2\pi f(n-\ell+m)} \\
&= \sum_{\ell=-\infty}^{\infty} h_\ell e^{j2\pi f \ell} \sum_{m=-\infty}^{\infty} e^{-j2\pi f m} \\
&\quad \times \sum_{n=-\infty}^{\infty} \phi_{XX}(n) e^{-j2\pi f n} \\
&= H^*(f) H(f) \Phi_{XX}(f) \\
&= |H(f)|^2 \, \Phi_{XX}(f) \ .
\end{aligned}
$$

5.7 CYCLOSTATIONARY RANDOM PROCESSES

Consider the random process

$$
X(t) = \sum_{n=-\infty}^{\infty} a_n \psi(t - nT)
$$

where $\{a_n\}$ is a sequence of complex random variables with mean μ_a and autocorrelation $\phi_{aa}(n)$, and $\psi(t)$, $0 \le t \le T$ is a real deterministic shaping function. Note that the mean of $X(t)$

$$
\mu_X = \mu_a \sum_{n=-\infty}^{\infty} \psi(t - nT)
$$

is periodic. The autocorrelation of $X(t)$ is

$$
\begin{aligned}
\phi_{XX}(t+\tau, t) &= \frac{1}{2} E[X(t+\tau) X^*(t)] \\
&= \frac{1}{2} E\left[\sum_{-\infty}^{\infty} a_n \psi(t+\tau - nT) \sum_{-\infty}^{\infty} a_m^* \psi(t - mT) \right] \\
&= \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} \phi_{aa}(n - m) \psi(t - mT) \psi(t + \tau - nT) \ .
\end{aligned}
$$

It is easy to show that

$$
\phi_{XX}(t + \tau + kT, t + kT) = \phi_{XX}(t + \tau, t) \ .
$$

Therefore, $\phi_{XX}(t + \tau, t)$ is periodic in t with period T.

The *time-averaged* psd of $X(t)$ can be computed by first determining the time-average autocorrelation

$$\phi_{XX}(\tau) = \frac{1}{T} \int_{-T/2}^{T/2} \phi_{XX}(t + \tau, t) dt$$

and then taking the Fourier transform in (A.66).

References

[1] M. Abramowitz and I. A. S. (ed.), *Handbook of Mathematical Functions.* New York, NY: Dover, 1965.

[2] F. Abrishamkar and E. Biglieri, "Suboptimum detection of trellis-coded CPM for transmission on bandwidth- and power-limited channels," *IEEE Trans. Commun.*, Vol. 39, pp. 1065–1074, July 1991.

[3] A. A. Abu-Dayya and N. C. Beaulieu, "Micro- and macrodiversity NCFSK (DPSK) on shadowed Nakagami-fading channels," *IEEE Trans. Commun.*, Vol. 42, pp. 2693–2702, August 1994.

[4] A. A. Abu-Dayya and N. C. Beaulieu, "Micro- and macrodiversity MDPSK on shadowed frequency-selective channels," *IEEE Trans. Commun.*, Vol. 43, pp. 2334–2343, August 1995.

[5] A. Abu-Dayya and N. Beaulieu, "Outage probabilities of cellular mobile radio systems with multiple Nakagami interferers," *IEEE Trans. Veh. Technol.*, Vol. 40, pp. 757–768, November 1991.

[6] A. S. Acampora, "Maximum likelihood decoding of binary convolutional codes on band-limited satellite channels," *IEEE Trans. Commun.*, Vol. 26, pp. 766–776, June 1978.

[7] A. S. Acampora, "Analysis of maximum likelihood sequence estimation performance for quadrature amplitude modulation," *Bell System Tech. J.,* Vol. 60, pp. 865-885, July 1981.

[8] Y. Akaiwa, "A conceptual design of microcellular radio communication system," in *IEEE Veh. Technol. Conf.*, Orlando, FL, pp. 156–160, May 1990.

[9] Y. Akaiwa and H. Andoh, "Channel segregation - a self organized dynamic channel allocation method: application to TDMA/FDMA microcellular systems," *IEEE J. Selec. Areas Commun.*, Vol. 11, pp. 949–954, August 1993.

[10] S. M. Alamouti, "A Simple Transmit Diversity Technique for Wireless Communications," *IEEE J. Selec. Areas Commun.*, Vol. 16, pp. 1451–1458, October 1998.

[11] J. B. Andersen, T. Rappaport, and S. Yoshida, "Propagation measurements and models for wireless communications channels," *IEEE Commun. Mag.*, Vol. 33, pp. 42–49, January 1995.

[12] J. B. Anderson, T. Aulin, and C. E. Sundberg, *Digital Phase Modulation.* New York, NY: Plenum, 1986.

[13] L. Anderson, "A simulation study of some dynamic channel assignment algorithms in a high capacity mobile telecommunications system," *IEEE Trans. Veh. Technol.*, Vol. 22, pp. 210–217, November 1973.

[14] S. Ariyavisitakul, "SIR-based power control in a CDMA system," in *IEEE Global Commun. Conf.*, Orlando, FL, pp. 868–873, December 1992.

[15] S. Ariyavisitakul and L. F. Chang, "Signal and interference statistics of a CMDA system with feedback power control," *IEEE Trans. Commun.*, Vol. 41, pp. 1626–1634, November 1993.

[16] T. Aulin, "A modified model for the fading signal at a mobile radio channel," *IEEE Trans. Veh. Technol.*, Vol. 28, pp. 182–203, August 1979.

[17] M. D. Austin and G. L. Stüber, "Velocity adaptive handoff algorithms for microcellular systems," in *IEEE Conf. Universal Personal Commun.*, Ottawa, Canada, pp. 793–797, October 1993.

[18] M. D. Austin and G. L. Stüber, "Velocity adaptive handoff algorithms for microcellular systems," *IEEE Trans. Veh. Technol.*, Vol. 43, pp. 549–561, August 1994.

[19] M. D. Austin and G. L. Stüber, "In-service signal quality estimation for TDMA cellular systems," in *IEEE Int. Symp. on Pers., Indoor and Mobile Radio Commun.*, Toronto, Canada, pp. 836–840, September 1995.

[20] M. E. Austin, "Decision-feedback equalization for digital communication over dispersive channels," Tech. Rep., MIT Lincoln Lab., Lexington, MA., August 1967.

[21] M. Austin and G. L. Stüber, "In-service Signal Quality Estimation for TDMA Cellular Systems," *Kluwer J. Wireless Personal Commun.*, Vol. 2, pp. 245–254, January 1996.

[22] M. D. Austin and G. L. Stüber, "Co-channel interference modeling for signal strength based handoff analysis," *Electronics Letters*, Vol. 30, pp. 1914–1915, November 1994.

[23] M. D. Austin and G. L. Stüber, "Direction biased handoff algorithms for urban microcells," in *IEEE Veh. Technol. Conf.*, Stockholm, Sweden, pp. 101–105, June 1994.

[24] M. D. Austin and G. L. Stüber, "Exact co-channel interference analysis for log-normal shadowed Rician fading channels," *Electronic Letters*, Vol. 30, pp. 748–749, May 1994.

[25] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate," *IEEE Trans. Inform. Theory*, Vol. 20, pp. 284–287, March 1974.

[26] G. Battail, "A Conceptual Framework for understanding Turbo codes," *IEEE J. Selec. Areas Commun.*, Vol. 16, pp. 245–254, February 1998.

[27] C. T. Beare, "The choice of the desired impulse response in combined linear-Viterbi algorithm equalizer," *IEEE Trans. Commun.*, Vol. 26, pp. 1301–1327, August 1978.

[28] N. C. Beaulieu, A. A. Abu-Dayya, and P. J. McLane, "Comparison of methods of computing lognormal sum distributions and outages for digital wireless applications," in *IEEE Int. Conf. on Commun.*, New Orleans, LA, pp. 1270–1275, May 1994.

[29] C. A. Belfiore and J. J.H. Park, "Decision-feedback equalization," in *Proc. IEEE*, pp. 1143–1156, August 1979.

[30] P. Bello, "Characterization of random time-variant linear channels," *IEEE Trans. Commun.*, Vol. 11, pp. 360–393, December 1963.

[31] S. Benedetto and G. Montorsi, "Design of parallel concatenated convolutional codes," *IEEE Trans. Commun.*, Vol. 44, pp. 591–600, May 1996.

[32] S. Benedetto and G. Montorsi, "Unveiling Turbo Codes: Some Results on Parallel Concatenated Coding Schemes," *IEEE Trans. Inform. Theory*, Vol. 42, pp. 409–428, March 1996.

[33] J.-E. Berg, R. Bownds, and F. Lotse, "Path loss and fading models for microcells at 900 MHz," in *IEEE Veh. Technol. Conf.*, Denver, CO, pp. 666–671, May 1992.

[34] R. C. Bernhardt, "Macroscopic diversity in frequency reuse radio systems," *IEEE J. Selec. Areas Commun.*, Vol. 5, pp. 862–870, June 1987.

[35] R. C. Bernhardt, "The use of multiple-beam directional antennas in wireless messaging systems," in *IEEE Veh. Technol. Conf.*, pp. 858–861, 1995.

[36] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding," in *IEEE Int. Conf. on Commun.*, Geneva, Switzerland, pp. 1064–1070, June 1993.

[37] E. Biglieri, "High-level modulation and coding for nonlinear satellite channels," *IEEE Trans. Commun.*, Vol. 32, pp. 616–626, May 1984.

[38] E. Biglieri, D. Divsalar, P. McLane, and M. Simon, *Introduction to Trellis-Coded Modulation with Applications.* New York, NY: McMillan, 1991.

[39] M. A. Birchler and S. C. Jasper, "A 64 kbps digital land mobile radio system employing M-16QAM," in *Proc. of 5th Nordic Sem. Land Mobile Radio*, Helsinki, Finland, pp. 237–241, December 1992.

[40] G. E. Bottomley and K. J. Molnar, "Interference Cancellation for improved channel estimation in array processing MLSE receivers," in *IEEE Veh. Technol. Conf.*, Phoenix, AZ, pp. 140–144, May 1997.

[41] G. E. Bottomley and K. Jamal, "Adpative Arrays and MLSE Equalization," in *IEEE Veh. Technol. Conf.*, Chicago, IL, pp. 50–54, July 1995.

[42] D. M. Brady, "An adaptive coherent diversity receiver for data transmission through dispersive media," in *IEEE Int. Conf. on Commun.*, San Francisco, CA, pp. 21.35–21.40, June 1970.

[43] A. R. Calderbank and J. Mazo, "A new description of trellis codes," *IEEE Trans. Inform. Theory*, Vol. 30, pp. 784–791, November 1984.

[44] A. R. Calderbank and N. J. Sloane, "New trellis codes based on lattices and cosets," *IEEE Trans. Inform. Theory*, Vol. 33, pp. 177–195, March 1987.

[45] J. Cavers and P. Ho, "Analysis of the error performance of trellis-coded modulation in Rayleigh-fading channels," *IEEE Trans. Commun.*, Vol. 40, pp. 74–83, January 1992.

[46] C. J. Chang and F. C. Ren, "Downlink power control in DS/CDMA cellular mobile radio networks," in *IEEE Conf. Universal Personal Commun.*, pp. 89–93, 1994.

[47] U. Charash, "Reception through Nakagami fading multipath channels with random delays," *IEEE Trans. Commun.*, Vol. 27, pp. 657–670, April 1979.

[48] P. Chevillat and J. D. J. Costello, "A multiple stack algorithm for erasure free decoding of convolutional codes," *IEEE Trans. Commun.*, Vol. 25, pp. 1460–1470, December 1977.

[49] P. Chevillat and E. Eleftheriou, "Decoding of trellis-encoded signal in the presence of intersymbol interference and noise," *IEEE Trans. Commun.*, Vol. 37, pp. 669–676, July 1989.

[50] P. Chevillat and E. Eleftheriou, "Decoding of trellis-encoded signals in the presence of intersymbol interference and noise," *IEEE Trans. Commun.*, Vol. 37, pp. 669–676, July 1989.

[51] S. Chia, R. Steele, E. Green, and A. Baran, "Propagation and bit-error ratio measurements for a microcellular system," *J. Inst. Electron. Radio Eng. (UK)*, Vol. 57, pp. 255–266, November 1987.

[52] S. T. S. Chia and R. J. Warburton, "Handover criteria for city microcellular systems," in *IEEE Veh. Technol. Conf.*, Orlando, FL, pp. 276–281, May 1990.

[53] K. M. Chugg and A. Polydoros, "MLSE for an Unknown Channel – Part I: Optimality Considerations," *IEEE Trans. Commun.*, Vol. 44, pp. 836–846, July 1996.

[54] L. Cimini, Jr., "Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing," *IEEE Trans. Commun.*, Vol. 33, pp. 665–675, July 1985.

[55] L. Cimini, Jr. and G. Foschini, "Distributed algorithms for dynamic channel allocation in microcellular systems," in *IEEE Veh. Technol. Conf.*, Denver, CO, pp. 641–644, May 1992.

[56] L. Cimini, Jr., G. J. Foschini, C.-L. I, and Z. Miljanic, "Call blocking performance of distributed algorithms for dynamic channel allocation in microcells," *IEEE Trans. Commun.*, Vol. 42, pp. 2600–2607, August 1994.

[57] L. Cimini, Jr., G. J. Foschini, and L. Shepp, "Single-channel user-capacity calculations for self-organizing cellular systems," *IEEE Trans. Commun.*, Vol. 42, pp. 3137–3143, December 1994.

[58] M. Cioffi and T. Kailath, "Fast, recursive-least-squares, transversal filters for adaptive filtering," *IEEE Trans. Acoutics, Speech and Signal Proc.*, Vol. 32, pp. 304–337, April 1984.

[59] A. P. Clark, ed., *Advanced Data Transmission Systems.* London, UK: Pentech Press, 1977.

[60] A. P. Clark, S. N. Abdullah, S. J. Jayasinghe, and K. Sun, "Pseudobinary and pseudoquatemary detection processes for linearly distorted multilevel QAM signals," *IEEE Trans. Commun.*, Vol. 33, pp. 639–645, July 1985.

[61] A. P. Clark and M. Clayden, "Pseudobinary Viterbi detector," *lEE Proc. Part F*, Vol. 131, pp. 208–218, April 1984.

[62] A. P. Clark and R. Harun, "Assessment of Kalman-filter channel estimator for an HF radio link," *IEE Proc.*, Vol. 133, pp. 513–521, October 1986.

[63] G. C. Clark, Jr. and J. B. Cain, *Error-Correction Coding for Digital Communications.* New York, NY: Plenum, 1981.

[64] R. Clarke, "A statistical theory of mobile radio reception," *Bell System Tech. J.*, Vol. 47, pp. 957–1000, 1968.

[65] G. R. Cooper and R. W. Nettleton, "A spread-spectrum technique for high-capacity mobile communications," *IEEE Trans. Veh. Technol.*, Vol. 27, pp. 264–275, November 1978.

[66] K. G. Cornett and S. B. Wicker, "Bit error rate estimation techniques for digital land mobile radios," in *IEEE Veh. Technol. Conf.*, Saint Louis, MO, pp. 543–548, May 1991.

[67] COST 207 TD(86)51-REV 3 (WG1):, "Proposal on channel transfer functions to be used in GSM tests late 1986," September 1986.

[68] COST 231 TD(91)109, "1800 MHz mobile net planning based on 900 MHz measurements," 1991.

[69] COST 231 TD(973)119-REV 2 (WG2):, "Urban transmission loss models for mobile radio in the 900- and 1,800-MHz bands," September 1991.

[70] D. J. Costello, *Construction of convolutional codes for sequential decoding.* 1969.

[71] D. J. Costello and G. Meyerhans, "Concatenated Turbo Codes," in *IEEE Inform. Theory Symp.*, pp. 571–574, October 1996.

[72] D. C. Cox and D. O. Reudnik, "A comparison of some channel assignment strategies in large-scale mobile communication systems," *IEEE Trans. Commun.*, Vol. 20, pp. 190–195, February 1972.

[73] D. C. Cox and D. O. Reudnik, "Increasing channel occupancy in large-scale mobile radio systems: dynamic channel reassignment," *IEEE Trans. Veh. Technol.*, Vol. 22, pp. 218–222, November 1973.

[74] D. C. Cox, "Cochannel interference considerations in frequency reuse small-coverage-area radio systems," *IEEE Trans. Commun.*, Vol. 30, pp. 135–142, January 1982.

[75] K. Daikoku and H. Ohdate, "Optimal channel reuse in cellular land mobile radio systems," *IEEE Trans. Veh. Technol.*, Vol. 32, pp. 217–224, August 1983.

[76] G. D'aria and V. Zingarelli, "Results on fast-Kalman and Viterbi adaptive equalizers for mobile radio with CEPT/GSM system characteristics," in *IEEE Global Commun. Conf.*, Hollywood, FL, pp. 815–819, December 1988.

[77] G. D'Avella, L. Moreno, and M. Sant'Agostino, "An adaptive MLSE receiver for TDMA digital mobile radio," *IEEE J. Selec. Areas Commun.*, Vol. 7, pp. 122–129, January 1989.

[78] W. B. Davenport and W. L. Root, *An Introduction to the Theory of Random Signals and Noise.* New York, NY: McGraw-Hill, 1987.

[79] G. W. Davidson, D. D. Falconer, and A. U. H. Sheikh, "An investigation of block adaptive decision feedback equalization for frequency selective fading channels," in *IEEE Int. Conf. on Commun.*, Philadelphia, PA, pp. 360–365, June 1988.

[80] F. de Jager and C. B. Dekker, "Tamed Frequency modulation, a novel method to achieve spectrum economy in digital transmission," *IEEE Trans. Commun.*, Vol. 26, pp. 534–542, May 1978.

[81] P. Dent, G. E. Bottomley, and T. Croft, "Jakes fading model revisited," *Electronic Letters*, Vol. 7, pp. 1162–1163, June 1993.

[82] D. Divsalar and M. Simon, "Trellis coded modulation for 4800-9600 bit/s transmission over a fading mobile satellite channel," *IEEE J. Selec. Areas Commun.*, Vol. 5, pp. 162–174, February 1987.

[83] D. Divsalar and M. Simon, "The design of trellis coded MPSK for fading channels: performance criteria," *IEEE Trans. Commun.*, Vol. 36, pp. 1004–1012, September 1988.

[84] D. Divsalar and M. Simon, "The design of trellis coded MPSK for fading channels: set partitioning for optimum code design," *IEEE Trans. Commun.,* Vol. 36, pp. 1013-1022, September 1988.

[85] D. Divsalar, M. Simon, and J. Yuen, "Trellis coding with asymmetric modulations," *IEEE Trans. Commun.*, Vol. 35, pp. 130–141, February 1987.

[86] R. *C.* Dixon, *Spread Spectrum Techniques.* New York, NY: IEEE Press, 1976.

[87] S. Dolinar and D. Divsalar, "Weight Distributions for Turbo Codes Using Random and Nonrandom Permutations," August 1995.

[88] J. Driscoll and N. Karia, "Detection process for V32 modems using trellis coding," *Proc. IEEE*, Vol. 135, pp. 143–154, April 1988.

[89] A. Duel-Hallen, *Detection for Channels with Intersymbol Interference.* Ph. D. thesis, Cornell University, Ithica, NY, 1987.

[90] A. Duel-Hallen and C. Heegard, "Delayed decision feedback sequence estimation," *IEEE Trans. Commun.*, Vol. 37, pp. 428–436, May 1989.

[91] A. Duel-Hallen, "Equalizers for Multiple Input/Multiple Output Channels and PAM Systems with Cyclostationary Input Sequences," *IEEE J. Selec. Areas Commun.*, Vol. 10, pp. 630–639, April 1992.

[92] A. Duel-Hallen, "Decorrelating Decision-Feedback Multiuser Detector for Synchronous Code-Division Multiple-Access Channel," *IEEE Trans. Commun.*, Vol. 41, pp. 285–290, February 1993.

[93] A. Duel-Hallen, "A Family of Multiuser Decision-Feedback Detectors for Asynchronous Code-Division Multiple-Access Channels," *IEEE Trans. Commun.,* Vol. 43, pp. 421–434, February 1995.

[94] F. Edbauer, "Performance of interleaved trellis-coded differential 8-PSK modulation over fading channels," *IEEE J. Selec. Areas Commun.*, Vol. 7, pp. 1340–1346, December 1989.

[95] EIA/TIA IS -54, "Cellular system dual-mode mobile station - base station compatibility standard,".

[96] EIA/TIA IS-95, "Mobile station – base station compatability standard for dual-mode wideband spread spectrum cellular system,".

[97] E. Eleftheriou and D. D. Falconer, "Restart methods for stabilizing FRLS adaptive equalizers in digital HF transmission," in *IEEE Global Commun.* Conf., Atlanta, GA, pp. 1558–1562, November 1984.

[98] E. Eleftheriou and D. D. Falconer, "Tracking properties and steady-state performance of RLS adaptive filter algorithms," *IEEE Tram. Acoutics, Speech and Signal Proc.*, Vol. 34, pp. 1097–1110, October 1986.

[99] E. Eleftheriou and D. D. Falconer, "Adaptive equalization techniques for HF channels," *IEEE J. Selec. Areas Commun.*, Vol. 5, pp. 238–247, February 1987.

[100] S. Elnoubi, R. Singh, and S. Gupta, "A new frequency channel assignment algorithm in high capacity mobile communications systems," *IEEE Trans. Veh. Technol.*, Vol. 31, pp. 125–131, August 1982.

[101] J. Engel and M. Peritsky, "Statistically-optimum dynamic server assignment in systems with interfering servers," *IEEE Trans. Veh. Technol.*, Vol. 22, pp. 203–209, November 1973.

[102] ETSI – European Telecommunications Standards Institute, *GSM Recommendation 05.08.* January 1991.

[103] W. V. Etten, "Maximum Likelihood Receiver for Multiple Channel Transmission Systems," *IEEE Trans. Commun.*, Vol. 24, pp. 276–283, February 1976.

[104] D. Everitt, "Traffic capacity of cellular mobile communications systems," *Computer Networks and ISDN Systems*, Vol. 20, pp. 447–454, December 1990.

[105] D. Everitt and D. Manfield, "Performance analysis of cellular mobile communication systems with dynamic channel assignment," *IEEE J. Selec. Areas Commun.*, Vol. 7, pp. 1172–1179, October 1989.

[106] D. E. Everitt and N. W. MacFadyen, "Analysis of multicellular mobile radio-telephone systems: a model and evaluation," *British Telecom Tech. J.*, Vol. 1, pp. 37–45, 1983.

[107] V. M. Eyuboğlu, "Detection of coded modulation signals on linear, severely distorted channels using decision-feedback noise prediction

with interleaving," *IEEE Trans. Commun.*, Vol. 36, pp. 401–409, April 1988.

[108] V.M. Eyuboğlu and D.Forney, Jr., "Trellis preceding: combining coding, preceding, and shaping for intersymbol interference channels," *IEEE Trans. Inform. Theory*, Vol. 38, pp. 301–314, March 1992.

[109] V.M. Eyuboğlu and S. U. Qureshi, "Reduced-state sequence estimation with set partitioning and decision feedback," *IEEE Trans. Commun.*, Vol. 36, pp. 13–20, January 1988.

[110] V. M. Eyuboğlu and S. U. Qureshi, "Reduced-state sequence estimation for coded modulation on intersymbol interference channels," *IEEE J. Selec. Areas Commun.*, Vol. 7, pp. 989–995, August 1989.

[111] G. Falciasecca, M. Frullone, G. Riva, M. Sentinelli, and A. M. Serra, "Investigation on a dynamic channel allocation for high capacity mobile radio systems," in *IEEE Vehicular Technology Conference*, Philadelphia, PA, pp. 176–181, 1988.

[112] D. D. Falconer and J. F. R. Magee, "Adaptive channel memory truncation for maximum likelihood sequence estimation," *Bell System Tech. J.*, Vol. 52, pp. 1541–1562, November 1973.

[113] D. D. Falconer and L. Ljung, "Application of fast Kalman estimation to adaptive equalization," *IEEE Trans. Commun.*, Vol. 26, pp. 1439–1446, October 1978.

[114] R. M. Fano, "A heuristic discussion of probabilistic decoding," *IEEE Trans. Inform. Theory*, Vol. 9, pp. 64–74, April 1963.

[115] K. Feher, *Advanced Digital Communications.* Englewood Cliffs, NJ: Prentice-Hall, 1987.

[116] K. Feher, *Wireless Digital Communications.* Englewood Cliffs, NJ: Prentice-Hall, 1995.

[117] W. Feller, *An Introduction of Probability Theory and Its Applications, Vol. I.* New York, NY: Wiley, 1968.

[118] L. F. Fenton, "The sum of log-normal probability distributions in scatter transmission systems," *IRE Trans. Commun.*, Vol. 8, pp. 57–67, March 1960.

[119] B. L. Floch, R. Halbert-Lassalle, and D. Castelain, "Digital sound broadcasting to mobile receivers," *IEEE Trans. Consumer Elect.*, Vol. 35, pp. 493–503, August 1989.

[120] G. J. Foschini, "A reduced-state variant of maximum-likelihood sequence detection attaining optimum performance for high signal-to-noise ratio performance," *IEEE Trans. Inform. Theory*, Vol. 24, pp. 605–609, September 1977.

[121] E. A. Frech and C. L. Mesquida, "Cellular models and hand-off criteria," *in IEEE Veh. Technol. Conf.*, San Francisco, CA, pp. 128–135, May 1989.

[122] R. C. French, "Error rate predictions and measurements in the mobile radio data channel," *IEEE Trans. Veh. Technol.*, Vol. 27, pp. 214–220, August 1978.

[123] R. C. French, "The effect of fading and shadowing on channel reuse in mobile radio," *IEEE Trans. Veh. Technol.*, Vol. 28, pp. 171–181, August 1979.

[124] B. Friedlander, "Lattice filters for adaptive processing," *Proc. IEEE*, Vol. 70, pp. 829–867, August 1982.

[125] H. Furukawa and A. Yoshihiko, "Self-organized reuse partitioning, a dynamic channel assignment method in cellular systems," in *IEEE Veh. Technol. Conf.*, Secaucus, NJ, pp. 524–527, May 1993.

[126] G. D. Forney, Jr., "Convolutional Codes I: Algebraic structure," *IEEE Trans. Inform. Theory*, Vol. 16, pp. 720–738, 1970.

[127] G. D. Forney, Jr., "Maximum likelihood sequence estimation of digital sequence in the presence of intersymbol interference," *IEEE Trans. Inform. Theory*, Vol. 18, pp. 363–378, May 1972.

[128] G. D. Forney, Jr., "Coset codes - part I: introduction to geometrical classification," *IEEE Trans. Inform. Theory*, Vol. 34, pp. 1123–1151, September 1988.

[129] P. Gaasvik, M. Cornefjord, and V. Svenson, "Different methods of giving priority to handoff traffic in a mobile telephone system with directed retry," in *IEEE Veh. Technol. Conf.*, Saint Louis, MO, pp. 549–553, May 1991.

[130] M. Gans, "A power-spectral theory of propagation in the mobile-radio environment," *IEEE Trans. Veh. Technol.*, Vol. 21, pp. 27–38, February 1972.

[131] G. J. Garrison, "A power spectral density analysis for digital FM," *IEEE Trans. Commun.*, Vol. 23, pp. 1228–1243, November 1975.

[132] E. A. Geraniotis and R. Mani, "Throughput analysis of a random access tree protocol for direct-sequence spread-spectrum packet radio networks," in *IEEE Military Commun. Conf.*, Washington, D. C., pp. 23.7.1–23.7.6, October 1987.

[133] E. A. Geraniotis, "Direct-sequence wpread-spectrum multiple-access communications over nonselective and frequency-selective Rician fading channels," *IEEE Trans. Commun.*, Vol. 34, pp. 756–764, August 1986.

[134] A. Gersho, "Adaptive equalization of highly dispersive channels," *Bell System Tech. J.*, Vol. 48, pp. 55–70, January 1969.

[135] A. Gersho and T. L. Lim, "Adaptive cancellation of intersymbol interference for data transmission," *Bell System Tech. J.*, Vol. 60, pp. 1997–2021, November 1981.

[136] K. S. Gilhousen, I. M. Jacobs, R. Padovani, A. J. Viterbi, L. A. Weaver, Jr., and C. E. Wheatley III, "On the capacity of a cellular CDMA system," *IEEE Trans. Veh. Technol.*, Vol. 40, pp. 303–312, May 1991.

[137] A. A. Giordano and F. M. Hsu, eds., *Least Square Estimation with Applications to Digital Signal Processing.* New York, NY: Wiley, 1985.

[138] K. Giridhar, S. Chari, J. J. Shynk, R. P. Gooch, and D. J. Artman, "Joint estimation algorithms for cochannel signal demodulation," in *IEEE Int. Conf. on Commun.*, Geneva, Switzerland, pp. 1497–1501, June 1993.

[139] R. D. Gitlin, J. E. Mazo, and M. G. Taylor, "On the design of gradient algorithms for digitally implemented adaptive filters," *IEEE Trans. Circuit Theory*, Vol. 20, pp. 125–136, March 1973.

[140] R. D. Gitlin and S. B. Weinstein, "Fractionally-spaced equalization: an improved digital transversal equalizer," *Bell System Tech. J.*, Vol. 60, pp. 275–296, February 1981.

[141] R. Gold, "Optimum binary sequences for spread-spectrum multiplexing," *IEEE Trans. Inform. Theory*, Vol. 13, pp. 619–621, October 1967.

[142] A. J. Goldsmith, L. J. Greenstein, and G. J. Foschini, "Error statistics of real time power measurements in cellular channels with multipath and shadowing," in *IEEE Veh. Technol. Conf.*, Secaucus, NJ, pp. 108–110, May 1993.

[143] A. J. Goldsmith, L. J. Greenstein, and G. J. Foschini, "Error statistics of real-time power measurements in cellular channels with multipath and

shadowing," *IEEE Trans. Veh. Technol.*, Vol. 43, pp. 439–446, August 1994.

[144] A. J. Goldsmith and L. J. Greenstein, "A Measurement-based model for predicting coverage areas of urban microcells," *IEEE J. Selec. Areas Commun.*, Vol. 11, pp. 1013–1023, September 1993.

[145] D. J. Goodman, S. A. Grandhi, and R. Vijayan, "Distributed dynamic channel assignment schemes," in *IEEE Veh. Technol. Conf.*, Secaucus, NJ, pp. 532–535, May 1993.

[146] D. N. Gordard, "Channel equalization using a Kalman filter for fast data transmission," *IBM Journal Research and Development*, Vol. 18, pp. 267–273, May 1974.

[147] I. Gradshteyn and I. Ryzhik, *Tables of Integrals, Series, and Products.* San Diego, CA: Academic Press, 1980.

[148] E. Green, "Path loss and signal variability analysis for microcells," in *5th Int. Conf. on Mobile Radio and Personal Commun.*, Coventry, UK, pp. 38–42, December 1989.

[149] O. Grimlund and B. Gudmundson, "Handoff strategies in microcellular systems," in *IEEE Veh. Technol. Conf.*, Saint Louis, MO, pp. 505–510, May 1991.

[150] M. Gudmundson, "Analysis of handover algorithms," in *IEEE Veh. Technol. Conf.*, Saint Louis, MO, pp. 537–541, May 1991.

[151] M. Gudmundson, "Analysis of handover algorithms in cellular radio systems," Report No. TRITA-TTT-9107, Royal Institute of Technology, Stockholm, Sweden, April 1991.

[152] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," *Electronics Letters*, Vol. 27, pp. 2145–2146, November 1991.

[153] J. C. Haartsen, "The Bluetooth Radio System," *IEEE Personal Commun. Mag.*, Vol. 7, pp. 28 – 36, February 2000.

[154] J. Hagenauer, "Decoding turbo codes with the soft-output Viterbi algorithm (SOVA)," *IEEE Int. Symp. Information Theory*, p. 64, January 1994.

[155] J. Hagenauer and P. Hoeher, "A Soft Output Viterbi Algorithm," in *IEEE Global Commun. Conf.*, pp. 47.1.1–47.1.7, 1989.

[156] J. Hagenauer, E. Offer, and L. Papke, "Iterative decoding of binary block and convolutional codes," *IEEE Trans. Inform. Theory*, pp. 429–445, March 1996.

[157] S. W. Halpern, "Reuse Partitioning in Cellular Systems," in *IEEE Veh. Technol Conf.*, Toronto, Ontario, Canada, pp. 322–327, 1983.

[158] K. Hamied, *Advanced Radio Link Design and Radio Receiver Design for Mobile Communications.* Ph. D. thesis, Georgia Institute of Technology, Atlanta, GA, 1994.

[159] K. Hamied and G. L. Stüber, "A fractionally-spaced MLSE Receiver," in *IEEE Int. Conf. on Commun.*, Seattle , WA, pp. 7–11, 1995.

[160] K. Hanabe, V. Tetsuro, and T. Otsu, "Distributed adaptive channel allocation scheme with variable C/I threshold in cellular systems," in *IEEE Veh. Technol. Conf.*, Secaucus, NJ, pp. 164–167, May 1993.

[161] P. Harley, "Short distance attenuation measurements at 900 MHz and 1.8 GHz using low antenna heights for microcells," *IEEE J. Selec. Areas Commun.,* Vol. 7, pp. 5–11, January 1989.

[162] M. Hata and T. Nagatsu, "Mobile location using signal strength measurements in cellular systems," *IEEE Trans. Veh. Technol.*, Vol. 29, pp.245–251, 1980.

[163] S. Haykin, ed., *Adaptive Filter Theory.* Englewood Cliff, NJ: Prentice-Hall, 1986.

[164] S. Haykin, ed., *Communication Systems, 3/e.* New York, NY: Wiley, 1994.

[165] M.-J. Ho and G. L. Stüber, "Co-channel interference of microcellular systems on shadowed Nakagami fading channels," in *IEEE Veh. Technol. Conf.*, Secaucus, NJ, pp. 568–571, May 1993.

[166] M.-J. Ho and G. L. Stüber, "Capacity and Power Control for CDMA Microcells," *ACM/Baltzer Journal on Wireless Networks*, Vol. 1, pp. 355–363, Oct 1995.

[167] P. Ho and D. Fung, "Error performance of interleaved trellis-coded PSK modulation in correlated Rayleigh fading channels," *IEEE Trans. Commun.*, Vol. 40, pp. 1800–1809, December 1992.

[168] J. M. Holtzman, "A simple, accurate method to calculate spread-spectrum multiple-access error probabilities," *IEEE Trans. Commun.*, Vol. 40, pp. 461–464, March 1992.

[169] J. Holtzman, "Adaptive measurement intervals for handoffs," in *IEEE Int. Conf. on Commun.*, Chicago, IL, pp. 1032–1036, June 1992.

[170] D. Hong and S. S. Rappaport, "Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures," *IEEE Trans. Veh. Technol.*, Vol. 35, pp. 77–92, August 1986.

[171] F. M. Hsu, A. A. Giordano, H. dePedro, and J. G. Proakis, "Adaptive equalization techniques for high speed transmission on fading dispersive HF channels," in *Nat. Telecommun. Conf.*, Houston, TX, pp. 58.1.1–58.1.7, November 1980.

[172] K. Imamura and A. Murase, "Mobile communication control using multi-transmitter simul/sequential casting (MSSC)," in *IEEE Veh. Technol. Conf.*, Dallas, TX, pp. 334–341, May 1986.

[173] W. C. Jakes, *Microwave Mobile Communication.* New York, NY: IEEE Press, 1993.

[174] H. Jiang and S. S. Rappaport, "CBWL: A new channel assignment and sharing method for cellular communication systems," *IEEE Trans. Veh. Technol.*, Vol. 43, pp. 313–322, May 1994.

[175] M. Kaji and A. Akeyama, "UHF-band propagation characteristics for land mobile radio," in *Int. Symp. Ant. and Prop.*, Univ. of British Columbia, Canada, pp. 835–838, June 1985.

[176] T. Kanai, "Autonomous reuse partitioning in cellular systems," in *IEEE Veh. Technol. Conf.*, Denver, CO, pp. 782–785, May 1992.

[177] T. Kasami, "Weight distribution of Bose-Chaudhuri-Hocquenghem codes," in *Combinatorial Mathematics and its Applications*, University of North Carolina Press, Chapel Hill, NC, pp. 335–357, 1967.

[178] T. Kasami, S. Lin, and W. Peterson, "Some results on cyclic codes which are invariant under the affine group and their applications," *Information and Control*, Vol. 11, pp. 475–496, November-December 1968.

[179] F. P. Kelly, *Reversibility and Stochastic Networks.* New York, NY: Wiley, 1979.

[180] F. P. Kelly, "Blocked probabilities in large circuit-switched networks," *Advances in Applied Probability*, Vol. 18, pp. 473–505, April-June 1986.

[181] R. W. Kerr and P. McLane, "Coherent detection of interleaved trellis encoded CPFSK on shadowed mobile radio channels," *IEEE Trans. Commun.*, Vol. 41, pp. 159–169, May 1992.

[182] S. Kozono, "Co-channel interference measurement method for mobile communication," *IEEE Trans. Veh. Technol.*, Vol. 36, pp. 7–13, January 1987.

[183] E. Kudoh and T. Matsumoto, "Effects of power control error on the system user capacity of DS/CDMA cellular mobile radios," *IEICE Transactions*, Vol. E75-B, pp. 524–529, June 1992.

[184] S. S. Kuek and W. C. Wong, "Ordered dynamic channel assignment scheme with reassignment in highway microcells," *IEEE Trans. Commun.*, Vol. 41, pp. 271–276, August 1992.

[185] G. Labedz, K. Felix, V. Lev, and D. Schaeffer, "Handover control issues in very high capacity cellular systems using small cells," in *Int. Conf. on Digital Land Mobile Radio Commun.*, Univ. of Warwick, Coventry, UK, 1987.

[186] H. J. Larson and B. Schubert, *Probabilistic Models in Engineering Sciences, Vol. I.* New York, NY: Wiley, 1979.

[187] B. Larsson, B. Gudmundson, and K. Raith, "Receiver performance for the North American digital cellular system," in *IEEE Veh. Technol. Conf.*, Saint Louis, MO, pp. 1–6, May 1991.

[188] W. C. Y. Lee, *Mobile Communications Engineering.* New York, NY: McGraw Hill, 1982.

[189] W. C. Y. Lee, "Estimate of local average power of a mobile radio signal," *IEEE Trans. Veh. Technol.*, Vol. 34, pp. 22–27, February 1985.

[190] W. C. Y. Lee, *Mobile Communications Design Fundamentals.* Indianapolis, IN: Sams, 1986.

[191] W. U. Lee and F. S. Hill, "A maximum-likelihood sequence estimator with decision-feedback equalization," *IEEE Trans. Commun.*, Vol. 25, pp. 971–979, June 1977.

[192] W. C. Y. Lee, "New cellular schemes for spectral efficiency," *IEEE Trans. Veh. Technol.*, Vol. 36, pp. 183–192, November 1987.

[193] W. C. Y. Lee, "Overview of cellular *CDMA "IEEE Trans. Veh. Technol.*, Vol. 40, pp. 291–302, May 1991.

[194] W. C. Y. Lee and Y. S. Yeh, "On the estimation of the second-order statistics of log-normal fading in mobile radio environment," *IEEE Trans. Commun.*, Vol. 22, pp. 809–873, June 1974.

[195] J. S. Lehnert and M. B. Pursley, "Error probability for binary direct-sequence spread-spectrum communications with random signature sequences," *IEEE Trans. Commun.*, Vol. 35, pp. 87–98, January 1987.

[196] A. Leon-Garcia, *Probability and Random Processes for Electrical Engineering.* Reading, MA: Addison-Wesley, 1989.

[197] J. Lin, F. Ling, and J. Proakis, "Joint data and channel estimation for TDMA mobile channels," in *IEEE Int. Symp. on Pers., Indoor and Mobile Radio Commun.*, Boston, MA, pp. 235–239, 1992.

[198] J. Lin, F. Ling, and J. Proakis, "Fading channel tracking properties of several adaptive algorithms for the North American digital cellular system," in *IEEE Veh. Technol. Conf.*, pp. 273–276, 1993.

[199] S. Lin and D. J. Costello, Jr., *Error Control Coding: Fundamentals and Applications.* Englewood Cliffs, NJ: Prentice-Hall, 1983.

[200] F. Ling and J. G. Proakis, "A generalized multichannel least-squares lattice algorithm based on sequential processing stages," *IEEE Trans. Acoutics, Speech and Signal Proc.*, Vol. 32, pp. 381–389, April 1984.

[201] F. Ling and J. G. Proakis, "Adaptive lattice decision-feedback equalizers – their performance and application to time-variant multipath channels," *IEEE Trans. Commun.*, Vol. 33, pp. 348–356, April 1985.

[202] J.-P. M. Linnartz, "Site diversity in land mobile cellular telephony network with discontinuous voice transmission," *Europ. Trans. Telecommun.*, Vol. 2, pp. 471–480, May 1991.

[203] J.-P. M. Linnartz, "Exact analysis of the outage probability in multiple-user mobile radio," *IEEE Trans. Commun.*, Vol. 40, pp. 20–23, January 1992.

[204] Y. Liu, I. Oka, and E. Biglieri, "Error probability for digital transmission over nonlinear channels with applications to TCM," *IEEE Trans. Inform. Theory*, Vol. 36, pp. 1101–1110, September 1990.

[205] E. M. Long, *Decision-aided sequential sequence estimation for inter-symbol interference channels.* Ph. D. thesis, Georgia Institute of Technology, 1989.

[206] E. M. Long and A. M. Bush, "Decision-aided sequential sequence estimation for intersymbol interference channels," in *IEEE Int. Conf. on Commun.*, Boston, MA, pp. 26.1.1–26.1.5, June 1989.

[207] F. Lotse and A. Wejke, "Propagation measurements for microcells in central Stockholm," in *IEEE Veh. Technol. Conf.*, Orlando, FL, pp. 539–541, May 1990.

[208] R. W. Lucky, "Automatic equalization for digital communication," *Bell System Tech. J.*, Vol. 44, pp. 547–588, April 1965.

[209] R. W. Lucky, "Techniques for adaptive equalization of digital communication systems," *Bell System Tech. J.*, Vol. 45, pp. 255–286, February 1966.

[210] R. Lucky, J. Salz, and E. Weldon, *Principles of Data Communication.* New York, NY: McGraw Hill, 1968.

[211] R. Lupas and S. Verdú, "Linear Multiuser Detectors for Synchronous Code-Division Multiple-Access Channels," *IEEE Trans. Veh. Technol.*, Vol. 35, pp. 123–136, January 1989.

[212] R. Lupas and S. Verdú, "Near-far Resistance of Multiuser Detectors in Asynchronous Channels," *IEEE Trans. Commun.*, Vol. 38, pp. 496–508, April 1990.

[213] V. H. MacDonald, "The cellular concept," *Bell System Tech. J.*, Vol. 58, pp. 15–49, January 1979.

[214] M. B. Madayam, P-C. Chen, and J. M. Holtzman, "Minimum Duration Outage for Cellular Systems: A Level Crossing Analysis," in *IEEE Veh. Technol. Conf.*, Atlanta, GA, pp. 879–883, June 1996.

[215] F. R. Magee and J. G. Proakis, "Adaptive maximum-likelihood sequence estimation for digital signaling in the presence of intersymbol interference," *IEEE Trans. Inform. Theory*, Vol. 19, pp. 120–124, January 1973.

[216] M. Marsan and G. Hess, "Shadow variability in an urban land mobile radio environment," *Electronics Letters*, Vol. 26, pp. 646–648, May 1990.

[217] M. Marsan and G. C. Hess, "Cochannel isolation characteristics in an urban land mobile environment at 900 MHz, in *IEEE Veh. Technol. Conf.*, Saint Louis, MO, pp. 600–605, 1991.

[218] S. J. Mason, "Feedback theory: further properties of signal flow graphs," *IRE*, Vol. 44, pp. 920–926, July 1956.

[219] J. L. Massey, "Coding and modulation in digital communications," in *Int. Zurich Sem. Digital Communications*, Zurich, Switzerland, pp. E2(l)–E2(4), March 1974.

[220] J. E. Mazo, "On the independence theory of equalizer convergence," *Bell System Tech. J.*, Vol. 58, pp. 963–993, May 1979.

[221] J. E. Mazo, "Analysis of decision directed equalizer convergence," *Bell System Tech. J.*, Vol. 59, pp. 1857–1876, December 1980.

[222] W. R. Mende, "Evaluation of a proposed handover algorithm for the GSM cellular system," in *IEEE Veh. Technol. Conf.*, Orlando, FL, pp. 264–269, May 1990.

[223] L. B. Milstein, D. L. Schilling, R. L. Pickholtz, V. Erceg, M. Kullback, E. G. Kanterakis, D. Fishman, W. H. Biederman, and D. C. Salerno, "On the feasibility of a CDMA overlay for personal communications networks," *IEEE J. Selec. Areas Commun.*, Vol. 10, pp. 655–668, May 1992.

[224] S. Mockford and A. M. D. Turkmani, "Penetration loss into buildings at 900 MHz," in *IEE Colloquium on Propagation Factors and Interference Modeling for Mobile Radio Systems*, London, UK, pp. 1/1–1/4, November 1988.

[225] S. Mockford, A. M. D. Turkmani, and J. D. Parsons, "Local mean signal variability in rural areas at 900 MHz," in *IEEE Veh. Technol. Conf.*, Orlando, FL, pp. 610–615, May 1990.

[226] P. E. Mogensen, P. Eggers, C. Jensen, and J. B. Andersen, "Urban area radio propagation measurements at 955 and 1845 MHz for small and micro cells," in *IEEE Global Commun. Conf.*, Phoenix, AZ, pp. 1297–1302, December 1991.

[227] P. E. Mogensen and S. Petersen, "Antenna configuration measurements for direct microcells," in *IEEE Int. Symp. on Pers., Indoor and Mobile Radio Commun.*, Den Hague, The Netherlands, pp. 1075–1080, September 1994.

[228] M. A. Mokhtar and S. C. Gupta, "Capacity for cellular CDMA PCS's in Nakagami fading log-normal shadowing channels," in *IEEE Conf. Universal Personal Commun.*, pp. 190–194, 1992.

[229] K. J. Molnar and G. E. Bottomley, "D-AMPS Performance in PCS Bands with Array Processing," in *IEEE Veh. Technol. Conf.*, Atlanta, GA, pp. 1496–1500, April 19%.

[230] P. Monsen, "Feedback equalization for fading dispersive channels," *IEEE Trans. Inform. Theory*, Vol. 17, pp. 56–64, January 1971.

[231] P. Monsen, "Theoretical and measured performance of DFE modem on a fading multipath channel," *IEEE Trans. Commun.*, Vol. 32, pp. 5–12, January 1984.

[232] M. Morf, A. Vieira, and D. T. Lee, "Ladder forms for identification and speech processing," in *IEEE Conf. Dec. and Contr.*, New Orleans, LA, pp. 1074–1078, December 1977.

[233] R. K. Morrow, Jr. and J. S. Lehnert, "Bit-to-bit error dependence in slotted DS/CDMA packet systems with random signature sequences," *IEEE Trans. Commun.*, Vol. 37, pp. 1052–1061, October 1989.

[234] R. Muammar, "Co-channel interference in microcellular mobile radio system," in *IEEE Veh. Technol. Conf.*, Saint Louis, MO, pp. 198–203, May 1991.

[235] R. Muammar and S. C. Gupta, "Cochannel interference in high-capacity mobile radio systems," *IEEE Trans. Commun.*, Vol. 30, pp. 1973–1978, August 1982.

[236] M. S. Mueller, "Least-squares algorithms for adaptive equalizers," *Bell System Tech. J.*, Vol. 60, pp. 1905–1925, October 1981.

[237] M. S. Mueller and J. Salz, "A unified theory of data-aided equalization," *Bell System Tech. J.*, Vol. 60, pp. 2023–2038, November 1981.

[238] A. Murase, I. C. Symington, and E. Green, "Handover criterion for macro and microcellular systems," in *IEEE Veh. Technol. Conf.*, Saint Louis, MO, pp. 524–530, May 1991.

[239] K. Murota and K. Hirade, "OMSK modulation for digital mobile radio telephony," *IEEE Trans. Commun.*, Vol. 29, pp. 1044–1050, July 1981.

[240] Y. Nagata, "Analysis for spectrum efficiency in single cell trunked and cellular mobile radio," *IEEE Trans. Veh. Technol.*, Vol. 35, pp. 100–113, August 1987.

[241] A. F. Naguib, V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time Coded Modulation for High Data Rate Wireless Communications," in *IEEE Global Commun. Conf.*, Phoenix, AZ, pp. 102–109, December 1997.

[242] A. F. Naguib, V. Tarokh, N. Seshadri, and A. R. Calderbank, "A Space-Time Coding Modem for High-Data-Rate Wireless Communications," *IEEE J. Selec. Areas Commun.*, Vol. 16, pp. 1459–1478, October 1998.

[243] M. Nakagami, "The *m* distribution; a general formula of intensity distribution of rapid fading," *Statistical Methods in Radio Wave Propagation, W.G. Hoffman, ed.*, pp. 3–36, 1960.

[244] S. Nanda and D. J. Goodman, "Dynamic Resource Acquisition: distributed carrier allocation for TDMA cellular systems," *Third Generation Wireless Information Networks*, pp. 99–124, 1992.

[245] S. Nanda, "Teletraffic models for urban and suburban microcells: cell sizes and handoff rates," *IEEE Trans. Veh. Technol.*, Vol. 42, pp. 673–682, November 1993.

[246] R. W. Nettleton and G. R. Schloemer, "A high capacity assignment method for cellular mobile telephone systems," in *IEEE Veh. Technol. Conf.*, San Francisco, CA, pp. 359–367, May 1989.

[247] P. Newson and M. R. Heath, "The capacity of a spread spectrum CDMA system for cellular mobile radio with consideration of system imperfections," *IEEE J. Selec. Areas Commun.*, Vol. 12, pp. 673–683, 1994.

[248] H. Nyquist, "Certain topics in telegraph transmission theory," *Trans. American Inst. of Elect. Eng.*, Vol. 47, pp. 617–644, March 1928.

[249] K. Okada, "A dynamic channel assignment strategy using information of speed and moving direction in microcellular systems," in *Int. Symp. Circuits and Systems*, Chicago, IL, pp. 2212–2215, May 1993.

[250] K. Okada and F. Kubota, "On dynamic channel assignment in cellular mobile radio systems," in *Int. Symp. Circuits and Systems*, Singapore, pp. 938–941, June 1991.

[251] K. Okada and F. Kubota, "A proposal of a dynamic channel assignment strategy with information of moving direction in microcellular systems," *Trans. IEICE*, Vol. E75-A, pp. 1667–1673, December 1992.

[252] K. Okanoue, A. Ushirokawa, H. Tomita, and Y. Furuya, "New MLSE receiver free from sample timing and input level controls," in *IEEE Veh. Technol. Conf.*, Secaucus, NJ, pp. 408–411, June 1993.

[253] Y. Okumura, E. Ohmuri, T. Kawano, and K. Fukuda, "Field strength and its variability in VHP and UHF land mobile radio service," *Rev. of the ECL*, Vol. 16, pp. 825–873, 1968.

[254] W. P. Osborne and M. B. Luntz, "Coherent and noncoherent detection of CPFSK," *IEEE Trans. Commun.*, Vol. 22, pp. 1023–1036, August 1974.

[255] H. Panzer and R. Beck, "Adaptive resource allocation in metropolitan area cellular mobile radio systems," in *IEEE Veh. Technol. Conf.*, Orlando, FL, pp. 638–645, May 1990.

[256] Papoulis, *Probability, Random Variables, and Stochastic Processes.* New York, NY: McGraw-Hill, 1984.

[257] J. D. Parsons, *The Mobile Radio Propagation Channel.* New York, NY: Wiley, 1992.

[258] J. D. Parsons and J. G. Gardiner, *Mobile Communication Systems.* New York, NY: Halsted Press, 1989.

[259] S. Pasupathy, "Nyquist's third criterion," *Proc. IEEE,* Vol. 62, pp. 860–861, June 1974.

[260] J. M. Perl, A. Shpigel, and A. Reichman, "Adaptive receiver for digital communication over HF channels," *IEEE J. Selec. Areas Commun.*, Vol. 5, pp. 304–308, February 1987.

[261] W. W. Peterson and E. J. Weldon, *Error Correcting Codes, 2/e.* Cambridge, MA: MIT Press, 1972.

[262] R. L. Pickholtz, D. L. Schilling, and L. B. Milstein, "Theory of spread-spectrum communications - a tutorial," *IEEE Trans. Commun.*, Vol. 30, pp. 855 – 884, May 1982.

[263] M. F. Pop and N. C. Beaulieu, "Limitations of Sum-of-Sinusoids Fading Channel Simulators," Submitted to IEEE J. Selec Areas in Commun.

[264] R. Prasad and J. C. Arnbak, "Comments on "analysis for spectrum efficiency in single cell trunked and cellular mobile radio"," *IEEE Trans. Veh. Technol.*, Vol. 37, pp. 220–222, November 1988.

[265] R. Prasad and A. Kegel, "Effects of Rician faded and log-normal shadowed signals on spectrum efficiency in microcellular radio," *IEEE Trans. Veh. Technol.*, Vol. 42, pp. 274–281, August 1993.

[266] R. Prasad, A. Kegel, and J. C. Arnbak, "Analysis of system performance of high-capacity mobile radio," in *IEEE Veh. Technol. Conf.*, San Francisco, CA, pp. 306–309, May 1989.

[267] R. Prasad and A. Kegel, "Improved assessment of interference limits in cellular radio performance," *IEEE Trans. Veh. Technol.*, Vol. 40, pp. 412–419, May 1991.

[268] R. Prasad and A. Kegel, "Spectrum efficiency of microcellular systems," *Electronic Letters*, Vol. 27, pp. 423–425, February 1991.

[269] R. Price and P. E. Green, "A communication technique for multipath channels," *Proc. IEEE*, Vol. 46, pp. 555–570, March 1958.

[270] J. G. Proakis, *Digital Communications, 3rd ed.* New York, NY: McGraw-Hill, 1995.

[271] J. G. Proakis and J. Miller, "An adaptive receiver for digital signaling through channels with intersymbol interference," *IEEE Trans. Inform. Theory*, Vol. 15, pp. 484–497, July 1969.

[272] M. B. Pursley, F. D. Garber, and J. S. Lehnert, "Analysis of generalized quadriphase spread-spectrum communications," in *IEEE Int. Conf. on Commun.*, Seattle, WA, pp. 15.3.1–15.3.6, June 1980.

[273] D. Qong and T. J. Lim, "Soft Handoff in CDMA Mobile Systems," *IEEE Personal Commun. Mag.*, pp. 6 – 17, December 1997.

[274] S. U. Qureshi, "Adaptive equalization," *Proc. IEEE*, Vol. 73, pp. 1349–1387, September 1985.

[275] S. U. Qureshi and E. E. Newhall, "Analysis of maximum likelihood sequence estimation performance for quadrature amplitude modulation," *IEEE Trans. Inform. Theory*, Vol. 19, pp. 448–457, July 1973.

[276] R. Raheli, A. Polydoros, and C.-K. Tzou, "The principle of per-survivor processing: a general approach to approximate and adaptive MLSE," in *IEEE Global Commun. Conf.*, pp. 33.3.1–33.3.6, 1991.

[277] P. A. Ranta, Z.-C. Honkasalo, and J. Tapaninen, "TDMA Cellular Network Application of an Interference Cancellation Technique," in *IEEE Veh. Technol. Conf.*, Chicago, IL, pp. 296–300, July 1995.

[278] T. Rappaport and L. Milstein, "Effects of path loss and fringe user distribution on CDMA cellular frequency reuse efficiency," in *IEEE Global Commun. Conf.*, San Diego, CA, pp. 500–506, December 1990.

[279] P. Raymond, "Performance analysis of cellular networks," *IEEE Trans. Commun.*, Vol. 39, pp. 1787–1793, December 1991.

[280] Research & Development Center for Radio Communications (RCR), "Digital cellular telecommunication systems," April 1991. RCR STD-27.

[281] D. O. Reudink, D. Jones, S. Meredith, and M. Reudink, "Narrow beam switched antenna experiment," First Workshop on Smart Antennas in Wireless Mobile Communications, 1994.

[282] S. Rice, "Statistical properties of a sine wave plus noise," *Bell System Tech. J.*, Vol. 27, pp. 109–157, January 1948.

[283] P. Robertson, "Illuminating the Structure of Code and Decoder of Parallel Concatenated Recursive Systematic (Turbo) Codes," in *IEEE Global Commun. Conf.*, San Fransisco, pp. 1298–1303, 1994.

[284] M. Rouanne and D. Costello, "An algorithm for computing the distance spectrum of trellis codes," *IEEE J. Selec. Areas Commun.*, Vol. 7, pp. 929–940, August 1989.

[285] A. Russell (aka E. Katz) and G. Stüber, "Sequential sequence estimation for trellis-coded modulation on multipath fading ISI channels," *IEEE Trans. Commun.*, pp. 2882–2885, December 1995.

[286] A. Rustako, N. Amitay, G. Owens, and R. Roman, "Radio propagation at microwave frequencies for line-of-sight microcellular mobile and personal communications," *IEEE Trans. Veh. Technol.*, Vol. 40, pp. 203–210, February 1991.

[287] D. Divsalar, S. Benedetto, G. Montorsi and F. Pollara, "Serial Concatenation of Interleaved Codes: Performance Analysis, Design and Iterative Decoding," 1996.

[288] D. Divsalar S. Benedetto, G. Montorsi and F. Pollara, "Soft-Output Decoding Algorithms in Iterative Decoding of Turbo Codes," 1996.

[289] J. Salz, "Optimum mean-square decision-feedback equalization," *Bell System Tech. J.*, Vol. 52, pp. 1341–1373, October 1973.

[290] A. Sampath and J. Holtzman, "Estimation of maximum Doppler frequency for handoff decisions," in *IEEE Veh. Technol. Conf.*, Secaucus, NJ, pp. 859–862, May 1993.

[291] E. H. Satorius and S. T. Alexander, "Channel equalization using adaptive lattice algorithms," *IEEE Trans. Commun.*, Vol. 27, pp. 899–905, June 1979.

[292] E. H. Satorius and J. D. Pack, "Application of least squares lattice algorithms to adaptive equalization," *IEEE Trans. Commun.*, Vol. 29, pp. 136–142, February 1981.

[293] D. Schleher, "Generalized Gram-Charlier series with application to the sum of lognormal variates," *IEEE Trans. Inform. Theory*, Vol. 23, pp. 275–280, March 1977.

[294] T. A. Schonhoff, "Symbol error probabilities for M-ary CPFSK: Coherent and noncoherent detection," *IEEE Trans. Commun.*, Vol. 24, pp. 644–652, June 1976.

[295] S. Schwartz and Y. S. Yeh, "On the distribution function and moments of power sums with log-normal components," *Bell System Tech. J.*, Vol. 61, pp. 1441–1462, September 1982.

[296] M. Serizawa and J. Murakami, "Phase tracking Viterbi demodulator," *Electronics Letters*, Vol. 40, pp. 792–794, 1989.

[297] N. Seshadri, "Joint data and channel estimation using fast blind trellis search techniques," in *IEEE Global Commun. Conf.*, San Diego, CA, pp. 1659–1663, 1990.

[298] N. Seshadri, V. Tarokh, and A. R. Calderbank, "Space-time codes for wireless communication: code construction," in *IEEE Veh. Technol. Conf.*, Phoenix, AZ, pp. 637–641, May 1997.

[299] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Tech. J.*, Vol. 27, pp. 379–423 and 623–656, 1948.

[300] W. H. Sheen and G. L. Stüber, "MLSE equalization and decoding for multipath-fading channels," *IEEE Trans. Commun.*, Vol. 39, pp. 1455–1464, October 1991.

[301] W. H. Sheen and G. L. Stüber, "Error probability for reduced-state sequence estimation," *IEEE J. Selec. Areas Commun.*, Vol. 10, pp. 571–578, April 1992.

[302] W. H. Sheen and G. L. Stüber, "Error probability of reduced-state sequence estimation for trellis-coded modulation on intersymbol interference channels," *IEEE Trans. Commun.*, *Vol.* 41, pp. 1265–1269, September 1993.

[303] W. H. Sheen and G. L. Stüber, "Error probability for maximum likelihood sequence estimation of trellis-coded modulation on ISI channels," *IEEE Trans. Commun.*, Vol. 42, pp. 1427–1430, April 1994.

[304] H. Shiino, N. Yamaguchi, , and Y. Shoji, "Performance of an adaptive maximum-likelihood receiver for fast fading multipath channel," in *IEEE Veh. Technol. Conf.*, Denver, CO, pp. 380–383, May 1992.

[305] M. K. Simon and D. Divsalar, "Maximum-likelihood block detection of moncoherent continuous phase modulation," *IEEE Trans. Commun.*, Vol. 41, pp. 90–98, January 1993.

[306] M. K. Simon, J. K. Omura, R. A. Scholtz, and B. K. Levitt, *Spread Spectrum Communications.* Rockville, MD: Computer Science Press, 1985.

[307] N. R. Sollenberger, "Architecture and implementation of an efficient and robust TDMA frame structure for digital portable communications," *IEEE Trans. Veh. Technol.*, Vol. 40, pp. 250–260, January 1991.

[308] F. G. Stremler, ed., *Introduction to Communication Systems, 3/e.* Reading, MA: Addison-Wesley, 1990.

[309] G. Stüber, "Soft decision direct-sequence DPSK receivers," *IEEE Trans. Veh. Technol.*, Vol. 37, pp. 151–157, August 1988.

[310] G. L. Stüber and C. Kchao, *Spread Spectrum Cellular Radio.* Georgia Institute of Technology, OCA Project No. E21-662 for Bell South Enterprises, December, 1990.

[311] H. Susuki, "A statistical model for urban radio propagation," *IEEE Trans. Commun.*, Vol. 25, pp. 673–680, July 1977.

[312] S. C. Swales, M. A. Beach, and D. J. Edwards, "Multi-beam adaptive base station antennas for cellular land mobile radio systems," in *IEEE Veh. Technol. Conf.*, San Francisco, CA, pp. 341–348, 1989.

[313] S. C. Swales, M. A. Beach, D. J. Edwards, and J. P. McGeehan, 'The performance enhancement of multibeam adaptive base station antennas for cellular land mobile radio systems," *IEEE Trans. Veh. Technol.*, Vol. 39, pp. 56–67, February 1990.

[314] E. W. Swokowski, *Calculus with Analytical Geometry.* New York, NY: Prindle, Weber, and Schmidt, 1979.

[315] S.Yoshida, A. Hirai, G. L. Tan, H. Zhou, and T. Takeuchi, "In-Service monitoring of multipath delay-spread and C/I for QPSK signal," in *IEEE Veh. Technol. Conf.*, Denver, CO, pp. 592–595, May 1992.

[316] J. Tajima and K. Imamura, "A strategy for flexible channel assignment in mobile communication systems," *IEEE Trans. Veh. Technol.,* Vol. 37, pp. 92–103,May1988.

[317] K. Takeo, M. Nishino, Y. Ameazwa, and S. Sato, "Adaptive traffic control scheme for non-uniform traffic distribution in microcellular mobile communication system," in *IEEE Veh. Technol. Conf.*, Orlando, FL, pp. 527–531, May 1990.

[318] S. Tekinay and B. Jabbari, "Handover and channel assignment in mobile cellular networks," *IEEE Commun. Mag.*, Vol. 29, pp. 42–46, 1991.

[319] H. Thaper, "Real-time application of trellis coding to high-speed voice-band data transmission," *IEEE J. Selec. Areas Commun.*, *Vol.* 2, pp. 648–658, September 1984.

[320] C. Tidestav, A. Ahlén, and M. Sternad, "Narrowband and Broadband Multiuser Detection Using a Multivariable DFE," in *IEEE Int. Symp. on Pers., Indoor and Mobile Radio Commun.*, Toronto, Canada, pp. 732–736, September 1995.

[321] R. J. Tront, J. K. Cavers, and M. R. Ito, "Performance of Kalman decision-feedback equalization in HF radio modems," in *IEEE Int. Conf. on Commun.*, Toronto, Canada, pp. 1617–1621, June 1986.

[322] A. M. D. Turkmani, "performance evaluation of a composite microscopic plus macroscopic diversity system," *IEE Proc. I.*, Vol. 138, pp. 15–20, January 1991.

[323] A. M. D. Turkmani, "Probability of error for M-branch selection diversity," *IEE Proc. I.*, Vol. 139, pp. 71–78, February 1992.

[324] A. M. D. Turkmani, J. D. Parsons, F. Ju, and D. G. Lewis, "Microcellular radio measurements at 900,1500, and 1800 MHz," in *5th Int. Conf. on Mobile Radio and Personal Communications*, Coventry, UK, pp. 65–68, December 1989.

[325] W. H. Tuttlebee, "Cordless personal communications," *IEEE Commun. Mag.*, Vol. 30, pp. 42–53, December 1992.

[326] M. Uesugi, S. Futagi, and K. Homma, "Interference Cancellation Method using DFE," in *IEEE Veh. Technol. Conf.*, Atlanta, GA, pp. 1190–1194, April 1996.

[327] G. Ungerboeck, "Theory on the speed of convergence in adaptive equalizers for digital communication," *IBM Journal Research and Development*, Vol. 16, pp. 546–555, November 1972.

[328] *G.* Ungerboeck, "Adaptive maximum likelihood receiver for carrier-modulated data transmission systems," *IEEE Trans. Commun.*, Vol. 22, pp. 624–636, May 1974.

[329] G. Ungerboeck, "Fractional tap-spacing equalizer and consequences for clock recovery in data modems," *IEEE Trans. Commun.*, Vol. 24, pp. 856–864, August 1976.

[330] G. Ungerboeck, "Channel coding with multilevel phase signals," *IEEE Trans. Inform. Theory*, Vol. 28, pp. 55–67, January 1982.

[331] G. Ungerboeck, "Trellis coded modulation with redundant signal sets - part I: introduction," *IEEE Commun. Mag.*, Vol. 25, pp. 5 – 11, February 1987.

[332] G. M. Vachula and J. F. S. Hill, "On optimal detection of band-limited PAM signals with excess bandwidth," *IEEE Trans. Commun.*, Vol. 29, pp. 886–890, June 1981.

[333] R. van Nee, G. Awater, M. Morikura, H. Takanashi, M. Webster, and K. W. Halford, "New High-Rate Wireless LAN Standards," *IEEE Commun. Mag.*, pp. 82 – 88, December 1999.

[334] M. Varanasi and B. Aazhang, "Multistage Detection in Asynchronous Code-Division Multiple-Access Communications," *IEEE Trans. Commun.*, Vol. 38, pp. 509–519, April 1990.

[335] M. Varanasi and B. Aazhang, "Near-Optimum Detection in Synchronous Code-Division Multiple-Access Systems," *IEEE Trans. Commun.*, Vol. 39, pp. 725–736, May 1991.

[336] S. Verdii, "Miminum probability of error for asynchronous Gaussian multiple-access channels," *IEEE Trans. Inform. Theory*, Vol. 32, pp. 85–96, January 1986.

[337] F. L. Vermuelen and M. E. Hellman, "Reduced-state Viterbi decoding for channels with intersymbol interference," in *IEEE Int. Conf. on Commun.*, Minneapolis, MN, pp. 37B.1-37B.4, June 1974.

[338] R. Vijayan and J. Holtzman, "Sensitivity of handoff algorithms to variations in the propagation environment," in *IEEE Conf. Universal Personal Commun.*, Ottawa, Canada, pp. 158–162, October 1993.

[339] R. Vijayan and J. M. Holtzman, "A model for analyzing handoff algorithms," *IEEE Trans. Veh. Technol.*, Vol. 42, pp. 351–356, August 1993.

[340] R. Vijayan and J. Holtzman, "Analysis of handoff algorithms using nonstationary signal strength measurements," in *IEEE Global Commun. Conf.*, Orlando, FL, pp. 1405–1409, December 1992.

[341] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inform. Theory*, Vol. 13, pp. 260–269, April 1967.

[342] A. J. Viterbi, "Convolutional codes and their performance in communication systems," *IEEE Trans. Commun.*, Vol. 19, pp. 751–772, October 1971.

[343] A. J. Viterbi, A. M. Viterbi, K. Gilhousen, and E. Zehavi, "Soft Handoff Extends CDMA Cell Coverage and Increases Reverse Channel Capacity," *IEEE J. Selec. Areas Commun.*, Vol. 12, pp. 1281–1288, October 1994.

[344] A. J. Viterbi, *CDMA Principles of Spread Spectrum Communication.* Reading, MA: Addison-Wesley, 1995.

[345] J.-F. Wagen, "Signal strength measurements at 881 MHz for urban microcells in downtown Tampa," in *IEEE Global Commun. Conf.*, Phoenix, AZ, pp. 1313–1317, December 1991.

[346] E. H. Walker, "Penetration of radio signals into buildings in cellular radio environments," *Bell System Tech. J.*, Vol. 62, September 1983.

[347] L.-C. Wang and C.-T. Lea, "Macrodiversity co-channel interference analysis," *Electronics Letters*, Vol. 31, pp. 614–616, 1995.

[348] L.-C. Wang and C.-T. Lea, "Performance gain of a S-macrodiversity in a lognormal shadowed Rayleigh fading channel," *Electronics Letters*, Vol. 31, pp. 1785–1787, September 1995.

[349] L.-C. Wang and C.-T. Lea, "Incoherent estimation on cochannel interference probability for microcellular systems," *IEEE Trans. Veh. Technol.*, Vol. 45, pp. 164–173, February 1996.

[350] L.-C. Wang, G. L. Stüber, and C.-T. Lea, "Architecture Design, Frequency Planning, and Performance Analysis for a Microcell/Macrocell Overlaying System," *IEEE Trans. Veh. Technol.*, Vol. 46, pp. 836–848, 1997.

[351] L.-C. Wang, G. L. Stüber, and C.-T. Lea, "Effects of Rician Fading and Branch Correlation on a Local-Mean-Based Macrodiversity Cellular System," *IEEE Trans. Veh. Technol.*, Vol. 2, pp. 429–436, March 1999.

[352] L. F. Wei, "Trellis-coded modulation with multidimensional constellations," *IEEE Trans. Inform. Theory*, Vol. 33, pp. 483–501, July 1987.

[353] L. F. Wei, "Coded M-DPSK with built-in time diversity for fading channels," *IEEE Trans. Inform. Theory*, Vol. 39, pp. 1820–1839, November 1993.

[354] K. Wesolowski, "On the performance and convergence of the adaptive canceler of intersymbol interference in data transmission," *IEEE Trans. Commun.*, Vol. 33, pp. 425–432, May 1985.

[355] K. Wesolowski, "An efficient DFE & ML suboptimum receiver for data transmission over dispersive channels using two-dimensional signal constellations," *IEEE Trans. Commun.*, Vol. 35, pp. 336–339, March 1987.

[356] K. A. West and G. L. Stüber, "An aggressive dynamic channel assignment strategy for a microcellular environment," *IEEE Trans. Veh. Technol.*, Vol. 43, pp. 1027–1038, November 1994.

[357] B. Widrow, J. M. McCool, M. G. Larimore, and J. C. R. Johnson, "Adaptive switching circuits," in *IRE Wescon Conv.*, Los Angeles, CA, pp. 96–104, August 1960.

[358] B. Widrow, J. M. McCool, M. G. Larimore, and C. R. J. Jr., "Stationary and nonstationary learning characteristics of the LMS adaptive filter," in *Proc. IEEE*, pp. 1151–1162, August 1976.

[359] C. Wijffels, H. Miser, and R. Prasad, "A micro-cellular CDMA system over slow and fast Ricean fading channels with forward error correction," *IEEE Trans. Veh. Technol.*, Vol. 42, pp. 570–580, November 1993.

[360] A. Williamson, B. Egan, and J. Chester, "Mobile radio propagation in Auckland at 851 MHz," *Electronic Letters*, Vol. 20, pp. 517–518, June 1984.

[361] J. J. Winters, "Optimum combining in digital mobile radio with cochannel interference," *IEEE Trans. Veh. Technol.*, Vol. 33, pp. 144–155, August 1984.

[362] J. J. Winters, "Signal Acquisition and Tracking with Adaptive Arrays in the Digital Mobile Radio System IS-54 with Flat Fading," *IEEE Trans. Veh. Technol.*, Vol. 42, pp. 377–384, November 1993.

[363] L. Wong and P. McLane, "Performance of trellis codes for a class of equalized ISI channels," *IEEE Trans. Commun.*, Vol. 36, pp. 1330–1336, December 1988.

[364] A. Wonjar, "Unknown bounds on performance in Nakagami channels," *IEEE Trans. Commun.*, Vol. 34, pp. 22–24, January 1986.

[365] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*. Prospect Heights, IL: Waveland Press, 1990.

[366] H. Xia, H. Bertoni, L. Maciel, and A. Landsay-Stewart, "Radio propagation measurements and modeling for line-of-sight microcellular systems," in *IEEE Veh. Technol Conf.*, Denver, CO, pp. 349–354, May 1992.

[367] Z. Xie, R. Short, and C. K. Rushforth, "A family of suboptimum detectors for coherent multiuser communications," *IEEE J. Selec. Areas Commun.*, Vol. 8, pp. 683–690, May 1990.

[368] F. Xiong, A. Zerik, and E. Shwedyk, "Sequential sequence estimation for channels with intersymbol interference of finite or infinite length," *IEEE Trans. Commun.*, Vol. 36, pp. 795–803, June 1990.

[369] Y. Yao and A. Sheikh, "Outage probability analysis for microcell mobile radio systems with cochannel interferers in Rician/Rayleigh fading environment," *Electronics Letters*, Vol. 26, pp. 864–866, June 1990.

[370] Y. Yao and U. Sheikh, "Investigation into cochannel interference in microcellular mobile radio systems," *IEEE Trans. Veh. Technol.*, Vol. 41, pp. 114–123, May 1992.

[371] Y. H. Yeh, J. W. Wilson, and S. C. Schwartz, "Outage probability in mobile telephone with directive antennas and macrodiversity," *IEEE J. Selec. Areas Commun.*, Vol. 2, pp. 507–511, July 1984.

[372] Y. Yeh and S. C. Schwartz, "Outage probability in mobile telephony due to multiple log-normal interferers," *IEEE Trans. Commun.,* Vol. 32, pp. 380–388, April 1984.

[373] K. L. Yeung and T.-S. P. Yum, "Compact pattern based channel assignment for cellular mobile systems," *IEEE Trans. Veh. Technol.*, Vol. 43, pp. 892–896, November 1994.

[374] J. Yokota, J. Horikoshi, and H. Suzuki, "MLSE Multiple Co-channel Interference Canceling with Channel Estimation for Future Mobile Radios," in *IEEE Int. Symp. on Pers., Indoor and Mobile Radio Commun.*, Toronto, Canada, pp. 753–757, September 1995.

[375] H. Yoshino, K. Fukawa, and H. Suzuki, "Interference Canceling Equalizer(ICE) for Mobile Radio Communication," in *IEEE Int. Conf. on Commun.*, New Orleans, LA, pp. 1427–1432, June 1994.

[376] W. P. Yung, "Probability of Bit Error for MPSK modulation with diversity reception in Rayleigh fading and lognormal shadowing channel," *IEEE Trans. Commun.*, Vol. 38, pp. 933–937, July 1990.

[377] L. A. Zadeh, "Frequency analysis of variable networks," *Institute Radio Engineers*, Vol. 38, pp. 291–299, 1950.

[378] M. Zhang and T. Yum, "Comparisons of channel-assignment strategies in cellular mobile telephone systems," *IEEE Trans. Veh. Technol.*, Vol. 38, pp. 211–215, November 1989.

[379] N. Zhang and J. Holtzman, "Analysis of handoff algorithms using both absolute and relative measurements," in *IEEE Veh. Technol. Conf.*, Stockholm, Sweden, pp. 82–86, June 1994.

[380] N. Zhang and J. Holtzman, "Analysis of a CDMA Soft Handoff Algorithm," in *IEEE Int. Symp. on Pers., Indoor and Mobile Radio Commun.*, Toronto, Canada, pp. 819–823, September 1995.

[381] R. Ziemer and R. Peterson, *Digital Communications and Spread Spectrum Systems.* New York, NY: MacMillan, 1985.

[382] M. Zorzi, "Simplified forward-link power control law in cellular CDMA," *IEEE Trans. Veh. Technol.*, Vol. 43, pp. 1088–1093, November 1994.

Index